Peter Deuflhard Jan Hermans
Benedict Leimkuhler Alan E. Mark
Sebastian Reich Robert D. Skeel (Eds.)

Computational Molecular Dynamics: Challenges, Methods, Ideas

Proceedings of the
2nd International Symposium on Algorithms
for Macromolecular Modelling,
Berlin, May 21–24, 1997

With 117 Figures and 22 Tables

Springer

Editors

Peter Deuflhard
Konrad-Zuse-Zentrum Berlin (ZIB)
Takustrasse 7
D-14195 Berlin-Dahlem, Germany
deuflhard@zib.de

Jan Hermans
Department of Biochemistry
and Biophysics
University of North Carolina
Chapel Hill, NC 27599-7260, USA
hermans@femto.med.unc.edu

Benedict Leimkuhler
Department of Mathematics
University of Kansas
405 Snow Hall
Lawrence, KS 66045, USA
leimkuhl@math.ukans.edu

Alan E. Mark
Laboratorium für Physikalische Chemie
ETH Zentrum
CH-8092 Zürich, Switzerland
mark@igc.phys.chem.ethz.ch

Sebastian Reich
Department of Mathematics
and Statistics
University of Surrey
Guildford, Surrey GU2 5XH,
United Kingdom
s.reich@surrey.ac.uk

Robert D. Skeel
Department of Computer Science
University of Illinois
1304 West Springfield Avenue
Urbana, IL 61801-6631, USA
skeel@cs.uiuc.edu

Front cover figure created by Thomas Steinke and Olaf Paetsch,
Konrad-Zuse-Zentrum (ZIB), Berlin-Dahlem, Germany

Mathematics Subject Classification (1991): 34C35, 65-06, 68-06, 70-06, 81-06, 82-06, 92-06

Cataloging-in-Publication Data applied for

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

Computational molecular dynamics: challenges, methods, ideas; proceedings of the 2nd International Symposium on Algorithms for Macromolecular Modelling, Berlin, May 21–24, 1997; with 22 tables / Peter Deuflhard... (ed.). – Berlin; Heidelberg; New York; Barcelona; Hong Kong; London; Milan; Paris; Singapore; Tokyo: Springer, 1999
(Lecture notes in computational science and engineering; 4)
ISBN 3-540-63242-5

ISBN 3-540-63242-5 Springer-Verlag Berlin Heidelberg New York

Cover Design: Friedhelm Steinen-Broo, Estudio Calamar, Spain
Cover production: *design & production* GmbH, Heidelberg
Typesetting: Camera-ready by Sebastian Reich

SPIN 10568830 46/3143 – 5 4 3 2 1 0 – Printed on acid-free paper

Preface

In May 21 - 24, 1997 the Second International Symposium on Algorithms for Macromolecular Modelling was held in the new building of the Konrad Zuse Zentrum on the attractive Science Campus of the Free University of Berlin. Organizers of the symposium were the editors of this book, plus Bernie Brooks and Wilfred van Gunsteren. The event brought together computational scientists in fields like biochemistry, biophysics, physical chemistry, or statistical physics and numerical analysts as well as computer scientists working on the advancement of algorithms, for a total of over 120 participants from 19 countries. In the course of the symposium, it was agreed not to write traditional proceedings, but rather to produce a representative volume that combines survey articles and original papers (all refereed) that would give an account of the current state of the art of Molecular Dynamics (MD).

At present, the main challenge of computational molecular dynamics stems from the huge discrepancy of timescales: phenomena of interest, such as protein folding or active site docking, occur on a micro- or millisecond time scale while we are routinely able to do computations on a scale of only one or a few nanoseconds. In order to bridge this gap, a drastic speedup of our algorithms and software appears necessary – besides any speedup originating from advances in computer technology. However, this will not be enough to achieve our goal. In addition, there is the need to explore further the potential for improved physical modelling and to develop both new theoretical concepts and new algorithmic ideas. That is why this volume deliberately allocates considerable space to new concepts and ideas from physics and mathematics.

With the main challenge and the general intentions of the editors in mind, the volume begins with an *Introductory Survey* by a longtime leader in the field, HERMAN BERENDSEN, drawing on his long experience and deep insight into the current status of molecular simulations and their future as an increasingly important method in structural biology and chemistry. With his unique personal insight, this article will be the beginning of many discussions as the book as a whole will serve as a forum for alternative views and further perspectives.

The remaining 28 articles have been grouped in five chapters that reflect the main topics of the Berlin meeting. As in any interdisciplinary volume, there is a degree of arbitrariness in the allocation of some of the articles.

The first chapter, on *Conformational Dynamics*, includes discussion of several rather recent computational approaches to treat the dominant slow modes of molecular dynamical systems. In the first paper, SCHULTEN and his group review the new field of "steered molecular dynamics" (SMD), in which "large" external forces are applied in order to be able to study unbinding of ligands and conformation changes on time scales accessible to MD

simulations. The second paper, by HELMS & MCCAMMON, surveys a wide range of different computational techniques for the exploration of conformational transitions of proteins, including the use of stochastic dynamics with the Poisson-Boltzmann approximation as a simple solvent model. The article by EICHINGER ET AL. combines several speedup techniques: multiple time stepping algorithms adapted to fit fast multipole methods (see also the last chapter of this book), the previously mentioned SMD technique, and GRUBMÜLLER's method of "computational flooding", which uses local potential modifications in order to successively drive the system to different low-energy basins. The novel approach taken by DEUFLHARD ET AL. employs ideas from the mathematics of dynamical systems to construct certain almost invariant sets in phase space, which can be interpreted as chemical conformations; their algorithm also supplies patterns and rates of conformational changes. In the last paper of this chapter, TOLSTUROKOV & VIRNIK describe another use of dynamical systems tools and propose a simplified set of differential equations for the description of an observed hysteresis behavior in water adsorption-desorption of nucleic acids.

The second chapter, on *Thermodynamic Modelling*, is devoted largely to methods for computing free energies and potentials of mean force. The paper by HERMANS ET AL. reviews experimental and theoretical techniques for studying the stability of protein-ligand complexes, including a new method for computing absolute free energies of binding with MD simulations, and summarizes recent applications from their laboratory. MARK ET AL. describe a new method to estimate relative binding free energies of a series of related ligands on the basis of a single simulated trajectory of a reference state in which a specially constructed, artificial ligand is modelled with a special "soft" potential function. KUCZERA describes a multiple-dimension approach by which conformation space is explored, while the potential of mean force is simultaneously computed. The joint paper from the groups of LESYNG and of MCCAMMON reviews an algorithm for the prediction of ionization constants in proteins; calculations of the relevant protein-solvent system are based on the already mentioned Poisson-Boltzmann equation. The paper by STRAUB & ANDRICIOAEI employs the Tsallis statistics to speed up phase space sampling. In the final article of this chapter, NEUMAIER ET AL. construct empirical potentials for possible use in off-lattice protein studies.

The third chapter, on *Enhanced Time-Stepping Algorithms*, opens with a personal account on long-timestep integration by SCHLICK. She assesses both the successes and the limitations of various algorithmic approaches including implicit discretization, harmonic/anharmonic separation of modes, and force splitting techniques combined with Langevin dynamics. The second paper, by ELBER ET AL., describes a large step-size approximation of stochastic path integrals arising from Langevin dynamics – requiring, however, knowledge about both initial and final states. On the basis of a detailed case study ASCHER & REICH argue that implicit discretizations should not be used with timesteps significantly larger than typical periods of the fast oscilla-

tions. In the paper by BERNE, the r-RESPA multiple timestepping (MTS) method is described and applied in the context of Hybrid Monte Carlo methods for sampling techniques such as J-Walking and S-Walking with the aim of a more rapid exploration of rugged energy landscapes. In the next paper, SKEEL & IZAGUIRRE advocate the use of MTS in a mollified impulse method to overcome resonance instabilities that are inherent in the standard impulse method. Yet another MTS-like approach can be found in the paper by JANEŽIČ & MERZEL, who suggest to split off a harmonic high frequency part of the motion and integrate that analytically. Finally, LEIMKUHLER demonstrates the stability of the recently proposed explicit symplectic integrators (with fixed timestep) in the numerical integration of rigid body motion over long time spans.

The fourth chapter, on *Quantum-Classical Simulations*, deals with the integration of molecular systems, parts of which are modelled in terms of quantum mechanics, where a full quantum mechanical treatment would be impossible. In the first paper, JUNGWIRTH & GERBER treat clusters of inert gases by calculating effective single-mode potentials from classical molecular dynamics which are then used in quantum calculations. An extension beyond the separability approximation is also suggested. The quality of the quantum-classical molecular dynamics (QCMD) model compared with full quantum mechanics (QM) and the Born-Oppenheimer approximation (BO) is considered by SCHÜTTE & BORNEMANN in terms of approximation theory. They also suggest an extended QCMD model that may open new perspectives in the case of energy level crossings, where BO is known to break down. Recently developed structure-preserving numerical integrators for this QCMD model are given by NETTESHEIM & SCHÜTTE. Symplectic multiple timestepping variants of these integrators are derived in the paper by NETTESHEIM & REICH. An alternative scheme is presented by HOCHBRUCK & LUBICH, who suggest that a type of mollified exponential integrators are especially well-suited for highly oscillatory systems such as QCMD and the Car-Parrinello approximation. The latter approximation is also used in the paper by MEIER ET AL. on ab-initio MD simulations of catalysis in a polymerization process. In the last paper of this chapter, IZVEKOV describes an algorithm for the calculation of absorption spectra based on exciton-phonon interactions.

The fifth and final chapter, on *Parallel Force Field Evaluation*, takes account of the fact that the bulk of CPU time spent in MD simulations is required for evaluation of the force field. In the first paper, BOARD and his coworkers present a comparison of the performance of various parallel implementations of Ewald and multipole summations together with recommendations for their application. The second paper, by PHILLIPS ET AL., addresses the special problems associated with the design of parallel MD programs. Conflicting issues that shape the design of such codes are identified and the use of features such as multiple threads and message-driven execution is described. The final paper, by OKUNBOR & MURTY, compares three force decomposition techniques (the checkerboard partitioning method,

the force-row interleaving method, and the force-stripped row method) in the context of a benchmark test problem.

August 31, 1998

Peter Deuflhard
Jan Hermans
Benedict Leimkuhler
Alan E. Mark
Sebastian Reich
Robert D. Skeel

Table of Contents

Introductory Survey

Molecular Dynamics Simulations: The Limits and Beyond*

Herman J.C. Berendsen

BIOSON Research Institute and Dept of Biophysical Chemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, the Netherlands

Abstract. This article reviews the present state of Molecular Dynamics (MD) simulations and tries to give an outlook into future developments. First an overview is given of methods, algorithms and force fields. After considering the limitations of the standard present-day techniques, developments that reach beyond the present limitations are considered. These concern three major directions: (a) inclusion of quantum dynamics, (b) reduction of complexity by reducing the number of degrees of freedom and averaging over interactions with less important degrees of freedom, (c) reduction to *mesoscopic dynamics* by considering particle densities rather than positions. It is concluded that MD is a mature technique for classical simulations of all-atom systems in the nanosecond time range, but is still in its infancy in reaching reliably into longer time scales.

1 Introduction and a Bit of History

This conference on Algorithms for Molecular Simulation is a good occasion to pause for a moment and consider where we are now and where we go in this field. The book by Allen and Tildesley [2] describes most of the techniques that are still in use today.

Molecular Dynamics (MD) simulations were first carried out in 1957 by Alder and Wainwright on a hard-sphere fluid [2]; the first fluid with soft interactions was simulated by Rahman in 1964 [3] and the first complex fluid (water) was simulated by Rahman and Stillinger in 1971 [4]. The first MD simulation of a protein was carried out in 1976 by Andrew McCammon [2], then postdoc in Martin Karplus' group, during a CECAM workshop in Orsay, France. That workshop [6], which brought about 20 physicists (one was the 'father' of MD, Anees Rahman) and protein specialists (one was Jan Hermans) together for two months, has been a seminal event in the development of simulation of biological macromolecules. Since then methods and force fields have improved and computers have become a thousandfold more powerful. Routine simulations now comprise fully hydrated systems with tens of thousands of atoms and extend over nanoseconds. Simulated systems include DNA, liquid crystals, polymers and lipid membranes.

* A contribution from the Groningen Biomolecular Sciences and Biotechnology Institute.

But the methods have not really changed. The Verlet algorithm to solve Newton's equations, introduced by Verlet in 1967 [7], and it's variants are still the most popular algorithms today, possibly because they are time-reversible and symplectic, but surely because they are simple. The force field description was then, and still is, a combination of Lennard-Jones and Coulombic terms, with (mostly) harmonic bonds and periodic dihedrals. Modern extensions have added many more parameters but only modestly more reliability. The now almost universal use of constraints for bonds (and sometimes bond angles) was already introduced in 1977 [8]. That polarisability would be necessary was realized then [9], but it is still not routinely implemented today. Long-range interactions are still troublesome, but the methods that now become popular date back to Ewald in 1921 [10] and Hockney and Eastwood in 1981 [11].

What has been developed within the last 20 years is the computation of thermodynamic properties including free energy and entropy [12, 13, 14]. But the ground work for free energy perturbation was done by Valleau and Torrie in 1977 [15], for particle insertion by Widom in 1963 and 1982 [16, 17] and for umbrella sampling by Torrie and Valleau in 1974 and 1977 [18, 19]. These methods were primarily developed for use with Monte Carlo simulations; continuous thermodynamic integration in MD was first described in 1986 [20].

Another topic that received increasing attention is the incorporation of quantum methods into dynamic simulations. True quantum dynamics for hundreds of particles is beyond any foreseeable computational capability, and only approximations are viable. We should distinguish:

(*i*) The application of quantum corrections to classical MD. An early example is the application of quantum corrections to water based on classical frequency distributions by Behrens *et al.* [21].

(*ii*) The use of quantum methods to derive potentials for the heavy particles in the Born-Oppenheimer approximation during the MD simulation. This is now a very active field, with important applications for the study of chemical reactions in the condensed phase. Pioneering work using semi-empirical QM was done by Warshel [22], and in the groups of Jorgensen [23], Kollman [24], and Karplus [25, 26]. A methodological breakthrough was the introduction in 1985 of DFT (quantum density functional theory) to solve instantaneous forces on the heavy particles in a method that is now called *ab initio molecular dynamics*, by Car and Parrinello [27]. This method solves the dynamical evolution of the ground-state electronic wave function described as a linear combination of plane waves. Selloni *et al.* [28] solved the dynamical evolution of the ground state wave function of a solvated electron on a 3-D grid. These methods are *adiabatic* in the sense that they do not incorporate excitations of the quantum particles.

(*iii*) The use of quantum methods to obtain correct statistical static (but not dynamic) averages for 'heavy' quantum particles. In this category path-integral methods were developed on the basis of Feynman's path

integral formulation [29] mostly by Chandler and Wolynes [30]. For a lucid description see [31].

(*iv*) The incorporation of quantum-dynamical evolution for selected particles or degrees of freedom in a classical environment. With the inclusion of non-adiabatic transfers to excited states this is a rather new field, with important applications to proton transfer processes in the condensed phase. We shall return to these methods in section 3.2.

The important enquiry into long time scales has also been a subject of interest over many years, but the progress has been slow. Computer capabilities have increased so rapidly that it has often been worthwhile to wait a few years to obtain the required increase in speed with standard methods rather than invent marginal improvements by faster algorithms or by using reduced systems. Many attempts to replace the time-consuming solvent molecules by potentials of mean force (see for example [32]) or to construct an appropriate outer boundary without severe boundary effects [43, 34] have been made, but none of these are fully satisfactory. Really slow events cannot be modeled by such simplifications: a drastic reduction in the number of degrees of freedom is needed. When events are slow because an identifiable barrier must be crossed, good results can be obtained by calculating the free energy at the barrier in one or a few degrees of freedom. However, when events are slow because a very large multidimensional configurational space must be explored (as in protein folding or macromolecular aggregation), the appropriate methods are still lacking. We shall return to this important topic in Section 3.3.

2 Where Are We Now?

With the danger of severe oversimplification, which unavoidably leads to improper underevaluation of important recent developments, I shall try to indicate where traditional, classical MD has brought us today, or will bring us tomorrow. This concerns the techniques rather than the applications, which cannot be reviewed in the present context. The main aspects to consider concern algorithms and force fields.

2.1 Algorithms

As remarked in the introduction, the reversible Verlet algorithm or any of its disguised forms as velocity-Verlet or leap-frog, has remained strong and sturdy. Non-reversible higher-order algorithms of the predictor-corrector type, such as Gear's algorithms, may be useful if very high accuracy is required, but offer little advantage in cases where the evaluation of forces is accompanied by noisy errors [35].

An elegant derivation of the Verlet-type algorithms has been given by Tuckerman et al. [36] and is useful in multiple timestep implementations,

called RESPA (REference System Propagator Algorithms) [37, 20]. Because of their elegance I cannot resist to quote the principle of these algorithms. They are based on the Trotter-Suzuki expansion of exponential operators [39] which is much used in quantum simulations:

$$e^{i(A+B)t} \approx e^{iAt/2} e^{iBt} e^{iAt/2},$$

where A and B are non-commuting operators. This expansion is obviously time reversible and gives an error of third order in t. If applied to the classical Liouville operator acting on phase space, and separating the components acting on coordinates and on momenta, equations for momenta and coordinate updates per time step are obtained. Let us make it simple: In cartesian coordinates a point in phase space is represented by a vector of positions x and velocities v. Evolving the vector $(x, v)^T$ over a time(step) t means applying both a *force propagator*

$$U_f(t) \begin{pmatrix} x \\ v \end{pmatrix} = \begin{pmatrix} x \\ v + tF(x)/m \end{pmatrix}$$

and a *velocity propagator*

$$U_v(t) \begin{pmatrix} x \\ v \end{pmatrix} = \begin{pmatrix} x + vt \\ v \end{pmatrix}$$

Each of these operators is unitary: $U(-t) = U^{-1}(t)$. Updating a time step with the propagator $U_f(\frac{1}{2}\Delta t)U_v(\Delta t)U_f(\frac{1}{2}\Delta t)$ yields the velocity-Verlet algorithm. Concatenating the force operator for successive steps yields the leap-frog algorithm:

$$v(t + \frac{1}{2}\Delta t) = v(t - \frac{1}{2}\Delta t) + F[x(t)]\Delta t/m$$
$$x(t + \Delta t) = x(t) + v(t + \frac{1}{2}\Delta t)\Delta t$$

A double-timestep algorithm with short- and long-range forces is obtained by applying the propagator [36]:

$$U_l(\frac{1}{2}\Delta t)\,[U_s(\frac{1}{2}\delta t)U_v(\delta t)U_s(\frac{1}{2}\delta t)]^n\,U_l(\frac{1}{2}\Delta t),$$

where U_s and U_l are the propagators for the short-range, resp. the long-range force, and $\Delta t = n\delta t$. These algorithms are not only time reversible, but they are sympletic and conserve volume in phase space [40].

In practice modifications are made to incorporate thermostats or barostats that may destroy the time-reversible and symplectic properties. While extended-system algorithms such as Nosé dynamics [41] can be designed on the principles of the reversible operators, methods that use proportional velocity or coordinate scaling [42] cannot. Such methods are very

convenient and practical, but – unless the time constant used for coupling with an external bath are long with respect to the intrinsic time constant for kinetic energy dispersal – they give undefined ensembles with unreliable fluctuations and may produce spurious transfer of kinetic energy to uncoupled degrees of freedom, such as the overall translation and rotation of the system. Standard programs correct for such effects.

The incorporation of holonomic constraints for covalent bond lengths (and sometimes bond angles) saves roughly a factor of 4 in the allowed time step for molecular systems and has been common practice for many years. Conserving constraints involves the solution of a set of nonlinear equations, which can be solved iteratively, either by solving a matrix equation after linearization, or by iteratively solving successive equations for each constraint. The latter method is employed in the widely used SHAKE program [8]. Recently a linear constraint solver LINCS has been introduced [43] which is much faster and more stable than SHAKE and is better suited for implementation in programs for parallel computers. It is built in our MD package GROMACS[1] [44].

The question whether constraints for covalent bonds give better physics than harmonic oscillators is not really resolved. A mathematical argument can be given that specific motions which occur in a frequency range clearly separated from all other motions can be considered uncoupled and can then be treated as constraints without affecting the overall motion. For molecular systems such a separation is valid for bond length constraints, but not generally for bond angle constraints; the latter should therefore not be constrained in large molecules [3]. One should in principle take care of corrections due to the Jacobian of the transformation when using constraints, related to the configuration dependence of the *extent* of phase space on a curved constrained surface [46, 47], but this effect is negligible if only bond lengths are constrained. A physical argument for using bond constraints is that real covalent bonds correspond to quantum-mechanical harmonic oscillators with frequencies well above $k_B T/h$. They are thus permanently in the stationary ground state and do not take up any additional kinetic or potential energy. Treating them as classical oscillators would provoke unphysical energy exchange and requires the application of quantum corrections to the energy. While this is true, the treatment of bonds as constraints denies the generation of low-frequency modes from coupled vibrations, and also prevents the oscillators to relax under the influence of an external force. The latter effect, if adiabatically imposed, absorbs an energy of $\frac{1}{2}F^2/k$ (where F is the external force and k is the harmonic force constant) from the environment, both for classical and quantum oscillators. For an OH stretch in water, with external forces around 10^{-9} N/m [48] this energy is about 0.5 kJ/mol. Finally, both classical vibrations and constraints neglect the configuration dependence of the zero-point energy of quantum oscillators; this effect amounts in water (with a 300 cm^{-1} shift in vibrational frequency) to 1.8 kJ/mol per OH oscillator.

[1] See `http://rugmd0.chem.rug.nl/ gromacs/`

These numbers are not negligible. At present such effects are on the average compensated by other force field terms through empirical parametrization.

We may conclude that the matter of optimal algorithms for integrating Newton's equations of motion is now nearly settled; however, their optimal and prudent use [28] has not been fully exploited yet by most programs and may still give us an improvement by a factor 3 to 5.

2.2 Force Fields

While simulations reach into larger time spans, the inaccuracies of force fields become more apparent: on the one hand properties based on free energies, which were never used for parametrization, are computed more accurately and discrepancies show up; on the other hand longer simulations, particularly of proteins, show more subtle discrepancies that only appear after nanoseconds. Thus force fields are under constant revision as far as their parameters are concerned, and this process will continue. Unfortunately the form of the potentials is hardly considered and the refinement leads to an increasing number of distinct atom types with a proliferating number of parameters and a severe detoriation of transferability. The increased use of quantum mechanics to derive potentials will not really improve this situation: *ab initio* quantum mechanics is not reliable enough on the level of kT, and on-the-fly use of quantum methods to derive forces, as in the Car-Parrinello method, is not likely to be applicable to very large systems in the foreseeable future.

This situation, despite the fact that reliability is increasing, is very undesirable. A considerable effort will be needed to revise the shape of the potential functions such that transferability is greatly enhanced and the number of atom types can be reduced. After all, there is only one type of carbon; it has mass 12 and charge 6 and that is all that matters. What is obviously most needed is to incorporate essential many-body interactions in a proper way. In all present non-polarisable force fields many-body interactions are incorporated in an average way into pair-additive terms. In general, errors in one term are compensated by parameter adjustments in other terms, and the resulting force field is only valid for a limited range of environments.

A useful force field should be accurate and simple. Therefore it is desirable that polarisability be incorporated by changing charges (positions or magnitudes) rather than by incorporating induced dipole moments, which involve dipole field gradient tensors to be computed. The best candidate is the shell model, which represents electron clouds by charges on a spring; a detailed study of nitrogen in all its phases by Jordan [50] could serve as a guide. The task of generating a new general force field with proper many-body interactions comprises a complete overhaul of the present force fields and a completely new parametrization involving not only static data but also free energy evaluations. This non-rewarding task is not likely to be accomplished without an internationally organized concerted effort.

2.3 Long-Range Interactions

The proper treatment of long-range interactions has not yet been settled in a quite satisfactory way. The use of a cut-off range for dispersion interactions with r^{-6} radial dependence does not present a severe problem, although continuum corrections for the range beyond the cut-off are often necessary, particularly for pressure calculations. But a simple cutoff for Coulombic terms can give disastreus effects, specially when ionic species are present or when dielectric properties are required. It has been observed (see e.g. [51]) that in electrolyte solutions the radial distribution of like ions strongly peaks around the cutoff radius, while also the short-range structure is severely distorted. In dipolar systems without explicit charges a cut-off is tolerable to compute structural and energetic properties, as long as dipoles are not broken in the cut-off treatment, but dielectric properties cannot be evaluated with any precision [52]. Another effect of (sharp) cut-offs, among other artefacts, is the introduction of additional noise into the system. Sufficiently smooth cut-offs, on the other hand, severely deviate from the correct Coulomb interaction. Several structural and dynamic artefacts have been described, see e.g. [45, 54, 55, 56, 57, 58, 59, 60, 61, 62]. Therefore it is recommended that at least some method to incorporate the long-range part of electrostatic interactions be included.

If the simulated system uses periodic boundary conditions, the logical long-range interaction includes a lattice sum over all particles with all their images. Apart from some obvious and resolvable corrections for self-energy and for image interaction between excluded pairs, the question has been raised if one really wishes to enhance the effect of the artificial boundary conditions by including lattice sums. The effect of the periodic conditions should at least be evaluated by simulation with different box sizes or by continuum corrections, if applicable (see below).

A survey of available methods has, among others, been given by Smith and Van Gunsteren [51]; see also Gilson [63]. Here a short overview of methods, with some comment on their quality and properties, will be given.

Cut-off Methods The use of an abrupt *potential cut-off* radius r_c for the evaluation of the (electrostatic) potential, while feasible for Monte Carlo simulations, implies a delta-function in the force at r_c. If a delta function is really implemented, it gives an unphysical force when particle distances hit the cutoff radius, and a dynamic behaviour that is very dependent on the cut-off range [64]. The use of an abrupt *force cut-off* is computationally more acceptable, although it introduces additional noise. For systems containing dipolar groups that are represented by charges it is mandatory that the cut-off is based on *charge groups* rather than on individual charges [65]. One should realize, however, that a force cut-off implies a shift in the potential, since the latter is continuous at r_c and zero beyond r_c. Therefore Monte Carlo and MD simulations with cut-offs are not expected to give the same equilibrium results. A useful extension

is the use of a *twin-range cut-off*, where the forces from a second shell are evaluated every 10 to 20 steps simultaneously with the construction of a neighbour list for the first shell. Smooth cut-offs can be accomplished by *switching functions* applied to the potential. They generally cause a better behaviour of the integration algorithm, but also disguise the errors that are nevertheless made. Ding et al. [66] have shown that traditional switching functions cause large errors in structure and fluctuations when applied to a dendrimeric polypeptide; they suggest a large smoothing range between $0.4289 r_c$ and r_c. Fun Lau et al. [61] found structural and dynamical influences of switching functions on water, and Perera et al. [62] found severe influences on the dynamic properties of ions in aqueous solution. All cut-off methods suffer from severe distortions in systems containing full charges and deny the evaluation of dielectric properties. The latter is due to the fact that fluctuations of the total dipole moment in a sphere are much reduced when the sphere is surrounded by vacuum [52].

Reaction Field Methods In order to alleviate the quenching of dipole moment fluctuations, a reaction field due to a polarizable environment beyond the cut-off can be incorporated in conjunction with cut-off methods. The reaction field [52, 51] is proportional to the total dipole moment within the cut-off sphere and depends on the dielectric constant and ionic strength of the environment. On the same level, a reaction potential (Born potential) should be applied, proportional to the total charge in the cut-off sphere. This is only applicable for homogeneous fluids, and therefore not generally useful. However, in the case the dielectric constant or ionic strength of the environment is taken to be infinite, conducting or *tin-foil* boundary conditions arise with simple expressions for the forces. Such reaction fields and potentials are of course in general also incorrect, but they produce well-behaved forces and allow better subsequent corrections based on continuum theory (see next item), especially in systems like macromolecules in aqueous solution. The expression for the electric field at particle r_i is

$$\boldsymbol{E}(\boldsymbol{r}_i) = (4\pi\epsilon_0)^{-1} \sum_j q_j \boldsymbol{r}_{ij} (r_{ij}^{-3} - r_c^{-3}); \qquad (r_{ij} < r_c),$$

which amounts to a well-behaved *shifted force*. By integration the total electrical interaction energy becomes:

$$V_{el} = \frac{1}{8\pi\epsilon_0} \sum_{i,j} q_i q_j \left(\frac{1}{r_{ij}} + \frac{r_{ij}^2}{2r_c^3} - \frac{3}{2r_c} \right); \qquad (r_{ij} < r_c).$$

This is close, but not equal to the tin-foil Born-corrected energy

$$V_{\text{Born}} = \frac{1}{8\pi\epsilon_0} \sum_{i,j \neq i} q_i q_j \left(\frac{1}{r_{ij}} - \frac{1}{r_c} \right) - (\sum_i q_i^2) \frac{1}{r_c}.$$

The discrepancy is not large and the last term is zero for a system without net charge. Thus we see that the use of a shifted Coulomb force is equivalent to a tin-foil reaction field and almost equivalent to a tin-foil Born condition.

Continuum Corrections If the geometry of the simulated system is not too complex, it is possible to make corrections for the 'incorrect' long-range treatment, based on continuum considerations. This has been convincingly shown by Wood [67] in a paper that has not received the attention it should. The idea is that 'the computational world' has its own physics (like cut-offs and periodic boundary conditions), and that the differences with the 'real world' are fairly smooth and therefore can be treated by continuum methods. Such corrections were made on earlier simulations by Straatsma [68] on the free energy of ionic hydration, using various cut-offs for the ion-water and the water-water interactions. While in the original paper the usual Born correction was made, a discrepancy remained due to the neglect of water-water interactions between pairs of molecules that are both correlated with the ion. Wood showed that all results fitted beautifully after correction based on the spatial distribution of solvent polarization. Such corrections could also be made had a tin-foil reaction field been used.

The same idea was actually exploited by Neumann in several papers on dielectric properties [52, 69, 70]. Using a tin-foil reaction field the relation between the (frequency-dependent) relative dielectric constant $\epsilon(\omega)$ and the autocorrelation function of the total dipole moment $\boldsymbol{M}(t)$ becomes particularly simple:

$$\epsilon(\omega) - 1 = \frac{1}{3\epsilon_0 V k T} \int_0^\infty \left(-\frac{d}{dt} \langle \boldsymbol{M}(0) \boldsymbol{M}(t) \rangle \right) e^{-i\omega t} \, dt$$

Hierarchical Methods Methods that group more particles together for increasing distances [32, 72, 73] scale roughly proportional to the number of particles N or to $N \log N$, rather than to N^2 (as for straightforward summation over pairs). For large system sizes such linear hierarchical methods should win out over other methods. Hierarchical methods, routinely applied in the simulation of star clusters and galaxies, have also been adopted for proteins [74] and implemented in simulation programs for large molecular clusters [75]. These methods have been extensively compared to each other and to Ewald summation [76], with the result that they only surpass Ewald summation for particle numbers in the hundred thousands. Since it has been known for a long time that Ewald summation is considerably more expensive than a Poisson solver on a grid [11, 77] (see next item), I conclude that there is not much point in pursuing these methods for molecular simulation.

Separation of Short- and Long-Range; Ewald and PPPM Methods If we split the total Coulomb interaction in a short- and a long-range contribution, chosen to be smooth functions of the distance, the two

contributions can be more efficiently handled by different techniques [65]. The short-range contribution is evaluated for each pair on the basis of a pair list. The long-range part can be recast in terms of a Poisson problem and then solved by an appropriate Poisson solver. The choice of Poisson solver depends on whether the system is periodic or not, on the preconditioning (i.e., is an approximate solution available from the previous step?) and on the wish to implement the algorithm on a parallel computer. It is possible to update the long-range part less frequently in a multiple time-step algorithm. The popular Ewald summation [10, 78, 79] and its variants [80], the efficient Ewald-mesh technique [40], and the particle-particle particle-mesh (PPPM) method [11] are all special cases of this class of techniques.

Consider[2] a short-range interaction defined by the electric field at position $\boldsymbol{r}_i$:

$$\boldsymbol{E}^s(\boldsymbol{r}_i) = \frac{1}{4\pi\epsilon_o} \sum_j q_j f(r_{ij}) \frac{\boldsymbol{r}_{ij}}{r_{ij}^3}$$

The function $f(r)$ is a force-switching function that goes smoothly from 1 ar $r = 0$ to 0 at $r = r_c$. The long-range part of the field, i.e., what remains from the complete Coulomb field:

$$\boldsymbol{E}^l(\boldsymbol{r}_i) = \frac{1}{4\pi\epsilon_o} \sum_j q_j [1 - f(r_{ij})] \frac{\boldsymbol{r}_{ij}}{r_{ij}^3},$$

can be considered to be generated by a charge density $\rho(\boldsymbol{r})$:

$$\rho(\boldsymbol{r}) = \epsilon_0 \nabla \cdot \boldsymbol{E}^l = \sum_j q_j g(\boldsymbol{r} - \boldsymbol{r}_j),$$

with

$$g(\boldsymbol{r}) = g(r) = -\frac{1}{4\pi r^2} \frac{df}{dr}.$$

Thus the long-range field (and potential) is generated by a charge density which is the convolution of the charges in the system with a radial *charge spread function* $g(r)$ dictated by the short-range force-switching function $f(r)$. The task is to solve for the long-range potential $\phi^l(\boldsymbol{r})$ (the negative gradient of which is the long-range field) from the Poisson equation

$$\nabla^2 \phi^l = -\rho/\epsilon_0.$$

If a gaussian function is chosen for the charge spread function, and the Poisson equation is solved by Fourier transformation (valid for periodic

[2] The force function $f(r)$ differs from that in ref. [65] by a factor r^2, yielding simpler expressions. Some errors in that reference have been corrected.

boundary conditions), the Ewald summation method is recovered. If the Fourier solution is obtained on a grid, allowing the use of FFT, the Ewald-mesh method is obtained. However, the charge spread function need not be gaussian and can be chosen such that the short-range field and its derivative go exactly to zero at the cut-off radius. The shifted force obtained with the tin-foil reaction field (see above) corresponds to a charge spread function that spreads each charge homogeneously over a sphere with radius r_c, which is not an optimal choice to avoid truncation errors. Iterative Poisson solvers on a regular grid using Gauss-Seidel iteration with successive overrelaxation (SOR) [82] are less efficient than FFT methods, but are also applicable to non-periodic systems and are more easily parallellized. In an MD run, the previous step provides a near-solution and only a few iterations are needed per step [83].

Summarizing, the most efficient way to handle electrostatic interactions correctly seems to be the appropriate splitting into smooth short-range and long-range parts, and handling the latter by an efficient Poisson solver, using the knowledge available from the history of the trajectory, and exploiting the fact that the long-range part fluctuates on a longer time scale.

3 The Limits and Beyond

3.1 Limits to Traditional MD

Despite all the shortcomings listed above, full particle classical MD can be considered mature [84]. Even when all shortcomings will be overcome, we can now clearly delineate the limits for application. These are mainly in the size of the system and the length of the possible simulation. With the rapidly growing cheap computer memory shear size by itself is hardly a limitation: several tens of thousands of particles can be handled routinely (for example, we report a simulation of a porin trimer protein embedded in a phospholipid membrane in aqueous environment with almost 70,000 particles [85]; see also the contribution of K. Schulten in this symposium) and a million particles could be handled should that be desired.

The main limitation is in simulated time, which at present is in the order of nanoseconds for large systems. We may expect computational capabilities to increase by at least an order of magnitude every five years, but even at that rate it will take 15 years before we reach routinely into the microseconds. While adequate for many problems, such time scales are totally inadequate for many more other problems, particularly in molecular processes involving biological macromolecules. The obvious example of this is *protein folding*, which in favourable cases occurs in the real world on a time scale of seconds, to be reached *in computero* by the year 2040 if force fields are improved accordingly. The latter condition will not automatically be fulfilled: an error in the total energy difference between native and unfolded state of only 8

kJ/mol (on a total energy difference of some 250 kJ/mol, and with system energies measured in hundreds of MJ/mol) would shift a predicted 'melting' temperature by 10 ^{0}C. It is instructive to realize that the Born correction in water for a univalent ion for a cut-off radius of 1 nm amounts to 70 kJ/mol, which clearly shows that force field improvements must be accompanied by very careful evaluation of long-range interactions. The incorporation of polarisability is imperative if we realize that the polarization energy of a single carbon atom at a distance of 0.4 nm from a unit charge amounts to about 5 kJ/mol.

The nanoseconds limit also indicates a limit in the *configurational sampling* that can be achieved by MD. Sufficient sampling of the configurational space accessible in an equilibrium condition is essential for the computation of thermodynamic properties that involve entropy, as the latter is a measure for the extent of accessible configurational space. Use of other equilibrium sampling techniques, like Monte Carlo simulation, does not really improve on the statistics. However, substantial improvements are obtained and still to be expected from multiconfigurational sampling, umbrella sampling, and other methods that bias sampling to include infrequently visited regions[3], and from methods that circumvent barriers in configurational space such as the modification ('softening') of potential functions and the introduction of a fourth spatial dimension [86, 14].

3.2 Inclusion of Quantum Dynamics

A limitation of classical force field-based MD is the restriction to covalent complexes, with exclusion of chemical reactions. The very important applications to reactions in the condensed phase, including enzyme reactions and catalysis in general, need extension with dynamic behaviour of non-covalent intermediates. The latter must be described by quantum-mechanical methods. Usually, except for fast electron transfer reactions, the Born-Oppenheimer approximation is valid for the electronic motion. Classical dynamics is generally sufficiently accurate for atoms heavier than hydrogen, but for proton transfer reactions explicit quantum-dynamical treatment of the proton is required which fully includes tunneling as well as non-adiabatic involvement of excited states. We shall separately consider the two aspects: (i) computing the reactive Born-Oppenheimer surface in large condensed systems, and (ii) predicting reaction dynamics, including the quantum behaviour of protons.

As *ab initio* MD for all valence electrons [27] is not feasible for very large systems, QM calculations of an *embedded* quantum subsystem are required. Since reviews of the various approaches that rely on the Born-Oppenheimer approximation and that are now in use or in development, are available (see Field [87], Merz [88], Åqvist and Warshel [89], and Bakowies and Thiel [90] and references therein), only some summarizing opinions will be given here.

[3] see for example the contribution by Grubmüller in this symposium.

The first point to remark is that methods that are to be incorporated in MD, and thus require frequent updates, must be both accurate and efficient. It is likely that only semi-empirical and density functional (DFT) methods are suitable for embedding. Semi-empirical methods include MO (molecular orbital) [90] and valence-bond methods [89], both being dependent on suitable parametrizations that can be validated by high-level *ab initio* QM. The quality of DFT has improved recently by refinements of the exchange density functional to such an extent that its accuracy rivals that of the best *ab initio* calculations [91]. DFT is quite suitable for embedding into a classical environment [92]. Therefore DFT is expected to have the best potential for future incorporation in embedded QM/MD.

The second aspect, predicting reaction dynamics, including the quantum behaviour of protons, still has some way to go! There are really two separate problems: the simulation of a slow activated event, and the quantum-dynamical aspects of a reactive transition. Only fast reactions, occurring on the pico- to nanosecond time scale, can be probed by direct simulation; an interesting example is the simulation by *ab initio* MD of metallocene-catalysed ethylene polymerisation by Meier et al. [93].

Most reactions are too slow on a time scale of direct simulation, and the evaluation of reaction rates then requires the identification of a transition state (saddle point) in a reduced space of a few degrees of freedom (reaction coordinates), together with the assumption of equilibration among all other degrees of freedom. What is needed is the evaluation of the potential of mean force in this reduced space, using any of the available techniques to compute free energies. This defines the probability that the system resides in the transition-state region. Then the reactive flux in the direction of the unstable mode at the saddle point must be calculated. If friction for the motion over the barrier is neglected, a rate according to Eyrings transition-state theory is obtained. In general, the rate is smaller due to unsuccessful barrier crossing, as was first described by Kramers [94]. The classical transition rate can be properly treated by the reactive flux method [95], see also the extensive review by Hänggi [96]. The reactive flux can be obtained from MD simulations that start from the saddle point. An illustrative and careful application of the computational approach to classical barrier crossing, including a discussion of the effects due to the Jacobian of the transformation to reaction coordinates, has recently been described by den Otter and Briels [47].

While the classical approach to simulation of slow activated events, as described above, has received extensive attention in the literature and the methods are in general well established, the methods for quantum-dynamical simulation of reactive processes in complex systems in the condensed phase are still under development. We briefly consider electron and proton quantum dynamics.

The proper quantumdynamical treatment of fast electronic transfer reactions and reactions involving electronically excited states is very complex, not only because the Born-Oppenheimer approximation brakes down but

also because a reliable description of potential surfaces for excited states in complex molecules is very difficult, even within the Born-Oppenheimer approximation. With electron transfer processes in proteins one resorts to simplified descriptions, e.g. for electron transfer paths [97, 98] or for the protein environment[99]. In some cases, such as solvated electrons, the time-dependent electron wave function can be computed in a mixed QM/MD simulation, given a suitably parametrized interaction function between electron and solvent atoms [28, 100]. The electron wave function is then usually defined on a grid, and its time evolution solved by a split-operator technique using fast Fourier transforms. In these adiabatic simulations the electron remains in its Born-Oppenheimer ground state, although it is possible to compute excited states (and thus spectra) as well. Non-adiabatic transitions between different states, driven by solvent fluctuations, are generally not important for such systems in equilibrium at normal temperatures.

For the case of intramolecular energy transfer from excited vibrational states, a mixed quantum-classical treatment was given by Gerber *et al.* already in 1982 [101]. These authors used a time-dependent self-consistent field (TDSCF) approximation. In the classical limit of TDSCF averages over wave functions are replaced by averages over bundles of trajectories, each obtained by SCF methods.

Proton transfer over hydrogen bonds, which is a rate-determining step in many enzyme reactions, is a case where both the quantum character of the particle (implying proton tunneling) and the non-adiabatic transitions between pure states are important [102]. A quantum-dynamical treatment of the proton(s) requires embedding of one or more quantum degrees of freedom (d.o.f.) in a classical dynamic environment (yielding QD/MD) and there are two main points of discussion concerning the embedding procedure.

One of the discussion points is how the quantum system reacts back on the classical d.o.f., i.e., how the forces on the classical system should be derived from the quantum system. One can use the gradient of the *effective* energy, i.e., of the expectation value of the total energy

$$\boldsymbol{F}_i = -\frac{\partial}{\partial \boldsymbol{R}_i}\langle \Phi | H(\boldsymbol{R}) | \Phi \rangle,$$

where $\boldsymbol{R}_i$ are the coordinates of the i-th classical particle and Φ is the wave function which is parametrically dependent on $\boldsymbol{R}$. These forces are considered to be the 'exact' forces [103]. Alternatively, the Hellmann-Feynman force, i.e., the expectation value of the gradient of the energy

$$\boldsymbol{F}_i = -\langle \Phi | \frac{\partial H(\boldsymbol{R})}{\partial \boldsymbol{R}_i} | \Phi \rangle.$$

can be used. These expressions are only equivalent if Φ is a pure eigenstate. In other cases, *extended* Hellmann-Feynman forces [104] have been derived from the $\boldsymbol{R}$-dependence of Φ. However, if Φ represents a *mixed quantum state*

as correct solution to the time-dependent Schrödinger equation (see below), the correct force is [105, 106]

$$\boldsymbol{F}_i = -\langle \Phi | \frac{\partial V(\boldsymbol{R})}{\partial \boldsymbol{R}_i} | \Phi \rangle$$

and *not* the gradient of the effective potential. This is imperative to produce conservation of momentum over the whole system as can be seen immediately from the Ehrenfest relation [107], which states that the rate of change of the expectation value of the momentum of a quantum particle is

$$\frac{d\langle p \rangle}{dt} = -\langle \Phi | \frac{\partial V(\boldsymbol{r})}{\partial \boldsymbol{r}} | \Phi \rangle,$$

where $\boldsymbol{r}$ is the quantum coordinate. Since for a potential energy term between two particles of the form $V(|\boldsymbol{r} - \boldsymbol{R}|)$ the gradients with respect to $\boldsymbol{r}$ and $\boldsymbol{R}$ are equal and opposite, the total momentum change will be zero. A proof that both energy and momentum are conserved has been given in [106]. The correctness of this force has also been shown by Bornemann et al. [108] who derived the equations of motion when one particle of two interacting quantum particles is systematically taken to the classical limit.

The reason that non-adiabatic transitions must be included for protons is that fluctuations in the potential for the quantum degrees of freedom due to the environment (e.g. solvent) contain frequencies comparable to the transition frequencies between protonic quantum states. In such cases pure quantum states do not persist.

The second discussion point is how the actual quantum system is to be described: should one follow the time evolution of the time-dependent Schrödinger equation (TDSE) that allows mixed states to evolve, or should one insist on selecting a pure state, taking care of (sudden) transitions between states by some additional action in order to satisfy the time evolution of probabilities of states as dictated by the TDSE? The former approach was followed, among others, by Bala et al. in wave packet dynamics applied to proton transfer in phospholipase A2 [109, 110] and by us in the Density Matrix Evolution (DME) method which describes the mixed time-dependent wave function on a simple, appropriately chosen, basis set. [105, 111, 112, 113, 114, 106, 115]. DME is obviously not capable of giving a correct response of the classical environment to quantum transitions, but is perfectly able to describe initial rate processes or quantum systems that only weakly influence their environment. In fact, DME is the common method used in the evolution of nuclear spin magnetization [116]. The latter approach has led to the *surface hopping* method pioneered by Pechukas [117], with a modern formulation by Tully [118]. The basic idea is that the dynamics of a pure quantum state is followed, simultaneous with the classical dynamics of the environment. At every step the probability of a transition to another quantum state is calculated and such transitions are realized on a stochastic basis. When a transition is made, velocities are scaled to conserve total energy. The method has been

applied to single and multiple proton transfer reactions by Hammes-Schiffer and collaborators [119, 120, 121, 122, 103].

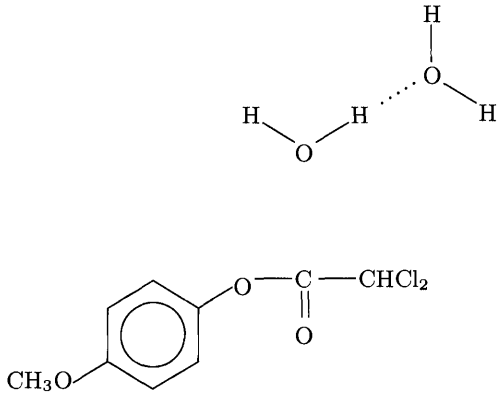
It seems that surface hopping (also called Molecular Dynamics with Quantum Transitions, MDQT) is a rather heavy tool to simulate proton dynamics. A recent and promising development is path integral centroid dynamics [123] that provides approximate dynamics of the centroid of the wavefunctions. Several improvements and applications have been published [123, 124, 125, 126, 127, 128].

We have applied the DME method to several proton transfer processes in aqueous solution (hydrogen malonate [111, 114], HIV protease [129, 130], and a slow neutral ester hydrolysis [131]. The latter case is a slow reaction with a rate constant of about once per minute, where the rate-limiting step is the transfer of a proton from a water molecule (which is in close proximity to the ester carbon, see Fig. 1a) to another water molecule, concerted with electron redistribution that leads to the tetrahedral intermediate (Fig. 1b). We used a combination of semi-empirical QM (to obtain the energy of the reactant state as a function of the distance between ester carbon and water oxygen and to obtain partial charges to be used in MD for intermediate states), thermodynamic integration (to obtain the free energy difference in solution between the equilibrium reactant state and the activated state from which proton transfer is probable), and DME (to obtain the initial rate of proton transfer). For the latter the proton potential along the transfer path was calculated during a biased MD simulation.

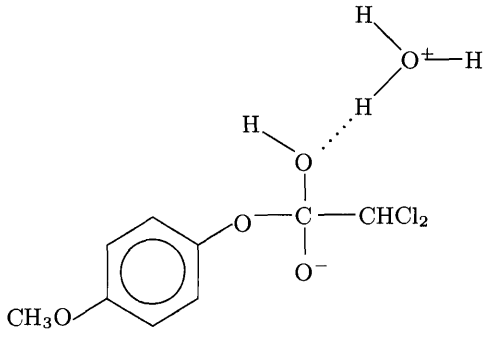
This proton potential, which is rapidly and heavily fluctuating (Fig. 2), drives the evolution of the proton wave function. The rate of transfer was then deduced from the coarse-grained average increase of proton population at the reactant side. Only the initial rate can be determined this way; this has the advantage that the protonic motion does not yet 'backreact' on the environment, but it has the disadvantage that no information is obtained about recrossing events. A good agreement with experimental rates was found, and a deuterium isotope effect of 3.9 was simulated, compared to the experimental value of 3.2. The picture is that proton transfer is driven by fluctuations due to solvent motion which transiently provide good tunneling conditions; the probability of transfer is however almost always negligibly low except when the carbon-water distance happens to be favourable for the reaction. The free energy of activation to reach this favourable state is one of the main reasons why the reaction is so slow. While the method is still crude, it can provide us with good estimates for reactions rates, including enzyme catalysis, when proton transfer is the rate-limiting step.

3.3 Reduction of Complexity: Averaging over Degrees of Freedom

Substantial headway towards longer time scales and larger systems can only be expected from reduction of system complexity. It is here where future



(a) Reactant state



(b) Product state

Fig. 1. The rate-determining step in the neutral hydrolysis of paramethoxy-phenyl dichloroacetate. In the reactant state (a) a water molecule is in proximity of the carbonyl carbon; after concerted proton transfer to a second water molecule and electron redistribution, a tetrahedral intermediate (b) is formed.

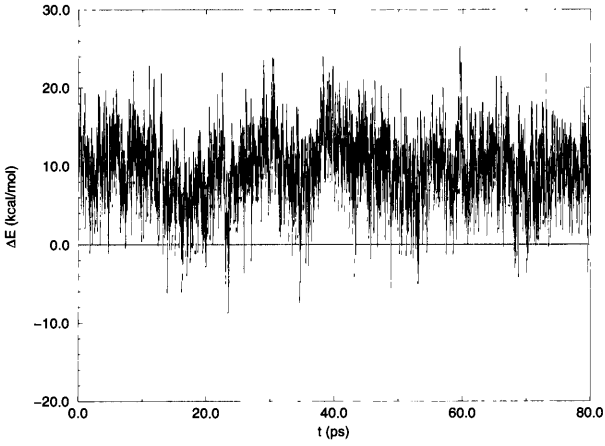


Fig. 2. The fluctuating difference between the proton potential at the product side relative to that at the reactant side (the difference between the two wells in a double-well proton potential). Whenever this difference is close to zero, tunneling conditions are favourable.

innovations are expected to be most fruitful. All such reductions concern in one way or another the omission of the explicit description of a (large) fraction of the degrees of freedom in the system.

The first requirement is the definition of a low-dimensional space of 'reaction coordinates' that still captures the essential dynamics of the processes we consider. Motions in the perpendicular 'null space' should have irrelevant detail and equilibrate fast, preferably on a time scale that is separated from the time scale of the 'essential' motions. Motions in the two spaces are separated much like is done in the Born-Oppenheimer approximation. The average influence of the fast motions on the 'essential' degrees of freedom must be taken into account: this concerns (*i*) correlations with positions expressed in a potential of mean force, (*ii*) correlations with velocities expressed in frictional terms, and (*iii*) an uncorrelated remainder that can be modeled by stochastic terms. Of course, this scheme is the general idea behind the well-known Langevin and Brownian dynamics.

In special cases (as in colloidal solutions) some particles can be considered as 'essential' and other particles as 'irrelevant', but in most cases the essential space will itself consist of collective degrees of freedom. A reaction coordinate for a chemical reaction is an example where not a particle, but some function of the distance between atoms is considered. In a simulation of the permeability of a lipid bilayer membrane for water [132] the 'reaction coordinate' was taken as the distance, in the direction perpendicular to the bilayer, between the center of mass of a water molecule and the center of mass of the rest of the system. In proteins (see below) a few collective degrees of freedom involving all atoms of the molecule, describe almost all the

macromolecular motion. One can also consider local densities of molecular species or molecular groups as essential degrees of freedom; in that case local flux densities are considered that respond to gradients of the thermodynamic potential (see below).

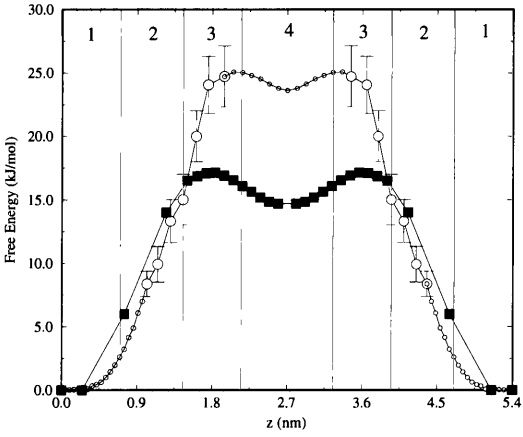


Fig. 3. The potential of mean force for a water molecule (open symbols) and an NH_3 molecule (filled symbols) penetrating into a lipid bilayer membrane. The aqueous phases are on both sides (regions 1), the headgroups are predominantly in regions 2, regions 3 contain the most ordered part of the hydrocarbon phase, and region 4 represents the middle of the hydrocarbon phase containing the end groups. The large circles with error bars were determined by integrating average constraint forces; the smaller points in regions 1 and 2 were determined directly from the observed water density and the smaller points in the middle were obtained by particle insertion (Marrink et al., J. Phys. Chem. **98** (1994) 4155 and *ibid.* **100** (1996) 16729).

Potentials of Mean Force In principle the potential of mean force in which the 'essential' coordinates move can be determined from detailed simulations including all degrees of freedom, in which the essential degrees of freedom are either treated as constraints or are restrained with an umbrella potential. In simple, low-dimensional cases this actually works: for example, in the simulation of the transport of water and other small molecules through a lipid bilayer [132, 133] the potential of mean force for a small molecule as function of the depth z in the membrane could be derived from integration of the mean force acting on a molecule constrained at a given z, using several constrained MD runs (see Fig. 3). The results could be checked in this case

by particle insertion in the middle of the bilayer. From analysis of the force fluctuations in the same simulations it was possible to derive the diffusion coefficient in the z-direction as a function of z, thus yielding a complete one-dimensional picture of the transfer path and allowing the computation of the permeation coefficient. A similar procedure was applied to a single file of hydrogen-bonded water molecules [134].

However, in many applications the essential space cannot be reduced to only one degree of freedom, and the statistics of the force fluctuation or of the spatial distribution may appear to be too poor to allow for an accurate determination of a multidimensional potential of mean force. An example is the potential of mean force between two ions in aqueous solution: the momentaneous forces are two orders of magnitude larger than their average which means that an error of 1% in the average requires a simulation length of 10^8 times the correlation time of the fluctuating force. This is in practice prohibitive. The errors do not result from incorrect force fields, but they are of a statistical nature; even an exact force field would not suffice.

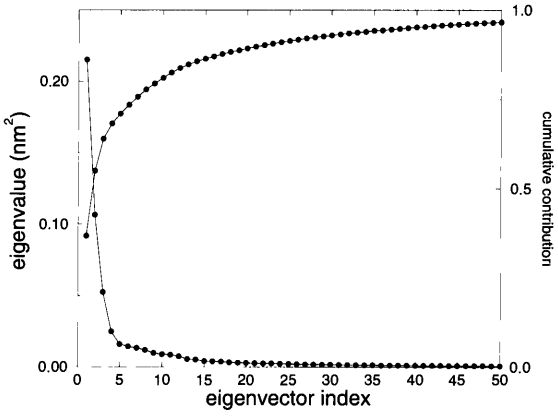
Thus one must rely on macroscopic theories and empirical adjustments for the determination of potentials of mean force. Such empirical adjustments use free energy data as solubilities, partition coefficients, virial coefficients, phase diagrams, etc., while the frictional terms are derived from diffusion coefficients and macroscopic theories for hydrodynamic interactions. In this whole field of enquiry progress is slow and much work (and thought!) will be needed in the future.

Essential Dynamics and Rigid Bodies in Proteins Proteins happen to be a very special kind of polymer, constructed such that they fold into specific structures, necessary for specific functions. But they are also flexible in a functional way, because functions as substrate binding, catalysis, allosteric effects and regulation require specific dynamical properties. It is therefore not too surprising that proteins are found to exhibit a collective motions with large amplitude in a low-dimensional subspace, described by 10 to 20 collective degrees of freedom, while the other 10,000 degrees of freedom seem to perform insignificant, low-amplitude and gaussian-distributed fluctuations [135, 137].

The 'essential degrees of freedom' are found by a principal component analysis of the position correlation matrix C_{ij} of the cartesian coordinate displacements x_i with respect to their averages $\langle x_i \rangle$, as gathered during a long MD run:

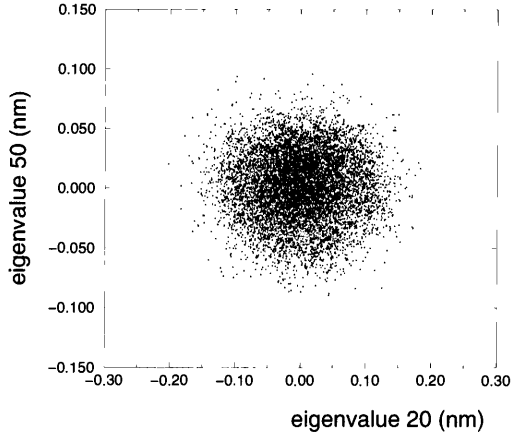
$$C_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle.$$

Each eigenvalue of this matrix indicates the contribution to the total fluctuation of the collective degree of freedom given by the corresponding eigenvector. If eigenvectors are ordered in descending order, it is generally observed [135, 136, 138] that the first 20 eigenvalues contribute some 80% of



m

(a) Eigenvector magnitudes in descending order (left scale) and cumulative contribution to total fluctuation (right scale)



m

(b) Fluctuation in the plane of the 20^{th} and 50^{th} eigenvector, showing virtually complete sampling of independent gaussian distributions.

Fig. 4. (a) Eigenvalues and distributions for the α-carbon atoms in the 56-residue B1-domain of streptococcal protein G, from a 1 ns MD simulation in water (courtesy of Bert de Groot, Groningen).

the total fluctuation. After the, say, 10^{th} eigenvector the distribution of the fluctuation is rather gaussian and uninteresting; such motions could be considered as 'near-constraints' and good candidates for reduction of complexity (see fig. 4).

With the very nonlinear force field expressed in all coordinates, it seems intractable to reformulate the potential energy surface as a function of the essential degrees of freedom, while treating the other d.o.f. as full constraints. What is possible is to *drive* the dynamics in a space of a few (2 or 3) degrees of freedom [139, 140] so as to expand the sampled volume. This is done by preventing a MD step to decrease the distance in that space with respect to a given reference point. In this way the borders of the accessible region are probed and a much more effective sampling is obtained than with normal MD; in the case of a cyclic polypeptide in solution the full conformational space could be sampled with this method [140]. The method is useful to characterize different type of motions, e.g. to distinguish between functional motions and motions that relate to instability and early unfolding processes [141].

If proteins are so full of internal 'near-constraints', we may take the analysis even a step further and investigate whether the protein is built from building blocks that can be approximated as rigid bodies. If there are n rigid building blocks, there are at most $6(n-1)$ internal degrees of freedom, most of which are likely to be additionally constrained. Recently Hayward [142, 143] has devised an automatic procedure that detects rigid bodies and characterizes the mutual motion of each pair, given at least two different conformations of the protein. These conformations can either be obtained from X-ray data, or from an essential dynamics analysis of a MD simulation.

The analysis (in the case of two structures) starts by a translational-rotational fit of the two structures and constructing the displacement vectors of all backbone atoms. Considering these as samples of a vector field, the curl of that vector field is computed, as sampled by each aminoacid. Thus a collection of rotation vectors is obtained. If a rigid body exists, the rotation vectors of all aminoacids in that body are equal, and different from rotation vectors in other rigid bodies. A standard cluster analysis on the rotation vectors, using the ratio of external to internal motion as a discrimination criterion, is then carried out. This yields a subdivision of the protein in semi-rigid bodies (if they exist) and identifies the links between them. The type of motion of one rigid body with respect to another is then analysed in terms of a unique axis, and such (hinge bending) motions can be objectively characterized as closing or twisting motions (Fig. 5).

Interestingly, there are many proteins with two domains that show a very clear hinge-bending motion with an obvious functional significance. Such domains have often been reported in the literature, but were never detected on an automated basis.

Fig. 5. Rigid-body analysis of citrate synthase, using two X-ray structures (after Hayward and Berendsen, Proteins **30** (1998) 144). The decomposition of the protein into two domains (dark gray and white) and two interconnecting regions (light gray) is shown, together with the hinge axis for the closing/opening motion between them.

Mesoscopic Dynamics After suitable collective degrees of freedom have been defined, their equations of motion can be systematically derived by averaging over equilibrium dynamics of the other degrees of freedom. There is a hierarchy of methods of the generalised Langevin dynamics type, with details that depend on the kind of averaging. Coarse-graining in time leads first to neglect of the time-dependent details of the stochastic contribution and then to the neglegt of inertial terms, resulting in *Brownian dynamics.* Methods of this type are still under development.

When a system contains gradients that are small over atomic distances, a big leap in time and space can be taken by describing the collective behaviour of *species s* (as a molecule or a collection of atoms in a molecule) that interact with neighbours through a simple hamiltonian (e.g. representing a gaussian chain) and feel the environment through a mean field defined by the local chemical potential $\mu_s(\boldsymbol{r})$. The variables are the space- and time-dependent densities of the species ρ_s. What in fact is done, is averaging locally over

nearby space and temporally over short times. The resulting method rests on the classical density functional formalism of Ginzburg and Landau [144]; its dynamical form [145, 149] is called *mesoscopic dynamics.*

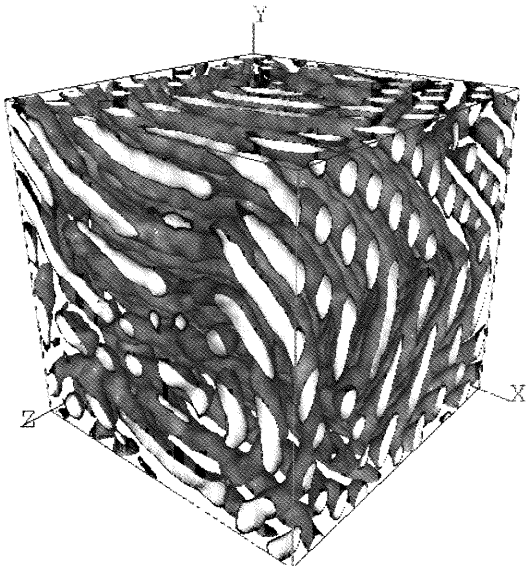


Fig. 6. Snapshot from a dynamic density functional simulation of the self-organisation of the block copolymer PL64 (containing 30 propylene oxide and 26 ethylene oxide units: $(EO)_{13}(PO)_{30}(EO)_{13}$) in 70% aqueous solution. The simulation was carried out during 6250 time steps on a $64 \times 64 \times 64$ grid (courtesy of B.A.C. van Vlimmeren and J.G.E.M. Fraaije, Groningen).

If there are no reactions, the conservation of the total quantity of each species dictates that the time dependence of ρ_s is given by minus the divergence of the flux $\rho_s \langle v_s \rangle$, where $\langle v_s \rangle$ is the drift velocity of the species s. The latter is proportional to the average force acting locally on species s, which is the *thermodynamic force*, equal to minus the gradient of the thermodynamic potential. In the local coupling approximation the mobility appears as a proportionality constant M. For spontaneous processes near equilibrium it is important that a noise term $\eta(t)$ is retained [146]. Thus dynamic equations of the form

$$\frac{\partial \rho}{\partial t} = M \nabla \cdot \rho \nabla \mu + \eta$$

are obtained (see [145, 147, 149] for details). The chemical potentials are derived from the functional derivative of the total free energy F, which itself is a functional of the density distribution:

$$\mu_s = \frac{\delta}{\delta \rho_s} F[\rho].$$

The method has severe limitations for systems where gradients on near-atomic scale are important (as in the protein folding process or in bilayer membranes that contain only two molecules in a separated phase), but is extremely powerful for (co)polymer mixtures and solutions [147, 148, 149]. As an example Fig. 6 gives a snapshot in the process of self-organisation of a polypropylene oxide-ethylene oxide copolymer PL64 in aqueous solution on its way from a completely homogeneous initial distribution to a hexagonal structure.

4 Conclusions

In this contribution I have tried to give a, necessarily subjective, review of the history, state-of-the-art, and future development of molecular dynamics simulations. The method is mature, but has not yet realized the full potential of the best algorithms, long-range interactions, and transferable, polarisable force fields. The near future will undoubtedly see the proper and efficient incorporation of quantum methods where needed, and an explosion in the simulation of reactive events, including enzyme reactions, is expected.

For the simulation of long-time and large-scale events a systematic hierarchy of models is in principle available, ranging from atomic detail to fluid dynamics, but there are many gaps to be filled and some of this territory is still unknown. On a mesoscopic scale promising methods have been developed, and I am optimistic that the gap between the atomic and mesoscopic scale will be filled in. But my optimism ends with the question whether one will ever be able to fold a protein reliably by *ab initio* simulation, starting from an aminoacid sequence alone and using only physical principles without resorting to database knowledge. For the most difficult questions one must be practical and exploit all available information, whether derived from theory or from experimental data in the broadest sense.

Acknowledgements

My initiation in the field of MD rests with the late Anees Rahman, whose insight and integrity has stimulated many of us, but who died far too young. Many CECAM activities would not have been realized without the stimulus provided by Carl Moser. The work in our laboratory has involved many collaborators over the years, all of whom I thank for their contributions, but with special thanks to the 'seniors' Wilfred van Gunsteren and Hans Fraaije. I have quoted work of Andrea Amadei, Bert de Groot, Steven Hayward, Berk Hess, Marc Lensink, Siewert-Jan Marrink, Janez Mavri, David van der Spoel, and Bernard van Vlimmeren: thank you all.

References

1. Allen, M.P., Tildesley, D.J.: Computer Simulation of Liquids. Clarendon Press, Oxford (1987)
2. Alder, B.J., Wainwright, T.E.: Phase transition for a hard sphere system. J. Chem. Phys. **27** (1957) 1208–1209.
3. Rahman, A.: Correlations in the motions of atoms in liquid argon. Phys. Rev. **136A** (1964) 405–411.
4. Rahman, A., Stillinger, F.H.: Molecular dynamics study of liquid water. J. Chem. Phys. **55** (1971) 3336–3359.
5. McCammon, J.A., Gelin, B.R., Karplus, M.: Dynamics of folded proteins. Nature **267** (1977) 585–590.
6. Berendsen, H.J.C. (ed.): Models for Proteins. CECAM Workshop Report 1976, Centre Européen de Calcul Atomique et Moléculaire, Orsay, France (present address: CECAM, ENS, Aile LR5, 46, Allée d'Italie, 69364 Lyon Cedex 07, France).
7. Verlet, L.: Computer 'experiments' on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. Phys. Rev. **165** (1967) 98–103.
8. Ryckaert, J.-P., Ciccotti,G., Berendsen, H.J.C.: Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. Comput. Phys. **23** (1977) 327-341.
9. Berendsen, H.J.C. (ed.): MC and MD of Water. CECAM Workshop Report 1972, Centre Européen de Calcul Atomique et Moléculaire, Orsay, France (present address: CECAM, ENS, Aile LR5, 46, Allée d'Italie, 69364 Lyon Cedex 07, France).
10. Ewald, P.P.: Die Berechnung optischer und elektrostatischer Gitterpotentiale. Ann. Phys. **64** (1921) 253–287.
11. Hockney, R.W., Eastwood, J.W.: Computer Simulation Using Particles. McGraw Hill, New York (1981).
12. Frenkel, D.: Free energy computation and first order phase transitions. In *Molecular Dynamic Simulation of Statistical Mechanical Systems*, Enrico Fermi Summer School, Varenna 1985, G. Ciccotti and W. Hoover, eds, North Holland, Amsterdam (1986) 43–65.
13. Free energy via molecular simulation: A primer. In: *Computer Simulations of Biomolecular Systems, Vol 2*, W.F. van Gunsteren, P.K. Weiner and A.J. Wilkinson, eds. Escom, Leiden (1993) 267–314.
14. Van Gunsteren, W.F., Beutler, T.C., Fraternali, F., King, P.M., Mark, A.E., Smith, P.E.: Computation of free energy in practice: Choice of approximations and accuracy limiting factors, in: *Computer Simulations of Biomolecular Systems, Vol 2*, W.F. van Gunsteren, P.K. Weiner and A.J. Wilkinson, eds. Escom, Leiden (1993) 315–348.
15. Valleau, J.P., Torrie, G.M.: A guide to Monte Carlo for statistical mechanics. 2. Byways, in *Statistical Mechanics A, Modern Theoretical Chemistry*, B.J. Berne, ed., Plenum Press, New York, **5** (1977) 169–194.
16. Widom, B.: Some topics in the theory of fluids. J. Chem. Phys. **39** (1963) 2808–2812.
17. Widom, B.: Potential distribution theory and the statistical mechanics of fluids. J. Phys. Chem **86** (1982) 869–872.

18. Torrie, G.M., Valleau, J.P.: Monte Carlo free energy estimates using non-Boltzmann sampling: application to the subcritical Lennard-Jones fluid. Chem. Phys. Lett. **28** (1974) 578–581.
19. Torrie, G.M., Valleau, J.P.: Nonphysical sampling distributions in Monte Carlo free energy estimation: umbrella sampling. J. Comput. Phys. **23** (1977) 187–199.
20. Berendsen, H.J.C., Postma, J.P.M., Van Gunsteren, W.F.: Statistical mechanics and molecular dynamics: The calculation of free energy. in *Molecular Dynamics and Protein Structure*, J. Hermans, ed., Polycrystal Book Service, PO Box 27, Western Springs, Ill., USA, (1985) 43–46.
21. Behrens, P.H., Mackay, D.H.J., White, G.M., Wilson, K.R.: Thermodynamics and quantum corrections from molecular dynamics for liquid water. J. Chem. Phys. **79** (1983) 2375–2389.
22. Warshel, A.: Computer Modeling of Chemical Reactions in Enzymes and Solutions. Wiley, New York (1992).
23. Chandrasekhar, J., Smith, S,F,m Jorgensen, W.L.: Theoretical examonation of the S_N2 reaction involving chloride ion and methyl chloride in the gas phase and aqueous solution. J. Amer. Chem. Soc. **107** (1985) 154–163.
24. Singh, U.C., Kollman, P.A.: A combined *ab initio* quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the $CH_3Cl + Cl^-$ exchange reaction and gas phase protonation of polyethers. J. Comput. Chem. **7** (1986) 718–730.
25. Bash, P.A., Field, M.J.,Karplus, M.: Free energy perturbation method for chemical reactions in the condensed phase: A dynamical approach based on a combined quantum and molecular dynamics potential. J. Am. Chem. Soc. **109** (1987) 8092–8094.
26. Field, M.J., Bash, P.A., Karplus, M.: A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. J. Comput. Chem. **11** (1990) 700–733.
27. Car, R., Parrinello,M.: Unified approach for molecular dynamics and density functional theory. Phys. Rev. Lett. **55** (1985) 2471–2474.
28. Selloni,A., Carnevalli, P., Car, R., Parrinello, M.: Localization, hopping, and diffusion of electrons in molten salts. Phys. Rev. Lett. **59** (1987) 823–826.
29. Feynman, R.P., Hibbs, A.R.: Quantum Mechanics and Path Integrals. McGraw-Hill, New York (1965).
30. Chandler, D., Wolynes, P.G.: Exploiting the isomorphism between quantum theory and classical statistical mechanics of polyatomic fluids. J. Chem. Phys. **74** (1981) 4078–4095.
31. Gillan, M.J.: The quantum simulation of hydrogen in metals. Philosoph. Mag. **A 58** (1988) 257–283.
32. Stouten, P.F.W., Frömmel, C., Nakamura, H., Sander, C.: An effective solvation term based on atomic occupancies for use in protein simulations. Mol. Simul. **10** (1993) 97–120.
33. Brooks III, C.L., Karplus, M.: Deformable stochastic boubdaries in molecular dynamics. J. Chem. Phys. **79** (1983) 6312–6325.
34. Juffer, A.H., Berendsen, H.J.C.: Dynamic surface boundary conditions: A simple boundary model for molecular dynamics simulations. Mol. Phys. **79** (1993) 623–644.

35. Berendsen. H.J.C., Van Gunsteren, W.F.: Practical algorithms for dynamic simulations, in *Molecular Dynamics Simulations of Statistical Mechanical Systems*, G. Ciccotti, ed., Soc. Italiana di Fisica, Bologna (1987) 43–65.
36. Tuckerman, M., Berne, B.J., Martyna, G.J.: Reversible multiple timescale molecular dynamics. J. Chem. Phys. **97** (1992) 1990–2001.
37. Tuckerman, M.E., Berne, B.J., Rossi, A.: Molecular dynamics algorithm for multiple time scales: systems with long range forces. J. Chem. Phys. **94** (1991) 6811–6815.
38. Tuckerman, M.E., Berne, B.J., Martyna, G.J.: Molecular dynamics algorithm for multiple time scales: systems with disparate masses. J. Chem. Phys. **94** (1991) 1465–1469.
39. De Raedt, H.: Product formula algorithms for solving the time-dependent Schrödinger equation. Comput. Phys. Rep. **7** (1987) 1–72.
40. Okunbor, D.I., Skeel, R.D.: Canonical numerical methods for molecular dynamics simulations. J. Comput. Chem. **15** (1994) 72–79.
41. Nosé, S.: A molecular dynamics method for simulations in the canonical ensemble. Mol. Phys. **52** (1984) 255–268; *ibid.* A unified formulation of the constant temperature molecular dynamics method. J. Chem. Phys. **81** (1984) 511–519.
42. Berendsen , H.J.C., Postma, J.P.M., Van Gunsteren, W.F., DiNola, A., Haak, J.R.: Molecular dynamics with coupling to an external bath. J. Chem. Phys. **81** (1984) 3684–3690.
43. Hess, B., Bekker, H., Berendsen, H.J.C., Fraaije, J.G.E.M.: LINCS: A linear constraint solver for molecular simulations. J. Comput. Chem. **18** (1997), 1463–1472.
44. Berendsen, Van der Spoel, D. Van Drunen, R.: GROMACS: A message-passing parallel molecular dynamics implementation. Comput. Phys. Comm. **91** (1995) 43–56.
45. Van Gunsteren, W.F., Berendsen, H.J.C.: Algorithms for macromolecular dynamics and constraint dynamics. Mol. Phys. **34** (1977) 1311-1327.
46. Berendsen, H.J.C. Van Gunsteren, W.F.: Molecular dynamics with constraints, in *The Physics of Superionic Conductors and Electrode Materials* ed. J.W. Perram, NATO ASI Series **B 92** (1983) 221–240 (Plenum, New York).
47. Den Otter, W.K., Briels, W.J.: The reactive flux method applied to complex reactions: using the unstable normal mode as a reaction coordinate. J. Chem. Phys. **106** (1997) 1–15.
48. Postma, J.P.M., Berendsen, H.J.C., Straatsma, T.P.: Intramolecular vibrations from molecular dynamics simulations of liquid water. Journal de Physique **C7** (1984) 31–40.
49. Procacci, P., Darden, T., Marchi, M.,: A very fast molecular dynamics method to simulate biomolecular systems with realistic electrostatic interactions. J. Phys. Chem. **100** (1996) 10464–10468.
50. Jordan, P.C., Van Maaren, P.J., Mavri, J., Van der Spoel, D., Berendsen, H.J.C.: Towards phase transferable potential functions: Methodology and application to nitrogen. J. Chem. Phys. **103** (1995) 2285.
51. Smith, P.E and Van Gunsteren, W.F.: Method for the evaluation of long range electrostatic forces in computer simulations of polar systems, in: *Computer Simulations of Biomolecular Systems, Vol 2*, W.F. van Gunsteren, P.K. Weiner, A.J. Wilkinson, eds. Escom, Leiden (1993) 182–212.
52. Neumann, M.: Dipole moment fluctuation formulas in computer simulations of polar systems. Mol. Phys. **50** (1983) 841–858.

53. Loncharich, R.J., Brooks, B.R.: The effects of truncating long-range forces on protein dynamics. Proteins **6** (1989) 32–45.
54. Kitchen, D.B., Hirata, F., Westbrook, J.D., Levy, R.: Conserving energy during molecular dynamics simulations of water, proteins, and proteins in water. J. Comput. Chem. **11** (1990) 1169–1180.
55. Schreiber, H., Steinhauser, O.: Cutoff size does strongly influence molecular dynamics results on solvated polypeptides. Biochem. **31** (1992) 5856–5860.
56. Schreiber, H., Steinhauser, O.: Molecular dynamic simulation studies of solvated polypeptides: why the cutoff scheme does not work. Chem. Phys. **168** (1992) 75–89.
57. Guenot, J., Kollman, P.A.: Conformational and energetic effects of truncating nonbonded interactions in an aqueous protein dynamics simulation. J. Comput. Chem. **14** (1993) 295–311.
58. Tasaki, K., McDonald, S., Brady, J.W.: Observations concerning the treatment of long range interactions in molecular dynamics simulations. J. Comput. Chem. **14** (1993) 278–284.
59. York, D.M., Wlodawer, A., Pederson, L.G., Darden, T.A.: Atomic-level accuracy in simulation of large protein crystals. Proc. Natl Acad. Sci. USA **91** (1994) 8715–8718.
60. Saito, M: Molecular dynamics simulations of proteins in solution: artefacts caused by the cutoff approximation. J. Chem. Phys. **101** (1994) 4055–4061.
61. Fun Lau, K., Alper, H.E., Thacher, T.S., Stouch, T.R.: Effects od switching functions on the behavior of liquid water in molecular dynamics. J. Phys. Chem. **98** (1994) 8785–8792.
62. Perera, L., Essmann, U., Berkowitz, M.: Effect of treatment of long-range forces on the dynamics of ions in aqueous solutions. J. Chem. Phys. **102** (1995) 450–456.
63. Gilson, M.K.: Theory of electrostatic interactions in macromolecules. Curr. Opinion Struct. Biol. **5** (1995) 216–223.
64. Teleman, O.: An efficient way to conserve the total energy in molecular dynamics simulations; boundary effects on energy conservation and dynamic properties. Mol. Simul. **1** (1988) 345–355.
65. Berendsen, H.J.C.: Electrostatic interactions, in: *Computer Simulations of Biomolecular Systems, Vol 2*, W.F. van Gunsteren, P.K. Weiner and A.J. Wilkinson, eds. Escom, Leiden (1993) 161–181.
66. Ding, H.Q., Karawasa, N., Goddard III, W.A.: Optimal spline cutoffs for Coulomb and van der Waals interactions. Chem. Phys. Lett. **193** (1992) 197–201.
67. Wood, R.H.: Continuum electrostatics in a computational universe with finite cut-off radii and periodic boundary conditions: Correction to computed free energies of ionic solvation. J. Chem. Phys. **103** (1995) 6177–6187.
68. Straatsma, T.P, Berendsen, H.J.C.: Free energy of ionic hydration: Analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulations. J. Chem. Phys. **89** (1988) 5876–5886.
69. Neumann, M., Steinhauser, O.: On the calculation of the frequency-dependent dielectric constant in computer simulations. Chem. Phys. Lett. **102** (1983) 508–513.
70. Neumann, M., Steinhauser, O., Pawley, G.S.: Consistent calculations of the static and frequency-dependent dielectric constant in computer simulations. Mol. Phys. **52** (1984) 97–113.

71. Appel, A.W.: SIAM J. Sci. Stat. Comput. **6** (1985) 85.
72. Barnes, J., Hut, P.: A hierarchical O(N log N) force calculation algorithm. Nature **324** (1986) 446–449.
73. Greengard, L., Rokhlin, V.J.: A fast algorithm for particle simulations. J. Comput. Phys. **73** (1987) 325–348.
74. Niedermeier, C, Tavan, P.: A structure-adapted multipole method for electrostatic interactions in protein dynamics. J. chem. Phys. **101** (1994) 734–748.
75. Nelson, M., Humphrey, W., Gursoy, A., Dalke, A., Kalé, L., Skeel, R.D., Schulten, K.: NAMD – A parallel, object-oriented molecular dynamics program. Int. J. Supercomputing Applications and High Performance Computing **10** (1996) 251–268.
76. Esselink, K.: A comparison of algorithms for long-range interactions. Comput. Phys. Comm. **87** (1995) 375–395.
77. Luty, B.A., Davis, M.E., Tironi, I.G., Van Gunsteren, W.F.: A comparison of particle-particle particle-mesh and Ewald methods for calculating interactions in periodic molecular systems. Mol. Simul. **14** (1994) 11–20.
78. Belhadj,M., Alper, H.A., Levy, R.M.: Molecular dynamics simulations of water with Ewald summation for the long-range electrostatic interactions. Chem. Phys. Lett. **179** (1991) 13–20.
79. Fincham, D.: Optimisation of the Ewald sum for large systems. Mol. Simul. **13** (1994) 1–9.
80. Heyes, D.M.: Electrostatic potentials and fields in infinite point charge lattices. J. Chem. Phys. **74** (1981) 1924–1929.
81. Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., Pedersen, L.G.: A smooth particle mesh Ewald method. J. Chem. Phys. **103** (1995) 8577–8593.
82. Press, W.H.,Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.': Numerical Recipes. The Art of Scientific Computing. Cambridge Univ. Press, 2nd ed. (1992)
83. Beckers, J. (Technical University, Delft): private communication.
84. Berendsen, H.J.C.: Bio-molecular dynamics comes of age. Science **271** (1996) 954.
85. Tieleman, D.P., Berendsen, H.J.C.: A molecular dynamics study of the pores formed by E. coli OmpF porin in a fully hydrated POPE bilayer. Biophys. J., in print (1998).
86. Van Schaik, R.C., Berendsen, H.J.C., Torda, A.E., Van Gunsteren, W.F.: A structure refinement method based on molecular dynamics in 4 spatial dimensions. J. Mol. Biol. **234** (1993) 751–762.
87. Field, M.: The simulation of chemical reactions, in: *Computer Simulations of Biomolecular Systems, Vol 2*, W.F. van Gunsteren, P.K. Weiner, A.J. Wilkinson, eds. Escom, Leiden (1993) 82–123.
88. Merz, K.M. Jr: Computer simulation of enzymatic reactions. Curr. Opinion Struct. Biol. **3** (1993) 234–240.
89. Åqvist, J., Warshel, A.: Simulation of enzyme reactions using valence bond force fields and other hybrid quantum/classical approaches. Chem. Rev. **93** (1993) 2523–2544.
90. Bakowies, D., Thiel, W.: Hybrid models for combined quantum mechanical and molecular mechanical approaches. J. Phys. Chem. **100** (1996) 10580–10594.
91. Handy, N.C.: Density functional theory. In *Quantum mechanical simulation methods for studying biological systems*, D. Bicout and M. Field, eds, Springer, Berlin (1996) 1–35.

92. Stanton, R.V., Hartsough, D.S., Merz, K.M. Jr: An Examination of a Density Functional-Molecular Mechanical Coupled Potential. J. Comput. Chem. **16** (1995) 113–128.
93. Meier, R.J., Van Doremaele, H.J., Iarlori, S., Buda, F.: Ab-initio molecular dynamics study of metallocene-catalysed ethylene polymerization. J. Amer. Chem. Soc. **116** (1994) 7274–7281.
94. Kramers, H.A.: Brownian motion in a field of force and the diffusion model of chemical reactions. Physica **VII** (1940) 284–304
95. Chandler, D.: Statistical mechanics of isomerization dynamics in liquids and the transition state approximation. J. Chem. Phys. **68** (1978) 2959–2970. See also Chandler, D.: J. Stat. Phys. **42** (1986) 49.
96. Hänggi, P., Talkner, P., Borkovec, M.: Reaction-rate theory: fifty years after Kramers. Revs Modern Phys. **62** (1990) 251–341.
97. Devault, D.: Quantum mechanical tunnelling in biological systems. Quart. Rev. Biophys. **13** (1980) 387–564.
98. Moser, C.C., Dutton, P.L.: Biological electron transfer: measurement, mechanism, engineering requirements. In *Quantum mechanical simulation methods for studying biological systems*, D. Bicout and M. Field, eds, Springer, Berlin (1996) 201–214.
99. Schulten, K.: Curve crossing in a protein: coupling of the elementary quantum process to motions of the protein. In *Quantum mechanical simulation methods for studying biological systems*, D. Bicout and M. Field, eds, Springer, Berlin (1996) 85–118.
100. Rossky, P.J., Schnitker, J.: The hydrated electron: quantum simulation of structure, spectroscopy and dynamics. J. Phys. Chem. **92** (1988) 4277–4285.
101. Gerber, R.B., Buch, V., Ratner, M.A.: Time-dependent self-consistent field approximation for intramolecular energy transfer. I. Formulation and application to dissociation of van der Waals molecules. J. Chem. Phys. **77** (1982) 3022–3030.
102. Berendsen, H.J.C., Mavri, J.: Simulating proton transfer processes: Quantum dynamics embedded in a classical environment. In *Theoretical Treatments of Hydrogen Bonding*, D. Hadži, ed., Wiley, New York (1997) 119–141.
103. Drukker, K., Hammes-Schiffer, S.: An analytical derivation of MC-SCF vibrational wave functions for the quantum dynamical simulation of multiple proton transfer reactions: Initial application to protonated water chains. J. Chem. Phys. **107** (1997) 363–374.
104. Bala, P., Lesyng, B., McCammon, J.A.: Extended Hellmann-Feynman theorem for non-stationary states and its application in quantum-classical molecular dynamics simulations. Chem. Phys. Lett. **219** (1994) 259–266.
105. Berendsen, H.J.C., Mavri, J.: Quantum simulation of reaction dynamics by Density Matrix Evolution. J. Phys. Chem. **97** (1993) 13464–13468.
106. Berendsen, H.J.C., Mavri, J.: Quantum dynamics simulation of a small quantum system embedded in a classical environment, In *Quantum mechanical simulation methods for studying biological systems*, D. Bicout and M. Field, eds, Springer, Berlin (1996) 157–179.
107. Ehrenfest, P.: Bemerkung über die angenäherte Gültigkeit der klassischen Mechanik innerhalb der Quantenmechanik. Z. für Physik **45** (1927) 455–457.
108. Bornemann, F.A., Nettesheim, P., Schütte, C.: Quantum-classical molecular dynamics as an approximation to full quantum dynamics. J. Chem. Phys. **105** (1996) 1074–1083.

109. Bala, P., Lesyng, B., McCammon, J.A.: Application of quantum-classical and quantum-stochastic molecular dynamics simulations for proton transfer processes. Chem. Phys. **180** (1994) 271–285.
110. Bala, P., Grochowsky, P., Lesyng, B., McCammon, J.A.: Quantum-classical molecular dynamics. Models and applications. In *Quantum mechanical simulation methods for studying biological systems*, D. Bicout and M. Field, eds, Springer, Berlin (1996) 119–156.
111. Mavri, J., Berendsen, H.J.C., Van Gunsteren, W.F.: Influence of solvent on intramolecular proton transfer in hydrogen malonate. Molecular dynamics study of tunneling by density matrix evolution and nonequilibrium solvation. J. Phys. Chem. **97** (1993) 13469–13476.
112. Mavri, J., Berendsen, H.J.C.: Dynamical simulation of a quantum harmonic oscillator in a noble-gas bath by density matrix evolution. Phys. Rev. **E 50** (1994) 198–204.
113. Mavri, J., Lensink, M., Berendsen, H.J.C.: Treatment of inelastic collisions of a particle with a quantum harmonic oscillator by density matrix evolution. Mol. Phys. **82** (1994) 1249–1257.
114. Mavri, J., Berendsen, H.J.C.: Calculation of the proton transfer rate using density matrix evolution and molecular dynamics simulations: Inclusion of the proton excited states. J. Phys. Chem. **99** (1995) 12711-12717.
115. Berendsen, H.J.C., Mavri, J.: Approach to non-adiabatic transitions by density matrix evolution and molecular dynamics simulation. Int. J. Quant. Chem. **57** (1996) 975–984.
116. Ernst, R.R., Bodenhausen, G., Wokaun, A.: Principles of Nuclear Magnetic Resonance in One and Two Dimensions. Clarendon Press, Oxford (1987).
117. Pechukas, P.: Time-dependent semiclassical scattering theory. II Atomic collisions. Phys. Rev. **181** (1969) 174–185.
118. Tully, J.C.: Molecular dynamics with electronic transitions. J. Chem. Phys. **93** (1990) 1061–1071.
119. Hammes-Schiffer, S., Tully, J.C.: Proton transfer in solution: Molecular dynamics with quantum transitions. J. Chem. Phys. **101** (1994) 4657–4667.
120. Hammes-Schiffer, S., Tully, J.C.: Vibrationally Enhanced Proton Transfer. J. Phys. Chem. **99** (1995) 5793–5797.
121. Hammes-Schiffer, S., Tully, J.C.: Nonadiabatic transition state theory and multiple potential energy surface molecular dynamics of infrequent events. J. Chem. Phys. **103** (1995) 8528–8537.
122. Hammes-Schiffer: Multiconfigurational molecular dynamics with quantum transitions: Multiple proton transfer reactions. J. Chem. Phys. **105** (1996) 2236–2246.
123. Cao, J., Voth, G.A.: The formulation of quantum statistical mechanics based on the Feynman path centroid density. I. Equilibrium properties. J. Chem. Phys. **100** (1994) 5093-5105; II Dynamical properties. J. Chem. Phys. **100** (1994) 5106–5117; III. Phase space formalism and nalysis of centroid molecular dynamics. J. Chem. Phys. **101** (1994) 6157–6167; IV. Algorithms for centroid molecular dynamics. J. Chem. Phys. **101** (1994) 6168–6183; V. Quantum instantaneous normal mode theory of liquids. J. Chem. Phys. **101** (1994) 6184–6192.
124. Cao, J., Voth, G.A.: A unified framework for quantum activated rate processes. I. General theory. J. Chem. Phys. **105** (1996) 6856–6870.

125. Martyna, G.J.: Adiabatic path integral molecular dynamics methods. I. Theory. J. Chem. Phys. **104** (1996) 2018–2027.
126. Cao, J., Martyna, G.J.: Adiabatic path integral molecular dynamics methods. II. Algorithms. J. Chem. Phys. **104** (1996) 2028–2035.
127. Lobaugh, J., Voth, G.A.: A path integral study of electronic polarization and nonlinear coupling effects in condensed phase proton transfer reactions. J.Chem. Phys. **100** (1994) 3039–3047.
128. Lobaugh, J., Voth, G.A.: The quantum dynamics of an excess proton in water. J.Chem. Phys. **104** (1996) 2056–2069.
129. Van der Spoel,D., Berendsen, H.J.C.: Determination of proton transfer rate constants using ab initio, molecular dynamics and density matrix evolution calculations. Pacific Symposium on Biocomputing, World Scientific, Singapore (1996) 1–14.
130. Mavri, J., Van der Spoel, D., Berendsen, H.J.C.: The rate of proton transfer in the active site of HIV-1 protease. Submitted, 1998.
131. Lensink, M., Mavri, J., Berendsen, H.J.C.: Simulation of a slow reaction with quantum character: Neutral hydrolysis of a carboxylic ester. Submitted (1998).
132. Marrink, S.-J., Berendsen, H.J.C.: Simulation of water transport through a lipid membrane. J. Phys. Chem. **98** (1994) 4155–4168.
133. Permeation process of small molecules across lipid membranes studied by molecular dynamics simulations. J. Phys. Chem. **100** (1996) 16729–16738.
134. Marrink, S.J., Jähnig, F., Berendsen, H.J.C.: Proton transport across transient single-file water pores in a lipid membrane studied by molecular dynamics simulations. Biophys. J. **71** (1996) 632–647.
135. Amadei, A., Linssen, A.B.M., Berendsen, H.J.C.: Essential Dynamics of Proteins. Proteins **17** (1993) 412–425.
136. Van Aalten, D.M.F., Amadei, A., Linssen, A.B.M., Eijsink, V.G.H., Vriend, G., Berendsen, H.J.C.: The essential dynamics of thermolysin: Confirmation of the hinge-bending motion and comparison of simulations in vacuum and water. Proteins **22** (1995) 45–54.
137. Amadei, A., Linssen, A.B.M., De Groot, B.L., Berendsen, H.J.C.: Essential degrees of freedom of proteins. In *Modelling of biomolecular structures and mechanisms*, A. Pullman *et al.*, eds, Kluwer (1995) 85–93.
138. Van Aalten, D.M.F., Findlay, J.B.C., Amadei, A., Berendsen,H.J.C.: Essential dynamics of the cellular retinol-binding protein. Evidence for ligand-induced conformational changes. Protein Engin. **8** (1995) 1129–1136.
139. Amadei, A., Linssen, A.B.M., De Groot, B.L., Van Aalten, D.M.F., Berendsen, H.J.C.: An efficient method for sampling the essential subspace of proteins. J. Biomol. Struct. Dyn. **13** (1996) 615–626.
140. De Groot, B.L., Amadei, A., Van Aalten, D.M.F., Berendsen, H.J.C.: Towards an exhaustive sampling of the configurational space of the two forms of the peptide hormone guanylin. J. Biomol. Struct. Dyn. **13** (1996) 741–751.
141. Creveld, L., Amadei, A., Van Schaik, C., Pepermans, R., De Vlieg, J., Berendsen, H.J.C.: Identification of functional and unfolding motions of cutinase as obtained from molecular dynamics computer simulations. Submitted (1998).
142. Hayward, S., Kitao, A., Berendsen, H.J.C.: Model-free methods to analyze domain motions in proteins from simulation: A comparison of normal mode analysis and molecular dynamics simulation of lysozyme. Proteins **27** (1997) 425–437.

143. Hayward, S., Berendsen, H.J.C.: Systematic analysis of domain motions in proteins from conformational change: New results on citrate synthase and T4 lysozyme. Proteins **30** (1998) 144–154.
144. Henderson, D.: Fundamentals of Inhomogeneous Fluids. Marcel Dekker, New York (1992).
145. Fraaije, J.G.E.M.: Dynamic density functional theory for microphase separation kinetics of block copolymer melts. J. Chem. Phys. **99** (1993) 9202–9212.
146. Van Vlimmeren, B.A.C., Fraaije, J.G.E.M.: Calculation of noise distribution in mesoscopic dynamics models for phase-separation of multicomponent complex fluids. Comput. Phys. Comm. **99** (1996) 21–28.
147. Maurits, N.M., Altevogt, P., Evers, O.A., Fraaije, J.G.E.M.: Simple numerical quadrature rules for Gaussian Chain polymer density functional calculations in 3D and implementation on parallel platforms. Comput. Theor. Polymer Sci. **6** (1996) 1–8.
148. Maurits, N.M., Fraaije, J.G.E.M.: Mesoscopic dynamics of copolymer melts: from density dynamics to external potential dynamics using nonlocal kinetic coupling. J. Chem. Phys. **107** (1997) 5879–5889.
149. Fraaije, J.G.E.M., Van Vlimmeren, B.A.C., Maurits, N.M., Postma, M., Evers, O.A., Hoffmann, C., Altevogt, P., Goldbeck-Wood, G.: The dynamic mean-field density functional method and its application to the mesoscopic dynamics of quenched block copolymer melts. J. Chem. Phys. **106** (1997) 4260–4269.

Part I

Conformational Dynamics

Steered Molecular Dynamics

Sergei Izrailev, Sergey Stepaniants, Barry Isralewitz, Dorina Kosztin, Hui Lu, Ferenc Molnar, Willy Wriggers, and Klaus Schulten

Beckman Institute, University of Illinois at Urbana-Champaign, 405 N. Mathews, Urbana, IL 61801, USA

Abstract. Steered molecular dynamics (SMD) induces unbinding of ligands and conformational changes in biomolecules on time scales accessible to molecular dynamics simulations. Time-dependent external forces are applied to a system, and the responses of the system are analyzed. SMD has already provided important qualitative insights into biologically relevant problems, as demonstrated here for applications ranging from identification of ligand binding pathways to explanation of elastic properties of proteins. First attempts to deduce potentials of mean force by discounting irreversible work performed on the system are summarized. The non-equilibrium statistical mechanics underlying analysis of SMD data is outlined.

1 Introduction

Molecular recognition and specific ligand-receptor interactions are central to many biochemical processes. The regulation of cellular signal-transduction pathways and gene expression, activity of enzymes, cell motility, molecular immunology and the action of hormones involve the triggering of functional responses by noncovalent associations of ligands with receptors. The prediction and design of ligands (inhibitors or substrates) for a given receptor is the main goal in rational drug design, and considerable effort is spent on the development of corresponding computational methods (Cohen et al., 1990; Colman, 1994; Marrone et al., 1997). New pharmaceuticals, e.g., the HIV protease inhibitors (Thaisrivongs et al., 1996; Lebon et al., 1996; Hanessian and Devasthale, 1996), derived in part from such methods, have made a major impact on clinical medicine, and computational modeling will be of increasing importance in the future.

Despite an abundance of modeling methods for ligand-receptor interactions and protein-protein docking (Strynadka et al., 1996) little is known about processes governed by adhesive interactions such as those occuring in the binding and unbinding of ligands. Presently, the prevailing point of view concerning computer simulations describing ligand binding and determining binding affinities is to strive for the ideal of *reversibility*, as in umbrella sampling and free energy perturbation (McCammon and Harvey, 1987; Ajay and Murcko, 1995; Gilson et al., 1997), with the hope that artifacts induced by the finite rate of conformational changes can be neglected. Reaching this ideal, however, requires extremely slow manipulation and, therefore, prohibitively expensive simulations. This chapter advocates a new computational method,

steered molecular dynamics (SMD), which accepts *irreversibility*, ceding for the present time accurate evaluation of binding affinities and potentials of mean force, but gaining access to biologically relevant information related to non-covalent bonding. We will demonstrate the wealth of such information using a broad range of examples. The concern that thermodynamic potentials cannot, even in principle, be obtained from irreversible processes has been proven unfounded by the remarkable identity derived by Jarzynski (1997a,b), $\langle \exp[-W/k_BT] \rangle = \exp[-\Delta F/k_BT]$. This identity connects the ensemble average of an exponential of the total work W performed on the system during a non-equilibrium transition from one state to another to the free energy difference ΔF between the two states.

Experimental techniques based on the application of mechanical forces to single molecules in small assemblies have been applied to study the binding properties of biomolecules and their response to external mechanical manipulations. Among such techniques are atomic force microscopy (AFM), optical tweezers, biomembrane force probe, and surface force apparatus experiments (Binning et al., 1986; Block and Svoboda, 1994; Evans et al., 1995; Israelachvili, 1992). These techniques have inspired us and others (see also the chapters by Eichinger et al. and by Hermans et al. in this volume) to adopt a similar approach for the study of biomolecules by means of computer simulations.

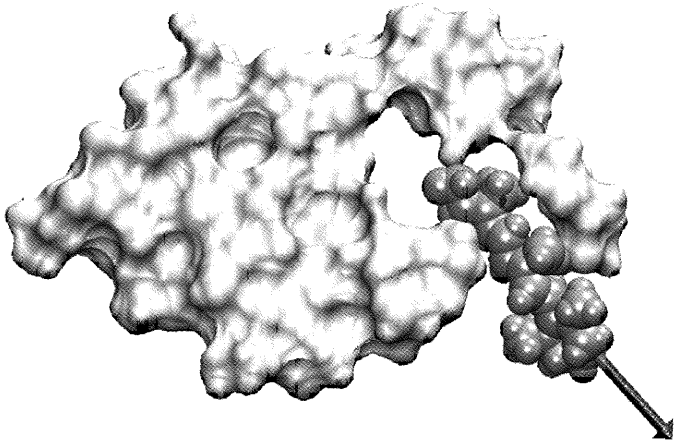


Fig. 1.
Extraction of a ligand from the binding pocket of a protein. The force (represented by an arrow) applied to the ligand (shown in van der Waals spheres) leads to its dissociation from the binding pocket of the protein (a slice of the protein represented as a molecular surface is shown).

In SMD simulations time-dependent external forces are applied, for example, to a ligand to facilitate its unbinding from a protein, as shown in Fig. 1. The analysis of the interactions of the dissociating ligand with the binding pocket, as well as the recording (as a function of time) of applied forces and ligand position, yields important structural information about the structure-function relationships of the ligand-receptor complex, binding pathways, and mechanisms underlying the selectivity of enzymes. SMD can also be applied to investigate the molecular mechanisms that determine elastic properties exhibited by proteins subjected to deformations in AFM and optical tweezer experiments, such as stretching of titin leading to unfolding of its immunoglobulin domains (Rief et al., 1997), or stretching of tenascin which results in unfolding of its fibronectin-III domains (Oberhauser et al., 1998).

Besides yielding *qualitative* information, these biologically and pharmaceutically motivated applications of SMD can also yield *quantitative* information about the binding potential of the ligand-receptor complex. A first advance in the reconstruction of the thermodynamic potential from SMD data by discounting irreversible work was made by Balsera et al. (1997) as outlined in Sect. "Reconstruction of the potential of mean force" below.

In the following we describe the methodology of SMD, illustrate applications of SMD through key examples, present the non-equilibrium statistical mechanical theory of SMD, and describe a method of reconstruction of a potential of mean force from SMD data. The applications include studies of the dissociation of biotin from avidin and streptavidin (Izrailev et al., 1997; Grubmüller et al., 1996) (see also the chapter by Eichinger et al. in this volume), the unbinding of retinal from bacteriorhodopsin (Isralewitz et al., 1997), the release of phosphate from actin (Wriggers and Schulten, 1998), the possible binding pathways of thyroid hormone to its receptor, the extraction of lipids from membranes (Stepaniants et al., 1997; Marrink et al., 1998), the unbinding of arachidonic acid from the prostaglandin H_2 synthase-1, and the force-induced unfolding of titin immunoglobulin domains (Lu et al., 1998). In the chapter by Hermans et al. in this volume another application of SMD, the extraction of bound xenon from mutant T4-lysozyme, is discussed.

2 Methods

One way to apply external forces to a protein-ligand complex is to restrain the ligand to a point in space (restraint point) by an external, e.g., harmonic, potential. The restraint point is then shifted in a chosen direction (Grubmüller et al., 1996; Isralewitz et al., 1997; Stepaniants et al., 1997; Marrink et al., 1998; Lu et al., 1998), forcing the ligand to move from its initial position in the protein and allowing the ligand to explore new contacts along its unbinding path. Assuming a single reaction coordinate x, and an external potential $U = K(x - x_0)^2/2$, where K is the stiffness of the restraint, and x_0 is the initial position of the restraint point moving with a constant velocity v, the

external force exerted on the system can be expressed as

$$F = K(x_0 + vt - x). \tag{1}$$

This force corresponds to the ligand being pulled by a harmonic spring of stiffness K with its end moving with velocity v. Alternatively, a fixed restraint point at a distance much larger than the length of the unbinding pathway may be chosen. In this case, the end of the spring does not move and its stiffness is linearly increased with time (Izrailev et al., 1997), i.e., $K = \alpha t$, and the force is

$$F = \alpha t(x_0 - x). \tag{2}$$

Other external forces or potentials can also be used, e.g., constant forces, or torques applied to parts of a protein to induce rotational motion of its domains (Wriggers and Schulten, 1997a).

SMD simulations require selection of a path, i.e., a series of directions of the applied force. In some cases a straight line path is sufficient, e.g., for avidin-biotin (Fig. 2), actin (Fig. 4), lipids in membranes (Fig. 6), or the unfolding of titin immunoglobulin domains (Fig. 8). Other biomolecular systems involve a ligand positioned at the bottom of a convoluted binding cleft, e.g., bacteriorhodopsin (Fig. 3), prostaglandin H_2 synthase (Fig. 7), and nuclear hormone receptors (Fig. 5). In the latter cases the forced unbinding of the ligand requires the direction of the force to be changed during the simulation to avoid distortion of the surrounding protein. The direction of the force can be chosen randomly (Lüdemann et al., 1997) or by guessing a direction on the basis of structural information. A force is then applied to the ligand in the chosen direction, and this direction is accepted or rejected based on factors such as conservation of secondary structure of the protein, deformation of the protein, the magnitude of the force applied, the average velocity of the ligand along the unbinding pathway, etc. (Isralewitz et al., 1997; Lüdemann et al., 1997). One possible protocol for selecting force directions in SMD defines a conical region of space around a preferred direction and selects new directions randomly within this region. A small cone angle strongly biases the chosen directions to the initial guess, whereas a large cone angle leads to exploration of more directions.

An initial and desired final configuration of a system can be used by the targeted molecular dynamics (TMD) method (Schlitter et al., 1993) to establish a suitable pathway between the given configurations. The resulting pathway can then be employed during further SMD simulations for choosing the direction of the applied force. TMD imposes time-dependent holonomic constraints which drive the system from one known state to another. This method is also discussed in the chapter by Helms and McCammon in this volume.

Other methods for identifying multi-dimensional reaction paths are based on stochastic dynamics. For example, a reaction path can be found by opti-

mization of the Onsager-Machlup action between the two end points of a trajectory (Olender and Elber, 1996) (see the chapter by Elber et al. in this volume). Alternatively, using the conformational flooding method (Grubmüller, 1995), one may sample the distribution of ligand conformations through principle component analysis and use forces derived from this analysis to drive the ligand away from the current distribution, as discussed in the chapter by Eichinger et al. in this volume.

3 Applications of SMD

The ultimate criterion for the value of a method such as SMD is how much can be learned from using it. In this section we provide examples of SMD applications yielding important insights into biological processes. First, we review the study of the biotin-avidin complex which served as a test bed for the method, then discuss three examples in which SMD identified binding pathways of ligands. Next, we demonstrate how SMD elucidated two key steps in fatty acid metabolism, namely, the extraction of lipids from membranes by phospholipase A_2 and the binding of arachidonic acid by prostaglandin H_2 synthase. Finally, we show how SMD revealed the mechanism behind the stretching of titin immunoglobulin domains.

3.1 Avidin-Biotin Complex as a Test Bed for SMD

The avidin-biotin complex, known for its extremely high affinity (Green, 1975), has been studied experimentally more extensively than most other protein-ligand systems. The adhesion forces between avidin and biotin have been measured directly by AFM experiments (Florin et al., 1994; Moy et al., 1994b; Moy et al., 1994a). SMD simulations were performed on the entire tetramer of avidin with four biotins bound to investigate the microscopic detail of unbinding of biotin from avidin (Izrailev et al., 1997).

In the simulations the rupture of biotin from avidin was induced by means of a soft harmonic restraint, as described by Eq. (2) with $K = \alpha t$ ranging from 0 to 120 pN/Å. The spatial range of thermal fluctuations of biotin associated with the restraint was $\delta x \simeq 3$ Å, i.e., on the order of the size of the binding pocket (about 10 Å). The fluctuations of the applied force, on the other hand, were small compared to its absolute value, and the force profiles exhibited nearly linear growth with time, similar to that reported in AFM experiments (see Fig. 2). The values of the rupture forces, i.e., the maximum measured force (450–800 pN), exceeded those measured in AFM experiments (160 pN). These SMD simulations did not exhibit any particular scaling of the rupture force with the pulling rate, which covered a span of almost two orders of magnitude. In SMD simulations of a similar streptavidin-biotin complex reported by Grubmüller et al. (1996) (see the chapter by Eichinger et al. in this volume), the scheme of Eq. (1) with a stiff spring ($K = 280$ pN/Å) was

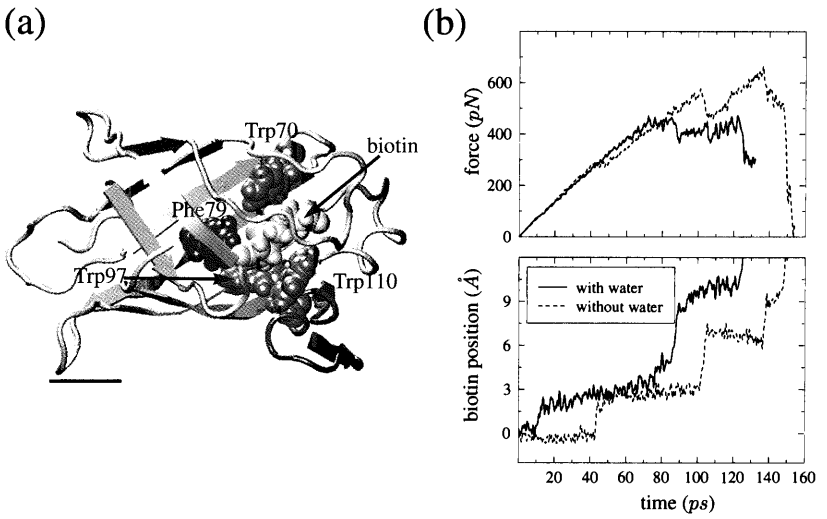


Fig. 2.
(a) Hydrophobic residues in the binding pocket of the avidin-biotin tetrameric complex (only one monomer shown): Phe79, Trp70, and Trp97, as well as Trp110 from the adjacent monomer, surround biotin tightly on all sides making the binding pocket impenetrable to water. (b) Biotin displacement and applied forces during the dissociation of the avidin-biotin complex with water molecules placed with the program DOWSER, and without water in the vicinity of the binding pocket.

employed, and the rupture force was found to scale linearly with the velocity v (cf. Sect. "Stochastic Modeling of SMD").

The simulations also revealed that flapping motions of one of the loops of the avidin monomer play a crucial role in the mechanism of the unbinding of biotin. The fluctuation time for this loop as well as the relaxation time for many of the processes in proteins can be on the order of microseconds and longer (Eaton et al., 1997). The loop has enough time to fluctuate into an open state on experimental time scales (1 ms), but the fluctuation time is too long for this event to take place on the nanosecond time scale of simulations. To facilitate the exit of biotin from its binding pocket, the conformation of this loop was altered (Izrailev et al., 1997) using the interactive molecular dynamics features of MDScope (Nelson et al., 1995; Nelson et al., 1996; Humphrey et al., 1996).

During unbinding, biotin was found to move in discrete steps (see Fig. 2). Each step can be identified with the formation and rupture of a network of hydrogen bonds which stabilize biotin in the binding pocket of avidin. The strongest bonds were formed between biotin and polar residues Tyr33, Ser16 and Thr35, consistent with experimental observations. Contacts of biotin with

nonpolar residues (see Fig. 2), especially with Trp110 of an adjacent avidin monomer (in the complete tetramer), are crucial for the unbinding process (Izrailev et al., 1997). These residues prevent water molecules from entering the binding pocket. To determine the effect of water molecules on the unbinding mechanism, 50 water molecules were placed in the avidin tetramer with the program DOWSER (Zhang and Hermans, 1996) (the algorithm for placing water in proteins is discussed in the chapter by Hermans et al. in this volume). The presence of four water molecules in the outer region of the binding pocket, i.e., close to biotin's valeryl side-chain carboxylate group, did not affect the stepwise motion of biotin, but reduced the rupture (maximum) force from 600 pN to 400 pN as shown in Fig. 2. The reduction of the rupture force resulted from the participation of water molecules in breaking the hydrogen bond networks between biotin and residues located near the exit of the binding pocket. Water did not penetrate the binding pocket on the time scale of the simulations.

3.2 Binding of Retinal to Bacterio-opsin

Bacteriorhodopsin (bR) (Oesterhelt et al., 1992; Schulten et al., 1995) is a light-driven vectorial proton pump found in the membrane of *Halobacterium salinarum*. The protein binds a retinal molecule through a Schiff base linkage to its Lys216 side group. Formation of bR from the apoprotein and retinal has been studied experimentally (Oesterhelt and Schumann, 1974; Chang et al., 1988; Booth et al., 1996), but the pathway of initial retinal entry during bR formation was not clearly understood. Despite its extremely poor solubility in water and a considerable affinity for lipid environment, retinal was generally believed to enter the protein through the solvent-exposed loops of the protein. However, a window on the lipid-exposed surface of bR located between helices E and F (see Fig. 3) which uncovers part of retinal (its β-ionone ring) can be an entry point for retinal. Inspection of the bR structure revealed that this window, in fact, is the only opening large enough to allow retinal entry and provide access to the Lys216 binding site of retinal. SMD simulations were performed to test this hypothesis by extracting retinal from bR with an external force along a path towards and out of the putative entry window (Isralewitz et al., 1997). If such an extraction could be carried out without significantly perturbing the protein on the time scale of MD simulations, then the extraction path could also constitute the binding path of retinal to the apoprotein on the much longer time scale of bR formation.

Due to the convoluted shape of the retinal binding site, retinal cannot be extracted from bR by application of a force along a single straight line. Therefore, the unbinding path was segmented, with the direction of the force determined anew for each of the ten segments (cf. Sect. "Methods"). For each segment the applied forces were described by Eq. (1) with $K = 10$ $k_BT/\text{Å}^2 \simeq 414$ pN/Å and $v = 0.125$ Å/ps at $T = 300$ K.

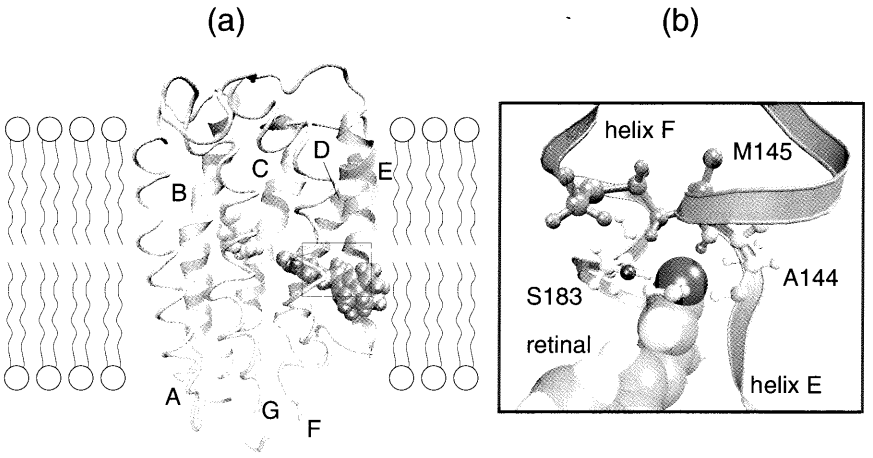


Fig. 3.
(a) Extraction of retinal from bacteriorhodopsin (bR). The backbone of bR is shown in tube and ribbon representation; the seven trans-membrane helices of bR are labeled A–G. The structure presents a snapshot of the SMD simulation with retinal (dark spheres) partially extracted from the protein bR through an opening between helices E and F into the hydrophobic phase of the surrounding membrane. (b) Magnification of boxed area in (a), at completion of retinal's exit from bR, representing the putative exit/entry window of retinal. The amino acids shown, Ala144, Met145, and Ser183, form a hydrogen bond network with retinal's carbonyl group attracting and guiding retinal into the protein.

It was found that extraction of retinal from bR along a path through the window between helices E and F (see Fig. 3) could be accomplished during a 0.2 ns simulation without disrupting the protein structure (Israelewitz et al., 1997). The maximum force applied was about 1000 pN and accounted for breaking of a strong hydrogen bond between retinal and Lys216.

Upon exit from the interior of bR, retinal formed a stable network of hydrogen bonds with residues Ala144, Met145 and Ser183 which line the putative exit/entry window (see Fig. 3). This suggests that retinal approaches the apoprotein from the hydrophobic phase of the membrane, binds to the stated residues, subsequently moves into bR forming a hydrogen bond with Lys216 and, finally, forms the Schiff base bond.

3.3 Actin's Back Door

Actin filaments are dynamic polymers whose assembly and disassembly in the cell cytoplasm drives shape changes (Small, 1989), cell locomotion (Theriot et al., 1992), and chemotactic migration (Theriot et al., 1992)Devreotes and

Zigmond, 1988). The ATP-hydrolysis that accompanies actin polymerization, ATP $\rightarrow$ ADP + P_i, and the subsequent release of the cleaved phosphate (P_i) are believed to act as a clock (Pollard et al., 1992; Allen et al., 1996), altering in a time-dependent manner the mechanical properties of the filament and its propensity to depolymerize. Molecular dynamics simulations suggested a so-called back door mechanism for the hydrolysis reaction ATP $\rightarrow$ ADP + P_i in which ATP enters the actin from one side, ADP leaves from the same side, but P_i leaves from the opposite side, the "back door" (Wriggers and Schulten, 1997b). This hypothesis can explain the effect of the toxin phalloidin which blocks the exit of the putative back door pathway and, thereby, delays P_i release as observed experimentally (Dancker and Hess, 1990).

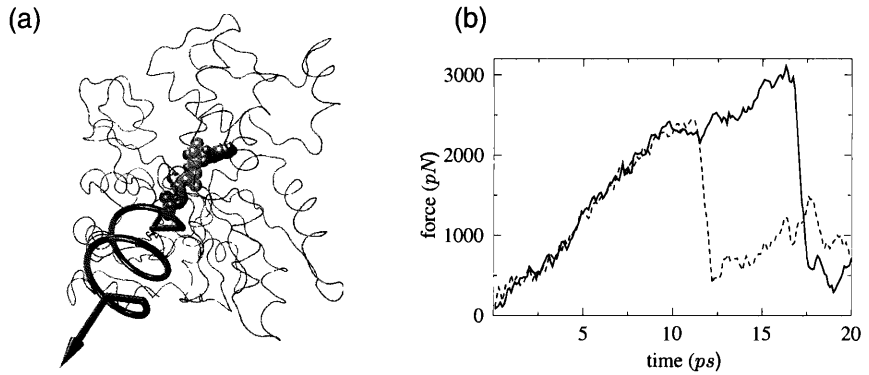


Fig. 4.
Phosphate release from actin. (a) Monomeric actin with ADP and P_i bound. The protein backbone (tube), ADP (grey spheres), and Ca^{2+}-P_i (black spheres) are shown. The orientation of the spring indicates the pulling direction during P_i unbinding. (b) Force exerted on the deprotonated (solid line) and protonated (dashed line) phosphate during the SMD simulations.

To reveal the microscopic processes underlying the unbinding of P_i, SMD simulations were carried out in which P_i was pulled along the back door pathway and the adhesion forces were measured (Wriggers and Schulten, 1998). The simulations revealed that the dissociation of P_i is likely to be controlled by its protonation. P_i, which is singly protonated (HPO_4^{2-}) after cleavage from ADP, needs to overcome a strong Coulomb energy barrier due to the presence of a Ca^{2+} ion associated with ADP. The resulting forces reached 3,000 pN as shown in Fig. 4; in case of protonated P_i ($H_2PO_4^-$) the maximal forces measured 2,400 pN. This suggests that protonation of P_i is required for unbinding from actin, consistent with kinetic measurements (Allen et al., 1996).

The P_i-coordinating amino acid residues and solvent molecules in the dissociation pathway exhibited a remarkable functional diversity (Wriggers and Schulten, 1998). A methylated histidine, highly conserved among actin species and believed to be functionally relevant, stabilized bound P_i through a rotation of its side chain relative to the crystal structure. The long side chain of an arginine remained attached to P_i for the most part of the unbinding, guiding the ligand to the protein surface. Other P_i-coordinating side chains were replaced by water molecules in the solvated back door channel. This hydration of P_i gave rise to a velocity-dependent unbinding force that reflects the mobility of the water molecules relative to the displaced P_i. A hydration step during unbinding has also been observed in other SMD simulations, e.g., in case of the extraction of retinal from bR (Isralewitz et al., 1997).

3.4 Binding of Hormones to Nuclear Hormone Receptors

Hormone binding to the thyroid hormone receptor initiates a series of molecular events culminating in the activation or repression of transcription of target genes. The transition between the bound and unbound form of the thyroid receptor is accompanied by a conformational change that enables the hormone–receptor complex to bind to specific sequences of DNA and other transcriptional coactivators or repressors (Brent et al., 1989; Damm et al., 1989; Andersson et al., 1992). SMD can determine likely pathways of hormone binding and unbinding, reveal components of the receptor involved in the unbinding, and thus contribute to the design of new ligands for hormone therapy.

An examination of the crystal structure of the rat α_1 thyroid hormone receptor (TR) ligand binding domain bound with a thyroid hormone agonist (Wagner et al., 1995) suggests three entry/exit points for the hormone as shown in Fig. 5a. By applying an external force to the ligand to facilitate its unbinding from the protein, the three possible pathways were explored. In the simulations, the protein-ligand system was surrounded by a water bath. One atom of the hormone was harmonically restrained (K = 10 kcal/molÅ^2 $\simeq$ 695 pN/Å) to a point moving with a constant velocity v = 0.08 Å/ps in a chosen direction. The investigation is still ongoing and presently only preliminary conclusions can be drawn from the SMD data.

During the unbinding process the force exerted on the hormone varied according to the interaction of the hormone with surrounding protein residues. The hormone was found to leave the binding pocket along all three pathways, but exerted the least effect on the protein conformation when pulled along path 1. The carboxylate group of the hormone formed direct hydrogen bonds with the guanidium of Arg228 and the amino nitrogen of Ser277. The external force exerted on the hormone increased until the Arg228 residue was dislodged and the associated hydrogen bond was broken. These events corresponded to the maximum values of the force shown in Fig. 5b. It should also be mentioned that path 1 was lined with flexible amino acid side groups,

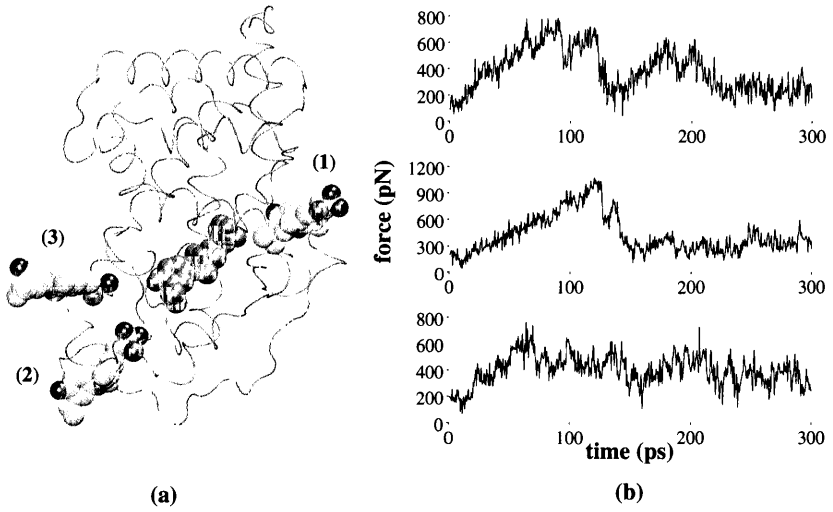


Fig. 5.
(a) Possible unbinding pathways for the dimit hormone from the thyroid hormone receptor: (1) the hormone leaves through the only discernible opening in the molecular surface; (2) the hormone moves underneath the last two helices of the protein (helices 11 and 12); (3) the hormone moves between helices 11 and 12. (b) Force profiles for the three unbinding pathways shown in (a) (top to bottom: path 1, path 2, path 3)

the flexibility of which was reflected by high experimentally observed temperature factors. Along paths 2 and 3, the hormone encountered Phe residues that needed to be moved out of the way. In both cases the force increased (see Fig. 5b) until the Phe residues changed their positions.

3.5 Extraction of Lipids from Membranes

SMD simulations were performed to investigate the extraction of a lipid from the dilauroyl-phosphatidyl-ethanolamin (DLPE) monolayer into the aqueous phase (see Fig. 6a) (Stepaniants et al., 1997). External forces described by Eq. (1) with $K = 10$ kcal/mol Å^2 $\simeq$ 695 pNÅ and $v = 0.014$ Å/ps were applied to the head group of the lipid, pulling it out from the membrane. The forces required to extract the lipid measured about 200 pN and remained constant within the range of fluctuations as shown in Fig. 6c. Analogous simulations were carried out by Marrink et al. (1998), extracting lipids from a dipalmitoyl-phosphatidyl-choline (DPPC) bilayer with pulling velocities of $v = 0.01$–0.5 Å/ps and resulting forces of 200-800 pN. In agreement with the results of SMD simulations of unbinding of biotin from streptavidin (Grubmüller et al., 1996) and avidin (Grubmüller et al., 1996)Izrailev et al.,

1997), performed with a range of pulling rates, the rupture force was found to decrease with decreasing v. The applied force as a function of distance increased up to the point of rupture, and then gradually decreased as the lipid proceeded into the solvent.

Although extraction of lipids from membranes can be induced in atomic force apparatus (Leckband et al., 1994) and biomembrane force probe (Evans et al., 1991) experiments, spontaneous dissociation of a lipid from a membrane occurs very rarely because it involves an energy barrier of about 20 kcal/mol (Cevc and Marsh, 1987). However, lipids are known to be extracted from membranes by various enzymes. One such enzyme is phospholipase A_2 (PLA_2), which complexes with membrane surfaces, destabilizes a phospholipid, extracts it from the membrane, and catalyzes the hydrolysis reaction of the *sn*-2-acyl chain of the lipid, producing lysophospholipids and fatty acids (Slotboom et al., 1982; Dennis, 1983; Jain et al., 1995). SMD simulations were employed to investigate the extraction of a lipid molecule from a DLPE monolayer by human synovial PLA_2 (see Fig. 6b), and to compare this process to the extraction of a lipid from a lipid monolayer into the aqueous phase (Stepaniants et al., 1997).

Due to the selection of a stiff restraint, the head group of the lipid was not allowed to fluctuate substantially and its motion essentially followed that of the restraint point. The forces measured during the extraction of the lipid exhibited large fluctuations on the order of 300 pN, as expected when a stiff restraint is employed (Izrailev et al., 1997; Balsera et al., 1997). The forces required to displace the lipid from the membrane into the binding pocket of PLA_2, shown in Fig. 6d, were larger than those required to displace the lipid from the membrane into the aqueous phase. This difference in the measured forces was due in part to the fact that the steric hindrance experienced by the lipid on its way out of the membrane into the active site of PLA_2 was larger than that for its movement into the aqueous phase; repositioning of PLA_2 could have reduced this hindrance. The results do not agree with the hypothesis of destabilization of the lipids by PLA_2 facilitating lipid extraction by the enzyme. The disagreement may have resulted from the steric effects mentioned above, an imperfect choice of the pulling direction for the lipid extraction into the enzyme, or insufficient sampling due to the short (500 ps) simulation time.

3.6 Binding of Arachidonic Acid to Prostaglandin H_2 Synthase-1

The enzyme prostaglandin H_2 synthase-1 (PGHS-1) catalyzes the transformation of the essential fatty acid, arachidonic acid (AA), to prostaglandin H_2 (Smith and DeWitt, 1996). This is the first committed step in the biosynthesis of prostanoids which modulate physiological processes such as platelet aggregation and inflammation. Aspirin, flurbiprofen, and other non-steroidal anti-inflammatory drugs directly target PGHS-1 by preventing the access of AA to its cyclooxygenase active site. This site involves a hydrophobic chan-

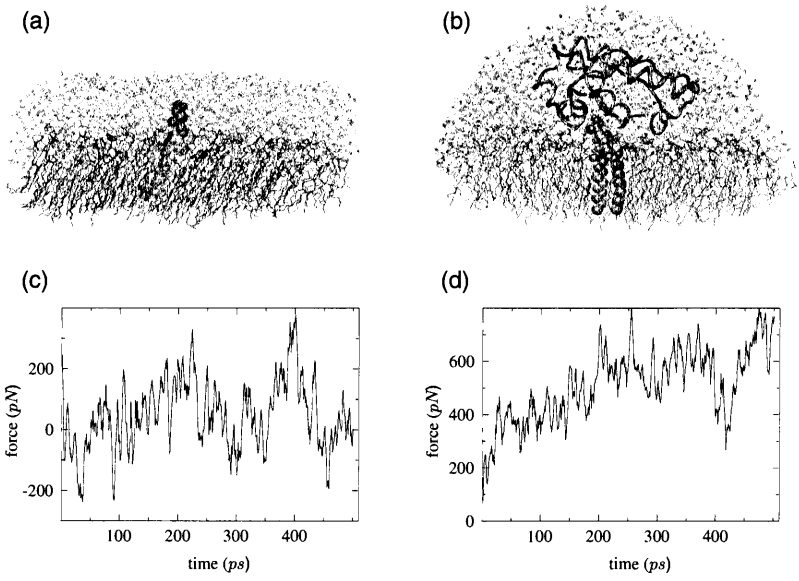


Fig. 6.
(a) Extraction of a lipid (black spheres) from the DLPE monolayer (lines) into the aqueous phase. (b) Extraction of a lipid (black spheres) from the DLPE monolayer (lines) into protein phospholipase A_2 (tube) solvated in water. (c) Force applied to the head of the lipid along the pulling direction during the extraction of the lipid into the aqueous phase. (d) Force applied to the head of the lipid along the pulling direction during the extraction of the lipid into the binding pocket of PLA_2.

nel, approximately 25 Å deep and 8 Å wide, which bends the fatty acid into a U-shape (Fig. 7). This shape is required for the catalyzed cyclooxygenation reaction.

Based on the crystal structure of PGHS-1, with flurbiprofen bound at the active site, a model for AA embedded in the enzyme was suggested, in which AA replaces the inhibitor (Picot et al., 1994). The aim of the investigation was to identify how AA folds itself into the required U-shape in the narrow binding channel, rather than entering the channel in a straight conformation. The simulations also sought to identify key residues guiding AA binding.

One monomer of the PGHS-1 homo-dimer (Picot et al., 1994), shown in Fig. 7 with AA bound in its putative cyclooxygenation site was used as a starting point for simulations enforcing the unbinding of AA. Pulling on the methyl end-group of the fatty acid with a harmonic spring as described by Eq. (1) and a range of force constants (K = 200–400 pN/Å) and pulling velocities (v = 10–0.1 Å/ps) led to the exit of the ligand from its narrow hydrophobic binding channel. During the forced unbinding a series of concerted torsional motions was observed (Fig. 7). AA contains four rigid *cis* double

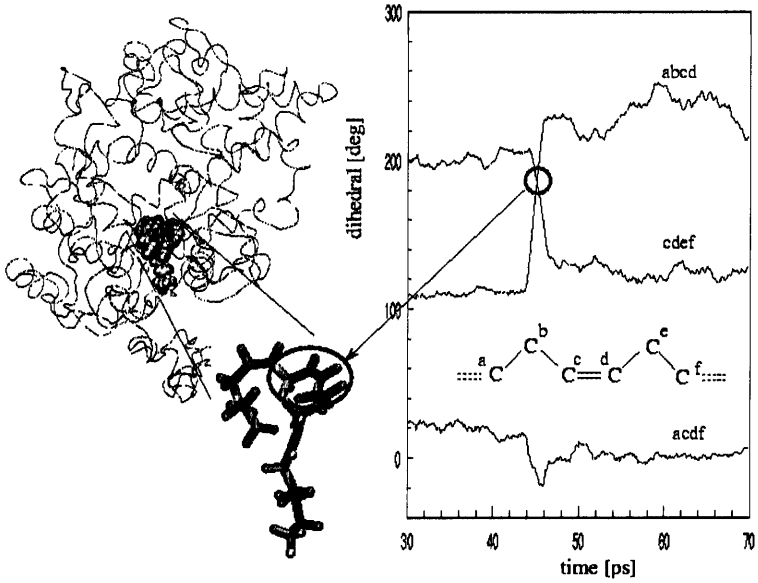


Fig. 7.
Binding site of arachidonic acid (AA) in prostaglandin H_2 synthase-1. An intermediate configuration observed during the unbinding event is depicted. AA contains four rigid *cis* double bonds connected to each other by a pair of conformationally flexible single bonds. The counter-rotation of atoms C^b and C^e around an axis defined by the $C^c = C^d$ bond was monitored through the change in dihedral angles defined by the C-atoms *abcd* and *cdef*. This rotation proceeded in a way which left the "backbone" of the ligand made of the conformationally rigid *cis* double bonds relatively unaffected (dihedral angle between C-atoms *acdf*).

bonds connected to each other by a pair of conformationally flexible single bonds. The unbinding mechanism can be described as a series of rotations around these single bonds which leave the "rigid backbone" of the fatty acid, formed by the conformationally inflexible *cis* double bonds, relatively unaffected (Fig. 7). This type of concerted motion was shown to be specific for the chemical structure of AA. If the all-*cis*-isomer was altered to an isomer with a *trans* double bond, the concerted motion became hindered and the unbinding of the molecule rendered energetically unfavorable.

Another set of simulations was carried out with the targeted molecular dynamics (TMD) method (Schlitter et al., 1993). The initial and final structures of an SMD simulation were used as input for the TMD simulations as discussed in "Methods". TMD trajectories were calculated in "both directions" between the input structures, simulating both the binding and the

unbinding events. A comparison of the SMD and TMD simulations revealed that the pathways generated by both methods show very similar modes of concerted rotations around single bonds during the unbinding of AA.

Both methods suggest that the chemical structure of AA (*cis* double bonds connected by two single bonds) allows the fatty acid to access the cyclooxygenase active site of PGHS-1 through a narrow hydrophobic channel and to bind in a shape favorable for the cyclooxygenation reaction.

3.7 Force-Induced Unfolding of Titin

The giant muscle protein titin, also known as connectin, is a roughly 30,000 amino acid long filament which plays a number of important roles in muscle contraction and elasticity (Labeit et al., 1997; Maruyama, 1997; Wang et al., 1993). The I-band region of titin, largely composed of immunoglobulin-like (Ig) domains, is believed to be responsible for the molecule's extensibility and passive elasticity. Recent AFM (Rief et al., 1997) and optical tweezer (Rief et al., 1997)Kellermayer et al., 1997; Tskhovrebova et al., 1997) experiments directly measured the force–extension profile of single titin molecules. In the AFM experiment, cloned sections of titin composed of adjacent I-band Ig domains were stretched at constant speed. The force–extension profile showed a pattern of sawtooth-shaped peaks, spaced 250–280 Å apart, with each force peak corresponding to the unfolding of a single Ig domain. The Ig domains were thus observed to unfold one by one under the influence of an applied external force. To examine in atomic detail the dynamics and structure-function relationships of this behavior, SMD simulations of force-induced titin Ig domain unfolding were performed (Lu et al., 1998).

The SMD simulations were based on an NMR structure of the Ig domain I27 of the cardiac titin I-band (Improta et al., 1996). The Ig domains consist of two β-sheets packed against each other, with each sheet containing four strands, as shown in Fig. 8b. After I27 was solvated and equilibrated, SMD simulations were carried out by fixing one terminus of the domain and applying a force to the other in the direction from the fixed terminus to the other terminus. Simulations were performed as described by Eq. (1) with $v = 0.5$ Å/ps and $K = 10\ k_BT/\text{Å}^2 \simeq 414$ pN/Å. The force–extension profile from the SMD trajectory showed a single force peak as presented in Fig. 8a. This feature agrees well with the sawtooth-shaped force profile exhibited in AFM experiments.

The simulation trajectory shown in Fig. 8b provides an explanation of how the force profile in Fig. 8a arises. During extension from 0 to 10 Å the two β-sheets slid away from each other, each maintaining a stable structure and its intra-sheet backbone hydrogen bonds. As the extension of the domain reached 14 Å, the structure within each sheet began to break: in one sheet, strands A' and G slid past each other, while in the other sheet, strands A and B slid past each other. The A'–G and A–B backbone hydrogen bonds broke nearly simultaneously, producing the large initial force peak seen in Fig. 8a.

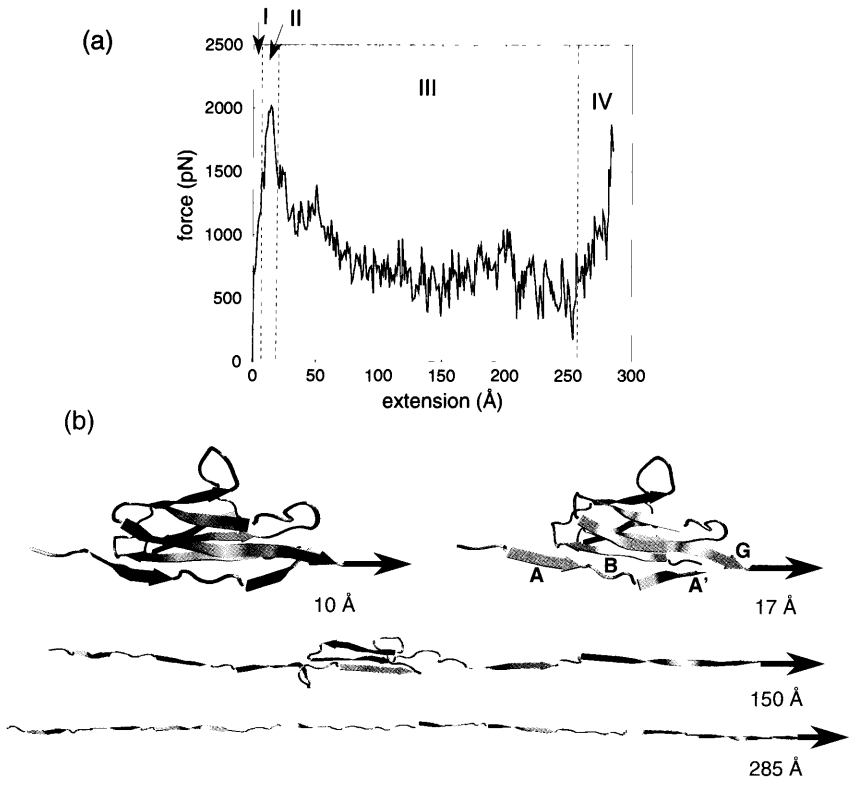


Fig. 8.
(a) Force extension profile from SMD simulations of titin I27 domain with a pulling velocity $v = 0.5$ Å/ps. The extension domain is divided into four sections: I. pre-burst, II. major burst, III. post-burst, IV. pulling of fully extended chain. (b) Intermediate stages of the force-induced unfolding. All I27 domains are drawn in the cartoon representation of the folded domain; solvating water molecules are not shown. The four figures at extensions 10 Å, 17 Å, 150 Å, and 285 Å correspond, respectively, to regions I to IV defined in (a).

These events marked the beginning of the Ig domain unfolding, after which the strands unraveled one at a time, accompanied by a large reduction in the recorded force. After an extension of 260 Å, the domain was completely unfolded; further stretching of the already extended polypeptide chain caused the force to increase dramatically.

The simulation (Lu et al., 1998) suggested how Ig domains achieve their chief design requirement of bursting one by one when subjected to external forces. At small extensions, the hydrogen bonds between strands A and B and between strands A' and G prevent significant extension of a domain, i.e.,

the domain maintains its β-sandwich structure. After these bonds break, resistance to unfolding becomes much smaller, and the domain unfolds rapidly. Only when a domain is fully extended does the force increase enough to begin the unfolding process in another domain.

4 Stochastic Modeling of SMD

In AFM experiments as well as in SMD simulations, the ligand extracted from the binding pocket is subjected to a time-dependent external force. The rate of change and the functional form of the applied force critically affect the behavior of the ligand and the information one can obtain from experiment and simulation (Evans and Ritchie, 1997; Izrailev et al., 1997; Balsera et al., 1997). To better understand the results of SMD simulations and how they compare to experimental measurements, it is helpful to consider an idealized one-dimensional stochastic model that captures the essence of unbinding phenomena and reveals the limit of the information that can be gained from SMD simulations about the binding potentials. The model assumes that the motion of the ligand proceeds in the strong friction limit along a reaction coordinate x, and is governed by a one-dimensional Langevin equation

$$\gamma \dot{x} \; = \; - \frac{dU}{dx} \; + \; F(x,t) \; + \; \sigma \, N(t) \; , \tag{3}$$

where γ is the time-independent coefficient of friction of the ligand in the binding pocket of the protein, $U(x)$ and $F(x,t)$ are respectively the potential surface governing dissociation and the external force applied to the ligand, and $\sigma N(t)$ is a stochastic force of amplitude σ and zero mean.

We assume that the unbinding reaction takes place on a time scale long compared to the relaxation times of all other degrees of freedom of the system, so that the friction coefficient can be considered independent of time. This condition is difficult to satisfy on the time scales achievable in MD simulations. It is, however, the most favorable case for the reconstruction of energy landscapes without the assumption of thermodynamic reversibility, which is central in the majority of established methods for calculating free energies from simulations (McCammon and Harvey, 1987; Elber, 1996) (for applications and discussion of free energy calculation methods see also the chapters by Helms and McCammon, Hermans et al., and Mark et al. in this volume).

In this section we describe the behavior of a ligand subjected to three types of external forces: a constant force, forces exerted by a moving stiff harmonic spring, and forces exerted by a soft harmonic spring. We then present a method of reconstruction of the potential of mean force from SMD force measurements employing a stiff spring (Izrailev et al., 1997; Balsera et al., 1997).

4.1 Unbinding Induced by a Constant Force

Unbinding processes can be viewed as taking place in several qualitatively different regimes (Izrailev et al., 1997; Marrink et al., 1998). These regimes can be illustrated by considering the simplest binding potential

$$U(x) = \begin{cases} +\infty & \text{for } x < a, \\ \Delta U \, \frac{x-a}{b-a} & \text{for } a \le x \le b, \\ \Delta U & \text{for } x > b. \end{cases} \tag{4}$$

We assume in the following that the ligand is bound in a binding pocket of depth $b-a = 7$ Å involving a potential barrier $\Delta U = 25$ kcal/mol, similar to that of streptavidin (Chilcotti et al., 1995). We also assume that the diffusion coefficient of the ligand is similar to the diffusion coefficient of the heme group in myoglobin ($D = 1$ Å^2/ns) as determined from Mößbauer spectra (Nadler and Schulten, 1984).

To unbind from a protein the ligand has to move from a, the minimum of the potential $U(x)$, to b, the maximum of $U(x)$. The mean first passage time $\tau(F)$ of such motion is (Izrailev et al., 1997)

$$\tau(F) = 2\tau_d \, \delta(F)^{-2} \left[e^{\delta(F)} - \delta(F) - 1 \right], \tag{5}$$

where $\tau_d = (b-a)^2/2D$ and

$$\delta(F) = \left[\Delta U - F(b-a) \right] / k_B T. \tag{6}$$

Activated Regime For a small applied force F corresponding to positive δ and $e^{\delta(F)} \gg \max(1, \delta)$, the mean time of unbinding is

$$\tau_{act} \approx 2\tau_d \, [\delta(F)]^{-2} \, e^{\delta(F)}. \tag{7}$$

This result reflects the Kramers' relation (Gardiner, 1985). A millisecond time of unbinding, i.e., $\tau_{act} \approx 1$ ms, corresponds in this case to a rupture force of 155 pN. For such a force the potential barrier ΔU is not abolished completely; in fact, a residual barrier of 9 kcal/mol is left for the ligand to overcome. The AFM experiments with an unbinding time of 1 ms are apparently functioning in the thermally activated regime.

Diffusive Regime In the case of a stronger force, such that $F \approx \Delta U/(b-a)$ and $\delta(F) \approx 0$, one obtains from (5)

$$\tau_{diff} = \tau_d \left(1 + \frac{\delta(F)}{3} \right). \tag{8}$$

In this regime the applied force completely overwhelms the binding potential and the ligand is subject to free diffusion. The mean free passage time in this regime is equal to τ_d and is on the order of 25 ns.

Drift Regime For still stronger forces corresponding to $\delta(F) < 0$ and $e^{\delta(F)} \ll 1 \ll |\delta(F)|$, using (5), one obtains

$$\tau_{drift} \approx 2\tau_d \, |\delta(F)|^{-1} . \tag{9}$$

This regime involves forces which are so strong that the ligand undergoes a drift motion governed by (3) in the limit that the fluctuating force $\sigma N(t)$ is negligible compared to the applied force. In this case a force of about 800 pN would lead to rupture within 500 ps.

These examples illustrate that SMD simulations operate in a different regime than existing micromanipulation experiments. Considerably larger forces (800 pN vs. 155 pN) are required to induce rupture, and the scaling behavior of the drift regime, characterized by (9), differs qualitatively from the activated regime as characterized by (7). Hence, SMD simulations of rupture processes can not be scaled towards the experimental force and time scales.

4.2 Unbinding Induced by Harmonic Springs

Assume now that the ligand is pulled by a harmonic spring, that is, subjected to an external force $F(x,t)$ of the form given by (1). The position of the ligand in the binding pocket fluctuates; according to the Boltzmann distribution of a harmonically bound particle, the position fluctuations associated with the spring are characterized by a variance $\delta x \sim (k_B T/K)^{1/2}$, and the corresponding variance in the applied force is related to K through $\delta F \sim (K k_B T)^{1/2}$. A stiff spring confines the ligand to fluctuate in a small region of the binding pocket, so that only local properties of the binding potential are sampled, while the fluctuations of the force are large. For a soft spring, on the other hand, the ligand is able to fluctuate in a large region of the binding pocket, and the fluctuations of the force are small.

Stiff Spring For a stiff spring, satisfying $K \gg |d^2U/dx^2|$, under the overdamped condition assumed in (3) the average force measured by the spring can be expressed as

$$\overline{F} = \frac{dU}{d\overline{x}} + \gamma v, \tag{10}$$

where $\overline{F} = K(vt - \overline{x})$, and where $\overline{F}$ and $\overline{x}$ denote a running time average of F and x, respectively (Balsera et al., 1997). Equation (10) implies that for a stiff restraint the average applied force measures the local slope of the binding potential plus a frictional contribution that depends linearly on the pulling velocity. This dependence was observed in the MD simulations of the biotin-streptavidin complex (Grubmüller et al., 1996).

Soft Spring For a soft spring, no linear scaling of the rupture force with the pulling velocity should result (Izrailev et al., 1997). For millisecond unbinding times and soft springs employed in AFM experiments ($K \simeq 6$ pN/Å) (Florin et al., 1994), thermal fluctuations facilitate the exit of the ligand from the binding pocket of the protein. This means that the unbinding is thermally activated and the unbinding time $\tau_R(F)$, according to Bell's relation, depends exponentially on the height of the energy barrier $\Delta U^\dagger$ reduced by the applied force F (Bell, 1978). The rupture force F_{AFM}, in this case, satisfies (Izrailev et al., 1997)

$$F_{AFM} = \frac{\Delta U^\dagger}{L} - \frac{k_B T}{L} \ln \frac{\tau_R(F)}{\tau_d}. \tag{11}$$

The rupture force measured in AFM experiments is given, therefore, by the average slope of the energy profile minus a correction related to the effects of thermal fluctuations. Equation (11) demonstrates that the rupture force measured in AFM experiments grows linearly with the activation energy of the system (Chilcotti et al., 1995). A comparison of (10) and (11) shows that the unbinding induced by stiff springs in SMD simulations, and that induced by AFM differ drastically, and that the forces measured by both techniques cannot be readily related.

4.3 Reconstruction of the Potential of Mean Force

Measurement of the unbinding force should not be the only goal of SMD simulations. Even if the value of that force corresponded to experimental observations it would still not yield sufficient information to understand the dynamics of association/dissociation. Knowledge of the free energy profile of the system along the unbinding coordinate is required. Balsera et al. (1997) showed that it is possible, under the idealized conditions of (3), to reconstruct a one-dimensional potential of mean force from SMD simulation data. For a stiff spring the frictional contribution to the applied force can be explicitly discounted. One can simply integrate (10) to obtain an estimate of the potential $U(x)$,

$$\overline{U}(x) - U(0) = \int_0^x dx' \left(\overline{F} - \gamma v\right) . \tag{12}$$

Dissipation, however, imposes limits on how precisely the potential can be reconstructed. With the introduction of the work performed by the frictional force W_{fr}, the uncertainty σ_U^2 in the reconstructed potential $U(x)$ can be presented as (Balsera et al., 1997)

$$\sigma_U^2 = 2\, k_B T\, W_{fr} \equiv 2\, k_B T\, \gamma v x\,. \tag{13}$$

Thus, the uncertainty in the potential U is determined by the irreversible work done on the system. This irreversible work is proportional to the pulling

velocity v and can be reduced with an increase in simulation time. In the avidin-biotin system, for example, the size of the avidin binding pocket is $x \sim 10$ Å. Assuming again a diffusion coefficient $D = 1$ Å^2/ns, simulation periods of 1 ns and 10 ns, corresponding to pulling velocities of the order of 10^{-2} and 10^{-3} Å/ps, yield $\sigma_U \sim 8$ and 3 kcal/mol, respectively. By contrast, an attempt to reconstruct the potential of mean force using a soft spring clearly fails.

5 Discussion

SMD is a novel approach to the study of the dynamics of binding/unbinding events in biomolecular systems and of their elastic properties. The simulations reveal the details of molecular interactions in the course of unbinding, thereby providing important information about binding mechanisms. The advantage of SMD over conventional molecular dynamics is the possibility of inducing relatively large conformational changes in molecules on nanosecond time scales. Other methods, such as umbrella sampling, free energy perturbation (McCammon and Harvey, 1987), and weighted histogram analysis (Kumar et al., 1992), aiming at the determination of the energy landscapes, typically involve small conformational changes and require extensive computations to achieve accuracy.

In cases where irreversible work done during unbinding can be attributed to a non-dispersive frictional force γv, a quantitative description of the thermodynamic potentials governing the binding and unbinding processes can be achieved by discounting the irreversible work from the calculated potential of mean force. However, the error in the reconstructed potentials is related to irreversible work done on the system and, therefore, may be unacceptably large. The estimate of the friction also presents a challenge, since it can be highly dispersive and may exhibit memory effects (Balsera et al., 1997).

Irreversibility of the unbinding process can also be accounted for by averaging over an ensemble of SMD trajectories according to the non-equilibrium equality for free energy differences (Jarzynski, 1997a; Jarzynski, 1997b) as described in the Introduction. This approach, however, requires averaging over multiple trajectories, and may be extremely sensitive to insufficient sampling of reaction pathways.

Irreversible work might also be discounted by forcing a conformational change in the system followed by the reverse conformational change, i.e., inducing hysteresis. Such an approach may yield a "model free" estimate of the irreversible work component from the hysteresis (Baljon and Robbins, 1996; Xu et al., 1996). Finally, lengthening the simulation time decreases the amount of irreversible work and the simulated process could, ideally, reach quasi-equilibrium in the limit of very long simulation times.

The simulations of the avidin-biotin complex (Izrailev et al., 1997) showed that a major difficulty involved in studies of the binding and flexibility of

macromolecules is the long time scale of motions such as the fluctuations of the avidin loop discussed above. These fluctuation times, ranging from several nanoseconds to seconds, are beyond the reach of SMD simulations that can presently be realized in a feasible amount of time (Balsera et al., 1996).

Solvation is likely to influence protein-ligand binding and, hence, the forces measured in SMD simulations. During the extraction of retinal from bacteriorhodopsin (see Fig. 3) water facilitated the breaking of a hydrogen bond between retinal and Lys216 (Israelewitz et al., 1997). In the simulations of the avidin-biotin complex, placement of several water molecules near the exit from the binding pocket reduced the measured binding force (see Fig. 2). In the simulations of the streptavidin-biotin complex (Grubmüller et al., 1996) the binding pocket was exposed to solvent due to exclusion of the ajacent streptavidin monomer. This allowed water molecules to enter the binding pocket and participate in breaking hydrogen bond networks between the ligand and the protein during the unbinding. The issue of how water molecules participate in the process of protein-ligand dissociation remains unclear and should be further investigated. For example, one may add water molecules to the binding pocket in the course of SMD simulations, as the retracting ligand frees up space in the binding pocket (Resat et al., 1996).

Binding and unbinding of non-covalently attached biomolecules are at the heart of many important processes and are the target of experimental investigations. SMD may serve to interpret measurements and suggest new experiments. The rapidly growing computer power available for simulations and increasing time resolution of experimental techniques will provide the basis for further advances in the method and will help bridge the gap in time scales between computer simulation and experiment.

6 Acknowledgments

The authors thank Y. Oono and M. Balsera for invaluable contributions to the joint development of SMD and J. Gullingsrud for many suggestions in the preparation of the manuscript. Images of molecular systems were produced with the program VMD (Humphrey et al., 1996). This work was supported by grants from the National Institute of Health (PHS 5 P41 RR05969-04), the National Science Foundation (BIR-9318159, BIR 94-23827 (EQ)), and the Roy J. Carver Charitable Trust.

References

[Ajay and Murcko, 1995] Ajay, and Murcko, M.: Computational methods to predict binding free energy in ligand-receptor complexes. J. Med. Chem. **38** (1995) 4953–4967

[Allen et al., 1996] Allen, P. G., Laham, L. E., Way, M., and Janmey, P. A.: Binding of phosphate, aluminum fluoride, or beryllium fluoride to F-actin inhibits severing by gelsolin. J. Biol. Chem. **271** (1996) 4665–4670

[Andersson et al., 1992] Andersson, M. L., Nordström, K., Demczuck, S., Harbers, M., and Vennström, B.: Thyroid hormone alters the DNA binding properties of chicken thyroid hormone receptors α and β. Nucl. Acids Res. **20** (1992) 4803–4810

[Baljon and Robbins, 1996] Baljon, R. C. A., and Robbins, M. O.: Energy dissipation during rupture of adhesive bonds. Science. **271** (1996) 482–484

[Balsera et al., 1997] Balsera, M., Stepaniants, S., Izrailev, S., Oono, Y., and Schulten, K.: Reconstructing potential energy functions from simulated force-induced unbinding processes. Biophys. J. **73** (1997) 1281–1287

[Balsera et al., 1996] Balsera, M. A., Wriggers, W., Oono, Y., and Schulten, K.: Principal component analysis and long time protein dynamics. J. Phys. Chem. **100** (1996) 2567–2572

[Bell, 1978] Bell, G. I.: Models for the specific adhesion of cells to cells. Science. **200** (1978) 618–627

[Binning et al., 1986] Binning, G., Quate, C. F., and Gerber, G.: Atomic force microscope. Phys. Rev. Lett. **56** (1986) 930–933

[Block and Svoboda, 1994] Block, S., and Svoboda, K.: Biological applications of optical forces. Ann. Rev. Biophys. Biomol. Struct. **23** (1994) 247–285

[Booth et al., 1996] Booth, P. J., Farooq, A., and Flitsch, S. L.: Retinal binding during folding and assembly of the membrane protein bacteriorhodopsin. Biochemistry. **35** (1996) 5902–5909

[Brent et al., 1989] Brent, G. A., Dunn, M. K., Harney, J. W., Gulick, T., and Larsen, P. R.: Thyroid hormone aporeceptor represses T_3 inducible promoters and blocks activity of the retinoic acid receptor. New Biol. **1** (1989) 329–336

[Cevc and Marsh, 1987] Cevc, G., and Marsh, D.: Phospholipid Bilayers: Physical Principles and Models. John Wiley & Sons, New York, 1987.

[Chang et al., 1988] Chang, C.-H., Jonas, R., Govindjee, R., and Ebrey, T.: Regeneration of blue and purple membranes for deionized bleached membranes of *halobacterium halobium*. Photochem. Photobiol. **47** (1988) 261–265

[Chilcotti et al., 1995] Chilcotti, A., Boland, T., Ratner, B. D., and Stayton, P. S.: The relationship between ligand-binding thermodynamics and protein-ligand interaction forces measured by atomic force microscopy. Biophys. J. **69** (1995) 2125–2130

[Cohen et al., 1990] Cohen, N., Blaney, J., Humblet, C., Gund, P., and Barry, D.: Molecular modeling software and methods for medicinal chemistry. J. Med. Chem. **33** (1990) 883–894

[Colman, 1994] Colman, P.: Structure-based drug design. Curr. Opinion Struct. Biol. **4** (1994) 868–874

[Damm et al., 1989] Damm, K., Thompson, C. C., and Evans, R. M.: Protein encoded by v-*erb*A functions as a thyroid-hormone receptor antagonist. Nature. **339** (1989) 593–597

[Dancker and Hess, 1990] Dancker, P., and Hess, L.: Phalloidin reduces the release of inorganic phosphate during actin polymerization. Biochim. Biophys. Acta. **1035** (1990) 197–200

[Dennis, 1983] Dennis, E. A.: Phospholipases. *In* The enzymes vol. XVI, 1983.

[Devreotes and Zigmond, 1988] Devreotes, P. N., and Zigmond, S. H.: Chemotaxis in eukaryotic cells: a focus on leukocytes and *Dictyostelium*. Ann. Rev. Cell Biol. 4 (1988) 649–686

[Eaton et al., 1997] Eaton, W. A., Muños, V., Thompson, P. A., Chan, C.-K., and Hofrichter, J.: Submillisecond kinetics of protein folding. Curr. Opinion Struct. Biol. **7** (1997) 10–14

[Elber, 1996] Elber, R.: Reaction path studies of biological molecules. *In* Recent developments in theoretical studies of proteins (Advanced series in physical chemistry, Vol. 7). R. Elber, editor. World Scientific, Singapore, 1996.

[Evans et al., 1991] Evans, E., Berk, D., and Leung, A.: Detachment of agglutinin-bonded red blood cells. Biophys. J. **59** (1991) 838–848

[Evans et al., 1995] Evans, E., Ritchie, K., and Merkel, R.: Sensitive force technique to probe molecular adhesion and structural linkages at biological interfaces. Biophys. J. **68** (1995) 2580–2587

[Evans and Ritchie, 1997] Evans, E., and Ritchie, K.: Dynamic strength of molecular adhesion bonds. Biophys. J. **72** (1997) 1541–1555

[Florin et al., 1994] Florin, E.-L., Moy, V. T., and Gaub, H. E.: Adhesion force between individual ligand-receptor pairs. Science. **264** (1994) 415–417

[Gardiner, 1985] Gardiner, C. W.: Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences. Springer-Verlag, New York, 1985.

[Gilson et al., 1997] Gilson, M., Given, J., Bush, B., and McCammon, J.: The statistical-thermodynamic basis for computation of binding affinities: A critical review. Biophys. J. **72** (1997) 1047–1069

[Green, 1975] Green, N. M.: Avidin. Advan. Prot. Chem. **29** (1975) 85–133

[Grubmüller, 1995] Grubmüller, H.: Predicting slow structural transitions in macromolecular systems: Conformational Flooding. Phys. Rev. E. **52** (1995) 2893–2906

[Grubmüller et al., 1996] Grubmüller, H., Heymann, B., and Tavan, P.: Ligand binding and molecular mechanics calculation of the streptavidin-biotin rupture force. Science. **271** (1996) 997–999

[Hanessian and Devasthale, 1996] Hanessian, S., and Devasthale, P.: Design and synthesis of novel, pseudo C2 symmetric inhibitors of HIV protease. Bioorg. Med. Chem. Lett. **6** (1996) 2201–2206

[Humphrey et al., 1996] Humphrey, W. F., Dalke, A., and Schulten, K.: VMD – Visual Molecular Dynamics. J. Mol. Graphics. **14** (1996) 33–38

[Improta et al., 1996] Improta, S., Politou, A., and Pastore, A.: Immunoglobulin-like modules from titin I-band: extensible components of muscle elasticity. Structure. **4** (1996) 323–337

[Israelachvili, 1992] Israelachvili, J. N.: Intermolecular and Surface Forces. Academic Press, London, 1992.

[Israelwitz et al., 1997] Israelwitz, B., Izrailev, S., and Schulten, K.: Binding pathway of retinal to bacterio-opsin: A prediction by molecular dynamics simulations. Biophys. J. **73** (1997) 2972–2979

[Izrailev et al., 1997] Izrailev, S., Stepaniants, S., Balsera, M., Oono, Y., and Schulten, K.: Molecular dynamics study of unbinding of the avidin-biotin complex. Biophys. J. **72** (1997) 1568–1581

[Jain et al., 1995] Jain, M. K., Gelb, M., Rogers, J., and Berg, O.: Kinetic basis for interfacial catalysis by phospholipase A_2. Methods in enzymology. **249** (1995) 567–614

[Jarzynski, 1997a] Jarzynski, C.: Equilibrium free-energy differences from nonequilibrium measurements: A master equation approach. Phys. Rev. E. **56** (1997a) 5018–5035

[Jarzynski, 1997b] Jarzynski, C.: Nonequilibrium equality for free energy differences. Phys. Rev. Lett. **78** (1997b) 2690–2693

[Kellermayer et al., 1997] Kellermayer, M., Smith, S., Granzier, H., and Bustamante, C.: Folding-unfolding transition in single titin modules characterized with laser tweezers. Science. **276** (1997) 1112–1116

[Kumar et al., 1992] Kumar, S., Bouzida, D., Swendsen, R. H., Kolman, P. A., and Rosenberg, J. M.: The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. J. Comp. Chem. **13** (1992) 1011–1021

[Labeit et al., 1997] Labeit, S., Kolmerer, B., and Linke, W.: The giant protein titin: emerging roles in physiology and pathophysiology. Circulation Research. **80** (1997) 290–294

[Lebon et al., 1996] Lebon, F., Vinals, C., Feytmans, E., and Durant, F.: Computational drug design of new HIV-1 protease inhibitors. Arch. Phys. Biochem. **104** (1996) B44.

[Leckband et al., 1994] Leckband, D. E., Schmitt, F. J., Israelachvili, J. N., and Knoll, W.: Direct force measurements of specific and nonspecific protein interactions. Biochemistry. **33** (1994) 4611–4624

[Lu et al., 1998] Lu, H., Isralewitz, B., Krammer, A., Vogel, V., and Schulten, K.: Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. Biophys. J. **75** (1998) 662–671

[Lüdemann et al., 1997] Lüdemann, S. K., Carugo, O., and Wade, R. C.: Substrate access to cytochrome P450cam: A comparison of a thermal motion pathway analysis with molecular dynamics simulation data. J. Mol. Model. **3** (1997) 369–374

[Marrink et al., 1998] Marrink, S.-J., Berger, O., Tieleman, P., and Jähnig, F.: Adhesion forces of lipids in a phospholipid membrane studied by molecular dynamics dimulations. Biophys. J. **74** (1998) 931–943

[Marrone et al., 1997] Marrone, T., Briggs, J., and McCammon, J.: Structure-based drug design: Computational advances. Ann. Rev. Pharm. Tox. **37** (1997) 71–90

[Maruyama, 1997] Maruyama, K.: Connectin/titin, a giant elastic protein of muscle. FASEB J. **11** (1997) 341–345

[McCammon and Harvey, 1987] McCammon, J. A., and Harvey, S. C.: Dynamics of Proteins and Nucleic Acids. Cambridge University Press, Cambridge, 1987.

[Moy et al., 1994a] Moy, V. T., Florin, E.-L., and Gaub, H. E.: Adhesive forces between ligand and receptor measured by AFM. Colloids and Surfaces. **93** (1994a) 343–348

[Moy et al., 1994b] Moy, V. T., Florin, E.-L., and Gaub, H. E.: Intermolecular forces and energies between ligands and receptors. Science. **266** (1994b) 257–259

[Nadler and Schulten, 1984] Nadler, W., and Schulten, K.: Theory of Mössbauer spectra of proteins fluctuating between conformational substates. Proc. Natl. Acad. Sci. USA. **81** (1984) 5719–5723

[Nelson et al., 1995] Nelson, M., Humphrey, W., Gursoy, A., Dalke, A., Kalé, L., Skeel, R., Schulten, K., and Kufrin, R.: MDScope – A visual computing environment for structural biology. Comput. Phys. Commun. **91** (1995) 111–134

[Nelson et al., 1996] Nelson, M., Humphrey, W., Gursoy, A., Dalke, A., Kalé, L., Skeel, R. D., and Schulten, K.: NAMD – A parallel, object-oriented molecular dynamics program. J. Supercomputing App. **10** (1996) 251–268

[Oberhauser et al., 1998] Oberhauser, A. F., Marszalek, P. E., Erickson, H., and Fernandez, J.: The molecular elasticity of tenascin, an extracellular matrix protein. Nature. In Press.

[Oesterhelt et al., 1992] Oesterhelt, D., Tittor, J., and Bamberg, E.: A unifying concept for ion translocation in retinal proteins. J. Bioenerg. Biomemb. **24** (1992) 181–191

[Oesterhelt and Schumann, 1974] Oesterhelt, D., and Schumann, L.: Reconstitution of bacteriorhodopsin. FEBS Lett. **44** (1974) 262–265

[Olender and Elber, 1996] Olender, R., and Elber, R.: Calculation of classical trajectories with a very large time step: Formalism and numerical examples. J. Chem. Phys. **105** (1996) 9299–9315

[Picot et al., 1994] Picot, D., Loll, P. J., and Garavito, M.: The X-ray crystal structure of the membrane protein prostaglandin H_2 synthase-1. Nature. **367** (1994) 243–249

[Pollard et al., 1992] Pollard, T. D., Goldberg, I., and Schwarz, W. H.: Nucleotide exchange, structure, and mechanical properties of filaments assembled from ATP-actin and ADP-actin. J. Biol. Chem. **267** (1992) 20339–20345

[Resat et al., 1996] Resat, H., Mezei, M., and McCammon, J. A.: Use of the grand canonical ensemble in potential of mean force calculations. J. Phys. Chem. **100** (1996) 1426–1433

[Rief et al., 1997] Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. M., and Gaub, H. E.: Reversible unfolding of individual titin immunoglobulin domains by AFM. Science. **276** (1997) 1109–1112

[Schlitter et al., 1993] Schlitter, J., Engels, M., Krüger, P., Jacoby, E., and Wollmer, A.: Targeted molecular dynamics simulation of conformational change - application to the t $\leftrightarrow$ r transition in insulin. Molecular Simulation. **10** (1993) 291–308

[Schulten et al., 1995] Schulten, K., Humphrey, W., Logunov, I., Sheves, M., and Xu, D.: Molecular dynamics studies of bacteriorhodopsin's photocycles. Israel Journal of Chemistry. **35** (1995) 447–464

[Slotboom et al., 1982] Slotboom, A. J., Verheij, H. M., and Haas, G. H. D.: On the mechanism of phospholipase A_2. *In* Phospholipids, . J. N. Hawthorne, and G. B. Ansell, editors. Elsevier Biomedical Press, New York. 359–435, 1982

[Small, 1989] Small, J. V.: Microfilament-based motility in non-muscle cells. Curr. Opinion Cell Biol. **1** (1989) 75–79

[Smith and DeWitt, 1996] Smith, W., and DeWitt, D.: Prostaglandin endoperoxide H synthases-1 and -2. Adv. Immunol. **62** (1996) 167–215

[Stepaniants et al., 1997] Stepaniants, S., Izrailev, S., and Schulten, K.: Extraction of lipids from phospholipid membranes by steered molecular dynamics. J. Mol. Model. **3** (1997) 473–475

[Strynadka et al., 1996] Strynadka, N., Eisenstein, M., Katchalski-Katzir, E., Shoichet, B., Kuntz, I., Abagyan, R., Totrov, M., Janin, J., Cherfils, J., Zimmerman, F., Olson, A., Duncan, B., Rao, M., Jackson, R., Sternberg, M., and James, M.: Molecular docking programs successfully predict the binding of a β-lactamase inhibitory protein to TEM-1 β-lactamase. Nature Struct. Biol. **3** (1996) 233–239

[Thaisrivongs et al., 1996] Thaisrivongs, S., Romero, D., Tommasi, R., Janakiraman, M., Strohbach, J., Turner, S., Biles, C., Morge, R., Johnson, P., Aristoff, P., Tomich, P., Lynn, J., Horng, M., Chong, K., Hinshaw, R., Howe, W., Finzel, B., and Watenpaugh, K.: Structure-based design of HIV protease inhibitors - 5,6-dihydro-4-hydroxy-2-pyrones as effective, nonpeptidic inhibitors. J. Med. Chem. **39** (1996) 4630–4642

[Theriot et al., 1992] Theriot, J. A., Mitchison, T. J., Tilney, L. G., and Portnoi, D. A.: The rate of actin-based motility of intracellular *Listeria monocytogenes* equals the rate of actin polymerization. Nature. **357** (1992) 257–260

[Tskhovrebova et al., 1997] Tskhovrebova, L., Trinick, J., Sleep, J., and Simmons, R.: Elasticity and unfolding of single molecules of the giant protein titin. Nature. **387** (1997) 308–312

[Wagner et al., 1995] Wagner, R., Apriletti, J. W., McGrath, M. E., West, B. L., Baxter, J. D., and Fletterick, R. J.: A structural role for hormone in the thyroid hormone receptor. Nature. **378** (1995) 690–697

[Wang et al., 1993] Wang, K., McCarter, R., Wright, J., Beverly, J., and Ramirez-Mitchell, R.: Viscoelasticity of the sarcomere matrix of skeletal muscles. Biophys. J. **64** (1993) 1161–1177

[Wriggers and Schulten, 1997a] Wriggers, W., and Schulten, K.: Protein domain movements: Detection of rigid domains and visualization of hinges in comparisons of atomic coordinates. Proteins: Struc. Func. and Genetics. **29** (1997a) 1–14

[Wriggers and Schulten, 1997b] Wriggers, W., and Schulten, K.: Stability and dynamics of G-actin: Back door water diffusion and behavior of a subdomain 3/4 loop. Biophys. J. **73** (1997b) 624–639

[Wriggers and Schulten, 1998] Wriggers, W., and Schulten, K.: Investigating a back door mechanism of actin phosphate release by steered molecular dynamics. Biophys. J. Submitted.

[Xu et al., 1996] Xu, D., Phillips, J. C., and Schulten, K.: Protein response to external electric fields: Relaxation, hysteresis and echo. J. Phys. Chem. **100** (1996) 12108–12121

[Zhang and Hermans, 1996] Zhang, L., and Hermans, J.: Hydrophilicity of cavities in proteins. Proteins: Struc. Func. and Genetics. **24** (1996) 433–438

Conformational Transitions of Proteins from Atomistic Simulations

Volkhard Helms and J. Andrew McCammon

Department of Chemistry and Biochemistry, University of California at San Diego,La Jolla CA 92093, USA

Abstract. The function of many important proteins comes from their dynamic properties, and their ability to undergo conformational transitions. These may be small loop movements that allow access to the protein's active site, or large movements such as those of motor proteins that are implicated with muscular extension. Yet, in spite of the increasing number of three-dimensional crystal structures of proteins in different conformations, not much is known about the driving forces of these transitions. As an initial step towards exploring the conformational and energetic landscape of protein kinases by computational methods, intramolecular energies and hydration free energies were calculated for different conformations of the catalytic domain of cAMP-dependent protein kinase (cAPK) with a continuum (Poisson) model for the electrostatics. In this paper, we will put the previous results into context and discuss possible extensions into the dynamic regime.

1 Introduction

Proteins are biopolymers formed by one or more continuous chains of covalently linked amino acids. Hydrogen bonds between non-adjacent amino acids stabilize the so-called elements of secondary structure, α-helices and $\beta-$sheets. A number of secondary structure elements then assemble to form a compact unit with a specific fold, a so-called domain. Experience has shown that a number of folds seem to be preferred, maybe because they are especially suited to perform biological protein function. A complete protein may consist of one or more domains.

Protein dynamics occurs on very different time scales ([McCammon and Harvey 1987, Jardetzky 1996]). Here, we are most interested in long time scale motions such as relative motion between secondary structure elements, and inter-domain motion.

Our present view of proteins is governed to a great deal by the information from X-ray crystallographic analysis of protein crystals. Here, a specific protein is described at atomic resolution as a static picture in an *average* conformation. Knowledge of this average conformation is sometimes sufficient to understand properties of proteins such as docking of protein modules to form larger protein complexes. Yet, the function of many proteins involves conformational transitions such as those in motor proteins that allow for muscular extension, or in enzymes where a surface loop may swing open to allow the entrance of a substrate to a buried active site. Insight into the structural

mechanisms of these transitions has been gained from the growing number of examples where multiple crystal structures have been determined of a protein in different conformations ([Gerstein et al. 1994]). Inter-domain motions can often be characterized as hinge-bending or shear movements.

Experimental information about protein conformational transitions has also been gained from a variety of other techniques such as gel-filtration, circular dichroism, intrinsic fluorescence spectroscopy, time-resolved fluorescence energy transfer, electron spin resonance with spin labels, neutron and small-angle X-ray scattering. Yet, the dynamic information from these techniques alone may be hard to link with specific structural transitions. The only techniques that allow for a simultaneous detection of protein structure and dynamics are time-resolved X-ray crystallography, and NMR spectroscopy ([Jardetzky 1996]). Time-resolved X-ray crystallography is undergoing a dramatic development, and it has been possible to collect structural data of protein mobility down to ns time scales. By combining these new developments with existing methods, our knowledge about the mechanisms and time scales of structural transitions in proteins will certainly expand enormously in the near future.

Theoretical investigations will certainly be essential in identifying the driving forces of these transitions and thus characterizing the underlying free energy landscape of folded proteins. Transitions between different conformational states of a folded protein have been studied with a wide variety of theoretical methods, but we limit ourselves to discussing a few important techniques in the following.

2 Calculations Based on Static Protein Structures

The first technique is very intuitive. Out of the few proteins that could be crystallized in a number of different conformations, adenylate kinase is probably the best-studied example. By combining nine observed crystal structures and interpolating between them, a movie was constructed that visualized a hypothetical path of its hinge-bending transition ([Vonrhein et al. 1995]).

We now turn towards theoretical techniques that involve the computation of macromolecular energies.

2.1 Accurate Energy Calculations of a Protein in Different Conformations

The interaction with the solvent is of similar importance as the intramolecular energy contributions and a correct representation of the solvent is therefore essential. If an explicit solvent description is chosen, averaging over many different solvent configurations is necessary in order to obtain converged statistical averages. Advantageous in this respect is describing the solvent as

a dielectric continuum. Especially accurate results for the electrostatic portion of the interaction with solvent and ions can be obtained by solving the Poisson-Boltzmann equation.

We have previously calculated conformational free energy differences for a well-suited model system, the catalytic subunit of cAMP-dependent protein kinase (cAPK), which is the best characterized member of the protein kinase family. It has been crystallized in three different conformations and our main focus was on how ligand binding shifts the equilibrium among these ([Helms and McCammon 1997]). As an example using state-of-the-art computational techniques, we summarize the main conclusions of this study and discuss a variety of methods that may be used to extend this study into the dynamic regime of protein domain motion.

Calculation of Conformational Free Energies for a Model of a Bilobal Enzyme Protein kinases catalyze the transfer of phosphate from adenosine triphosphate (ATP) to protein substrates and are regulatory elements of most known pathways of signal transduction.

The catalytic subunit of cAPK contains two domains connected by a peptide linker. ATP binds in a deep cleft between the two domains. Presently, crystal structures showed cAPK in three different conformations, (1) in a closed conformation in the ternary complex with ATP or other tight-binding ligands and a peptide inhibitor PKI(5-24), (2) in an intermediate conformation in the binary complex with adenosine, and (3) in an open conformation in the binary complex of mammalian cAPK with PKI(5-24). Fig.1 shows a superposition of the three protein kinase configurations to visualize the type of conformational movement.

While the angle formed by the two domains opens by ca. 15 deg between the crystal structures of the closed and open forms, data from low-angle X-ray scattering experiments of cAPK alone in solution indicated an opening of as much as 39 deg. The transition apparently not only involves the hinge movement of the two domains relative to each other but also a number of more subtle structural changes when analyzed in detail. Still, both domains superimpose quite well, and to a first approximation, the domain movement can be regarded as a relative displacement of rigid bodies.

Structures Used in this Study Instead of performing energy calculations for different crystal structure conformations, a strategy was chosen that involved mapping the domains of one highly-resolved structure onto the intermediate and open frameworks. By comparing these chimeric structures with the native closed structure, the calculated energy differences are solely due to the opening and closing mechanism.

As a template for an 'intermediate' conformation of protein kinase, the crystal structure of the binary complex of cAPK with adenosine (1bkx.pdb in the Protein Data Bank) was used. As templates for 'open' conformations

Fig. 1. Superposition of three crystal structures of cAMP-dependent protein kinase that show the protein in a closed conformation (straight line), in an intermediate conformation (dashed line), and in an open conformation (broken line). The structures were superimposed on the large lobe. In three locations, arrows identify corresponding amino acid positions in the small lobe.

of cAPK, we used two structures of binary complexes of mammalian cAPK with PKI(5-24) (1ctp.pdb and 1cmk.pdb). Three different crystal structures that contain cAPK in a closed conformation were modelled into putative 'intermediate' and 'open' conformations by superimposing the two domains on those of the intermediate or open template structures and then constructing appropriate peptide linkages.

Free Energy Function The conformational free energy was estimated by the following energy expression:

$$\Delta G_{confE} = \Delta E_{bond} + \Delta E_{angle} + \Delta E_{prop.dih} + \Delta E_{improp.dih} + \Delta E_{vdW} + \Delta E_{coul} + \Delta G_{PB} + \Delta G_{surf} .$$

Here, ΔE_{bond}, ΔE_{angle}, $\Delta E_{prop.dih}$, and $\Delta E_{improp.dih}$ are equal to the covalent intramolecular energies for bond stretching, angle bending, and for the out-of plane bending of proper and improper dihedral angles. ΔE_{vdW} and ΔE_{coul} are the noncovalent intrasolute van der Waals and coulombic energies. These intramolecular energy terms were all evaluated by the molecular mechanics force field CHARMm22 after a short energy minimization to remove local strain. ΔG_{PB} is the electrostatic free energy of hydration evaluated by the UHBD program ([Madura et al. 1995])as a finite- difference solution of the Poisson equation. The apolar solvation free energy ΔG_{surf} is calculated as the solvent-accessible surface area of the protein times a constant.

The continuum description of the solvent yields a free energy including entropic solvent effects. However, intra-molecular entropy differences between different conformational states of the protein are not accounted for, nor is the entropic contribution for possible displacement of solvent molecules bound to the protein during conformational transitions. Energy calculations were performed for the unbound protein, for binary complexes with ATP and with PKI(5-24), and for a ternary complex with ATP and with PKI(5-24). By comparing relative differences of conformational energies, it is assumed that many of the neglected contributions to the free energy will cancel out.

The results of the calculations for the three systems were averaged and are displayed in Fig. 2. For unbound cAPK as well for cAPK complexed with ATP or PKI, the total conformational free energies increase monotonically in the sequence closed - 'intermediate' - 'open'. Binding is an interplay of opposing effects of unfavourable Poisson free energy of hydration that favors the solvation of buried protein groups, and the favourable van der Waals energy and coulombic energy that favour association of both domains.

The ranking of conformational free energies indicated that the closed state of cAPK is favored even in the absence of ligands, which is in contrast to experimental data that showed a preferred population of the open conformation. One reason for this discrepancy could be that the modelled 'intermediate'

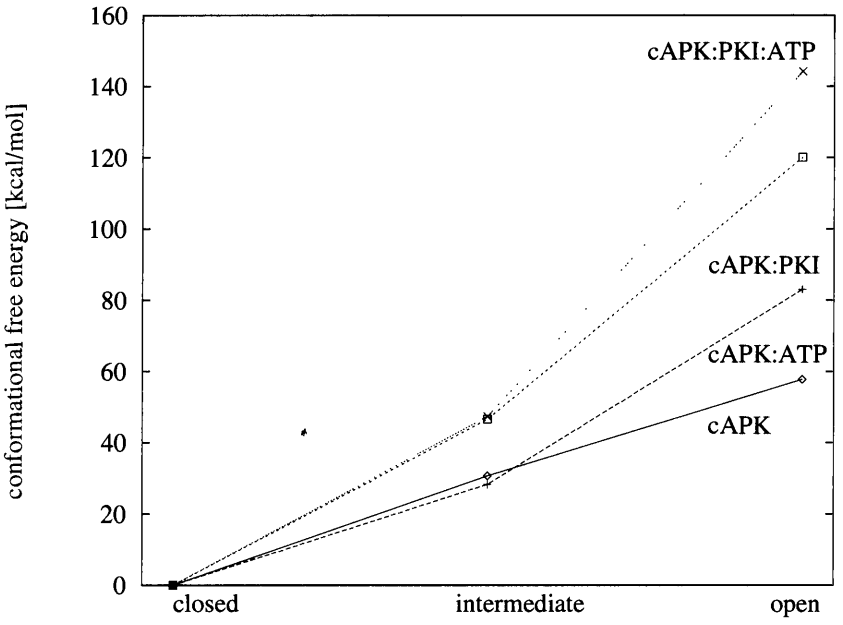


Fig. 2. Conformational free energy of closed, intermediate and open protein kinase conformations. 'cAPK' indicates the unbound form of cAMP-dependent protein kinase, 'cAPK:ATP' the binary complex of cAPK with ATP, 'cAPK:PKI' the binary complex of cAPK with the peptide inhibitor PKI(5-24), and 'cAPK:PKI:ATP' the ternary complex of cAPK with ATP and PKI(5-24). Shown are averaged values for the three crystal structures 1ATP.pdb, 1CDKA.pdb, and 1CDKB.pdb. All values have been normalized with respect to the free energy of the closed conformations.

and 'open' states do not correspond to real states. Another more likely reason might be the omission of important contributions to the conformational free energy in our calculations as mentioned above.

It was therefore advisable to focus on the analysis of relative effects such as how the energy ranking of states is affected by the binding of the peptide inhibitor and ATP. The calculated conformational free energies indicate that both ATP and PKI(5-24) disfavour the open state with regard to intermediate and closed state. Binding of PKI(5-24) seems to drive the complete closing of the binding cleft from the intermediate to the closed state. This is in accord with data from small-angle neutron scattering that showed that the binding of inhibitor peptide, but not ATP alone, was sufficient to cause the reduction in the protein radius of gyration.

The following two techniques were developed to expand such static calculations into a pseudo-dynamic regime by calculating higher derivatives of the potential energy and by introducing an additional degree of freedom.

2.2 Normal Mode Analysis (NMA)

The basic idea of NMA is to expand the potential energy function $U(\mathbf{x})$ in a Taylor series expansion around a point $\mathbf{x}_0$ where the gradient of the potential vanishes ([Case 1996]). If third and higher-order derivatives are ignored, the dynamics of the system can be described in terms of the normal mode directions and frequencies $\mathbf{q}_i$ and ω_i which satisfy:

$$\underline{M}^{1/2}\,\underline{F}\,\underline{M}^{-1/2}\,\mathbf{q_i} = \omega_i^2\,\mathbf{q_i}$$
$$\mathbf{q_i}\,\mathbf{q_j} = \delta_{ij}\,.$$

The matrix $\underline{M}$ contains atomic masses on its diagonal, and the Hessian matrix $\underline{F}$ contains the second derivatives of the potential energy evaluated at $\mathbf{x}_0$.

The influence of solvent can be incorporated in an implicit fashion to yield so-called langevin modes. Although NMA has been applied to allosteric proteins previously, the predictive power of normal mode analysis is intrinsically limited to the regime of fast structural fluctuations. Slow conformational transitions are dominantly found in the regime of anharmonic protein motion.

2.3 Langevin Dynamics on an Adiabatic Surface

In an early study of lysozyme ([McCammon et al. 1976]), the two domains of this protein were assumed to be rigid, and the hinge-bending motion in the presence of solvent was described by the Langevin equation for a damped harmonic oscillator. The angular displacement Θ from the equilibrium position is thus governed by

$$I\frac{d^2\Theta}{dt^2} + f\frac{d\Theta}{dt} + k\Theta = N_r(t)\,.$$

The effective moment of inertia I and the friction coefficient f could easily be estimated. The force constant k associated with the relative motion of the lobes was determined from an empirical energy function. To do so, the molecule was opened in a step-wise fashion by manipulating the hinge region and each resulting structure was energy minimized. Then, the interaction energy between the two domains was measured, and plotted versus Θ.

3 Simulations of Protein Dynamics

Simulations of the dynamic motion of proteins aim at sampling relevant portions of the conformational space accessible to the proteins under the influence of environmental variables such as temperature, pressure, and pH. We

will limit ourselves here to applications of atomistic molecular dynamics simulations. Techniques that employ a reduced number of simulation particles e.g. one bead per amino acid side chain ([Head-Gordon and Brooks 1991]) have large potential applicability, but their success in the description of protein dynamics has been limited. Apparently, atomistic detail is very important to provide the necessary flexibility and small-scale interactions on the length-scale of hydrogen bonds.

3.1 Molecular Dynamics Simulations

Molecular dynamics simulations ([McCammon and Harvey 1987]) propagate an atomistic system by iteratively solving Newton's equation of motion for each atomic particle. Due to computational constraints, simulations can only be extended to a typical time scale of 1 ns currently, and conformational transitions such as protein domains movements are unlikely to be observed.

To facilitate conformational transitions in the before-mentioned adenylate kinase, Elamrani and co-workers scaled all atomic masses by a large factor thus allowing the use of a high effective simulation temperature of 2000K ([Elamrani et al. 1996]). To prevent protein unfolding, elements of secondary structure had to be constrained.

The important underlying components of protein motion during a simulation can be extracted by a Principal Component Analysis (PCA). It stands for a diagonalization of the variance-covariance matrix $\underline{R}$ of the mass-weighted internal displacements during a molecular dynamics simulation.

$$\underline{R} = \frac{1}{N} \underline{X} \ \underline{X}^T .$$

$\underline{X}$ is a matrix whose elements $\mathbf{x}_{it}$ give the mass-weighted internal displacements of each atomic coordinate i from its average position at a given time step t. N is the total number of integration steps.

The essential slow modes of a protein during a simulation accounting for most of its conformational variability can often be described by only a few principal components. Comparison of PCA with NMA for a 200 ps simulation of bovine pancreatic trypsic inhibitor showed that the variation in the first principal components was twice as high as expected from normal mode analysis ([Hayward et al. 1994]). The so-called essential dynamics analysis method ([Amadei et al. 1993]) is a related method and will not be discussed here.

An interesting approach has recently been chosen in the MBO(N)D program ([Moldyn 1997]). Structural elements of different size varying from individual peptide planes up to protein domains can be defined to be rigid. During an atomistic molecular dynamics simulation, all fast motion orthogonal to the lowest normal modes is removed. This allows use of ca. 20 times longer time steps than in standard simulations.

3.2 Conformational Flooding

Grubmüller described a method to induce conformational transitions in proteins and derived rate constants for these ([Grubmüller 1994]). The method employs subsequent modifications of the original potential function based on a principal component analysis of a short MD simulation. It is discussed in more detail in the chapter of Eichinger *et al.* in this volume.

3.3 Targeted Dynamics

If both starting structure and target structure are known, the method of targeted molecular dynamics simulation can be used to enforce a conformational transition towards the given final structure during a given simulation time t_s ([Schlitter et al. 1994]).

Starting at an initial conformation $\mathbf{x}_i$ with a distance ρ_0 from the final structure $\mathbf{x}_f$, the allowed distance ρ is decreased at each integration step δt by

$$\delta\rho = (\rho_0 - \rho_f)\frac{\delta t}{t_s}\,.$$

In order to fulfil the following condition at each instant,

$$|\mathbf{x} - \mathbf{x}_f|^2 - \rho^2 = 0\,.$$

an additional constraint force is applied to the system by a Lagrangian multiplier while taking a time step.

3.4 Approximate Simulation Between Two Structures with Large Time Steps

Related to the previous method, a simulation scheme was recently derived from the Onsager-Machlup action that combines atomistic simulations with a reaction path approach ([Oleander and Elber 1996]). Here, time steps up to 100 times larger than in standard molecular dynamics simulations were used to produce approximate trajectories by the following equations of motion:

$$\begin{aligned} \underline{M}\,\frac{d^2\mathbf{x}}{dt^2} + \frac{dU}{d\mathbf{x}} &= \mathbf{Q} \\ \underline{M}\,\frac{d^2\mathbf{Q}}{dt^2} + \frac{d^2U}{d\mathbf{x}^2}\,\mathbf{Q} &= 0\,. \end{aligned}$$

The trajectories with the highest probability are those for which the Onsager-Machlup action

$$S = \int dt \, (\mathbf{Q} \cdot \mathbf{Q})$$

is minimal. The boundary conditions employed were of fixed end points and of zero $\mathbf{Q}$ at the edges of the path connecting starting and end points. It was shown that stable molecular dynamics simulations could be computed by optimization of S.

3.5 Non-Equilibrium MD Simulations

By applying a pulling force at a portion of the solute molecule in a specific direction (see chapters of Eichinger *et al.* and Schulten in this volume), conformational transitions can be induced in specific directions. In order to reconstruct information about the underlying potential function governing protein motion, the irreversible work performed on the system by these forces must be discounted ([Balsera et al. 1997]).

3.6 Poisson-Boltzmann Stochastic Dynamics (PBSD)

This method was introduced recently as an efficient technique to accurately model solvent and salt effects in an implicit fashion ([Gilson et al. 1993]). The forces are calculated as a sum of three terms:

$$\mathbf{F} = \rho \mathbf{E} - \frac{1}{2} E^2 - \frac{1}{2} \epsilon \kappa^2 \phi^2 \nabla \lambda \, .$$

The first term represents the forces due to the electrostatic field, the second describes forces that occur at the boundary between solute and solvent regime due to the change of dielectric constant, and the third term describes ionic forces due to the tendency of the ions in solution to move into regions of lower dielectric. Applications of the so-called PBSD method on small model systems and for the interaction of a stretch of DNA with a protein model have been discussed recently ([Elcock et al. 1997]). This simulation technique guarantees equilibrated solvent at each state of the simulation and may therefore avoid some of the problems mentioned in the previous section. Due to the smaller number of particles, the method may also speed up simulations potentially. Still, to be able to simulate long time scale protein motion, the method might ideally be combined with non-equilibrium techniques to enforce conformational transitions.

A number of issues need to be addressed before this method will become a routine tool applicable to problems as the conformational equilibrium of protein kinase. E.g. the accuracy of the force field, especially the combination of Poisson-Boltzmann forces and molecular mechanics force field, remains to be assessed. The energy surface for the opening of the two kinase domains in Fig. 2 indicates that intramolecular noncovalent energies are overestimated compared to the interaction with solvent.

4 Conclusions

The study of slow protein dynamics is a fascinating field with still many unknowns. We have presented a number of computational techniques that are currently being used to tackle those questions. Most promising for our case seems the development of methods that combine an implicit solvent description with techniques to induce conformational transitions.

5 Acknowledgements

This work was supported by a postdoctoral NATO fellowship to VH by the Deutscher Akademischer Austauschdienst, and by grants from NIH, NSF, and the NSF Supercomputer Centers MetaCenter program. VH is also a fellow of the Program in Mathematics and Molecular Biology and of the La Jolla Interfaces in Sciences Program.

References

[Amadei et al. 1993] Amadei, A., Linssen, A.B.M., Berendsen, H.J.C.: Essential Dynamics of Proteins. Proteins **17** (1993) 412–425

[Balsera et al. 1997] Balsera, M., Stepaniants, S., Izrailev, S., Oono, Y., Schulten, K.: Reconstructing Potential Energy Functions from Simulated Force-Induced Unbinding Processes. Biophys. J. **73** (1997) 1281–1287

[Case 1996] Case, D.A.: Normal mode analysis of protein dynamics. Curr. Op. Struct. Biol. **4** (1994) 285–290

[Elamrani et al. 1996] Elamrani, S., Berry, M.B., Phillips Jr., G.N., McCammon, J.A.: Study of Global Motions in Proteins by Weighted Masses Molecular Dynamics: Adenylate Kinase as a Test Case. Proteins **25** (1996) 79–88

[Elcock et al. 1997] Elcock, A.H., Potter, M.J., McCammon, J.A.: Application of Poisson-Boltzmann Solvation Forces to Macromolecular Simulations. In "Computer Simulation of Biomolecular Systems," Vol. 3, A.J. Wilkinson et al. eds., ESCOM Science Publishers B.V., Leiden

[Gerstein et al. 1994] Gerstein, M., Lesk, A.M., Chothia, C.: Structural Mechanisms for Domain Movements in Proteins. Biochemistry **33** (1994) 6739–6749

[Gilson et al. 1993] Gilson, M.K., Davis, M.E., Luty, B.A., McCammon, J.A.: Computation of Electrostatic Forces on Solvated Molecules Using the Poisson-Boltzmann Equation. J. Phys. Chem. **97** (1993) 3591–3600

[Grubmüller 1994] Grubmüller, H.: Predicting Slow Structural Transitions in Macromolecular Systems - Conformational Flooding. Phys. Rev. E **52** (1994) 2893–2906

[Hayward et al. 1994] Hayward, S., Kitao, A., Gō, N.: Harmonic and anharmonic aspects in the dynamics of BPTI: A normal mode analysis and principal component analysis. Prot. Sci. **3** (1994) 936–943

[Head-Gordon and Brooks 1991] Head-Gordon, T., Brooks, C.L.: Virtual rigid body dynamics. Biopol. **31** (1991) 77–100

[Helms and McCammon 1997] Helms, V., McCammon, J.A.: Kinase Conformations: A computational study of the effect of ligand binding. Prot. Sci. **6** (1997) 2336–2343

[Jardetzky 1996] Jardetzky, O.: Protein dynamics and conformational transitions in allosteric proteins. Prog. Biophys. Mol. Biol. **65** (1996) 171–219

[Madura et al. 1995] Madura, J.D., Briggs, J.M., Wade, R.C., Davis, M.E., Luty, B.A., Ilin, A., Antosiewicz, J., Gilson, M.K., Bagheri, B., Scott, L.R., McCammon, J.A.: Electrostatics and Diffusion of Molecules in Solution: Simulations with the University of Houston Brownian Dynamics Program. Comp. Phys. Comm. **91** (1995) 57–95

[McCammon et al. 1976] McCammon, J.A., Gelin, B.R., Karplus, M., Wolynes, P.G.: The hinge-bending mode in lysozyme. Nature **262** (1976) 325–326

[Moldyn 1997] Moldyn Inc., 955 Massachusetts ave, 5th Floor, Cambridge, MA 02139-3180, USA

[McCammon and Harvey 1987] McCammon, J.A., Harvey, S.C.: Dynamics of Proteins and Nucleic Acids. Cambridge University Press, Cambridge (1987).

[Oleander and Elber 1996] Oleander, R., Elber, R.: Calculation of classical trajectories with a very large time step: Formalism and numerical examples. J. Chem. Phys. **105** (1996) 9299–9315

[Schlitter et al. 1994] Schlitter, J., Engels, M., Krüger, P.: Targeted molecular dynamics: A new approach for searching pathways of conformational transitions. J. Mol. Graph. **12** (1994) 84–89

[Vonrhein et al. 1995] Vonrhein, C., Schlauderer, G.J., Schulz, G.E.: Movie of the structural changes during a catalytic cycle of nucleoside monophosphate kinases. Structure **3** (1995) 483–490.

Conformational Dynamics Simulations of Proteins

Markus Eichinger, Berthold Heymann, Helmut Heller, Helmut Grubmüller, and Paul Tavan

Institut für Medizinische Optik, Theoretische Biophysik,
Ludwig-Maximilians-Universität München,
Oettingenstraße 67, D-80538 München, Germany

Abstract. Molecular dynamics (MD) simulations of proteins provide descriptions of atomic motions, which allow to relate observable properties of proteins to microscopic processes. Unfortunately, such MD simulations require an enormous amount of computer time and, therefore, are limited to time scales of nanoseconds. We describe first a fast multiple time step structure adapted multipole method (FAMUSAMM) to speed up the evaluation of the computationally most demanding Coulomb interactions in solvated protein models, secondly an application of this method aiming at a microscopic understanding of single molecule atomic force microscopy experiments, and, thirdly, a new method to predict slow conformational motions at microsecond time scales.

1 Introduction

In many cases the detailed knowledge of dynamic processes at the atomic level is essential to understand protein function, e.g., ligand binding or enzymatic reactions. Through a microscopic description of interatomic forces [1] and atomic motions, molecular dynamics (MD) simulations [2, 3] can serve as a tool to interpret experimental data and to make predictions, which can guide future experiments. In such simulations, the motions are computed by numerically solving Newton's equations. Here, the forces are derived from an empirical energy function accounting for chemical binding forces as well as van der Waals and electrostatic interactions between partially charged atoms.

For the study of protein dynamics quite large simulation systems — typically comprising several 10,000 atoms — are required. The system must be that large because the native protein environment (water or lipids) strongly affects the dynamics of the protein [4, 5, 6, 7, 8] and, therefore, has to be included into the simulation system. The large number of atoms provides a first reason why MD simulations of proteins pose a computational challenge. A second reason is that femtosecond integration time steps are necessary to enable sufficiently smooth descriptions of the fastest degrees of freedom. Thus, MD simulations of such systems are currently limited to nanoseconds (i.e., a few million integration steps) even if the most powerful supercomputers and efficient algorithms are used. Although there are a number of biochemically important processes which occur at such very fast time scales and have been

successfully studied by MD simulations [9, 10], most biochemical processes occur at much slower scales and, therefore, are currently inaccessible to conventional MD methods. This technical limitation motivates substantial efforts taken by many groups to determine suitable approximations which ideally should allow more efficient simulations without seriously affecting relevant features of the system, which may be grouped into specialized integration schemes and multiple time stepping [11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31] (see also the chapter by Schlick and Berne within this book), multipole methods [32, 33, 34, 35, 36, 37, 38], as well as grid and Ewald methods [39, 40, 41, 42]. Most of the efforts focus on the efficient computation of the electrostatic interactions within the protein and between protein and solvent, since, typically, this is the computationally most demanding task.

As an example for an efficient yet quite accurate approximation, in the first part of our contribution we describe a combination of a structure adapted multipole method with a multiple time step scheme (FAMUSAMM — fast multistep structure adapted multipole method) and evaluate its performance. In the second part we present, as a recent application of this method, an MD study of a ligand-receptor unbinding process enforced by single molecule atomic force microscopy. Through comparison of computed unbinding forces with experimental data we evaluate the quality of the simulations. The third part sketches, as a perspective, one way to drastically extend accessible time scales if one restricts oneself to the study of conformational transitions, which are ubiquitous in proteins and are the elementary steps of many functional conformational motions.

2 Efficient MD-Simulation Methods

In order to solve the classical equations of motion numerically, and, thus, to obtain the motion of all atoms the forces acting on every atom have to be computed at each integration step. The forces are derived from an energy function which defines the molecular model [1, 2, 3]. Besides other important contributions (which we shall not discuss here) this function contains the Coulomb sum

$$U \propto \sum_{i}^{N} \sum_{j<i} \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j|}$$

over all pairs of atoms (i, j) with partial charges q_i at positions $\mathbf{r}_i$. The evaluation of this sum dominates the computational effort in MD simulations as it scales quadratically with the number N of charged particles.

A very simple — and in fact quite widely used — approximation completely neglects long range electrostatic interactions beyond a certain cut-off distance [43] of typically $8 - 15$ Å. For systems which are significantly larger

than this cut-off distance the computation of the remaining Coulomb interactions then scales with N instead of N^2. However, such truncation leads to serious artifacts concerning the description of the structure and dynamics of proteins [44, 24, 45], and more accurate methods which include the long range interactions should be preferred. Multipole methods and multiple-time-step methods are well established and widely used for this purpose. We briefly sketch both methods and subsequently show how their combination allows highly efficient simulations.

2.1 Efficient Multipole Methods

Multipole methods approximate the long-range forces originating from a group of point charges by truncated multipole expansions of their electrostatic potential. Using a hierarchy of grids for subdivision of space, nested at multiple scales, and a corresponding hierarchical organization of charge groups and multipole expansions [33] a computational complexity of $O(NlogN)$ can be achieved. By additionally using a hierarchy of local Taylor expansions for the evaluation of the electrostatic potential in the vicinity of a group of particles Greengard and Rokhlin have constructed the so-called *fast multipole method* (FMM) that even scales with $O(N)$ for large systems [34, 35].

For MD simulations of biomolecules the FMM-type grouping of charges, defined by a fixed and regular subdivision of space, requires multipole expansions of rather high order (more than 6 terms of the expansion) as to achieve sufficient numerical accuracy [34]. If, instead, as shown in Figure 1, the charge grouping is adapted to specific structural and dynamical properties of the simulated biomolecules, the multipole expansions can be truncated at quite low orders, e.g., after the second order, while maintaining sufficient accuracy [36, 37, 38].

In the FAMUSAMM framework, e.g., we have grouped locally stable groups of typically three to ten covalently bound atoms into so-called structural units (level 1 in Fig. 1). By construction, these structural units either carry integer elementary charges or are uncharged, but dipolar. Test simulations show that for distances > 10 Å already the lowest non-vanishing multipole moments of these structural units provide a sufficiently accurate description of the electrostatic forces within biomolecules with an error below 2%. The objects of the next hierarchy level (level 2 in the figure) are formed by grouping structural units into clusters. For interaction distances > 15 Å the electrostatic potential of those objects, again, can be approximated by their lowest multipole moment. Extending this scheme to higher hierarchy levels, such a structure adapted multipole method (SAMM) provides a substantial speed-up for MD simulations as compared to the conventional, grid-based methods [38, 46].

The performance of this first version of SAMM [36] can be further enhanced by additionally utilizing FMM-strategies [34, 38]. Here, in the vicinity

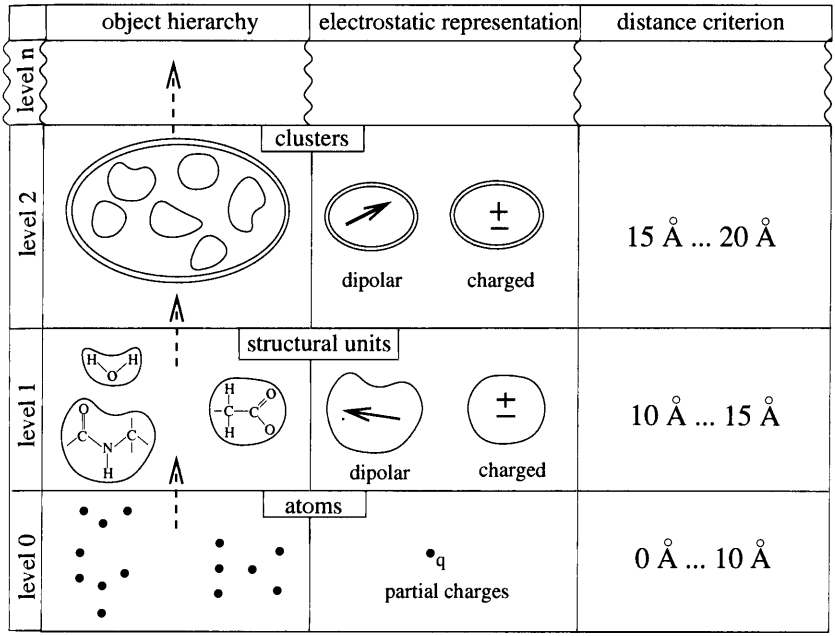


Fig. 1. Structure adapted hierarchical description of Coulomb interactions in biological macromolecules. Filled circles (level 0) represent atoms, structural units (level 1) are surrounded by a single-line border, and clusters (level 2) are surrounded by a double-line border.

of a given object (e.g., a structural unit or a cluster) the electrostatic potential originating from distant charge distributions is approximated by a local Taylor expansion. Specifically, the basic tasks involved in the FMM aspect of SAMM are:

Task 1: Calculate the first non-vanishing multipole moment of the electrostatic potential of composed objects (i.e., structural units and clusters).

Task 2: Add up electrostatic potential contributions to local Taylor expansions of all objects on each hierarchy level. (Contributions to the local Taylor expansion of a selected object arise from all other objects on the same hierarchy which fulfill the distance criterion given in Fig. 1.)

Task 3: Transform ("inherit") local Taylor expansions from a upper hierarchy level to the next lower hierarchy level.

Task 4: Explicitly calculate the Coulomb interactions between atoms which are closer than about 10 Å.

In the next section we will illustrate how to further speed up the SAMM method by introducing multiple-time-stepping.

2.2 SAMM with Multiple-Time-Stepping

In general, multiple-time-step methods increase computational efficiency in a way complementary to multipole methods: The latter make use of regularities *in space*, whereas multiple-time-stepping exploits regularities *in time*. Figure 2 illustrates the general idea:

As sketched in the right part of the figure, forces between distant atoms generally exhibit slower fluctuations than forces between close atoms. Therefore, without significant loss of accuracy, the more slowly fluctuating forces may be computed using larger integration step sizes. As shown in the left part of the figure, the required classification of forces can be implemented, e.g., by grouping atom pairs into distance classes. The slowly fluctuating forces arising from outer distance classes may then be evaluated less frequently (filled squares) than the fast ones and, instead, are extrapolated (open squares) from previously computed forces at the time steps in between.

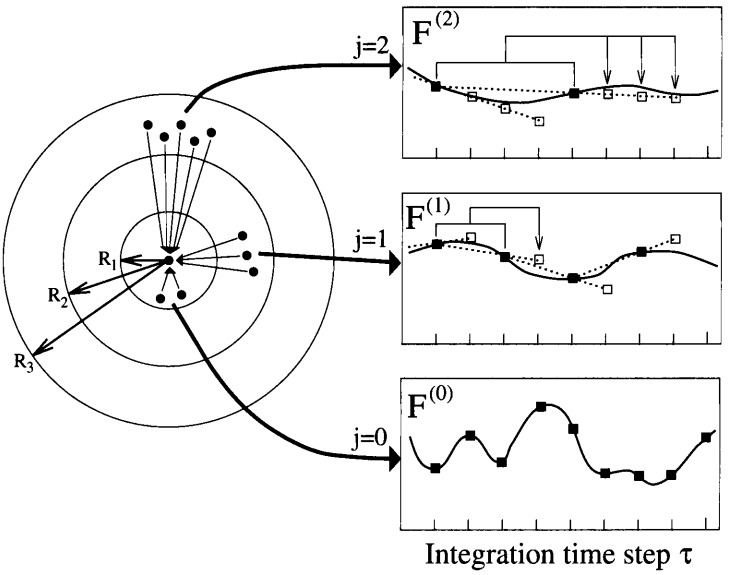


Fig. 2. Distance classes $j = 0, 1, 2, \ldots$ (left) are defined for an atom (central dot) by a set of radii R_{j+1}; the right curves sketch the temporal evolution of the total force $\mathbf{F}^{(j)}$ acting on the selected atom originating from all atoms in distance class j; shown are the exact forces (solid line), their exact values to be computed within the multiple time step scheme (filled squares), linear force extrapolations (dotted lines), and resulting force estimates (open squares).

This hierarchical extrapolation procedure can save a significant amount of computer time as it avoids a large fraction of the most time consuming step, namely the exact evaluation of long range interactions. Here, computational

speed is gained at the cost of an increased demand for memory, e.g., for each atom and each distance class two previously computed forces have to be kept in memory.

In the framework of the fast structure adapted multipole method the memory demand could be drastically reduced. This was achieved by applying the multiple-time-step scheme (we used the so-called DC-1d scheme [24]) to the interactions between charge groups (structural units and clusters) rather than to the forces acting on individual atoms. In the following we give a short description of this tight and efficient combination. We termed the resulting algorithm FAMUSAMM (*multiple-time-step structure-adapted multipole method*) [46, 47].

A detailed analysis of fast SAMM has shown [47] that the most time consuming tasks are task 2 and task 4 described above. In task 2 for each hierarchy level (except for level 0) a local Taylor expansion is calculated for each object. Note that here we refer to expansions which comprise only contributions from objects of the same hierarchy level which, in addition, fulfill the distance criterion given in Fig. 1. From each of these local expansions, *approximated* electrostatic forces $\tilde{\mathbf{F}}^{(j)}$ acting on the atoms contained in the associated object could be computed and, in analogy to the exact forces $\mathbf{F}^{(j)}$ used in the multiple time step scheme described above (see Fig. 2), the multipole-derived forces $\tilde{\mathbf{F}}^{(j)}$ could be extrapolated by multiple time stepping. We further improved that obvious scheme, however, in that we applied multiple time step extrapolations to the *coefficients* of the local Taylor expansions instead. That strategy reduces memory requirements by a significant factor without loss of accuracy, since the number of local Taylor coefficients that have to be kept for the extrapolation is smaller than the number of forces acting on all atoms of the respective object.

Additionally, to optimize task 4, we applied a conventional, atom pair interaction based multiple-time-step scheme to the force computation within the innermost distance class. Here, for atom pairs closer than 5 Å, the Coulomb sum is calculated every step, and for all other atom pairs the Coulomb sum is extrapolated every second step from previously explicitly calculated forces.

This completes the outline of FAMUSAMM. The algorithm has been implemented in the MD simulation program EGO_VIII [48] in a sequential and a parallelized version; the latter has been implemented and tested on a number of distributed memory parallel computers, e.g., IBM SP2, Cray T3E, Parsytec CC and ethernet-linked workstation clusters running PVM or MPI.

2.3 Computational Performance

Here we want to document that FAMUSAMM actually provides an enhanced computational efficiency both as compared to SAMM as well as to the reference method which is characterized by exact evaluation of the Coulomb sum. To that aim we have carried out a series of test simulations for systems of

varying size ranging from 500 to 40,000 atoms. We used the sequential version of EGO_VIII. All simulations were executed on a DEC-ALPHA 3300L (175 MHz) workstation equipped with 96 MB RAM. Figure 3 shows that the average computation time required for one MD integration step scales linearly with system size for systems comprising more than about 1,000 atoms.

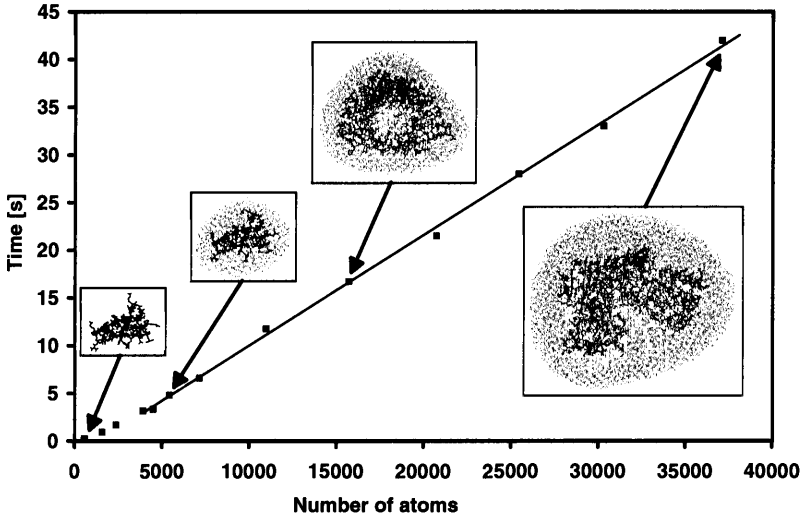


Fig. 3. Average computation time for one step using EGO_VIII on a DEC-Alpha 3300L workstation (175 MHz) for simulation systems of varying size. The insets show some of the protein-water systems used for the benchmark simulations.

For large systems comprising 36,000 atoms FAMUSAMM performs *four* times faster than SAMM and as fast as a cut-off scheme with a 10 Å cut-off distance while completely avoiding truncation artifacts. Here, the speed-up with respect to SAMM is essentially achieved by the multiple-time-step extrapolation of local Taylor expansions in the outer distance classes. For this system FAMUSAMM executes by a factor of 60 faster than explicit evaluation of the Coulomb sum. The subsequent Section describes, as a sample application of FAMUSAMM, the study of a ligand-receptor unbinding process.

3 Microscopic Interpretation of Atomic Force Microscope Rupture Experiments

That simulation study [49] aimed at a microscopic interpretation of single molecule atomic force microscope (AFM) experiments [50], in which unbinding forces between individual protein–ligand complexes have been measured

(Fig. 4, top). In particular we asked, what interatomic interactions cause the experimentally observed unbinding forces.

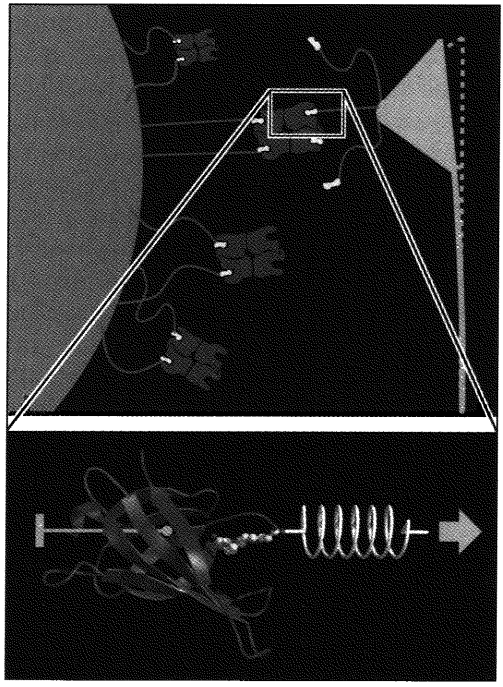


Fig. 4. Typical AFM rupture experiment (top): Receptor molecules are fixed via linker molecules to a surface (left); in the same way, ligand molecules are connected to the AFM cantilever (right). When pulling the cantilever towards the right, the pulling force applied to the ligand can be measured. At the point of rupture of the ligand–receptor complex the measured force abruptly drops to zero so that the rupture force can be measured.
Computer rupture simulation (bottom): In the course of an MD simulation of the ligand–receptor complex at atomic detail the ligand is pulled towards the right with a 'computer spring', while the receptor (drawn as a ribbon model) is kept in place. From the elongation of the 'spring' the pulling force during the unbinding process is computed, and, thereby, a 'force profile' is obtained. The rupture force is interpreted as the maximum of this force.

Both the AFM rupture experiments as well as our simulation studies focussed on the streptavidin-biotin complex as a model system for specific ligand binding. Streptavidin is a particularly well-studied protein and binds its ligand biotin with high affinity and specificity [51]. Whereas previous experiments (see references in Ref. [49]) and simulation studies [52] referred only to bound/unbound states and the associated kinetics, the recent AFM

rupture experiments have provided a new and complementary perspective on ligand binding by focussing at atomic details of binding/unbinding *pathways*: The former were described in terms of binding free energies as thermodynamic quantities, which are independent of the particular reaction pathway; the latter relate to *forces*, which actually depend on details of the unbinding reaction path and, therefore, can provide new insights into these details.

To enable an atomic interpretation of the AFM experiments, we have developed a molecular dynamics technique to simulate these experiments [49]. From such 'force simulations' rupture models at atomic resolution were derived and checked by comparisons of the computed rupture forces with the experimental ones. In order to facilitate such checks, the simulations have been set up to resemble the AFM experiment in as many details as possible (Fig. 4, bottom): the protein–ligand complex was simulated in atomic detail starting from the crystal structure, water solvent was included within the simulation system to account for solvation effects, the protein was held in place by keeping its center of mass fixed (so that internal motions were not hindered), the cantilever was simulated by use of a harmonic 'spring potential' and, finally, the simulated cantilever was connected to the particular atom of the ligand, to which in the AFM experiment the linker molecule was connected.

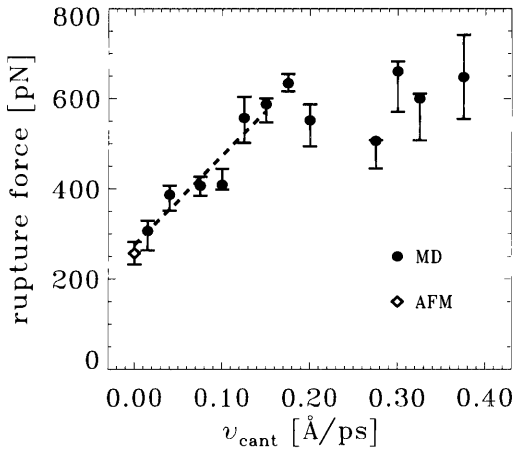


Fig. 5. Theory vs. experiment: rupture forces computed from rupture simulations at various time scales (various pulling velocities v_{cant}) ranging from one nanosecond ($v_{\mathrm{cant}} = 0.015$ Å/ps) to 40 picoseconds ($v_{\mathrm{cant}} = 0.375$ Å/ps) (black circles) compare well with the experimental value (open diamond) when extrapolated linearly (dashed line) to the experimental time scale of milliseconds.

However, one significant difference between the AFM experiment and its simulations cannot be avoided at present: Whereas the AFM experiment

takes place at a millisecond time scale, our simulations had to be completed within the nanosecond time scale. So, in fact, in the simulation the pulling velocity had to be chosen about six orders of magnitude larger than in the AFM experiment!

In a first attempt to bridge these six orders of magnitude, we performed a series of rupture force simulations using pulling velocities ranging from 0.375 to 0.015 Å/ps. As can be seen in Fig. 5, we observed a linear dependency of the computed rupture forces in the velocity range between 0.15 and 0.015 Å/ps. This suggests that simple friction dominates the non-equilibrium effects in this regime described by a friction coefficient of 20 pN s/m. A simple linear extrapolation of the computed rupture forces to the experimental time scale shows agreement between theory and experiment. Clearly, this first step has not yet solved the question how to bridge the six orders of magnitude gap between theory and experiment (cf. also [53]). To answer that question, a better understanding of the physics of rupture experiments using simplified models on the one hand (cf., e.g., Ref. [54]) and, on the other hand, a careful analysis of the atomic processes which cause the velocity dependent rupture forces is necessary.

One of the results of an MD rupture simulation is the pulling force as a function of time or cantilever position $z_{\text{cant}}(t)$, called the *force profile*. Figure 6 shows an example, derived from an extended 1 ns-simulation, where a pulling velocity of 0.015 Å/ps was used. The apparent multitude of force maxima mirrors the complexity of the energy landscape traversed by the biotin on its way out of the binding pocket. The peaks of this force profile can be attributed to the rupture and formation of individual hydrogen bonds and water bridges shown in the snapshots of Fig. 7, which characterize the main steps of the rupture process. The rupture forces in Fig. 5 are the maxima of the corresponding force profiles.

We will not discuss here in detail our atomic model of the unbinding process derived from our simulations and sketched in Fig. 7, but restrict ourselves to two unexpected features. One is that the rupture of the initially very strong hydrogen bonds between the ligand and the residues of the binding pocket (Fig. 7 A) does not entail immediate unbinding. Rather, the complex is stabilized by a transient network of water bridges and other transient hydrogen bonds, which form during the unbinding process (Fig. 7 B and C). Only after subsequent rupture of these hydrogen bonds the maximum force — the rupture force — is reached and the biotin rapidly moves out of the entry of the binding pocket (Fig. 7 D). As another feature we observed, towards the end of the unbinding process, a second force maximum, which we attribute to a strong transient hydrogen bond and several water bridges between biotin and the entry of the binding pocket (Fig. 7 E). Crossing of that second barrier, which cannot yet be resolved in the AFM experiment, completes the unbinding process.

In summary, our simulations provided detailed insight into the complex mechanisms of streptavidin-biotin rupture. They attribute the binding force

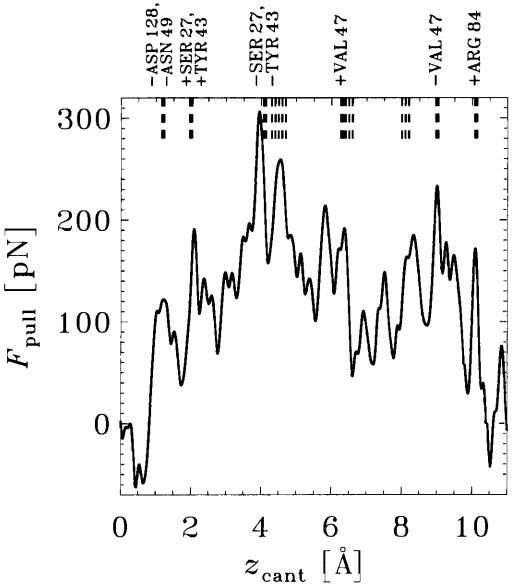


Fig. 6. Force profile obtained from a one nanosecond simulation of streptavidin–biotin rupture showing a series of subsequent force peaks; most of these can be related to the rupture of individual microscopic interactions such as hydrogen bonds (bold dashed lines indicate their time of rupture) or water bridges (thin dashed lines).

to a network of hydrogen bonds between the ligand and the binding pocket and show that water bridges substantially enhance the stability of the complex. Good agreement with experimental results was obtained. Further 'force simulations' of various systems, e.g., an antigen-antibody complex, are in progress.

4 Conformational Flooding

The previous application — in accord with most MD studies — illustrates the urgent need to further push the limits of MD simulations set by todays computer technology in order to bridge time scale gaps between theory and either experiments or biochemical processes. The latter often involve conformational motions of proteins, which typically occur at the microsecond to millisecond range. Prominent examples for functionally relevant conformational motions

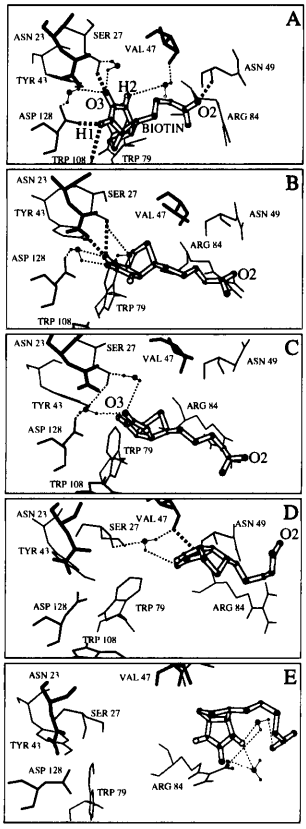


Fig. 7. 'Snapshots' of rupture taken (A) at the start of the simulation ($z_{\text{cant}} = 0$), (B) at $z_{\text{cant}} = 2.8\,\text{Å}$, (C) at $z_{\text{cant}} = 4.1\,\text{Å}$, (D) at $z_{\text{cant}} = 7.1\,\text{Å}$, and (E) at $z_{\text{cant}} = 10.5\,\text{Å}$. The biotin molecule is drawn as a ball-and-stick model within the binding pocket (lines). The bold dashed lines show hydrogen bonds, the dotted lines show selected water bridges.

are the opening and closing of ion channels or, as proposed by Griffith [55] and Prusiner [56], pathogenic conformational transitions in prion proteins, the putative agents of mad cow and Creutzfeldt-Jacob diseases. Conformational motions often involve a complex and concerted rearrangement of many atoms in a protein from its initial state into a new conformation. These rearrangements, called conformational transitions, exhibit a multi-rate behaviour, which is is captured by the concept of "hierarchical conformational substates" introduced by Hans Frauenfelder [57]. According to that concept the free energy landscape of a protein exhibits a large number of nearly isoenergetic minima corresponding to the *conformational substates*, which are separated by barriers of different height [58].

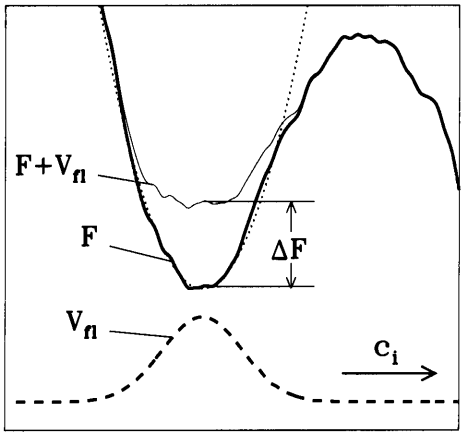


Fig. 8. 'Conformational flooding' lowers free energy barriers of conformational transitions and thus accelerates such transitions. The figure shows a one dimensional cut through the high dimensional free energy landscape F (bold line) along a particular conformational coordinate c_i. During an MD simulation the protein remains in the initial configuration (local minimum in the free energy), since the high barrier to the right cannot be overcome on an MD time scale. However, the MD simulation can serve to approximate the free energy harmonically in the vicinity of the initial configuration (dotted line) in order to derive an artificial 'flooding potential' V_{fl} (dashed line). Inclusion of this potential (thin line) in subsequent MD simulations reduces the barrier height by an amount of ΔF and thereby destabilizes the initial configuration.

Figure 8 shows a one-dimensional sketch of a small fraction of that energy landscape (bold line) including one conformational substate (minimum) as well as, to the right, one out of the typically huge number of barriers separating this local minimum from other ones. Keeping this picture in mind the conformational dynamics of a protein can be characterized as "jumps" between these local minima. At the MD time scale below nanoseconds only very low barriers can be overcome, so that the studied protein remains in or close to its initial conformational substate and no predictions of slower conformational transitions can be made.

In order to make such predictions possible, we have developed the *conformational flooding* (CF) method, which *accelerates* conformational transitions [59] and thereby brings them into the scope of MD simulations ("flooding simulations"). The method is a generalization of the "local elevation method" [60] in that it rests on a quasi harmonic model for the free energy landscape in the vicinity of the minimum representing the initial (known) conformational state. This model is derived from an ensemble of structures generated by a conventional MD simulation as will be described below and is shown in Fig. 9. From that model a "flooding potential" V_{fl} is constructed (dashed line in Fig. 8), which, when subsequently included into the poten-

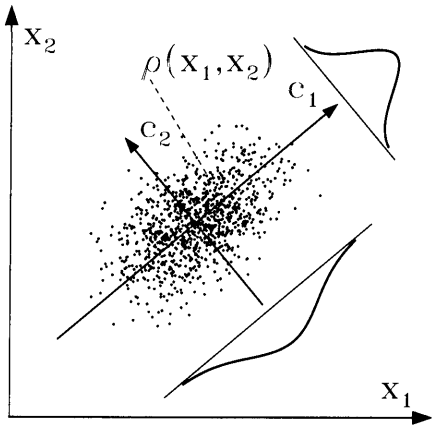


Fig. 9. Two-dimensional sketch of the 3N-dimensional configuration space of a protein. Shown are two Cartesian coordinates, x_1 and x_2, as well as two conformational coordinates (c_1 and c_2), which have been derived by principle component analysis of an ensemble ("cloud" of dots) generated by a conventional MD simulation, which approximates the configurational space density ρ in this region of configurational space. The width of the two Gaussians describe the size of the fluctuations along the configurational coordinates and are given by the eigenvalues λ_i.

tial energy function of the system, raises the minimum under consideration (thin line in Fig. 8) and thereby lowers the surrounding free energy barriers by an amount ΔF without severely modifying the barriers themselves. As a result, transitions over these barriers are accelerated by approximately the Boltzmann factor $\exp(\frac{\Delta F}{k_B T})$. In detail, the following steps are necessary to perform a CF simulation:

Step 1: A short conventional MD simulation (typically extending over a few 100 ps) is performed to generate an ensemble of protein structures $\{\mathbf{x} \in \mathcal{R}^{3N}\}$ (each described by N atomic positions), which characterizes the initial conformational substate. The 2-dimensional sketch in Fig. 9 shows such an ensemble as a cloud of dots, each dot $\mathbf{x}$ representing one "snapshot" of the protein.

Step 2: This ensemble is subjected to a "principal component analysis" (PCA) [61] by diagonalizing the covariance matrix $\mathbf{C} \in \{\mathcal{R}^{3N} \times \mathcal{R}^{3N}\}$,

$$\mathbf{C} := \langle (\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T \rangle \quad \text{with} \quad \bar{\mathbf{x}} = \langle \mathbf{x} \rangle$$

i.e.,

$$\mathbf{C} = \mathbf{Q}^T \Lambda^{-1} \mathbf{Q}$$

with orthonormal $\mathbf{Q}$ and $\Lambda = (\delta_{ij}\lambda_i) \in \{\mathcal{R}^{3N} \times \mathcal{R}^{3N}\}$, where $\langle \ldots \rangle$ denotes an average over the ensemble $\{\mathbf{x}\}$.

Step 3: The eigenvectors of $\mathbf{C}$ define $3N - 6$ collective coordinates (quasi

particles) $\mathbf{q} := \mathbf{Q}(\mathbf{x} - \bar{\mathbf{x}})$, where we have eliminated the six rotational and translational degrees of freedom. From these $3N - 6$ degrees of freedom we select a number $m < 3N - 6$ *conformational coordinates* $\mathbf{c} = (c_1, \ldots, c_m)^T$ associated to the largest eigenvalues. Thus, the conformational coordinates cover most of the atomic fluctuations occurring at the 100 ps time scale. These m degrees of freedom are expected to dominate (not necessarily exclusively) conformational motion also at slower time scales [62, 63, 64].

Step 4: This PCA defines a *multivariate Gaussian model* $\tilde{\rho}^{\mathbf{c}}$,

$$\tilde{\rho}^{\mathbf{c}}(\mathbf{c}) \propto \exp\left[-\mathbf{c}^T \Lambda_c \mathbf{c}/2\right]$$

of the conformational space density $\rho(\mathbf{c})$, from which the quasi harmonic approximation of the energy landscape,

$$\tilde{F}(\mathbf{c}) = -k_B T \ln\left[\tilde{\rho}(\mathbf{c})\right] = \frac{1}{2} k_B T \mathbf{c}^T \Lambda_c \mathbf{c}$$

is derived (see Ref. [59]).

Step 5: From that model of the current substate we construct the *flooding potential* V_{fl} of strength E_{fl},

$$V_{\mathrm{fl}} = E_{\mathrm{fl}} \ \exp\left[-\frac{1}{2}\frac{k_B T \mathbf{c}^T \Lambda_c \mathbf{c}}{E_{\mathrm{fl}}}\right],$$

which is included in a subsequent MD simulation within the energy function used in the conventional MD simulation before (see Fig. 8), thereby causing the desired acceleration of transitions.

As a sample application we describe simulations suggesting possible conformational transitions of the protein BPTI (Bovine Pancreatic Trypsin Inhibitor) at a time scale of several 100 nanoseconds (see Fig. 10). First we carried out a conventional MD simulation of 500 ps duration (no explicit solvent included), during which the protein remained in its initial conformational substate CS 1. The upper left part of the figure shows several snapshots of the backbone taken from that simulation; the lower left shows a projection of the 500 ps trajectory onto the two conformational coordinates with largest eigenvalues (corresponding to Fig. 9). From that ensemble we constructed a flooding potential as described above (dashed contour lines, superimposed to the CS 1-trajectory, bottom right). The flooding potential was subsequently switched on and rapidly induced a conformational transition (to the right in the figure) into another energy minimum, CS 2. After switching off the flooding potential the new conformational state of the protein remained stable, indicating that, indeed, the new minimum is separated from CS 1 by a large energy barrier. Using multi-dimensional transition state theory [59] we could estimate that in an conventional (i.e., unperturbed) MD simulation that conformational transition would have been observed only after several hundred

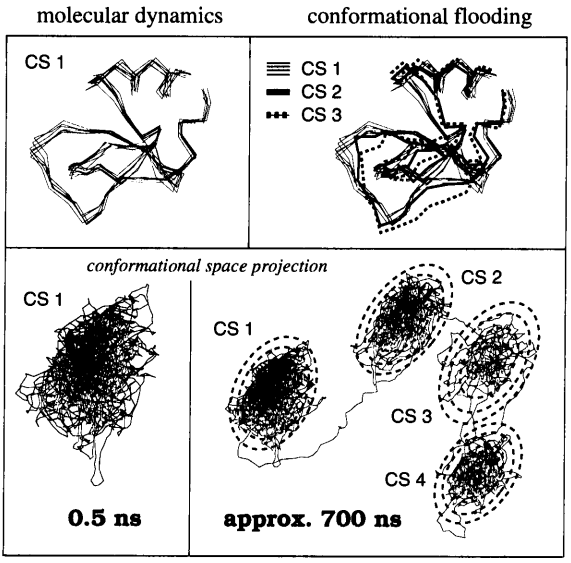


Fig. 10. 'Conformational flooding' accelerates conformational transitions and makes them accessible for MD simulations. Top left: snapshots of the protein backbone of BPTI during a 500 ps-MD simulation. Bottom left: a projection of the conformational coordinates contributing most to the atomic motions shows that, on that MD time scale, the system remains in its initial configuration (CS 1). Top right: 'Conformational flooding' forces the system into new conformations after crossing high energy barriers (CS 2, CS 3, ...). Bottom right: The projection visualizes the new conformations; they remain stable, even when the applied flooding potentials (dashed contour lines) is switched off.

nanoseconds. As shown in Fig. 10, the CF method can be applied iteratively to systematically search for further conformational substates, CS 3, CS 4 etc. The upper right part of the figure shows the backbone configuration of BPTI corresponding to the new substates.

MD simulations are valuable tools if one wants to gain detailed insight into *fast* dynamical processes of proteins and other biological macromolecules at atomic resolution. But since conventional MD simulations are confined to the study of very fast processes, conformational flooding represents a complementary and powerful tool to predict and understand also slow conformational motions. Another obvious application is an enhanced refinement of Xray- or NMR-structures.

References

1. M. Levitt and Shneior Lifson. Refinement of protein conformation using a macromolecular energy minimization procedure. *J. Mol. Biol.*, 46:269–279, 1969.

2. J. A. McCammon, B. R. Gelin, and M. Karplus. Dynamics of folded proteins. *Nature (London)*, 267:585–590, 1977.
3. W. F. van Gunsteren and H. J. C. Berendsen. Algorithms for macromolecular dynamics and constraint dynamics. *Mol. Phys.*, 34(5):1311–1327, 1977.
4. Olle Edholm, Oliver Berger, and Fritz Jähnig. Structure and fluctuations of bacteriorhodopsin in the purple membrane. *J. Mol. Biol.*, 250:94–111, 1995.
5. M. Levitt and R. Sharon. Accurate simulation of protein dynamics in solution. *Proc. Natl. Acad. Sci. USA*, 85:7557–7561, 1988.
6. Walter Nadler, Axel T. Brünger, Klaus Schulten, and Martin Karplus. Molecular and stochastic dynamics of proteins. *Proc. Natl. Acad. Sci. USA*, 84:7933–7937, Nov. 1987.
7. H. Kovacs, A.E. Mark, J. Johansson, and W.F. van Gunsteren. The effect of environment on the stability of an integral membrane helix: Molecular dynamics simulations of surfactant protein C in chloroform, methanol and water. *J. Mol. Biol.*, 247:808–822, 1995.
8. G.H. Peters, D.M.F van Aalten, O. Edholm, S. Toxvaerd, and R. Bywater. Dynamics of proteins in different solvent systems: Analysis of essential motion in lipases. *Biophys. J.*, 71:2245–2255, 1996.
9. M. C. Nuss, W. Zinth, W. Kaiser, E. Kölling, and D. Oesterhelt. Femtosecond spectroscopy of the first events of the photochemical cycle in bacteriorhodopsin. *Chem. Phys. Lett.*, 117(1):1–7, 1985.
10. Feng Zhou, Andreas Windemuth, and Klaus Schulten. Molecular-dynamis study of the proton pump cycle of bacteriorhodopsin. *Biochem.*, 32(9):2291–2306, 1993.
11. B. Leimkuhler and R. D. Skeel. *J. Comp. Phys.*, 112:117, 1994.
12. B.J. Leimkuhler, S. Reich, and R. D. Skeel. Integration methods for molecular dynamics. In *Mathematical approaches to biomolecular structure and dynamics*, Seiten 161–185, New York, 1996. Springer.
13. R. D. Skeel, G. H. Zhang, and T. Schlick. A family of symplectic integrators: Stability, accuracy, and molecular dynamics applications. *SIAM J. Scient. COMP.*, 18:203–222, 1997.
14. A. Ahmad and L. Cohen. A numerical integration scheme for the N-body gravitational problem. *J. Comp. Phys.*, 12:389–402, 1973.
15. W. B. Streett, D. J. Tildesley, and G. Saville. Multiple time step methods in molecular dynamics. *Mol. Phys.*, 35:639–648, 1978.
16. R. C. Y. Chin, G. W. Hedstrom, and F. A. Howes. *Considerations on Solving Problems with Multiple Scales.* Academic Press, Orlando, Florida, 1985.
17. Andreas Windemuth. Dynamiksimulation von Makromolekülen. Diplomarbeit, Technical University of Munich, Physics Department, T 30, James-Franck-Street, 8046 Garching, August 1988.
18. Mark E. Tuckerman, Glenn J. Martyna, and Bruce J. Berne. Molecular dynamics algorithm for condensed systems with multiple time scales. *J. Chem. Phys.*, 93(2):1287–1291, Jul. 1990.
19. Helmut Grubmüller, Helmut Heller, Andreas Windemuth, and Klaus Schulten. Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Mol. Sim.*, 6:121–142, 1991.
20. Mark E. Tuckerman and Bruce J. Berne. Molecular dynamics algorithm for multiple time scales: Systems with disparate masses. *J. Chem. Phys.*, 94(2):1465–1469, January 1991.

21. Mark E. Tuckerman, Bruce J. Berne, and Glenn J. Martyna. Molecular dynamics algorithm for multiple time scales: Systems with long range forces. *J. Chem. Phys.*, 94(10):6811–6815, May 1991.
22. Robert D. Skeel, Jeffrey J. Biesiadecki, and Daniel Okunbor. Symplectic integration for macromolecular dynamics. In *Proceedings of the International Conference Computation of Differential Equations and Dynamical Systems.* World Scientific Publishing Co., 1992. in press.
23. Robert D. Skeel and Jeffrey J. Biesiadecki. Symplectic integration with variable stepsize. *Ann. Num. Math.*, 1:191–198, 1994.
24. Helmut Grubmüller. *Molekulardynamik von Proteinen auf langen Zeitskalen.* Doktorarbeit, Technische Universität München, Germany, Jan. 1994.
25. D. Okunbor and R. D. Skeel. Explicit canonical methods for Hamiltonian systems. Working document, Numerical Computing Group, University of Illinois at Urbana-Champaign, 1991.
26. Andreas Windemuth. *Advanced Algorithms for Molecular Dynamics Simulation: The Program PMD.* ACS Books, 1995.
27. D. D. Humphreys, R. A. Friesner, and B. J. Berne. Simulated annealing of a protein in a continuum solvent by multiple-time-step molecular dynamics. *J. Phys. Chem.*, 99:10674–10685, 1995.
28. P. Procacci, T. Darden, and M. Marchi. A very fast molecular dynamics method to simulate biomolecular systems with realistic electrostatic interactions. *J. Phys. Chem.*, 100:10464–10468, 1996.
29. S. J. Stuart, R. Zhou, and B. J. Berne. Molecular dynamics with multiple time scales: The selection of efficient reference system propagators. *J. Chem. Phys.*, 105:1426–1436, 1996.
30. R. Zhou and B. J. Berne. A new molecular dynamics method combining the reference system propagator algorithm with a fast multipole method for simulating proteins and other complex systems. *J. Phys. Chem.*, 103:9444–9459, 1995.
31. T. Schlick, E. Bartha, and M. Mandziuk. Biomolecular dynamics at long timesteps: Bridging the timescale gap between simulation and experimentation. *Ann. Rev. Biophys. Biom. Structure*, 26:181–222, 1997.
32. Andrew. W. Appel. An efficient program for many-body simulation. *SIAM J. Sci. Stat. Comput.*, 6(1):85–103, January 1985.
33. Josh Barnes and Piet Hut. A hierarchical $o(n \log n)$ force-calculation algorithm. *Nature (London)*, 324:446–449, December 1986.
34. L. Greengard and V. Rokhlin. On the evaluation of electrostatic interactions in molecular modeling. *Chem. Scr.*, 29A:139–144, 1989.
35. James F. Leathrum and John A. Board. The parallel fast multipole algorithm in three dimensions. Technical report, Dept. of Electrical Engineering, Duke University, Durham, 1992.
36. C. Niedermeier and P. Tavan. A structure adapted multipole method for electrostatic interactions in protein dynamics. *J. Chem. Phys.*, 101:734–748, 1994.
37. Christoph Niedermeier. *Modellierung elektrostatischer Wechselwirkungen in Proteinen: Eine strukturadaptierte Multipolmethode.* Doktorarbeit, Ludwig-Maximilians-Universität, München, Germany, 1995.
38. C. Niedermeier and P. Tavan. Fast version of the structure adapted multipole method — efficient calculation of electrostatic forces in protein dynamics. *Mol. Sim.*, 17:57–66, 1996.

39. B. A. Luty, I. G. Tironi, and W. F. van Gunsteren. Lattice-sum methods for calculating electrostatic interactions in molecular simulations. *J. Chem. Phys.*, 103:3014–3021, 1995.
40. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. The smooth particle mesh ewald method. *J. Chem. Phys.*, 103:8577, 1995.
41. Brock A. Luty, Ilario G. Tironi, and Wilfried F. van Gunsteren. Lattice-sum methods for calculating electrostatic interactions in molecular simulations. *J. Chem. Phys.*, 103:3014–3021, 1995.
42. Brock A. Luty and Wilfried F. van Gunsteren. Calculating electrostatic interactions using the particle–particle particle–mesh method with nonperiodic long-range interactions. *J. Phys. Chem.*, 100:2581–2587, 1996.
43. Bernhard R. Brooks, Robert E. Bruccoleri, Barry D. Olafson, David J. States, S. Swaminathan, and Martin Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.*, 4(2):187–217, 1983.
44. Charles L. Brooks III, B. Montgomery Pettitt, and Martin Karplus. Structural and energetic effects of truncating long ranged interactions in ionic and polar fluids. *J. Chem. Phys.*, 83(11):5897–5908, December 1985.
45. Richard J. Loncharich and Bernard R. Brooks. The effects of truncating long-range forces on protein dynamics. *Proteins*, 6:32–45, 1989.
46. M. Eichinger, H. Grubmüller, H. Heller, and P. Tavan. FAMUSAMM: An algorithm for rapid evaluation of electrostatic interaction in molecular dynamics simulations. *J. Comp. Chem.*, 18:1729–1749, 1997.
47. M. Eichinger. Paralleler schneller Multipolalgorithmus mit Mehrschrittverfahren für Molekulardynamiksimulationen. Diplomarbeit, Ludwig-Maximilians-Universität, München, Germany, 1995.
48. M. Eichinger, H. Grubmüller, and H. Heller. *User Manual for EGO_VIII, Release 2.0.* Theoretische Biophysik, Institut für Medizinische Optik, Ludwig-Maximilians-Universität, Theresienstr. 37, D-80333 München, Germany, 1995. Electronic access: http://www.imo.physik.uni-muenchen.de/ego.html.
49. Helmut Grubmüller, Berthold Heymann, and Paul Tavan. Ligand binding: Molecular mechanics calculation of the streptavidin-biotin rupture force. *Science*, 271(5251):997–999, 1996.
50. E.-L. Florin, V. T. Moy, and H. E. Gaub. Adhesion forces between individual ligand-receptor pairs. *Science*, 264:415–417, Apr. 15 1994.
51. N. M. Green. Avidin. *Adv. Protein Chem.*, 29:85, 1975.
52. S. Miyamoto and P. A. Kollman. Absolute and relative binding free energy calculations of the interaction of biotin and its analogs with streptavidin using molecular dynamics/free energy perturbation approaches. *Proteins*, 16:226–245, 1993.
53. S. Izrailev, S. Stepaniants, M. Balsera, Y. Oono, and K. Schulten. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys. J.*, 72:1568–1581, 1997.
54. Evan Evans and Ken Ritchie. Dynamic strength of molecular adhesion bonds. *Biophys. J.*, 72:1541, 1997.
55. J. S. Griffith. *Nature (London)*, 215:1043–1044, 1967.
56. S. B. Prusiner. *Science*, 252:1515–1522, 1991.
57. Hans Frauenfelder, Sthephen G. Sligar, and Peter G. Wolynes. The energy landscape and motions of proteins. *Science*, 254:1598–1603, 1991.

58. R. H. Austin, K. W. Beeson, L. Eisenstein, H. Frauenfelder, and I.C. Gunsalus. Dynamics of ligand binding to myoglobin. *Biochem.*, 14(24):5355–5373, 1975.
59. Helmut Grubmüller. Predicting slow structural transitions in macromolecular systems: conformational flooding. *Phys. Rev. E*, 52:2893, 1995.
60. T. Huber, A. E. Torda, and W. F. van Gunsteren. Local elevation: A method for improving the searching properties of molecular dynamics simulation. *J. of Computer-Aided Molecular Design*, 8:695–708, 1994.
61. J. C. Gower. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53:325, 1966.
62. A. Amadei, A. B. M. Linssen, and H. J. C. Berendsen. Essential dynamics of proteins. *Proteins*, 17:412–425, 1993.
63. Steven Hayward, Akio Kitao, and Nobuhiro Gō. Harmonic and anharmonic aspects in the dynamics of BPTI: A normal mode analysis and principal component analysis. *Physica Scripta*, 3:936–943, 1994.
64. H. Grubmüller, N. Ehrenhofer, and P. Tavan. Conformational dynamics of proteins: Beyond the nanosecond time scale. In M. Peyard, editor, *Proceedings of the Workshop 'Nonlinear Excitations in Biomolecules', May 30–June 4, 1994, Les Houches (France)*, Seiten 231–240. Centre de Physique des Houches (France), Springer-Verlag, 1995.

Computation of Essential Molecular Dynamics by Subdivision Techniques

Peter Deuflhard[1,2], Michael Dellnitz[3], Oliver Junge[3], and Christof Schütte[1,2]

[1] Konrad Zuse Zentrum Berlin, Takustr. 7, 14195 Berlin, Germany
[2] Freie Universität Berlin, Fachbereich Mathematik, Arnimallee 2–6, 14195 Berlin, Germany
[3] Mathematisches Institut, Universität Bayreuth, 95440 Bayreuth

Dedicated to Prof. Dr. Dr. h.c. Roland Bulirsch on the occasion of his 65th birthday.

Abstract. The paper presents basic concepts of a new type of algorithm for the numerical computation of what the authors call the *essential dynamics* of molecular systems. Mathematically speaking, such systems are described by Hamiltonian differential equations. In the bulk of applications, individual trajectories are of no specific interest. Rather, time averages of physical observables or relaxation times of conformational changes need to be actually computed. In the language of dynamical systems, such information is contained in the natural invariant measure (infinite relaxation time) or in almost invariant sets ("large" finite relaxation times). The paper suggests the direct computation of these objects via eigenmodes of the associated Frobenius-Perron operator by means of a multilevel subdivision algorithm. The advocated approach is different from both Monte-Carlo techniques on the one hand and long term trajectory simulation on the other hand: in our setup long term trajectories are replaced by short term sub-trajectories, Monte-Carlo techniques are connected via the underlying Frobenius-Perron structure. Numerical experiments with the suggested algorithm are included to illustrate certain distinguishing properties.

1 Introduction

The classical microscopic description of molecular processes leads to a mathematical model in terms of Hamiltonian differential equations. In principle, the discretization of such systems permits a simulation of the dynamics. However, as will be worked out below in Section 2, both forward and backward numerical analysis restrict such simulations to only short time spans and to comparatively small discretization steps. Fortunately, most questions of chemical relevance just require the computation of *averages* of physical observables, of *stable conformations* or of *conformational changes*. The computation of averages is usually performed on a *statistical physics* basis. In the subsequent Section 3 we advocate a new computational approach on the basis of the mathematical theory of *dynamical systems*: we directly solve a

discretized eigenvalue problem for the Frobenius–Perron operator without any physical a-priori assumptions. In this framework, the computation of averages corresponds to the approximation of *invariant measures* and *invariant sets* associated with eigenvalue $\lambda = 1$ – a fact, which is well-known for quite a while. In a similar way, conformations of molecular systems correspond to *almost invariant sets* related to eigenmodes of the Frobenius–Perron operator for $\lambda \approx 1$ – a fact, which has just recently been discovered [6]. On the basis of this insight, a multilevel box discretization (or subdivision technique) is applied to the Frobenius-Perron eigenvalue problem, which only requires short term subtrajectory computations. Finally, an illustrative numerical example is included in Section 4. This paper is a shortened and improved version of the more elaborate Technical Report [8].

2 Trajectory Simulation

In classical MD (cf. textbook [1]) a molecule is modelled by a Hamiltonian function

$$H(q,p) = \tfrac{1}{2} p^T M^{-1} p + V(q),$$

where q and p are the corresponding positions and momenta of the atoms, M the diagonal mass matrix, and V a differentiable potential. The Hamiltonian H is defined on the phase space $\Gamma \subset \mathbb{R}^{6N}$. The corresponding canonical equations of motion

$$\dot{q} = M^{-1} p, \qquad \dot{p} = -\operatorname{grad} V \tag{1}$$

describe the dynamics of the molecule. The formal solution of (1) with initial state $x_0 = (q(0), p(0))$ is given by $x(t) = (q(t), p(t)) = \Phi^t x_0$, where Φ^t denotes the flow. In the process of one-step numerical integration of (1) we replace the flow by a *discrete flow* Ψ^τ, so that

$$x_{k+1} = \Psi^\tau x_k \quad \Rightarrow \quad x_k = \left(\Psi^\tau\right)^k x_0,$$

with stepsize τ (assumed to be constant, for the time being). An important feature of molecular processes is that *long term predictions* over periods much longer than the applied time steps are required.

Forward Analysis In this type of analysis, we are interested in the propagation of initial perturbations δx_0 along the flow Φ^t of (1), i.e., in the growth of the perturbations $\delta x(t; x_0) = \Phi^t(x_0 + \delta x_0) - \Phi^t x_0$. The *condition number* $\kappa(t)$ may be defined as the worst case error propagation factor (cf. textbook [4]), so that, in first order perturbation analysis and with a suitable norm $|\cdot|$:

$$|\delta x(t; x_0)| < \kappa(t) |\delta x_0| \quad \text{for all } x_0.$$

Note that this number $\kappa(t)$ is *independent of any discretization.* From this point of view, numerical integration is reasonable only over time intervals $[0, T]$ with $\kappa(T)$ sufficiently small compared to expected input errors. Integrable systems (such as the popular Kepler problem) are characterized by $\kappa(T) \sim T$ [10], which allows for quite long term simulations. In real life MD problems, however, κ is exponentially increasing. This fact is illustrated in Fig. 1, where test simulations for the Butane molecule are presented. Note that here we are discussing a property of the *continuous* Hamiltonian system. Therefore, in order to avoid any discretization error effects unusually small time steps ($\tau = 0.005$ fs) have been chosen within the Verlet scheme. As can be seen, global propagation of physically negligible initial perturbations totally spoils any information after a time span, which is significantly shorter than time spans of physical interest.

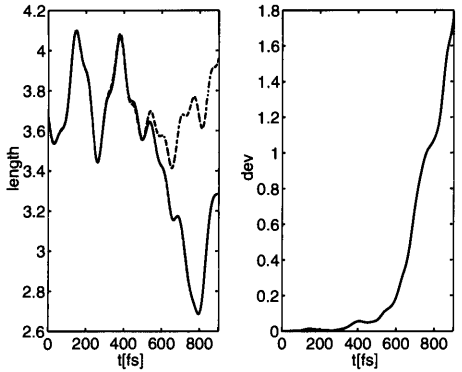


Fig. 1. Comparison of two different dynamical simulations for the Butane molecule: Verlet discretization with stepsize $\tau = 0.005$fs. Initial spatial deviation: 10^{-4}Å. Left: Evolutions of the total length (=distance between the first and the last carbon atom) of the molecule (in Å). Right: Spatial deviation (in Å) of the two trajectories versus time.

Backward Analysis In this type of analysis, the *discrete* solution is regarded as an exact solution of a perturbed problem. In particular, backward analysis of *symplectic discretizations* of Hamiltonian systems (such as the popular Verlet scheme) has recently achieved a considerable amount of attention (see [17, 8, 3]). Such discretizations give rise to the following feature: the discrete solution of a Hamiltonian system is "exponentially close" to the exact solution of a *perturbed* Hamiltonian system, in which, for consistency order p and stepsize τ, the perturbed Hamiltonian has the form [11, 3]

$$\tilde{H} = H + \sum_{k=0}^{N} \tau^{p+k} H_k \qquad (2)$$

This means that the discrete solution nearly conserves the Hamiltonian $\tilde{H}$ and, thus, conserves H up to $\mathcal{O}(\tau^p)$. If H is analytic, then the truncation index N in (2) is arbitrary. In general, however, the above formal series diverges as $N \to \infty$. The term "exponentially close" may be specified by the following theorem.

Theorem 1 ([8]). *Let H be analytic. There exists some $\tau_* > 0$, so that for all $\tau < \tau_*$ the numerical solution $x_k = (\Psi^\tau)^k x_0$ and the exact solution $\tilde{x}$ of the perturbed system $\tilde{H}$ (the sum being truncated after $N = \mathcal{O}(1/\tau)$ terms) with $\tilde{x}(0) = x_0$ remain exponentially close in the sense that*

$$x_k - \tilde{x}(k\tau) = \mathcal{O}(e^{-c/\tau})$$

with some problem-dependent constant $c > 0$ over a time interval $T = \mathcal{O}(1/\tau)$, i.e., for all $k\tau < T$.

In applications, one is often interested in approximating time averages over a time interval $[0, T]$ via associated mean values of x_k, $k = 1 \ldots T/\tau$. For T (or τ) small enough, the above backward analysis may lead to much better error estimates than the worst case estimates of forward analysis.

In fact, numerical observations show that the average of the *total energy* is nearly constant over rather long time spans for large stepsizes, say $\tau \approx 1$ fs. However, this desirable property does not carry over to other averages, where stepsizes much smaller than desirable ($\tau \ll 1$ fs) may be required. This phenomenon is illustrated in Fig. 2. Here the time average of the molecular length of Butane over $T = 200$ ps requires a discretization with $\tau_* \approx 10^{-2}$fs in comparison with physically interesting stepsizes $\tau \approx 1 \ldots 10$fs. This illustrates, that —even though the discrete system is an only slightly perturbed Hamiltonian system— a successful control of energy is not sufficient for the reliable computation of other observables.

Summarizing, from a mathematical point of view, both forward and backward analysis lead to the insight that long term trajectory simulation should be avoided even with symplectic discretizations. Rather, in the spirit of *multiple* as opposed to *single* shooting (cf. BULIRSCH [4, 18]), only *short term* trajectories should be used to obtain reliable information.

Remark. The authors are well aware of the commonly shared belief within the physics MD community that the apparent chaos does not destroy all long term information, but only irrelevant information; in particular, average quantities are supposed to be not sensitive to numerical perturbations. We do not aim at deciding about this fundamental question here.

Essential Dynamics In most applications details of individual MD trajectories are of only minor interest. An illustrative example due to GRUBMÜLLER [10] is documented in Figure 3. It describes the dynamics of a polymer chain of 100 CH_2 groups. Possible stepsizes for numerical integration are confined

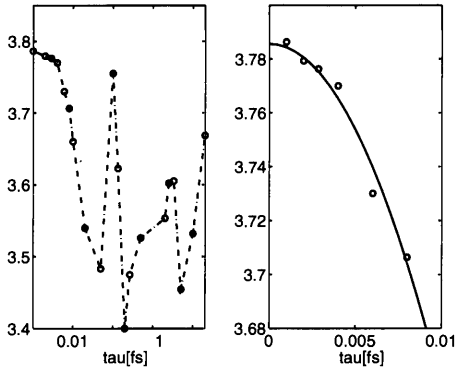


Fig. 2. Left: Time average (over $T = 200$ps) of the molecular length of Butane versus discretization stepsize τ for the Verlet discretization. Right: Zoom of the asymptotic domain ($\tau < 10^{-2}$ fs) and quadratic fit.

by the fast oscillations ($\tau < 10$fs). Time scales of physical interest range between 10^3 and 10^5 ps, which is a factor $10^5 - 10^7$ larger. The figure presents six different zoom levels, each of which scales up in time by a factor of 10. On the small time scales (upper levels) the dynamical behavior is characterized by nonlinear *oscillations* around certain vague "equilibrium positions". On larger and larger time scales these oscillations become less and less important. On the largest time scale (lowest level) we observe an "essential" dynamical behavior as a kind of flip–flop between two "conformations".

As a consequence of this observation, the *essential* dynamics of the molecular process could as well be modelled by *probabilities* describing mean durations of stay within different conformations of the system. This idea is not new, cf. [10]. Even the phrase "essential dynamics" has already been coined in [2]: it has been chosen for the reformulation of molecular motion in terms of its "almost invariant" degrees of freedom. But unlike the former approaches, which aim in the same direction, we herein advocate a different line of method: we suggest to *directly* attack the computation of the conformations and their stability time spans, which means some global approach clearly differing from any kind of statistical analysis based on long term trajectories.

3 Dynamical Systems Approach

We restrict our attention to symplectic one-step discretizations of (1), which leads to discrete dynamical systems of the form

$$x_{j+1} = f(x_j), \quad j = 0, 1, 2, \ldots, \tag{3}$$

where m denotes the number of time steps between two updates of the discrete mapping $f = (\Psi^\tau)^m : \Gamma \to \Gamma$.

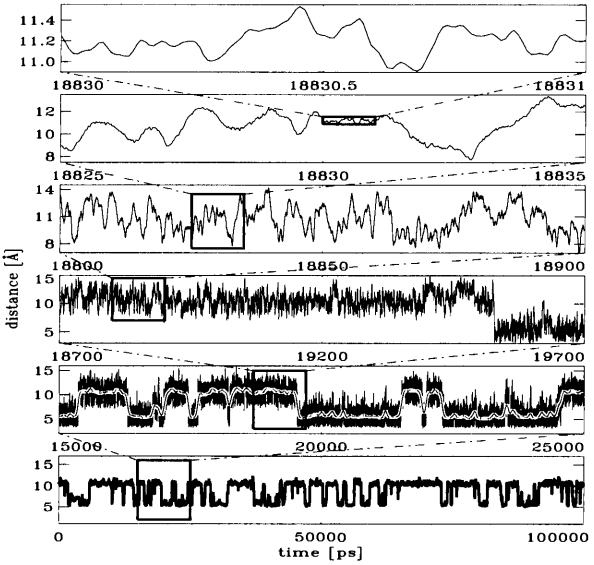


Fig. 3. MD simulation of a polymer chain of 100 CH_2 groups due to [10]. The dynamics of the distance between two CH_2–groups (# 12 and # 36). The series of plots illustrates the oscillations of the distance at time scales increasing by a zoom factor of 10 at each level.

The long term behavior of any system (3) is described by so-called *invariant measures*: a probability measure μ is invariant, iff $\mu(f^{-1}(B)) = \mu(B)$ for all measurable subsets $B \subset \Gamma$. The associated *invariant sets* are defined by the property that $B = f^{-1}(B)$. Throughout the paper we will restrict our attention to so-called SBR-measures (cf. [16]), which are robust with respect to stochastic perturbations. Such measures are the only ones of physical interest. In view of the above considerations about modelling in terms of probabilities, the following interpretation will be crucial: given an invariant measure μ and a measurable set $B \subset \Gamma$, the value $\mu(B)$ may be understood as the probability of finding the system within B.

3.1 Frobenius-Perron Eigenvalue Problem

A key observation for our purposes here is that the numerical computation of invariant measures is equivalent to the solution of an *eigenvalue problem* for the so-called *Frobenius-Perron operator* $P : \mathcal{M} \to \mathcal{M}$ defined on the set $\mathcal{M}$ of probability measures on Γ by virtue of

$$(P\mu)(B) = \mu(f^{-1}(B)) \quad \text{for all measurable } B \subset \Gamma \text{ and arbitrary } \mu \in \mathcal{M}.$$

Invariant measures correspond to fixed points of P which means that $P\mu = \mu$ iff $\mu \in \mathcal{M}$ is invariant. In what follows, we will advocate to discretize the operator P in such a way that its (matrix) approximation P_d has an eigenvector

v_d satisfying $P_d v_d = v_d$, which means that v_d is an approximation of an invariant measure. For an invariant measure, any numerical discretization may be interpreted as a stochastic perturbation of the original problem.

Chemical Conformations From a chemical point of view, biomolecular systems are characterized by different "conformations" – a term, which simultaneously describes both distinguishable geometric configurations and the associated chemical "functionality". In a conformation, the large scale geometric structure of the molecule is understood to be conserved, whereas on smaller scales the system may well rotate, oscillate or fluctuate. For a conformation to be an object of chemical interest, the duration of stay within that comformation should be long enough (*stable* conformation) or, equivalently, it should make a significant contribution to any (statistical) averages. Conformational *changes* are therefore rare events, which will show up only in long–term simulations.

From a mathematical point of view, conformations are special subsets of phase space: a) *invariant sets* of MD systems, which correspond to infinite durations of stay (or relaxation times) and contain all subsets associated with different conformations, b) *almost invariant sets*, which correspond to finite relaxation times and consist of conformational subsets. In order to characterize the dynamics of a system, these subsets are the interesting objects. As already mentioned above, invariant measures are fixed points of the Frobenius-Perron operator or, equivalently, eigenmodes of the Frobenius-Perron operator associated with eigenvalue exactly 1. In view of this property, almost invariant sets will be understood to be connected with eigenmodes associated with (real) eigenvalues close (but not equal) to 1 – an idea recently developed in [6].

Almost Invariant Sets Due to Frobenius-Perron theory, the eigenmode for $\lambda = 1$ has only positive components, which permits a direct interpretation as an eigenmeasure – in this case the invariant measure. The situation is different for eigenmodes associated with $\lambda \neq 1$, which cannot directly be interpreted as measures. In order to clarify the connection between eigenmodes to eigenvalues $\lambda \approx 1$ but $|\lambda| < 1$ with almost invariant sets, we now perform some "Gedankenexperiment": the basic idea is that almost invariant sets should allow for an interpretation as being *continuous* perturbations of invariant sets. This interpretation will lead us to an appropriate criterion for the identification of almost invariant sets exploiting the eigenmodes to eigenvalues $\lambda \approx 1$. Suppose, that the system under investigation depends on a suitable parameter γ (e.g. the total energy or a parameter in the potential, cf. example (15) in Section 4). Let $B = B_1 \cup B_2$ with B_1, B_2 subsets to be defined.

In the *unperturbed* case, say $\gamma = \gamma_0$, let the system have two *disjoint invariant* sets B_1 and B_2 associated with two eigenmodes and invariant mea-

sures μ_1 and μ_2, respectively. In this situation, we have

$$\mu_1(B_2) = \mu_2(B_1) = 0 \quad \text{and} \quad \mu_1(B_1) = \mu_2(B_2) = 1. \tag{4}$$

For $B = B_1 \cup B_2$, both measures are eigenmeasures of P associated with the eigenvalue $\lambda = 1$. Hence, there is no unique invariant measure μ. In fact, any linear combination

$$\mu = \alpha\mu_1 + (1-\alpha)\mu_2, \quad \alpha \in [0,1]$$

will be an invariant measure. In order to span the thus arising two-dimensional eigenspace by orthogonal eigenmodes, say μ, ν, we are directly led to define

$$\nu = \alpha\mu_2 - (1-\alpha)\mu_1.$$

From (4) we then find

$$\nu(B_2) = \alpha, \quad \nu(B_1) = \alpha - 1, \tag{5}$$

and thus

$$\nu(B) = 2\alpha - 1, \quad |\nu|(B) = \mu(B) = 1. \tag{6}$$

In the *perturbed* case, assume now a small change in γ, which induces a small intersection of B_1 and B_2. Let $B = B_1 \cup B_2$ remain to be invariant. Then, we have a unique eigenmeasure $\tilde{\mu}$ with eigenvalue $\lambda_1 = 1$ and another eigenvector $\tilde{\nu}$ associated with $\lambda_2 \neq 1$. Under some continuity assumption λ_2 should be close to 1 and thus we have $\tilde{\nu}(B) = 0$. In view of (6), continuity for $\gamma = \gamma_0$ then requires $\alpha = 1/2$. Therefore (5) implies that

$$\tilde{\nu}(B_1) = -1/2, \quad \tilde{\nu}(B_2) = 1/2, \tag{7}$$

while $|\tilde{\nu}|$ approximates a probability measure. From this insight, we may infer a *construction principle* for almost invariant sets: decompose B into two almost invariant sets B_1, B_2 such that conditions (7) and (6) hold simultaneously. Given a (spatial) discretization (cf. Section 3.2) this means to collect all positive components within the one and all negative components within the other subset (cf. Section 4) — thus defining a unique identification of almost invariant sets. Note, that the number $1/2$ does *not* arise from any symmetry argument. This simplified treatment only covers the identification of two almost invariant sets. For a generalization along these lines of thought see [9, 6].

After these preliminaries we are now ready for a mathematically precise definition of an almost invariant set. Let $\rho \in \mathcal{M}$ be any probability measure. We say that the set B is *δ-almost invariant with respect to ρ* if

$$\frac{\rho(f^{-1}(B) \cap B)}{\rho(B)} = \delta. \tag{8}$$

In words: δ is the probability that points in B are mapped into B under f. In particular, if B is an invariant set, that is $f^{-1}(B) = B$, then $\delta = 1$ independent of the choice of ρ. Once $\delta = \delta_B$ has been computed for a given set $B \subset \Gamma$ using the stepsize τ in the discrete dynamical system with $f = (\Psi^\tau)^m$, the system's probability of staying within B for time T can be estimated to be

$$p_B(T) = \delta_B^{T/(m\tau)}. \tag{9}$$

This is just the kind of information required from the chemical point of view.

As it turns out, there exists a relationship between those probabilities, by which sets are almost invariant, and associated eigenvalues λ (cf. [6]).

Proposition 2. *Let ν be an eigenmode to P with respect to a real eigenvalue $|\lambda| < 1$. Furthermore, let $B \subset \Gamma$ be a set with $\nu(B) = \frac{1}{2}$. If B is δ_1-almost invariant and $\Gamma - B$ is δ_2-almost invariant with respect to $|\nu|$, then*

$$\delta_1 + \delta_2 = \lambda + 1. \tag{10}$$

For more than two almost invariant sets one has to consider all eigenmeasures corresponding to eigenmodes for eigenvalues close to one. In this case, the following lemma will be helpful.

Lemma 3. *Let $\rho \in \mathcal{M}$ be a probability measure and let X and Y be disjoint sets which are δ_X- resp. δ_Y-almost invariant with respect to ρ. Moreover suppose that $f^{-1}(X) \cap Y = \emptyset$ and $f^{-1}(Y) \cap X = \emptyset$. Then $X \cup Y$ is $\delta_{X\cup Y}$-almost invariant with respect to ρ where*

$$\delta_{X\cup Y} = \frac{\rho(X)\delta_X + \rho(Y)\delta_Y}{\rho(X)+\rho(Y)}. \tag{11}$$

In (10), both δ_1 and δ_2 appear as independent constants. If, in addition, the dynamical system possesses some symmetry, then these numbers may satisfy a further relation. To illustrate this fact, let us consider the simplest case where we have a symmetry transformation κ in the problem with $\kappa^2 = id$. Then one can show (see again [6]):

Corollary 4. *In addition to the assumptions in Proposition 2 suppose that*

(i) the set B satisfies $\kappa B = \Gamma - B$, and
(ii) the measure $|\nu|$ is κ-symmetric, that is $\kappa^*|\nu| = |\nu|$.

Then $\Gamma - B$ is δ-almost invariant with respect to $|\nu|$ if and only if B is δ-almost invariant. In particular

$$\delta = \tfrac{1}{2}(\lambda + 1). \tag{12}$$

These relations will be exploited in our subsequent Section 4.

The literature on ergodic theory contains an interesting theorem concerning the spectrum of the Frobenius-Perron operator P. In order to state this result, we have to reformulate P as an operator on the Hilbert space $L^2(\Gamma)$ of all square integrable functions on the phase space Γ. Since Φ^τ and, therefore, f are volume preserving, this operator $P : L^2(\Gamma) \to L^2(\Gamma)$ is *unitary* (cf. [20], Thm. 1.25). As a consequence, its spectrum lies on the unit circle.

However, we consider the Frobenius-Perron operator acting on measures rather than on $L^2(\Gamma)$. The reason is that we are working on energy cells approximating energy surfaces (see Section 3.2) and not in the entire space Γ – and on an energy surface it may be the case that there is no invariant measure with an L^2-density. For instance, consider the situation that (part of) the energy surface is given by a homoclinic trajectory connecting an equilibrium p with itself. Then it is easy to see that the only invariant measure supported on this homoclinic orbit is the Dirac measure δ_p supported in the equilibrium p. In particular, there is no invariant measure possessing an L^2-density. Hence the space of measures is certainly more appropriate than $L^2(\Gamma)$ in our context.

On the other hand, also measure spaces have significant disadvantages. Roughly speaking, the reason is that a measure cannot take into account the dynamical behavior in the neighborhood of its support. Hence there is some need for a new functional analytic setting for the Frobenius-Perron operator which allows to represent correctly the dynamical behavior. Indeed, this is currently under investigation [15]. We would also like to emphasize that our numerical approach seems to reproduce precisely the correct results from the dynamical point of view. In fact, we obtain the correct almost invariant sets together with reasonable corresponding eigenvalues.

However, by constructing a *nested* sequence of successively larger discrete spaces and approximations therein we hope to end up with some approximation of a unique invariant measure, which is then implicitly defined via the constructed sequence of subspaces. An expression of this mathematical consideration is the *multilevel* structure of the suggested algorithm – details see below (Section 3.2). In physical terms, we hope that the perturbations introduced by discretization induce a unique and smooth invariant measure but are so weak that they do *not* destroy the essential *physical* structure of the problem.

3.2 Subdivision Techniques

This section deals with the question of how to approximate the essential features of the flow Φ^τ for given energy E. Recall that the flow ϕ^τ conserves energy, i.e., it maps the *energy surface* $\Gamma_0(E) = \{x \in \Gamma : H(x) = E\}$ onto itself. In the language of statistical physics, we want to approximate the *microcanonical ensemble*. However, even for a symplectic discretization, the discrete flow $f = (\Psi^\tau)^m$ does *not* conserve energy exactly, but only on

average. Therefore, rather than discretizing the energy surface, we will have to discretize an *energy cell* defined by

$$\Gamma_{\delta E}(E) = \{x \in \Gamma, |H(x) - E| \leq \delta E\}$$

in terms of some perturbation parameter δE. On the basis of the above considerations, δE must be small enough to approximate $\Gamma_0(E)$, but also large enough to allow for energy perturbations due to the discretization. In the following we use the notation $\delta E(\tau)$ for a suitable choice. In view of subdivision techniques applied to hyperbolic dynamical systems (cf. [5]), the main algorithmic steps for Hamiltonian systems are as follows:

1. Construction of an approximate covering of the energy cell.
2. Setting up the Frobenius-Perron operator with respect to this subset.

We now describe each of these steps in more detail.

Covering of Energy Cells Assume that the energy cells under consideration are compact sets and the stepsize τ is fixed. We want to construct a collection $\mathcal{B}$ of boxes in phase space such that the union Q of these subsets is a covering of $\Gamma_{\delta E(\tau)}(E)$, the energy cell we focus on. To this end, consider a sequence of energy cells $\Gamma_{\delta E_k}(E)$ that shrink to $\Gamma_{\delta E(\tau)}(E)$, i.e., consider $\delta E_k \geq \delta E_{k+1} \geq \cdots \geq \delta E(\tau)$. The collection $\mathcal{B}$ is constructed via a sequence of collections $\mathcal{B}_k$ by the following *recursive subdivision algorithm*: $\hat{\mathcal{B}}_k$ results from $\mathcal{B}_{k-1}$ by subdivision and gives $\mathcal{B}_k$ by the following *selection step*

$$\mathcal{B}_k = \{B \in \hat{\mathcal{B}}_k : B \cap \Gamma_{\delta E_k}(E) \neq \emptyset\}. \tag{13}$$

Thus, this multi-level process produces a finer and finer covering of $\Gamma_{\delta E(\tau)}(E)$. Up to now, the parameters δE_k are adapted to the size of the boxes according to some heuristics. Recall that an approximation of the energy surface $\Gamma_0(E)$ would only be possible in the limit $\tau \to 0$ which implies $\delta E(\tau) \to 0$.

Discretization of the Frobenius-Perron Operator The previous step led to a collection $\mathcal{B} = \{B_1, \ldots, B_d\}$ covering some energy cell $\Gamma_{\delta E}(E)$. Considering now the dynamics inside this energy cell, we are only interested in the subsets $G_i = B_i \cap \Gamma_{\delta E}(E)$ of the boxes $B_i \in \mathcal{B}$. We may now use this covering for a discretization of the Frobenius-Perron operator. The discretization is realized via a Galerkin type ansatz as described in [6]. Consider the case of locally constant basis functions

$$\varphi_i = \chi_{G_i}, \quad i = 1, 2, \ldots, d.$$

Then, the discretized Frobenius-Perron operator $v = P_d u$ can be written componentwise as

$$v(G_i) = \sum_{j=1}^{d} p_{ij} u(G_j), \quad p_{ij} = \frac{m(f^{-1}(G_i) \cap G_j)}{m(G_j)}, \quad i = 1, \ldots, d, \tag{14}$$

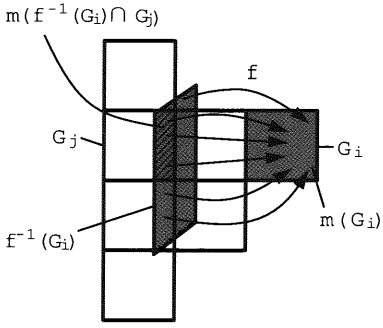


Fig. 4. Computation of the stochastic matrix P_d via mapping of discretization boxes.

where m denotes the Lebesgue measure (i.e., $m(G_j)$ is the volume of the box G_j). For approximating the "volume"

$$m(f^{-1}(G_i) \cap G_j) = \int_{G_j} \chi_{G_i}(f(x))dx$$

we use a Monte-Carlo approximation of the integral. That is, we approximate the *transition probabilities* p_{ij} via a Monte-Carlo discretization,

$$p_{ij} = \frac{1}{m(G_j)} \int_{G_j} \chi_{G_i}(f(x))dx \approx \frac{1}{N} \sum_{n=1}^{N} \chi_{G_i}(f(x_n)),$$

where the x_n are chosen randomly and uniformly distributed in G_j (cf. Fig. 4). Since the algorithmic control guarantees that each point is mapped to exactly one box G_j, the matrix P_d is certainly singly stochastic (w.r.t. columns). The map f is symplectic and hence volume preserving. Therefore, if the union $Q = \cup_i G_i$ of all subsets were invariant under f, i.e., $f(Q) = Q$, the matrix P_d would also be stochastic with respect to its rows. However, such an effect would be unwanted, since then, we would always end up with an approximate invariant measure being equidistributed on Q (which is not always correct). Fortunately, $f(Q)$ in general deviates from Q, which reflects the fact that Q approximates the energy surfaces of the original problem instead of that of the symplectic discretization Ψ. But $f(Q) \not\subset Q$ implies that there are some points $x \in Q$ with $f(x) \notin Q$. Let $x \in G_j$ be one of these points. We then assign $f(x)$ to the nearest box G_i, i.e., the transition $x \to f(x)$ is counted for the transition probability p_{ij}. The same is done in the case when the covering Q of the energy cell $\Gamma_{\delta E}(E)$ is eventually incomplete. These perturbations of f are artificial but necessary in order to assure that the matrix approximation P_d is stochastic (w.r.t. columns) thus inheriting the most important property of the operator P.

After the assembling of the stochastic matrix P_d we have to solve the associated *non-selfadjoint* eigenvalue problem. Our present numerical results have been computed using the code **speig** by RADKE AND SØRENSEN in MATLAB,

which allow for a simultaneous subspace iteration to compute eigenmodes associated with the eigenvalue cluster around $\lambda = 1$.

The basic scheme of this algorithm is similar to "cell-to-cell mapping" techniques [14] but differs substantially in one important aspect: If applied to larger problems, a direct cell-to-cell approach quickly leads to tremendous computational effort. Only a proper exploitation of the multi-level structure of the subdivision algorithm (also for the eigenvalue problem) may allow for application to molecules of real chemical interest. But even this more sophisticated approach suffers from combinatorial explosion already for moderate size molecules. In a next stage of development [19] this restriction will be circumvented using certain hybrid Monte-Carlo methods.

4 Illustrative Numerical Example

Let us introduce a suitably simple example in order to illustrate the notion of almost invariant sets and the performance of our algorithm for Hamiltonian systems. For $p = (p_1, p_2), q = (q_1, q_2) \in \mathbb{R}^2$ consider the potential

$$V_4(q) = \left(\tfrac{3}{2}q_1^4 + \tfrac{1}{4}q_1^3 - 3q_1^2 - \tfrac{3}{4}q_1 + 3\right) \cdot \left(2q_2^4 - 4q_2^2 + \gamma\right) \quad \text{with } \gamma = 3. \quad (15)$$

As illustrated in Fig. 5, this potential comprises four local minima at the points $(\pm 1, \pm 1)$ (named A, B, C, D), which are separated by four saddle-points. The energy barrier between A and B is significantly higher than the other three ones. The dynamical behavior of the system consists of oscillations around the local minima and, if the total energy is large enough for the system to cross the barriers, of motions from one minimum to the other. If the energy is not too large, there will be two kinds of "long term" dynamical behavior:

(a) oscillations in the neighborhood of the four different minima,
(b) back and forth oscillations between two different minima: $A \leftrightarrow D$, $B \leftrightarrow C$, and $C \leftrightarrow D$.

This is observed in simulations of the dynamics. Fig. 5 presents a solution which starts with an oscillation between A and D, followed by an oscillation around C, a long period of oscillations between A and D and so on. The similarity of the trajectories shown in Fig. 5 (right) and Fig. 3 illustrates that we are actually looking at the same kind of phenomena. Thus, for the case presented in Fig. 5, the neighborhoods of the different minima should turn out to be "almost invariant sets" as well as neighborhoods of the pairs of minima (A, D), (C, D) and (B, C) together with regions around the corresponding saddle-points "between" them.

Inefficiency of Direct Simulation Suppose we want to compute the corresponding invariant measure μ by direct simulation. Direct long term simulation by symplectic discretization of (1) yields the discrete solution $(x_k)_{1,\ldots,N}$.

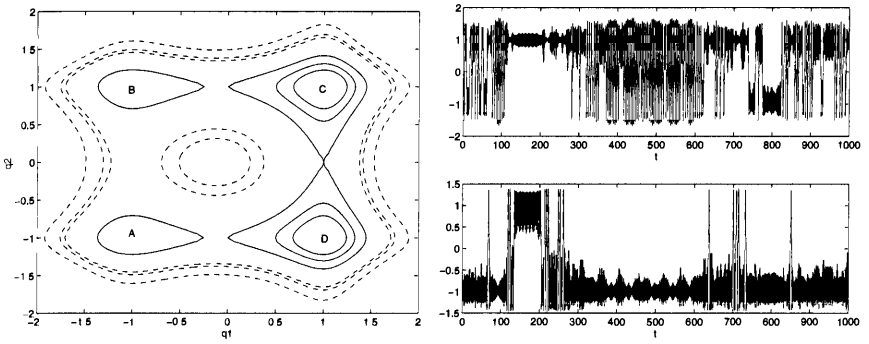


Fig. 5. The left hand side figure shows a contour plot of the potential energy landscape due to V_4 with equipotential lines of the energies $E = 1.5, 2, 3$ (solid lines) and $E = 7, 8, 12$ (dashed lines). There are minima at the four points $(\pm 1, \pm 1)$ (named A to D), a local maximum at $(0, 0)$, and saddle-points in between the minima. The right hand figure illustrates a solution of the corresponding Hamiltonian system with total energy $E = 4.5$ (positions q_1 and q_2 versus time t).

For N large enough and a box $B \subset \Gamma$ one takes the relative frequency of $x_j \in B$ as an approximation of the invariant density in B. If the system were ergodic, the convergence of this algorithm would be guaranteed. Even in this case the convergence could be arbitrarily slow, when the iteration gets trapped within an almost invariant set of the system.

Our global subdivision approach is not sensitive to such a situation. Over sufficiently long run times of direct simulation both methods eventually yield roughly the same results, see Fig. 6.

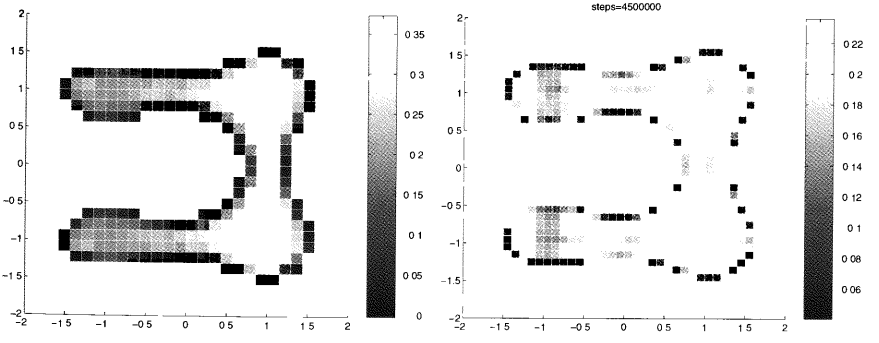


Fig. 6. The density of the invariant measure of the potential V_4 for total energy $E = 4.5$. Results of the subdivision approach (left) and a direct simulation with about 4.5 million steps for stepsize $\tau = 1/30$ (right).

Almost Invariant Sets Recall that the relevant almost invariant sets correspond to eigenvalues $\lambda \approx 1$ with $|\lambda| < 1$ of the associated Frobenius–Perron operator.

Based on observations concerning the dynamical behavior we already conjectured that there exist seven almost invariant sets – a conjecture that we now want to check numerically. We employ the subdivision algorithm for subtrajectories of length $m\tau = 0.1$. The final box-collection corresponding to the total energy $E = 4.5$ after 18 subdivision steps consists of 18963 boxes.

A simultaneous computation of the four largest eigenvalues $\lambda_1, \ldots, \lambda_4$ leads to the following table:

Number	Eigenvalue
1	1.0000
2	0.9963
3	0.9891
4	0.9782

The invariant measure ν_1 corresponding to $\lambda_1 = 1$ has already been shown in Fig. 6. Next, we discuss the information provided by the eigenmeasure ν_2 corresponding to λ_2. The box coverings in the two parts of Fig. 7 approximate two sets B_1 and B_2, where the discrete density of ν_2 is positive resp. negative. We observe, that for $\gamma > 4.5$ in (15) the energy $E = 4.5$ of the system would not be sufficient to move from B_1 to B_2 or vice versa. That is, in this case B_1 and B_2 would be invariant sets. Thus, we are exactly in the situation illustrated in our Gedankenexperiment in Section 3.1.

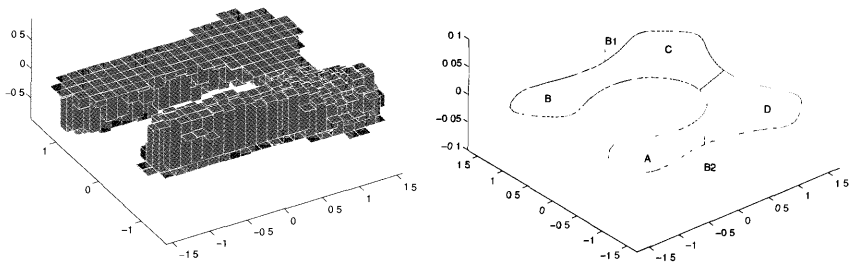


Fig. 7. Eigenmeasure ν_2 of the Frobenius–Perron operator to the second largest eigenvalue $\lambda_2 = 0.9963$ for the test system (15) with $\gamma = 3$. ν_2 was computed via our new subdivision algorithm (cf. Section 4).

Moreover, our Hamiltonian system possesses an additional symmetry — it is equivariant under the transformation $(q_2, p_2) \to -(q_2, p_2)$. In other words each of these sets is a candidate for a set B mentioned in the assumptions of Corollary 4. Thus, by this result, both of these sets are almost invariant with

probability $\delta = (\lambda_2 + 1)/2 = (0.9963 + 1)/2 = 0.9981$. Observe that these almost invariant sets confirm the observation made above that dynamically there exist "long term" oscillations between the minima $A \leftrightarrow D$ and $B \leftrightarrow C$.

The third eigenmeasure ν_3 corresponding to λ_3 provides information about three additional almost invariant sets: on the left hand side in Fig. 8 we have the set corresponding to the oscillation $C \leftrightarrow D$, whereas on the right hand side the two almost invariant sets around the equilibria A and B are identified. Again the boxes shown in the two parts of Fig. 8 approximate two sets where the discrete density of ν_3 is positive resp. negative. In this case we can use Proposition 2 and the fact that A and B are symmetrically related to conclude that for all these almost invariant sets $\delta \geq \lambda_3 = 0.9891$.

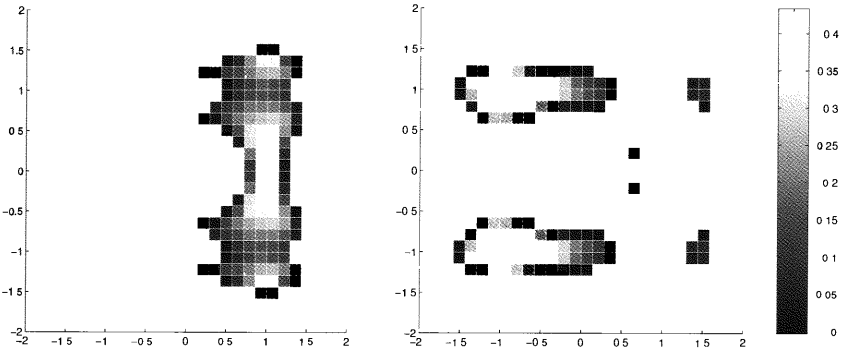


Fig. 8. Illustration of three almost invariant sets with respect to the probability measure $|\nu_3|$. The coloring is done according to the magnitude of the discrete density.

Finally, the information on the remaining almost invariant sets in the neighborhood of the equilibria C and D can be extracted using the eigenvalue λ_4 with the eigenmeasure ν_4 (see Fig. 9). In the two parts of Fig. 9 we show again the boxes, which approximate two sets, where the discrete density of ν_4 is positive resp. negative. Let us denote by Y the union of the boxes around equilibrium B in the first part of the figure and by X the boxes around D. (We ignore the isolated box in the left lower corner, which we regard as a numerical artifact.) We now use Lemma 3 to derive a lower bound for δ_X. Numerically we obtain the values $|\nu_4(X)| = 0.3492$ and $|\nu_4(Y)| = 0.1508$. Note that $|\nu_4(X \cup Y)| = 0.5$ and $\lambda_4 + 1 = 2\delta_{X \cup Y}$ (using again the symmetry and Corollary 4) which leads to the estimate

$$\delta_X = \frac{0.5\delta_{X\cup Y} - |\nu_4|(Y)\delta_Y}{|\nu_4|(X)} = \frac{\lambda_4 + 1 - 4|\nu_4|(Y)\delta_Y}{4|\nu_4|(X)} \geq \frac{\lambda_4 + 1 - 4|\nu_4|(Y)}{4|\nu_4|(X)} = 0.9844.$$

In all calculations done so far a fixed stepsize $\tau = 0.1$ has been used. Hence an application of formula (9) leads to the following table concerning flip- flop probabilities between different conformations.

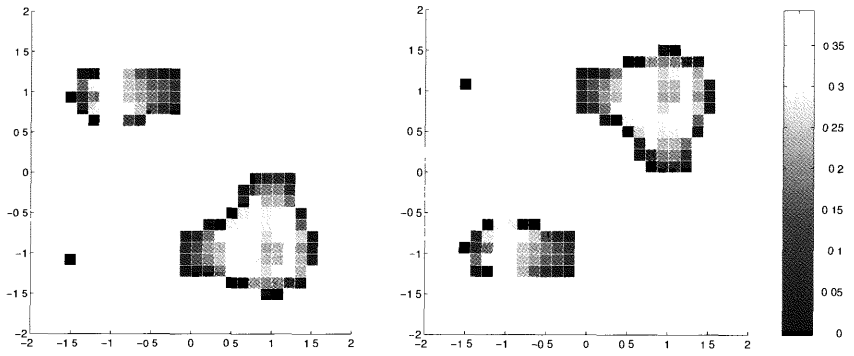


Fig. 9. Illustration of four almost invariant sets with respect to the probability measure $|\nu_4|$. The coloring is done according to the magnitude of the discrete density.

conformation	probability to stay within for			
	0.1 sec.	1 sec.	10 sec.	100 sec.
$A \leftrightarrow D$, $B \leftrightarrow C$	0.9981	0.9812	0.8268	0.1493
$C \leftrightarrow D$, A, B	0.9891	0.8962	0.3342	< 0.0002
C,D	0.9844	0.8545	0.2076	$< 10^{-6}$

These numbers indicate that it is very unlikely for the system to stay in C and D for more than 100 seconds, whereas for an oscillation $A \leftrightarrow D$ or $B \leftrightarrow C$ this may well be the case. In particular, these results are in nice agreement with Fig. 5 (right): there we observe an oscillation $A \leftrightarrow D$ for about 200 seconds, whereas the longest stay in the neighborhood of the minimum C only lasts about 60 seconds.

5 Conclusion

The paper suggests a novel concept for computing essential features of Hamiltonian systems arising in molecular dynamics. The concept involves the determination of invariant and almost invariant sets via eigenmodes of the associated Frobenius-Perron operator. The numerical findings in an illustrative example are intriguing, but a sound theoretical basis of the approach is still missing. Moreover, the algorithm as it stands now is suitable only for small systems. However, extensions to avoid the "curse of dimension" are already under development. The value of the suggested concept and its range of applicability to MD will prove or disprove in the years to come.

Acknowledgements. We thank Sebastian Reich for helpful discussions. Research partly supported by the Deutsche Forschungsgemeinschaft under Grant De 448/5-2 and Grant De 293/2-1.

References

1. M.P. Allen and D.J. Tildesley. Computer Simulations of Liquids. Clarendon Press, Oxford (1990)
2. A. Amadei, A.B.M. Linssen, and H.J.C. Berendsen. Essential dynamics on proteins. Proteins **17** (1993) 412-425
3. G. Benettin and A. Giorgilli. On the Hamiltonian interpolation of near to the identity symplectic mappings with applications to symplectic integration algorithms. J. Stat. Phys. **74** (1994)
4. R. Bulirsch. Die Mehrzielmethode zur numerischen Lösung von nichtlinearen Randwertproblemen und Aufgaben der optimalen Steuerung. Tech. Rep., Carl-Cranz-Gesellschaft (1971)
5. M. Dellnitz and A. Hohmann. A subdivision algorithm for the computation of unstable manifolds and global attractors. Numerische Mathematik **75** (1997) 293–317
6. M. Dellnitz and O. Junge. On the approximation of complicated dynamical behavior. To appear in SIAM J. Num. Anal., 1998. Also available as Preprint SC 96-35, Konrad Zuse Zentrum, Berlin (1996)
7. P. Deuflhard and F. Bornemann. Numerische Mathematik II — Integration gewöhnlicher Differentialgleichungen. Walter de Gruyter, Berlin, New York (1994)
8. P. Deuflhard, M. Dellnitz, O. Junge, and Ch. Schütte. Computation of essential molecular dynamics by subdivision techniques I: Basic concept. Preprint SC 96-45, Konrad Zuse Zentrum, Berlin (1996)
9. P. Deuflhard, W. Huisinga, A. Fischer, Ch. Schütte. Identification of Almost Invariant Aggregates in Nearly Uncoupled Markov Chains. Preprint, Preprint SC 98-03, Konrad Zuse Zentrum, Berlin (1998)
10. H. Grubmueller and P. Tavan. Molecular dynamics of conformational substates for a simplified protein model. J. Chem. Phys. **101** (1994)
11. E. Hairer. Backward analysis of numerical integrators and symplectic methods. Annals of Numerical Mathematics **1** (1994)
12. E. Hairer and Ch. Lubich. The life-span of backward error analysis for numerical integrators. Numer. Math. **76** (1997) 441–462
13. E. Hairer and D. Stoffer. Reversible long–term integration with variable step sizes. Report (1995)
14. H. Hsu. Global analysis by cell mapping. Int. J. Bif. Chaos **2** (1992) 727–771
15. G. Keller. Private communication (1998)
16. D. Ruelle. A measure associated with axiom A attractors. Amer. J. Math. **98** (1976) 619–654
17. J.M. Sanz-Serna and M.P. Calvo. Numerical Hamiltonian Systems. Chapman and Hall, London, Glasgow, New York, Tokyo (1994)
18. J. Stoer and R. Bulirsch. Introduction to Numerical Analysis. Springer Verlag, Berlin, Heidelberg, New York, Tokyo (1980)
19. Ch. Schütte, A. Fischer, W. Huisinga, P. Deuflhard. A Hybrid Monte-Carlo Method for Essential Molecular Dynamics. Preprint, Preprint SC 98-04, Konrad Zuse Zentrum, Berlin (1998)
20. P. Walters. An Introduction to Ergodic Theory. Springer Verlag, Berlin, Heidelberg, New York, Tokyo (1981)

Mathematical Model of the Nucleic Acids Conformational Transitions with Hysteresis over Hydration–Dehydration Cycle

Michael Ye. Tolstorukov[1] and Konstantin M. Virnik[2]

[1] Chair of Molecular and Applied Biophysics, Kharkov State University, Svobody Sq. 4, Kharkov 310077, Ukraine
[2] Biophysics Division, Institute of Radiophysics and Electronics, Nat. Acad. Sci. of Ukraine, Akad. Proskura Str. 12, Kharkov 310085, Ukraine

Abstract. A model of the conformational transitions of the nucleic acid molecule during the water adsorption–desorption cycle is proposed. The nucleic acid–water system is considered as an open system. The model describes the transitions between three main conformations of wet nucleic acid samples: A–, B– and unordered forms. The analysis of kinetic equations shows the non–trivial bifurcation behaviour of the system which leads to the multistability. This fact allows one to explain the hysteresis phenomena observed experimentally in the nucleic acid–water system. The problem of self–organization in the nucleic acid–water system is of great importance for revealing physical mechanisms of the functioning of nucleic acids and for many specific practical fields.

1 Introduction

At present it is clear that the conformational dynamics of the nucleic acids (NA) is of importance for the functioning of the living cell [1]. There is a number of the conformations, spatial configurations with definite structural parameters, which NAs adopt under the wide range of the experimental conditions. These are the families of the canonical B–conformation, A–conformation, Z–conformation (ordered forms) and unordered or P–form (observed in the samples with low water content) with no base stacking and no melting when heated (for reviews see [2], [3], [4], [5]). Their appearance is dependent on the primary structure of the NA and the experimental conditions. The NA conformations are stable, but not static. The transitions between NA conformations are possible on changing of the properties of the NA surroundings such as the temperature, the ionic strength of the NA solution, the water content of the sample, etc.

For modelling conformational transitions and nonlinear dynamics of NA a phenomenological approach is often used. This allows one not just to describe a phenomenon but also to understand the relationships between the basic physical properties of the system. There is a general algorithm for modelling in the frame of the phenomenological approach: determine the dominant motions of the system in the time interval of the process treated and then write

down and solve the equations describing these motions [6]. There is a great number of works dealing with the modelling of NA conformational dynamics [[7], [8], [9], [10] and references therein].

As the conformation of NAs is dependent on the properties of their environment the next step to understanding the physical basis of the conformational dynamics of the NAs could be the direct consideration of the interaction of macromolecule with its nearest surrounding ion-hydrate. In the present study we restrict our consideration to the hydration shell of the NA, assuming that the countreion content is constant. A convenient sample to investigate the interaction of NA with water is wet NA fibers and films. The hydration shell stabilizes the ordered forms of NA by means of forming regular structures due to H–bonds between adsorbed water molecules, "water backbone", and is one of the main factors influencing the NA conformational state. Such "water bridges" considerably contribute to the stabilization of the helical structure up to 70 % of the melting enthalpy of B–DNA in wet films and gels [5]. The regular hydration shells of the distinct nucleic acid conformations are known to be considerably different and distinct groups of the adsorbed water molecules play different roles in the stabilization of the distinct NA conformations [2], [11], [3]. The dry NA is known to be unordered [5].

The hydration shell is formed with the increasing of the water content of the sample and the NA transforms from the unordered to A– and then to B–form, in the case of DNA and DNA–like polynucleotides and salt concentrations similar to in vivo conditions. The reverse process, dehydration of NA, results in the reverse conformational transitions but they take place at the values of relative humidity (r.h.) less than the forward direction [12]. Thus, there is a conformational hysteresis over the hydration–dehydration loop. The adsorption isotherms of the NAs, i.e. the plots of the number of the adsorbed water molecules versus the r.h. of the sample at constant temperature, also demonstrate the hysteresis phenomena [13]. The hysteresis is reproducible and its value does not decrease for at least a week.

A possible explanation of the hysteresis could be the non–equilibrium of the DNA hydration. In that case the value of hysteresis has to depend on the size of the experimental sample. However, such a dependence is not observed in the wide range of DNA film thicknesses (0.05–0.2 μm) [14], [12]. Thus, hysteresis cannot be a macroscopic phenomenon and does reflect the molecular interaction of water and the biopolymer.

Taking into account the hydration shell of the NA and the possibility of the water content changing we are forced to consider the water + nucleic acid as an open system. In the present study a phenomenological model taking into account the interdependence of hydration and the NA conformation transition processes is offered. In accordance with the algorithm described above we consider two types of the basic processes in the system and thus two time intervals: the water adsorption and the conformational transitions of the NA, times of the conformational transitions being much more greater

than the time of the water adsorption. The relaxation time of the adsorbed water molecules is about 10^{-10}–10^{-9} s [15], [16], while the time of the conformational equilibrium in the case of the wet NA samples is of the order of hours and even days [17].

2 The Model

2.1 Conformational Transitions

We consider a finite space, which contains the NA sample and is in contact with a bath of water or water vapor. That allows one to maintain the r.h. in the experimental space at a constant level and change it when necessary. Such a scheme corresponds to the real experiments with wet NA samples. A NA molecule is simulated by a sequence of units of the same type. Thus, in the present study, we consider the case of a homogeneous NA or the case where averaging over the unit type is possible. Every unit can be found in the one of three conformational states: unordered, A– or B– conformations. The units can reversibly change their conformational state. A unit corresponds to a nucleotide of a real NA. We assume that the NA strands do not diverge during conformational transitions in the wet NA samples [18]. The conformational transitions are considered as cooperative processes that are caused by the unfavorable appearance of an interface between the distinct conformations.

The processes of the conformational transitions of NA can be illustrated with the scheme:

$$U \underset{k_{\mathrm{A}}^-}{\overset{k_{\mathrm{A}}^+}{\rightleftharpoons}} A, \quad A \underset{k_{\mathrm{B}}^-}{\overset{k_{\mathrm{B}}^+}{\rightleftharpoons}} B, \quad B \underset{k_{\mathrm{U}}^-}{\overset{k_{\mathrm{U}}^+}{\rightleftharpoons}} U \tag{1}$$

where U represents unordered state; A represents the A–form and B – represents the B–form; $k_{\mathrm{A}}^+, k_{\mathrm{A}}^-, k_{\mathrm{B}}^+, k_{\mathrm{B}}^-, k_{\mathrm{U}}^+, k_{\mathrm{U}}^-$ – the probabilities of the occurrence of the corresponding conformational transition per time unit.

Let us introduce the variables which are the probabilities of finding an arbitrary unit in a certain conformational state: U for unordered state, A for A– and B for B–form of the NA. There is the natural relationship between the variables:

$$U + A + B = 1 \tag{2}$$

The temporal evolution of the variables can be described with the set of differential equations, which corresponds to the scheme (1):

$$\begin{aligned} \frac{dA}{dt} &= k_{\mathrm{A}}^+ U - k_{\mathrm{A}}^- A + k_{\mathrm{B}}^- B - k_{\mathrm{B}}^+ A \\ \frac{dB}{dt} &= k_{\mathrm{B}}^+ A - k_{\mathrm{B}}^- B + k_{\mathrm{U}}^- U - k_{\mathrm{U}}^+ B \\ \frac{dU}{dt} &= k_{\mathrm{U}}^+ B - k_{\mathrm{U}}^- U + k_{\mathrm{A}}^- A - k_{\mathrm{A}}^+ U \end{aligned} \tag{3}$$

The NAs such as DNA usually used in the experiments consist of 10^4–10^5 nucleotides. Thus, they should be considered as macrosystems. Moreover, in experiments with wet NA samples macroscopic quantities are measured, so averaging should also be performed over all nucleic acid molecules in the sample. These facts justify the usage of the macroscopic equations like (3) in our case and require the probabilities of finding macromolecular units in the certain conformational state as variables of the model.

In accordance with the principle of detailed balance the set (3) with regard to (2) after some mathematics can be rewritten as:

$$\begin{aligned} \frac{dA(t)}{dt} &= k_{\mathrm{A}}^{+}\left[1-\left(1+\exp(\Delta F_{\mathrm{UA}}^{\mathrm{p}}+2\sigma(1-2A-B)+\Delta F_{\mathrm{UA}}^{\mathrm{w}})\right)A-B\right] \\ \frac{dB(t)}{dt} &= k_{\mathrm{B}}^{+}\left[A-B\exp(\Delta F_{\mathrm{AB}}^{\mathrm{p}}+2\sigma\,(A-B)+\Delta F_{\mathrm{AB}}^{\mathrm{w}})\right] \end{aligned} \tag{4}$$

Here the following relationships have been used:

$$\begin{aligned} \frac{k_{\mathrm{A}}^{-}}{k_{\mathrm{A}}^{+}} &= \exp\left(\Delta F_{\mathrm{UA}}^{\mathrm{p}}+2\sigma\,(1-2A)-2\sigma B+\Delta F_{\mathrm{UA}}^{\mathrm{w}}\right) \\ \frac{k_{\mathrm{B}}^{-}}{k_{\mathrm{B}}^{+}} &= \exp\left(\Delta F_{\mathrm{AB}}^{\mathrm{p}}+2\sigma\,(A-B)+\Delta F_{\mathrm{AB}}^{\mathrm{w}}\right) \end{aligned}$$

where $\Delta F_{\mathrm{UA}}^{\mathrm{p}}$ and $\Delta F_{\mathrm{AB}}^{\mathrm{p}}$ are the free energy changes (per NA units' mole) in the NA transition from the unordered form to the A–form and from A– to the B– form respectively; σ is the free energy cost of creating an interface between the distinct conformational "phases", it is clear that the parameter σ has different values in the cases of interfaces between the unordered state and the A–form, the A–form and the B–form, etc., but we made all values equal to simplify the model; $\Delta F_{\mathrm{UA}}^{\mathrm{w}}, \Delta F_{\mathrm{AB}}^{\mathrm{w}}$ are the changes of the free energy of the hydration shell of the NA (per NA units' mole) in the U→A and A→B conformational transitions. All parameters are expressed in units of RT.

The expressions appearing in the exponents are the free energy change of the NA–water system per unit mole in the U→A and A→B conformational transitions. The terms $\Delta F_{\mathrm{UA}}^{\mathrm{w}}, \Delta F_{\mathrm{AB}}^{\mathrm{w}}$ are introduced to take into account the effect of the hydration shell on the NA conformational state. It is clear that the terms $\Delta F_{\mathrm{UA}}^{\mathrm{w}}, \Delta F_{\mathrm{AB}}^{\mathrm{w}}$ are functions of the degree of completeness of the NA hydration shell and, as it is formed with the increasing of the r.h. of the sample, are the functions of the r.h. The form of these functions obviously depends on the details of the NA hydration. As it has been shown in Sect. 1 there is a lot of data on this matter. However, in the present qualitative study, at this stage of our research, it is more important to describe only the main features of the NA hydration. Moreover, as water adsorption time is much more less than times of the NA conformational transitions we can use equations describing the equilibrium values of the amount of water adsorbed by NA instead of the differential equations describing the water adsorption process itself. Thus, we have to find an equation describing the water adsorption isotherms of the NA.

2.2 Water Adsorption

The problem of the theoretical description of biopolymer water adsorption isotherms has drawn the attention of researchers for a long time. In the works [19], [20] a rigorous statistical basis for equations describing the isotherms for the case of homogeneous adsorption surfaces and noninteracting adsorption sites of N different types has been suggested. The general equation is:

$$\sum_{j=1}^{N} v_j^{\mathrm{m}} \frac{\partial}{\partial X_{\mathrm{e}}} \ln\left(f_j(X_{\mathrm{e}})\right) = \frac{X(X_{\mathrm{e}})}{X_{\mathrm{e}}} \tag{5}$$

where $X(X_{\mathrm{e}})$ is the total adsorption, i.e. the number of adsorbed water molecules on the NA nucleotide, at the r.h. of the sample X_{e}; v_j^{m} are the primary site capacities of the N different types of the sites whose respective partition functions are $f_j(X_{\mathrm{e}})$. From this equation we can obtain the well known B.E.T. equation (introduced by Brunauer, Emmett and Teller as early as 1938) and its modifications [19] most conventionally used for the modelling of adsorption isotherms and as it is shown in [20] the isotherm suggested by D'Arcy and Watt which is the sum of the Langmuir's, Henry's and multilayer terms [21]. Such a separation of the total adsorption is of importance in the case of NAs since distinct groups of water molecules in the NA hydration shell play different roles in the stabilization of distinct conformation of NAs. However, the isotherm of d'Arcy and Watt can be derived rigorously (5) only if it is assumed that there are $(N-1)$ types of primary sites and the sets of the secondary and other sites, i.e. that monolayer and multilayer adsorptions are independent. Obviously, as it has been noted in the work [20], these assumptions are not valid.

In the present study we try to obtain the isotherm equation in the form of a sum of the three terms: Langmuir's, Henry's and multilayer adsorption, because it is the most convenient and is easily physically interpreted but, using more a realistic assumption. Namely, we take the partition functions as in the case of the isotherm of d'Arcy and Watt [20], but assume that the value of v^{m} for the multilayer adsorption appearing in the (5) is equal to the sum of the number of adsorbed water molecules on the Langmuir's and Henry's sites:

$$\begin{aligned} &f_j(X_{\mathrm{e}}) = 1 + a_j X_{\mathrm{e}}; && \text{for } j = 1, 2, ..., (N-1) \\ &f_j(X_{\mathrm{e}}) = 1 + bX_{\mathrm{e}} + (bX_{\mathrm{e}})^2 + (bX_{\mathrm{e}})^3 + ...; && \text{for } j = N \\ &v_N^{\mathrm{m}} = X_{\mathrm{L}}(X_{\mathrm{e}}) + X_{\mathrm{H}}(X_{\mathrm{e}}) \end{aligned} \tag{6}$$

Substituting (6) to (5) and assuming there are $M < N-1$ types of primary sites for which $a_j v_j^{\mathrm{m}} << 1$ (sites without saturation, Henry's sites) and neglecting the heterogeneity of the Langmuir's and Henry's sites we can

write the total adsorption in the following form:

$$X_{\mathrm{L}}(X_{\mathrm{e}}) = \frac{v_{\mathrm{L}}^{\mathrm{m}} a_{\mathrm{L}} X_{\mathrm{e}}}{1 + a_{\mathrm{L}} X_{\mathrm{e}}}; \quad X_{\mathrm{H}} = a_{\mathrm{H}} X_{\mathrm{e}}; \quad X_{\mathrm{ML}} = (X_{\mathrm{L}} + X_{\mathrm{H}}) \frac{b X_{\mathrm{e}}}{1 - b X_{\mathrm{e}}}$$
$$X(X_{\mathrm{e}}) = X_{\mathrm{L}} + X_{\mathrm{H}} + X_{\mathrm{ML}} \tag{7}$$

The adsorption isotherm in the form (7) should be considered as a reasonable approximation, the more so, as our aim here is not to obtain faithful values of the adsorption parameters but to describe as completely as possible the qualitative behaviour of the NA–water system. It is worth to note here that the use of a polynomial of the form:

$$f(X_{\mathrm{e}}) = 1 + \alpha X_{\mathrm{e}} + \beta X_{\mathrm{e}}^2 + \gamma X_{\mathrm{e}}^3 + \ldots$$

to make an exact fit to the experimental data using a Chebyshev orthogonal curve fitting technique resulted in physically unrealizable activity coefficients $\alpha, \beta, \gamma, \ldots$

However, the B.E.T. and modificated B.E.T as well as isotherm of d'Arcy and Watt fit the experimental data only in some range of the relative humidities up to about 80–85%. At the same time the adsorption in the interval 90–100% is of great interest for in this interval the A→B conformational transition, which is of biological importance, takes place [17], [18]. This disagreement can be the result of the fact that the adsorbed water molecules can form a regular lattice, structure of which depends on the conformation of the NA. To take into account this fact we assume that the water binding constants depend on the conformational variables of the model, i.e:

$$\begin{aligned} a_{\mathrm{L}} &= a_{\mathrm{L}}^0 \exp(\lambda^{\mathrm{A}} A + \lambda^{\mathrm{B}} B) \\ a_{\mathrm{H}} &= a_{\mathrm{H}}^0 \exp(\eta^{\mathrm{A}} A + \eta^{\mathrm{B}} B) \\ b &= b^0 \exp(\beta^{\mathrm{A}} A + \beta^{\mathrm{B}} B) \end{aligned} \tag{8}$$

where $\lambda^{\mathrm{A}}, \eta^{\mathrm{A}}, \beta^{\mathrm{A}}, \lambda^{\mathrm{B}}, \eta^{\mathrm{B}}, \beta^{\mathrm{B}}$ are the differences between the free energies (per unit mole) of the hydration shells of the Langmuir's, Henry's centers and multilayer in the A– or B–forms of the nucleic acid (indicated with superscripts) and unordered state; $a_{\mathrm{L}}^0, a_{\mathrm{H}}^0, b^0$ are the adsorption constants if a biopolymer unit is in the unordered form.

The expressions for the changes in the free energy of the hydration shell of the NA (per NA units' mole) in the U→A and A→B conformational transitions should be written in the form:

$$\begin{aligned} \Delta F_{\mathrm{UA}}^{\mathrm{w}} &= -\left(\lambda^{\mathrm{A}} X_{\mathrm{L}} + \eta^{\mathrm{A}} X_H + \beta^{\mathrm{A}} X_{\mathrm{ML}}\right) \\ \Delta F_{\mathrm{AB}}^{\mathrm{w}} &= -\left((\lambda^{\mathrm{B}} - \lambda^{\mathrm{A}}) X_{\mathrm{L}} + (\eta^{\mathrm{B}} - \eta^{\mathrm{A}}) X_{\mathrm{H}} + (\beta^{\mathrm{B}} - \beta^{\mathrm{A}}) X_{\mathrm{ML}}\right) \end{aligned} \tag{9}$$

Equations (4) and (9) along with (8) and (7) form the a set of the differential–algebraic equations dependent on X_{e} which describes the behaviour of the NA water system, namely the conformational transitions in

the system and the adsorption process. It should be stressed that in (9) the stationary values of the adsorption variables appear, while the values of the conformational variables in (8) are the instantaneous ones that correspond to the different time scales of the adsorption and conformational transitions.

3 Results and Discussion

The typical dependence of the stable stationary solutions to (4) on the control parameter of the model X_e is presented in Fig. 1. These results have been obtained as numerical solutions of (4) with $\frac{dA(t)}{dt}$ and $\frac{dB(t)}{dt}$ equal to zero. The stability of these solutions has been determined by using the linear analysis [22]. The multistability of the model in a certain interval of a values of the control parameter allows one to explain the conformational and adsorption hysteresis phenomena observed in the real system over a r.h. loop. The theoretical curves are in the good agreement with the experimental ones, obtained for the natrium salt of thymus calf DNA using gravimetry (circles in Fig. 1a).

In the Figs. 2, 3 the calculations and experimental (circles in Figs. 2a, 3a) results obtained for the limited r.h. ranges 0–60%, 0–90% respectively are presented. At each value of r.h. we have found the stationary stable solution of (4) which is closest to the one obtained at the previous calculation step. Such a procedure allows us to simulate the real experiment when the hydration and dehydration of the NA sample is performed sequentially. In the range 0–60% r.h. no conformational transition of the NA molecule occurs (Figs. 2b, c) and, therefore, neither conformational nor adsorption–desorption hysteresis phenomena are observed (Figs. 2a, b, c). In the range 0–90% r.h. only U→A conformational transition occurs (Fig. 3b), therefore the hysteresis loop shortens (Figs. 3a, b).

These results allows us to connect the observed hysteresis to the conformational changes in the NA molecule and consider it not as a macroscopic phenomenon like capillary hysteresis, but as natural property of the NA–water system. Our experimental and numerical results are in agreement with the data of other authors [13], [12], [14].

The adsorption–desorption hysteresis does not disappear or decrease during at least a week of exposure of the NA sample to a r.h. of 56%, this value being chosen because the adsorption hysteresis is the greatest at this r.h. The hysteresis lifetime is great enough to consider the hysteresis as a permanent phenomenon for the processes of the cellular regulation.

The probability of finding a nucleic acid unit in the certain conformation according to our results is never equal to the unit. It agrees with the idea that NAs are not static but fluctuating, "breathing", objects [23]. For example, in RNA molecule with 10^6 base pairs at the room temperature about 510 base pairs do not take part in the stacking and are not connected with H-bonds [2].

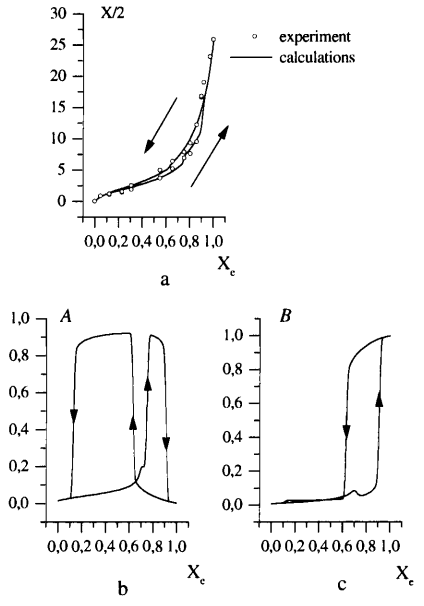


Fig. 1. The dependence of the stable stationary values of the adsorption and conformational variables on the control parameter, X_e. a–total adsorption per the mole of the nucleotides, b–the probability of finding of an arbitrary NA unit in the A form, c–the probability of finding of an arbitrary NA unit in the B–form. Parameters' values used to obtain numerical results: $v_{mL} = 3, a_L = 15.4, a_H^0 = 3.24, b^0 = 0.72$, $\Delta F_{UA}^p = 1.15, \Delta F_{AB}^p = 0.7075, \sigma = 1.575, \lambda^A = 0.165, \eta^A = 0.325, \beta^A = 0.065, \lambda^B = 0.1, \eta^B 0.3, \beta^B = 0.177$

The conformational transitions in the presented model take place according to the all–or–nothing law, i.e. they occur at the certain r.h. value. The same behaviour has been observed, for example, for the helix–coil transition of the model double–stranded structure $A(pA)_{17}{\cdot}U(pU)_{17}$ [24]. It is worth noting that this structure is homogeneous, the same is supposed in our model.

The model suggested allows one to investigate the possible autowave processes in the system. The autowave processes such as the trigger waves (propagating fronts) of a new conformation propagating along the NA molecule in the conformational transitions have been described in the frame of a more simple trigger model where only two conformational states of the NA have been distinguished and only Langmuir's adsorption has been taken into account [25]. Such propagating fronts agree with the well known "Zipper" model of the NA helix–coil transition [26]. Also, stochastic analysis based on the set of nonlinear master equations has been performed for the simple model. It has been shown that the multistability can be connected to the metastable states of the NA water system [25]. It seems highly plausible that the dy-

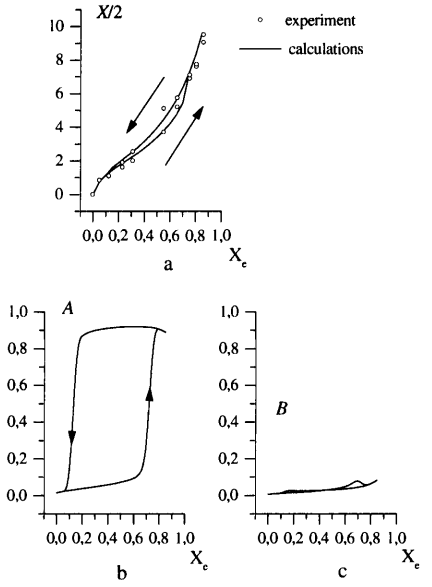


Fig. 2. The dependence of the stable stationary values of the adsorption and conformational variables on the control parameter, for $0 < X_e < 0.9$. a–total adsorption per the mole of the nucleotides, b–the probability of finding of an arbitrary NA unit in the A–form, c–the probability of finding of an arbitrary NA unit in the B–form.

namical behaviour of the simple model will be an intrinsic feature of more complete model of the three states suggested in the present study.

The model suggested can be easily extended to the case of inhomogeneous NAs by means of introducing the dependence of the model parameters on the number of the NA unit in the chain and solving (4) and (7) for every NA unit. This seems important as natural NAs such as DNA and RNA are inhomogeneous. The extension of the model on the case of more than three conformations also can be done easily.

References

1. Miyazawa, T.: Conformational aspects and biological functions of biomolecules. J. Mol. Struct. **126** (1985) 493–508
2. Saenger, W.: Prinsiples of nucleic acid structure. Springer–Verlag, Berlin (1984) 584
3. Goodfellow, J.M., Cruzeiro-Hansson, L., Norberto de Souza, O., Parker, K., Sayle, T., Umrania, Y.: DNA structure, hydration and dynamics. Int. J. of Radiation Biology **66** (1994) 471–478
4. Zehfus, M.H., Johnson, W.C.: Conformation of P–form DNA. Biopolymers **23** (1984) 1269–1281

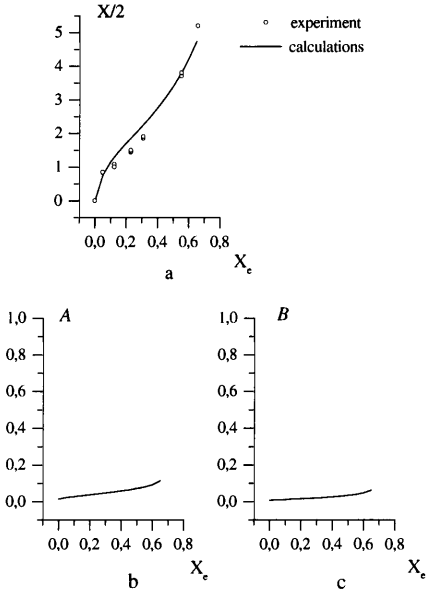


Fig. 3. The dependence of the stable stationary values of the adsorption and conformational variables on the control parameter, for $0 < X_e < 0.6$. a–total adsorption per the mole of the nucleotides, b–the probability of finding of an arbitrary NA unit in the A–form, c–the probability of finding of an arbitrary NA unit in the B–form.

5. Maleyev, V. Ya., Semenov, M. A., Gasan, A.. I., Kashpur, V. A.: Physical propoties of the DNA-water system. Biopysics **38** (1993) 789–811
6. Krumhansl, J.A., Wysin, G.M., Alexander, D.M., Garcia, A., Lohmdahl, P.S., Layane, S.P.: In: Structure and motions: membranes, nucleic acids and proteins. (Clementi, E., Corongin, G., Sarma, M.H., Sarma, R.H., eds) Adenine Press, New York (1985) 407–415
7. Christiansen, P.L., Muto, V.: Nonlinear models of DNA dynamics. Physica D **68** (1993) 93-96
8. Volkov, S.N.: Conformational transitions and the mechanism of transmission of long–range effects in DNA. Preprint ITP-88-12E, Kiev (1988) 22
9. Krumhansl, J.A., Alexander, D.M.: Nonlinear dynamics and conformational exitations in biomolecular materials. In: Structure and dynamics: nucleic acids and proteins. (Clementi, E., Sarma, R.H., eds) Adenine Press, New York (1983) 61–80
10. Theory of helix–coil transitions in biopolymers. (Poland, D., Scheraga, H.A., eds) Academic Press, New York (1970)
11. Eisenstein, M., Shakked, Z.: Hydration patterns and intermolecular interactions in A-DNA crystal structures. Implications for DNA recognition. J. Mol. Biol. **248** (1995) 662–678
12. Wetzel, R., Zirwer, D., Becker, M.: Optical anisotropy of oriented deoxyribonucleic acid films of different water content. Biopolymers **8** (1969) 391–401

13. Falk, M., Hartman, K.A., Lord, R.C.: Hydration of deoxyribonucleic acid. I. A gravimetric study. J. Am. Chem. Soc. **84** (1962) 3843–3846
14. Volkov, V.V., Gasan, A.I., Maleev, V.Ya.: The effect of glycerol on hysteresis phenomena in DNA. Preprint IRE–386, Kharkov (1989) 16 (in Russian)
15. Lahajnar, G., Zupancic, I., Rupprecht, A.: Proton NMR relaxation and diffusion study of water sorbed in oriented DNA and hyaluronic acid samples. In: Biophysics of Water (Franks, F., Mathias, S., eds) Wiley, New York (1982) 231–234
16. Kochoyan, M., Leroy, J.L.: Hydration and solution structure of nucleic acids. Current Opinion in Structural Biology **5** (1995) 329–333
17. Lindsay, S. M., Lee, S. A., Powell, J. M., Weidlich, T., DeMarko, C., Lewen, G. D., Tao, N. J. Rupprecht, A.: The origin of the A to B transition in DNA fibers and films. Biopolymers **27** (1988) 1015–1043
18. Ivanov, V.I., Minchenkova, L.E., Minyat, E.E., Schyolkina, A.K.: Cooperative transitions in DNA with no separation of strands. Cold Spring Harbor Symposia on Quantitative Biology **47** (1983) 243–250
19. Gascoyne, P.R.C., Pethig, R.: Experimental and theoretical aspects of hydration isotherms for biomolecules. J. Chem. Soc. Faradey Trans. **1** (1977) 171–180
20. Starikov, E.B., Bolbukh, T.V., Semenov M.A.: The study of nucleic acid hydration isotherms. Preprint IRE–359, Kharkov (1987) 19 (in Russian)
21. d'Arcy, R.L., Watt, J.C.: Analysis of sorption isotherm of nonhomogeneous sorbents. Trans. Faraday Soc. **66** (1970) 1236–1245
22. Nicolis, G., Prigogine, I.: Self–organization in nonequilibrium systems. John Willey & Sons, New York (1977) 512
23. Hippel, P.H. von, Wang, K.-Y.: Dynamic aspects of native DNA structure: kinetics of the formaldehyde reaction with calf thymus DNA. J. Mol. Biol. **61** (1971) 587–613
24. Porschke, D., Eigen, M.: Cooperative nonenzymic base recognition. III. Kinetics of the helix–coil transition of the oligoribouridylic oligoriboadenylic acid system and of oligoriboadenylic acid alone at acidic pH. J. Mol. Biol. **62** (1971) 361–381
25. Tolstorukov, M.Ye., Gatash, S.V., Maleev, V.Ya.: Self–organization and nonlinear dynamics of nucleic acid–water system. Special Issue of Int. J. Bif. Chaos (in press)
26. Porschke D.: Elementary steps of base recognition and helix–coil transitions in nucleic acids. Mol. Biol. Biochem. Biophys. **24** (1977) 191–218

Part II

Thermodynamic Modelling

Simulation Studies of Protein-Ligand Interactions*

Jan Hermans[1], Geoffrey Mann[1], Lu Wang[1,2], and Li Zhang[1,3]

[1] Department of Biochemistry and Biophysics, School of Medicine, University of North Carolina, Chapel Hill, NC 27599-7260, USA
[2] current address: Department of Pharmaceutical Chemistry, UCSF, San Francisco, CA, USA
[3] current address: Scripps Institute, La Jolla, CA, USA

Abstract. Protein-ligand interactions control a majority of cellular processes and are the basis of many drug therapies. First, this paper summarizes experimental approaches used to characterize the interactions between proteins and small molecules: equilibrium measurement of binding constant and standard free energy of binding and the dynamic approach of ligand extraction via atomic force microscopy. Next, the paper reviews ideas about the origin of different component terms that contribute to the the stability of protein-ligand complexes. Then, theoretical approaches to studying protein-small molecule interactions are addressed, including forced extraction of ligand and perturbation methods for calculating potentials of mean force and free energies for molecular transformation. Last, these approaches are illustrated with several recent studies from our laboratory: (1) binding of water in cavities inside proteins, (2) calculation of binding free energy from "first principles" by a new application of molecular transformation, and (3) extraction of a small ligand (xenon) from a hydrophobic cavity in mutant T4-lysozyme L99A.

1 Introduction

Protein-ligand interactions control a majority of cellular processes, as, enzyme catalysis, assembly of organelles, energy transduction, signaling, diverse control functions, and replication, expression and storage of the genetic material. Furthermore, protein-small molecules interactions provide the mechanism of many drug therapies, and are, therefore, a focus of interest in the pharmaceutical industry. Simulation studies have been used to address both fundamental and practical issues of protein-ligand interactions. If the goal is to reach a full understanding of structure, dynamics and structural stability, highly accurate simulations are preferable and computational cost can be high, as a single simulation may provide new insight. In the practical setting of a drug discovery process, many compounds must be screened in a short time, and

* This work has been supported by the U.S. National Science Foundation (grant MCB-9314854) and the U.S. National Institutes of Health's National Center for Research Resources (grant RR08102 to the UNC/Duke Computational Structural Biology Resource).

approximate models that can be evaluated more rapidly are preferred. In this paper, we discuss first some background and current ideas and then summarize results of some recent studies of protein-ligand interactions done in our laboratory.

2 Measures of the Stability of Protein-Ligand Complexes

Equilibria in Solution The stability of a protein-ligand complex in solution is measured in terms of the equilibrium constant and the standard free energy of association based on it. For association of species P and L in solution to form a complex PL, i.e., for

$$P+ \rightleftharpoons PL \tag{1}$$

the equilibrium constant (for association) is related to the concentrations of the three species under equilibrium conditions

$$K_a = \frac{c_{PL}}{c_P c_L} \tag{2}$$

and the difference in free energy for converting one mole of P plus one mole of L into one mole of PL under standard conditions is given by

$$\Delta A^\circ{}_a = -k_B T \ln K_a \tag{3}$$

Equilibrium constants for protein-small molecule association usually are easily measured with good accuracy; it is normal for standard free energies to be known to within ± 0.5 kcal/mol. Standard conditions define temperature, pressure and unit concentration of each of the three reacting species. It is to be expected that the standard free energy difference depends on temperature, pressure and solvent composition; $\Delta A^\circ{}_a$ also depends on an *arbitrary* choice of standard unit concentrations.

Ligand Extraction by Atomic Force Microscopy The recently developed atomic force microscope (AFM) allows measurement of the force required to extract a ligand molecule from a single protein-ligand complex, and provides an entirely new kind of information about the binding process. In a typical experiment, protein molecules are attached to a substrate, and ligand molecules are attached to the tip of the AFM probe. First, the probe is brought in contact with the substrate; then it is again moved away, while position and force are recorded. If in the contact phase a ligand molecule on the probe has managed to bind to a protein molecule on the substrate, a force will be required in the second phase in order to extract the ligand from the binding site. Experimental results on unbinding of biotin from its complex with streptavidin are among the first such results to have been obtained [9, 26].

3 Components of the Affinity

A decomposition of the binding free energy into component terms is useful as a means of gaining insight into the factors that contribute to the stability of the complex, or lack thereof [3]. Two important terms, the protein-ligand energy and the solvation free energy can be calculated rather easily: the former can be computed with the molecular mechanics force field, and averaged over a molecular dynamics simulation, and the latter can be estimated in terms of solvent polarization computed on the basis of a continuum dielectric model and an empirical hydrophobic free energy which is proportional to the molecular surface. However, some other component terms are very difficult to estimate. Consequently, this approach is most useful when the latter terms can be assumed to be constant, as for a series of related ligands, for which differences in binding free energy can then be estimated from the remaining components, and lends itself to high-volume tasks, such as screening a library of small molecules to find likely inhibitors of an enzyme when the structure of the binding site is known, e.g. [20].

Furthermore, an actual or conceptual decomposition is useful because it can lead to a better appreciation of factors underlying the binding energetics. We consider here the following four components.

1. Direct ligand-protein interactions. Van der Waals and Coulomb energy of interaction of atoms of ligand with atoms on protein.

2. Solvation. Difference in solvation of protein and ligand molecules in the complex and in isolation.

3. Confinement. The ligand molecule is confined to a small volume and a narrow range of orientations, that correspond to the freedom of the ligand in the bound state.

4. Conformational adjustments. The conformations of the protein and the ligand are adjusted to meet the requirements of the complex.

One should realize that this decomposition is, for more than one reason, fundamentally flawed: it focuses on single conformations, not taking into account proper averaging over an equilibrium distribution; it tends to ignore cross terms, such as a dependence of the first two terms on conformational adjustments; quite generally, a valid decomposition should be specified, at least conceptually, in terms of a sequence of steps in the reaction scheme (such as, transfer from solution to vacuum, confinement, conformational adjustment and placement of the ligand in the binding site). In fact, the following discussion shows that some of the terms can be computed, but that for others no computation has yet been developed. Thus, this decomposition may be useful as an aid to understanding, but does not provide a useful route to calculation of binding free energies [23].

Ligand-Protein Interactions The energy of formation of ligand-protein contacts can be computed as a sum of non-bonded (Lennard-Jones and electrostatic) terms similar to those used in a molecular dynamics simulation.

The direct ligand-protein interactions and the net solvation-desolvation term together should give an energy contribution that strongly favors formation of the complex (large and negative), since the other two components favor its dissociation.

Solvation-Desolvation In the formation of the complex, the individual components are desolvated, and the complex is solvated; the net result is a decrease of the solvent-exposed surface. The corresponding free energy changes are often large; each can be treated as the sum of two components, one for the creation of a solute-solvent interface, and the other for the polarization of the solvent. It has become common practice to estimate the first of these as the product of a constant factor and the molecular surface area, and to compute the second with a model in which a continuum of high dielectric constant surrounds a molecular volume of low dielectric constant, containing explicit atomic charges. The energetics of this model are computed by a numerical solution of the Poisson-Boltzmann equation [32, 33]; the charge distribution is the same as that used in a molecular dynamics simulation, and the molecular surface is defined with the aid of a set of standard atomic radii [4]. This calculation gives an estimate of the free energy, i.e. contains both enthalpic and entropic contributions. These estimates of solvation free energy are reasonably accurate, and require a modest computational effort. A drawback is that the solvation free energy is calculated for a single conformation of protein and ligand. Optimization of the complex with inclusion of the gradient of the solvation free energy is possible, but time consuming. It is much faster to equilibrate the complex in a molecular dynamics simulation with explicit solvation, which can also produce a sample of representative solvated conformations [36].

Confinement of the Ligand A large free energy contribution results from the effective confinement of the ligand in the narrow space of the binding site. In the free state, the center of mass of the ligand molecule has available a volume of 1 L/mol = 1660 Å^3, while in a tightly bound state the center of mass may be restricted to within ± 0.5 Å, so that the available volume is only 1 Å^3. Such a de-facto compression by a factor of 1660 corresponds to a change in free energy of $k_B T \ln 1660 = 4.4$ kcal/mol. Similarly, the free ligand has completely free orientation with respect to three degrees of freedom, while the orientation of a tightly bound ligand is limited to within a few degrees. If the range is reduced to within $\pm 10° = \pm 0.18$ rad for each of the three degrees of freedom, then this would give a further change of free energy by $k_B T \ln(8\pi^2/0.36^3) = 4.5$ kcal/mol. One can determine the positional and orientational freedom of the ligand with a simulation of the dynamics of the ligand in a *rigid* binding site, and from that compute an estimate of the confinement free energy term. However, this is not possible if the protein is flexible, because then changes in position and orientation of the ligand may

correspond either to displacement of the ligand in the binding site, or to displacement of the binding site as an intrinsic part of the dynamics of the protein without displacement of the ligand relative to the binding site, and it is not clear how to separate these two effects.

The confinement term is unique because it alone causes a dependence of the binding free energy on the choice of unit concentration in the standard state: the volume available per ligand molecule in the free state, and hence the compression factor, depend on the unit concentration.

The importance of the confinement term was recognized early by Jencks, who attempted to obtain at least a qualitative estimate [28]. However, Jencks estimated this "cratic" contribution with an expression in terms of mole fractions, which has turned out to be unsuited [31]. Jencks realized correctly that the interaction energy of a protein with even a weakly bound ligand must be considerable, because the cratic free energy term must be overcome. This led him to the following important principle that is useful in rational design of enzyme inhibitors: if two weak ligands are combined into a single molecule with the correct geometry for each part to interact favorably with a different partner in the binding site, then this new molecule has unexpectedly high affinity for the protein, because the two favorable energetic contributions together more than outweigh the single confinement term [29].

Conformational Adjustments The conformations of protein and ligand in the free state may differ from those in the complex. The conformation in the complex may be different from the most stable conformation in solution, and/or a broader range of conformations may be sampled in solution than in the complex. In the former case, the required adjustment raises the energy, in the latter it lowers the entropy; in either case this effect favors the dissociated state (although exceptional instances in which the flexibility increases as a result of complex formation seem possible). With current models based on two-body potentials (but not with force fields based on polarizable atoms, currently under development), separate intra-molecular energies of protein and ligand in the complex are, in fact, definable. However, it is impossible to assign separate entropies to the two parts of the complex.

4 Simulations of Protein-Ligand Interaction

Free Dynamics In simulations one usually represents a single protein molecule and one or a few ligand molecules. In principle, one might then obtain an estimate of the binding constant by monitoring the state of the protein during a long simulation in which ligand were observed to bind and unbind many times, and determining the fraction of time, x_{PL} during which a ligand molecule was bound, and then, by equating time-average with ensemble-average properties, write $K_a = x_{PL}/(x_P c_L)$, with $x_P = 1 - x_{PL}$. However, the association and dissociation rates will nearly always be too slow to make

this an option; especially if the complex is tight, it might, once formed, not ever dissociate during a simulation. Instead, alternative strategies have been developed to assess the stability of protein-ligand complexes with theoretical models. In these methods, the binding free energy is obtained from simulations in which the force field is *perturbed* with terms that force dissociation or formation of the complex; the free energy change can be found from the work done by the external force required to maintain the perturbation.

Potentials of Mean Force The conceptually simplest way to force the formation of a complex is to constrain the distance between the ligand and the binding site, while gradually reducing this distance. Then, in order for the ligand to approach the binding site along a path, s, an external force, F_s must be applied that balances the net internal force on the ligand. The work, W_s performed by the external force includes the free energy change and a contribution, W_F required to overcome friction

$$W_s = \Delta A_s + W_F \tag{4}$$

In the limit of very slow change (quasi-static process) the frictional component is zero and then the work done by the external force equals the free energy change, i.e.,

$$\Delta A_s = \lim_{t_s \to \infty} \int F_s \, ds \tag{5}$$

Calculations at increasingly longer simulation times, t_s are done to verify convergence [13]. In the slow change method, the integral is approximated in a simulation in which s is changed by a small amount, δs after each integration step,

$$\Delta A_s \cong \sum F_s \, \delta s \tag{6}$$

Alternatively, the free energy can be found as the integral of the mean force

$$\Delta A_s = \int < F_s > ds \tag{7}$$

Such a free energy is called a potential of mean force. Average values of F_s can be computed in dynamics simulations (which sample a Boltzmann distribution), and the integral can be estimated from a series of calculations at several values of s. A third method computes the free energy for perturbing the system by a finite step in s, for example, from s_1 to s_2, with

$$\Delta A_{1,2} = \Delta A_2 - \Delta A_1 = -k_B T \ln < \exp[-\Delta U_{1,2}/k_B T] >_{s_1} \tag{8}$$

the averaging being done for the configurations sampled in a simulation which is run for $s = s_1$, and $\Delta U_{1,2}$ the energy for perturbing the system from its current configuration. Typically, the value of ΔA_s is found as the sum of many contributions $\Delta A_{1,2}$, for successive small increments of s.

Forced Extraction of Ligand Potential of mean force calculations approximately mimic ligand extraction processes that can be studied with AFM. One may subject the ligand to a constant force and observe its displacement out of the binding site versus time, subject the ligand to a steady displacement and monitor the required force, or apply the force via a spring, of which one end is attached to the ligand, and the other end is moved at a steady rate. (The third option most closely approximates the mechanics of the AFM experiment.) Two groups have published the results of simulations of the forced extraction of biotin from its complex with streptavidin [7, 19]. (See also separate contributions in this volume by Helmut Grubmüller and by Stepaniants, Izrailev, Wriggers and Schulten.)

Molecular Transformation In a second method to compute the changes in free energy that result from binding, the perturbation is applied to the molecular mechanics force field. The reasoning behind this approach is that if the forces between ligand and protein and solvent are reduced to zero, the protein interacts only with the solvent and hence is represented in the state without ligand bound, while the ligand molecule is then in a vacuum. In other words, in the process of reducing the forces between protein and ligand, the ligand is transferred from the bound state to an ideal gas state. The free energy change can be found by the same methods as used to compute a potential of mean force, if an effective force driving the transformation is defined with

$$F_{tr} = -\partial U/\partial\lambda \tag{9}$$

where λ is a parameter that controls the value of the forces between protein and ligand: for $\lambda = 0$ these are zero, and for $\lambda = 1$ they are applied at full strength. The free energy can be found as the limiting value for long simulation time, t_s , with

$$\Delta A_{tr} = \lim_{t_s \to \infty} \int_0^1 -\frac{\partial U}{\partial \lambda}\, d\lambda \tag{10}$$

and can be evaluated with any of the three methods (6-8). The transformation process corresponds to the following binding reaction,

$$P + L^{(v)} \rightleftharpoons PL \tag{11}$$

where the superscript (v) indicates the ideal gas state. (The standard free energy for this equilibrium differs from that for the equilibrium in water by an amount equal to the free energy to transfer the ligand from water to vacuum.)

Molecular transformation calculations can be used for computing *differences in binding free energy* for two ligands by a well-established scheme

based on a reaction cycle. Let K and L be two different ligands binding in the same site, then we can write the following cycle of four reactions, in clockwise order, binding of L to P, transformation of PL to PK, dissociation of PK and transformation of K to L

$$\begin{array}{ccc} P+L & \Rightarrow & PL \\ \Uparrow & & \Downarrow \\ P+K & \Leftarrow & PK \end{array}$$

Since the net free energy change for the cycle is zero, the difference between the *computable* free energies for transforming L to K, when bound to protein and when dissolved in water, is equal to the difference between the *measurable* standard free energies of binding L and K to protein, i.e., the difference in affinity

$$\Delta A_{tr,P} - \Delta A_{tr,w} = \Delta A^{\circ}{}_{a,L} - \Delta A^{\circ}{}_{a,K} \tag{12}$$

This principle has been applied in a contribution by Mark, Schäfer, Liu and van Gunsteren to this volume, and in section 6 of this article. For a review of free energy perturbation methods see [8].

5 Binding of Water in Cavities Inside Protein Molecules

This section summarizes the results of a study of internal hydration of protein molecules, based on a very simple approach, in which only the intermolecular energy of protein and ligand was considered, and also describes the Dowser tool that was developed as a result of that study [37]. Water molecules inside cavities in proteins constitute integral parts of the structure. In most of the filled cavities, the internal water molecules are held with two or more hydrogen bonds, while cavities without hydrogen bonding groups on the surface are empty. Due to experimental error and interpretative uncertainty of electron density maps, internal water positions cannot always be unequivocally assigned; the problem is worse for structures determined at lower resolution. We have sought a quantitative measure of the hydrophilicity of the cavities by calculating the energy of introducing a water molecule into a cavity, using the known structure of the protein and standard molecular mechanics energies. In a survey of a number of proteins, it was found that a threshold value of the water-protein interaction energy at -12 kcal/mol distinguished hydrated from empty cavities. In one instance of two independent crystallographic determinations of the same structure [2, 24], we were able to conclude on the basis of these energies and additional crystallographic information (occupancy and B-factor) that in one structure many more buried water sites had been assigned than were, in fact, physically present.

Clearly, the value of -12 kcal/mol for the threshold energy is not accidental. Buried water molecules are in equilibrium with water molecules in the

bulk solvent. Therefore, the protein must provide to a buried water molecule an attractive environment at least as good as the environment of liquid water or ice, with two hydrogen bonds per water molecule. The intermolecular energy of ice is -12 kcal/mol and for water it is -10 kcal/mol. (Even at room temperature the free energy of water is very close to that of ice.) To have similar or lower free energy, an internal water molecule can have an energy below -12 kcal/mol and then it may have as low an entropy as in ice. To be bound with energy above -12 kcal/mol, it would have to have greater freedom. In most cases, the energy is provided by formation of two or more hydrogen bonds to groups on the protein, and the bound water molecule has little freedom to move. Only in rare instances a water molecule may be held internally with higher entropy and energy and stabilized by other than hydrogen-bonding interactions, as proposed for interleukin-1β [5].

Dowser Program On the basis of these results, we have developed a method for locating likely binding sites for water molecules in the interior of protein molecules of known structure. Internal cavities (and optionally crevices) are located with Connolly's molecular surface program [4]; in each cavity, the placement of a water molecule that gives lowest interaction energy is determined. If the energy is above a limit, the site is taken to be empty. By iteration it is also possible to locate more than one water molecule in a cavity [16]. The Dowser tool has been used to locate water molecules in cytochrome-c oxidase. This protein is a redox-driven pump which couples the reduction of oxygen to water to the translocation of protons across a lipid bilayer membrane. Dowser was able to locate water channels that constitute the route of entrance and exit of the pumped proton; on the basis of this structure, a model for the protein's function could be proposed [17].

6 Simulation of Small Ligands Bound in T4-lysozyme L99A

6.1 Absolute Binding Constants from "First Principles"

The problems that occur when one tries to estimate affinity in terms of component terms do not arise when perturbation methods are used with simulations in order to compute potentials of mean force or free energies for molecular transformations; simulations use a simple physical force field and thereby implicitly include all component terms discussed earlier. We have used the molecular transformation approach to compute binding affinities from these "first principles" [14]. The basic approach had been introduced in early work, in which we studied the affinity of xenon for myoglobin [11]. The procedure was to gradually decrease the interactions between xenon atom and protein, and compute the free energy change by standard perturbation methods, cf. (10). An essential component is to impose a restraint on the

position of the xenon, which changes concurrently with the interactions between xenon and protein; the restraint is in full effect when the interactions are absent, and zero when the interactions are at full strength. The restraint is essential for two reasons: by keeping the non-interacting ligand molecule in the binding site, it makes the process reversible, and, by confining the isolated ligand molecule to a small volume, it defines a standard state whose free energy (chemical potential) can be related to that of a more common standard state by a simple formula.

Our recent work on the affinity of benzene for a cavity inside the structure of T4 lysozyme mutant L99A [6] has extended this method to polyatomic molecules; this is achieved by coupling the transformation to a "body restraint" potential which restrains not only the position but also the orientation of the ligand molecule in the absence of interactions with the protein; i.e. all six external degrees of freedom of the ligand (with simple extension to internal degrees of freedom, if needed) are restrained, rather than the three external degrees of freedom needed for a xenon atom [14] (See also discussion by Gilson et al. [10]). In our work, the following restraint potential has been used

$$\begin{aligned}(1-\lambda)\,U_r &= (1-\lambda)\,(U_{r,x}+U_{r,\theta}+U_{r,\chi}) \\ &= (1-\lambda)\,[\frac{K_x}{2}(\mathbf{X}_{r,l}-\mathbf{X}_{r,p})^2+\frac{K_\theta}{2}(1-\cos\theta)+\frac{K_\chi}{2}\chi^2] \quad (13)\end{aligned}$$

The term $U_{r,x}$ restrains a point in the ligand, $\mathbf{X}_{r,l}$ to a reference position, $\mathbf{X}_{r,p}$ defined in terms of the protein coordinates, the term $U_{r,\theta}$ restrains the angle, θ between a vector in the ligand and a reference vector in the protein, and the term $U_{r,\chi}$ restrains the angle, χ between a plane formed by three points in the ligand and a reference plane formed by three points in the protein. A "point" can be an atomic position or the center of mass of several atoms (of ligand or protein); instead of a point on the protein, one may also use a fixed reference position. As a result of multiplication by $(1-\lambda)$, the restraint is applied when the ligand-protein energy is "off", and not applied when it is "on", cf. (9) and (10).

Clearly, a free energy of binding computed with (9), (10) and (13) refers to a highly restricted state of the dissociated ligand. In order to convert such a free energy to a free energy relative to a normal standard state with volume per molecule V_o and no restriction on the molecular orientation, the following term must be added

$$\Delta A_r = -k_B T\,[\ln(z_x/V_o)+\ln(z_\theta/4\pi)+\ln(z_\chi/2\pi)] \quad (14)$$

where

$$z_x = \int \exp(-\frac{U_{r,x}}{k_B T})\,d\mathbf{X}, \quad (15)$$

and similar expressions for z_θ and z_χ. (A correction term is required for symmetrical molecules).

The computed free energy of binding depends on the choice of standard state for the ligand as a result of the explicit dependence of ΔA_r on V_o. In our original paper [14] we have cited a number of articles in the recent literature in which the binding free energy had been estimated incorrectly with omission of this term. The agreement between theoretical and experimental values of the free energy was very good in both of our studies with this method [11, 14], which is generally applicable to produce estimates of the binding affinity of enzymes for typical substrates and inhibitors.

6.2 Estimates of Component Term for Conformation Adjustment and "Cratic" Contribution for Binding Benzene

Good agreement between experimental and theoretical values is strictly necessary as a validation of the methodology, however, relatively little insight is gained from the ability, per se, to compute the binding free energy. However, we were able to gain considerable insight into the factors determining the affinity, by a series of simulations of binding in which the protein's conformation was held fixed. The contribution of protein flexibility was in evidence from the differences in affinity for a series of different conformations of the protein (Table 1). The affinity is very low if the energy of the protein has been minimized in the absence of the ligand, intermediate in the structure that has been fit to the ligand by energy minimization and is highest for conformations that have adapted to the presence of the ligand in a *dynamic* situation, i.e., either the experimental structure or a series of snapshots of a molecular dynamics simulation of the complex. Comparison of the latter values of around -8.5 kcal/mol with the net value of -5.5 kcal/mol for binding to a dynamic protein shows that the protein's conformation is altered to adapt itself to the presence of the ligand, at a cost of 3 kcal/mol. We believe this to be the first such estimate.

Table 1. Computed free energies of binding of benzene to different rigid models of the protein.

Protein conformation	$\Delta A^{\circ}_{a,1\,M\ \text{gas}}$ (kcal/mol)
Experimental (crystal) coordinates	−8.5
Energy minimum without benzene	−3.5
Energy minimum with benzene	−6.3
Dynamics snapshot at time t_o	−6.8
Dynamics snapshot at time $t_o + 2$ ps	−8.2
Dynamics snapshot at time $t_o + 4$ ps	−8.9
Dynamics snapshot at time $t_o + 6$ ps	−8.3
Dynamics snapshot at time $t_o + 8$ ps	−7.5

Also by using the results of dynamics simulations with a rigid protein molecule, we were able to quantitate the "cratic" component, as approximately 8 kcal/mol, and attribute this roughly to a contribution of 5 kcal/mol for positional confinement and 3 kcal/mol for loss of rotational freedom (relative to a molecule in 1 M gas). This was found to correspond to remaining freedom of ± 0.2 Å in the three translations and $\pm 6°$ in the three rotations (taking into account also a free energy contribution in favor of binding of $-k_B T \ln 12$ for the high molecular symmetry of benzene). These numbers were found to agree closely with the deviations of position and orientation from the mean accessed by the benzene molecule in a dynamics simulation. Here, these deviations were estimated from the cratic free energy; conversely, the cratic free energy can be estimated accurately from observed deviations. (However, note that this analysis was possible thanks to use of a rigid protein model in these simulations.)

The method described here becomes less suitable as the size of the ligand molecule increases. The problem of computing the "cratic" term for formation of complexes of two protein molecules have been discussed by others [34][35].

6.3 Comparison of Affinities of Several Ligands

We have applied simulations to the binding of several other molecules as ligands in the cavity in T4 lysozyme mutant L99A (Table 2). No binding data were available at the time for these molecules (except benzene); we were, however, aware that binding of the noble gases had been observed, and that structure determination by x-ray crystallography was in progress. The computed binding free energies agree well with observations that binding is observed at 1 atm pressure of xenon (concentration 0.0044 M), but that higher pressures are needed to observe binding of krypton and argon [30]. The crystallographic structures showed that two atoms of each noble gas were bound in the cavity, something the simulations had not taken into account. Cyclohexane is not expected to bind inside the cavity, and its binding also has not been reported; the high protein-ligand energy in the equilibrated complex indicates that the puckered cyclohexane ring is too bulky for the cavity. The water-protein energy in a complex with a water molecule in the cavity is well above the threshold that was established in our study of buried water molecules, and observation of a water molecule bound inside this apolar cavity is not to be expected. In fact, after a 1 ps simulation, the water molecule escaped from the cavity into the solution.

As an approximation, one may equate the ligand-protein energy with the binding energy; by subtracting this from the free energy of binding, one then obtains a remainder which is dominated by the cratic free energy component. (There is no contribution from solvation in this case.) One sees in table 2 that this is much smaller for the noble gases; the difference corresponds to the fact that of benzene orientation and position are restricted in the binding site, while of xenon, krypton and argon only the position is restricted. An

Table 2. Table Ligand-protein interaction energies and free energies (kcal/mol).

Ligand	$< E_{intra} >$	$\Delta A^\circ{}_a$ (simulation)	$\Delta A^\circ{}_a - < E_{intra} >$	$\Delta A^\circ{}_a$ (experiment)
benzene	−13.2	−5.5	7.7	−5.2
xenon	−4.7	−2.5	2.2	< -1.9
argon	−2.3	−0.5	1.8	> -1.9
krypton	−2.4	−0.1	2.3	> -1.9
cyclohexane	+4.9			"too fat"
water	−3.			"too polar"

entropic term of about 2 kcal/mol corresponds to a rather considerable free volume of circa 60 Å^3 available to a noble gas atom bound in the cavity, which is not unreasonable, since the volume of the cavity has been estimated at circa 150 Å^3. It is also understandable that there is room for a second noble gas atom to bind; binding of the second xenon atom (in a smaller space) will be stabilized by a favorable interaction with the first.

7 Extraction of Bound Xenon from Mutant T4-Lysozyme

A problem with comparing the results of ligand extraction simulations and experiments is the large difference in rate of extraction: in the computer simulations the extraction is performed in hundreds of ps, but in an experimental setting the time is much longer, of the order of ms. As a result, the contribution of irreversible effects (i.e., friction) to the extraction forces will be much larger in the simulations, and it is not clear if the simulation results can be extrapolated to the much longer time scale of the experiments. Two papers reporting simulations of extraction of biotin from streptavidin differ in interpretation: in the first an extrapolation of the simulation results to longer times gave apparent agreement with the experimental results [7], while the second paper concluded that such a long extrapolation was not feasible [19].

We have chosen to study the extraction of the xenon atom from its binding site inside the hydrophobic cavity in mutant T4 lysozyme as a simple system in which to model the ligand extraction process. The internal binding site in this mutant is hydrophobic and excludes water; as a result, an important source of friction in the extraction of a ligand (the simultaneous entry of water molecules) is absent. On the other hand, this system shares with the avidin-biotin system the requirement for a distortion of the geometry at the exit point in order to permit the ligand to escape. With long, but feasible, simulations it may therefore be possible to approach conditions of very slow extraction and hence small friction, in which the extraction force is dominated by the change in free energy (Cf. eq. 4). We describe first the interactive simulations in which we located an exit path for the xenon atom, and then the results of a series of extractions performed at different rates.

7.1 Steered Dynamics Tool (SMD)

SMD refers to a system for interactively steering molecular dynamics calculations by adding user-specified external forces into the computation on the fly [21]. (We call these forces "tugs".) SMD consists of two software components. The computational component is a version of the SigmaX molecular dynamics program and is responsible for computing the dynamics of the system under study, including its response to user-defined external forces [15]. The display component is VMD, a molecular graphics code written as part of the MDScope project [27] at the University of Illinois [18]. (The MDScope developers have also created a steered dynamics interface to their NAMD dynamics code using VMD; cf. an article by Stepaniants, Izrailev, Wriggers and Schulten in this volume.)

The purpose of introducing external forces is to help the molecular system overcome energy barriers between states. One application of this is to steer the system to a proposed new geometric conformation, which can be further studied. In other cases, the forces model physical experiments in which molecular complexes are pulled apart via atomic-force microscopy. The experiments in turn provide some validation of the computed model. The difficulty of changing the conformations in the computational model gives insight into physically realizable pathways. SMD provides a new tool for biochemists to use in exploring the structure of proposed designs, as well as in more general applications such as exploring the molecular dynamics model itself. Its primary use is in modeling single large biomolecules in a bath of water acting as the solvent.

In our implementation of SMD, modified versions of VMD and Sigma communicate with each other using a customized, lightweight protocol. Sigma sends atomic positions resulting from each molecular dynamics time step to VMD for display. When the user specifies restraints on parts of the displayed model, VMD sends them to Sigma, where they are converted into potential-well restraints added to the force field [21].

A likely exit path for the xenon was identified as follows. Different members of our research group placed the exit path in the same location and were able to control extraction of the xenon atom with the "tug" feature of the steered dynamics system without causing exaggerated perturbations of the structure. The exit path is located between the side chains of leucines 84 and 118 and of valine 87; the flexible side chain of lysine 83 lies just outside the exit and part of the time is an obstacle to a linear extraction (Fig. 1).

7.2 Extractions at Different Rates

The interactive studies were followed by a series of batch calculations in which the xenon atom was dragged at a constant rate from the cavity to a location well into the solvent (total distance of approximately 1.8 nm, depending on the starting location of the xenon atom) in fixed time intervals between 1

Fig. 1. Exit route of xenon in simulations of the extraction process. The xenon atom is solid black. The atoms of the residues surrounding the exit path are shown as spheres, and the protein backbone is shown as a thin curve. On the left, the xenon is viewed "through" the exit between residues; on the right, the view is from the side and the direction of the "tug" is marked with a line.

and 500 ps [22]. (Such a simulation approximates an AFM experiment with a very stiff lever.) For each extraction time, calculations were done starting from different instants in a prior 2-ns long simulation of the complex. These simulations were done with a force field that included explicit solvation with SPC water molecules, periodic boundary conditions and Ewald summation.

The work required in order to move the xenon atom was computed by the slow-change method, according to (6). Results for 100 ps extractions are shown in Fig. 2. Near the starting point, i.e. in the bound state, the xenon is moved relatively easily, considerable work is required to move it through the layer of hydrophobic side chains, and there is evidence for a small energy barrier before the xenon leaves the protein at approximately 0.8 nm; the work performed beyond that point averages to that required according to Stokes' law to drag a sphere of radius 0.22 nm through a continuum with the viscosity of water.

The dependence of the mean work performed in the extraction for different extraction times is shown in Fig. 3.[1] One sees that in very rapid extractions a very large amount of work is required to overcome the friction, which decreases as the extraction is done more slowly; ultimately, only a small amount of work remains to be done to compensate the free energy change for transferring the hydrophobic ligand into the solvent. In a simple system, the friction

[1] Results are reported in energy units that correspond to those in experimental papers: molar units for binding equilibria (kcal/mol) and molecular units for ligand extraction experiments (zJ = pN nm). 1 kcal/mol = 6.9 zJ, and at 300 K, $k_B T = 4.2$ zJ.

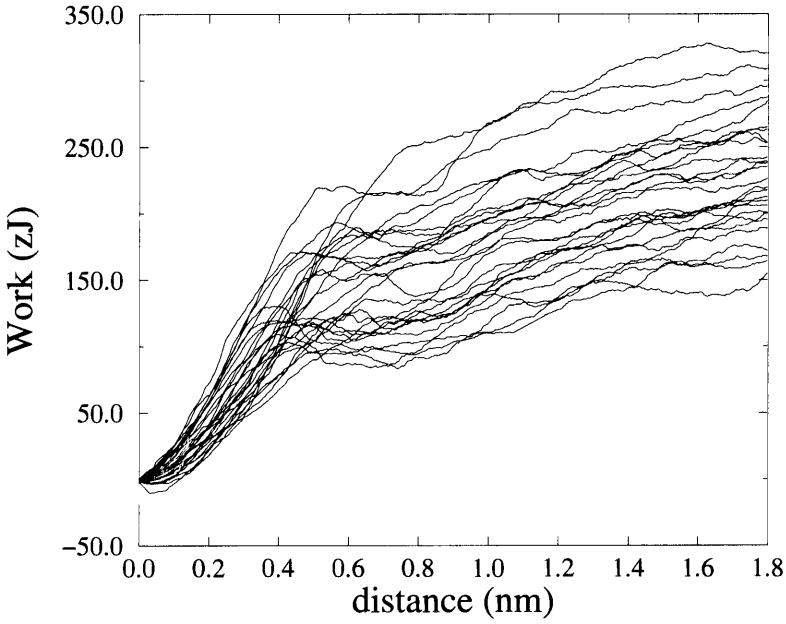


Fig. 2. Work performed to extract xenon from T4 lysozyme L99A in 27 independent simulations of 100 ps each.

is expected to be proportional to the rate of displacement; in this system, the dependence is apparently more complex.

A linear dependence approximately describes the results in a range of extraction times between 1 ps and 50 ps, and this extrapolates to a value of W_s not far from that observed for the 100 ps extractions. However, for the simulations with extraction times, $t_s > 50$ ps, the work decreases more rapidly with $1/t_s$, which indicates that the 100 ps extractions still have a significant frictional contribution. As additional evidence for this, we cite the statistical error in the set of extractions from different starting points (Fig. 2). As was shown by one of us in the context of free energy calculations[12], and more recently again by others specifically for the extraction process [1], the statistical error in the work and the frictional component of the work, W_F are related. For a simple system obeying the Fokker-Planck equation, both friction and mean square deviation are proportional to the rate, and

$$< (\delta W)^2 >= 2k_B T \times W_F \tag{16}$$

A lack of simplicity of the present system follows directly from our results, according to which $(< (\delta W)^2 >)^{1/2}$, and not $< (\delta W)^2 >$, varies approximately as the extraction rate (results not shown). Hence, it is not possible to

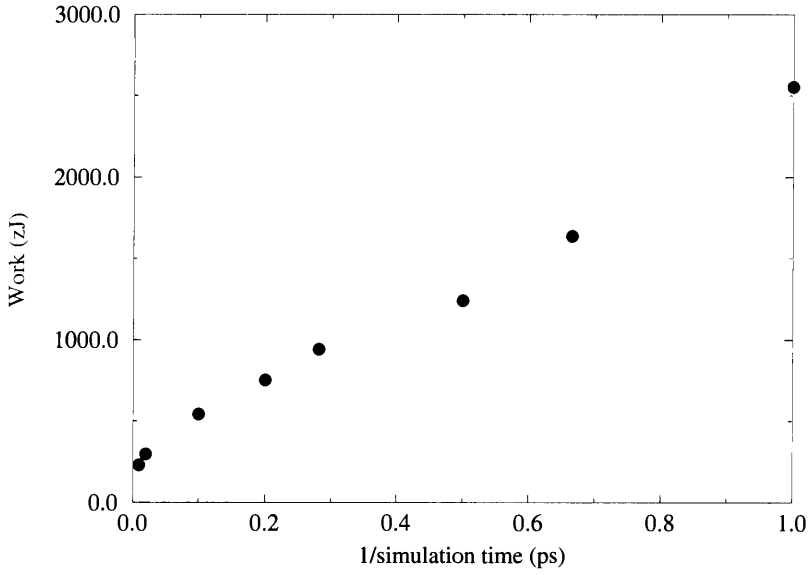


Fig. 3. Work required to extract xenon from T4 lysozyme L99A. Each value is a mean over several independent simulations.

calculate the frictional component of the work from the mean square deviation with (16); however, it is still obvious that the mean square deviation of 2000 $(zJ)^2$ observed over 27 extraction experiments each of 100 ps duration corresponds to a significant frictional component. [A rough estimate of the free energy for extracting the xenon atom is the sum of (minus) the protein-xenon energy observed in simulations (-5 kcal/mol) and the (experimental) free energy to transfer xenon from vacuum to water (1.5 kcal/mol), i.e. 6.5 kcal/mol = 45 zJ.]

Thus, we have found unexpected complexities and even in this simple system have not yet been unable to accurately extrapolate the results of simulations done over periods varying from 1 to several hundred ps, to the low-friction conditions of extraction experiments performed in times on the order of ms. The present results indicate that one should not expect agreement between extraction experiments and simulations in more complex situations typically found in experiments, involving also a reverse flow of water molecules to fill the site being evacuated by the ligand, unless the simulation times are prolonged well beyond the scope of current computational resources, and thereby strengthen the conclusion reached in the second theoretical study of extraction of biotin from its complex with avidin [19].

References

1. Balsera, M., Stepaniants, S., Izrailev, S., Oono, Y., Schulten, K.: Reconstructing potential energy functions from simulated force-induced unbinding processes. Biophys. J. **73** (1997) 1281-1287
2. Bode, W., Papamokos, E., Musil, D.: The high-resolution X-ray crystal structure of the complex formed between subtilisin Carlsberg and eglin c, an elastase inhibitor from the leech Hirudo medicinalis. Eur. J. Biochem. **166** (1987) 673-692
3. Boresch, S., Archontis, G., Karplus, M.: Free energy simulations: The meaning of the individual contributions from component analysis. Proteins: Str. Funct. Genet., **20** (1994) 25-33
4. Connolly, M. L.: Analytical molecular surface calculation. J. Appl. Crystallogr. **16** (1983) 548-558
5. Ernst, J. A., Clubb, R. T., Zhou, H.-X., Gronenborn, A. M., Clore, G. M. Demonstration of positionally disordered water within a protein hydrophobic cavity by NMR. Science **267** (1995) 1813-1815
6. Eriksson, A. E., Baase, W. A., Wozniak, J. A., Matthews, B. W.: A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. Nature **355** (1992) 371-373
7. Grubmüller, H., Heymann, B., Tavan, P. Ligand binding and molecular mechanics calculation of the streptavidin-biotin rupture force. Science **271** (1996) 997-999
8. van Gunsteren, W. F.: Methods for calculation of free energies and binding constants: Successes and problems. In Computer simulations of biomolecular systems. ESCOM, Leiden. (1989) 27-59
9. Florin, E. V., Moy, T. V., Gaub, H. E.: Adhesion forces between individual ligand-receptor pairs. Science **264** (1994) 415-417
10. Gilson, M. K., Given, J. A., Bush, B. L., McCammon, J. A.: The statistical-thermodynamic basis for computation of binding affinities: A critical review. Biophys. J. **72** (1997) 1047-1069
11. Hermans, J., Subramaniam, S.: The free energy of xenon binding to myoglobin from molecular dynamics simulation. Isr. J. Chem. **27** (1986) 225-227
12. Hermans, J.: A simple analysis of noise and hysteresis in free energy simulations. J. Phys. Chem. **95** (1991) 9029-9032
13. Hermans, J., Yun, R. H., Anderson, A. G.: Precision of free-energies calculated by molecular dynamics simulations of peptides in solution. J. Comp. Chem. **13** (1992) 429-442
14. Hermans, J., Wang, L.: Inclusion of loss of translational and rotational freedom in theoretical estimates of free energies of binding. Application to a complex of benzene and mutant T4-lysozyme. J. Am. Chem. Soc. **119** (1997) 2707-2714
15. Hermans, J. Sigma documentation. University of North Carolina. (1994) http://femto.med.unc.edu/SIGMA
16. Hermans, J., Zhang, L., Xia, X. Dowser documentation. University of North Carolina. (1994) http://femto.med.unc.edu/DOWSER
17. Hofacker, I., Schulten, K.: Oxygen and proton pathways in cytochrome-c oxidase. Proteins: Str. Funct. Genet. **29** (1998) 100-107
18. Humphrey, W.F., Dalke, A., Schulten, K.: VMD - Visual molecular dynamics. J. Mol. Graphics **14** (1996) 33-38

19. Izrailev, S., Stepaniants, S., Balsera, M., Oono, Y., Schulten, K.: Molecular dynamics study of unbinding of the avidin-biotin complex. Biophys. J. **72** (1997) 1568-1581
20. Jedrzejas, M. J., Singh, S. Brouillette, W. J. Air, G. M. Luo, M. A.: 1995. Strategy for theoretical binding constant, Ki calculation for neuraminidase aromatic inhibitors, designed on the basis of the active site structure of influenza virus neuraminidase. Proteins: Struct. Funct. Genet. **23** (1995) 264-277
21. Leech, J., Prins, J., Hermans, J.: SMD: Visual steering of molecular dynamics for protein design. IEEE Computational Science & Engineering **3**(4): (1996) 38–45
22. Mann, G., Prins, J., Hermans, J.: Energetics of forced extraction of ligand: Simulation studies of Xe in mutant T4 lysozyme as a simple test system. Bioohys. J., in preparation (1998)
23. Mark, A. E., van Gunsteren, W. F.: Decomposition of the free energy of a system in terms of specific interactions. Implications for theoretical and experimental studies. J. Mol. Biol. **240** (1994) 167-176
24. McPhalen, C. A., James, M. N. G.: Structural comparison of two serine proteinase-protein inhibitor complexes: Eglin-C-Subtilisin Carlsberg and CI-2-subtilisin novo. Biochemistry **27** (1988) 6582-6598
25. Morton, A., Baase, W. A., Matthews, B. W.: Energetic origins of specificity of ligand binding in an interior nonpolar cavity of T4 lysozyme. Biochemistry **34** (1995) 8564-8575.
26. Moy, V. T., Florin, E. V., Gaub, H. E.: Intermolecular forces and energies between ligands and receptors. Science 266 (1994) 257-259
27. Nelson, M., Humphrey, W., Gursoy, A., Dalke, A., Kalé, L., Skeel, R., Schulten, K., Kufrin, R.: MDScope: A visual computing environment for structural biology. Comp. Phys. Comm. **91** (1995) 111-134
28. Page, M. L., Jencks, W. P.: Entropic contributions to rate accelerations in enzymic and intramolecular interactions and the chelate effect. Proc. Natl. Acad. Sci. USA **68** (1971) 1678-1683
29. Jencks, W. P.: On the attribution and additivity of binding energies. Proc. Natl. Acad. Sci. USA **78** (1981) 4046-4050
30. Quillin, M. L., Baase, W. A., Matthews, B. W.: Binding of small electron-dense ligands in large protein cavities. In International union of crystallography. XII congress and general assembly. American crystallographic association, Seattle, WA, USA. (1996) abstract C215.
31. Holtzer, A.: The 'cratic' correction and related fallacies. Biopolymers **34** (1995) 595-602
32. Sharp, K. A., Honig, B.: Electrostatic interactions in macromolecules: Theory and applications. Ann. Rev. Biophys. Biophys. Chem. **19** (1990) 301-332
33. Sitkoff, D., Sharp, K. A., Honig, B.: Accurate calculation of hydration free energies using macroscopic solvent models. J. Phys. Chem. **98** (1994) 1978-1988
34. Steinberg, I. Z., Scheraga, H. A.: Entropy changes accompanying association reactions of proteins. J. Biol. Chem. **238** (1963)172-181.
35. Tidor, B., Karplus, M.: The contribution of vibrational entropy to molecular association. J. Mol. Biol. **238** (1994) 405-414.
36. Vorobjev, Y., Almagro, J. C., Hermans, J.: Conformational free energy calculated by a new method from dynamics simulation and continuum dielectric

model discriminates between native and intentionally misfolded conformations of proteins. Proteins: Struct. Funct. Genet., submitted for publication (1998)
37. Zhang, L., Hermans, J.: Hydrophilicity of cavities in proteins. Proteins: Str. Funct. Genet. **24** (1996) 433-438

Estimating Relative Free Energies from a Single Simulation of the Initial State

Alan E. Mark[1], Heiko Schäfer[1], Haiyan Liu[2], and Wilfred van Gunsteren[1]

[1] Laboratorium für Physikalische Chemie, ETH Zentrum, CH-8092 Zürich, Switzerland

[2] Department of Biology, University of Science and Technology of China, Hefei, 230026, China.

Abstract. To estimate free energy differences from a single simulation of the initial state one may either, use a series expansion of the free energy around the initial state, make an assumption in regard to the functional form of the free energy or treat the mutation as a single step perturbation. Of these the perturbation approach holds the greatest promise. The perturbation approach is fast, easy to implement and does not depend on empirically derived parameters or assumptions. Given an appropriate reference state the perturbation approach can be used to rapidly estimate solvation or binding free energies of a wide range of related compounds for use in force field development or structure based drug design.

1 Introduction

All equilibrium properties of a system directly or indirectly depend on the difference in free energy between alternate states. As a result, there is great interest in efficient algorithms to estimate relative free energies from atomic simulations. This is especially true for applications such as the estimation of relative binding free energies in structure based drug design where the time and cost required to screen hundreds or even thousands of related compounds is a critical factor. The essential difficulty is that the free energy is a global property of a system. The total free energy of a system, F, expressed in terms of the canonical partition function, Z, is given by [1, 2]

$$F = -k_B T \ln Z \tag{1}$$
$$Z = \left[h^{3N} N!\right]^{-1} \int\int \exp\left[\frac{-\mathcal{H}(p,q)}{k_B T}\right] dpdq$$

where k_B is Boltzmann's constant, T is the absolute temperature and h is Planck's constant. $\mathcal{H}(p,q)$ is the Hamiltonian of the system expressed in terms of the positions, p, and the conjugate momenta, q of the N particles. The total free energy is dependent on an integral over all phase space. As it is not possible to completely sample all phase space, it is never possible to determine the absolute free energy of any real system. It is possible, nevertheless, to determine the relative free energy between closely related states of a single

system either from the work required to go from an initial to a final state via a reversible path or from the relative probability of finding the system in one state as opposed to another.

Normally, we are concerned with free energy differences between chemically distinct states. That is, states for which the Hamiltonians differ. In such cases the so-called coupling parameter approach is used. The Hamiltonian of the system is made dependent upon an arbitrary coupling parameter λ, $\mathcal{H}(p, q, \lambda)$, such that when $\lambda = \lambda_A$ the Hamiltonian of the system corresponds to state A, and when $\lambda = \lambda_B$ the Hamiltonian of the system corresponds to state B. If the Hamiltonian of the system is a function of λ the free energy of the system is also a function of λ and using basic statistical mechanics it may be easily shown that the difference in the configurational free energy between the two states is given by [1, 2],

$$\Delta F_{AB} = F_B - F_A = \int_{\lambda_A}^{\lambda_B} \frac{\partial F(\lambda)}{\partial \lambda} d\lambda = \int_{\lambda_A}^{\lambda_B} \left\langle \frac{\partial V}{\partial \lambda} \right\rangle d\lambda \tag{2}$$

where V is the potential energy of the system. This, the so-called Thermodynamic Integration formula, expresses the free energy as an integral of the work along a specific path. Alternatively the so-called Thermodynamic Perturbation formula may be used [1, 2],

$$\Delta F_{AB} = -k_B T \ln \left\langle e^{-(V_B - V_A)/k_B T} \right\rangle_A \tag{3}$$

in which the difference in free energy is expressed in terms of the probability of finding configurations representative (low energy) of state B in an ensemble generated using the Hamiltonian of state A. Formally, the perturbation approach encompasses any mutation of the system. The ensemble average in (3), however, only converges if there is a strong overlap between the low energy configurations of the two states, that is, the effect of the mutation on the configurations sampled is small. For this reason the perturbation formula is generally expressed as a sum over a series of small steps in λ [1, 2],

$$\Delta F_{AB} = \sum_{\lambda=\lambda_A}^{\lambda=\lambda_B-\delta\lambda} -k_B T \ln \left\langle e^{-[V(\lambda+\delta\lambda)-V(\lambda)]/k_B T} \right\rangle_\lambda . \tag{4}$$

For many applications, however, these standard approaches are inappropriate. Although potentially highly accurate, Thermodynamic Integration and Thermodynamic Perturbation are both computationally very intensive. This makes them too slow and expensive for the rapid screening of compounds such as is required for structure based drug design [3]. In addition, as the free energy is calculated along a specific path, determined by the dependence of the Hamiltonian on λ, separate calculations must be performed for each compound.

An alternative to Thermodynamic Integration or Thermodynamic Perturbation calculations in which simulations at a series of intermediate states are performed, is to estimate the difference in free energy between two states based only on a simulation of the initial state. In principle, the free energy difference between a reference state and any other state of a system can be determined if the fluctuations of the reference state are completely known. The free energy difference can in effect be extrapolated from the behaviour of the system in the reference state. For structure based drug design, estimating free energy differences relative to a single reference state has many advantages. Most importantly, the behaviour of the system in the initial state is independent of the mutation considered. Thus, the same simulation can be used to estimate the difference in free energy to multiple alternate states potentially making such methods highly efficient.

2 Free Energy Extrapolation

Three basic approaches can be used to estimate the difference in free energy between two states of a system based on an initial state ensemble. They are:

(i) a series expansion of the free energy around the initial state,
(ii) an assumption in regard to the functional form of the free energy or,
(iii) the mutation may be treated as a single step perturbation.

Series expansion: Smith and van Gunsteren [4] investigated the first approach expanding the free energy as a function of the coupling parameter λ into a Taylor series around a given reference state, $\lambda = 0$,

$$\Delta F(\lambda) = F(\lambda) - F(0) = F'|_{\lambda=0}\lambda + \frac{1}{2!}F''|_{\lambda=0}\lambda^2 + \frac{1}{3!}F'''|_{\lambda=0}\lambda^3 + \ldots \quad (5)$$

and computed the values of the higher-order derivatives, $F'', F''', \ldots$ as averages over the reference state ensemble. Using a 1 ns simulation of a dipolar diatomic molecule in water they predicted, with reasonable precision, the change in free energy associated with substantial charge rearrangement ($\pm 0.25e$) truncating the series beyond the 2nd or 3rd order [4]. Higher-order derivatives, however, converged slowly. The approach was later generalised by Hummer and Szabo [5] to include simulations of both the initial and final states. This has the advantage that larger mutations can be considered but requires separate simulations for each compound. Other variations on this general theme have been proposed. Gerber et al. [6] calculated the first derivative of the free energy with respect to λ for individual terms of the force field independently and used a linear combination of these derivatives to predict differences in the binding affinity of trimethoprin derivatives to dihydrofolate reductase. This first order approximation, however, proved insufficient. A similar approach has been investigated by Radmer and Kollman

[7]. A potential advantage of methods based on a series expansion of the free energy is that the convergence of the series is determined by the λ dependence of the potential energy function meaning that the efficiency of the approach could be enhanced by a judicious choice of coupling scheme.

Fixed functional dependence: An alternative to a series expansion is to assume a particular functional dependence of the free energy on λ. Jayaram and Beveridge [8], for example, derived an expression for the free energy of the system assuming the fluctuations of the potential energy obeyed a Gaussian distribution. This approach performed well when used to estimate the excess free energy of water but performed less well for hydration free energies of simple compounds. Similar expressions have been recently proposed by Amadei et al. [9]. A special case of this class of method is Linear Response theory. The basic premise in Linear Response theory is that the response of the environment to any given perturbation is linear, the classic example being the response of a system of constant dielectric to the introduction of a charge. The Linear Response assumption is equivalent to assuming that the fluctuations in the energy of interaction between a molecule and its surroundings are Gaussian distributed. If true, the difference in free energy is determined by the 1st and 2nd derivatives of the free energy with respect to the change. The approach is, therefore, the same as a Taylor expansion truncated after the second term assuming linear coupling between the initial and final states [10, 4, 11]. In the Linear Response limit the difference in free energy between two states may also be expressed as

$$\Delta F_{AB} = \frac{1}{2}\left(\langle V_B - V_A\rangle_A - \langle V_A - V_B\rangle_B\right) \tag{6}$$

where $< \ldots >_A$ and $< \ldots >_B$ refer to ensemble averages over the initial and final states. Equation (6) expresses Linear Response in its pure form. Aqvist and others have, however, demonstrated considerable success predicting relative binding free energies and relative hydration free energies using a semi-empirical approach in which the change in free energy is approximated as [11, 12],

$$\Delta F_{AB} \approx \frac{1}{2}\left(\langle V^{ele}\rangle_B - \langle V^{ele}\rangle_A\right) + \alpha\left(\langle V^{vdw}\rangle_B - \langle V^{vdw}\rangle_A\right) \tag{7}$$

where V^{ele} and V^{vdw} respectively refer to the electrostatic and van der Waals interaction energies between the solute and its environment, and α is an empirically derived scaling factor. In effect a Linear Response is assumed for the electrostatic contributions, but the van der Waals contributions are empirically parameterised based on a set of related compounds. Carlson and Jorgensen [13] took this approach a step further by introducing empirical scaling parameters for both the electrostatic and van der Waals interactions

and adding an additional term for the free energy of cavity formation. Although they refer to this as an Extended Linear Response method the approach is in effect an empirical model in which the interaction energies in the initial and final states enter as variables. This is not to say that the approach does not perform well for some systems [13, 14]. Separate simulations are, however, again required for each compound of interest. The basic difficulty is that while the response of a dielectric medium to a change in charge is essentially linear the response of a dense medium to the creation or deletion of an atom or atoms is not. The same problem is encountered when attempting to use a series expansion for mutations involving the creation or deletion of atoms [15].

Single step perturbation: If the series converges, equation (5) is formally equivalent to the direct application of the perturbation formula. This was noted by Liu et al. [15] who showed, for the rearrangement of charge on a diatomic molecule, that results comparable to those of Smith and van Gunsteren [4] could be obtained more simply using a perturbation approach. Further studies by Lui and others showed that given an appropriately chosen reference state the creation or deletion of atoms or even groups of atoms could be also treated as a single step perturbation [15, 16, 17].

The three approaches are closely related. They are in fact equivalent in the limit that the fluctuations in the energy of interaction between a molecule and its surrounding obey a Gaussian distribution. All perform well for cases involving small changes in partial charge [4, 5, 10, 15]. Where they differ is in the manner by which higher order terms are treated. In practice, it is always necessary to truncate a series expansion. This effectively sets all higher order derivatives of the free energy to zero. In methods that assume a particular functional dependence for the free energy higher order terms are again fixed. In contrast, all terms are implicitly included in the perturbation approach. Thus, the perturbation approach remains formally exact while the other approaches must be considered approximations. Higher order terms, however, converge slowly and using (3) the free energy may be determined by rarely sampled configurations.

3 Extrapolation Based on a Biased Ensemble

To estimate the difference in free energy between two states of a system using the perturbation approach the reference ensemble must contain representative low energy configurations for the alternate state of interest. Using Molecular Dynamics (MD) or Monte Carlo (MC) simulation techniques to generate the reference ensemble only low energy configurations of the reference state will be sampled. It is well known that low energy regions of the reference state do not correspond to low energy regions of the end state in cases where

atoms are created or deleted. For this reason, when using the Thermodynamic Perturbation formula, mutations are standardly broken into a series of small steps as in (4). The essential difficulty is that of sampling. This is illustrated diagrammatically in Fig. 1 for the mutation of *p*-methylphenol to *p*-methoxyphenol, one of the cases considered by Liu et al. [15]. The mutation fails as a single step perturbation because no configurations containing a cavity large enough to accommodate the methoxy group will be sampled in a simulation of *p*-methylphenol. The reference state ensemble need not, however, correspond to a physically meaningful state. By applying an appropriate biasing function, the range of configurations sampled may be spread to include configurations with and without appropriate cavities as is illustrated in Fig. 2.

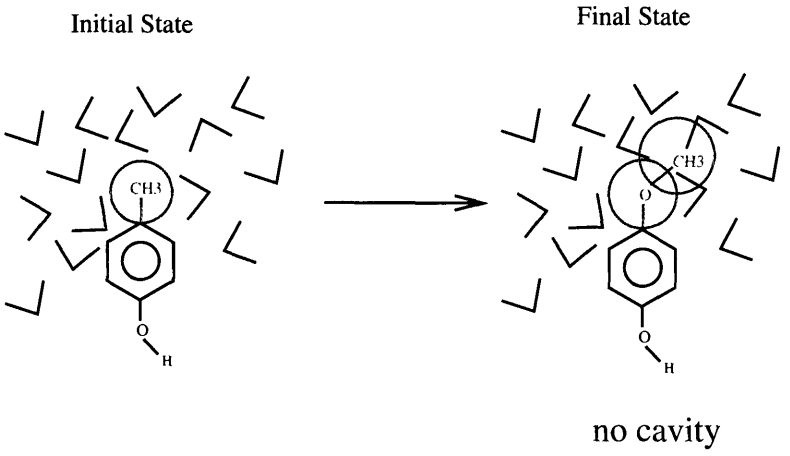


Fig. 1. A single step perturbation fails for the mutation of *p*-methylphenol to *p*-methoxyphenol because no configuration with a cavity large enough to accommodate the methoxy group is sampled in a simulation of *p*-methylphenol in water.

3.1 Choice of Biasing Function

The biasing function is applied to spread the range of configurations sampled such that the trajectory contains configurations appropriate to both the initial and final states. For the creation or deletion of atoms a softcore interaction function may be used. The standard Lennard-Jones (LJ) function used to model van der Waals interactions between atoms is strongly repulsive at short distances and contains a singularity at $r = 0$. This precludes two atoms from occupying the same position. A so-called softcore potential in contrast approaches a finite value at short distances. This removes the sin-

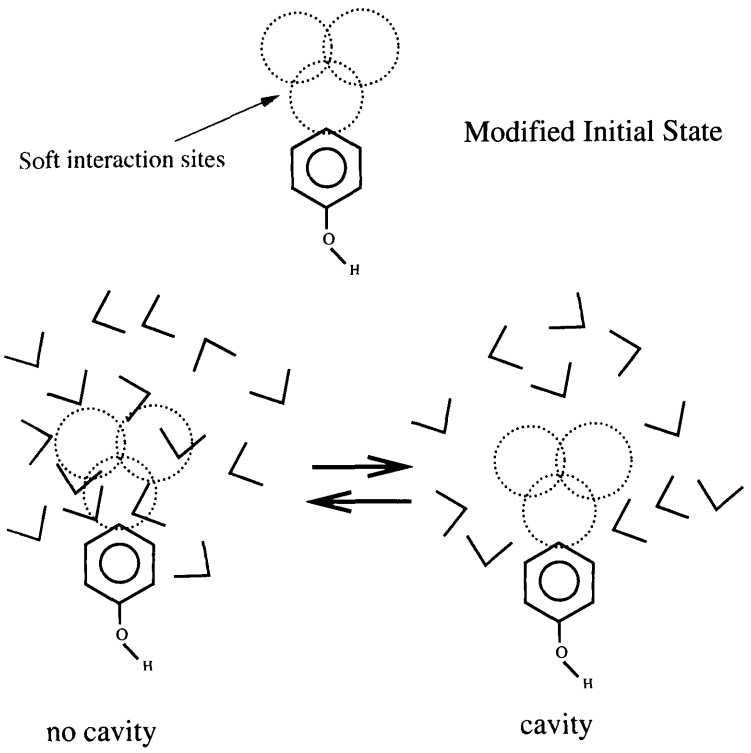


Fig. 2. By replacing the methyl group in *p*-methlyphenol with a series of softcore interaction sites it is possible to spread the range of configurations sampled to include configurations with and without a cavity in the *para* position.

gularity in the potential and allows atoms to pass through each other. The softcore function used by Liu et al. [15] was of the form

$$V_{i,j}(r) = \frac{1}{\alpha\sigma_{i,j}^6 + r^6}\left(\frac{4\epsilon_{i,j}\sigma_{i,j}^{12}}{\alpha\sigma_{i,j}^6 + r^6} - 4\epsilon_{i,j}\sigma_{i,j}^6\right) \tag{8}$$

where $\sigma_{i,j}$ and $\epsilon_{i,j}$ are the Lennard Jones interaction parameters for atoms i and j and α determines the height of the barrier at $r = 0$. Curves calculated using (8) for a series of α values are displayed in Fig. 3. As seen in Fig. 3 this form has the advantage that for $r > \sigma$ the normal LJ potential energy function is recovered. Softcore potentials are increasingly used in free energy calculations to avoid sampling problems associated with the singularity in the LJ interaction as atoms are created or destroyed [18]. As a biasing function this form is effective, but not necessarily optimal. Figure 4 shows radial distribution functions (rdf's) between the centre of such a soft interaction site and the oxygen atom of the surrounding water. The interaction between the

cavity and the water was described by the potential energy curves displayed in Fig. 3. These correspond to barrier heights of approximately 2, 3, 4 and 5 k_BT respectively. The rdf's show considerable structure. In particular there is a maximum at approximately 0.4 nm. In order to generate the broadest possible sampling such structure is undesirable and suggests that alternative biasing functions may perform better. From Fig. 4 it evident that the optimal barrier height for this system is between 3 and 4 k_BT as this results in the greatest spread of configurations. Using a barrier of 2 k_BT the cavity is filled by solvent. Using a barrier of 5 k_BT the cavity is devoid of solvent. Both extremes are undesirable.

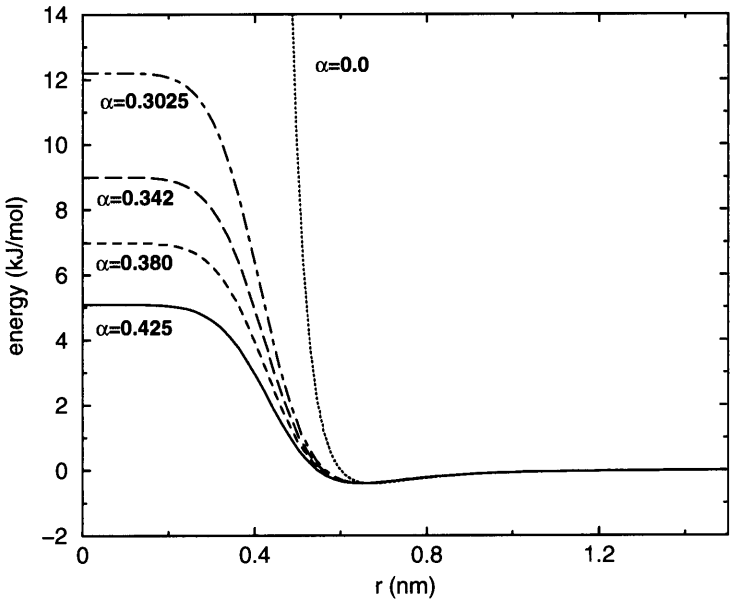


Fig. 3. Curves calculated using (8) for a series of increasing α values. The curves were calculated using $\sigma = 0.6$ nm and $\epsilon = 0.4$ kJ/mol. Note that for $\alpha = 0.0$ the normal 6-12 Lennard Jones potential energy function is recovered.

3.2 Implementation and Efficiency

To generate a trajectory of configurations using using Molecular Dynamics or Monte Carlo simulations techniques all interactions defined by the Hamiltonian must be considered. In addition, to integrate the equations of motion in Molecular Dynamics, or obtain an appropriate acceptance ratio using Monte Carlo, the step size must be small. The configurations sampled are thus highly

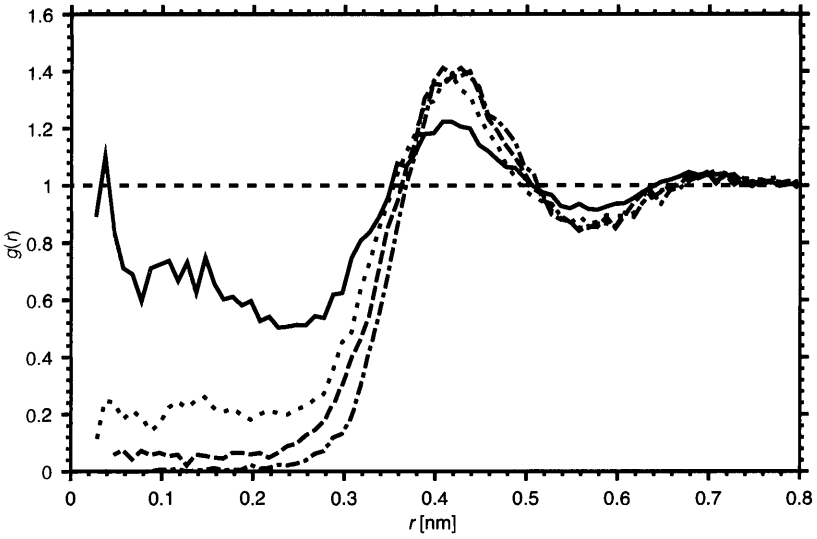


Fig. 4. Radial distribution functions between the centre of a test cavity and the oxygen atom of the surrounding water. The curves correspond to the different barrier heights for the softcore interaction illustrated in Fig. 3

correlated. To estimate the difference in free energy using (3), however, only those interactions that change between the initial and final states need be considered and, in addition, the configurations over which the average is taken should be uncorrelated. The average in (3) is over an unperturbed ensemble. It is thus independent of the mutation considered. This means that the reference ensemble need only be generated once and a stored trajectory may then used to estimate free energy differences for specific compounds. This approach is highly efficient. As illustrated in Fig. 5 only a small number of pair interaction need to be considered for most mutations. If a cutoff is used, all solvent molecules which lie outside this cutoff with respect to any site of interest may be removed. As the configurations used should be uncorrelated the trajectory may be sampled sparsely. For these reasons estimating changes in free energy using a pre-existing trajectory is 10^2 to 10^3 faster than standard free energy methods.

3.3 Creation and Placements of Atoms and Molecules

Atoms not explicitly included in the trajectory must be generated. The position at which an atom may be placed is in some sense arbitrary, the approach being analogous to the insertion of a test particle. Chemically meaningful end states may be generated by placing atoms based on internal coordinates. It is required, however, that an atom be sampled in the same relative location in every configuration. An isolated molecule can, for example, be inserted into

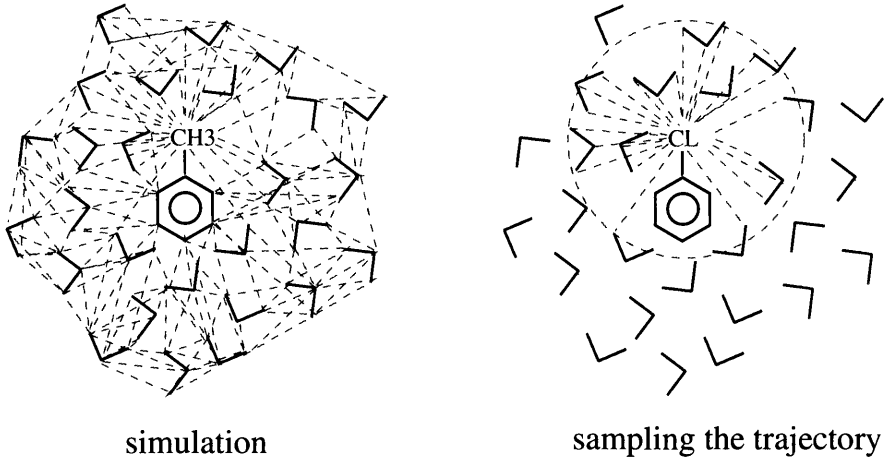


Fig. 5. To generate an ensemble using Molecular Dynamics or Monte-Carlo simulation techniques the interaction between all pairs of atoms within a given cutoff radius must be considered. In contrast, to estimate changes in free energy using a stored trajectory only those interactions which are perturbed need be determined making the approach highly efficient.

a given cavity in any number of orientations to improve statistics, but the orientation may not be systematically optimised for each configuration.

3.4 Internal Degrees of Freedom

When atoms not explicitly included in the trajectory are created based on internal coordinates the generated molecule or part thereof is effectively rigid. Contributions to the free energy from internal degrees of freedom are ignored. To estimate contributions from internal degrees of freedom the free energy must be averaged over a correctly weighted series of conformations which cover the range of potential motion. There will be a net contribution to the free energy only if the probability of sampling a given configuration is different in the initial and final states. Thus, only those degrees of freedom affected by the environment need be considered. To weight the configurations appropriately we must separate the difference in the potential energy between the reference and alternate state, ΔV ($\Delta V = V_B - V_A$) into a sum of inter- and intramolecular terms, $\Delta V = \Delta V_{inter} + \Delta V_{intra}$. Note, if the molecule is rigid the intramolecular term is a constant and can be ignored. If not, configurations must be assigned Boltzmann weights based exclusively on the intramolecular term as this has not been included in the Hamiltonian of the initial state. Configurations could, for example, be generated using a Monte Carlo procedure considering only ΔV_{intra}. In the extrapolation itself, double counting must be avoided by considering only the intermolecular

contribution, ΔV_{inter}. A simple alternative is to systematically generate configurations covering a range of motion and normalise the ensemble average by ΔV_{intra},

$$\Delta F = -k_B T \ln \frac{\left\langle e^{-(V_B - V_A)/k_B T} \right\rangle_A}{\left\langle e^{-\Delta V_{intra}/k_B T} \right\rangle_A}. \quad (9)$$

In essence this is equivalent to treating $-\Delta V_{intra}$ as an additional umbrella function.

3.5 Combining Biased Trajectories

No biasing function will be appropriate for all mutations. The greater the spread of configurations generated, the broader the range of mutations the extrapolation can encompass. The greater the spread of configurations generated, however, the more difficult it is to obtain sufficient sampling for any given mutation. In general, it will be most efficient to choose a biasing function appropriate for a specific class of mutation. For ligand design, trajectories generated with several different biasing functions will probably be required to test a range of derivatives of a lead compound. If appropriate configurations are sampled in each, the calculated difference in free energy will be independent of the trajectory used. If no appropriate (low energy) configurations are sampled in a given trajectory the difference in free energy using (3) is in general overestimated. In this case the lowest estimate for the free energy difference will be closest to the desired result, barring statistical fluctuations. Alternatively, trajectories generated using different biasing functions may be combined by weighting the configurations in each trajectory by a weight factor, W,

$$W = e^{-\Delta F_{biased}/k_B T} \quad (10)$$

where ΔF_{biased} is the difference in free energy between a common reference state and the particular biased state considered.

3.6 Range of Application

The critical factor for any method involving an approximation or an extrapolation is its range of application. Liu et al. [15] demonstrated that the approach performed well for mutations involving the creation or deletion of single atoms. The method has also been successfully applied to the prediction of the relative binding affinities of benzene, toluene and *o*-, *p*-, and *m*-xylene to a mutant of T4-lysozyme [16]. In both cases, however, the perturbation to the system was small. To investigate range over which the extrapolation may

be applicable we have applied the approach to the determination of hydration free energies for a series of polar and non-polar solutes of varying size. The reference state was a 1.6 ns trajectory of a non-polar softcore cavity in water. The radius and barrier height of the cavity were approximately 0.3 nm and 7.0 kJ/mol respectively and correspond to the short-dashed curve in Figs. 3 and 4. The non-polar solutes were methane, ethane, propane, butane, isobutane, pentane and cyclopentane, and the polar solutes were chloroform, water, methanol and ethanol. The free energy of hydration was estimated from the difference between perturbing to a dummy cavity which has no interactions with the surrounding solvent and perturbing to each of the compounds of interest. The extrapolation using a non-polar reference state performs well for ail the non-polar solutes with the exception of pentane which does not fit within the cavity. The results from the extrapolations are plotted in Fig. 6 against values obtained using Thermodynamic Integration. The solid line in Fig. 6 indicates the ideal one-to-one correlation, the dashed line corresponds to a line of best fit. There is essentially a one-to-one correspondance between the results obtained using a single step perturbation and Thermodynamic Integration for mutations ranging from the complete removal of the cavity to the insertion of cyclopentane. The largest deviation of 3 kJ/mol occurs for butane. No allowance for internal degrees of freedom has, however, been made. The extrapolation performed less well for the polar solutes. Although, there is a linear correlation between the extrapolated and Thermodynamic Integration results, the extrapolation systematically overestimates the free energy. This indicates that solvent configurations appropriate to the polar solutes are not being sampled using a non-polar reference state and an alternative to the non-polar reference state needs to be considered. Nevertheless, the results demonstrate that results comparable to that obtained using Thermodynamics Integration is possible for mutations involving the creation or removal from 1 to 5 atoms.

4 Conclusions

Estimating free energy differences from an initial state ensemble has the advantage that a large number of potential modifications of a given compound can be investigated in a single calculation. Because extrapolation is based on an unperturbed ensemble the nature of the perturbation does not have to be predefined. The change in free energy associated with a particular mutation can be obtained quickly and efficiently by reanalysis of stored trajectories. Extrapolation based on the application of the perturbation formula to an appropriately biased ensemble holds considerable promise for use as a rapid, non-empirical means of estimating relative binding affinities for a wide range of possible chemical modifications of a lead compound. As such the approach could be used to guide experimental planning in drug design a task not practical using normal free energy calculations.

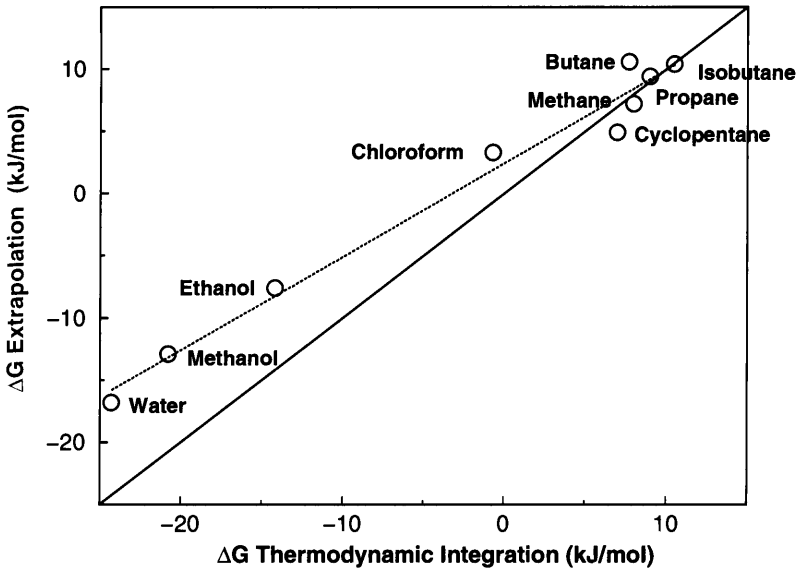


Fig. 6. Free energies of hydration calculated, for a series of polar and non-polar solute molecules by extrapolating using (3) from a 1.6 ns trajectory of a softcore cavity in water plotted against values obtained using Thermodynamic Integration. The solid line indicates an ideal one-to-one correspondence. The broken line is a line of best fit through the calculated points.

References

1. van Gunsteren, W. F., Beutler, T. C., Fraternali, F., King, P. M., Mark, A. E., Smith, P. E.: Computation of free energy in practice: Choice of approximations and accuracy limiting factors. In: "Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications", Vol. 2, van Gunsteren, W. F., Weiner, P. K., Wilkinson A. J., Eds., Escom Science Publishers, Leiden, The Netherlands, (1993), pp. 315–348
2. Mark, A.E.: Free energy perturbation calculations. Encyclopaedia of Computational Chemistry, Wiley, New York, (1998) (in press).
3. van Gunsteren, W. F., King, P. M., Mark, A. E.: Fundamentals of drug design from a biophysical viewpoint. Quart. Rev. Biophysics **27** (1994) 435–481
4. Smith, P. E., van Gunsteren, W. F.: Predictions of free energy differences from a single simulation of the initial state. J. Chem. Phys. **100** (1994) 577–585
5. Hummer, G., Szabo, A.: Calculation of free energy differences from computer simulations of initial and finial states. J. Chem. Phys. **105** (1996) 2004–2010
6. Gerber, P. R., Mark, A. E., van Gunsteren, W. F.: An approximate but efficient method to calculate free energy trends by computer simulation: Application to dihydrofolate reductase-inhibitor complexes. J. Comp. Aid. Mol. Desgn **7** (1993) 305–323
7. Radmer, R. J., Kollman, P. A.: Approximate free energy calculation methods and structure based ligand design. J. Comp. Aid. Mol. Desgn (in press)

8. Jayaram, B., Beveridge, D. L.: A simple method to estimate free energy from molecular simulation: Renormalization on the unit interval. J. Phys. Chem. **94** (1990) 7288-7293
9. Amadei, A., Apol, M. E. F., Di Nola, A., Berendsen, H. J. C.: The quasi-Gaussian entropy theory: Free energy calculations based on the potential energy distribution function. J. Chem. Phys. **104** (1996) 1560–1574
10. Levy, R. M., Belhadj, M., Kitchen, D. B.: Gaussian fluctuation formula for electrostatic free energy changes. J. Chem. Phys. **95** (1991) 3627-3633
11. Aqvist, J., Medina, C., Samuelsson, J. -E.: A new method for predicting binding affinity in computer-aided drug design. Prot. Eng. **7** (1994) 385-391
12. Hansson, T., Aqvist, J.: Estimation of binding free energies for HIV proteinase inhibitors by molecular dynamics simulations. Prot. Eng. **8** (1995) 1137-1144
13. Carlson, H. A., Jorgensen, W. L.: An extended linear response method for determining free energies of hydration. J. Phys. Chem. **99** (1995) 10667-10673
14. McDonald, N. A., Carlson, H. A., Jorgensen, W. L.: Free energies of solvation in chloroform and water from a linear response approach. J. Phys. Org. Chem. **10** (1997) 563–576
15. Liu, H., Mark, A. E., van Gunsteren, W. F.: Estimating the relative free energy of different molecular states with respect to a single reference state. J. Phys. Chem. **100** (1996) 9485–9494
16. Mark, A. E., Xu, Y., Liu, H., van Gunsteren, W. F.: Rapid non-empirical approaches for estimating relative binding free energies. Acta Biochim. Polonica **42** (1995) 525–536
17. Schäfer, H., van Gunsteren, W. F., Mark, A. E.: Estimating relative binding free energies from an initial state: Hydration free energies. J. Phys. Chem. (submitted)
18. Beutler, T. C., Mark, A. E., van Schaik, R. C., Gerber, P. R., van Gunsteren, W. F.: Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. Chem. Phys. Letters **222** (1994) 529–539

Exploration of Peptide Free Energy Surfaces

Krzysztof Kuczera[1]

Departments of Chemistry and Biochemistry, Cell and Molecular Biology, University of Kansas 2010 Malott Hall, Lawrence, KS 66045, kuczera@tedybr.chem.ukans.edu

Abstract. The Conformational Free Energy Thermodynamic Integration (CFTI) method, a new multidimensional approach for conformational free energy simulations, is presented. The method is applied to two problems of biochemical interest: exploration of the free energy surfaces of helical alanine (Ala) and α-methylalanine (Aib) homopeptides in vacuum and the cost of pre-organization of the opioid peptide Tyr-D-Pen-Gly-Phe-D-Pen (DPDPE) peptide for disulfide bond formation. In the CFTI approach a single molecular dynamics simulation with all ϕ and ψ dihedrals kept fixed yields the complete conformational free energy gradient for the studied peptides. For regular structures of model peptides $(Ala)_n$ and $(Aib)_n$ where n=6,8,10 and Aib is α-methylalanine in vacuum, free energy maps in the helical region of $\phi - \psi$ space are calculated, and used to roughly locate stable states. The locations of the free energy minima are further refined by the novel procedure of free energy optimization by steepest descent down the gradient, leading to structures in excellent agreement with experimental data. The stability of the minima with respect to deformations is studied by analysis of second derivatives of the free energy surface. Analysis of free energy components and molecular structures uncovers the molecular mechanism for the propensity of Aib peptides for the 3_{10}-helix structure in the interplay between the quality and quantity of hydrogen bonds. For the linear form of the DPDPE peptide in solution, free energy differences are calculated between four conformers: Cyc, representing the structure adopted by the linear peptide prior to disulfide bond formation, β_C and β_E, two slightly different β-turns previously identified as representative, stable structures of the peptide, and Ext, an extended structure. The simulations indicate that β_E is the most stable of the studied conformers of linear DPDPE in aqueous solution, with β_C, Cyc and Ext having free energies higher by 2.3, 6.3, and 28.2 kcal/mol, respectively. The free energy differences of 4.0 kcal/mol between β_C and Cyc, and 6.3 kcal/mol between β_E and Cyc, reflect the cost of pre-organizing the linear peptide into a conformation conducive for disulfide bond formation. Such a conformational change is a pre-requisite for the chemical reaction of S–S bond formation to proceed.

1 Introduction

Conformational free energy simulations are being widely used in modeling of complex molecular systems [1]. Recent examples of applications include study of torsions in n-butane [2] and peptide sidechains [3, 4], as well as aggregation of methane [5] and a helix bundle protein in water [6]. Calculating free energy differences between molecular states is valuable because they are observable thermodynamic quantities, related to equilibrium constants and

rate constants of processes. Most current simulation methods follow a similar procedure, in which constraints or restraints are used to force the system along a reaction coordinate, enabling the determination of the associated free energy profile, or potential of mean force [1, 7]. This leads to two common drawbacks. First, although behavior of the full system is simulated, the free energy profile is determined only in one direction - along the reaction path. Second, although significant barriers along the reaction path may be overcome, sampling of conformations in the orthogonal directions is limited and incomplete leading to problems with convergence of results.

This paper presents the theoretical background and some practical applications of a new conformational free energy simulation approach, aimed at correcting the above shortcomings. The new method, called Conformational Free energy Thermodynamic Integration (CFTI), is based on the observation that it is possible to calculate the conformational free energy gradient with respect to an arbitrary number of conformational coordinates from a single simulation with all coordinates in the set kept fixed [2, 8]. The availability of the conformational gradient makes possible novel techniques of multidimensional conformational free energy surface exploration, including locating free energy minima by free energy optimization and analysis of structural stability based on second derivatives of the free energy. Additionally, by performing simulations with all "soft" degrees of freedom of the system kept fixed, free energy averages converge very quickly, effectively overcoming the conformational sampling problem.

The subject of this work is the application of the CFTI technique to study of the conformational free energy surfaces of several peptides. The first part involves simulations of helical structures of model peptides – $(Ala)_n$ and $(Aib)_n$, where n=6,8,10 and Aib is α-methylalanine. These simple peptides serve as models for studying structure and dynamics of more general peptide and protein systems. Extensive studies of $(Ala)_n$ and $(Aib)_n$ in vacuum demonstrate the feasibility of multidimensional CFTI calculations, including locating free energy minima and determination of intrinsic and relative stability of different regular helical structures - the 3_{10}-helix , α-helix and π-helix . The simulation results reproduce the experimentally observed stabilization of the 3_{10}-helix relative to the α-helix in the α-methylated peptides, and provide a microscopic model for this interesting effect [8].

The second application of the CFTI approach described here involves calculations of the free energy differences between conformers of the linear form of the opioid pentapeptide DPDPE in aqueous solution [9, 10]. DPDPE (Tyr-D-Pen-Gly-Phe-D-Pen, where D-Pen is the D isomer of β,β-dimethylcysteine) and other opioids are an interesting class of biologically active peptides which exhibit a strong correlation between conformation and affinity and selectivity for different receptors. The cyclic form of DPDPE contains a disulfide bond constraint, and is a highly specific δ opioid [11]. Our simulations provide information on the cost of pre-organizing the linear peptide from its stable solution structure to a cyclic-like precursor for disulfide bond formation. Such

a conformational change is a pre-requisite for the chemical reaction of S–S bond formation to proceed. Use of the CFTI approach with all 14 conformationally important dihedrals constrained leads to quickly converging averages, allowing efficient evaluation of the free energy differences between different DPDPE structures [9, 10].

2 The CFTI Method

The free energy profile or potential of mean force along a conformational coordinate ξ may be defined as

$$A_c(\xi') \;=\; -\beta^{-1}\,\ln\,Z(\xi') \tag{1}$$

where $\beta = 1/\mathrm{kT}$, k is the Boltzmann constant, T the absolute temperature, and

$$Z(\xi') \;=\; \int dq\; e^{-\beta U(q)}\;\delta(\xi(q)\;-\;\xi') \tag{2}$$

is the contribution to the classical configurational integral Z of the "slice" of configuration space with fixed $\xi = \xi'$, and $\xi(q)$ expresses the generalized coordinate ξ in terms of Cartesian coordinates. This leads to the basic formula of the thermodynamic integration (TI) conformational free energy method [2, 12]:

$$\frac{\partial A_c(\xi')}{\partial \xi'} \;=\; =\; \langle\; \frac{\partial U}{\partial \xi'} \;-\; \beta^{-1}\;\frac{\partial \ln|J|}{\partial \xi'}\;\rangle_{\xi'} \tag{3}$$

where $\langle ... \rangle_{\xi'}$ denotes an average over conformations with fixed $\xi = \xi'$ and J is the Jacobian of the coordinate transformation between the Cartesian coordinates q and the generalized coordinate set containing ξ. The term involving J has been found to be negligibly small in previous practical applications [2, 8, 12].

The CFTI method extends the standard TI to an arbitrary number of dimensions [2, 8]. Analogously to Eqs. (1) and (2) the free energy surface is defined as

$$A_c(\xi_1', \xi_2', ..., \xi_m') \;=\; -\beta^{-1}\;\ln Z(\xi_1', \xi_2', ..., \xi_m') \tag{4}$$

where $Z(\xi_1', \xi_2', ..., \xi_m')$ is the configurational integral corresponding to fixed values $\xi_k = \xi_k'$ for all coordinates in the set ξ_k, k=1,...,m. The free energy gradient is then [2] :

$$\frac{\partial A_c(\xi_1', \xi_2', ..., \xi_m')}{\partial \xi_k'} \;=\; \langle\; \frac{\partial U}{\partial \xi_k'} \;-\; \beta^{-1}\;\frac{\partial \ln|J|}{\partial \xi_k'}\;\rangle_{\xi'} \tag{5}$$

where $\langle ... \rangle_{\xi_i'}$ denotes an average over conformations with fixed values $\xi_i = \xi_i'$ for i = 1,...,m. In the CFTI approach, the gradient of the conformational

free energy with respect to an arbitrary set of conformational coordinates is obtained from a simulation with all coordinates of the set kept fixed. Formulae for multidimensional conformational entropy gradients and for second derivatives of the free energy have been developed [8, 9].

The CFTI method is highly efficient, has improved convergence properties and enables new ways of exploring energy landscapes of flexible molecules. The efficiency is due to the fact that calculation of the free energy gradient with respect to an arbitrary number of coordinates may be performed at essentially the same cost as a standard one-dimensional TI simulation under the same conditions [2]. This is because the most expensive terms to evaluate, $\partial U/\partial \xi_k$, may be expressed in terms of simple algebraic transformations of the Cartesian gradient $\partial U/\partial q_j$, which is known at each step of a simulation [2, 8]. A single simulation yields derivatives of free energy with respect to all conformational degrees of interest, yielding a complete local characterization of conformational space, not just the derivative along a one-dimensional reaction path [2, 8]. This enables the determination of stability of structures with respect to perturbations, location of minima on the free energy surface, and finding minimum free energy paths connecting different states. The accelerated convergence may be achieved by selecting all "soft" degrees of freedom as the fixed coordinates. In the case of peptides these would be the backbone ϕ, ψ, and some of the sidechain dihedrals [8, 9, 10]. The sampling of the restricted conformational space of remaining hard degrees of freedom and solvent is very fast - simulations of 20-50 ps were sufficient to obtain precise gradient values in the studied cases. Simulations of similar length are sometimes used in the standard approach to free energy profiles, where only the reaction coordinate is constrained. However in these methods, because of the size of the available conformational space, the convergence of thermodynamic averages is often assumed rather than actually achieved.

Additionally, as in all TI-based approaches, the free energy differences are linear functions of the potential. Thus non-rigorous decompositions may be made into contributions from the different potential energy terms, parts of system and individual coordinates, providing valuable insight into the molecular mechanisms of studied processes [8, 9, 10].

3 Model Peptides: $(Ala)_n$ and $(Aib)_n$ in Vacuum

The $(Ala)_n$ and $(Aib)_n$ peptides, n=10,8,6 were simulated in vacuum, with all 2(n-1) ϕ and ψ dihedrals fixed at constant values:

$$\phi_i = \phi, \quad i = 2, ..., n; \qquad \psi_i = \psi, \ i = 1, ..., n-1$$

where i is the residue number. This corresponds to regular structures with identical conformations of each residue, identified by the pair of values (ϕ,ψ). At each point the system coordinates were generated from model geometric parameters, energy minimized and equilibrated for 20 ps. A 40 ps trajectory

at 300 K was then generated. At all stages all the ϕ and ψ dihedrals were kept fixed at the required values. The components with respect to the individual ϕ and ψ dihedrals were added up to yield a reduced two-dimensional gradient $(\partial A/\partial\phi, \partial A/\partial\psi)$ [8].

Fig. 1. Free energy gradient map of $(Ala)_{10}$ in vacuum

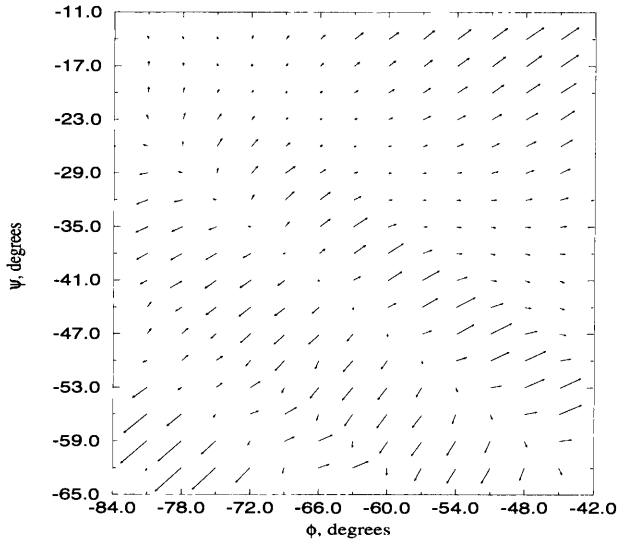
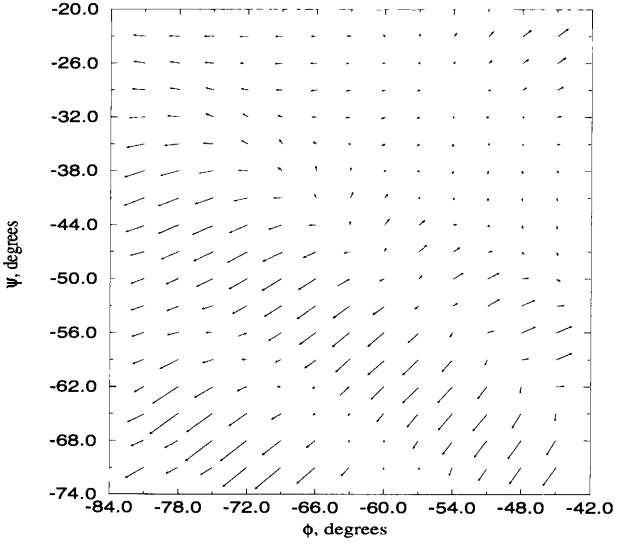


Fig. 2. Free energy gradient map of $(Aib)_{10}$ in vacuum



The free energy gradient maps for $(Ala)_{10}$ and $(Aib)_{10}$ are shown in Figs. 1 and 2. It is clear that $(Ala)_{10}$ has two minima in the studied region, corresponding to and α-helix and a π-helix . Aib_{10} has three minima - a 3_{10}-helix , an α-helix and a π-helix . The stable states occupy diagonal valleys, as is expected for helices where anticorrelated changes of ϕ and ψ on successive residues leave the hydrogen bonding pattern relatively unchanged [13]. A more detailed discussion, including free energy surfaces obtained by integrating the gradient maps is presented in [8].

The locations of some of the free energy minima were refined by free energy optimization, performed as a sequence of steps of down the free energy gradient. For $(Ala)_{10}$ the α-helix minimum at (-66^o,-41^o) is in good agreement with crystallographic data. In the case of $(Aib)_{10}$ peptides, the α-helical minimum was at (-57^o,-52^o), while the 3_{10}-helix was at (-53^o,-30^o), again in accord with crystallographic data. The π-helix positions were not refined, they lie in the (-75^o,-56^o) and (-69^o,-65^o) areas for $(Ala)_{10}$ and $(Aib)_{10}$, respectively.

From the free energy gradient maps, the free energy second derivative matrix $H_{ij} = \partial^2 A/\partial\xi_i\partial\xi_j$ was evaluated numerically on the two-dimensional (2D) (ϕ,ψ) surface. It is simplest to describe this matrix by its two eigenvalues. In all cases the eigenvalues were positive at the studied minima, with the higher eigenvalue corresponding to correlated changes in ϕ and ψ, and the lower one to anti-correlated changes. The helices are thus easier to deform along the "helical troughs", with motions corresponding to cylindrical breathing, which leaves the helix rise and turn unchanged. The "hard modes" of deformation involve changes of both helical rise and turn, and correspond approximately to reaction coordinates for concerted transitions between different helix types [8].

Free energy differences between α-helix and 3_{10}-helix states for the $(Ala)_n$ and $(Aib)_n$ systems were also calculated (Table 1). For $(Ala)_n$ the α-helix was the more stable helical state for n=10,8,6, with the free energy difference decreasing with peptide length. For $(Aib)_n$ the α-helix was more stable for n=10, and the 3_{10}-helix was more stable for n=6; for n=8 the two helical forms had comparable stability. The main conclusion is that the simulation results agree with the experimentally observed trends - that α-methylation stabilizes the 3_{10}-helix relative to the α-helix , and that this effect becomes more pronounced in shorter peptides. Additionally, a microscopic mechanism for this phenomenon may be determined by comparing structural data in Table 2 with the free energies in Table 1. There is interplay of two effects - internal strain and nonbonded interactions tend to favor the α-helix , while electrostatic interactions tend to favor the 3_{10}-helix . Destabilization by internal strain and nonbonded repulsion is systematically greater in the more crowded $(Aib)_n$ systems. The behavior of the electrostatic contribution is more complicated, reflecting changes in both quantity and quality of hydrogen bonds. The 3_{10}-helix contains one more hydrogen bond than the α-helix . However, Table 2 shows while the α-helix has better hydrogen bond-

Table 1. Energetics of the concerted α-helix $\rightarrow$ 3_{10}-helix transition in $(Aib)_n$. For $(Ala)_n$ profiles connect α-helix minima to (ϕ, ψ) = (-60°,-29°). For $(Aib)_n$ profiles connect α-helix minima to 3_{10}-helix minima. Statistical errors of the total free energy and the potential energy components were below 0.1 kcal/mol in all cases. Units: kcal/mol.

System	Potential energy components[a]				Energy-entropy components	
ΔA	Total	Int	Elec	vdW	ΔU	-TΔS
		$\alpha \rightarrow 3_{10}$				
$(Ala)_{10}$	14.2	2.8	3.0	8.4	17.7 ± 0.4	-3.5 ± 0.4
$(Ala)_8$	8.3	1.9	-2.0	8.4	11.5 ± 0.3	-3.2 ± 0.3
$(Ala)_6$	3.7	0.8	-4.7	7.6	3.9 ± 0.2	-0.2 ± 0.2
$(Aib)_{10}$	3.9	3.9	-13.5	13.5	6.0 ± 0.7	-2.1 ± 0.7
$(Aib)_8$	0.3	2.9	-13.1	10.5	2.0 ± 0.5	-1.7 ± 0.5
$(Aib)_8$	-2.8	1.6	-12.0	7.6	-2.8 ± 0.2	0.0 ± 0.2

[a] Int: internal strain component; Elec: electrostatic interaction component; vdW: van der Waals interaction component.

ing geometry than the 3_{10}-helix in $(Ala)_{10}$, the reverse is true for $(Aib)_{10}$. The synergistic effect of greater number of hydrogen bonds and improved interactions within each bond strongly stabilizes the $(Aib)_n$ 3_{10}-helix , making it the favored structure for short peptides.

Qualitatively similar results were obtained in preliminary calculations of several $(Ala)_{10}$ and $(Aib)_{10}$ helical structures in water.

4 DPDPE Pre-Organization for Disulfide Bond Formation

For the linear form of DPDPE (Tyr-D-Pen-Gly-Phe-D-Pen, where D-Pen is the D isomer of β,β-dimethylcysteine) free energy differences were calculated between four structures, denoted as Cyc, Ext, β_C and β_E [9, 10]. The Cyc or cyclic-like conformer corresponds to an experimental structure of the cyclic form with the disulfide bond removed, Ext - to an extended structure. The two type IV β turns, β_C and β_E, have been identified as the stable, representative structure of linear DPDPE in solution based on previous unconstrained simulations [14]. Three conformational free energy simulations were performed: Cyc $\rightarrow$ β_C, Ext $\rightarrow$ β_E, and β_C $\rightarrow$ β_E. Each simulation involved generation of a series of intermediate states lying along a straight line connecting the initial and final state in dihedral angle space (29 for Cyc $\rightarrow$ β_C, 13 for Ext $\rightarrow$ β_E, and 15 for β_C $\rightarrow$ β_E). At each state 20 ps equilibration

Table 2. Geometric description of MD average helical structures

System	Helix structure			H-bond geometry			
	Rise	Turn	Radius	$R_{O...H}$	$R_{O...N}$	⟨(N-H-O)	⟨(H-O-C)
	Å	^{o}	Å	Å	Å	^{o}	^{o}
$(Ala)_{10}$							
α-helix	1.50	99.4	2.24	2.08	3.03	161.1	148.0
3_{10}-helix	1.93	109.2	1.84	2.25	3.17	156.8	121.8
$(Aib)_{10}$							
α-helix	1.58	96.9	2.34	2.26	3.24	170.8	155.9
3_{10}-helix	2.02	113.9	1.82	2.16	3.13	161.3	131.2

and 40 ps trajectory generation was performed for the peptide in a truncated cubic box with 875 water molecules with periodic boundary conditions at 300 K. The overall simulation times were thus 1.74, 0.78 and 0.90 ns for the three free energy simulations. In this initial work we focused on proving the feasibility of the approach, and did not try to minimize the simulation time by choosing some minimal number of points and their spacing. In the thermodynamic integration approach it is possible, in principle, to have the simulation windows quite widely separated from each other. All 14 dihedrals necessary to determine the conformation of DPDPE were kept fixed (these included all backbone ϕ and ψ dihedrals and most sidechain dihedrals, only dihedral angles corresponding to rotations of the peptide bonds, bonds involving hydrogen atoms and termini were left unconstrained). A more detailed description of the methodology may be found in [9, 10].

The calculated overall free energy changes ΔA were -4.0 ± 0.8 kcal/mol for the Cyc → β_C "unfolding" transition, -28.2 ± 0.9 kcal/mol for the Ext → β_E transition and -2.3 ± 0.8 kcal/mol for the β_C → β_E transition (Table 3). The most stable of the examined conformers is thus β_E, with β_C, Cyc and Ext having free energies 2.3, 6.3 and 28.2 kcal/mol higher, respectively. The predicted conformer population ratios are thus 42,000:860:1 for β_E:β_C:Cyc; the population of the extended conformer should be negligible.

The two β-turn structures, β_C and β_E are the most stable among those considered. This is in accord with the unconstrained nanosecond simulations of linear DPDPE, which converged to these conformers [14]. Because the cyclic form is relatively rigid, it is assumed that the conformation it adopts in solution is the biologically active one, responsible for its high affinity and specificity towards the δ opioid receptor. The relatively low population of the cyclic-like structure for the linear peptide thus agrees qualitatively with the

Table 3. Conformational free energy simulation of linear DPDPE. Changes in free energy and its components. Units: kcal/mol

Component	Cyc → β_C	Ext → β_E	β_C → β_E
Free energy ΔA	-4.0 ± 0.7	-28.2 ± 0.9	-2.3 ± 0.8
Entropy -TΔS	-66.1 ± 31.2	10.4 ± 36.4	39.2 ± 33.6
Potential energy ΔU	62.1 ± 31.2	-38.6 ± 36.4	-41.5 ± 33.7
	Free energy components		
Internal strain of solute	3.5 ± 0.1	-3.3 ± 0.1	-1.3 ± 0.1
Solute-solute interaction	-46.8 ± 0.1	-8.9 ± 0.1	-0.6 ± 0.1
Solute-solvent interaction	39.3 ± 0.9	-16.0 ± 1.2	-0.4 ± 1.1

observed lower potency and different receptor specificity of this form of the DPDPE peptide as compared to the cyclic form.

The Cyc conformer represents the structure adopted by the linear peptide prior to disulfide bond formation, while the two β-turns are representative stable structures of linear DPDPE. The free energy differences of 4.0 kcal/mol between β_C and Cyc, and 6.3 kcal/mol between β_E and Cyc, reflect the cost of pre-organizing the linear peptide into a conformation conducive for disulfide bond formation. Such a conformational change is a pre-requisite for the chemical reaction of S–S bond formation to proceed.

In the Cyc → β_C simulation the contributions are: 3.5 kcal/mol from internal deformations, -46.8 kcal/mol from solute-solute and 39.3 kcal/mol from solute-solvent interactions. Thus our results indicate that the preference of the linear peptide for the β-turn is due to significantly more favorable solute-solute interactions in that conformer. The sum of the solute-solute and internal strain terms, -43.3 kcal/mol, represents the conformational strain energy released after the peptide is allowed to relax from the cyclic-like to the β_C structure. This provides a quantitative measure for the intuitive concept of release of strain energy after removal of the S–S conformational constraint. Interactions with the solvent water provide a compensating effect, preferentially stabilizing the cyclic-like structure. Strong solute-solvent interactions have been seen previously in standard MD simulations of cyclic DPDPE [14], which exhibits parallel orientation of carbonyl groups, aggregation of hydrophobic groups and an exceptionally high dipole moment. The two large effects of internal strain and solvation have have opposite signs and mainly cancel, leading to a moderate conformational free energy difference between the cyclic-like and β_C structures. This appears to be a general property of solution thermodynamics of biological systems, where large contributions hidden in overall free energy change are revealed by theoretical analysis [15].

The energy-entropy decomposition results are presented in Table 3. Due to significantly higher statistical errors, these results are much less reliable than the overall ΔA. It appears that the CFTI approach, while leading to quickly converging free energy values, does not provide improved energy-entropy results compared to other methods [15, 16]. In the Ext $\rightarrow \beta_E$ and $\beta_C \rightarrow \beta_E$ simulations the energetic and entropic contributions are lower in magnitude or comparable to their respective errors, and no reliable conclusions can be drawn from these results. In the Cyc $\rightarrow \beta_C$ simulation entropic and energetic terms were -TΔS = -66 $\pm$ 31 kcal/mol and ΔU = 62 $\pm$ 31 kcal/mol, respectively, suggesting that the β_C conformer is entropically favored over Cyc, i.e. that the pre-organization for disulfide bond formation carries an entropic penalty.

The free energy differences obtained from our constrained simulations refer to strictly specified states, defined by single points in the 14-dimensional dihedral space. Standard concepts of a molecular conformation include some region, or volume in that space, explored by thermal fluctuations around a transient equilibrium structure. To obtain the free energy differences between conformers of the unconstrained peptide, a correction for the thermodynamic state is needed. The volume of explored conformational space may be estimated from the covariance matrix of the coordinates of interest, $C_{ij} = \langle(\xi_i - \langle\xi_i\rangle)(\xi_j - \langle\xi_j\rangle)\rangle$ [13, 10]. For each of the four selected conformers, three 20 *ps* simulations were performed with the dihedrals, which were kept fixed in the free energy simulation, constrained to their initial value with a harmonic restraint potential $\frac{1}{2}k(\phi - \phi_0)^2$, with decreasing force constants k = 3, 2, and 1 $kcal/(mol \cdot rad^2)$. The determinants $|C_I|$ of the covariance matrices were calculated (as products of eigenvalues) and extrapolated to k=0 for each conformer I. The correction to the free energy difference between conformers I and J was then calculated as $(-1/2)kT\ln(|C_J|/|C_I|)$ at $k = 0$ [9, 10]. The corrections obtained were -0.2 kcal/mol for Cyc $\rightarrow \beta_C$, 0.8 kcal/mol for Ext $\rightarrow \beta_E$, and 0.2 kcal/mol for $\beta_C \rightarrow \beta_E$, respectively. These correction are comparable to the errors of the calculated free energy differences and do not affect the conclusions [9, 10]. Thus it appears that the multidimensional free energy gradient obtained by fixing all flexible dihedrals in a pentapeptide may be integrated to obtain free energy differences between conformers without introducing significant bias.

Using the CFTI protocol, we have calculated directly both the derivative of the free energy with respect to the reaction path $\partial A/\partial\lambda$ and the 14 individual derivatives $\partial A/\partial\xi_k$, k = 1,...,14 with respect to all fixed coordinates along the path:

$$\frac{\partial A}{\partial \lambda} = \langle \frac{\partial U}{\partial \lambda} \rangle = \langle \sum_{k=1}^{14} \frac{\partial U}{\partial \xi_k} \frac{\partial \xi_k}{\partial \lambda} \rangle \quad ; \quad \frac{\partial A}{\partial \xi_k} = \langle \frac{\partial U}{\partial \xi_k} \rangle \tag{6}$$

Using this information, the overall free energy change may thus be decomposed into contributions from individual dihedrals. As with all free energy

decomposition schemes, this is not thermodynamically rigorous, but may lead to useful insights into peptide structure and function. Details of the decomposition results are discussed elsewhere [9, 10].

5 Conclusions

Initial results of the CFTI method, a new conformational free energy simulation approach, are presented. The main idea of the method is the generalization of standard thermodynamic integration from one to many dimensions. By performing a single simulation with a set of conformational coordinates held fixed, free energy derivatives with respect to all coordinates in the set may be obtained. The availability of the conformational free energy gradient opens the door to new ways of exploring free energy surfaces of flexible molecules.

The feasibility of the new method is first demonstrated for the model $(\text{Ala})_n$ and $(\text{Aib})_n$ peptides in vacuum. Free energy gradient maps are calculated, providing a powerful new way of visualizing the variation of free energy with conformation. A completely new procedure, free energy optimization by steepest descent, is performed for these systems to locate free energy minima. Second derivatives of the free energy are obtained to describe the stability of free energy minima with respect to small deformations. The experimentally observed trend of enhanced 3_{10}-helix stability relative to the α-helix in short Aib-containing peptides is reproduced. Structural analysis and free energy decompositions provide insight into the microscopic mechanism of this effect, revealing an underlying interplay between the quality and quantity of hydrogen bonds.

The second application of the CFTI protocol is the evaluation of the free energy differences between four states of the linear form of the opioid peptide DPDPE in solution. Our primary result is the determination of the free energy differences between the representative stable structures β_C and β_E and the cyclic-like conformer Cyc of linear DPDPE in aqueous solution. These free energy differences, 4.0 kcal/mol between β_C and Cyc, and 6.3 kcal/mol between β_E and Cyc, reflect the cost of pre-organizing the linear peptide into a conformation conducive for disulfide bond formation. Such a conformational change is a pre-requisite for the chemical reaction of S–S bond formation to proceed. The predicted low population of the cyclic-like structure, which is presumably the biologically active conformer, agrees qualitatively with observed lower potency and different receptor specificity of the linear form relative to the cyclic peptide.

The DPDPE simulations are the first example of use of CFTI to accelerate convergence of thermodynamic averages. This approach has two important advantages. First, because all "soft" degrees of freedom of the solute are fixed, the simulations do not suffer from conformational sampling problems common in other free energy simulation protocols, and all averages converge

very quickly. Second, the method yields not only the derivative of the free energy along the chosen reaction path between the initial and final state, but also the free energy gradients with respect to all the fixed coordinates at each path point. This provides more information about the free energy surface of the studied system at no increase in computational cost over standard thermodynamic integration approaches. The CFTI thermodynamic averages converged quite quickly, over tens of picoseconds. This is contrast to standard one-dimensional reaction path simulations – e.g. using the peptide end-to-end distance as the reaction coordinate, which generally require much longer sampling at each window. By averaging over the hard degrees of freedom of the solute and over the solvent distribution around the constrained solute, well defined thermodynamic states are generated at the ends and at the intermediate steps of the path, and the internal strain energy term is included in the free energy evaluation. Our results indicate that the CFTI method is a powerful, useful tool for simulating flexible molecules in solution. In our future work we plan to use the multidimensional free energy gradient to perform free energy optimization of stable states and reaction paths for such systems.

Acknowledgments. This work was supported by the Petroleum Research Fund of the American Chemical Society (grant #29566-G4), by a K*STAR EPSCoR FIRST Award, and by the University of Kansas General Research Fund. The use of the Origin 2000 computer at the Kansas Center for Advanced Scientific Computing is gratefully acknowledged.

References

1. T. P. Straatsma and J. A. McCammon. Computational alchemy. *Annu. Rev. Phys. Chem.*, 43:407–435, 1992.
2. K. Kuczera. One- and multidimensional conformational free energy simulations. *J. Comp. Chem.*, 17:1726–1749, 1996.
3. T.P. Straatsma and J.A. McCammon. Multiconfiguration thermodynamic integration. *J. Chem. Phys.*, 95:1175–1188, 1991.
4. P. E. Smith and B. M. Pettitt. Amino acid side–chain populations in aqueous and saline solution: Bis(penicillamine) enkephalin. *Biopolymers*, 32:1623–1629, 1992.
5. D. E. Smith and A. D. J. Haymet. Free energy, entropy and internal energy of hydrophobic interactions: computer simulations. *J. Chem. Phys.*, 98:6445–6454, 1993.
6. E. M. Boczko and C. L. Brooks III. First principles calculation of the folding free energy of a three helix bundle protein. *Science*, 269:393–396, 1995.
7. M. Mezei and D. L. Beveridge. Free energy simulations. *Ann. N. Y. Acad. Sci.*, 482:1–23, 1986.
8. Y. Wang and K. Kuczera. Exploration of conformational free energy surface of helical Ala_n and Aib_n peptides. *J. Phys. Chem. B*, 101:5205–5213, 1997.
9. Y. Wang. *Computational Approach to the Influence of the Disulfide Bond on Peptide Properties.* PhD thesis, University of Kansas, 1997.

10. Y. Wang and K. Kuczera. Conformational free energy surface of the linear DPDPE peptide: Cost of pre-organization for disulfide bond formation. *J. Am. Chem. Soc.*, submitted, 1997.
11. H. I. Mosberg, R. Hurst, V. J. Hruby, K. Gee, H. I. Yamamura, J. J. Galligan, and T. F. Burks. Bis-penicillamine enkephalins possess highly improved specificity toward δ opioid receptors. *Proc. Natl. Acad. Sci. U.S.A.*, 80:5871–5874, 1983.
12. E. A. Carter, G. Ciccotti, J. T. Hynes, and R. Kapral. Constrained reaction coordinate dynamics for the simulation of rare events. *Chem. Phys. Lett.*, 156:472–477, 1989.
13. C. L. Brooks III, M. Karplus, and B. M. Petitt. *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics.* John Wiley and Sons, New York, 1988.
14. Y. Wang and K. Kuczera. Molecular dynamics simulations of cyclic and linear DPDPE: Influence of the disulfide bond on peptide flexibility. *J. Phys. Chem.*, 100:2555–2563, 1996.
15. J. Gao, K. Kuczera, B. Tidor, and M. Karplus. Hidden thermodynamics of mutant proteins: A molecular dynamics analysis. *Science*, 244:1069–1072, 1989.
16. K. Kuczera, J. Gao, B. Tidor, and M. Karplus. Free energy of sickling: A simulation analysis. *Proc. Natl. Acad. Sci. USA*, 87:8481–8485, 1990.

Prediction of pK_as of Titratable Residues in Proteins Using a Poisson-Boltzmann Model of the Solute-Solvent System

Jan Antosiewicz[1], Elżbieta Błachut-Okrasińska[1], Tomasz Grycuk[1], James M. Briggs[2], Stanisław T. Włodek[3], Bogdan Lesyng[1,4], and J. Andrew McCammon[2,5]

[1] Department of Biophysics, University of Warsaw, 02-089 Warsaw, Poland
[2] Department of Pharmacology, University of California at San Diego, La Jolla, CA 92093-0365, USA
[3] Texas Center for Advanced Molecular Computation, University of Houston, Houston, TX 77204-5502, USA
[4] Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw, 02-106 Warsaw, Poland
[5] Department of Chemistry and Biochemistry, University of California at San Diego, La Jolla, CA 92093-0365, USA

Abstract. This article provides an overview of an algorithm used for the prediction of ionization constants of titratable residues in proteins. The algorithm is based on an assumption that the difference in protonation behavior of a given group in an isolated state in solution, for which the ionization constant is assumed to be known, and the protonation behavior in the protein environment is purely electrostatic in origin. Calculations of the relevant electrostatic free energies are based on the Poisson-Boltzmann (PB) model of the protein-solvent system and the finite-difference solution to the corresponding PB equation. The resultant multiple site titration problem is treated by one of two methods. The first is a hybrid approach, based on collecting ionizable groups into clusters. The second method is a Monte Carlo approach based on the Metropolis algorithm for extracting a sufficient number of low-energy ionization states out of all possible states, to obtain a correct estimation of thermodynamic properties of the system.

As examples of applications, we present the overall accuracy of predicted ionization constants for about 50 groups in 4 proteins, changes in the average charge of bovine pancreatic trypsin inhibitor at pH 7 along a molecular dynamics trajectory, and finally, we discuss some preliminary results obtained for protein kinases and protein phosphatases.

1 Introduction

Studies of electrical interactions in proteins, polypeptides, and amino acids started over 60 years ago [1]. To a large extent, electrostatic properties of proteins are determined by the ability of certain amino acids to exchange protons with their environment and the dependence of these processes on pH. The proton occupies a special position as a promoter and mediator in

chemical reactions occuring in solution. Many reactions in organic chemistry are catalyzed by acids or bases; likewise, most enzymes contain active groups which promote acid-base catalysis [2]. The activity of such enzymes requires that catalytic residues exist in the appropriate protonation state. Moreover, protonation states of ionizable residues surrounding the active site are also important since their electrostatic field can stabilize transition states along the reaction path. Finally, the enzyme's overall charge distribution, with the main contribution resulting from ionizable amino acids, can play a significant role in steering its substrates toward the active site.

The PB model for macromolecular electrostatics treats the solute as a low-dielectric region bounded by the molecular surface and containing atomic charges. The atomic charges are typically drawn from a molecular mechanics force field. The charges are located at atomic positions determined by X-ray crystallographic methods or high-resolution NMR spectroscopy. The solute is surrounded by high-dielectric aqueous solvent which may contain a dissolved electrolyte. The mobile charge distribution in the solvent is described by a Boltzmann distribution. Long-range electrostatic interactions are explicitly accounted for, while the electronic polarizability is implicitly included in the assumed dielectric constants of the media.

In this article, we present the basic principles underlying our algorithm for predicting ionization constants of residues in proteins based on the PB model. As examples of applications, we present the overall accuracy of computed ionization constants for about 50 groups in 4 proteins, changes in the average net charge of bovine pancreatic trypsin inhibitor (BPTI) at pH 7 along a molecular dynamics trajectory, and finally, we discuss some results obtained for protein kinases and protein phosphatases, proteins controlling a variety of fundamental cellular processes.

2 Theory

2.1 Protonation Equilibria in Simple Systems

Consider first a proton dissociation reaction for a group AH being a part of a small molecule

$$\mathrm{AH} \rightleftharpoons \mathrm{A}^- + \mathrm{H}^+ \ . \tag{1}$$

The dissociation constant, K_a, and standard free energy change, ΔG^o, for this reaction are related by the equation,

$$K_\mathrm{a} = \frac{[A^-][H^+]}{[AH]} = \exp(\frac{-\Delta G^o}{\mathrm{R}T}) \ , \tag{2}$$

where $[X]$ refers to the concentration of species X; R is the gas constant; and T is the absolute temperature. ΔG^o is the standard free energy for the state with the dissociated molecule and proton in solution relative to the state with

the proton bound to the molecule. Introducing the well known definitions of pH and pK_a

$$pH = -\log_{10}[H^+] \qquad pK_a = -\log_{10} K_a \ , \tag{3}$$

we can convert (2) to the form

$$\log_{10} \frac{[A^-]}{[AH]} = pH - pK_a \ . \tag{4}$$

Assuming that the concentrations are proportional to the probabilities of being in a given state, and using the following distribution,

$$\frac{[A^-]}{[AH]} = \exp(\frac{-\Delta G}{RT}) \ , \tag{5}$$

and choosing the neutral state of the group as the reference state, we finally obtain

$$\Delta G = 2.303RT\gamma(pH - pK_a) \ , \tag{6}$$

where γ is $+1$ for bases, and -1 for acids. ΔG can be considered as the free energy of the ionized state relative to the neutral one [3]. For a protein with M ionizable groups, the free energy can be expressed as a sum of terms shown by (6). One should distinguish between the group as a part of an isolated amino acid and the group as a subunit of a macromolecule. Referring to the first as the model compound environment, and to the second as the protein environment, one can define for each group i the two following pK_a values: the first $pK_{i,\mathrm{model}}$ and the second $pK_{i,\mathrm{protein}}$. The model compound used for each amino acid is generated by excising the residue from the protein without completing the N- and C-termini (e.g. HNCα(R)CO, where R corresponds to the side chain starting at Cβ). This partial backbone is neutral, as it is in the polypeptide chain in the protein. The experimental pK_a for each model compound was derived from molecules that most closely resemble the entire model compound. In most cases, this means that the model compound pK_a was derived from a small organic molecule (e.g. MeSH for Cys). However, in some cases the model compound pK_as were adjusted to better represent experimentally determined pK_as for the blocked and native amino acids. Thus, the free energy of the ionization state $(x_1, ..., x_M)$ of the whole protein is:

$$\Delta G(x_1, ..., x_M) = 2.303RT \sum_i x_i \gamma_i (pH - pK_{i,\mathrm{protein}}) \ , \tag{7}$$

where x_i is 1 when the group i is ionized, and 0 when it is neutral. Using (2) one can write

$$pK_{i,\mathrm{model}} = \frac{\Delta G^o_{i,\mathrm{model}}}{2.303RT} \ , \tag{8}$$

and

$$\mathrm{pK}_{i,\mathrm{protein}} = \frac{\Delta G^o_{i,\mathrm{protein}}}{2.303\mathrm{R}T} \quad . \tag{9}$$

Therefore, we can define a standard free energy difference $\Delta\Delta G^o_{i,\mathrm{m}\to\mathrm{p}} \equiv \Delta G^o_{i,\mathrm{protein}} - \Delta G^o_{i,\mathrm{model}}$ and rewrite (7) in the form:

$$\Delta G(x_1, ..., x_M) = 2.303\mathrm{R}T \sum_i x_i \gamma_i (\mathrm{pH} - \mathrm{pK}_{i,\mathrm{model}} - \frac{\Delta\Delta G^o_{i,\mathrm{m}\to\mathrm{p}}}{2.303\mathrm{R}T}) \quad , \tag{10}$$

Thus, one can calculate the free energy of the macromolecule in a given ionization state, provided that $\mathrm{pK}_{i,\mathrm{model}}$, and $\Delta\Delta G^o_{i,\mathrm{m}\to\mathrm{p}}$ are known.

2.2 Protonation Equilibria for Interacting Sites

The basic assumption underlying the method described here for predicting ionization equilibria in proteins is that each $\Delta\Delta G^o_{i,\mathrm{m}\to\mathrm{p}}$ is electrostatic in origin. Within the framework of the PB model for the solute-solvent system, $\Delta\Delta G^o_{i,\mathrm{m}\to\mathrm{p}}$ can be computed from the 4 electrostatic free energies based on the following sets of point charges representing the residue in its isolated state and as a titratable group in the protein: a) a set of charges representing the neutral form of the residue; and b) a set of charges representing the charged form of the residue. The energy necessary to assemble the whole set of n point charges in an arbitrary dielectric body of ϵ_i immersed in an infinite medium with another dielectric constant ϵ_s (provided there is no field dependence of the permittivity of the dielectric) is [4]:

$$W = \frac{1}{2} \sum_{i=1}^{n} q_i \phi_i \quad , \tag{11}$$

where ϕ_i is the electrostatic potential at the location of charge q_i, and is considered to be the sum of the Coulombic potentials due to all other charges except the charge on i, plus the total reaction field at the location of the charge i. In principle, each $\Delta G_{i,\mathrm{m}\to\mathrm{p}}$ depends on the actual ionization states of all other titratable residues. However, assuming additivity of the electrostatic potentials created by fixed charges representing the atoms of the protein, one can write (7) in the form:

$$\Delta G(pH, x_1, ..., x_M) = 2.303\ \mathrm{R}T \sum_{i=1}^{M} x_i \gamma_i (\mathrm{pH} - \mathrm{pK}_{i,\mathrm{intrinsic}}) + \\ + \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} x_i x_j \gamma_i \gamma_j \Psi_{ij}, \tag{12}$$

where $\mathrm{pK}_{i,\mathrm{intrinsic}}$ is the intrinsic pK_a of group i [5]; and Ψ_{ij} is the absolute value of the interaction energy of groups i and j. The intrinsic pK_a of group i is given by:

$$\mathrm{pK}_{i,\mathrm{intrinsic}} = \mathrm{pK}_{i,\mathrm{model}} - \gamma_i \Delta\Delta G_i / 2.303 \mathrm{R}T, \tag{13}$$

where

$$\Delta\Delta G_i \equiv \Delta G^{\mathrm{electrostatic}}_{i,\mathrm{protein}} - \Delta G^{\mathrm{electrostatic}}_{i,\mathrm{model}}, \tag{14}$$

is the difference between the free energy change of ionization for group i in the otherwise unionized protein, and in a model compound of $\mathrm{pK}_{i,\mathrm{model}}$,

$$\Delta G^{\mathrm{electrostatic}}_{i,\mathrm{model}} = W^{(\mathrm{ionized})}_{i,\mathrm{model}} - W^{(\mathrm{neutral})}_{i,\mathrm{model}}, \tag{15}$$

and

$$\Delta G^{\mathrm{electrostatic}}_{i,\mathrm{protein}} = W^{(\mathrm{ionized})}_{i,\mathrm{protein}} - W^{(\mathrm{neutral})}_{i,\mathrm{protein}}. \tag{16}$$

The intrinsic pK_a thus represents the pK_a of the group in the protein with all other titratable amino acids in their neutral state. Finally, the interaction energy, between the sites i and j, is

$$\Delta\Delta G_{i,j} = \Delta G^{\mathrm{electrostatic}}_{i,j,\mathrm{protein}} - (\Delta G^{\mathrm{electrostatic}}_{i,\mathrm{protein}} + \Delta G^{\mathrm{electrostatic}}_{j,\mathrm{protein}}) \equiv \gamma_i \gamma_j \Psi_{ij}, \tag{17}$$

where $\Delta G^{\mathrm{electrostatic}}_{i,j,\mathrm{protein}}$ has a similar meaning as $\Delta G^{\mathrm{electrostatic}}_{i,\mathrm{protein}}$, but there are two sites, i and j, ionized.

2.3 The Poisson-Boltzmann Equation

The Poisson equation relates spatial variation of the potential ϕ at position $\mathbf{r}$ to the density of the charge distribution, ρ, in a medium with a dielectric constant ϵ

$$\nabla \cdot (\epsilon(r) \nabla \phi(r)) = -4\pi\rho(r) \ . \tag{18}$$

Equation (18) is valid when the polarizability of the dielectric is proportional to the electrostatic field strength [4]. The operator ∇ in the Cartesian coordinate system has the form $\nabla \equiv (\partial/\partial x, \partial/\partial y, \partial/\partial z)$.

When this equation is applied to a system composed of a macromolecule immersed in an aqueous medium containing a dissolved electrolyte, the fixed partial charges of each atom of the macromolecule result in a charge density described by ρ^{f}, and the mobile charges of the dissolved electrolyte are described by ρ^{m}, which i derived from a Boltzmann distribution of the ions and coions.

$$\rho(r) = \rho^f(r) + \rho^m(r) = \rho^f(r) + \lambda(r) \sum_i q_i n_i \exp(-q_i \phi(r)/\mathrm{k}T) \ , \tag{19}$$

where n_i is the bulk number density of ions of type i and λ is 1 in mobile ion-accessible regions and 0 everywhere else. Linearizing the exponential terms in the mobile charge distribution into the form

$$\sum_i q_i n_i \exp(-q_i\phi(r)/\mathrm{k}T) \approx \sum_i q_i n_i - (1/\mathrm{k}T)\sum_i n_i q_i^2 \phi(r) \ , \tag{20}$$

and introducing a modified Debye-Hückel parameter κ (modified to be dielectric independent)

$$\bar{\kappa}^2 = \frac{4\pi \sum_i n_i q_i^2}{\mathrm{k}T} \tag{21}$$

one obtains the linearized PB equation

$$\nabla \cdot (\epsilon(r)\nabla\phi(r)) = -4\pi\rho^f(r) + \lambda(r)\bar{\kappa}^2\phi(\mathbf{r}) \ . \tag{22}$$

Full derivation of this equation was presented by Garrett and Poladian [6], see also [7, 8, 9].

3 Models

3.1 Structures

The method described here can be applied to proteins whose structures have been solved crystallographically or by high-resolution multidimensional NMR spectroscopy [10]. The present calculations require that coordinates be established for hydrogen atoms. Polar and aromatic hydrogens are added with the `HBUILD` command [11] of `CHARMM` [12]. The `CHARMM` forcefields allow two possible models for participation of hydrogen atoms in the molecular model: the first includes only polar hydrogens and the second one all hydrogens. For the purpose of the PB approach to the problem of ionization equilibria in proteins, an intermediate hydrogen model was introduced; this includes all polar and all aromatic ring hydrogens. This allows for inclusion of potentially important interactions of ionizable groups with a charge distribution on aromatic rings of such amino acids like phenylalanine and tyrosine, although at present the high polarizability of the charge distribution of such rings is not included in the model. Hydrogens are added in such a way that resulting structures are fully protonated, i.e. they correspond to low pH conditions. The positions of the hydrogens are further optimized by 500 steps of steepest descent energy minimization. By default, the protonation site for all carboxylic acids is the second oxygen atom in the structure file. No default exists for histidines and so the user is required to specify which nitrogen, **ND1** or **NE2**, is the one that deprotonates to generate the neutral form. In all calculations reported here, **NE2** was taken to be deprotonatable.

3.2 Modeling Ionization Processes

Ionization refers to the process of losing or accepting a proton. This influences the total charge and the distribution of charges. Both effects are reflected in our model. An example of this is shown below for aspartic acid (Asp). When the residue is considered neutral, the atomic charges and radii have values corresponding to a *neutral* parameter set, when it is considered ionized, the atomic charges and radii have values taken from a *charged* parameter set. In the charged form of the Asp residue both the charge and the radius of the HD atom are set to zero. The UHBD program [13, 14] used for the electrostatic calculations, does not include atoms with zero radii in construction of the dielectric map for the system. Therefore, not only changes in the charge distribution accompanying the ionization process are taken into account but also the changes in the dielectric boundary between the solvent and the solute are included.

		neutral				charged	
		q	r			q	r
ASP	N	-0.400	1.630	ASP	N	-0.400	1.630
ASP	H	0.250	0.713	ASP	H	0.250	0.713
ASP	CA	0.100	2.018	ASP	CA	0.100	2.018
ASP	CB	0.000	1.991	ASP	CB	0.000	1.991
ASP	CG	0.700	1.666	ASP	CG	0.140	1.666
ASP	OD1	-0.550	1.354	ASP	OD1	-0.570	1.479
ASP	OD2	-0.400	1.381	ASP	OD2	-0.570	1.479
ASP	HD	0.250	0.713	ASP	HD	0.000	0.000
ASP	C	0.600	1.666	ASP	C	0.600	1.666
ASP	O	-0.550	1.381	ASP	O	-0.550	1.381

Because changes in the point charges of some atoms of the titratable amino acids upon protonation or deprotonation, are taken into account, together with the changes in the dielectric boundary accompanying the process, we call this method a "full-group" titration model. In this respect it is similar to the method described by Yang et al. [15], and possibly also to that of Bashford and Gerwert [16]. However, the present method is particularly efficient: four PB calculations per ionizable group provide all of the necessary energy terms. In addition, the procedure takes advantage of the scripting features of UHBD, and is therefore highly automated.

3.3 Parameters

The solvent dielectric constant, ionic strength and temperature are chosen to fit the conditions of the experimental studies. The protein dielectric constant is assigned some small value, e.g. 4. The PB calculations are currently carried out with the atomic charges and radii of the PARSE parameter set, developed by Honig and coworkers [17] or that for CHARMM [12]. The PARSE parameter set

is attractive because, when used with a solvation model based upon the PB equation, it yields accurate solvation energies for a variety of small molecules. The CHARMM parameter set has the advantage of including parameters for more molecules and is also more easily accessible. Values for $pKa_{model,i}$ are tabulated in [18]. The protein dielectric boundary is taken to be the Richards probe-accessible (i.e. molecular) surface [19], computed with a probe of radius of 1.4 Å, and an initial dot-density of 500 per atom [20].

4 Methods

4.1 The Finite-Difference Method

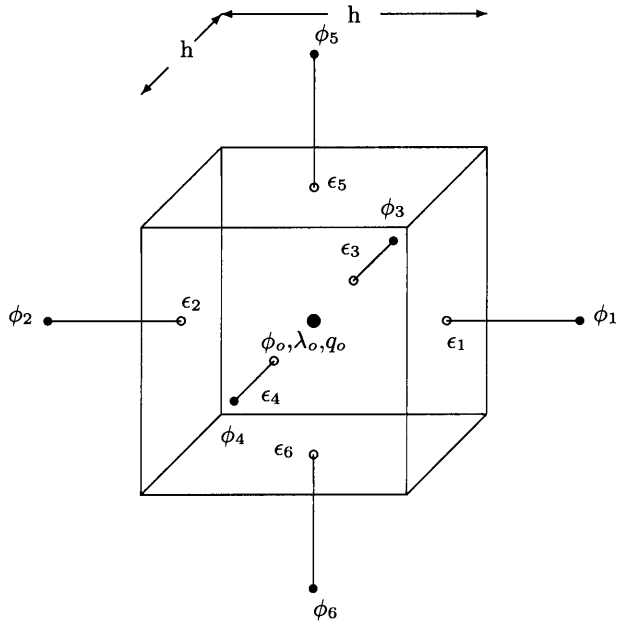


Fig. 1. Explanation of the principles of the finite-difference method for solution of the Poisson-Boltzmann equation

The UHBD program used in our calculations [13, 14], generates the electrostatic potentials for investigated systems by solving the PB equation using the finite-difference method. The principle of this method is as follows. Consider a small cube of side length h centered at a certain point, say $\mathbf{r}_o$ (see Fig. 1). Integrating (22) over the volume occupied by the cube and applying Gauss' theorem ($\int_V \nabla \cdot \mathbf{A}) \mathrm{d}v = \oint_\Sigma \mathbf{A} \cdot \mathbf{n} \mathrm{d}\sigma$), approximating continuous functions by distinct values at indicated points inside and outside the cube, and

finally approximating derivatives by the ratio of the differences, we get the following equation relating the potentials, charges and dielectric constants at the grid points

$$h^2 \sum_{i=1}^{6} \frac{\epsilon_i(\phi_i - \phi_o)}{h} = -4\pi q_o + \lambda(\mathbf{r}_o)\bar{\kappa}^2 \phi_o h^3 \ . \tag{23}$$

Equation (23) represents the essence of the finite-difference method [21, 22, 23, 24].

In the finite-difference method, the macromolecule(s) and a region of the surrounding solvent are mapped onto a cubic lattice; each lattice point represents a small region of either the molecule(s) or the solvent. At each point, values for the charge density, dielectric constant, and the ionic strength parameters are assigned. The whole procedure can be represented by the following steps:

1. Map all fixed charges onto a regular cubic grid of mesh size h: if a, b, and c are the fractional distances along the grid axes that the actual location of the charge q_o is from the grid point, then the fractional charge assigned to that grid point is $q = q_o(1-a)(1-b)(1-c)$ [22, 13].
2. Determine the dielectric constants for each of the faces between the grid points.
   a. Each grid point is assigned the solvent dielectric constant, ϵ_s, unless it is within the selected radius of any atom, where it is then assigned the interior value, ϵ_i.
   b. intermediate points between the given pair of two grid points is assigned a dielectric constant according to a smoothing function,

$$\epsilon_i \epsilon_s / (\epsilon_i d + \epsilon_s(1-d)) \tag{24}$$

   where d is the fraction of the grid line that is in the solvent [25].
3. Assign λ value of 1 or 0 to each grid point to identify points that are or are not accessible to mobile ions, respectively.
4. Solve the set of linear equations

$$\begin{aligned} 4\pi q(i,j,k)/h = {} & \epsilon(i-\tfrac{1}{2},j,k)(\phi(i,j,k) - \phi(i-1,j,k)) + \\ & \epsilon(i+\tfrac{1}{2},j,k)(\phi(i,j,k) - \phi(i+1,j,k)) + \\ & \epsilon(i,j-\tfrac{1}{2},k)(\phi(i,j,k) - \phi(i,j-1,k)) + \\ & \epsilon(i,j+\tfrac{1}{2},k)(\phi(i,j,k) - \phi(i,j+1,k)) + \\ & \epsilon(i,j,k-\tfrac{1}{2})(\phi(i,j,k) - \phi(i,j,k-1)) + \\ & \epsilon(i,j,k+\tfrac{1}{2})(\phi(i,j,k) - \phi(i,j,k+1)) + \\ & \lambda(i,j,k)\bar{\kappa}^2 \phi(i,j,k) h^2 \ . \end{aligned} \tag{25}$$

with an appropriate boundary condition for the electrostatic potential at the grid edges, using one of the available iteration methods [21, 22, 23, 24].

The model presented above requires a minimum of four UHBD calculations for each titratable group: two for the protein environment and two for the model compound environment. Each calculation is carried out with charges on the titratable group in question appropriate to either the neutral or ionized state, and with all other charges set to zero. Although the form of the equation for the interaction energy suggests that this interaction has to be separately calculated for each pair of the sites, the following equation explains the approximation used in order to avoid this apparent necessity,

$$\Delta G_{ij} = \sum_{k_i=1}^{m_i'} \sum_{k_j=1}^{m_j'} q'_{k_j} \phi'^{p}_{k_i k_j} - \sum_{k_i=1}^{m_i'} \sum_{k_j=1}^{m_j} q_{k_j} \phi'^{p}_{k_i k_j} - \sum_{k_i=1}^{m_i} \sum_{k_j=1}^{m_j'} q'_{k_j} \phi^{p}_{k_i k_j} + \sum_{k_i=1}^{m_i} \sum_{k_j=1}^{m_j} q_{k_j} \phi^{p}_{k_i k_j} \quad , \tag{26}$$

where charges q are the charges of the group j and the potentials ϕ are the potentials at the location of the charges of the group j, arising from the presence of all charges of the group i; the primed symbols refer to the ionized state and the unprimed symbols refer to the neutral state of the group i or j.

Because of the use of the "focusing" method [18], more than four calculations are actually carried out for each group. However, the focusing method saves computer time by permitting the use of less extensive finite-difference grids.

This set of calculations results in an output file containing all of the energies required to solve for the pH-dependent properties of interest. The form and content of the file are as follows:

M
$pK_{1,model}$ γ_1 $\Delta\Delta G_1$ 1
$\Psi_{1,2}$
⋮
$\Psi_{1,M}$
$pK_{2,model}$ γ_2 $\Delta\Delta G_2$ 2
$\Psi_{2,3}$
⋮
$\Psi_{2,M}$
⋮
$pK_{M,model}$ γ_M $\Delta\Delta G_M$ M

The first line contains the number of ionizable sites, M. Subsequent lines are organized in blocks whose first line contains information about the group

itself, and followed by information about its interactions with other groups. The first line of each block contains the model compound pK_a of the group ($pK_{i,model}$), the type of the group (γ_i), the electrostatic free energy difference for ionization of the group in the protein with all other sites in the neutral state, relative to the same change in the model compound ($\Delta\Delta G_i$), and the index of the group. The remaining $M - i$ lines of each block contain the absolute values of the effective interaction potentials (Ψ_{ij}) with the remaining $M - i$ sites ($j = i + 1, ..., M$).

The data in this file may be used in the computation of the electrostatic free energy of a given protein in any of its 2^M ionization states. Evaluation of average pH-dependent properties of a protein would require computation of the energies of all ionization states and subsequent evaluation of the average property, based on a Boltzmann's distribution. However, the total number of ionization states is prohibitively large for many systems of interest. Therefore, instead of enumerating all possible states, some other methods must be used in order to predict pH-dependent properties of proteins. In the subsequent section two possible approaches are described. The first is implemented in the `Hybrid` program [26] for rapid evaluation of mean charges, pK_as and electrostatic free energies for proteins, and the second is based on a Monte Carlo procedure as implemented in the `DOPS` (Distribution Of Protonation States) program [27]. The latter method is much slower, but provides results not only for mean charges and pK_as but also for fluctuations of the charges. It also provides a list of a predefined number of the lowest energy protein protonation states found during the Monte Carlo search, which can help to identify (nearly) degenerate states. Both programs use the above energy matrix as input.

4.2 Treatment of Multiple Ionizations by a Hybrid Approach

In an elegant paper, Gilson [26] presented a treatment of multiple ionization equilibria based on separation of ionizable groups into clusters. Each cluster is distinguished by strong interactions between its members and as a consequence possible strong correlations between ionization states of individual residues. Interactions between groups belonging to different clusters are much smaller, and therefore intercluster correlations are small. A full partition function ($\Sigma_i e^{-G_i/RT}$) is evaluated for each cluster which is then used to determine fractional ionizations of titratable groups belonging to each cluster. Contribution of cluster-cluster interactions to energies of titratable groups of a given cluster are treated by a mean field approach [28]. In the mean field approximation, the ionization equilibrium of group i is influenced by group j according to $\theta_j G_{ij}$, where θ_j is fractional ionization of group j.

The method is implemented by initially selecting a state in which each group is fully ionized. The partition function of the first cluster is evaluated using the initial-guess charges for groups in other clusters with higher indices, and fractional ionizations of its groups are obtained. The partition

function and fractional ionizations for the second cluster are calculated, using the previously evaluated fractional ionizations for the first cluster and the initial-guess charges for groups in other clusters with higher indices, and so on. All clusters are thus subsequently updated, and the process is iterated to self-consistency, where the criterion for convergence is that changes in fractional charges fall below some small value. This procedure is the same as that of Tanford and Roxby [29] except that the iterative loop is over clusters of groups, instead of individual groups. The iterations proceed reasonably quickly for clusters up to about 10 groups, which is the maximal allowed size of any cluster. The method of dividing ionizable groups into clusters and other details of the procedure have been fully described previously [26].

4.3 Treatment of Multiple Ionizations by a Monte Carlo Method

The DOPS program [27, 30] extracts the most probable ionization states of the protein from 2^M total states, using a Metropolis Monte Carlo (MC) algorithm [31]. The initial state is the protein with all sites protonated. Given a starting state, the Gibbs free energy, G1, is computed by use of (6). A random number in the range $\{0-1\}$ is then picked for each ionizable group in the protein, and the ionization state is switched for each group whose random number is greater than some cutoff value s. Thus, an average of $(1.0-s)*100\%$ of the ionizable groups are toggled at each step. If s is 0.99, then ca. 1% of the ionizable residues will be changed at each MC step. The second ionization state is then accepted or rejected depending upon its free energy, G2, again calculated by use of (6): when G2$\leq$G1, the new state is accepted; when G2>G1, the new state is accepted with a probability of $e^{-(G2-G1)/RT}$. For this another random number in the range $\{0-1\}$ is picked - if it is smaller than $e^{-(G2-G1)/RT}$ the new state is accepted. Since the starting state is always far from equilibrium (all residues ionized), the initial *equilibration* steps are discarded, and only the subsequent *production* steps are used in computing thermodynamic averages. The number of *equilibration* and *production* steps should be chosen depending on the size of the system. The value of s is adjusted so that approximately 50% of new states are accepted.

The Monte Carlo approach, although much slower than the Hybrid method, makes it possible to address very large systems quite efficiently. It should be noted that the Monte Carlo approach gives a correct estimation of thermodynamic properties even though the number of *production* steps is a tiny fraction of the total number of possible ionization states.

5 Algorithm

In order to summarize the procedures used for computing ionization constants of titratable residues in proteins, the steps used in our algorithm will be enumerated below:

1. Preparation of a Brookhaven Protein Data Bank (PDB)-formatted [10] file containing the coordinates and appropriate names of all atoms, including all polar and aromatic hydrogens.
2. Use of the finite-difference PB (FDPB) method to calculate the self- and interaction-energies of the ionizable groups in the protein and solvent.
3. Use of a Monte Carlo or a "cluster" (`Hybrid`) algorithm to calculate ionization constants of the titratable groups, net average charges, and electrostatic free energies as functions of pH.

The procedure is computationally efficient. For example, for the catalytic subunit of the mammalian cAMP-dependent protein kinase and its inhibitor, with 370 residues and 131 titratable groups, an entire calculation requires 10 hours on an SGI O2 workstation with a 175 MHz MIPS R10000 processor. The bulk of the computer time is spent on the FDPB calculations. The speed of the procedure is important, because it makes it possible to collect results on many systems and with many different sets of parameters in a reasonable amount of time. Thus, improvements to the method can be made based on a broad sampling of systems.

6 Sample Applications

6.1 Overall Efficiency of the Algorithm

The presented algorithm was applied to 4 proteins (lysozyme, ribonuclease A, ovomucid and bovine pancreatic trypsin inhibitor) containing 51 titratable residues with experimentally known pK_as [32, 33]. Fig. 2 shows the correlation between the experimental and calculated pK_as. The linear correlation coefficient is r = 0.952; the slope of the line is A = 1.028; and the intercept is B = -0.104. This shows that the overall agreement between the experimental and predicted pK_as is good.

However, some significant deviations between the experimental and predicted pK_a values are visible and some improvements and modifications of the procedure are possible.

6.2 Total Charge of BPTI Along MD Trajectory

In molecular mechanics and molecular dynamics studies of proteins, assigment of standard, non-dynamical ionization states of protein titratable groups is a common practice. This assumption seems to be well justified because proton exchange times between protein and solution usually far exceed the time range of the MD simulations. We investigated to what extent the assumed protonation state of a protein influences its molecular dynamics trajectory, and how often our titration algorithm predicted ionization states identical to those imposed on the groups, when applied to a set of structures derived from a molecular dynamics trajectory [34]. As a model we took the bovine

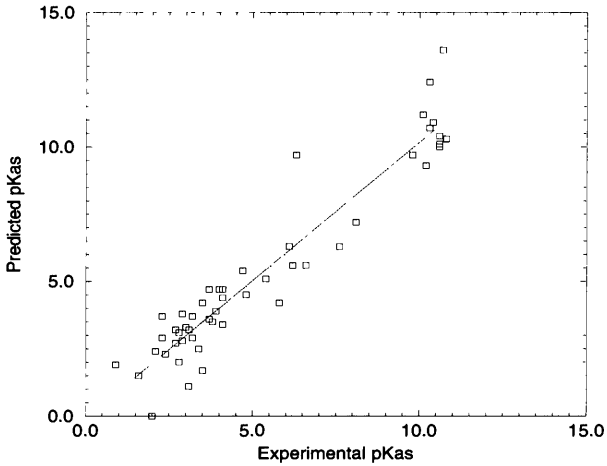


Fig. 2. Correlation between predicted and experimental pK$_a$s in 4 globular proteins

pancreatic trypsin inhibitor, a 58-residue protein which is used as the prototypical small protein (BPTI) and is the subject of many experimental and theoretical studies.

Fig. 3 shows the evolution of the charge distribution calculated with the PB equation of a set BPTI structures generated from the 200 ps MD simulation in solution. In this simulation the total charge of the protein was set to 5 e, by neutralization of the protein N-terminal group and the use of standard protonation states for all other protonation sites. However, the starting x-ray structure (PDB entry code [12] 4pti) is consistent with fractional ionization of the N-terminal group of 0.6 which indicates that for this structure the probability for this group of being neutral is slightly smaller than that of being charged. It is seen that at the initial stage of simulation (0-100 ps) the maximum net charge was about 5.8 e, in agreement with the fact that the starting structure taken for the simulation favored a larger charge than 5 e. The total charge becomes, however, very rapidly centered at the value of 5 e, so the distribution at 100-200 ps is nearly unimodal around that value, indicating an excellent agreement with the imposed charge during the MD simulation. The result presented in Fig. 3 shows that even a single charge mutation can substantially change the protein conformations explored during the MD simulation. This result not only emphasizes the need for the careful protein charge assigment, but also points out the direction of future development of the MD simulation techniques which would allow for modifying the titration state of a protein during the simulations.

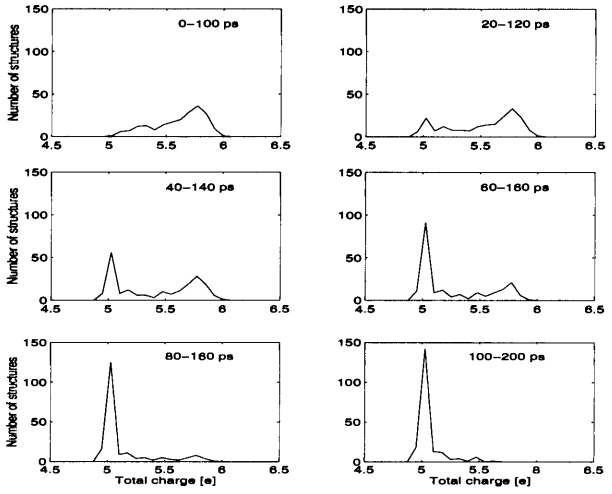


Fig. 3. Charge distributions of the BPTI structures sampled from the indicated subranges of the MD trajectory for the BPTI X-ray structure; from ref. [34]

6.3 Protein Kinases and Phosphatases

The last part of this account will be devoted to protein kinases and protein phosphatases and some recent results we have obtained for them. Protein kinases and phosphatases are signaling biomolecules that control the level of phosphorylation and dephosphorylation of tyrosine, serine or threonine residues in other proteins, and by this means regulate a variety of fundamental cellular processes including cell growth and proliferation, cell cycle and cytoskeletal integrity.

Left side of Fig. 4 shows a ribbon model of the catalytic (C-) subunit of the mammalian cAMP-dependent protein kinase. This was the first protein kinase whose structure was determined [35]. Figure 4 includes also a ribbon model of the peptide substrate, and ATP (stick representation) with two manganese ions (CPK representation). All kinetic evidence is consistent with a preferred ordered mechanism of catalysis with ATP binding preceeding substrate binding.

The catalytic subunit then catalyzes the direct transfer of the γ-phosphate of ATP (visible as small beads at the end of ATP) to its peptide substrate. Catalysis takes place in the cleft between the two domains. Mutual orientation and position of these two lobes can be classified as either *closed* or *open*, for a review of the structures and function see e.g. [36]. The presented structure shows a closed conformation. Both the apoenzyme and the binary complex of the porcine C-subunit with di-iodinated inhibitor peptide represent the crystal structure in an open conformation [37] resulting from an overall rotation of the small lobe relative to the large lobe.

The explicitly shown residues in the ribbon models of the catalytic subunit and the substrate are the phosphorylated threonine residue at position 197 (P-Thr197) (CPK representation) and the serine residue at position 17 of the substrate (Ser17) (stick representation). Phosphorylation of Thr197 is essential for activity of the kinase [38]. Our simulations (for mammalian protein kinase) suggest that at pH 7 the preferred form of the phosphate group of P-Thr197 in the closed form is doubly ionized, whereas in the open form the single ionization state is strongly preferred. Another interesting feature observed was that Ser17 of the model of substrate bound to the closed and open forms of protein kinases is very hard to ionize since, for no case considered, was the pK_a of the residue found to be below 20. However, as is shown in Table 1, effects of neutralization of important acidic residues (Asp166, Asp184 and Glu91) is significant - decreasing of the pK_a of Ser17 by 6 pH units is observed for neutralization of Asp166. On the other hand, forcing the ionized state of Ser17 of the substrate increases pK_a of Asp166 to 7.9 in the closed form and to 10.9 in the open form. This indicates strong correlations between the protonation equilibria of Asp166 and Ser17 in accordance with the expected enzymatic mechanism. Finally, it should be mentioned that Glu170, Glu203 and Glu230 play a significant role in binding of the peptide substrate, mainly *via* their interactions with Arg residues of the substrate. According to our estimate, the total electrostatic interaction with Arg residues of the substrate in the closed form is about 20 kcal/mol stronger than in the open one.

The right side of Fig. 4 shows a ribbon model of the protein tyrosine phosphatase (PTPase) from *Yersinia* with atoms of the catalytic Cys403 residue shown using a CPK representation. This enzyme catalyzes a reversible dephosphorylation of tyrosine residues in proteins, but the mechanism is not well understood. One way of approaching this problem is to investigate the role of the conserved residues which are present in the PTPase signature motif. Site directed mutagenesis studies, besides selecting the biologically relevant reaction pathway, have also focused on determining the pK_a of the active site Cys403 in the wild–type enzyme and selected mutants. Knowledge of the protonation states of the residues located in the neighborhood of Cys403 is essential for proposing a relevant mechanism. The pK_a of the active site thiol group was experimentally determined to be 4.67 [39]. This is an unusually low value because the pK_a of a Cys residue is usually 8.5 [39]. It suggests that strong electrostatic interactions are responsible for stabilizing this anion [39].

Table 2 shows the results of our preliminary calculations of the pK_a of the Cys403 residue, for several different models of the enzyme, based on two structures available from the PDB. In the case of the YPT structure, a crystal water molecule is close to Cys403 and was included in some of the calculations as part of the protein (i.e. it was treated with the same internal dielectric as that of the protein). Simulations denoted as $+H_2O$ in Table 2, include a crystallographically resolved, buried water molecule, situated 3.21Å from

Fig. 4. Ribbon model of protein kinase with peptide substrate and Mn_2ATP (left) and protein phosphatase (right)

Table 1. Effects of neutralization of some acidic residues on the pK_a of Ser17 residue of the substrate

mutation	shifts in pK_a of serine in the substrate	
	the present algorithm	in ref. [40]
Glu-91 →Gln	-1.9	-3.0
Asp-166→Asn	-5.5	-6.0
Asp-184→Asn	-2.7	-5.0

the sulphur atom of Cys403. In the third row of Table 2, +min1, means that minimization of the structure was performed by the CHARMM program [12] for a spherical region of radius 5 Å around the SG atom of Cys403, and +min2 means that the minimization was performed for a 12 Å radius spherical region. During the minimization Glu290, Asp356 and His402, i.e. residues important for the catalysis or protonation state of Cys403, were kept protonated. Finally, +27 a.a. means that 27 N-terminal amino acids lacking from the PDB file were added to the structure using the Biopolymers module of the InsightII software [41]. In the case of the YTS structure, it should be mentioned that an SO_4^{-2} ion is present and its S atom is 3.45Å away from the sulphur atom of Cys403.

Table 2. Predicted intrinsic and apparent pK_a values for the Cys403 residue in *Yersinia* phosphatase for different models of the structure; the data refer to a temperature of 293 K and an ionic strength corresponding to 150 mM of monovalent salt. See the text for the detailed description of the conditions under which each pK estimation was made. The experimentally determined value is 4.67 [39]

structure	$pK_{a,int}$	$pK_{a,app}$
open structure – YPT		
wild-type	7.3	4.1
$+H_2O$	10.1	5.4
+min1	8.6	3.4
+min2	8.5	3.3
+27 a.a.	9.3	4.7
closed structure – YTS		
wild-type	10.5	1.6
$+SO_4^{-2}$	30.1	28.6

It can be seen from Table 2 that the intrinsic values of the pK_as are close to the model compound value that we use for Cys(8.3), and that interactions with surrounding titratable residues are responsible for the final apparent values of the ionization constants. It can also be seen that the best agreement with the experimental value is obtained for the YPT structure suplemented with the 27 N-terminal amino acids, although both the original YPT structure and the one with the crystal water molecule give values close to the experimentally determined one. Minimization, however, makes the agreement worse, probably because it was done without the presence of any solvent molecules, which are important for the residues on the surface of the protein. For the YTS structure, which refers to the protein crystallized with an SO_4^{-2} ion, the results with and without the ion included in the calculations, are far from the experimental value. This may indicate that con-

formational change corresponding to the transition from the YPT structure to the YTS structure is accompanied by a stronger tendency of Cys403 to be ionized. However, the presence of a negative ion, possibly that of the substrate, in the vicinity of Cys403 leads to a significant increase of its predicted pK_a. These aspects are currently under investigation.

Acknowledgments. The work described in this article was supported by Fogarty NIH Award TW00768 (JAM and JA), NIH grants GM31749 (JAM) and GM56553 (JAM and JMB), by a grant from the State Committee for Scientific Research, Poland (8T11F00609; to BL), and by a grant of computer time from the San Diego Supercomputing Center (JAM and JMB).

References

1. Edsall, J. T.: George Scatchard, John G. Kirkwood, and the electrical interactions of amino acids and proteins. Trends Biochem. Sci. **7** (1982) 414–416.
2. Eigen, M.: Proton transfer, acid-base catalysis, and enzymatic hydrolysis. Angew. Chem. Int. Ed. Engl. **3** (1964) 1–19.
3. Poland, D.: Cooperative equilibria in physical biochemistry. Clarendon Press, Oxford, 1978.
4. Böttcher, C. J. F.: Theory of electric polarization, Volume I. Elsevier, Amsterdam, 1973.
5. Tanford, C., Kirkwood, J. G.: Theory of protein titration curves. I. General equations for impenetrable spheres. J. Am. Chem. Soc. **79** (1957) 5333–5339.
6. Garrett, A. J. M., Poladian, L.: Refined derivation, exact solutions, and singular limits of the Poisson-Boltzmann equation. Ann. Phys. **188** (1988) 386–435.
7. Sharp, K. A., Honig, B.: Electrostatic interactions in macromolecules. Theory and applications. Ann. Rev. Biophys. Chem. **19** (1990) 301–332.
8. Sharp, K. A., Honig, B.: Calculating total electrostatic energies with the nonlinear Poisson-Boltzmann equation. J. Phys. Chem. **94** (1990) 7684–7692.
9. Zhou, H.-X.: Macromolecular electrostatic energy within the nonlinear Poisson-Boltzmann equation. J. Chem. Phys. **100** (1994) 3152–3162.
10. Bernstein, F. C., Koettzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., Tasumi, M. J.: The protein data bank: A computer-based archival file for molecular structures. J. Mol. Biol. **123** (1977) 557–594.
11. Brunger, A. T., Karplus, M.: Polar hydrogen positions in proteins: Empirical energy placement and neutron diffraction comparison. Proteins: Struct. Func. Genet. **4** (1988) 148–156.
12. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., Karplus, M.: CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J. Comp. Chem. **4** (1982) 187–217.
13. Davis, M. E., Madura, J. D., Luty, B. A., McCammon, J. A.: Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics program. Comp. Phys. Commun. **62** (1991) 187–197.
14. Madura, J. D., Briggs, J. M., Wade, R. C., Davis, M. E., Luty, B. A., Ilin, A., Antosiewicz, J., Gilson, M. K., Bagheri, B., Scott, L. R., McCammon,

J. A.: Electrostatics and disffusion of molecules in solution: simulations with the university of houston brownian dynamics program. Comp. Phys. Commun. **91** (1995) 57–95.

15. Yang, A. S., Gunner, M. R., Sampogna, R., Sharp, K., Honig, B.: On the calculation of pKas in proteins. Proteins: Struct. Func. Genet. **15** (1993) 252 - 265.
16. Bashford, D., Gerwert, K.: Electrostatic calculations of the pKa values of ionizable groups in bacteriorhodopsin. J. Mol. Biol. **224** (1992) 473-486.
17. Sitkoff, D., Sharp, K. A., Honig, B.: Accurate calculation of hydration free energies using macroscopic solvent models. J. Phys. Chem. **98** (1994) 1978–1988.
18. Antosiewicz, J., McCammon, J. A., Gilson, M. K.: Prediction of pH-dependent properties of proteins. J. Mol. Biol. **238** (1994) 415–436.
19. Richards, F. M.: Areas, volumes, packing and protein structure. Ann. Rev. Biophys. Bioeng. **6** (1977) 151–176.
20. Gilson, M. K., Sharp, K. A., Honig, B. H.: Calculating the electrostatic potential of molecules in solution: Method and error assessment. J. Comp. Chem. **9** (1988) 327–335.
21. Warwicker, J., Watson, H. C.: Calculation of the electric potential in the active site cleft due to α-helix dipoles. J. Mol. Biol. **157** (1982) 671–679.
22. Klapper, I., Hagstrom, R., Fine, R., Sharp, K., Honig, B.: Focusing of electric fields in the active site of cu,zn superoxide dismutase. Proteins: Struct. Func. Genet. **1** (1986) 47–79.
23. Davis, M. E., McCammon, J. A.: Solving the finite difference linearized Poisson-Boltzmann equation: A comparison of relaxation and conjugate gradients methods.. J. Comp. Chem. **10** (1989) 386-394.
24. Nicholls, A., Honig, B.: A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. J. Comp. Chem. **12** (1991) 435–445.
25. Davis, M. E., McCammon, J. A.: Dielectric boundary smoothing in finite difference solutions of the poisson equation: An approach to improve accuracy and convergence. J. Comp. Chem. **12** (1991) 909–912.
26. Gilson, M. K.: Multiple-site titration and molecular modeling: Two rapid methods for computing energies and forces for ionizable groups in proteins. Proteins: Struct. Funct. Genet. **15** (1993) 266–282.
27. Antosiewicz, J., Porschke, D.: Electrostatics of hemoglobins from measurements of the electric dichroism and computer simulations. Biophys. J. **68** (1995) 655-664.
28. Bashford, D., Karplus, M.: Multiple-site titration curves of proteins: an analysis of exact and approximate methods for their calculation. J. Phys. Chem. **95** (1991) 9556-9561.
29. Tanford, C., Roxby, R.: Interpretation of protein titration curves: Application to lysozyme. Biochem. **11** (1972) 2192–2198.
30. Antosiewicz, J.: Computation of the dipole moments of proteins. Biophys. J. **69** (1995) 1344–1354.
31. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E.: Equation of state calculations by fast computing machines. J. Chem. Phys. **21** (1953) 1087 1092.

32. Antosiewicz, J., Briggs, J. M., Elcock, A. E., Gilson, M. K., McCammon, J. A.: Computing the ionization states of proteins with a detailed charge model. J. Comp. Chem. **17** (1996) 1633–1644.
33. Antosiewicz, J., McCammon, J. A., Gilson, M. K.: The determinants of pK_as in proteins. Biochem. **35** (1996) 7819–7833.
34. Wlodek, S. T., Antosiewicz, J., McCammon, J. A.: Prediction of titration properties of structures of a protein derived from molecular dynamics trajectories. Protein Sci. **6** (1997) 373–382.
35. Knighton, D. R., Zheng, J., Eyck, L. F. T., Ashford, V. A., Xuong, N., Taylor, S. S., Sowadski, J. M.: Crystal structure of the catalytic subunit of cAMP-dependent protein kinase. Science **253** (1991) 407–420.
36. Taylor, S. S., Radzio-Andzelm, E.: Cyclic AMP-dependent protein kinase. In Protein Kinases, Woodgett, J. R., editor, IRL Press, Oxford, 1994.
37. Karlsson, R., Zheng, J., Zheng, N.-H., Taylor, S. S., Sowadski, J. M.: Structure of the mamalian catalytic subunit of cAMP-dependent protein kinase and an inhibitor peptide displays an open conformation. Acta Cryst. D **49** (1993) 381-388.
38. Steinberg, R. A.: A kinase-negative mutant of s49 mouse lymphoma cells is defective in posttranslational maturation of catalytic subunitof cyclic amp-dependent protein kinase. Mol. Cell Biol. **11** (1991) 705-712.
39. Zhang, Z.-Y., Dixon, J. E.: Active site labeling of the *yersinia* protein tyrosine phosphatase: The determination of the pK_a of active site cysteine and the function of the conserved histidine 402. Biochem. **32** (1993) 9340-9345.
40. Tsigelny, I., Grant, B. D., Taylor, S. S., Ten Eyck, L. F.: Catalytic subunit of cAMP-dependent protein kinase: Electrostatic features and peptide recognition. Biopolymers **39** (1996) 353–365.
41. Molecular Simulations, Inc.: InsightII. Molecular Simulations, Inc., Waltham, MA, 1992.

Exploiting Tsallis Statistics

John E. Straub and Ioan Andricioaei

Department of Chemistry, Boston University, Boston, Massachusetts 02215

Abstract. Two incarnations of the canonical ensemble probability distribution based on the generalization of statistical mechanics of Tsallis are described. A generalization of the law of mass action is used to derive equilibrium constants. Reaction rate constants for barrier crossing are derived using the transition state theory approximation. Monte Carlo and Molecular Dynamics algorithms which can be used to sample Tsallis statistical distributions are defined. The results are used to demonstrate that MC and MD algorithms which sample the Tsallis statistical distributions can be expected to enhance the rate of phase space sampling in simulations of many body systems.

1 An Introduction to Tsallis Statistics

Nearly ten years ago, Tsallis proposed a possible generalization of Gibbs-Boltzmann statistical mechanics.[1] He built his intriguing theory on a re-expression of the Gibbs-Shannon entropy $S = -k \int p(\Gamma) \ln p(\Gamma) d\Gamma$ written

$$S = \lim_{q \to 1} S_q = \lim_{q \to 1} \frac{k}{q-1} \int p_q(\Gamma)(1 - [p_q(\Gamma)]^{q-1}) d\Gamma \tag{1}$$

where $d\Gamma = d\mathbf{r}^N d\mathbf{p}^N$ is a phase space increment. On the right, the "replica trick" identity $\ln x = \lim_{n \to 0}(x^n - 1)/n$ has been used; q is a real number.[1, 2]

Tsallis studied the properties of S_q which he referred to as a "generalized entropy." He noted that much of the standard mathematical structure of Gibbs-Boltzmann statistical mechanics remained intact *before* the limit is taken; that is, for $S_{q \neq 1}$. This prompted the use of a generalized formalism based on the non-additive entropy S_q to re-derive, for non-extensive systems, a variety of results of the standard statistical mechanics (see [3] and references therein). For example, it is possible to define the probability of finding the system at a given point in phase space $\Gamma = (\mathbf{r}^N, \mathbf{p}^N)$ by extremizing S_q subject to the constraints

$$\int p_q(\Gamma) d\Gamma = 1 \qquad \text{and} \qquad \int [p_q(\Gamma)]^q H(\Gamma) d\Gamma = \mathrm{E_q} \tag{2}$$

where $H(\Gamma)$ is the system Hamiltonian for N distinguishable particles in d dimensions. The result is

$$p_q(\Gamma) = \frac{1}{Z_q h^{dN}} \left[1 - (1-q)\beta H(\Gamma)\right]^{\frac{1}{1-q}} \tag{3}$$

where

$$Z_q = \frac{1}{h^{dN}} \int [1 - (1-q)\beta H(\Gamma)]^{\frac{1}{1-q}} d\Gamma \tag{4}$$

plays the role of the canonical ensemble partition function. Using the identity $\lim_{n\to 0}(1+an)^{1/n} = \exp(a)$, in the limit that $q = 1$, the standard probability of classical Gibbs-Boltzmann statistical mechanics

$$p(\Gamma) = \frac{1}{Zh^{dN}} \exp(-\beta H(\Gamma)) \tag{5}$$

is recovered.

1.1 Surely You're Joking, Mr. Tsallis

Before the limit is taken, the properties of the probability distribution appear to be strange in at least five ways.

(1) In the $q \neq 1$ regime, for certain points in phase space the probability $p_q(\Gamma)$ may be negative or even imaginary. We then say that the probability of the system accessing that point in phase space is zero. This may be the case even when the energy is finite.

(2) Equally foreign is the property that when $q \neq 1$ the relative probability of two points in phase space depends on the choice of the zero of energy.[1] By defining

$$\bar{H}(\Gamma) = \frac{1}{\beta(q-1)} \ln [1 - (1-q)\beta H(\Gamma)] \tag{6}$$

the probability of being at a point Γ in phase space can be written in the familiar form

$$p(\Gamma) = \frac{1}{Zh^{dN}} \exp(-\beta \bar{H}(\Gamma)) \tag{7}$$

However, for a constant potential shift ϵ the relative probability

$$\frac{p_q(\Gamma_{\text{new}})}{p_q(\Gamma_{\text{old}})} = \left[\frac{1 - (1-q)\beta(H(\Gamma_{\text{new}}) + \epsilon)}{1 - (1-q)\beta(H(\Gamma_{\text{old}}) + \epsilon)}\right]^{\frac{1}{1-q}} \tag{8}$$

depends on ϵ. This ratio can be rewritten as

$$\frac{p_q(\Gamma_{\text{new}})}{p_q(\Gamma_{\text{old}})} = \left[\frac{1 - (1-q)\beta' H(\Gamma_{\text{new}})}{1 - (1-q)\beta' H(\Gamma_{\text{old}})}\right]^{\frac{1}{1-q}} \tag{9}$$

where potential shift ϵ has been absorbed in an effective "temperature"

$$\frac{1}{\beta'} = \frac{1}{\beta} + (q-1)\epsilon. \tag{10}$$

In the $q = 1$ limit, the effective temperature equals the standard temperature. Otherwise, adding a constant shift to the potential energy is equivalent to rescaling the temperature at which the canonical probability distribution is computed.

(3) Equilibrium average properties are calculated using a statistical weighting of the probability $p_q(\Gamma)$ of Eq. (3) raised to the power of q as required by the generalized statistical mechanics. The so-called q-expectation value is written

$$\langle \dots \rangle_q = \int [p_q(\Gamma)]^q \dots d\Gamma. \tag{11}$$

In general, $\langle 1 \rangle_q \neq 1$ for $q \neq 1$. Clearly, this is an odd "average!" It is also inconvenient as it requires evaluation of Z_q.

(4) Is the "temperature" $1/\beta$ related to the variance of the momentum distribution as in the classical equipartition theorem? It happens that there is no simple generalization of the equipartition theorem of classical statistical mechanics. For the 2N dimensional phase space $\Gamma = (x_1 \dots x_N, p_1, \dots p_N)$ the ensemble average for a *harmonic* system is

$$\frac{\langle p_k^2 \rangle_q}{\langle 1 \rangle_q} = \frac{\int [p_q(\Gamma)]^q p_k^2 d\Gamma}{\int [p_q(\Gamma)]^q d\Gamma} = \frac{1}{\beta} \frac{1}{1-(q-1)N} \tag{12}$$

where we assume unit mass. For the case of $q = 1$ we find the standard result that $\langle p_k^2 \rangle = 1/\beta$. In general we find that the average is proportional to $1/\beta$ but not equal to it. The situation is equally strange for the unnormalized "multifractal" average where

$$\langle p_k^2 \rangle_q = \text{stuff} \times \frac{1}{\beta^{1+(1-q)N/2}} \tag{13}$$

but the "stuff" is a q-dependent constant that may be negative or imaginary!

The distribution of momenta cannot be written as a product of single particle distributions and we find that

$$\langle p_k^2 + p_k^2 \rangle_q \neq 2 \langle p_k^2 \rangle_q. \tag{14}$$

There is no simple linear scaling of the variance of the momentum with the number of degrees of freedom.

(5) When $q = 1$ the extensivity of the entropy can be used to derive the Boltzmann entropy equation $S = k \ln W$ in the microcanonical ensemble. When $q \neq 1$, it is the odd property that the generalization of the entropy S_q is not extensive that leads to the peculiar form of the probability distribution. The non-extensivity of S_q has led to speculation that Tsallis statistics may be applicable to gravitational systems where interaction length scales comparable to the system size violate the assumptions underlying Gibbs-Boltzmann statistics.[4]

1.2 Ideal Systems

Let's focus on the $q > 1$ regime for systems with a Hamiltonian of the form

$$H(\Gamma) = \sum_k^N \frac{1}{2m}\mathbf{p}_k^2 + U(\mathbf{r}^N), \tag{15}$$

in d-dimensional space, the partition function can be written as

$$Z_q = \left(\frac{1}{\Lambda\sqrt{q-1}}\right)^{dN} \frac{\Gamma(\frac{1}{q-1} - \frac{dN}{2})}{\Gamma(\frac{1}{q-1})} \int \left[1 - (1-q)\beta U(\mathbf{r}^N)\right]^{\frac{1}{1-q}+\frac{dN}{2}} d\mathbf{r}^N \tag{16}$$

where $\Lambda = \sqrt{h^2\beta/2\pi m}$ is the thermal de Broglie wavelength of the kth oscillator. For an ideal gas of N particles where $U = 0$ we find

$$Z_q = \left(\frac{L}{\Lambda\sqrt{q-1}}\right)^{dN} \frac{\Gamma(\frac{1}{q-1} - \frac{dN}{2})}{\Gamma(\frac{1}{q-1})} \tag{17}$$

where L is the length of a side of a cubic box containing the ideal gas. In the limit that $q \to 1$ we can use the asymptotic approximation $\Gamma(x+a)/\Gamma(x+b) = x^{a-b}$, good for large x, to show that the standard partition function for an ideal gas is recovered

$$\lim_{q\to 1} Z_q = \left(\frac{L}{\Lambda}\right)^{dN}. \tag{18}$$

Now consider a system of N one-dimensional harmonic oscillators with the Hamiltonian

$$H(\Gamma) = \sum_k^N \left[\frac{1}{2m}p_k^2 + \frac{1}{2}m\omega_k^2 x_k^2\right]. \tag{19}$$

The canonical ensemble partition function is the phase space integral

$$Z_q = \frac{1}{h^N} \int dx^N \int dp^N \left[1 - (1-q)\beta H(\Gamma)\right]^{\frac{1}{1-q}}. \tag{20}$$

The configuration integral can be evaluated and the partition function is

$$Z_q = \left(\frac{2\pi}{h\beta(q-1)}\right)^N \frac{\Gamma(\frac{1}{q-1} - N)}{\Gamma(\frac{1}{q-1})} \prod_k^N \frac{1}{\omega_k} \tag{21}$$

Note that there is not a unique separation of the partition function as $Z_q = Z_q^{\text{trans}} Z_q^{\text{vib}}$. However, using the result for the ideal gas translational partition function

$$Z_q^{\text{vib}} = \left(\frac{2\pi}{m\beta(q-1)}\right)^{N/2} \frac{1}{L^N} \frac{\Gamma(\frac{1}{q-1} - N)}{\Gamma(\frac{1}{q-1} - \frac{N}{2})} \prod_k^N \frac{1}{\omega_k} \tag{22}$$

In the limit that $q \to 1$ we see that

$$\lim_{q\to 1} Z_q = \left(\frac{2\pi}{\beta h}\right)^N \prod_k^N \frac{1}{\omega_k} \tag{23}$$

and the canonical ensemble partition function for N harmonic oscillators in classical Gibbs-Boltmann statistics is recovered.

1.3 Ensemble Averages and the "q-Expectation Value"

For certain values of q and a harmonic potential, the distribution $p_q(\Gamma)$ can have infinite variance and higher moments. This fact has motivated the use of the q-expectation value to compute the average of an observable A

$$\langle A\rangle_q = \frac{1}{(Z_q h^N)^q} \int A(\Gamma)\ [1-(1-q)\beta H(\Gamma)]^{\frac{q}{1-q}} d\Gamma. \tag{24}$$

Since the averaging operator is not normalized and in general $\langle 1\rangle_q \neq 1$ for $q \neq 1$, it is necessary to compute Z_q to determine the average. To avoid this difficulty, we employ a different generalization of the canonical ensemble average

$$\langle A\rangle_q = \frac{\int A(\Gamma)[1-(1-q)\beta H(\Gamma)]^{\frac{q}{1-q}} d\Gamma}{\int [1-(1-q)\beta H(\Gamma)]^{\frac{q}{1-q}} d\Gamma}, \tag{25}$$

which is obviously normalized.

Consider a system of N particles in d dimensions. Using the standard procedure of integrating over the momenta in Cartesian coordinates, we can write the average of a mechanical property $A(\mathbf{r}^N)$ as

$$\langle A\rangle_q = \frac{\int A(\mathbf{r}^N)\left[1-(1-q)\beta U(\mathbf{r}^N)\right]^{\frac{q}{1-q}+\frac{dN}{2}} d\mathbf{r}^N}{\int \left[1-(1-q)\beta U(\mathbf{r}^N)\right]^{\frac{q}{1-q}+\frac{dN}{2}} d\mathbf{r}^N} \tag{26}$$

This definition is based on and proportional to the q-expectation value. However, it is more useful since it is not necessary to evaluate the partition function to compute an average.

1.4 Monte Carlo Methods for Pure Tsallis Statistics

A configurational Monte Carlo algorithm based on uniform random trial moves and the acceptance probability

$$p = \min\left[1, \left(\frac{p_q(\mathbf{r}^N_{\text{new}})}{p_q(\mathbf{r}^N_{\text{old}})}\right)^{q+\frac{dN}{2}(1-q)}\right], \tag{27}$$

where $p_q(\mathbf{r}) \propto [1-(1-q)\beta U(\mathbf{r})]^{\frac{1}{1-q}}$, will sample the configurational distribution dictated by Eq. (26). Such a Monte Carlo algorithm may be used to compute equilibrium averages in the Tsallis statistical distribution.

In the thermodynamic limit of large N for $q \neq 1$ we find that the acceptance probability is unity for even the largest positive change in potential energy ΔU. Effectively, the Monte Carlo sampling reduces to a random walk on the potential energy landscape. This feature can be used to devise a sampling scheme to overcome broken ergodicity. Only part of the time the acceptance would be that in Eq. (27), while the rest of the time will be spent sampling according to the standard Metropolis criterion.

Alternatively, one may use a phase space Monte Carlo method with uniform random trial moves and an acceptance probability

$$p = \min\left[1, \left(\frac{p_q(\Gamma_{\text{new}})}{p_q(\Gamma_{\text{old}})}\right)^q\right] \tag{28}$$

to sample the phase space distribution $[p_q(\Gamma)]^q$. In the thermodynamic limit, this form of the acceptance probability does not suffer from the perculiar behavior exhibited by the form above. Moreover, an algorithm of this sort can be used to calculate standard, Gibbs-Boltzmann ($q = 1$) equilibrium thermodynamic averages

$$\langle A \rangle = \left\langle \frac{A\, \mathrm{e}^{-\beta H(\Gamma)}}{[1-(1-q)\beta H(\Gamma)]^{\frac{q}{1-q}}} \right\rangle_q \left\langle \frac{\mathrm{e}^{-\beta H(\Gamma)}}{[1-(1-q)\beta H(\Gamma)]^{\frac{q}{1-q}}} \right\rangle_q^{-1}. \tag{29}$$

Using this expression, the standard $q = 1$ equilibrium average properties may be calculated over a trajectory which samples the generalized statistical distribution for $q \neq 1$ with the advantage of enhanced sampling for $q > 1$.

1.5 Chemical Equilibrium

Consider the chemical equilibrium

$$\alpha \rightleftharpoons \gamma. \tag{30}$$

The general form of the equilbrium constant is

$$K_{eq} = \frac{X_\gamma}{X_\alpha} = \frac{\int [p_q]^q \theta_\gamma(\Gamma) d\Gamma}{\int [p_q]^q \theta_\alpha(\Gamma) d\Gamma} \tag{31}$$

where X_α and X_γ are mole fractions of reactants and products, and the Heaviside functions $\theta_\alpha(\Gamma)$ and $\theta_\gamma(\Gamma)$ are unity for phase space points in well α and γ, respectively, and zero otherwise. We will restrict our evaluation to one dimension; extension to many dimensions is straightforward.

In the limit that the barrier height is large compared with the thermal energy, it is standard practice to expand the potential near the reactant well

minimum to quadratic order and approximate the integral by the contribution near the well minimum. This approximation is well justified when the thermal distribution is Gaussian and relatively short-ranged. However, for $q > 1$ the approximation is not so easily justified. For $q = 2$ the equilibrium distribution for a harmonic well is a Cauchy-Lorentz distribution so that it is likely that significant contributions from anharmonicity far from the well minimum will contribute to the integral. Aware of this limitation, we carry out the integral using a harmonic approximation to the α and γ potential wells. For the α state well

$$U(x) \simeq U(x_\alpha) + m\omega_\alpha^2 (x - x_\alpha)^2/2 \tag{32}$$

We follow this with a low temperature approximation to the integral over the well's phase space population. The resulting equilibrium constant is

$$K_{eq}(\beta; q) = \frac{X_\gamma}{X_\alpha} = \frac{\omega_\alpha}{\omega_\gamma} \left[\frac{1 - (1-q)\beta U(x_\alpha)}{1 - (1-q)\beta U(x_\gamma)} \right]^{\frac{1}{1-q}}. \tag{33}$$

In the limit that $q \to 1$, the equilibrium distributions are more delocalized and the low temperature approximation may not be well justified.

1.6 Transition State Theory for Rates of Barrier Crossing

The transition state theory estimate of the rate constant for barrier crossing provides an upper bound on the rate of transition between well-defined reactant and product states. An assumption of the theory is that once a reactant acquires enough energy to cross the barrier, it will cross the transition state and be deactivated as product. Dynamical recrossings of the transition state, associated with weak or strong damping, or nonadiabatic transitions, lead to reductions in the rate of barrier crossing from the transition state theory estimate. In this section, we examine the transition state theory rate constant for Tsallis statistics. This discussion is restricted to one-dimension, but the generalization to many dimensions is straightforward.

There is still some debate regarding the form of a dynamical equation for the time evolution of the density distribution in the $q \neq 1$ regime. Fortunately, to evaluate the rate constant in the transition state theory approximation, we need only know the form of the equilibrium distribution. It is only when we wish to obtain a more accurate estimate of the rate constant, including an estimate of the transmission coefficient, that we need to define the system's dynamics.

For a one-dimensional bistable potential with the transition state positioned along the reaction coordinate x at $x = x^\ddagger$, the TST rate for forward reaction is defined as

$$k_{TST}(\beta; q) = \frac{\int \delta(x - x^\ddagger) v \Theta(v) [p_q(\Gamma)]^q d\Gamma}{\int \Theta(x^\ddagger - x) [p_q(\Gamma)]^q d\Gamma} \tag{34}$$

where $p_q(\Gamma) = p_q(x,v)$ is the generalized statistical distribution

$$p_q(x,v) = \frac{1}{Z_q h}\left[1-(1-q)\beta H(x,v)\right]^{\frac{1}{1-q}} \tag{35}$$

where

$$Z_q = \frac{1}{h}\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dv[1-(1-q)\beta H(x,v)]^{\frac{1}{1-q}} \tag{36}$$

is the generalized partition function and $H(x,v) = mv^2/2 + U(x)$ is the Hamiltonian.

For $q < 1$ there can be difficulties which arise from distributions which have zero probability in the barrier region and zero rate constant. In our analysis we assume that for any q the zero of energy is chosen such that the probability $p_q(\Gamma)$ is positive and real for all Γ. The transition state theory rate constant as a function of the temperature and q is

$$k_{TST}(\beta;q) = \frac{\Gamma(\frac{q}{q-1})}{\Gamma(\frac{q}{q-1}-\frac{1}{2})}\left(\frac{q-1}{2\pi m\beta}\right)^{1/2}\left[1-(1-q)\beta U(x^\ddagger)\right]^{\frac{1}{1-q}}\frac{1}{\chi(\beta;q)} \tag{37}$$

where

$$\chi(\beta;q) = \int_{-\infty}^{x^\ddagger}\left[1-\beta(1-q)U(x)\right]^{\frac{q}{1-q}+\frac{1}{2}}dx. \tag{38}$$

We can approximate this fraction of states in the reactant well, by expanding the potential in a harmonic approximation and assuming that the temperature is low compared with the barrier height. This leads to an estimate for the rate constant

$$k_{TST}(\beta;q) = \frac{\omega_\alpha}{2\pi}\left[\frac{1-(1-q)\beta U(x^\ddagger)}{1-(1-q)\beta U(x_\alpha)}\right]^{\frac{1}{1-q}}. \tag{39}$$

As we expect, in the limit that $q \to 1$ the standard transition state theory result

$$k_{TST}(\beta;q=1) = \frac{\omega_\alpha}{2\pi}\exp[-\beta(U(x^\ddagger)-U(x_\alpha))] \tag{40}$$

is recovered.

Returning to the more general expression, in the low temperature limit we find that the rate

$$k_{TST}(\beta;q) \to \frac{\omega_\alpha}{2\pi}\left[\frac{U(x^\ddagger)}{U(x_\alpha)}\right]^{\frac{1}{1-q}} \tag{41}$$

independent of the temperature (even when T=0)! For the special case of $q = 2$ we find

$$k_{TST}(\beta;q=2) = \frac{\omega_\alpha}{2\pi}\left[\frac{1+\beta U(x_\alpha)}{1+\beta U(x^\ddagger)}\right] \tag{42}$$

1.7 Temperature Scaling in Simulated Annealing

Using these results, we can derive a scaling relation for the optimal cooling schedule in a simulated annealing optimization protocol. We suppose that the relavent energy scales of $U(\mathbf{r}^N)$ are bounded by ΔU, the difference in energy between the ground and first excited state minima, and $U^\ddagger$, the highest barrier on the potential surface accessed from the global energy minimum. The final temperature (maximum β) reached in a simulated annealing run must be small enough so that at equilibrium the mole fraction in the global energy minimum basin is significant. In other words, based on Eq. (33) we demand that $K_{eq}^{\max} = K_{eq}(\beta_{\max}; q)$.

The time that the trajectory must spend at $\beta_{\max}$ to ensure that the equilibrium distribution is sampled is at least τ_{min}, the time required to surmount the largest barrier separating the global energy minimum from other thermodynamically important states. Using Eq. (39) we find

$$\tau_{\min} = \frac{2\pi}{\omega_\alpha}\left[1 - (1 - \frac{1}{\eta})\frac{U^\ddagger}{\Delta U}\right]^{\frac{1}{1-q}}. \tag{43}$$

where

$$\eta = \left(\frac{\omega_\alpha}{\omega_\gamma} K_{eq}^{\max}\right)^{q-1}. \tag{44}$$

$K_{eq}^{\max}$ is the maximum allowable equilibrium constant for the ground and first excited state populations at the final and lowest temperature reached in the annealing run, $\beta_{\max}$. For most cases of interest, we expect that $\eta << 1$.

In the limit $q \to 1$ of Gibbs-Boltzmann statistics, using the fact that $\lim_{x\to 0}[1 - a(1 - b^x)]^{1/x} = b^a$, we find that

$$\tau_{min} = \left(\frac{2\pi}{\omega_\alpha}\right)\eta^{-\frac{U^\ddagger}{\Delta U}}. \tag{45}$$

The time for classical simulated annealing increases exponentially as a function of the ratio of the energy scales $U^\ddagger/\Delta U$. However, for $q > 1$ the situation is qualitatively different. As a result of the weak temperature dependence in the barrier crossing times, the time for simulated annealing increases only weakly as a power law.

2 Maxwell-Tsallis Statistics

We have developed Monte Carlo algorithms based on sampling Tsallisian distributions. Using a uniform random trial move and the acceptance probability

$$p = \min\left[1, \left(\frac{p_q(\mathbf{r}_{\text{new}}^N)}{p_q(\mathbf{r}_{\text{old}}^N)}\right)^q\right]. \tag{46}$$

the detailed balance condition

$$[p_q(x)]^q W(x \to x') = [p_q(x')]^q W(x' \to x) \tag{47}$$

is satisfied where $W(x, x')$ is an element of the transition matrix. The walk generated by Eq. (28) will sample the distribution

$$[p_q(\mathbf{r}^N)]^q = \frac{1}{Z_q}\left[1 - (1-q)\beta U(\mathbf{r}^N)\right]^{\frac{q}{1-q}} \tag{48}$$

This probability distribution can be found by extremizing the generalization of the entropy Eq. (1) subject to the constraints

$$\int p_q(\mathbf{r}^N) d\mathbf{r}^N = 1 \qquad \text{and} \qquad \int [p_q(\mathbf{r}^N)]^q U(\mathbf{r}^N) d\mathbf{r}^N = \mathrm{U_q} \tag{49}$$

We might then assume a Maxwell distribution of momenta so that the overall phase space distribution is that of a Maxwell-Tsallis statistics.

2.1 Hybrid Monte Carlo Algorithm

We have implemented the generalized Monte Carlo algorithm using a hybrid MD/MC method composed of the following steps.

1. Velocities are randomly chosen from a Maxwell distribution at a given temperature.
2. The positions and velocities are updated for a time step Δt according to Newton's equation of motion using the force deriving from $\bar{U}$.
3. The point (phase space point or configuration) is accepted or rejected according to the criterion

$$p = \min\left[1, \exp[-\beta(\Delta K + \Delta \bar{U})]\right] \tag{50}$$

where ΔK is the change in standard classical kinetic energy and $\Delta \bar{U}$ is the change in the effective potential energy

$$\bar{U}(\mathbf{r}^N) = \frac{1}{\beta(q-1)} \ln\left[1 - (1-q)\beta U(\mathbf{r}^N)\right]. \tag{51}$$

4. Return to 1.

When the integrator used is reversible and symplectic (preserves the phase space volume) the acceptance probability will exactly satisfy detailed balance and the walk will sample the equilibrium distribution $[p_q(\mathbf{r}^N)]^q$.

A similar algorithm has been used to sample the equilibrium distribution $[p_q(\mathbf{r}^N)]^q$ in the conformational optimization of a tetrapeptide[5] and atomic clusters at low temperature.[6] It was found that when $q > 1$ the search of conformational space was greatly enhanced over standard Metropolis Monte Carlo methods. In this form, the velocity distribution can be thought to be Maxwellian.

2.2 Molecular Dynamics with an Effective Potential

For a given potential energy function $U(\mathbf{r}^N)$, the corresponding generalized statistical probability distribution which is generated by the Monte Carlo algorithm is proportional to

$$P_q(\mathbf{r}^N) = [1 - (1-q)\beta U(\mathbf{r}^N)]^{\frac{q}{1-q}}. \tag{52}$$

Consider the generalized distribution $P_q(\mathbf{r}^N)$ to be generated in the Gibbs-Boltzmann canonical ensemble ($q = 1$) by an effective potential $\mathcal{W}_q(\mathbf{r}^N;\beta)$ which is defined

$$P_q(\mathbf{r}^N) = \exp(-\beta \mathcal{W}_q(\mathbf{r}^N;\beta)) \tag{53}$$

such that when $q = 1$, $\mathcal{W}(\mathbf{r}^N;\beta) = U(\mathbf{r}^N)$. For $q \neq 1$, the effective potential will depend on temperature as well as the coordinates.

Given this effective potential, it is possible to define a constant temperature molecular dynamics algorithm such that the trajectory samples the distribution $P_q(\mathbf{r}^N)$. The equation of motion then takes on a simple and suggestive form

$$m_k \frac{d^2\mathbf{r}_k}{dt^2} = -\nabla_{\mathbf{r}_k}\bar{U} = -\nabla_{\mathbf{r}_k}U(\mathbf{r}^N)q[1-(1-q)\beta U(\mathbf{r}^N)]^{-1} \tag{54}$$

for a particle of mass m_k and position $\mathbf{r}_k$ and $\bar{U}$ defined by Eq. (51). It is known that in the canonical ensemble a constant-temperature molecular dynamics algorithm generates samples from the configuration space according to the Boltzmann probability. As a result, the molecular dynamics with the effective potential $\mathcal{W}_q(\mathbf{r}^N;\beta)$ will sample from the $P_q(\mathbf{r}^N)$ distribution.

The effective force derived from the effective potential $\mathcal{W}(\mathbf{r}^N)$ has a number of interesting properties. It is of the form $F_q(\mathbf{r}^N;\beta) = -\nabla_{\mathbf{r}_k}\bar{U} = F_1(\mathbf{r}^N)\alpha_q(\mathbf{r}^N;\beta)$ where $F_1(\mathbf{r}^N)$ is the "exact" force for standard molecular dynamics ($q = 1$) and $\alpha_q(\mathbf{r}^N;\beta)$ is a scaling function which is unity when $q = 1$ but can otherwise have a strong effect on the dynamics.

Assume that the potential is defined to be a positive function. In the regime $q > 1$, the scaling function $\alpha_q(\mathbf{r}^N,\beta)$ is largest near low lying minima of the potential. In barrier regions, where the potential energy is large, the scaling function $\alpha_q(\mathbf{r}^N,\beta)$ is small. It may surprise you that for the function to be well defined, βU must be greater than zero.

Use of the effective potential has the effect of reducing the magnitude of the force in the barrier regions. Therefore, a particle attempting to pass over a potential energy barrier will meet with less resistance when $q > 1$ than when $q = 1$. At equilibrium, this leads to more delocalized probability distributions with an increased probability of sampling barrier regions. This argument demonstrates that when $q > 1$ the generalized molecular dynamics or Monte Carlo trajectories will cross barriers more frequently and explore phase space more efficiently (for a review of recent methods for enhanced phase-space sampling see [7]).

2.3 Rate and Equilibrium Constants

We once again carry out the integral in the low temperature approximation and find

$$k_{TST}(\beta;q) = \frac{\omega_\alpha}{2\pi}\frac{\Gamma(\frac{q}{q-1})}{\Gamma(\frac{q}{q-1}-\frac{1}{2})}\frac{1}{\sqrt{1-(1-q)\beta U(x_\alpha)}}\left[\frac{1-(1-q)\beta U(x^\ddagger)}{1-(1-q)\beta U(x_\alpha)}\right]^{\frac{q}{q-1}} \quad (55)$$

Within the same approximation, we estimate that the equilibrium constant is

$$K_{eq}(\beta;q) = \frac{X_\beta}{X_\alpha} = \frac{\omega_\alpha}{\omega_\beta}\left[\frac{1-(1-q)\beta U(x_\alpha)}{1-(1-q)\beta U(x_\beta)}\right]^{\frac{1}{1-q}+\frac{1}{2}}. \quad (56)$$

As in the case of pure Tsallis statistics and Eq. (33), in the limit that $q \to 1$ the standard transition state theory result

$$k_{TST}(\beta;q=1) = \frac{\omega_\alpha}{2\pi}\exp[-\beta(U(x^\ddagger)-U(x_\alpha))] \quad (57)$$

is recovered.

Returning to the more general expression, in the low temperature limit we find that the transition state theory estimate of the rate is

$$k_{TST}(\beta;q) \to \frac{\omega_\alpha}{2\pi}\left[\frac{U(x^\ddagger)}{U(x_\alpha)}\right]^{\frac{1}{1-q}}\frac{1}{\sqrt{\beta U(x_\alpha)}} \quad (58)$$

which scales as $1/\sqrt{\beta}$ at low temperatures for all q.

For the special case of $q = 2$ we find

$$k_{TST}(\beta;q=2) = \frac{\omega_\alpha}{\pi}\left[\frac{1+\beta U(x_\alpha)}{1+\beta U(x^\ddagger)}\right]^2\frac{1}{\sqrt{1+\beta U(x_\alpha)}} \quad (59)$$

For $q = 1$, the normal transition state theory rate constant is independent of temperature at high temperatures and varies exponentially with temperature in the limit of low temperatures (kT small compared with the barrier height $U^\ddagger$) as

$$k_{TST} \sim \exp(-\beta U^\ddagger). \quad (60)$$

For $q \neq 1$, at high temperature the rate is independent of temperature $k_{TST} = \omega_\alpha/\pi$ which is a factor of two larger than the result for $q = 1$. At low temperature

$$k_{TST} \sim \beta^{-1/2} \quad (61)$$

independent of q! As q approaches unity the exponential Arrhenius temperature dependence is recovered. However, for larger values of q the temperature scaling of the rate is a weak inverse power law. At all temperatures, the transition state theory rate constant is significantly larger for $q > 1$ than for $q = 1$.

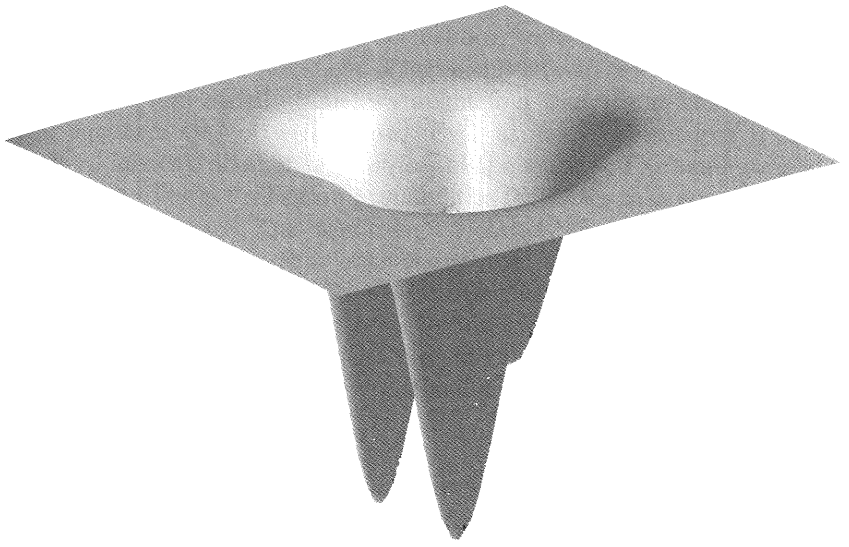


Fig. 1 The two-dimensional potential considered in the text has two deep holes, seen in the front and left sides of the plot, and a less deeper hole to the right.

2.4 Master Equations and Relaxation to Equilibrium

The relaxation of a system to equilibrium can be modeled using a master equation

$$\frac{dP_i}{dt} = \sum_{j \neq i} \left[L_{ij} P_j(t) - L_{ji} P_i(t) \right] . \tag{62}$$

The elements of the transition matrix from state j to state i can be estimated in the transition state theory approximation

$$L_{ij} = \frac{\omega_j}{2\pi} \left[\frac{1 + (q-1)\beta U(x_{ij}^{\ddagger})}{1 + (q-1)\beta U(x_j)} \right]^{\frac{1}{1-q}} , \tag{63}$$

where the total phase space probability of the jth state is proportional to

$$M_j = \frac{2\pi}{\beta \omega_j} \left[1 + (q-1)\beta U(x_j) \right]^{\frac{1}{1-q}} . \tag{64}$$

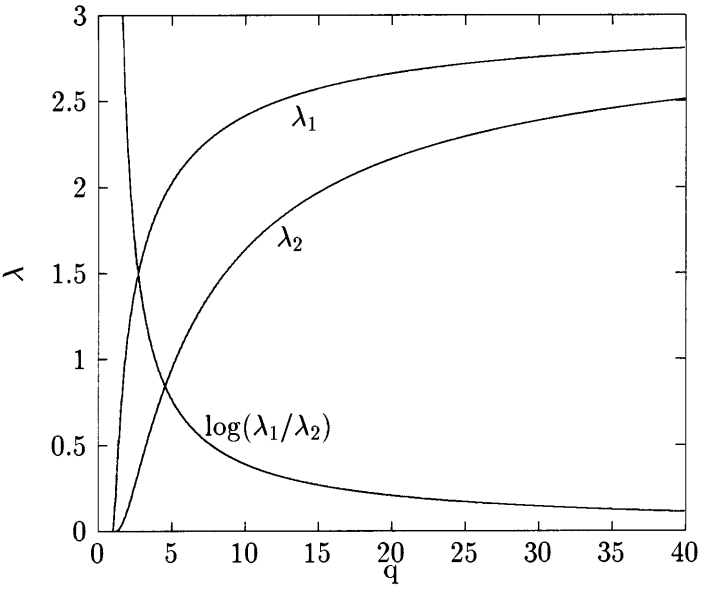


Fig. 2 The dependence on q of the non-zero eigenvalues of the two-dimesional, three-hole model described in the text shows the rapid onset of escape from wells as soon as q exceeds unity and the saturation at higher values of q. Also shown is the logarithm of the ratio of the two non-zero eigenvalues.

The symmetric transmission coefficients are defined $B_{ij} = L_{ij}M_i$. The general solutions are of the form

$$P_i(t) = \sum_n a_n \phi_i^{(n)} \exp(-\lambda_n t) \tag{65}$$

in terms of the eigenfunctions $\phi_i^{(n)}$ and eigenvalues λ_n of the transmission matrix $\hat{L}$.

We have calculated the eigenvalues for a two-dimensional model system described by a potential function consisting of three holes, two deeper holes of equal depth, and a more shallow hole. The barrier between the deeper holes is higher than the other two barriers, which have equal height.[8] One of the three eigenvalues is zero, for any q, for reasons of conservation of probability. In the figure, we plot the other two as a function of q, together with their ratio. There is to be observed, for a thermal energy of a tenth of the well depth, the quick increase in the magnitude of the eigenvalues for $q > 1$. This implies fast relaxation to equilibrium even at low temperatures, due to the delocalized character of the Tsallisian distributions.

Do we expect this model to be accurate for a dynamics dictated by Tsallis statistics? A jump diffusion process that randomly samples the equilibrium canonical Tsallis distribution has been shown to lead to anomalous diffusion and Lévy flights in the $5/3 < q < 3$ regime.[3] Due to the delocalized nature of the equilibrium distributions, we might find that the microstates of our master equation are not well defined. Even at low temperatures, it may be difficult to identify distinct microstates of the system. The same delocalization can lead to large transition probabilities for states that are not adjacent in configuration space. This would be a violation of the assumptions of the transition state theory - that once the system crosses the transition state from the reactant microstate it will be deactivated and equilibrated in the product state. Concerted transitions between spatially far-separated states may be common. This would lead to a highly connected master equation where each state is connected to a significant fraction of all other microstates of the system.[9, 10]

JES gratefully acknowledges the Alfred P. Sloan Foundation for support and the National Science Foundation for support (CHE-9632236) and computational resources at the National Center for Supercomputing Applications (CHE-960010N).

References

1. C. Tsallis. *J. Stat. Phys.*, 52:479, 1988.
2. E.M.F. Curado and C. Tsallis. *J. Phys. A: Math. Gen*, 24:L69, 1991.
3. C. Tsallis. *Phys. Rev. Lett.*, 75:3589, 1995.
4. J. Maddox. *Nature*, 365:103, 1993.
5. I. Andricioaei and J. E. Straub. *Phys. Rev. E*, E 53:R3055, 1996.
6. I. Andricioaei and J. E. Straub. *J. Chem. Phys.*, 107:9117, 1997.
7. B. J. Berne and J. E. Straub. *Curr. Opin. Struc. Bio.*, 7:181, 1997.
8. S. Huo and J.E. Straub. *J. Chem. Phys.*, 107:5000, 1993.
9. R. Zwanzig. *J. Chem. Phys.*, 103:9397, 1995.
10. Such "global" master equations have recently gained attention in the study of protein folding. See for example C. De Dominicis, H. Orland and F. Lainee, *J. Phys. Lett.*, 46:L463, 1985; E.I. Shakhnovich and A. M. Gutin, *Eurphys. Lett.*, 9: 569, 1989; J. G. Saven, J. Wang and P. G. Wolynes, *J. Chem. Phys.*, 101:11037, 1994.

New Techniques for the Construction of Residue Potentials for Protein Folding

Arnold Neumaier, Stefan Dallwig, Waltraud Huyer, and Hermann Schichl *

Institut für Mathematik, Universität Wien
Strudlhofgasse 4, A-1090 Wien, Austria

Abstract. A smooth empirical potential is constructed for use in off-lattice protein folding studies. Our potential is a function of the amino acid labels and of the distances between the C_α atoms of a protein. The potential is a sum of smooth surface potential terms that model solvent interactions and of pair potentials that are functions of a distance, with a smooth cutoff at 12 Ångström. Techniques include the use of a fully automatic and reliable estimator for smooth densities, of cluster analysis to group together amino acid pairs with similar distance distributions, and of quadratic programming to find appropriate weights with which the various terms enter the total potential. For nine small test proteins, the new potential has local minima within 1.3–4.7Å of the PDB geometry, with one exception that has an error of 8.5Å.

Keywords. protein folding, tertiary structure, potential energy surface, global optimization, empirical potential, residue potential, surface potential, parameter estimation, density estimation, cluster analysis, quadratic programming

1991 MSC Classification. primary 92C40; secondary 62H30, 90C20

1 Overview

The protein folding problem is the task of understanding and predicting how the information coded in the amino acid sequence of proteins at the time of their formation translates into the 3-dimensional structure of the biologically active protein. A thorough recent survey of the problems involved from a mathematical point of view is given by NEUMAIER [22].

The forces in a protein molecule are modeled by the gradient of the potential energy $V(s,x)$ in dependence on a vector s encoding the amino acid sequence of the molecule and a vector x containing the Cartesian coordinates of all essential atoms of a molecule. In an equilibrium state x, the forces $\nabla V(s,x)$ vanish, so x is stationary; and for stability reasons we must have a local minimizer. The most stable equilibrium state of a molecule is usually the

* The authors gratefully acknowledge support of this research by the Austrian Fond zur Förderung der wissenschaftlichen Forschung (FWF) under grant P11516-MAT.

native (tertiary) state. Thus, finding the native state of a protein molecule with sequence s is considered to be more or less equivalent to finding the global minimizer $\hat{x}$ of $V(s, x)$.

Therefore, modeling a protein molecule amounts to deciding on the atoms considered to be essential and to specifying the contribution of the various interactions to the potential. Since the work to find the global minimizer increases drastically (and possibly exponentially) with the dimension of x, it is customary to use for larger proteins a reduced description that treats only very few atoms in each amino acid as essential.

LEVITT & WARSHEL [17, 18] were the first to show that reduced representations may work; they used C_α atoms and virtual atoms at side chain centroids. OOBATAKE & CRIPPEN [24] simplified further by only considering the C_α atoms. This is sufficient since there are reasonably reliable methods (HOLM & SANDER [11, 12]) that compute a full atom geometry from the geometry of the C_α atoms. (All atom representations are used as well, but limited to the prediction of tiny systems such as enkephalin.)

Our potential is a sum of smooth *surface potentials* that model amino acid-solvent interactions and of smooth *pair potentials* that model amino acid-amino acid interactions. As in [24], we take as essential only the C_α atoms.

Traditionally, pair potentials are determined by assuming that a set of known structures, generally taken from the Brookhaven Protein Data Bank (PDB) [3, 35, 30], is an equilibrium ensemble of structures, so that, up to a constant factor determined by the temperature, the energy can be calculated from Boltzmann's law and statistics on the known structures. In order to obtain useful statistics, the protein structures used must be carefully selected; see, e.g., HOBOHM et al. [9]. A more detailed overview can be found in SIPPL [28]. Other empirical potential construction techniques are discussed in BAUER & BEYER [1] and ULRICH et al. [34]. The fact that the potential is directly derived from geometric data implies that it automatically takes account of solvation and entropy corrections; on the other hand, one only gets a mean potential of low resolution. Reconstructions using mean potentials are reported by SUN [31] for apamin (18 residues) and mellitin (26 residues) using genetic algorithms, by SUN [32] for mellitin, APPI (36 residues) and crambin (46 residues) using simulated annealing, by GUNN et al. [6] for myoglobin (153 residues) using a combination of simulated annealing and genetic algorithms, and by SIPPL et al. [29] for myoglobin and lysozyme (129 residues) by an assembly process using a fragment database. In all these papers, results are only 'native-like' when compared with the experimental structures.

To find appropriate empirical pair potentials from the known protein structures in the Brookhaven Protein Data Bank, it is necessary to calculate densities for the distance distribution of C_α-atoms at given bond distance d and given residue assignments a_1, a_2. Up to a constant factor that is immaterial for subsequent structure determination by global optimization, the potentials then emerge as the negative logarithm of the densities. Since

a huge number of pair potentials is required, fully automatic and reliable density estimators are necessary.

The only density estimators discussed in the protein literature are histogram estimates. However, these are nonsmooth and thus not suitable for global optimization techniques that combine local and global search. Moreover, histogram estimates have, even for an optimally chosen bin size, the extremely poor accuracy of $O(n^{-1/10})$ only, for a sample of size n. The theoretically attainable accuracy of the densities is much better, namely $O(n^{-1/2})$. (See HALL & MARRON [7] and the survey by JONES et al. [14].)

We therefore use smooth density estimation techniques that are more reliable than the histogram estimates. To improve the reliability for rare amino acid pairs, we use clustering techniques that identify 'similar' pairs that can be modeled by the same density.

Since the precise form of the potential and the estimation procedure are still in flux, we only discuss the main features but not all the details.

2 Empirical Residue Potentials

We assume that proteins are coded by a sequence

$$s = (s_1, \ldots, s_n)$$

of n amino acids labelled by $s_i \in \{1, \ldots, 20\}$ in one-to-one correspondence with the names of the 20 natural amino acids, and that

$$x_s = (x_1, \ldots, x_n)$$

is a vector listing the Cartesian coordinates $x_i \in \mathbb{R}^3$ of the C_α atoms in the native geometry of a protein molecule with sequence s.

An ideal empirical potential function on the residue level is a function V that assigns to each sequence-coordinates pair (s, x) an *energy* $V(s, x)$ such that

$$x_s \in \operatorname{argmin}_x \; V(s, x), \tag{1}$$

i.e., x_s is a solution of the global optimization problem associated with the sequence s.

In practice, we have a finite database of known pairs with limited accuracy, and we want to satisfy (1), at least approximately, for the pairs in the data base. To make best use of current optimization technology, it is desirable to have a smooth (i.e., twice continuously differentiable with respect to x) potential. This allows for robust local optimization (e.g., [5, 25]), and can be combined with global search techniques such as simulated annealing (e.g., [15, 32]), genetic algorithms (e.g., [10, 31]), smoothing methods (e.g., [21, 16]) or branch and bound techniques (e.g., [20]) to approach the global minimizer for sequences s with unknown native geometry.

Unfortunately, the approach of determining empirical potentials from equilibrium data is intrinsically limited, even if we assume complete knowledge of all equilibrium geometries and their energies. It is obvious that statistical potentials cannot define an energy scale, since multiplication of a potential by a positive, constant factor does not alter its global minimizers. But for the purpose of tertiary structure prediction by global optimization, this does not not matter.

A more serious limitation comes from a nonuniqueness theorem (stated and proved in NEUMAIER [23]) that shows that the set of local and global minimizers and stationary points of any family of potentials $V(s,x)$ can be reproduced by an infinite family of other potentials. Thus, no set of equilibrium geometries can determine the true effective potential energy function. This complements the findings of THOMAS & DILL [33] (through a simulation study) and BRYNGELSON [4] (by theoretical arguments) that statistical potentials do not quantitatively reflect the true energies. In particular, empirical potentials solely derived from databases of equilibrium data will probably never be useful for dynamical studies. Moreover, the "energies" computed by an empirical potential may have little to do with real energies, as has been recently pointed out by BEN-NAIM [2].

While this is disappointing, the nonuniqueness theorem also shows that if *some* empirical potential is able to predict correct protein folds then many other empirical potentials will do so, too. Thus, the construction of empirical potentials for fold prediction is much less constrained than one might think initially, and one is justified in using additional qualitative theoretical assumptions in the derivation of an appropriate empirical potential function.

Our potential,

$$V(s,x) = \sum_a z_a V_a(s,x) + \sum_\gamma z_\gamma V_\gamma(s,x), \tag{2}$$

is a weighted sum of smooth *surface potentials* $V_a(s,x)$ that model the total contribution of the interaction of all amino acids with label a with the solvent, and of smooth *pair potentials* $V_\gamma(s,x)$ that model the total contribution of the interactions between all pairs of amino acids classified to be of the same class γ. The constants z_a and z_γ are positive weights scaling these contributions in a manner adapted to the available data.

(While smooth pair potentials are the rule in the literature, surface terms have traditionally been discontinuous; the only potential using smooth surface terms seems to appear in LUND et al. [19], where the surface term is a function of a smooth approximation to the number of neighbors of a C_α atom.)

The class of a pair of amino acids in positions i, k of a sequence s depends on the labels s_i and s_k of the amino acids and the *residue distance* $i-k$, and is specified through a suitably constructed class table, and

$$V_\gamma(s,x) = \sum_{(s_i,s_k,i-k) \text{ of class } \gamma} U_\gamma(r_{ik}),$$

where $r_{ik} = \|x_i - x_k\|$ is the Euclidean distance between the positions x_i and x_k of two C_α atoms, and the *pair potentials* $U_\gamma(r)$ are smooth functions of the distance. Instead of the more customary Lennard-Jones like potentials we use pair potentials that are constant for $r \geq r_{\max}$, with a cutoff at $r_{\max}$ at 12 Ångström, since amino acids at most three residues apart are at distance < 12Å in known proteins. All pair potentials are normalized such that their global minimum has the value 0.

To make the pair potentials flexible but fast to evaluate we chose the $U_\gamma(r_{ik}) = W_\gamma(q_{ik})$ as low degree polynomials in

$$q_{ik} = \left(\frac{\max(c_{\max} - c_{ik}, 0)}{2c_{\max} - c_{ik}}\right)^3,$$

with vanishing constant coefficient, where $c_{\max} = \frac{1}{2}r_{\max}^2$ and

$$c_{ik} = \frac{1}{2}\|x_i - x_k\|^2 = \frac{1}{2}(\langle x_i, x_i\rangle + \langle x_k, x_k\rangle) - \langle x_i, x_k\rangle$$

is the *half squared distance* between two C_α atoms. q_{ik} is bounded in [0,0.125] and smooth, has compact support $[0, r_{\max}]$, and can be evaluated cheaply from inner products, without taking square roots. The coefficients of the polynomials $W_\gamma(q)$ are found adaptively to fit empirical potentials derived from the data; see below.

For similar reasons we chose the surface potentials to be

$$V_a(s, x) = \sum_{s_i = a} W_a(q_i),$$

with low degree polynomials $W_a(q_i)$ in some expression q_i that is larger on the surface than in the interior. Since q_{ik} is large only when the ith and the kth residue are close, and since there are fewer residues close to a residue at the surface, the expression

$$q_i = \frac{1.41}{\sum_{k \neq i} q_{ik}} - \frac{2.65}{n^{1/3}}$$

was used, which has this property. The constants were chosen such that q_i has approximately mean 1 and is reasonably independent of the protein size n. q_i (and hence each surface potential) is a smooth function of the coordinates, except when some position x_k $(k \neq i)$ coincides with x_i; then q_i is infinite. Thus, by enforcing $W_a(q) \to \infty$ as $q \to \infty$, the surface potentials also serve the task to keep the amino acids apart.

Given the pair and surface potentials, the weights are then constructed by solving the convex bound constrained quadratic program

$$\begin{aligned} \min \quad & \sum_{\text{proteins } s} \|\nabla V(s, x_s)\|^2 \\ \text{s.t.} \quad & \text{all } z_a \geq 1, \quad \text{all } z_\gamma \geq 1. \end{aligned}$$

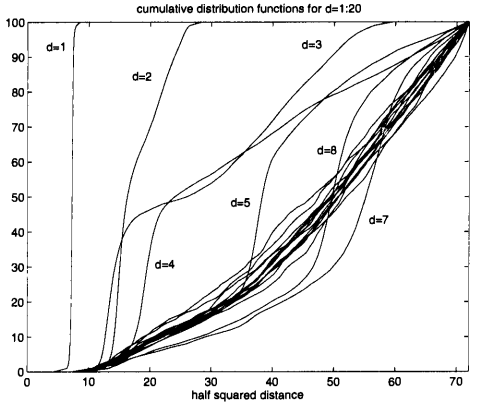


Fig. 1. Cumulative distribution function of half squared distances of amino acid pairs at residue distance d ($d = 1, \ldots, 20$), truncated at $c_{\max} = 72$ (12Åcutoff)

The objective function is a nonnegative, convex quadratic in z. It should vanish exactly for an ideal potential; hence minimizing the objective can be expected to give a good approximation to the best potential. The constraints are inspired by the independence assumption of SIPPL [28], which amounts to the particular choice

$$\text{all } z_a = 1, \quad \text{all } z_\gamma = 1;$$

our constraint relaxes this assumption in a natural way. It turned out that in the solution, $z_a > 1$ for 9 of the 20 amino acids, and $z_\gamma > 1$ for 15 of the 305 pair classes γ used. (For quadratic programming in general, see, e.g., [5].)

3 Density Estimation

The polynomials $W_a(q)$ ($q = q_{ik}$) and $W_\gamma(q)$ ($q = q_i$) needed to specify the pair and surface potentials are constructed from the set of such q realized in a data base of 266 proteins with a total of 46100 residues by means of density estimation techniques.

Boltzmann's classical formula

$$\langle f(q) \rangle = \frac{1}{Z} \int e^{-\beta W(q)} f(q) dq$$

for the expectation of functions of a random variable q can be used to justify the use of some multiple $W(q)$ of the negative logarithm of the density $\rho(q) = Z^{-1} e^{-\beta W(q)}$ of q as a useful definition of a potential contribution involving any interesting function q of the coordinates. (For the case when q is an atomic distance, the pros and cons of this recipe are discussed in more detail by SIPPL [28].)

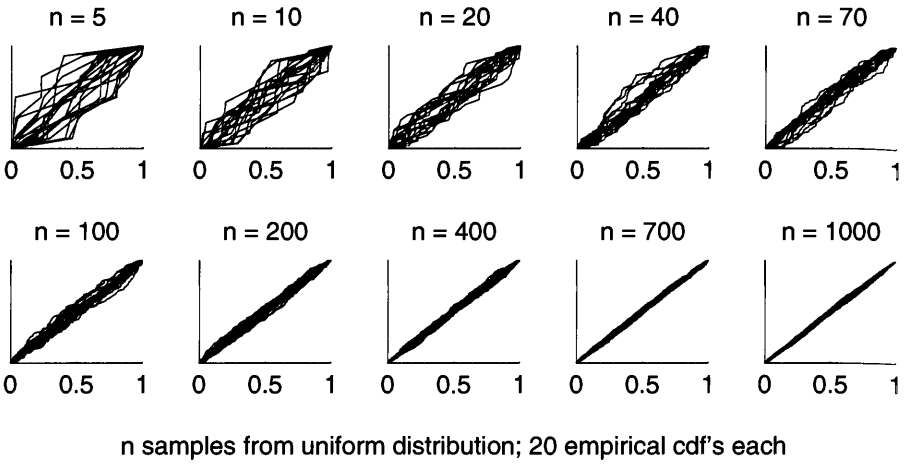


Fig. 2. Empirical cdf's for many samples of increasing size from a uniform distribution

For the robust estimation of the pair potentials, some obstacles had to be overcome. There are a huge number of different triples $(s_i, s_k, i-k)$, and to find densities, we needed a way to group them in a natural way together into suitable classes. A look at the cumulative distribution functions (cdf's) of the half squared distances c_{ik} at residue distance $d = i - k$ (w.l.o.g. > 0), displayed in Figure 1, shows that the residue distances 8 and higher behave very similarly; so in a first step we truncated all residue distances larger than 8 to 8.

This left $20 \times 20 \times 8 = 3200$ classes, with some classes being very sparsely populated. For such classes, the error term $O(n^{-1/2})$ is unacceptably large, and density estimators are intrinsically unstable under variations of small samples. A Monte Carlo test with samples from a uniform distribution displayed in Figure 2 shows that a sample size of at least about 100 is needed to reproduce a cdf and hence a density with a reasonable accuracy.

Therefore it was essential to have an additional mechanism that groups together data for 'similar' amino acid pairs. We performed this task by means of a weighted mean square cluster analysis. Our clustering procedure used linearly interpolated empirical cdf's of the initial classes together with a statistical estimate of their accuracy. We then repeatedly joined the smallest remaining class to the class with the most compatible cdf and recomputed the cdf of the merged class until all classes contained at least 100 sample points. Additional correction phases allowed outliers in some class to migrate to a more suitable class. Applied to the 3200 initial classes, this produced a list of 305 classes with significantly different distributions. (Different clustering procedures are used in some recent residue potential constructions: ULRICH et al. [34] use clustering to group the 210 unordered amino acid

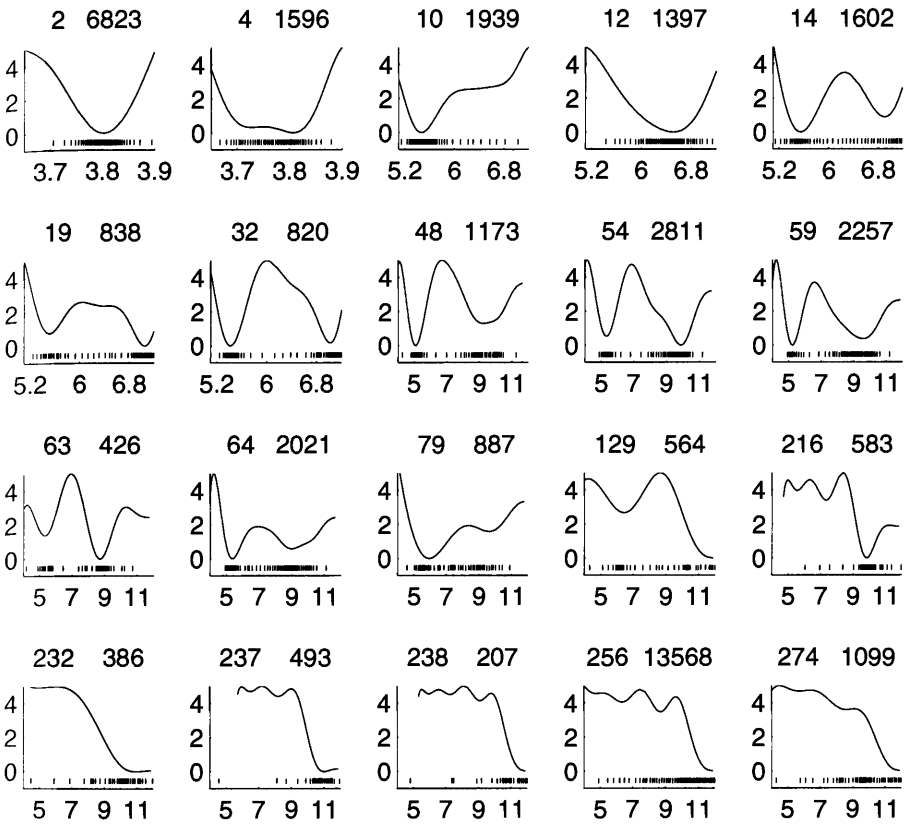


Fig. 3. Some representative pair potentials $U_\gamma(r)$, scaled to move their interesting range to [0,5]. The numbers above each potential denote the class label γ and the number of data points available for the fit. (For example, class 63 gives distance 3 potentials for the amino acid pairs Lys-Asp, Arg-Lys and Glu-Tyr.) The spectrum below each potential consists of 50 lines picked uniformly from the data.

pairs of residue distance 3 into 36 classes based on some energy measure, and HUBER & TORDA [13] apply clustering to estimated parameters.)

To compute densities for such a large number of distributions, reliable and fully automatic density estimators are necessary. The only density estimators discussed in the protein literature are histogram estimates. However, these are nonsmooth and thus not suitable for global optimization techniques that combine local and global search. Moreover, for a sample of size n and an optimally chosen bin size, histogram estimates have an accuracy of $O(n^{-1/10})$. This is an extremely poor accuracy, far away from the theoretically attainable accuracy $O(n^{-1/2})$ of other density estimators. (To reach $n^{-1/10} = 0.1$ one needs $n = 10^{10}$, while $n^{-1/2} = 0.1$ holds already for $n = 100$.)

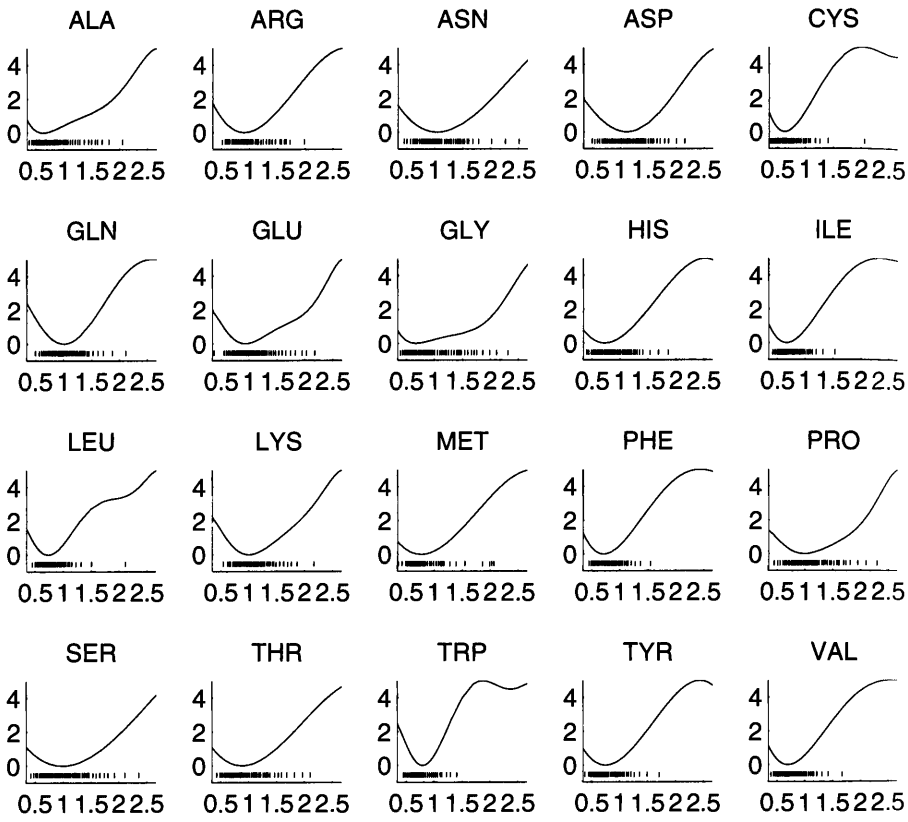


Fig. 4. Surface potentials $W(q)$

One therefore needs a smooth density estimation techniques that is more reliable than the histogram estimates. The automatic estimation poses additional problems in that the traditional statistical techniques for estimating densities usually require the interactive selection of some smoothing parameter (such as the bin size). Some publicly available density estimators are available[1], but these tended to oversmooth the densities. So we tried a number of ideas based on numerical differentiation of the empirical cdf to devise a better density estimator.

The best recipe we found so far is based on the statistical model

$$W(q_l) \approx w_l := \log(q_{l+1} - q_{l-1}) + \log(n/1.5),$$

where q_l is the lth sample point in increasing order. $W(q)$ is estimated from this relation by polynomial regression, and the Bayes criterion of SCHWARZ [26] is used to select the correct polynomial degree.

[1] see, e.g., `http://solon.cma.univie.ac.at/~neum/stat.html#density`

We tested our recipe on many trial densities by Monte Carlo simulation, e.g., on the normal mixture target densities of JONES et al. [14]. Examples of pair potentials $U_\gamma(r) = W_\gamma(q(r))$ reconstructed in this way are given in Figure 3.

We believe that this recipe gives a reasonably reliable density estimator. But is not yet perfect in that it suffers occasionally from picking an unsuitable polynomial order for the potential, and it tends to give small spurious oscillations at small densities when the original density is not of the form $e^{-W(q)}$ for some low degree polynomial $W(q)$. Unfortunately, these translate into significant oscillations for large values of the corresponding potentials; see Fig. 3 for classes 216, 237, 238, 256. (We are working on replacing polynomials by splines that should suppress these oscillations.)

For the surface potentials, sufficiently many data were available, and no further problems appeared. The resulting potentials are shown in Figure 4. The hydrophobic amino acids are easily recognized as those for which the potential well is at small values of q.

4 Results and Future Work

We tested our new potential by applying a local optimization procedure to the potential of some proteins, starting with the native structure as given in the Brookhaven Protein Data Bank, and observing how far the coordinates moved through local optimization. For a good potential, one expects the optimizer to be close to the native structure. As in ULRICH et al. [34], we measure the distance between optimizer B and native structure A by the distance matrix error

$$DME = \sqrt{\binom{n}{2}^{-1} \sum_{i<k} (r_{ik}^B - r_{ik}^A)^2};$$

it is usually a little larger than the root mean square (RMS) error that is based on optimal superposition.

The results of the optimization for 9 small test proteins, both for the potential with constant weights 1 and with the optimized weights, are given in Table 1. The optimized weights lead to smaller errors; the resulting potentials have minima within 1.3–4.7Å of the PDB geometry, with one exception that has an error of 8.5Å.

At present, the data base used for the fit was not specially selected to avoid homologous proteins. Thus, a further improvement can be expected from using data for one of the specially prepared lists of PDB files (cf. HOBOHM et al. [9]). We also expect further improvements from replacing the polynomial fits in the potential estimation procedure by piecewise cubic fits; though at the moment it is not clear how to select the number of nodes needed to get a good but not overfitting approximation to the density. Finally, we are considering

Table 1. Distance matrix errors DME (in Å) between optimizers and native structures

PDB code	$z = 1$	optimized z
1cti	2.2	4.7
1gcn	0.6	1.3
1mhu	3.4	3.0
1mrb	3.2	2.4
1mrt	3.3	2.3
2eti	11.8	8.5
2mhu	2.3	2.5
2mrt	4.0	2.4
3znf	5.5	2.3

adding chirality terms to further enhance the quality of our potential. More extensive testing, e.g., using the threading test of HENDLICH et al. [8] will be done after all these enhancements have been made.

References

1. A. Bauer and A. Beyer, An improved pair potential to recognize native protein folds, Proteins: Struct. Funct. Gen. 18 (1994), 254-261.
2. A. Ben-Naim, Statistical potentials extracted from protein structures: Are these meaningful potentials? J. Chem Phys. 107 (1997), 3698-3706.
3. F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, E. Meyer, M.D. Bryce, J.R. Rogers, O. Kennard, T. Shikanouchi and M. Tasumi, The protein data bank: A computer-based archival file for macromolecular structures, J. Mol. Biol. 112 (1977), 535-542.
4. J.D. Bryngelson, When is a potential accurate enough for structure prediction? Theory and application to a random heteropolymer model of protein folding, J. Chem. Phys. 100 (1994), 6038–6045.
5. P.E. Gill, W. Murray and M.H. Wright, Practical optimization, Acad. Press, London 1981.
6. J. R. Gunn, A. Monge, R.A. Friesner and C.H. Marshall, Hierarchical algorithm for computer modeling of protein tertiary structure: folding of myoglobin to 6.2A resolution, J. Phys. Chem. 98 (1994), 702-711.
7. P. Hall and J.S. Marron, Lower bounds for bandwidth selection in density estimation, Probab. Th. Rel. Fields 90 (1991), 149-173.
8. M. Hendlich, P. Lackner, S. Weitckus, H. Floeckner, R. Froschauer, K. Gottsbacher, G. Casari and M. J. Sippl, Identification of native protein folds amongst a large number of incorrect models, J. Mol. Biol. 216 (1990), 167-180.
9. U. Hobohm, M. Scharf, R. Schneider and C. Sander, Selection of representative protein data sets, Protein Sci. 1 (1992), 409-417.
10. J. Holland, Genetic algorithms and the optimal allocation of trials, SIAM J. Computing 2 (1973), 88-105.

11. L. Holm and C. Sander, Database algorithm for generating protein backbone and side-chain co-ordinates from a C^{α} trace, J. Mol. Biol. 218 (1991), 183-194.
12. L. Holm and C. Sander, Fast and simple Monte Carlo algorithm for side chain optimization in proteins. Proteins 14 (1992), 213-223.
13. T. Huber and A.E. Torda, Protein fold recognition without Boltzmann statistics or explicit physical basis, submitted to Protein Sci. (1997).
14. M.C. Jones, J.S. Marron and S.J. Sheather, Progress in data-based bandwidth selection for kernel density estimation, Comput. Statist. 11 (1996), 337-381.
15. S. Kirkpatrick, C.D. Geddat, Jr., and M.P. Vecchi, Optimization by simulated annealing, Science 220 (1983), 671-680.
16. J. Kostrowicki and H.A. Scheraga, Application of the diffusion equation method for global optimization to oligopeptides, J. Phys. Chem. 96 (1992), 7442-7449.
17. M. Levitt, A simplified representation of protein confomations for rapid simulation of protein folding, J. Mol. Biol. 104 (1976), 59-107.
18. M. Levitt and A. Warshel, Computer simulation of protein folding, Nature 253 (1975), 694-698.
19. O. Lund, J. Hansen, S. Brunak and J. Bohr, Relationship between protein structure and geometrical constraints, Protein Sci. 5 (1996), 2217-2225.
20. C.D. Maranas, I.P. Androulakis and C.A. Floudas, A deterministic global optimization approach for the protein folding problem, pp. 133-150 in: Global minimization of nonconvex energy functions: molecular conformation and protein folding (P. M. Pardalos et al., eds.), Amer. Math. Soc., Providence, RI, 1996.
21. J.J. Moré and Z. Wu, Global continuation for distance geometry problems, SIAM J. Optimization 7 (1997), 814-836.
22. A. Neumaier, Molecular modeling of proteins and mathematical prediction of protein structure, SIAM Rev. 39 (1997), 407-460.
23. A. Neumaier, A nonuniqueness theorem for empirical protein potentials, in preparation.
24. M. Oobatake and G.M. Crippen, Residue-residue potential function for conformational analysis of proteins, J.Phys. Chem. 85 (1981), 1187-1197.
25. T. Schlick and A. Fogelson, TNPACK – A truncated Newton minimization package for large scale problems, ACM Trans. Math. Softw. 18 (1992), 46-70; 71-111.
26. G. Schwarz, Estimating the dimension of a model, Ann. Statistics 6 (1978), 461–464.
27. D. Shortle, Y. Wang, J. Gillespie and J.O. Wrabl, Protein folding for realists: a timeless phenomenon, Prot. Sci. 5 (1996), 991-1000.
28. M.J. Sippl, Boltzmann's principle, knowledge based mean fields and protein folding, J. Comp. Aided Mol. Design 7 (1993), 473-501.
29. M.J. Sippl, M. Hendlich and P. Lackner, Assembly of polypeptide and protein backbone conformations from low energy ensembles of short fragments, Protein Sci. 1 (1992), 625-640.
30. D.R. Stampf, C.E. Felser and J.L. Sussman, PDBBrowse – a graphics interface to the Brookhaven Protein Data Bank, Nature 374 (1995), 572-574.
31. S. Sun, Reduced representation model of protein structure prediction: statistical potential and genetic algorithms, Protein Sci. 2 (1993), 762-785.
32. S. Sun, Reduced representation approach to protein tertiary structure prediction: statistical potential and simulated annealing, J. Theor. Biol. 172 (1995), 13-32.

33. P.D. Thomas and K.A. Dill, Statistical potentials extracted from protein structures: How accurate are they? J. Mol. Biol. 257 (1996), 457-469.
34. P. Ulrich, W. Scott, W.F. van Gunsteren and A. Torda, Protein structure prediction force fields: parametrization with quasi Newtonian dynamics, Proteins 27 (1997), 367-384.
35. L.L. Walsh, Navigating the Brookhaven Protein Data Bank, Cabos Communication 10 (1994), 551-557.

Part III

Enhanced Time-Stepping Algorithms

Some Failures and Successes of Long-Timestep Approaches to Biomolecular Simulations

Tamar Schlick

Department of Chemistry and Courant Institute of Mathematical Sciences, New York University and The Howard Hughes Medical Institute, 251 Mercer Street, New York, NY 10012, U.S.A.

Abstract. A personal account of work on long-timestep integration of biomolecular dynamics is presented, emphasizing the limitations, as well as success, of various approaches. These approaches include implicit discretization, separation into harmonic and anharmonic motion, and force splitting; some of these techniques are combined with stochastic dynamics. A Langevin/force-splitting approach for biomolecular simulations termed LN (for its origin in a Langevin/normal-modes scheme) is also described, suitable for general thermodynamic and sampling questions. LN combines force linearization, stochastic dynamics, and force splitting via extrapolation so that the timestep for updating the slow forces can be increased beyond half the period of the fast motions (i.e., 5 fs). This combination of strategies alleviates the severe stability restriction due to resonance artifacts that apply to symplectic force-splitting methods and can yield significant speedup (with respect to small-timestep reference Langevin trajectories). Extensions to sampling problems are natural by this approach.

1 Introduction

Looking ahead into the future bimolecular simulations, it is difficult to pinpoint the hurdles that will remain in 30 to 50 years. Will the quality of the molecular mechanics force fields still limit the accuracy of answers that can be attained regarding macromolecular structure and kinetics? Will algorithms aimed at capturing all thermally-accessible states remain a hurdle? Certainly, the enormous range of spatial and temporal scales associated with biologically interesting motions of macromolecules is daunting to computationalists today. However, the effects of the technical improvements in computer architecture, speed, processor communication, and parallel programming on the sampling problem cannot be predicted.

The second international symposium on Algorithms for Macromolecular Modeling (Konrad Zuse-Zentrum, Berlin, May 21-24, 1997) brought together many theoreticians to ponder at these questions. Covering the most exciting developments in structure determination, timestepping algorithms in molecular dynamics (MD), free energy methods, quantum/classical dynamics, and other practical simulation topics, the presenters triggered many lively though not always congenial — discussions. Many of these advances are

presented in this volume. In this contribution, we offer a historical perspective of our group's work on molecular dynamics schemes, experiences of which provide useful insights into the limitations of long-timestep integration methods. These methods lead to the competitive force-linearization/force-splitting Langevin algorithm LN. This is possible by combining a stochastic formulation with force splitting by extrapolation (rather than impulses) to alleviate stability problems at large outer timesteps. The Langevin trajectories differ from Newtonian trajectories, but are suitable for certain thermodynamic and conformational problems relating to macromolecules. The stochastic coupling parameter can also be made small (i.e., just sufficient for numerical stability) to make trajectories closer to Newtonian.

2 Long-Timestep Approaches

The difficulty in simulating long-time processes of biomolecular systems modeled atomistically using standard force fields is well recognized: the timescales associated with molecular motion span an enormous range. Thus, while the fastest, high-frequency modes dictate a timestep of 1 fs or less in standard explicit schemes for acceptable resolution, this value is short by more than ten orders of magnitude than the slow and large-amplitude processes of major biological interest. Given the high cost of each iteration — which requires a force evaluation for a large system, a calculation dominated by long-range nonbonded interactions — this timestep restriction severely limits the total time that can be simulated for a large polymer. In addition, the continuous range of the biomolecular vibrational spectrum and the strong coupling among modes rule out general methods that work well on separable problems. See [1], for example, for an introduction into this timescale dilemma, and [2, 3, 4, 5] for reviews of the associated numerical problem.

The chaotic nature of individual MD trajectories has been well appreciated. A small change in initial conditions (e.g., a fraction of an Ångstrom difference in Cartesian coordinates) can lead to exponentially-diverging trajectories in a relatively short time. The larger the initial difference and/or the timestep, the more rapid this Lyapunov instability. Fig. 1 reports observed behavior for the dynamics of a butane molecule. The governing Newtonian model is the following set of two first-order differential equations:

$$\begin{aligned} \mathbf{M}\dot{V}(t) &= -\nabla E(X(t))\,, \\ \dot{X}(t) &= V(t), \end{aligned} \tag{1}$$

where X and V are the collective position and velocity vectors, respectively; $\mathbf{M}$ is the diagonal mass matrix, $\nabla E(X(t))$ is the gradient vector of the potential energy E, and the dot superscripts denote differentiation with respect to time, t.

The top part of Fig. 1 shows the time evolution of the central dihedral angle of butane, τ (defined by the four carbon atoms), for trajectories

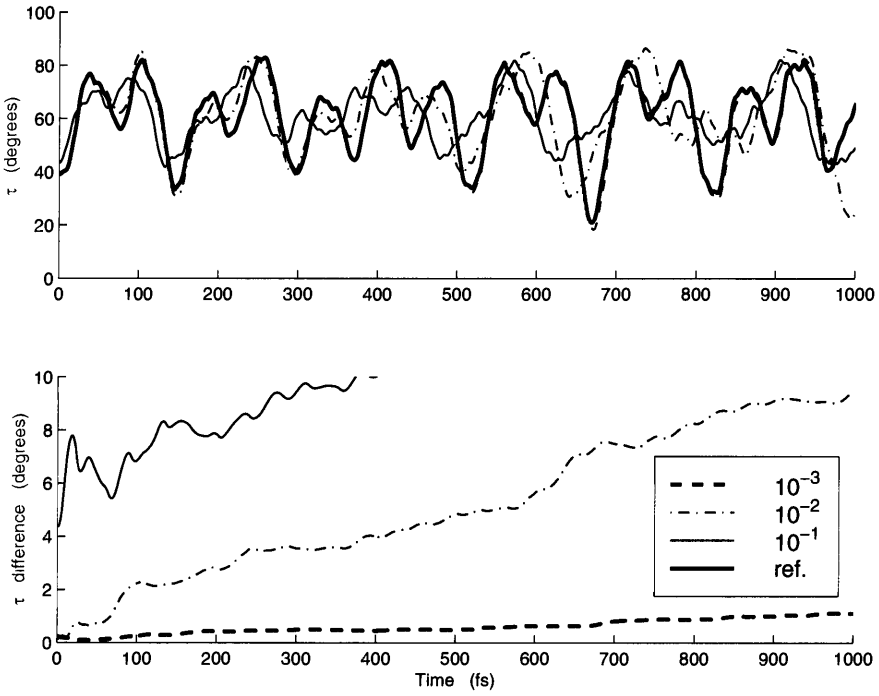


Fig. 1. The time evolution (top) and average cumulative difference (bottom) associated with the central dihedral angle of butane τ (defined by the four carbon atoms), for trajectories differing initially in 10^{-3}, 10^{-2}, and 10^{-1} Ångstoms of the Cartesian coordinates from a reference trajectory. The leap-frog/Verlet scheme at the timestep $\Delta t = 1$ fs is used in all cases, with an all-atom model comprised of bond-stretch, bond-angle, dihedral-angle, van der Waals, and electrostatic components, as specified by the AMBER force field within the INSIGHT/Discover program.

differing initially in a fraction (10^{-3}, 10^{-2}, 10^{-1}) of an Ångstom (in the Cartesian coordinates) for $\Delta t = 1$ fs. The plotted average cumulative error $\sum_t |\tau(t) - \tau'(t)| / t$ for these trajectories (bottom) shows divergence of these trajectories from the reference trajectory and from one another. However, the divergence seen here is not exponential but rather linear in appearance, most likely because of the compactness (boundedness) of biomolecular systems.

For many reasons, including the approximate nature of the governing force fields, MD simulations are not intended to animate the life of a biomolecule faithfully, as if by an unbiased video camera; rather, our experimental-by-nature computer 'snapshots' aim at predicting meaningful statistical properties of a complex system in the ultimate goal of relating structure to function. Both conformational sampling algorithms (such as Monte Carlo) and dynamic simulations can be used to generate such molecular ensembles for analysis. While sampling methods are much cheaper computationally, only

the dynamic approach offers continuous pathway information within the limitations described above. Unlike applications in astrophysics, where precise trajectory pathways are sought (e.g., planetary orbits), biomolecular trajectories are generated more in the spirit of 'numerical statistical mechanics', where ensemble molecular averages are replaced by temporal averages of a single system.

To increase the reliability of such simulations of inherently-chaotic systems, special care is needed to formulate efficient numerical procedures for generating dynamic trajectories. This reliability, or accuracy in a loose sense, must be measured with respect to the precise simulation questions posed. Mathematically, there are classes of methods for conservative Hamiltonian systems termed *symplectic* that possess favorable numerical properties in theory and practice [6]. Essentially, these schemes preserve volumes in phase space (as measured rigorously by the Jacobian of the transformation from one set of coordinates and momenta to the next). This preservation in turn implies certain physical invariants for the system. Another view of symplectic integration is that the computed trajectory remains close in time to the solution of a nearby Hamiltonian $\tilde{H}$ (i.e., one which is order $O(\Delta t)^p$ away from the initial value of the true Hamiltonian $H \equiv \frac{1}{2}(V^0)^T \mathbf{M}(V^0) + E(X^0)$, where p is the order of the integrator). This property translates to good long-time behavior in practice: small fluctuations about the initial (conserved in theory) value of H, and no systematic drift in energy, as might be realized by a nonsymplectic method. Fig. 2 illustrates this behavior for a simple nonlinear system, a cluster of four water molecules, integrated by Verlet and by the classical fourth-order Runge-Kutta method. A clear damping trend is seen by the latter, nonsymplectic integrator, especially at the larger timestep.

Besides compatibility of resolution for the chosen combination algorithm and timestep with the answers sought, we also seek to minimize the required computational time. Thus, practical considerations demand that the frequency of calculations associated with the most expensive part of the Newtonian forces be as small as possible. We associate this frequency with the 'outermost timestep' of the method. This desire to lengthen the outermost timestep has led to a variety of protocols for saving computer time: from constraining the fastest degrees of freedom via algebraic constraints [7, 3, 16, 10], to updating the nonbonded pair list infrequently, and formulating multiple-timestepping (MTS) schemes [11] that have become very successful in biomolecular dynamics [12, 13, 14, 15, 16, 17].

Accuracy, however, in biomolecular trajectories, must be defined somewhat subjectively. In the absence of exact reference data (from experiment or from an analytical solution), the convention has been to measure 'accuracy' with respect to reference trajectories by a Verlet-like integrator [18, 19] at a timestep of 1 or 0.5 fs (about one tenth or one twentieth the period, respectively, of the fastest period: an O–H or N–H stretch). As pointed out by Deuflhard *et al.* [20], these values are still larger than those needed to

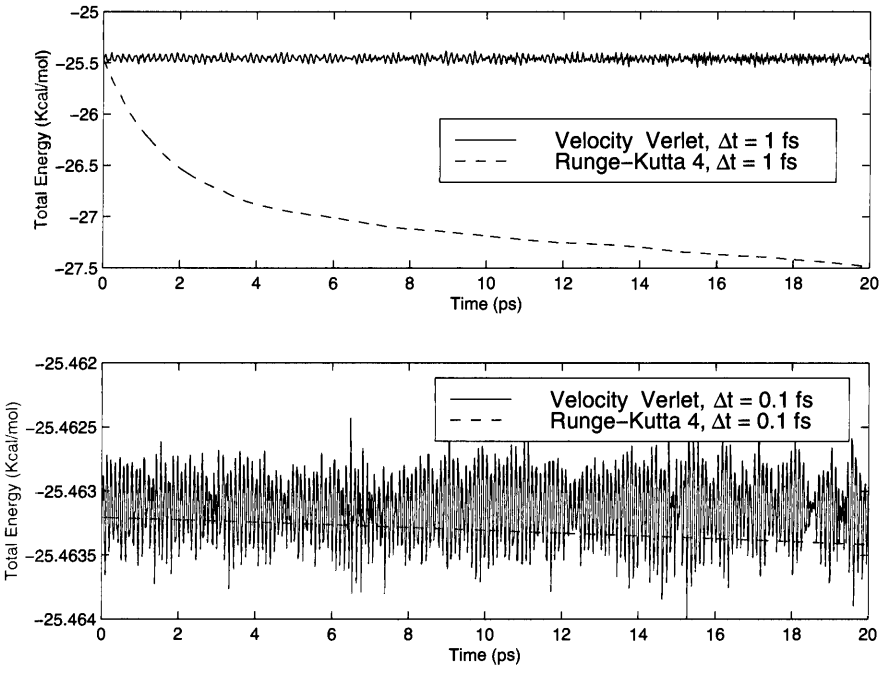


Fig. 2. The time evolution of the total energy of four water molecules (potential-energy details are given in [48]) as propagated by the symplectic Verlet method (solid) and the nonsymplectic fourth-order Runge-Kutta method (dashed pattern) for Newtonian dynamics at two timestep values.

resolve well the motion of the fastest degrees of freedom and lead to reliable values of slowly-converging quantities.

Shown in Fig. 3 is the time evolution of the end-to-end-distance of the butane molecule for different choices of timesteps (0.02, 0.2, 1.0 and 2.0 fs). Clearly, different paths are realized. Indeed, the obtained average end-to-end distance as a function of the simulation timestep in Fig. 4, inspired by [20], shows convergence only for small timesteps, less than 0.2 fs. In fact, our experiments with standard ordinary-differential-equation packages for MD integration reveal rejection of the 1 fs timestep on the basis of error criteria. The data shown in Fig. 4 are taken from single-timestep, as well as multiple-timestep (MTS), Verlet simulations. In the latter, three timesteps are used: $\Delta\tau$, $2\Delta\tau$, and $4\Delta\tau$, with the outermost timestep considered Δt for the figure. Thus the triple-timestep data point shown for $\Delta t = 2$ fs has an inner timestep of 0.5 fs and a medium timestep of 1 fs. Bond-length and angle terms are updated every $\Delta\tau$; dihedral-angle terms are updated every $2\Delta\tau$; and nonbonded interactions are recalculated every $4\Delta\tau$. As seen in the figure, accuracy is determined by the innermost timestep. This is the computational

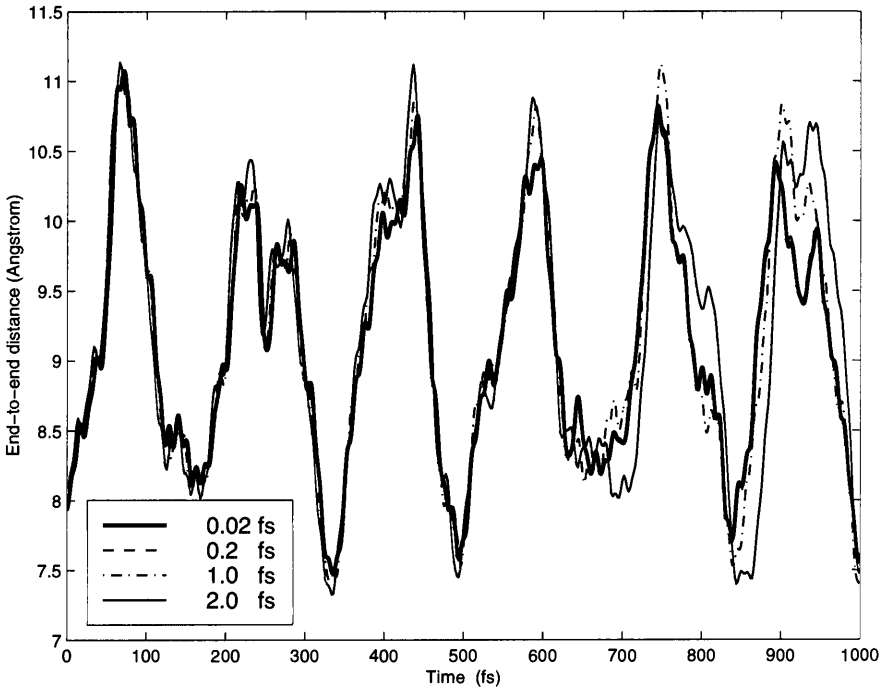


Fig. 3. The time evolution of the end-to-end distance of butane for different timesteps (see Fig. 1 caption).

advantage exploited in MTS schemes [16, 17] since the work for a large system is dominated by the outer timestep. For the small butane system, the cost of the MTS simulation with outer timestep Δt is about double that of the single-timestep simulation at Δt.

This discussion suggests that even the 'reference trajectories' used by symplectic integrators such as Verlet may not be sufficiently accurate in this more rigorous sense. They are quite reasonable, however, if one requires, for example, that trajectories capture the spectral densities associated with the fastest motions in accord to the governing model [13, 15]. Furthermore, other approaches, including nonsymplectic integrators and trajectories based on stochastic differential equations, can also be suitable in this case when carefully formulated.

3 Stochastic Dynamics

An alternative framework to Newtonian dynamics, namely Langevin dynamics, can be used to mask mild instabilities of certain long-timestep approaches. The Langevin model is phenomenological [21] — adding friction and random

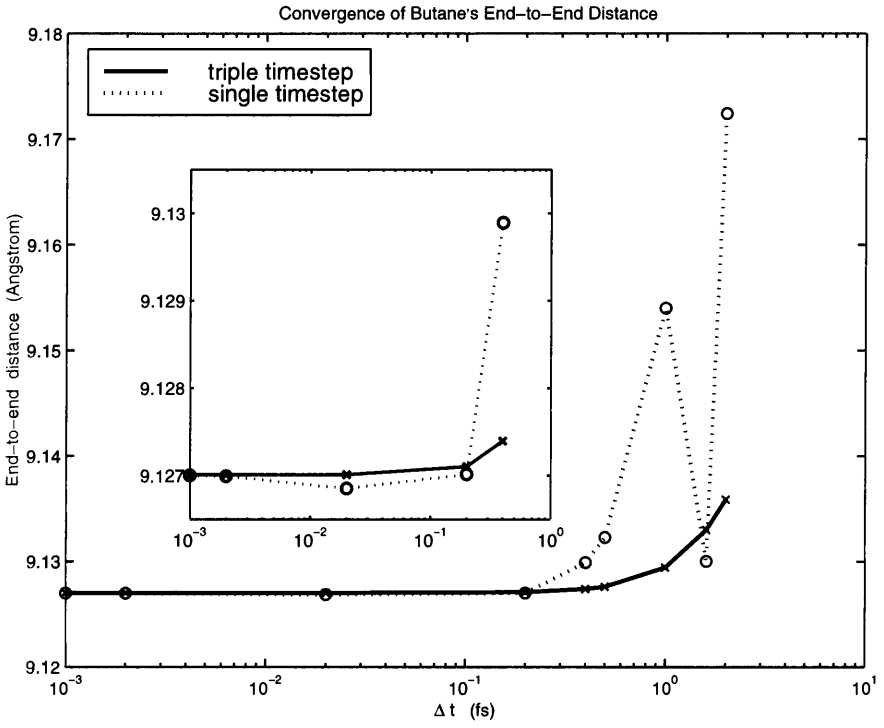


Fig. 4. The average end-to-end-distance of butane as a function of timestep (note logarithmic scale) for both single-timestep and triple-timestep Verlet schemes. The timestep used to define the data point for the latter is the outermost timestep Δt (the interval of updating the nonbonded forces), with the two smaller values used as $\Delta t/2$ and $\Delta t/4$ (for updating the dihedral-angle terms and the bond-length and angle terms, respectively).

forces to the systematic forces — but with the physical motivation to represent a simple heat bath for the macromolecule by accounting for molecular collisions. The continuous form of the simplest Langevin is given by:

$$\begin{aligned} \mathbf{M}\dot{V}(t) &= -\nabla E(X(t)) - \gamma \mathbf{M} V(t) + R(t), \\ \dot{X}(t) &= V(t), \end{aligned} \tag{2}$$

where γ is the collision parameter (in reciprocal units of time), or damping constant. The random-force vector R is a stationary Gaussian process with statistical properties given by:

$$\langle R(t) \rangle = 0, \qquad \langle R(t) R(t')^T \rangle = 2\gamma k_B T \mathbf{M}\, \delta(t - t'), \tag{3}$$

where k_B is the Boltzmann constant, T is the target temperature, and δ is the usual Dirac symbol. Fig. 5 illustrates the effects of increasing γ on the trajectories and phase diagrams of a harmonic oscillator.

The Langevin model has been employed extensively in the literature for various numerical and physical reasons. For example, the Langevin framework has been used to eliminate explicit representation of water molecules [22], treat droplet surface effects [23, 24], represent hydration shell models in large systems [25, 26, 27], or enhance sampling [28, 29, 30]. See Pastor's comprehensive review [22].

Enhanced sampling, in particular, is possible in some cases because it can be shown on the basis of classical Kramers transition-rate theory that a critical γ value can lead to faster convergence of equilibrium distributions (e.g., for key geometric and energetic properties). See [31] and [32] for recent applications of this idea to the slow macromolecular processes of DNA supercoiling and protein folding, respectively. To illustrate, the fluctuations of the writhing number of supercoiled DNA, a geometric characteristic of global shape, is shown in Fig. 6 as a function of γ over a large range. This fluctuation envelope exhibits a maximum at a certain range of γ which can be considered 'optimal' for sampling purposes. Equilibrium writhe distributions (as shown in Fig. 7) should in theory be the same for all γ.

In the large γ limit, we enter the Brownian, or diffusive regime. This physical range can also be used (with suitable algorithms, other than those described here) to explore configuration spaces of floppy systems efficiently. See [33, 34], for instance, for applications to the large-scale opening/closing lid motion of the enzyme triosephosphate isomerase (TIM), and to juxtaposition of linearly-distant segments in long DNA systems, respectively.

Since the stochastic Langevin force mimics collisions among solvent molecules and the biomolecule (the solute), the characteristic vibrational frequencies of a molecule in vacuum are dampened. In particular, the low-frequency vibrational modes are overdamped, and various correlation functions are smoothed (see Case [35] for a review and further references). The magnitude of such disturbances with respect to Newtonian behavior depends on γ, as can be seen from Fig. 8 showing computed spectral densities of the protein BPTI for three γ values. Overall, this effect can certainly alter the dynamics of a system, and it remains to study these consequences in connection with biomolecular dynamics.

A physical value for γ for each particle can be chosen according to Stokes law* (with stick boundary conditions):

$$\gamma = 6\pi\eta/m\,, \tag{4}$$

* An appropriate value of γ for a system modeled by the simple Langevin equation can also be determined so as to reproduce observed experimental translation diffusion constants, D_t; in the *diffusive limit*, D_t is related to γ by $D_t = k_B T/\sum m\gamma$. See [22, 36], for example.

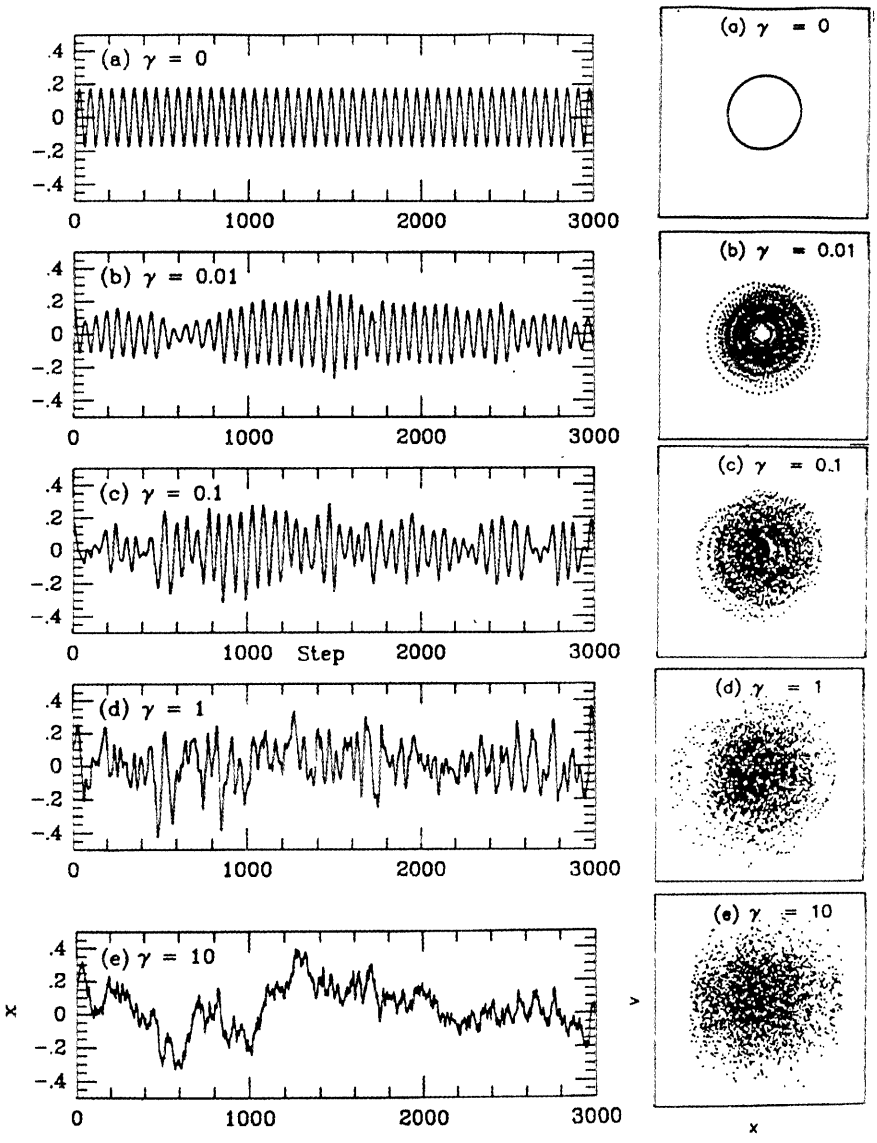


Fig. 5. Langevin trajectories for a harmonic oscillator of angular frequency $\omega = 1$ and unit mass simulated by a Verlet-like method (extended to Langevin dynamics) at a timestep of 0.1 (about 1/60 the period) for various γ. Shown for each γ are plots for position versus time and phase-space diagrams.

where m is the particle's mass and η is the solvent viscosity. For example, $\gamma = 50$ ps^{-1} is a typical collision frequency for protein atoms exposed to

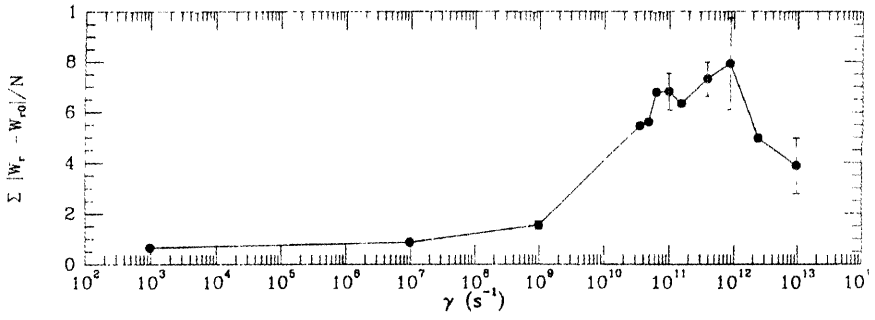


Fig. 6. The envelope of the writhing number of closed circular DNA subject to torsional stress as a function of γ, as computed from Langevin trajectories. Data are from [31].

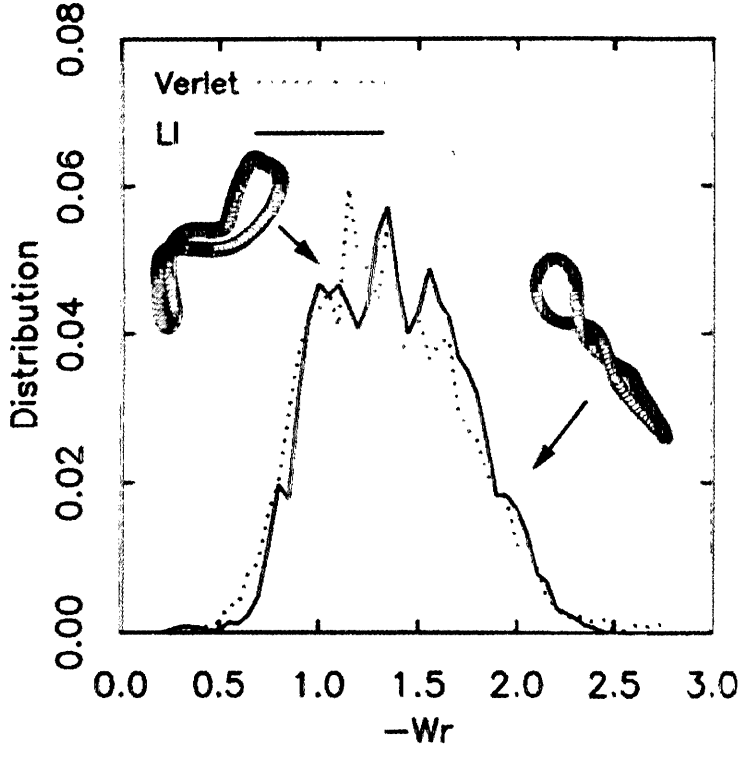


Fig. 7. Writhe distributions for closed circular DNA as obtained by LI (see Section 4.1) versus explicit integration of the Langevin equations. Data are from [36].

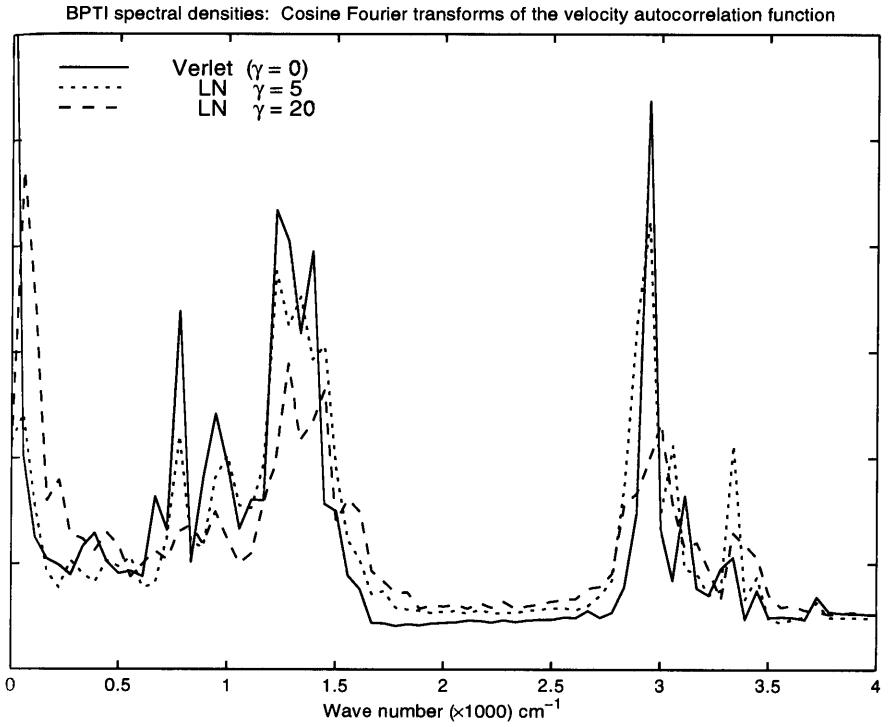


Fig. 8. Spectral densities for BPTI as computed by cosine Fourier transforms of the velocity autocorrelation function by Verlet ($\gamma = 0$) and LN ($\gamma = 5$ and 20 ps^{-1}). Data are from [88].

solvent having a viscosity of 1 cp at room temperature [37]; it is also in the range of the estimated value for water (γ= 54.9 ps^{-1}). A much smaller value, however, than such a physically-appropriate value, will lead to less physical damping and hence closer normal modes corresponding to the Newtonian system. This is our goal in setting γ for the LN method. The spectral densities of BPTI shown in Fig. 8 for nonzero γ were computed by the LN algorithm.

For future reference, the Verlet algorithm [18] can be generalized to include the friction and stochastic terms above, and is typically used in the following form described by Brooks, Brünger and Karplus, known as BBK [23, 37]:

Generalized Verlet Algorithm for Langevin Dynamics

$$
\begin{aligned}
V^{n+1/2} &= V^n + \mathbf{M}^{-1}\frac{\Delta t}{2}(-\nabla E(X^n) - \gamma \mathbf{M} V^n + R^n) \\
X^{n+1} &= X^n + \Delta t V^{n+1/2} \\
V^{n+1} &= V^{n+1/2} + \mathbf{M}^{-1}\frac{\Delta t}{2}(-\nabla E(X^{n+1}) - \gamma \mathbf{M} V^{n+1} + R^{n+1}).
\end{aligned} \tag{5}
$$

The third equation above is implicit for V^{n+1}, but the linear dependency allows solution for V^{n+1} in closed form. Here, the superscripts ^{n} refer to the difference-equation approximation to the solution at time $n\Delta t$. The superscript used for R is not important, as it is chosen independent at each step; when the Dirac delta function of eq. (3) is discretized, $\delta(t-t')$ is replaced by $\delta_{nm}/\Delta t$.

The popular Verlet method is recovered by setting γ and R^n to zero in the above propagation formulas.

4 Implicit Integrators and Resonance Artifacts

A reasonable approach for achieving long timesteps is to use *implicit* schemes [38]. These methods are designed specifically for problems with disparate timescales where explicit methods do not usually perform well, such as chemical reactions [39]. The integration formulas of implicit methods are designed to increase the range of stability for the difference equation. The experience with implicit methods in the context of biomolecular dynamics has not been extensive and rather disappointing (e.g., [40, 41]), for reasons discussed below.

4.1 Implicit Euler (IE)

The *implicit-Euler* (IE) scheme, for example, discretizes system (1) as:

$$
\begin{aligned}
\mathbf{M}(V^{n+1} - V^n)/\Delta t &= -\nabla E(X^{n+1}), \\
(X^{n+1} - X^n)/\Delta t &= V^{n+1}.
\end{aligned} \tag{6}
$$

The IE scheme is nonconservative, with the damping both frequency and timestep dependent [42, 43]. However, IE is unconditionally stable or *A-stable*, i.e., the stability domain of the model problem $y'(t) = q\,y(t)$, where q is a complex number (exact solution $y(t) = \exp(qt)$), is the set of all $\{q\Delta t\}$ satisfying $\mathrm{Re}\,(q\Delta t) \leq 0$, or the left-half of the complex plane. The discussion of IE here is only for future reference, since the application of the scheme is faulty for biomolecules.

Because of the inherent damping, the application of IE to biomolecular dynamics only makes sense in the context of a model with a restoring force.

Discretizing the Langevin equation (2,3) by IE produces the following system which implicitly, rather than explicitly, defines X^{n+1} in terms of quantities known from previous timesteps:

$$\begin{aligned} \mathbf{M}(V^{n+1} - V^n)/\Delta t &= -\nabla E(X^{n+1}) - \gamma \mathbf{M} V^{n+1} + R^{n+1}, \\ (X^{n+1} - X^n)/\Delta t &= V^{n+1}, \end{aligned} \tag{7}$$

where R is specified at each step by eq. (3).

We have shown that instead of solving a nonlinear system, the solution of X^{n+1} from system (7) can be obtained by minimizing the "dynamics function", $\Phi(X)$, where

$$\begin{aligned} \Phi(X) &= \frac{1}{2}(1 + \gamma \Delta t)\,(X - X_0^n)^T \mathbf{M}(X - X_0^n) + (\Delta t)^2 E(X), \\ X_0^n &= X^n + [\Delta t/(1 + \gamma \Delta t)]\,(V^n + \Delta t \mathbf{M}^{-1} R^{n+1}). \end{aligned} \tag{8}$$

Thus, each step of dynamics by the Langevin/Implicit-Euler scheme ("LI") consists of solving for X^{n+1} by minimizing $\Phi(X)$ and then computing V^{n+1} from the second equation of system (7). The starting candidate for minimization of Φ can be X^n, $X^n + \Delta t V^n$, or X_0^n, and efficient minimization can be accomplished using our truncated Newton package [44, 45, 46, 47], recently made available in the comprehensive molecular mechanics and dynamics package CHARMM. Note that the LI method behaves like a potential energy minimizer at each iteration when the timestep is large. This explains intuitively the stability of the method at large timesteps.

There are three issues of concern regarding the application of LI to biomolecular dynamics: (1) the governing Langevin model, (2) the implications of numerical damping, and (3) the CPU performance, given that nonlinear minimization is required at each, albeit longer, timestep.

The Langevin model and the differences between Newtonian and Langevin systems are general problems and were briefly mentioned in the previous section in both physical and numerical terms. Essentially, the Langevin framework is adequate for some conformational and sampling questions. The issues of numerical damping and CPU performance have turned to be disappointing for LI except for special cases. We noted that the random force is insufficient to counter the strong and frequency-dependent damping of the IE scheme. Since the high-frequency bond-stretch modes are coupled to lower-frequency motions (angle bending, dihedral rotation, etc.), which in turn trigger other deformations, damping these high frequencies alters molecular motion substantially. We have shown that LI tends to preserve 'local' structure (as measured by ice-like features of liquid water [48] and much slower decay of autocorrelation functions [49, 29]) and exhibit an effective lower temperature for the system. In addition to this damping problem, there is no net computational advantage to the scheme. Though the timestep can be set to much larger values than used by explicit schemes (e.g., $\Delta t = 40$ fs [29] or more),

we gain little overall from this timestep increase due to the high cost of the minimization.

Though LI failed for general biomolecular applications [50], it has been found to be a useful ingredient in two other contexts: macroscopic separable models, and enhanced sampling.

In macroscopic DNA applications, long double-stranded DNA can be modeled by B-spline [51, 52, 53] or Fourier [54, 55] curve-fitting techniques and subjected to elastic-based energy functions [56]. Since the biologically-relevant high-frequency modes are conveniently absent in these systems, computational gain (e.g., factor of 3) without significant difference in the results can be achieved with LI over explicit integrators (see Fig. 7) [36]; in this context, explicit formulations are limited for stability by a restraint term for the total contour length of the curve, a computational device.

The minimization-like behavior of LI in the long-timestep regime, combined with the kinetic-like first term of Φ, can be exploited for enhanced sampling purposes, a challenge at least as great as long-timescale MD simulations. Formulation of an efficient sampling method can be accomplished by adding appropriate perturbations to the minimum of the dynamics function. Such sampling methods based on LI have been developed independently by Scheraga and co-workers, to suggest unfolding protein pathways [30], and by Derreumaux, to generate insight into the folding/unfolding dynamics of model peptides and and proteins [29, 57].

The algorithm of Hao *et al.* [30] essentially minimizes $\Phi(X)$, eq. (8), at each step to obtain X^{n+1} and then adds to it a rescaled velocity term to maintain uniform spatial perturbations to the protein at each step. This procedure succeeds in unfolding and refolding the enzyme BPTI (bovine pancreatic trypsin inhibitor) and capturing a unique and reproducible pathway for a specific protein intermediate lacking one of the native-structure disulfide bonds. Interestingly, the researchers find that many expanded conformations can be refolded closely to the native structure and therefore suggest that these partially-folded forms would be favorable folding intermediates of BPTI.

In the "dynamics driver approach" (DA) [29], a vector $X_{DA} = X^n + P$ is generated at each step n as the initial guess for minimization of the dynamics function Φ. The perturbation vector P is chosen to have magnitude λ (e.g., $\lambda = 0.35$ Å) and three components set as:

$$P = [\lambda \ \sin\alpha\cos\beta, \ \lambda \ \sin\alpha\sin\beta, \ \lambda \ \cos\alpha] \,, \tag{9}$$

where the angles α and β are randomly chosen in the interval $[-\pi,\pi]$. In addition to specification of a perturbation vector P, each step of DA involves rescaling of the velocity vector V^n by a factor related to λ and application of two acceptance criteria in the spirit of Metropolis/Monte Carlo. These criteria involve both the energetic components associated with the high frequency modes and the total kinetic energy [57]. The DA method does not yield continuous dynamic trajectories but enhances sampling considerably, spending more time in energetically-favorable regions, as shown in [29]. Our recent

application of DA to the large-scale conformational change of the active site in TIM produced several pathways between the open and closed form of this enzyme for analysis [57]. On the basis of these energetic-pathway analyses, we have suggested that the timescale for transition in TIM is much slower than previously assumed, namely microsecond rather than nanosecond range.

4.2 Implicit Midpoint (IM)

The *implicit-midpoint* (IM) scheme differs from IE above in that it is symmetric and symplectic. It is also special in the sense that the transformation matrix for the model linear problem is unitary, partitioning kinetic and potential-energy components identically. Like IE, IM is also A-stable. IM is therefore a more reasonable candidate for integration of conservative systems, and several researchers have explored such applications [58, 59, 60, 61].

The IM discretization applied to system (2) is:

$$\begin{aligned} \mathbf{M}\,(V^{n+1} - V^n)/\Delta t &= -\nabla E\,([X^n + X^{n+1}]/2) - \gamma \mathbf{M}(V^n + V^{n+1})/2 + R^n\,, \\ (X^{n+1} - X^n)/\Delta t &= (V^n + V^{n+1})/2\,. \end{aligned} \tag{10}$$

This nonlinear system can be solved, following [42], by obtaining X^{n+1} as a minimum of

$$\begin{aligned} \Phi(X) &= \frac{\widetilde{\gamma}}{2}\,(X - X_0^n)^{\mathrm{T}}\,\mathbf{M}\,(X - X_0^n) + (\Delta t)^2\,E\left(\frac{X + X^n}{2}\right), \\ X_0^n &= X^n + \Delta t\,V^n/\widetilde{\gamma} + \mathbf{M}^{-1}(\Delta t)^2 R^n/(2\widetilde{\gamma})\,, \quad \widetilde{\gamma} = 1 + (\gamma \Delta t)/2\,. \end{aligned} \tag{11}$$

Hence for IM applied to Newtonian dynamics $\widetilde{\gamma} = 1$ and the R^n term in X_0^n is absent. Following minimization of the IM dynamics function to obtain X^{n+1}, V^{n+1} is obtained from the second equation of system (10).

An examination of the application of IM to MD shows very good numerical properties (e.g., energy conservation and stability) for moderate timesteps, larger than Verlet [62, 41]. However integrator-induced *resonance* artifacts limit the application of this approach to larger integration stepsizes. Essentially, resonance occurs at special timesteps that are related in a complex way (via the stepwise propagation transformation) to the various timescales of the motion [63, 64, 6, 65]. At those timesteps, a concerted effect stemming from one component of the motion (e.g., heating of a bond-stretch vibrational mode) leads to very large energetic fluctuations or instability (e.g., bond rupture). Thus, resonance problems lead in general to erratic, rather than systematic, error patterns as a function of timestep. They are also method and system dependent [63, 66, 67], occur for both implicit and explicit schemes (e.g., Verlet and IM [63, 62]), and depend strongly on the fastest frequency in the system and possibly on the coupling strength to other vibrational modes.

Resonant timesteps can be estimated on the basis of one-dimensional analysis [65, 62] from the propagating rotation matrices in phase-space for

symplectic methods. Namely, it can be shown that the *effective rotation angle* θ_{eff} in phase space is a function of the timestep and the integrator [65, 66]. The *effective angular frequency* $\omega_{\text{eff}} = \theta_{\text{eff}}/\Delta t$, can be expressed in terms of the natural frequency ω of the model oscillator (as a function of Δt and the method). For example, we can estimate these resonant timesteps for our α-parameterized family of symplectic methods ($\alpha \geq 0$) which uses the following discrete approximation to the force vector F at time $n\Delta t$:

$$F^n = F(X^n + \alpha\, \Delta t^2\, \mathbf{M}^{-1} F^n)\,. \tag{12}$$

Note that the method is implicit unless $\alpha = 0$; symplecticness is proven in [66, 67]. For these schemes, we obtain

$$\omega_{\text{eff}}\, \Delta t = \theta_{\text{eff}} = 2 \arcsin\left(\omega\, \Delta t\, \sqrt{\phi}/2\right), \tag{13}$$

where

$$\phi = 1/(1 + \alpha(\omega \Delta t)^2)\,. \tag{14}$$

The values $\alpha = 0, 1/12, 1/4$ and $1/2$ correspond, respectively [66], to the Verlet, Störmer-Cowell/Numerov, implicit-midpoint, and "LIM2" methods, the latter introduced in [41]. All integrators are second-order, except for Störmer-Cowell/Numerov, which is fourth-order accurate.

The resonant timesteps of order $n : m$ (where n and m are integers), correspond to a sampling of n phase-space points in m revolutions. This condition implies that

$$n\, \theta_{\text{eff}} \;=\; 2\pi\, m\,. \tag{15}$$

By substituting the method and Δt-dependent formula for ω_{eff} in this expression, we can estimate the resonant timesteps on the basis of linear analysis [62]:

$$\Delta t_{n:m} \;=\; 2\pi m/n\, \omega_{\text{eff}}\,. \tag{16}$$

The smaller the value of n (the *resonance order*), the larger the timestep of disturbance. For example, the linear stability for Verlet is $\omega\Delta t < 2$ for second-order resonance, while IM has no finite limit for stability of this order. Third-order resonance is limited by $\sqrt{3}$ (≈ 1.72) for Verlet compared to about double, or $2\sqrt{3}$ (≈ 3.46), for IM. See Table 1 for limiting values of $\omega\Delta t$ corresponding to interesting combinations of α and n. This table also lists timestep restrictions relevant to biomolecular dynamics, assuming the fastest motion has period of around 10 fs (appropriate for an O–H stretch, for example).

These estimates concur with simulation observations for a more complex nonlinear system, a blocked alanine model [67]. Specifically, Verlet becomes

Table 1. Stability Limits and Resonant Timesteps for the Verlet and Implicit Midpoint Schemes

n, resonance order	Limit on $\omega\Delta t$		$\Delta t_{n:1}$ for MD[a]	
	Verlet	IM	Verlet	IM
2	2.0	∞	3.2	∞
3	1.7	3.5	2.8	5.5
4	1.4	2.0	2.3	3.2
5	1.2	1.5	1.9	2.3

[a] based on $\omega = 0.63$ fs^{-1}, which corresponds to the period $P = 2\pi\omega = 10$ fs, appropriate for an O-H stretch

unstable for a timestep around 2.8 fs, roughly half the value yielding instability for IM, though IM exhibits large energy fluctuations when the timestep exceeds 5 fs (see Fig. 9). This resonance problem, combined with the CPU requirements for minimization, dampen the likelihood of using implicit methods effectively for biomolecular dynamics.

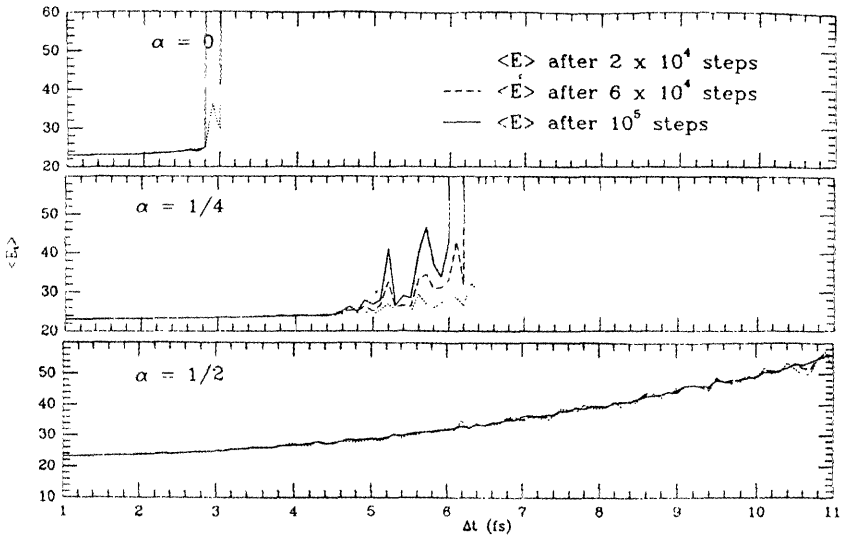


Fig. 9. Mean total energy vs. Δt for the Verlet ($\alpha = 0$), IM ($\alpha = 1/4$) and LIM2 ($\alpha = 1/2$) schemes for the blocked alanine model. The three lines correspond to averaging energies over an increasing number of steps: 2×10^4 (dotted line), 6×10^4 (dashed line), and 10^5 (solid line).

4.3 Other Implicit Methods

The IE and IM methods described above turn out to be quite special in that IE's damping is extreme and IM's resonance patterns are quite severe relative to related symplectic methods. However, success was not much greater with a symplectic implicit Runge-Kutta integrator examined by Janežič and coworkers [40].

The symplectic method LIM2 [41] was further explored in [67] with the thought that it might alleviate resonance in comparison to IM. It turns out that the parameter α affects the relationship between the numerical frequency and the actual frequency of the system. Specifically, the maximum possible phase angle change per timestep *decreases* as the parameter α *increases*. Hence, the angle change can be limited by selecting a suitable α. The choice $\alpha \geq \frac{1}{2}$ restricts the phase angle change to less than one quarter of a period and thus is expected to eliminate notable disturbances due to fourth-order resonance. The requirement that $\alpha \geq \frac{1}{3}$ guarantees that the phase angle change per timestep is less than one third of a period and therefore should also avoid third-order resonance for the model problem. This was found to hold in our application to a representative nonlinear system, a blocked alanine residue [67]. Namely, the energy averages increase with Δt but exhibit no erratic resonance patterns for LIM2 as did IM (Fig. 9). Unfortunately, these energetic increases are not acceptable (e.g., approximately 30% and 100% of the small-timestep value, respectively, for $\Delta t = 5$ and 9 fs for this system). Part of this behavior is also due to an error constant for LIM2 that is greater that of leap-frog/Verlet.

4.4 CPU Time

The CPU time required for implementation of implicit schemes is clearly a practical concern. Simply put, it is difficult to beat an explicit method which only requires one force evaluation per timestep, even on the most advanced computer architectures [68]. CPU costs have not been detailed in general for the implicit schemes attempted for MD, since the systems examined for feasibility were often simple and not representative of macromolecules. Solution performance of the nonlinear system, or equivalently the nonlinear optimization, subproblem also depends on the method used for this task, the timestep value of the integrator, and other factors. However, we have estimated based on our experience for CPU considerations alone (i.e., ignoring the quality of results obtained at large timesteps) that the timestep in implicit schemes must be at least 20 times greater that used in Verlet just for breaking even in total CPU time. It is difficult to expect reasonable resolution in this regime given the stability limit and resonance problems outlined above. Perhaps semi-implicit [2] or cheaper implementations of implicit schemes [69] will better handle this problem: it might be possible to treat the local terms implicitly and the nonlocal terms explicitly. Exploitation of parallel machine

architecture has potential for further speedup but, if experience to date on parallelization of linear algebra codes can be considered representative, parallel computers tend to favor explicit methods.

5 Force Linearization/Normal Modes

A separation into harmonic and anharmonic motion is a reasonable strategy for attacking the timescale disparity in biomolecules. This is because the high-frequency motion is largely harmonic; thus if this component is filtered out in some way, a large timestep may be used for the residual motion. This concept was first explored by Tuckerman *et al.* [70] but abandoned early because of the large cost of the reference solution for realistic systems; originally this approach was advocated because of greater claimed accuracy than multiple-timestep methods. We too found our LIN approach based on normal modes [71, 72] to be very expensive, but LIN led eventually to the competitive LN approach. This experience suggests that in general methods that might initially appear computationally demanding or even intractable are still worth considering, since competitive variations or further developments, in both software and hardware, cannot be anticipated.

The LIN method (described below) was constructed on the premise of filtering out the high-frequency motion by NM analysis and using a large-timestep implicit method to resolve the remaining motion components. This technique turned out to work when properly implemented for up to moderate timesteps (e.g., 15 fs) [73] (each timestep interval is associated with a new linearization model). However, the CPU gain for biomolecules is modest even when substantial work is expanded on sparse matrix techniques, adaptive timestep selection, and fast minimization [73]. Still, LIN can be considered a true long-timestep method.

Following a brief discussion of normal mode (NM) techniques, we will describe the LIN method and summarize the results obtained.

5.1 Normal Modes (NM)

The normal mode approach describes the motion of a system from its collection of harmonic vibrations. The frequencies are determined by the shape of the potential energy surface only near energy minima, and once these modes are determined at a given configuration, equilibrium and dynamic properties can be approximated. To date, NM has been used mainly an analysis tool for studying dynamic properties of biological systems. Quasi-harmonic extensions can partially treat larger-amplitude motions [74, 75, 35], but they cannot explicitly account for multiple minima in configuration space. The required eigenvalue calculations for full determination of the normal-mode displacements and frequencies are prohibitively expensive for biomolecules (as they scale with the third power of the number of independent variables),

but approximate procedures for large systems have been developed for selected (typically low-frequency) modes (e.g., [74, 76, 77, 78, 79, 80]).

The merging of NM analysis and MD is attractive because of the complementarity of the two techniques. The former provides an accurate description for an equilibrium reference system, while the latter can yield in theory a complete description of nonequilibrium events. Such a union has been mainly pursued in the context of 'essential dynamics' [81, 82] in which major features of macromolecules are approximately described by following a small set of low-frequency vibrational modes. The problem in this description arises from the need to identify this low-frequency subset from dynamic simulations (via correlation matrices): a dominant set of low-frequency modes cannot, in general, be reliably estimated from the short simulations feasible today [83], given the long relaxation times of proteins in solution. Therefore, techniques based on projecting the low-frequency motion on the Newton equations of motion [81, 82], or splitting harmonic and anharmonic motion [84] with limited harmonic-model updating, cannot work for general biomolecular applications. Deviations from the harmonic approximation can emerge within 15 fs [73]. For the same reasons, a realistic description of a protein's long-time dynamics based on a low-frequency vibrational subset might only be possible through substantial incorporation of information from enhanced sampling techniques [82].

A rather different approach based in part on normal modes is the substructuring of Turner *et al.* [85]. This technique, originating in aerospace dynamics, partitions a multibody system into a collection of rigid and flexible particles. The motion of the atoms within these bodies is then propagated via selected low-frequency normal-mode components; the dynamic interactions between bodies are modeled rigorously. Large overall computational gains might be possible, but significant work is needed to devise system-dependent substructuring protocols. Although it is difficult to show general agreement with small timestep dynamic simulations by this approach, slow-scale motions might be more easily captured than with traditional methods.

5.2 The LIN Method

The LIN method ("Langevin/Implicit/Normal-Modes") combines frequent solutions of the linearized equations of motions with anharmonic corrections implemented by implicit integration at a large timestep. Namely, we express the collective position vector of the system as $X(t) = X_h(t) + Z(t)$. (In LN, $Z(t)$ is zero). The first part of LIN solves the linearized Langevin equation for the 'harmonic' reference component of the motion, $X_h(t)$. The second part computes the residual component, $Z(t)$, with a large timestep.

Linearized Reference System We first formulate the following *linearized* Langevin system at some reference point X_r (e.g., X^n, $X^n + \frac{\Delta t}{2} V^n$):

$$\begin{aligned} \mathbf{M}\dot{V} &= -\nabla E(X_r) - \widetilde{\mathbf{H}}(X_r)\,(X - X_r) \; - \gamma \mathbf{M} V + R, \\ \dot{X} &= V\,. \end{aligned} \tag{17}$$

The matrix $\widetilde{\mathbf{H}}$ is a sparse approximation to the Hessian of E at X_r [73], such as the Hessian resulting from short (e.g., 4.5 Å) cutoffs [73] or the second derivatives coming from the bond-length, bond-angle, dihedral-angle, and the 1–4 electrostatic components. The goal is to formulate cheaply a reasonable description for the harmonic motion. Our experience has shown that the second $\widetilde{\mathbf{H}}$ choice (local energy components) is much easier to implement (to exploit sparsity), though the cutoff matrix might be preferred if cost were not an issue since the interval over which the linearization is retained can be lengthened [73].

The solution $X_h(t)$ of the linearized equations of motion can be solved by standard NM techniques or, alternatively, by explicit integration. We have experimented with both and found the second approach to be far more efficient and to work equally well. Its handling of the random force discretization is also more straightforward (see below). For completeness, we describe both approaches here.

Analytic Solution

The standard analytic procedure involves calculating the orthogonal transformation matrix $\mathbf{T}$ that diagonalizes the mass weighted Hessian approximation $\mathbf{H}' \equiv \mathbf{M}^{-\frac{1}{2}} \widetilde{\mathbf{H}} \mathbf{M}^{-\frac{1}{2}}$, namely

$$\mathbf{D} = \mathbf{T}\widetilde{\mathbf{H}}'\mathbf{T}^{-1}. \tag{18}$$

Eigenvalues of the diagonal matrix $\mathbf{D}$ will be denoted as λ_i. With the transformations

$$Q = \mathbf{T}\mathbf{M}^{1/2}(X - X_r) \quad \text{and} \quad F = \mathbf{T}\mathbf{M}^{-1/2}R, \tag{19}$$

applied to the NM-displacement coordinates Q and random force F, system (17) is reduced to the set of decoupled, scalar differential equations for $\{V_{q_i}\, \dot{V}_{q_i}\}$:

$$\begin{aligned} \dot{V}_{q_i} &= -\mathbf{D}Q - \gamma V_{q_i} + F, \\ \dot{Q} &= V_{q_i}\,. \end{aligned} \tag{20}$$

Here, the force F is a linear combination of the components of R; it also has a Gaussian distribution and autocorrelation matrix that satisfies the same properties of $R(t)$ as shown in eq. (3), with $\mathbf{I}$ (the $n \times n$ unit matrix) replacing $\mathbf{M}$ [71]:

$$\langle F(t) \rangle = 0, \qquad \langle F(t) F(t')^T \rangle = 2\gamma k_B \mathrm{T}\, \mathbf{I}\, \delta(t - t')\,. \tag{21}$$

The initial conditions of system (20) coincide with those for the original equations: $X_h(0) = X^n$ and $V_h(0) = V^n$. Appropriate treatments, as discussed in [72], are essential for the random force at large timesteps to maintain thermal equilibrium since the discretization $\delta(t - t') \Rightarrow \delta_{nm}/\Delta t$ is poor for large Δt. This problem is alleviated by the numerical approach below because the relevant discretization of the Dirac function is the inner timestep $\Delta\tau$ rather than a large Δt.

The decoupled set of equations in system (20) can be solved for all the Q_i and associated velocities V_{q_i} by closed-form formulas that depend on the eigenvalues [71]. The harmonic position and velocity vectors at time $n\Delta t$ can then obtained from the expressions:

$$\begin{aligned} X_h^{n+1} &= X^n + \mathbf{M}^{-1/2}\mathbf{T}^{-1}Q\,, \\ V_h^{n+1} &= \mathbf{M}^{-1/2}\mathbf{T}^{-1}\dot{Q}\,. \end{aligned} \tag{22}$$

Numeric Solution

The alternative numerical procedure involves solving system (17) by integration with a small 'inner timestep', $\Delta\tau$, as required for traditional MD (0.5 or 1 fs). In theory, the procedure can become unstable for negative eigenvalues λ since the associated solutions, $\exp(-i\sqrt{\lambda}t)$ (here only i is the complex number $\sqrt{-1}$), will grow in magnitude. However, for reasonable small timesteps these instabilities appear mild and require no special treatment in our context. The cost of this solution is also minimal, with each iteration dominated by sparse-Hessian/vector products and *no new force evaluations*, as in each MD step. For this integration, we use the second-order partitioned Runge-Kutta method (Lobatto IIIa,b) [86], which reduces to the velocity Verlet method when $\gamma = 0$. We take X_r as the 'midpoint' $X^n + \frac{\Delta t}{2}V^n$ and define the inner iteration process as:

$$\begin{aligned} X_{i+\frac{1}{2}} &= X_i + \frac{\Delta\tau}{2}V_i, \\ V_{i+1} &= V_i - \Delta\tau\mathbf{M}^{-1}[\nabla E(X_r) + \tilde{\mathbf{H}}(X_r)(X_{i+\frac{1}{2}} - X_r) + \gamma\mathbf{M}V_{i+1} - R], \\ X_{i+1} &= X_{i+\frac{1}{2}} + \frac{\Delta\tau}{2}V_{i+1}. \end{aligned} \tag{23}$$

This loop is iterated k_1 times to cover the interval Δt ($k_1\Delta\tau = \Delta t$) to produce X_h^{n+1}. Note the Hessian/vector products in the second equation of system (23). The random force is updated according to eq. (3) at every $\Delta\tau$ substep, so there is no problem of thermal equilibrium as for larger timesteps [72].

Residual (Anharmonic) Correction Once the next $X_h(t)$, X_h^{n+1}, is obtained as a solution to system (17) by either the analytic or numeric procedures as above, the residual motion component, $Z(t)$, can be determined by

solving the new set of equations which Z satisfies. These are determined by using $Z = X - X_h$ and the origin of X and X_h as solutions of systems (2) and (17) respectively. This leads to the system [71]:

$$\begin{aligned} \mathbf{M}\dot{W}(t) &= -\nabla E(X_h + Z(t)) - \gamma \mathbf{M} W(t) + \nabla E(X_r(t)) + \widetilde{\mathbf{H}}(X_h - X_r), \\ \dot{Z}(t) &= W(t), \end{aligned} \tag{24}$$

where W is the time derivative of Z. The initial conditions for system (24) are: $Z(0) = 0$ and $W(0) = 0$.

To solve this system, we apply the implicit midpoint scheme (see system (10)) to system (24) and follow the same algebraic manipulation outlined in [71, 72] to produce a nonlinear system $\nabla\Phi(Y) = 0$, where $Y = (X + X^n)/2$. This system can be solved by reformulating this solution as a minimization task for the dynamics function Φ:

$$\Phi(Y) = 2\widetilde{\gamma}\,(Y - Y_0^n)^T\,\mathbf{M}\,(Y - Y_0^n) + (\Delta t)^2\,E(Y), \quad \widetilde{\gamma} = 1 + (\gamma\Delta t)/2, \tag{25}$$

where

$$Y_0^n = \frac{X_h^{n+1} + X^n}{2} + \frac{\Delta t^2}{4\widetilde{\gamma}}\mathbf{M}^{-1}\left[\nabla E(X_r) + \widetilde{\mathbf{H}}\left((X_h^{n+1} + X^n)/2 - X_r\right)\right] \tag{26}$$

Thus, each correction step of LIN requires nonlinear minimization of Φ. The initial approximate minimizer of Φ can be X_h^{n+1} or $(X_h^{n+1} + X^n)/2$ (we use the latter). The new coordinate and velocity vectors for timestep $n+1$ are then obtained from the relations

$$X^{n+1} = 2Y - X^n, \qquad V^{n+1} = V_h^{n+1} + 2\,(X^{n+1} - X_h^{n+1})/\Delta t. \tag{27}$$

LIN Results Computational performance of LIN was enhanced by using sparse Hessian/vector products in the first part and using more lenient convergence criteria than the default values in the minimization subproblem. With 0.5-fs inner timestep values used for solution of the linearized equations of motion, we found the work required for the linearized problem overall very small (< 20 %). Comparison of LIN results to explicit simulations (BBK algorithm, system 5) at that timestep showed excellent agreement of thermodynamic properties for a blocked alanine model for which the LIN outer timestep, Δt, was 15–30 fs [73]. For BPTI, LIN results with $\Delta t = 15$ fs gave very good agreement with the corresponding explicit simulations with respect to energetic and structural properties.

The LIN outer timestep is limited by the range of validity of the harmonic approximation. The value 15 fs appears to be the largest for achieving essentially the same global behavior for LIN and the reference explicit trajectories. This can be seen from Fig. 10 showing for the blocked alanine model the differences in potential energy components (bond length, bond angle, dihedral angle, van der Waals, and electrostatic) before and after the residual corrections corresponding to LIN trajectories at timesteps of 2 (yellow), 5 (red),

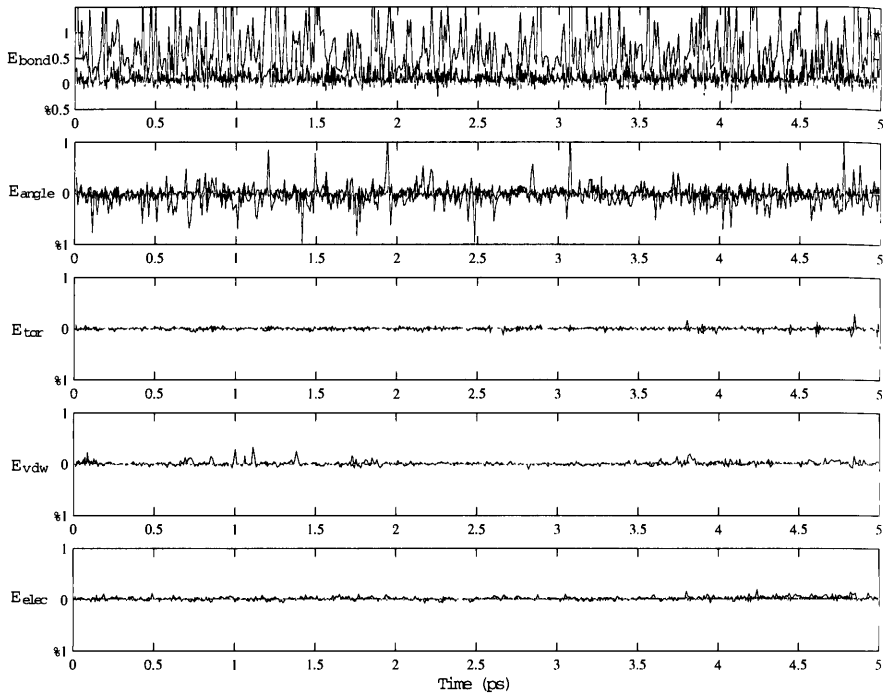


Fig. 10. Differences in potential energy components for the blocked alanine model (for bond length, bond angle, dihedral angle, van der Waals, and electrostatic terms, shown top to bottom) before and after the residual corrections in LIN trajectories at timesteps of 2 fs (yellow), 5 fs (red), and 10 fs (blue).

and 10 (blue) fs. Thus, the 2-fs curves (yellow) are essentially flat; the red curves (5 fs) show some differences for the bond lengths and angles, and the blue patterns (10 fs) indicate larger differences in these two components.

With the improvements outlined here, LIN becomes a true long timestep method in the definition of García-Archilla *et al.* [87]. Unfortunately, despite substantial optimization work, speedup is only modest on a single processor — a factor of 1.4 for BPTI — since minimization still consumes more than 80% of the total CPU time, The speedup factor should increase with system size, and further routes for optimization might be identified, but we did not pursue this line of research. Instead, our examination of the magnitude of the anharmonic correction vector (or, similarly, the range of validity of the harmonic approximation) led to development of LN, which includes LIN's linearization, but not the expensive, correction step [73]. Even though the frequency of harmonic updating in LN can only reach 5 fs for good agreement with small-timestep simulations, the speedup by this approach is much better than LIN and similar to the factor obtained by MTS methods (e.g., 5,

with respect to simulations using a constant inner timestep). With the additional incorporation of force splitting into LN, the relevant outer timestep can be increased much further than this harmonic-updating value, with greater overall speedup.

6 Force Linearization + Force Splitting: The LN Method

The idea of LN is to eliminate LIN's expensive implicit integration component and, concomitantly, reduce the interval length over which the harmonic model is followed. This view is reasonable given that the anharmonic corrections are small when $\Delta t < 5$ fs (Fig. 10). Our first implementation of LN was without force splitting [73]. The method was verified in this form for model proteins and shown to yield a speedup factor of around 5 with respect to reference (BBK) explicit trajectories ($\Delta\tau = 0.5$ fs). This is possible because the linearization component is relatively cheap. As found by many others who introduced MTS schemes, the expensive component is the work required for the gradient evaluation of large systems (e.g., more than 80% of the total CPU for systems of 2000 atoms and more). In the context of the LN method, we found that an extrapolative force-splitting approach, together with the Langevin formulation, can alleviate severe resonances [88, 89].

6.1 LN Without Force Splitting

The skeletal LN procedure is a dual timestep scheme, $\{\Delta\tau, \Delta t_m\}$, of two practical tasks: (a) constructing the Hessian $\widetilde{\mathbf{H}}$ in system (17) every Δt_m interval, and (b) solving system (17), where R is given by eq. (3), at the timestep $\Delta\tau$ by procedure (23) outlined for LIN above. When a force-splitting procedure is also applied to LN, a value $\Delta t > \Delta t_m$ is used to update the slow forces less often than the linearized model. A suitable frequency for the linearization is 1–3 fs (the smaller value is used for water systems), and the appropriate inner timestep is 0.5 fs, as in LIN. This inner timestep parallels the update frequency of the fast motions in force splitting approaches, and the linearization frequency (Δt_m) is analogous to the medium timestep used in such three-class schemes (see below).

Computational speed can be achieved since the subintegration process does not require new force evaluations, as in every step of standard MD integration, and the linearization phase is extremely cheap when sparse-Hessian/vector multiplication routines are employed. Our experience has shown that a sparse $\widetilde{\mathbf{H}}$ resulting from 4.5 Å cutoffs evaluated every 5 fs gives very similar results in comparison to explicit trajectories at 0.5 fs [73]. A sparser $\widetilde{\mathbf{H}}$ including only bonded interactions (i.e., bond-length, bond-angle and dihedral-angle) can be used together with a smaller value, 1–3 fs [5], for similar computational gains [73]. As the sparsity of the bonded $\widetilde{\mathbf{H}}$

is less than 5% for systems with more than 6000 atoms, one calculation of $\widetilde{\mathbf{H}}$ requires about one fourth the CPU of the gradient calculation. This fact, combined with the cheapness of evaluating sparse-Hessian/vector products, explains the method's efficiency.

6.2 LN With Force Splitting

The small CPU percentages for the sparse Hessian evaluation and the linearization in LN as described above also imply that the majority of the work comes from gradient computations. Indeed, the percentage of CPU spent on gradient computations is about 80% for lysozyme and 99% for a large water droplet with more than 12,000 atoms (91% for this water system when nonbonded cutoffs at 12 Å are enforced) [88]. This suggests that additional force splitting, as introduced by Street and coworkers in the 1970s [11] and developed further by Berne and colleagues [70, 13, 16, 17], can yield further speedups. Modern symplectic versions of these MTS methods were pioneered by Schulten and co-workers who developed the first parallel machine for MD computations [14] and independently by Berne, Tuckerman, Martyna [90] and later co-workers. Typically, the force splitting is extended into three classes: fast bonded interactions, local nonbonded interactions, and nonlocal interactions. The splitting of the nonbonded interaction into short and long-range parts can be defined by a spherical range and accomplished using a smooth switching function [12, 88].

In LN, the bonded interactions are treated by the approximate linearization, and the local nonbonded interactions, as well as the nonlocal interactions, are treated by constant extrapolation over longer intervals (Δt_m and Δt, respectively). We define the integers $k_1, k_2 > 1$ by their relation to the different timesteps as $\Delta t_m = k_1 \, \Delta\tau$ and $\Delta t = k_2 \, \Delta t_m$. This extrapolation as used in LN contrasts the modern 'impulse' MTS methods which only add the contribution of the slow forces at the time of their evaluation. The impulse treatment makes the methods symplectic, but limits the outermost timestep due to resonance (see figures comparing LN to impulse-MTS behavior as the outer timestep is increased in [88]). In fact, the early versions of MTS methods for MD relied on extrapolation and were abandoned because of a notable energy drift. This drift is avoided by the phenomenological, stochastic terms in LN.

Formally, we describe the LN method with the above force splitting below for the triplet protocol $\{\Delta\tau, \; \Delta t_m, \; \Delta t\}$. The fast, medium, and slow force components are distinguished by subscripts; we take the medium forces as those nonbonded interactions within a 6 Å region.

LN Algorithm (with Force Splitting via Extrapolation)

$X_r = X + \frac{\Delta t_m}{2} V$
$F_s = -\nabla E_{\text{slow}}(X_r)$
For $j = 1$ to k_2
  $X_r = X + \frac{\Delta t_m}{2} V$
  $\widetilde{\mathbf{H}} = \widetilde{\mathbf{H}}(X_r)$
  $F_m = -\nabla E_{\text{mid}}(X_r) - \nabla E_{\text{fast}}(X_r)$
  $F = F_m + F_s$
  For $i = 1$ to k_1
    evaluate the Gaussian force R
    $X = X + \frac{\Delta \tau}{2} V$
    $V = \left(V + \mathbf{M}^{-1} \Delta \tau \left(F - \widetilde{\mathbf{H}} [X - X_r] + R \right) \right) / (1 + \gamma \Delta \tau)$
    $X = X + \frac{\Delta \tau}{2} V$
  end
end

6.3 LN Results

A detailed examination of LN behavior is available [88] for the blocked alanine model, the proteins BPTI and lysozyme, and a large water system, compared to reference Langevin trajectories, in terms of energetic, geometric, and dynamic behavior. The middle timestep in LN can be considered an adjustable quantity (when force splitting is used), whose value does not significantly affect performance but does affect accuracy with respect to the reference trajectories. For example, we have used $\Delta t_m = 3$ fs for the proteins in vacuum, but 1 fs for the water system, where librational motions are rapid.

In Tables 2 and 3 we show the error percentages of the LN energy components (and kinetic temperature) with respect to the explicit Langevin trajectories at $\Delta \tau = 0.5$ fs for BPTI and lysozyme simulations. The "Reference" column shows energy means and variances for the explicit trajectory (produced by the BBK scheme), and the LN columns (each corresponding to a different $k2$ value, as indicated in the heading) show the *percentage error* of each entry (mean energy component and associated variance) *with respect to the reference values.* Here, $\Delta t_m = 3$ fs, so the LN variants shown ($k2 = 1$, 3, 6, 12, 24, 48, and 96) correspond to outer timesteps Δt of 3, 9, 18, 36, 72, 144, and 288 fs, respectively.

We first note errors in total energy means that are not greater than 0.5% for all LN versions tested. Individual energy components show errors that are generally less than 1%, with the exception of the van der Waals energy that can reach 4% for large k_2. Of course, this discussion of relative errors reflects practical rather than mathematical considerations, since constants can be added to individual terms without affecting the dynamics. The relative errors

Table 2. Percentage error for LN compared to reference Langevin trajectories (at 0.5 fs) for energy means and associated variances for BPTI over 60 ps at $\gamma = 20$ ps^{-1} ($\Delta\tau = 0.5$ fs, $\Delta t_m = 3$ fs, and $\Delta t = k_2 \Delta t_m$, where k_2 ranges from 1 for LN 1 to 96 for LN 96.

	Reference[a] (Δt = 0.5 fs)		LN 1 (Δt = 3 fs)		LN 3 (Δt = 9 fs)		LN 6 (Δt = 18 fs)		LN 12 (Δt = 36 fs)		LN 24 (Δt = 72 fs)		LN 48 (Δt = 144 fs)		LN 96 (Δt = 288 fs)	
E[b]	1620.2	32.4 $_{2\%}$	0.41	1.54	0.33	1.54	0.3	1.85	0.36	2.47	0.3	3.09	0.37	4.32	0.35	5.25
E_k	809.0	22.3 $_{3\%}$	0.43	0.9	0.36	0.45	0.38	0.45	0.42	0.45	0.43	8.52	0.44	0.45	0.44	0.9
E_p	811.1	23.0 $_{3\%}$	0.42	1.3	0.32	1.3	0.22	2.61	0.32	3.91	0.18	5.22	0.32	8.26	0.27	10.43
E_{bond}	322.4	14.6 $_{5\%}$	0.09	0	0.09	0	0.06	0	0.03	0	0.06	0.68	0.03	0.68	0.03	0.68
E_{angle}	456.2	16.0 $_{4\%}$	0.33	0	0.39	0.62	0.53	0.62	0.53	0.62	0.24	0.62	0.28	0.62	0.24	0.63
E_{tor}	353.6	8.93 $_{3\%}$	0.54	0.11	0.42	0.22	0.31	0.11	0.25	0.11	0.4	1.34	0.31	0.34	0.23	0
E_{vdw}	−119.0	12.9 $_{11\%}$	1.43	1.55	1.51	1.55	1.26	1.55	1.26	1.55	3.28	2.33	3.78	3.10	1.68	13.95
E_{elec}	−1958.4	16.0 $_{1\%}$	0.02	9.38	0.02	9.38	0.04	11.87	0.08	15.00	0.16	23.75	0.16	28.75	0.09	32.50
T	300.2	8.28 $_{3\%}$	0.43	0.60	0.37	0.60	0.40	0.48	0.43	0.60	0.43	0.72	0.47	0.48	0.47	0.72
Speedup[c]	(1.0)		4.0		7.4		9.3		10.8		11.7		12.6		13.0	

[a]The variance percentage of the mean energy is shown as a subscript to the variance value recorded for the BBK scheme.

[b]Energy [in Kcal/mol] is given for the total, kinetic, potential (with respect to the initial values −1664.96 corresponding to a local minimum near the initial configuration), bond length, bond angle, torsion angle, van der Waals, and electrostatic components.

[c]The speedup is relative to the reference trajectory, which takes 14 hours.

seen for different terms thus correspond to magnitudes assigned in practice in the current force field.

The errors in the variance values (reflecting the fluctuations about the means) are larger: for the total energy, variance errors can be as large as 7% for large k_2 (the potential energy is the source rather than the kinetic energy); most other entries for energy components are less than 3%, except for two van der Waals values (LN 96 for BPTI and LN 3 for lysozyme) and all electrostatic entries. Note, however, that for the electrostatic energy the variance of the reference trajectory is a very small percentage of the mean value, namely 1% for BPTI. Thus, for example, the LN 96 variance (worst case for BPTI) for the electrostatic energy is still 1% of the reference energy mean although the value in the table is 33% (indicating an absolute energy variance of 16×1.33 kcal/mol.). Thus, the values shown in Tables 2 and 3 still reflect a satisfactory agreement between LN trajectories and small-timestep analogs of the same Langevin equation. See [88] for many other examples of thermodynamic and geometric agreement.

It was also interesting to compare LN behavior as k_2 increases to trajectories that use nonbonded cutoffs: for very large k_2, behavior of the LN trajectory begins to resemble the cutoff trajectory [88]. This observation suggests that the model itself, rather than the numerical scheme *per se*, is responsible for the deviations.

Table 3. Percentage error for LN compared to reference Langevin trajectories (at 0.5 fs) for energy means and associated variances for lysozyme over 60 ps at $\gamma = 20$ ps^{-1} ($\Delta\tau = 0.5$ fs, $\Delta t_m = 3$ fs, and $\Delta t = k_2 \Delta t_m$, where k_2 ranges from 1 for LN 1 to 96 for LN 96.

	Reference[a] (Δt = 0.5 fs)			LN 1 (Δt = 3 fs)		LN 3 (Δt = 9 fs)		LN 6 (Δt = 18 fs)		LN 12 (Δt = 36 fs)		LN 24 (Δt = 72 fs)		LN 48 (Δt = 144 fs)		LN 96 (Δt = 288 fs)	
E[b]	3605.2	46.4	1%	0.42	1.94	0.33	3.23	0.45	2.80	0.43	3.88	0.49	4.96	0.50	5.60	0.46	6.90
E_k	1814.9	32.8	2%	0.42	0.61	0.40	0.61	0.46	0.61	0.49	0.61	0.51	0.61	0.50	0.61	0.49	0.30
E_p	1790.3	33.6	2%	0.42	1.79	0.26	3.27	0.44	3.27	0.37	5.36	0.47	7.44	0.51	9.52	0.42	13.10
E_{bond}	714.6	21.5	3%	0.08	0	0.07	0	0.04	0	0.10	0.47	0.07	0.47	0.06	0.47	0.04	0.47
E_{angle}	1010.3	23.8	2%	0.32	0	0.42	0	0.36	0	0.49	0.42	0.37	0.42	0.36	0.42	0.28	0
E_{tor}	719.4	12.7	2%	0.60	1.57	0.58	2.36	0.67	1.57	0.72	1.57	0.61	2.36	0.68	2.36	0.47	1.57
E_{vdw}	−426.7	20.2	5%	0.80	1.98	1.24	6.44	1.10	1.49	1.78	1.49	0.94	1.98	1.10	2.48	1.48	0.50
E_{elec}	−5080.2	27.2	1%	0.03	2.21	0.05	6.99	0.01	4.78	0.06	5.88	0.03	9.93	0.02	13.97	0.03	22.06
T	299.9	5.4	2%	0.43	1.85	0.40	0	0.47	0	0.50	1.85	0.50	0	0.50	1.85	0.50	0
Speedup[c]	(1.0)			4.9		9.3		12.7		15.4		17.2		18.6		19.1	

[a]The variance percentage of the mean energy is shown as a subscript to the variance value recorded for the BBK scheme.

[b]Energy [in Kcal/mol] is given for the total, kinetic, potential (with respect to the initial values −4637.85 corresponding to a local minimum near the initial configuration), bond length, bond angle, torsion angle, van der Waals, and electrostatic components.

[c]The speedup is relative to the reference trajectory, which takes 72 hours.

Another view of this theme was our analysis of spectral densities. A comparison of LN spectral densities, as computed for BPTI and lysozyme from cosine Fourier transforms of the velocity autocorrelation functions, revealed excellent agreement between LN and the explicit Langevin trajectories (see Fig. 5 in [88]). Here we only compare the spectral densities for different γ: Fig. 8 shows that the Langevin patterns become closer to the Verlet densities ($\gamma = 0$) as γ in the Langevin integrator (be it BBK or LN) is decreased.

Finally, we show in Fig. 11 the speedup of LN at increasing outer timesteps for BPTI, lysozyme, and the large water droplet (with all nonbonded interactions included). We can see speedups exceeding an order of magnitude, reaching an asymptotic value at fairly small k_2. This suggests that the best compromise between efficiency and accuracy in LN is a moderate value of k_2. The asymptotic limit (analyzed in [88]) can be explained by the increasing cost of the medium forces. Note that in LN 3 for lysozyme, evaluation of the medium forces consumes 32% of the total CPU time (53% for slow forces). In LN 48, this value doubles, though the slow-force work is reduced to only 6% of the total time.

Very recently, we have developed and incorporated into the CHARMM molecular mechanics program a version of LN that uses direct-force evaluation, rather than linearization, for the fast-force components [91]. The scheme can be used in combination with SHAKE (e.g., for freezing bond lengths) and with periodic boundary conditions. Results for solvated protein and nucleic-

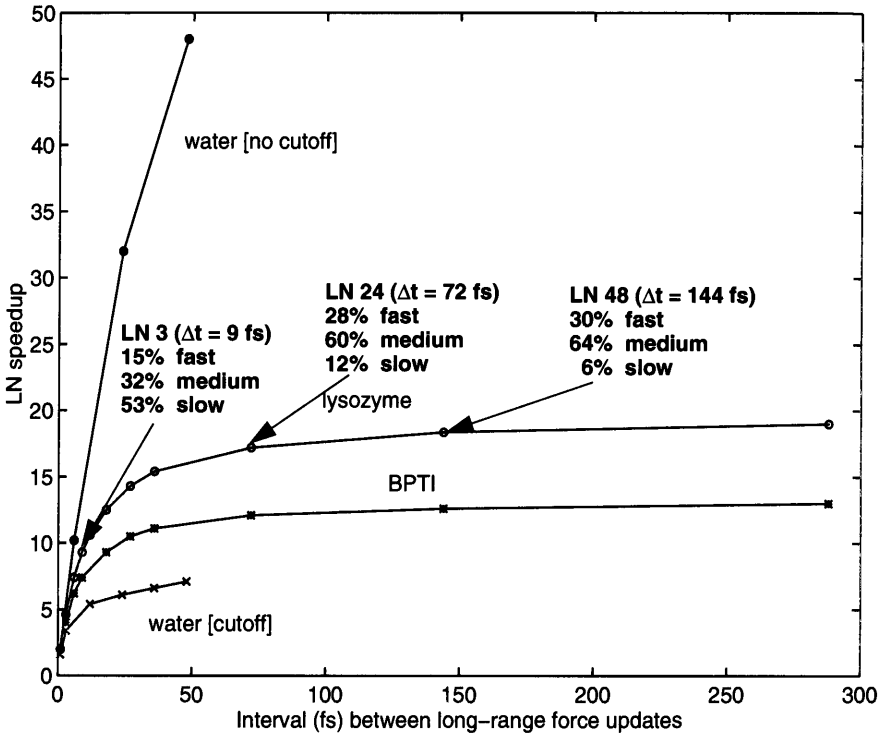


Fig. 11. The Speedup of LN at increasing outer timesteps for BPTI (2712 variables), lysozyme (6090 variables), and a large water system (without nonbonded cutoffs; 37179 variables). For lysozyme, the CPU distribution among the fast, medium, and slow forces is shown for LN 3, 24, and 48.

acid systems [work of A. Sandu and D. Strahs] are very similar to those reported in [88]. A more extensive analysis of resonance artifacts of impulse versus extrapolative Verlet-based force-splitting schemes is described in [91]; the work also offers further insights into LN's good performance.

7 Conclusions

The many approaches to the challenging timestep problem in biomolecular dynamics have achieved success with similar final schemes. However, the individual routes taken to produce these methods — via implicit integration, harmonic approximation, other separating frameworks, and/or force splitting into frequency classes — have been quite different. Each path has encountered different problems along the way which only increased our understanding of the numerical, computational, and accuracy issues involved. This contribution reported on our experiences in this quest. LN has its roots in LIN, which

originally combined normal modes with implicit integration. LIN in turn was developed to remedy nonconservative implicit schemes (like implicit-Euler) or symplectic implicit schemes (like implicit-midpoint) which suffer severely from resonance artifacts.

The heightened appreciation of resonance problems, in particular, has been quite recent [63, 62], and contrasts the more systematic error associated with numerical stability that grows systematically with the discretization size. Ironically, resonance artifacts are worse in the modern impulse multiple-timestep methods, formulated to be symplectic and reversible; the earlier extrapolative variants were abandoned due to energy drifts.

Ultimately, the compromise between the realized speedup and the accuracy obtained for the governing dynamic model should depend on the applications for which the dynamic simulations are used for. For very detailed dynamic pathways, only the Newtonian approach is probably adequate. For general conformational sampling questions, many other simulation methodologies can work well. In particular, if a weak coupling to a phenomenological heat bath, as in the LN method, is tolerated, the general efficiency of force splitting methods can be combined with the long-timestep stability of methods that resolve harmonic and anharmonic motions separately (such as LIN) to alleviate severe resonances and yield speedup. The speedup achieved in LN might be exploited in general thermodynamic studies of macromolecules, with possible extensions into enhanced sampling methods envisioned.

Acknowledgments

I thank Eric Barth, Margaret Mandziuk, Adrian Sandu, Xiaoliang Qian, and Zhongwei Zhu for kindly providing figures and data for this article. I am indebted to them and to other excellent colleagues — Philippe Derreumaux, Hongmei Jian, Bimal Mishra, Gomathi Ramachandran, Bob Skeel, and Guihua Zhang — for sharing the labor and joys of this work. Support from the National Science Foundation (*Grand Challenge Award* ASC-9318159) and the National Institutes of Health (NYU subcontract for the Parallel Computing Resource for Structural Biology award RR08102 to Jan Hermans) is acknowledged. T.S. is an investigator of the Howard Hughes Medical Institute.

References

1. H. S. Chan and K. A. Dill. The protein folding problem. *Physics Today*, 46:24–32, 1993.
2. J. A. McCammon, B. M. Pettitt, and L. R. Scott. Ordinary differential equations of molecular dynamics. *Computers Math. Applic.*, 28:319–326, 1994.
3. B. J. Leimkuhler, S. Reich, and R. D. Skeel. Integration methods for molecular dynamics. In J. P. Mesirov, K. Schulten, and D. W. Sumners, editors, *Mathematical Approaches to Biomolecular Structure and Dynamics*, volume 82 of *IMA Volumes in Mathematics and Its Applications*, pages 161–186, New York, New York, 1996. Springer-Verlag.

4. R. Elber. Novel methods for molecular dynamics simulations. *Curr. Opin. Struc. Biol.*, 6:232–235, 1996.
5. T. Schlick, E. Barth, and M. Mandziuk. Biomolecular dynamics at long timesteps: Bridging the timescale gap between simulation and experimentation. *Ann. Rev. Biophys. Biomol. Struc.*, 26:179–220, 1997.
6. J. M. Sanz-Serna and M. P. Calvo. *Numerical Hamiltonian Problems.* Chapman & Hall, London, 1994.
7. J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comp. Phys.*, 23:327–341, 1977.
8. H.C. Andersen. Rattle: a 'velocity' version of the SHAKE algorithm for molecular dynamics calculations. *J. Comp. Phys.*, 52:24–34, 1983.
9. B. Leimkuhler and R. D. Skeel. Symplectic numerical integrators in constrained Hamiltonian systems. *J. Comp. Phys.*, 112:117–125, 1994.
10. E. Barth, K. Kuczera, B. Leimkuhler, and R. D. Skeel. Algorithms for constrained molecular dynamics. *J. Comp. Chem.*, 16:1192–1209, 1995.
11. W. B. Streett, D. J. Tildesley, and G. Saville. Multiple time step methods in molecular dynamics. *Mol. Phys.*, 35:639–648, 1978.
12. M. E. Tuckerman, B.J. Berne, and G. J. Martyna. Molecular dynamics algorithm for multiple time scales: Systems with long range forces. *J. Chem. Phys.*, 94:6811–6815, 1991.
13. M. E. Tuckerman and B. J. Berne. Molecular dynamics in systems with multiple time scales: Systems with stiff and soft degrees of freedom and with short and long range forces. *J. Comp. Chem.*, 95:8362–8364, 1992.
14. H. Grubmüller, H. Heller, A. Windemuth, and K. Schulten. Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Mol. Sim.*, 6:121–142, 1991.
15. M. Watanabe and M. Karplus. Dynamics of molecules with internal degrees of freedom by multiple time-step methods. *J. Chem. Phys.*, 99:8063–8074, 1993.
16. D. E. Humphreys, R. A. Friesner, and B. J. Berne. A multiple-time-step molecular dynamics algorithm for macromolecules. *J. Phys. Chem.*, 98(27):6885–6892, 1994.
17. R. Zhou and B. J. Berne. A new molecular dynamics method combining the reference system propagator algorithm with a fast multipole method for simulating proteins and other complex systems. *J. Chem. Phys.*, 103:9444–9459, 1995.
18. L. Verlet. Computer 'experiments' on classical fluids: I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.*, 159(1):98–103, July 1967.
19. M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids.* Oxford University Press, New York, New York, 1990.
20. P. Deuflhard, M. Dellnitz, O. Junge, and Ch. Schütte. Computation of essential molecular dynamics by subdivision techniques: I. Basic concepts. Technical Report SC 96–45, Konrad-Zuse-Zentrum für Informationstechnik Berlin, Takustraβe 7, D-14195, Berlin-Dahlem, December 1996.
21. D. A. McQuarrie. *Statistical Mechanics.* Harper & Row, New York, New York, 1976. Chapters 20–21.
22. R. W. Pastor. Techniques and applications of Langevin dynamics simulations. In G. R. Luckhurst and C. A. Veracini, editors, *The Molecular Dynamics of Liquid Crystals*, pages 85–138. Kluwer Academic, Dordrecht, The Netherlands, 1994.

23. A. Brünger, C. L. Brooks, III, and M. Karplus. Stochastic boundary conditions for molecular dynamics simulations of ST2 water. *Chem. Phys. Lett.*, 105:495–500, 1982.
24. T. Simonson. Accurate calculation of the dielectric constant of water from simulations of a microscopic droplet in vacuum. *Chem. Phys. Lett.*, 250:450–454, 1996.
25. D. Beglov and B. Roux. Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations. *J. Chem. Phys.*, 100:9050–9063, 1994.
26. D. Beglov and B. Roux. Dominant solvations effects from the primary shell of hydration: Approximation for molecular dynamics simulations. *Biopolymers*, 35:171–178, 1994.
27. D. Beglov and B. Roux. Numerical solutions of the hypernetted chain equation for a solute of arbitrary geometry in three dimensions. *J. Chem. Phys.*, 103:360–364, 1995.
28. R. J. Loncharich, B. R. Brooks, and R. W. Pastor. Langevin dynamics of peptides: The frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide. *Biopolymers*, 32:523–535, 1992.
29. P. Derreumaux and T. Schlick. Long-time integration for peptides by the dynamics driver approach. *Proteins: Struc. Func. Gen.*, 21:282–302, 1995.
30. M. H. Hao, M. R. Pincus, S. Rackovsky, and H. A. Scheraga. Unfolding and refolding of the native structure of bovine pancreatic trypsin inhibitor studied by computer simulations. *Biochemistry*, 32:9614–9631, 1993.
31. G. Ramachandran and T. Schlick. Solvent effects on supercoiled DNA dynamics explored by Langevin dynamics simulations. *Phys. Rev. E*, 51:6188–6203, 1995.
32. D. K. Klimov and D. Thirumalai. Viscosity dependence of the folding rates of proteins. *Phys. Rev. Lett.*, 79:317–320, 1997.
33. R. C. Wade, M. E. Davis, B. A. Luty, J. D. Madura, and J. A. McCammon. Gating of the active site of triose phosphate isomerase: Brownian dynamics simulations of flexible peptide loops in the enzyme. *Biophys. J.*, 64:9–15, 1993.
34. Hongmei Jian. *A Combined Wormlike-Chain and Bead Model for Dynamic Simulations of Long DNA*. PhD thesis, New York University, Department of Physics, New York, New York, October 1997.
35. D. A. Case. Normal mode analysis of protein dynamics. *Curr. Opin. Struc. Biol.*, 4:385–290, 1994.
36. G. Ramachandran and T. Schlick. Beyond optimization: Simulating the dynamics of supercoiled DNA by a macroscopic model. In P. M. Pardalos, D. Shalloway, and G. Xue, editors, *Global Minimization of Nonconvex Energy Functions: Molecular Conformation and Protein Folding*, volume 23 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 215–231, Providence, Rhode Island, 1996. American Mathematical Society.
37. R. W. Pastor, B. R. Brooks, and A. Szabo. An analysis of the accuracy of Langevin and molecular dynamics algorithms. *Mol. Phys.*, 65:1409–1419, 1988.
38. C. W. Gear. *Numerical Initial Value Problems in Ordinary Differential Equations*. Prentice Hall, Englewood Cliffs, New Jersey, 1971.
39. E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems*, volume 14 of *Springer Series in Computational Mathematics*. Springer-Verlag, New York, New York, second edition, 1996.
40. D. Janežič and B. Orel. Implicit Runge-Kutta method for molecular dynamics integration. *J. Chem. Info. Comp. Sci.*, 33:252–257, 1993.

41. G. Zhang and T. Schlick. Implicit discretization schemes for Langevin dynamics. *Mol. Phys.*, 84:1077–1098, 1995.
42. C. S. Peskin and T. Schlick. Molecular dynamics by the backward Euler's method. *Comm. Pure App. Math.*, 42:1001–1031, 1989.
43. T. Schlick. Pursuing Laplace's vision on modern computers. In J. P. Mesirov, K. Schulten, and D. W. Sumners, editors, *Mathematical Applications to Biomolecular Structure and Dynamics*, volume 82 of *IMA Volumes in Mathematics and Its Applications*, pages 219–247, New York, New York, 1996. Springer-Verlag.
44. T. Schlick and M. L. Overton. A powerful truncated Newton method for potential energy functions. *J. Comp. Chem.*, 8:1025–1039, 1987.
45. T. Schlick and A. Fogelson. TNPACK — A truncated Newton minimization package for large-scale problems: I. Algorithm and usage. *ACM Trans. Math. Softw.*, 14:46–70, 1992.
46. P. Derreumaux, G. Zhang, B. Brooks, and T. Schlick. A truncated-Newton method adapted for CHARMM and biomolecular applications. *J. Comp. Chem.*, 15:532–552, 1994.
47. D. Xie and T. Schlick. Efficient implementation of the truncated Newton method for large-scale chemistry applications. *SIAM J. Opt.*, 1997. In Press.
48. T. Schlick, S. Figueroa, and M. Mezei. A molecular dynamics simulation of a water droplet by the implicit-Euler/Langevin scheme. *J. Chem. Phys.*, 94:2118–2129, 1991.
49. A. Nyberg and T. Schlick. Increasing the time step in molecular dynamics. *Chem. Phys. Lett.*, 198:538–546, 1992.
50. T. Schlick and C. S. Peskin. Comment on: The evaluation of LI and LIN for dynamics simulations. *J. Chem. Phys.*, 103:9888–9889, 1995.
51. T. Schlick and W. K. Olson. Supercoiled DNA energetics and dynamics by computer simulation. *J. Mol. Biol.*, 223:1089–1119, 1992.
52. T. Schlick and W. K. Olson. Trefoil knotting revealed by molecular dynamics simulations of supercoiled DNA. *Science*, 257:1110–1115, 1992.
53. T. Schlick, B. Li, and W. K. Olson. The influence of salt on DNA energetics and dynamics. *Biophys. J.*, 67:2146–2166, 1994.
54. G. Liu, W. K. Olson, and T. Schlick. Application of Fourier methods to computer simulation of supercoiled DNA. *Comp. Polymer Sci.*, 5:7–27, 1995.
55. G. Liu, T. Schlick, A. J. Olson, and W. K. Olson. Configurational transitions in Fourier-series-represented DNA supercoils. *Biophys. J.*, 73:1742–1762, 1997.
56. T. Schlick. Modeling superhelical DNA: Recent analytical and dynamic approaches. *Curr. Opin. Struc. Biol.*, 5:245–262, 1995.
57. P. Derreumaux and T. Schlick. The loop opening/closing motion of the enzyme triosephosphate isomerase. *Biophys. J.*, 74:72–81, 1998.
58. F. Kang. The Hamiltonian way for computing Hamiltonian dynamics. In R. Spigler, editor, *Applied and Industrial Mathematics*, pages 17–35. Kluwer Academic, Dordrecht, The Netherlands, 1990.
59. J. C. Simo, N. Tarnow, and K. K. Wang. Exact energy-momentum conserving algorithms and symplectic schemes for nonlinear dynamics. *Computer Methods in Applied Mechanics and Engineering*, 100:63–116, 1994.
60. J. C. Simo and N. Tarnow. The discrete energy-momentum method. Conserving algorithms for nonlinear elastodynamics. *ZAMP*, 43:757–793, 1992.

61. O. Gonzales and J. C. Simo. On the stability of symplectic and energy-momentum conserving algorithms for nonlinear Hamiltonian systems with symmetry. *Comp. Meth. App. Mech. Engin.*, 134:197, 1994.
62. M. Mandziuk and T. Schlick. Resonance in the dynamics of chemical systems simulated by the implicit-midpoint scheme. *Chem. Phys. Lett.*, 237:525–535, 1995.
63. J. J. Biesiadecki and R. D. Skeel. Dangers of multiple-time-step methods. *J. Comp. Phys.*, 109:318–328, 1993.
64. D. I. Okunbor and R. D. Skeel. Canonical numerical methods for molecular dynamics simulations. *J. Comp. Chem.*, 15:72–79, 1994.
65. V. I. Arnold. *Mathematical Methods of Classical Mechanics.* Springer-Verlag, New York, Heidelberg, Berlin, 1989. second edition.
66. R. D. Skeel, G. Zhang, and T. Schlick. A family of symplectic integrators: Stability, accuracy, and molecular dynamics applications. *SIAM J. Sci. Comp.*, 18(1):202–222, January 1997.
67. T. Schlick, M. Mandziuk, R.D. Skeel, and K. Srinivas. Nonlinear resonance artifacts in molecular dynamics simulations. *J. Comp. Phys.*, 139:1–29, 1998.
68. D. Janežič and B. Orel. Parallelization of an implicit Runge-Kutta method for molecular dynamics integration. *J. Chem. Info. Comp. Sci.*, 34:641–646, 1994.
69. M. Zhang and R. D. Skeel. Cheap implicit symplectic integrators. *Applied Numerical Mathematics*, 25:297–302, 1997.
70. M. E. Tuckerman, G. J. Martyna, and B. J. Berne. Molecular dynamics algorithm for condensed systems with multiple time scales. *J. Chem. Phys.*, 93:1287–1291, 1990.
71. G. Zhang and T. Schlick. LIN: A new algorithm combining implicit integration and normal mode techniques for molecular dynamics. *J. Comp. Chem.*, 14:1212–1233, 1993.
72. G. Zhang and T. Schlick. The Langevin/implicit-Euler/Normal-Mode scheme (LIN) for molecular dynamics at large time steps. *J. Chem. Phys.*, 101:4995–5012, 1994.
73. E. Barth, M. Mandziuk, and T. Schlick. A separating framework for increasing the timestep in molecular dynamics. In W. F. van Gunsteren, P. K. Weiner, and A. J. Wilkinson, editors, *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*, volume III, chapter 4, pages 97–121. ESCOM, Leiden, The Netherlands, 1997.
74. R. M. Levy, A. R. Srinivasan, W. K. Olson, and J. A. McCammon. Quasi-harmonic method for studying very low frequency modes in proteins. *Biopolymers*, 23:1099–1112, 1984.
75. B. R. Brooks, D. Janežič, and M. Karplus. Harmonic analysis of large systems. I. Methodology. *J. Comp. Chem.*, 16:1522–1542, 1995.
76. R. M. Levy, O. de la Luz Rojas, and R. A. Friesner. Quasi-harmonic method for calculating vibrational spectra from classical simulations on multidimensional anharmonic potential surfaces. *J. Phys. Chem.*, 88:4233–4238, 1984.
77. B. R. Brooks and M. Karplus. Normal modes for specific motions of macromolecules: Application to the hinge-bending mode of lysozyme. *Proc. Natl. Acad. Sci. USA*, 82:4995–4999, 1985.
78. M. Hao and S. C. Harvey. Analyzing the normal mode dynamics of macromolecules by the component synthesis method. *Biopolymers*, 32:1393–1405, 1992.

79. P. Dauber-Osguthorpe and D. J. Osguthorpe. Partitioning the motion in molecular dynamics simulations into characteristic modes of motion. *J. Comp. Chem.*, 14:1259–1271, 1993.
80. D. Perahia and L. Mouawad. Computation of low-frequency normal modes in macromolecules: Improvements to the method of diagonalization in a mixed basis and application to hemoglobin. *Comput. Chem.*, 19:241–246, 1995.
81. A. Amadei, A. B. M. Linssen, and H. J. C. Berendsen. Essential dynamics of proteins. *Proteins: Struc. Func. Gen.*, 17:412–425, 1993.
82. A. Amadei, A. B. M. Linssen, B. L. deGroot, D. M. F. van Aalten, and H. J. C. Berendsen. An efficient method for sampling the essential subspace of proteins. *J. Biomol. Struct. Dynam.*, 13:615–625, 1996.
83. M. A. Balsera, W. Wriggers, Y. Oono, and K. Schulten. Principal component analysis and long time protein dynamics. *J. Phys. Chem.*, 100:2567–2572, 1996.
84. D. Janežič and F. Merzel. An efficient symplectic integration algorithm for molecular dynamics simulations. *J. Chem. Info. Comp. Sci.*, 35:321–326, 1995.
85. J. D. Turner, P. K. Weiner, H. M. Chun, V. Lupi, S. Gallion, and U. C. Singh. Variable reduction techniques applied to molecular dynamics simulations. In W. F. van Gunsteren, P. K. Weiner, and A. J. Wilkinson, editors, *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*, volume 2, chapter 24. ESCOM, Leiden, The Netherlands, 1993.
86. E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems*, volume 14 of *Springer Series in Computational Mathematics*. Springer-Verlag, New York, New York, 1991.
87. B. García-Archilla, J.M. Sanz-Serna, and R.D. Skeel. Long-time-step methods for oscillatory differential equations. *SIAM J. Sci. Comp.*, 1996. To appear, [Also Tech. Rept. 1996/7, Dep. Math. Applic. Comput., Univ. Valladolid, Valladolid, Spain].
88. E. Barth and T. Schlick. Overcoming stability limitations in biomolecular dynamics: I. combining force splitting via extrapolation with Langevin dynamics in LN. *J. Chem. Phys.*, 109:1617–1632, 1998.
89. E. Barth and T. Schlick. Extrapolation versus impulse in multiple-timestepping schemes: II. linear analysis and applications to Newtonian and Langevin dynamics. *J. Chem. Phys.*, 109:1632–1642, 1998.
90. M. E. Tuckerman, B. J. Berne, and G. J. Martyna. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.*, 97:1990–2001, 1992.
91. A. Sandu and T. Schlick. Further analysis of impulse and extrapolation force splitting in molecular dynamics. *J. Comp. Phys.*, 1998. Submitted

Application of a Stochastic Path Integral Approach to the Computations of an Optimal Path and Ensembles of Trajectories*

Ron Elber[1], Benoit Roux[2], and Roberto Olender[1,3]

[1] Department of Physical Chemistry, and Department of Biological Chemistry, The Fritz Haber Research Center, and the Wolfson Center for Applied Structural Biology, The Hebrew University, Jerusalem 91904, Israel
[2] Department de Physique, Department de Chimie, Universite de Montreal, Case Postale 6128, Succursale Centre-Ville, Montreal, Quebec, Canada H3C 3J7
[3] Peptor Ltd., Kiryat Weizmann 16, Rehovot 76326, Israel

Abstract. A stochastic path integral is used to obtain approximate long time trajectories with an almost arbitrary time step. A detailed description of the formalism is provided and an extension that enables the calculations of transition rates is discussed.

1 Introduction

In numerous cases an atomically detailed picture is required to understand function of biological molecules. The wealth of atomic information that is provided by the Molecular Dynamics (MD) method is the prime reason for its popularity and numerous successes. The MD method offers: (a) qualitative understanding of atomic processes by detailed analysis of individual trajectories, and (b) comparison of computations to experimental data by averaging over a representative set of sampled trajectories.

Nevertheless, the technique suffers from a severe time scale problem - the trajectories are computed for (at most) a few nanoseconds. This is far too short compared to times required for many processes in biophysics. For example, the R to T conformational transition in hemoglobin lasts tens of microseconds [1], and the typical time for ion migration through the gramicidin channel is hundreds of nanoseconds. This limits (of course) our ability to make a meaningful comparison to experiments, using MD.

Extending time scales of Molecular Dynamics simulations is therefore one of the prime challenges of computational biophysics and attracted considerable attention [2–5]. Most efforts focus on improving algorithms for solving the initial value differential equations, which are in many cases, the Newton's equations of motion.

Here we suggest an alternative route to the problem in which the equations of motion are formulated as a boundary value problem. This limits the

* This research was supported by grants from the Israel Science Foundation, Israel Science Ministry and the National Institutes of Health to RE.

present approach to systems in which we know the start and the end points. Knowing the end points leads to problems that are different and complementary to the cases in which only the initial values are available. For example, conformational transitions between two known states of the protein are best approached by the proposed methodology. Conformational searches for low energy structures are best investigated by the methods based on initial values.

To perform the boundary value computations it is necessary to provide the length (in time) of the trajectory. The energy of the trajectory is not specified. It is therefore possible to compute a trajectory, which is very short in time but with high energy. Such high-energy trajectories are easy to detect and can be employed as starting paths for a refinement in which a longer trajectory (with hopefully lower energy) is calculated.

We further discuss how quantities typically measured in the experiment (such as a rate constant) can be computed with the new formalism. The computations are based on stochastic path integral formulation [6]. Two different sources for stochasticity are considered. The first (A) is randomness that is part of the mathematical modeling and is built into the differential equations of motion (e.g. the Langevin equation, or Brownian dynamics). The second (B) is the uncertainty in the approximate numerical solution of the *exact* equations of motion.

The two sources of stochasticity are conceptually and computationally quite distinct. In (A) we do not know the exact equations of motion and we solve instead phenomenological equations. There is no systematic way in which we can approach the exact equations of motion. For example, rarely in the Langevin approach the friction and the random force are extracted from a microscopic model. This makes it necessary to use a rather arbitrary selection of parameters, such as the amplitude of the random force or the friction coefficient. On the other hand, the equations in (B) are based on atomic information and it is the solution that is approximate. For example, to compute a trajectory we make the ad-hoc assumption of a Gaussian distribution of numerical errors. In the present article we also argue that because of practical reasons it is not possible to ignore the numerical errors, even in approach (A).

The difference between the two "philosophies" will become clearer as we continue (so we hope!). Nevertheless, it is useful to point out that a similar mathematical model is suggested for both conceptual approaches.

To exemplify both aspects of the formalism and for illustration purposes, we divide the present manuscript into two major parts. We start with calculations of trajectories using approximate solution of atomically detailed equations (approach B). We then proceed to derive the equations for the conditional probability from which a rate constant can be extracted. We end with a simple numerical example of trajectory optimization. More complex problems are (and will be) discussed elsewhere [7].

A lengthy and detailed description of the present methodology as applied to the solution of the Newton's equations of motion was published [7]. A

related discussion on Brownian trajectories can also be found [8]. Readers interested in further information are encouraged to use these references.

We recently received a preprint from Dellago et al. [9] that proposed an algorithm for path sampling, which is based on the Langevin equation (and is therefore in the spirit of approach (A) [8]). They further derive formulas to compute rate constants that are based on correlation functions. Their method of computing rate constants is an alternative approach to the formula for the state conditional probability derived in the present manuscript.

2 Method

2.1 General Considerations

We consider the computation of a trajectory $-X(t)$, where $X(t)$ is a vector of variables that evolve in time $-t$. The vector includes all the coordinates of the particles in the system and may include the velocities as well. Unless specifically indicated otherwise $X(t)$ includes coordinates only. The usual way in which such vectors are propagated numerically in time is via a sequence of short time solutions to a differential equation. One of the differential equations of prime concern is the Newton's equation of motion:

$$m\frac{d^2X(t)}{dt^2} = \frac{dU}{dX(t)} \tag{1}$$

where m is the mass (matrix), U is the potential energy and the derivative with respect to a vector denotes a gradient ($\frac{dU}{dX(t)} \equiv \frac{dU}{dX}|_{X=X(t)}$). An alternative equation (which we shall also consider) is of Brownian dynamics

$$\gamma\frac{dX(t)}{dt} = -\frac{dU}{dX(t)} + R(t) \tag{2}$$

where γ is the friction coefficient and $R(t)$ is a random force which is typically sampled from a Gaussian distribution so that the averages – $\langle \dots \rangle$ – are [10–12]

$$\langle R(t)\rangle = 0 \; and \; \langle R(t)R(t')\rangle = 2\gamma k_B T\delta(t-t') = \sigma^2\delta(t-t') \tag{3}$$

The Boltzmann constant is k_B and T the absolute temperature. $\delta(t-t')$ is the Dirac delta function. Below we assume for convenience (equation (5)) that the delta function is narrow, but not infinitely narrow. The random force has a zero mean and no correlation in time. For simplicity we further set the friction to be a scalar which is independent of time or coordinates.

We further note that the Langevin equation (which will not be discussed in detail here) is an intermediate between the Newton's equations and the Brownian dynamics. It includes in addition to an inertial part also a friction and a random force term:

$$m\frac{d^2X(t)}{dt^2} = -\frac{dU}{dX(t)} - \gamma\frac{dX(t)}{dt} + R(t)$$

It is useful to point out at present the important differences between equation (1) and equation (2). Equation (1) is using only microscopic information while in equation (2) a phenomenological variable is employed – the friction. Care must be used in the application of the phenomenological equation if the goal is the computations of the *microscopic rate*. The *microscopic friction* [13,14] is difficult (at the least) to compute exactly. Moreover, even the potential used in the Langevin equation may be phenomenological. The potential must be an effective energy averaged over (for example) the solvent coordinates. The averaging is replaced in many cases by a phenomenological expression. Therefore, determination of the *microscopic rate* from equations that employ friction may be problematic.

Another difference is related to the mathematical formulation. Equation (1) is deterministic and does not include explicit stochasticity. In contrast, the equations of motion for a Brownian particle include noise. Nevertheless, similar algorithms are adopted to solve the two differential equations as outlined below. The most common approach is to numerically integrate the above differential equations using small time steps and preset initial values.

There is a number of algorithms to solve equations (1) and (2) that differ appreciably in their properties which are beyond the scope of the present article. In the discussion below we use the velocity Verlet algorithm. However, better approaches can be employed [2–5]. We define "a rule" – $F(X(t), \Delta t)$ that modifies $X(t)$ to $X(t + \Delta t)$ and repeat the application of this rule as desired. For example the velocity Verlet algorithm ("rule") is:

$$\begin{aligned} X(t + \Delta t) &= X(t) + V(t)\delta t - \frac{\Delta t^2}{2} M^{-1} \frac{dU}{dX(t)} \\ V(t + \Delta t) &= V(t) - \frac{\Delta t}{2} M^{-1} \left(\frac{dU}{dX(t)} + \frac{dU}{dX(t + \Delta t)} \right) \end{aligned} \tag{4}$$

The coordinate vector is X and V is the velocity part. Δt is expected to be small.

3 The Onsager-Machlup Action for a Brownian Trajectory

A related algorithm can be written also for the Brownian trajectory [10]. However, the essential difference between an algorithm for a Brownian trajectory and equation (4) is that the Brownian algorithm is not deterministic. Due to the existence of the random force, we cannot be satisfied with a single trajectory, even with pre-specified coordinates (and velocities, if relevant). It is necessary to generate an ensemble of trajectories (sampled with different values of the random force) to obtain a complete picture. Instead of working with an ensemble of trajectories we prefer to work with the conditional probability. I.e., we ask "what is the probability that a trajectory being at

$X(t)$ at time t will be at $X(t+\Delta t)$ at time $t+\Delta t$. We denote the conditional probability by $P(X(t))|X(t+\Delta t);\Delta t$. For a Brownian trajectory the displacement is:

$$\Delta X(t) = X(t+\Delta t) - X(t) = \frac{1}{\gamma}\left[-\int_t^{t+\Delta t}\frac{dU}{dX(t')}dt' + \int_t^{t+\Delta t} R(t')dt'\right] \quad (5)$$

The second integral defines a new random variable $\eta(\Delta t) = \int_0^{\Delta t} R(t')dt'$ with a mean of zero. The probability distribution for $\eta(t)$ is a Gaussian and its variance is computed as

$$\begin{aligned}\langle\eta(\Delta t)\cdot\eta(\Delta t)\rangle &= \langle\int_0^{\Delta t} R(t')dt' \int_0^{\Delta t} R(t'')dt''\rangle = \int_0^{\Delta t}\int_0^{\Delta t}\langle R(t')R(t'')dt'dt'' \\ &= \int_0^{\Delta t}\int_0^{\Delta t}\sigma^2\delta(t'-t'')dt'dt'' = \sigma^2\Delta t\end{aligned} \quad (6)$$

The "$\cdot$" denotes a scalar product. $\eta(t)$ is also equal to $X(t+\Delta t) - X(t) + \frac{1}{\gamma}\int_t^{t+\Delta t}\frac{dU}{dX(t')}dt'$ (equation (5)). J – the Jacobian of transformation from $\eta(t)$ back to $X(t)$ is an issue of a possible concern (we thank Professor David Chandler for initiating the discussion on the Jacobian). The calculations of the Jacobian depend on the stochastic process that we would like to model and are performed on the discrete time representation (see next paragraph). It is possible to use two different kinds of stochastic calculus (the Ito or the Stratonovich calculus [11–12,15]) and to obtain a Jacobian, which is a constant, or not.

The stochastic differential equation and the second moment of the random force are insufficient to determine which calculus is to be preferred. The two calculus correspond to different physical models [11,12]. It is beyond the scope of the present article to describe the difference in details. We only note that the Ito calculus consider $\eta(t)$ to be a function of the edge of the interval while the Stratonovich calculus takes an average value. Hence, in the Ito calculus using a discrete representation $\eta(t)$ becomes $\eta(t_n) \approx \eta_n = \gamma(X_n - X_{n-1}) + \frac{dU}{dX_{n-1}}\Delta t$. Developing the determinant of the Jacobian – $J_{nm} \equiv \frac{\partial\eta_n}{\partial X_m}$ from the time origin (for example, the element $\frac{\partial\eta_n}{\partial X_n}$ is γ) we find it to be independent of X [15]. The constant Jacobian cancels when the normalized probability is considered. We therefore ignore the Jacobian in the rest of the article.

The conditional probability is

$$P(X_{n-1}|X_n;\Delta t) \propto e^{-\left(\left[\gamma(X_n - X_{n-1}) + \Delta t\frac{dU}{dX_{n-1}}\right]^2/2\Delta t\sigma^2\right)} \quad (7)$$

The definition of the above conditional probability for the case of Brownian trajectories can be found in textbooks [12]. However, the definition of the conditional probability for the Newton's equations of motion is subtler than that.

4 The Onsager Machlup Action for "Newtonian" Mechanics

Consider a numerical solution of the Newton's differential equation with a finite time step – Δt. In principle, since the Newton's equations of motion are deterministic the conditional probability should be a delta function

$$P(X(t)|X(t+\Delta t);\Delta t) = \delta[X(t+\Delta t) - F(X(t),\Delta t)] \tag{8}$$

We have in mind trajectory calculations in which the time step Δt is large and therefore the computed trajectory is unlikely to be the exact solution. Let $X_{num}(t)$ be the numerical solution as opposed to the "true" solution $X_{exact}(t)$. A plausible estimate of the errors in $X_{num}(t)$ can be obtained by plugging it back into the differential equation.

$$m\frac{d^2X_{num}(t)}{dt^2} + \frac{dU}{dX_{num}} \equiv Q(t) = \varepsilon(t) \tag{9}$$

The vector of errors at time t is $\varepsilon(t)$. The essential assumption which we made is that $\varepsilon(t)$ is a Gaussian random number, and that the errors are not correlated in time.

$$\langle\varepsilon(t)\rangle = 0 \; and \; \langle\varepsilon(t')\cdot\varepsilon(t)\rangle = \sigma^2\delta(t-t') \approx \delta_{t,t'}/\Delta t \tag{10}$$

The last "approximation" is for finite Δt. When the equations of motions are solved exactly, the model provides the correct answer ($\sigma^2 = 0$). When the time step is sufficiently large we argue below that equation (10) is still reasonable. The essential assumption is for the intermediate range of time steps for which the errors may maintain correlation. We do not consider instabilities of the numerical solution which are easy to detect, and in which the errors are clearly correlated even for large separation in time. Calculation of the correlation of the errors (as defined in equation (9)) can further test the assumption of no correlation of $\langle Q(t)Q(t')\rangle$.

We (of course) have no proof that the errors satisfy the above suggestions for arbitrary Δt. Nevertheless, qualitative arguments in favor of our assumption can be provided for relatively large time steps. Suppose that a solution $X_{num}(t)$ is obtained using time steps $\Delta t/(N_1\cdot N_2)$, where N_1 and N_2 are integers. If the time difference $\Delta t/N_1$ is large we may expect that the errors $\varepsilon(t)$ and $\varepsilon(t+\Delta t/N_1)$ are not correlated. The central limit theorem suggests that the errors separated by Δt (which are the sums of all the intermediate uncorrelated "random" numbers): $\varepsilon(t+\Delta) = \varepsilon(t)+\varepsilon(t+\Delta t/N_1)+\varepsilon(t+2\Delta t/N_1)+\ldots$ are sampled from a Gaussian distribution.

It remains to be seen, if the approximation using large time steps is reasonable. We shall show later the effect of the approximation on the power spectrum of the trajectory. More specifically, we shall demonstrate that large time steps filter out high frequency motions.

The probability of sampling $\varepsilon(t)$ is

$$P(\varepsilon(t)) \propto e^{\varepsilon(t)^2/2\langle\varepsilon^2\rangle} \propto e^{-\left[m\frac{d^2X_{num}(t)}{dt^2}+\frac{dU}{dX_{num}}\right]^2/2\langle\varepsilon^2\rangle} \tag{11}$$

Hence, we use the trajectory that was obtained by numerical means to estimate the accuracy of the solution. Of course, the smaller the time step is, the smaller is the variance, and the probability distribution of errors becomes narrower and concentrates around zero. Note also that the Jacobian of transformation from ε to X *must* be such that $\log[J]$ is independent of X at the limit of $\varepsilon \to 0$. Similarly to the discussion on the Brownian particle we consider the Ito Calculus [10–12] by a specific choice of the discrete time [15] – $\varepsilon_n = \frac{m}{\Delta t^2}[X_n + X_{n-2} - 2X_{n-1}] + \frac{dU}{dX_{n-2}}$. This choice is simple since the Jacobian is a constant. Since our knowledge on the properties of the errors is limited, our guess better be simple.

We assume that the sequential errors are not correlated in time, we can write the probability of sampling a sequence of errors as the product of the individual probabilities. We further use the finite time approximation for the delta function and have:

$$\Pi P(\varepsilon_i(t)) \propto e^{-\Sigma\left[m\frac{d^2X_{num}}{dt^2}+\frac{dU}{dX_{num}}\right]^2\Delta t/2\sigma^2} \tag{12}$$

In the continuum limit we define the probability of a path – $X_{num}(\tau)$. The probability is a functional of the path.

$$P[X_{num}(\tau)] \propto e^{-\frac{1}{2\sigma^2}\int_0^t\left(m\frac{d^2X(\tau)}{d\tau^2}+\frac{dU}{dX}\right)^2 d\tau} \equiv e^{-\frac{S}{2\sigma^2}} \approx e^{-\frac{1}{2\sigma^2}\Sigma\left(m\frac{d^2X(\tau)}{d\tau^2}+\frac{dU}{dX}\right)^2\Delta t} \tag{13}$$

S is the Onsager-Machlup action [6] that was derived originally from a different perspective of phenomenological and irreversible thermodynamics. Note also that this functional was used by Gauss to solve the equations of motion [16] (we thank Prof. David Shalloway for bringing the work of Gauss to our attention). We deliberately wrote the discrete form of the integral, which will be used in the optimization. We further comment that S is never negative, which makes it a convenient function for Monte-Carlo sampling or optimization. With pre- specified initial and final coordinates, it is now possible to ask, "What is the most probable trajectory that will connect the reactant and the product?" The most probable trajectory will be the one that minimizes S.

Clearly, a single trajectory does not necessarily provide a good representation of the system configurations and properties. Trajectory sampling

is therefore a more appropriate approach. Nevertheless, the most probable trajectory may be a useful tool in qualitative analysis of "typical" trajectories and in providing a sampling bias for domains that correspond to more probable paths.

In the language of conditional probabilities that was used in equation (6) we can use (12) to compute (unnormalized) conditional probability of going from $X(0)$ to $X(t)$ that is, $P(X(0)|X(t);t)$

$$P(X(0)|X(t);t) \propto \int_{X(0)}^{X(t)} D[X(\tau)] e^{-S[X(\tau)]/2\sigma^2} \tag{14}$$

$D[X(\tau)]$ is used to denote a path integral. Hence, equation (14) corresponds to a summation of all paths leading from $X(0)$ to $X(t)$. The same expression is used for the Brownian trajectories and for Newtonian's trajectories with errors. The action is of course different in both cases.

A saddle point approximation to the above integral provides the definition for optimal trajectories. The computations of most probable trajectories were discussed at length [1]. We consider the optimization of a discrete version of the action.

$$S \cong \sum_{i=1}^{N} \left[\frac{m}{\Delta t^2}(X_{i+1} + X_{i-1} - 2X_i) + \frac{dU}{dX_{i-1}} \right]^2 \Delta t \tag{15}$$

A finite difference formula is used to estimate the second derivatives of the coordinate vector with respect to time and S is now a *function* of all the intermediate coordinate sets. An optimal value of S can be found by a direct minimization, by multi-grid techniques, or by an annealing protocol [7]. We employed in the optimization analytical derivatives of S with respect to all the X_j-s.

Since S is a function of *all* the intermediate coordinates, a large scale optimization problem is to be expected. For illustration purposes consider a molecular system of 100 degrees of freedom. To account for 1000 time points we need to optimize S as a function of 100,000 independent variables (!). As a result, the use of a large time step is not only a computational benefit but is also a necessity for the proposed approach. The use of a small time step to obtain a trajectory with accuracy comparable to that of Molecular Dynamics is not practical for systems with more than a few degrees of freedom. For small time steps, ordinary solution of classical trajectories is the method of choice.

5 Filtering High Frequency Motions

It is also interesting to examine the behavior of the optimized solution as a function of the step size. As discussed below, the proposed algorithm is very

stable with respect to the step size. Numerical algorithms that propagate initial values of coordinates and velocities loose stability quite quickly as Δt increases. The present optimization protocol maintains its stability for almost an arbitrary choice of the time step. Nevertheless, the optimized set of coordinates provides only an approximation to the true trajectory. Can we suggest how the approximate solution is related to the true solution?

Consider a three-point approximation to the path:

$$S \cong \left[\frac{m}{\Delta t^2}(X_2 + X_0 - 2X_1) + \frac{dU}{dX_0}\right]^2 \Delta t \tag{16}$$

S is a non-linear function of the vector X_1. (X_0 and X_2 are held fixed during the variation). Clearly the three vectors are insufficient to describe the path accurately if the time step is large. Nevertheless, the solution with a large Δt has interesting characteristics discussed below. We expand the difference between the exact solution and the three-point approximation in a Fourier series. The solution to the three-point approximation is represented as a parabola in time, and a Fourier series represents the remainder of the path. We now have for $X(t)$

$$X(t) = a + bt + \frac{1}{2}ct^2 + \sum_k d_k \sin(\omega_k t); \quad \left(\omega = \frac{k\pi}{\Delta t}\right) \tag{17}$$

We use the sine series since the end points are set to satisfy exactly the three-point expansion [7]. The Fourier series with the pre-specified boundary conditions is complete. Therefore, the above expansion provides a trajectory that can be made exact. In addition to the parameters a, b and c (which are determined by X_0, X_1 and X_2) we also need to calculate an infinite number of Fourier coefficients – $\{d_k\}$. In principle, the way to proceed is to plug the expression for $X(t)$ (equation (17)) into the expression for the action S as defined in equation (13), to compute the integral, and optimize the Onsager-Machlup action with respect to all of the path parameters.

The problem is (of course) the computation of these integrals. The integral over the time derivatives can be performed analytically. However, this is not the case for integrals that include the force, which is a general and complex function of the coordinates. We therefore use the same type of approximation that we employ in the discrete optimization of the path. That is, we assume a constant value of the force during the time interval Δt. Within the constant force approximation, the integral can be evaluated analytically and then optimized. It is evident however, that our analysis is not a "true" estimate of the errors, since we employ an approximation. The approximation is used also in the computations. Therefore, rather than providing a true estimate of the accuracy of the solution, it tells us something on the properties

of the approximate trajectory.

$$
\begin{aligned}
S &= \int_0^{\Delta t} \left[m\frac{d^2X(t)}{dt^2} + \frac{dU}{dX(t)} \right]^2 dt \\
&= \int_0^{\Delta t} \left[m\frac{d^2X(t)}{dt^2} \right]^t \left[m\frac{d^2X(t)}{dt^2} \right] dt + 2\int_0^{\Delta t} \left[m\frac{d^2X(t)}{dt^2} \right]^t \left[\frac{dU}{dX(t)} \right] dt \\
&\quad + \int_0^{\Delta t} \left[\frac{dU}{dX(t)} \right]^t \left[\frac{dU}{dX(t)} \right] dt
\end{aligned}
\tag{18}
$$

The square brackets denote a vector, and $[\,]^t$ a transposed vector. The exact expression for the Onsager-Machlup action is now approximated by

$$
\begin{aligned}
S \approx\ & [cm]^t[cm]\Delta t + \sum_{k=1}^{\infty} [d_k m]^t [d_k m] \left(\frac{k\pi}{\Delta t} \right)^4 \\
& + 2[cm]^t \left[\frac{dU}{dX_0} \right] \Delta t + \left[\frac{dU}{dX_0} \right]^t \left[\frac{dU}{dX_0} \right] \Delta t
\end{aligned}
\tag{19}
$$

In the derivation we used the exact expansion for $X(t)$, but an approximate expression for the last two integrals, in which we approximate the potential derivative by a constant at X_0. The optimization of the action S with respect to all the Fourier coefficients, shows that the action is optimal when all the d_k are zero. These coefficients correspond to frequencies larger than $\approx \pi/\Delta t$. Therefore, the optimal solution does not contain contributions from these modes. Elimination of the fast modes from a trajectory, which are thought to be less relevant to the long time scale behavior of a dynamical system, has been the goal of numerous previous studies.

For example, the SHAKE algorithm [17] freezes out particular motions, such as bond stretching, using holonomic constraints. One of the differences between SHAKE and the present approach is that in SHAKE we have to know in advance the identity of the fast modes. No such restriction is imposed in the present investigation. Another related algorithm is the Backward Euler approach [18], in which a Langevin equation is solved and the slow modes are constantly cooled down. However, the Backward Euler scheme employs an initial value solver of the differential equation and therefore the increase in step size is limited.

One may argue that the high frequency motions are "filtered" since the time step is too large and the discrete trajectory could not possibly follow the fast modes; hence, that the "filtering" is trivial. Of course we cannot follow fast motions if the time step we use is larger than their typical period. Nevertheless, these fast motions are the prime problem in attempts to increase the time step. We cannot just ignore them in the initial value approach since their presence introduces instabilities to the solution. The optimization of the Onsager-Machlup action is considerably more stable. We

provide in Fig. 1 a simple demonstration of the filtering effect, examining results for the harmonic oscillator. The filtering makes it possible for us to use almost arbitrary time steps, while keeping in mind that elimination of high frequency motions, (which is intuitively appealing), does not necessarily result in the same trajectory for the slow modes.

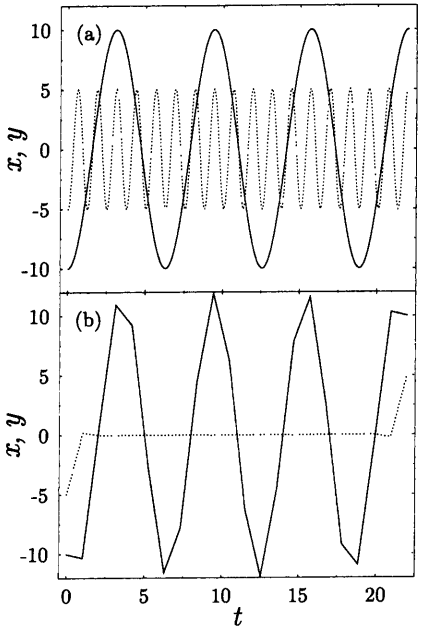


Fig. 1. Optimization of the Onsager-Machlup action for the two dimensional harmonic oscillator. The potential energy is $U(x,y) = \frac{1}{2}(x^2 + 25y^2)$, the mass is 1 and 3.5 slow oscillations are considered. The boundary conditions are $X(0) = -X(t_{end}) = -10$, and $y(0) = -y(t_{end}) = -5$, where t_{end} is the end time of the trajectory. Two trajectories are shown. The first consists of 500 intermediate points (dotted lines) which provide an essentially exact solution. The second trajectory employed only 20 intermediate points (solid lines). Note that for the solution with 20 points the fast oscillations are frozen while the slow oscillations are reasonably well reproduced.

To improve the accuracy of the solution, the size of the time step may be decreased. The smaller is the time step, the smaller are the assumed errors in the trajectory. Hence, in contrast (for example) to the Langevin equation that includes the friction as a phenomenological parameter, we have here a systematic way of approaching a microscopic solution. Nevertheless, some problems remain. For a very large time step, it is not clear how relevant is the optimal trajectory to the reality, since the path variance also becomes large. Further-

more, the trajectories computed for biological systems are usually chaotic, and the results are sensitive to the initial conditions. It is therefore necessary to generate an ensemble of trajectories with different initial conditions, taking into account the width of a "pipe" that includes plausible paths. Only an ensemble of trajectories will enable the computations of experimentally observed phenomena.

6 Weights of Individual Trajectories

We shift the discussion from computations of optimal trajectories to sampling of trajectories with a given weight. Hence rather than seeking a single trajectory with a maximal $e^{-S/2\sigma^2}$, we consider in addition other trajectories with potentially lower weights but non-negligible contributions to the conditional probability. For that purpose we need to provide an estimate to $2\sigma^2$. This is the goal of the present section.

As discussed above the errors in the trajectory are correlated with the missing rapid motions. In contrast to the friction approach of estimating the variance, which may affect long time phenomena, we identify our errors as the missing ("filtered") high frequency modes. We therefore attempt to account approximately for the fast motions by choosing the trajectory variance accordingly.

The errors in the present stochastic path formalism reflect short time information rather than long time information. Short time data are easier to extract from atomically detailed simulations. We set the second moment of the errors in the trajectory – σ^2 – to a constant (i.e. the same value for all degrees of freedom). This is an approximation similar in spirit to the application of a constant scalar friction in the Langevin dynamics.

We proposed [7] two possible approaches to estimate these errors. Here we discuss them only briefly. Trajectories that are not "too far" from the optimal trajectory will have a significant weight. We denote by $X_{opt}(t)$ the optimized trajectory, and by $X_{exact}(t)$ the exact trajectory. The optimal trajectory is not the same as the exact trajectory, since it was computed with a large time step. $S/2\sigma^2$ is expanded up to a second order near the optimal trajectory

$$S_{opt}[X_{exact}(t)] \approx \frac{1}{2}[X_{exact}(t) - X_{opt}(t)]^t \frac{\delta^2 S}{\delta X_{opt}^2(t)}[X_{exact}(t) - X_{opt}(t)] \quad (20)$$

We then require that the exact trajectory will have a weight of at least $\frac{1}{e}$, which will make its sampling possible in a search biased by the weight.

$$\frac{P[X_{opt}(t)]}{P[X_{exact}(t)]} = e^{\frac{-1}{2\sigma^2}\sum \frac{1}{2}[X_{exact}(t_i)-X_{opt}(t_i)]^t \frac{\delta^2 S}{\delta X_{opt}^2(t_i)}[X_{exact}(t_i)-X_{opt}(t_i)]\Delta t} \approx \frac{1}{e}$$

$$2\sigma^2 \approx \sum \frac{1}{2}[X_{exact}(t_i) - X_{opt}(t_i)]^t \frac{\delta^2 S}{\delta X_{opt}^2(t)}[X_{exact}(t_i) - X_{opt}(t_i)]\Delta t \quad (21)$$

To compute the above expression, short molecular dynamics runs (with a small time step) are calculated and serve as exact trajectories. Using the exact trajectory as an initial guess for path optimization (with a large time step) we optimize a discrete Onsager-Machlup path. The variation of the action with respect to the optimal trajectory is computed and used in the above formula.

It is also possible to use normal mode analysis [7] to estimate the difference between the exact and the optimal trajectories. Yet another formula is based on the difference between the optimal and the exact actions: $2\sigma \approx [S[X_{exact}(t)] - S[X_{opt}(t)]]$. The action is computed (of course), employing a large time step and the exact solution as a starting point.

In spite of the three methods of estimating σ^2 that were briefly described above, it is clear that the computation of the variance of the path is still an open question that needs to be investigated further in the future. The above expressions are order of magnitude estimates and not exact formulas.

7 Computation of the State Conditional Probability

Once numerical estimates of the weight of a trajectory and its variance ($2\sigma^2$) are known we are able to use sampled trajectories to compute observables of interest. One such quantity on which this section is focused is the rate of transitions between two states in the system. We examine the transition between a domain A and a domain B, where the A domain is characterized by an inverse temperature – β. The weight of an individual trajectory which is initiated at the A domain and of a total time length – $N\Delta t$ is therefore

$$P[X(t)] \approx e^{\beta E(X_{0A}(t),V_{0A}(t))} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\left[\frac{m}{\Delta t^2}(X_{i+1}+X_{i-1}-2X_i)+\frac{dU}{dX_{i-1}}\right]^2\Delta t} \tag{22}$$

The initial energy – $E(X_{0A}(t), V_{0A}(t))$ – is a function of the coordinates and the velocities. In principle, the use of momenta (instead of velocities) is more precise, however, we are using only Cartesian coordinates, making the two interchangeable. We need to sample many paths to compute ensemble averages. Perhaps the most direct observable that can be computed (and measured experimentally) is the *state* conditional probability – $P(A|B,t)$ defined below:

$$P[A_0|B,t] = \int_{X_0\in A} dX_0\, dV_0\, e^{-\beta E_0} \frac{\int_{X\in B} dX[P(X_0|X,t)]}{\int_{X\in A+B} dX[P(X_0|X,t)]} \Big/ \int_{X_0\in A} dX_0\, dV_0\, e^{-\beta E_0} \tag{23}$$

This is the conditional probability that the system which was in state A at time zero will be in state B at time t. Note that we use the normalized conditional probability since the trajectory must end either at A or at B.

This function can be used to compute many quantities besides the rate, and is formally "cleaner" than the rate constant discussed below.

The difficulties in performing a numerical calculation of a "rate constant" are two fold: First, similarly to other sampling procedures by Monte-Carlo or Molecular Dynamics approaches, we do not know the normalization of the weight defined in equation (22). Second, we do not have a rigorous formula for the rate. This is similar to previous microscopic expressions for the rate constant, which were derived assuming the existence of a phenomenological rate law [19]. The existence of the exponential decay (the first order rate law) cannot be proven in general, since it is system dependent. We provide below a short description of the phenomenology that we employ in order to extract expressions that are *related* to the rate. We derive exact expressions for some quantities of interest, and the results can be related to a rate constant. However, it is not necessary to do so.

We consider a two state system, state A and state B. A state is defined as a domain in phase space that is (at least) in local equilibrium since thermodynamic variables are assigned to it. We assume that A or B are described by a local canonical ensemble. There are no dark or hidden states and the probability of the system to be in either A or in B is one. A phenomenological rate equation that describes the transitions between A and B is

$$\begin{aligned} & A \underset{k_2}{\overset{k_1}{\leftrightarrow}} B \\ & \frac{dP(A)}{dt} = -k_1 P(A) + k_2 P(B) \\ & \frac{dP(B)}{dt} = k_2 P(B) - k_1 P(A) \end{aligned} \tag{24}$$

$P(A)$ is the probability of observing the system in state A, and $P(B)$ is the probability of observing state B. In this model, the space is divided exactly into A and B. The dividing hyper-surface between the two is employed in Transition State Theory for rate calculations [19]. The identification of the dividing surface, which is usually assumed to depend on coordinates only, is a non-trivial task. Moreover, in principle, the dividing surface is a function of the whole phase space – coordinates and velocities, and therefore the exact calculation of it can be even more complex. Nevertheless, it is a crucial ingredient of the Transition State Theory and variants of it.

The dividing surface is not a major concern of the methodology below which makes it different and perhaps complementary to the usual picture for barrier crossing (the dividing surface is required in order to identify the A and the B states). For example, the present approach is not limited to a single barrier crossing, and multiple crossings are treated by the same protocol. Moreover, even a very rough energy surface with a multitude of minima and barriers is accessible to the present methodology. A possible disadvantage of the present approach is that fast time scales with a period smaller than Δt

are filtered out. If anything interesting happens quickly, we may miss it and will observe only the start and the end points.

In physical systems, in the absence of external forces, A and B approach equilibrium:

$$\frac{dP_{eq}(A)}{dt} = 0 \qquad \frac{dP_{eq}(B)}{dt} = 0 \qquad \frac{P_{eq}(A)}{P_{eq}(B)} = \frac{k_2}{k_1} \tag{25}$$

The initial conditions are $P(A, t = 0) = A_0$ and $P(B, t = 0) = 0$

In the "conditional probability" language we write for the above phenomenological equation:

$$\begin{aligned} P(A_0|A,t) &= P(A_{eq}) + (1 - P(A_{eq}))e^{-kt} \\ P(A_0|B,t) &= P(B_{eq})(1 - e^{-kt}) \\ k &= k_1 + k_2 \end{aligned} \tag{26}$$

The above phenomenological equations are assumed to hold in our system as well (after appropriate averaging). Below we derive formulas for $P(A_0|B,t)$, which start from a microscopic model and therefore makes it possible to compare the same quantity with the above phenomenological equation. We also note that the formulas below are, in principle, exact. Therefore tests of the existence of a "rate constant" and the validity of the above model can be made. We rewrite the state conditional probability with the help of a step function – $H_B(X)$. $H_B(X)$ is zero when X is in A and is one when X is in B.

$$P[A_0|B,t] = \int\limits_{X_0 \in A} dX_0\, e^{-\beta E_0} \frac{\int\limits_{X \in A+B} dX[H_B(X)P(X_0|X,t)]}{\int\limits_{X \in A+B} dX[P(X_0|X,t)]} \Big/ \int\limits_{X_0 \in A} dX_0\, e^{-\beta E_0} \tag{27}$$

We set the initial distribution to be canonical, however, other distributions may be used as well. We further define the average with respect to the path:

$$\langle H_B(X)\rangle_s = \frac{\iint\limits_{X \in A+B} e^{-S/2\sigma^2} H_B(X) D[X(\tau)]dX}{\iint\limits_{X \in A+B} e^{-S/2\sigma^2} D[X(\tau)]dX} \tag{28}$$

The average of the step function, using the action for "a Boltzmann weight" can be pursued by standard statistical mechanics. It may require more elaborate sampling techniques such as the Umbrella sampling [20].

The state conditional probability is therefore written in terms of the computable, average step function:

$$P(A_0|B,t) = \frac{\int dX_0 e^{-\beta E_0}\langle H_B\rangle_S}{\int dX_0 e^{-\beta E_0}} \tag{29}$$

The additional integration in the final equation is similar to integration of potential of mean force. Equation (29) is more difficult to compute since X_0 is a multidimensional vector.

A single calculation of the discrete path integral with a fixed length of time t can be employed to compute the state conditional probability at many other times. It is possible to use segments of the path of time length $\Delta t, 2\Delta t, \ldots, N\Delta t$ sampled in trajectories of total length of $N\Delta t$ and to compute the corresponding state conditional probabilities. The result of the calculations will make it possible to explore the exponential relaxation of $P(A_0|B,t)$ for times between 0 and t.

8 A Numerical Example

We describe a simple computational example to demonstrate two key features of the new protocol: Stability with respect to a large time step and filtering of high frequency modes. In the present manuscript we do not discuss examples of rate calculations. These calculations will be described in future publications.

The model consists of a two dimensional harmonic oscillator with mass 1 and force constants of 1 and 25. In Fig. 1 we show trajectories of the two oscillators computed with two time steps. When the time step is sufficiently small compared to the period of the fast oscillator an essentially exact result is obtained. If the time step is large then only the slow vibration persists, and is quite accurate. The filtering effect is consistent (of course) with our analytical analysis. Similar effects were demonstrated for more complex systems [7].

The optimization of the action was pursued in the above cases using straightforward conjugate gradient minimization. The initial guess of the path was a straight line, interpolating the two end points. For systems with a linear force (as the harmonic oscillators), there is no need for more sophisticated approaches, since the functional is parabolic in the intermediate coordinate sets. It is important to stress however that for a more general path that includes non-linear forces the optimization is a non-trivial computation. We have tried and employed different tools (discussed below) in cases that require more global optimization. However, it is very likely that the existing set of tools can be expanded and enhanced further.

For more demanding surfaces of actions we also used multi-grid techniques [21] and global optimization methods such as simulated annealing [22]. The multi-grid approach made a profound improvement in the speed and accuracy of the optimization of the path and we therefore describe below the protocol that we used. We employed a V cycle for the multi-grid refinement. First an initial guess for a path with the highest density of points was briefly minimized. In the next step some intermediate points were removed and the path spatial resolution was decreased. An optimization of the lower resolution path followed, and the result was used for further path refinements. The

minimizations focus at first on a sequence of paths with reduced spatial resolution and then on a sequence of paths with an increased density of grid points ("V" cycle). Empirically this procedure was shown to provide significantly better results compared to one time optimization with a single and fixed spatial resolution.

9 Summary

We outlined the principles of a novel approach of computing approximate trajectories with a very large time step. Perhaps, the most interesting feature of the suggested protocol is its stability with respect to variations of the step size. In contrast to initial value problems, we do not integrate differential equations but employ an optimization procedure of the whole trajectory. Since the whole trajectory is studied simultaneously, the present protocol is considerably more demanding on computer memory. The significantly larger system (as compared to a single coordinate set or a few coordinate sets required by common integrators) is another significant computational burden that needs to be addressed. Nevertheless, the possibility of obtaining approximate trajectories with profoundly larger time steps than was possible before is intriguing and encouraging, and is therefore worth further investigation.

The new formalism is especially useful for parallel and distributed computers, since the communication intensity is exceptionally low and excellent load balancing is easy to achieve. In fact, we have used cluster of workstations (Silicon Graphics) and parallel computers – Terra 2000 and IBM SP/2 – to study dynamics of proteins.

We further comment that reactive trajectories that successfully pass over large barriers are straightforward to compute with the present approach, which is based on boundary conditions. The task is considerably more difficult with initial value formulation.

The essential assumption of this manuscript is the existence of a constant variance of Gaussian errors along the trajectory. While we attempted to correlate the variance with the high frequency motions, many uncertainties and questions remain. These are topics for future research.

If the above assumption is reasonable, then the modeling of most probable trajectories and of ensembles of trajectories is possible. We further discussed the calculations of the state conditional probability and the connection of the conditional probability to rate constants and phenomenological models.

References

1. J. Hofrichter, J.H. Sommer, E.R. Henry and W.A. Eaton, Proc. Natl. Acad. Sci. USA 80, 2235 (1983)
2. M.E. Tuckerman, G.J. Martyna and B.J. Berne, J. Chem. Phys., 97, 1990-2001 (1992); F. Figueirido, R. Zhou, B.J. Berne and Ronald M. Levy, J. Chem. Phys. 106, 9835 9849 (1997)

3. P. Derreumaux and T. Schlick, Proteins 21, 282 (1995)
4. D. Okunbor and R.D. Skeel, J. Computational Chemistry, 15, 72–79 (1994)
5. H. Grubmuller, H. Heller, A. Windemuth and K. Schulten, Molecular Simulations, 6, 121–142 (1991)
6. L. Onsager and S. Machlup, Phys. Rev. 91, 1505 (1953); S. Machlup and L. Onsager, ibid., 91, 1512 (1953)
7. R. Olender and R. Elber, J. Chem. Phys. 105, 9299 (1996)
8. R. Olender and R. Elber, "Yet another look at the steepest descent path", J. Mol. Struct. Theochem and the proceeding of the WATOC symposium, 398–399, 63–72 (1997)
9. C. Dellago, P.G. Bolhuis, F.S. Csajka, and D. Chandler, "Transition path sampling and the calculation of rate constants", a preprint.
10. B. Oksendal, "Stochastic differential equations: An introduction with applications", Springer-Verlag, Berlin, 1995
11. H. Risken, "The Fokker-Planck equation: Methods of solution and applications", Springer- Verlag, Berling, 1984, chapter 3.
12. C.W. Gardiner, "Handbook of stochastic methods for physics, chemistry and natural sciences", Springer-Verlag, Berlin, 1990
13. B. J. Berne, M. E. Tuckerman, J.E. Straub and A.L.R. Bug, J. Chem. Phys. 93, 5084 (1990)
14. R. Elber, D. P. Chen, D. Rojewska, and R. S. Eisenberg, Biophys. J. 68, 906–924 (1995)
15. H. Kleinert, "Path integrals in quantum mechanics, statistics and polymer physics", World Scientific, Singapore, 1995, chapters 18.5 and 18.6.
16. C. Lanczos, "The Variational Principles of Mechanics", University of Toronto Press, Toronto, 1970
17. J.P. Ryckaert, G. Ciccotti and H.J.C. Berendsen, J. of Comput. Physics 23, 327 (1977)
18. C.S. Peskin and T. Schlick, Communications on Pure and Applied Mathematics, XLII, 1001 (1989); G. Zhang and T. Schlick, J. Comp. Chem. 14, 121 (1993); G. Zhang and T. Schlick, J. Chem. Phys., 101, 4995 (1994)
19. S. Glasstone, K.J. Laidler and H. Eyring, "The theory of rate processes", McGraw-Hill, New York, 1961
20. J.P. Valleau and G.M. Torrie, "A Guide to Monte Carlo for Statistical Mechanics: 2. Byways", in "Statistical Mechanics", Ed. B. Berne, Plenum Press, New York, 1977
21. S.F. McCormick, Editor, "Multigrid Methods", SIAM, Philadelphia, 1987
22. S. Kirkpatrick, Jr., C.D. Gelatt and M.P. Vecchi, Science 220, 671 (1983)

On Some Difficulties in Integrating Highly Oscillatory Hamiltonian Systems

Uri M. Ascher[1]* and Sebastian Reich[2]

[1] Institute of Applied Mathematics and Department of Computer Science, University of British Columbia, Vancouver, B.C., Canada V6T 1Z4 (ascher@cs.ubc.ca)
[2] Konrad-Zuse-Zentrum, Takustr. 7, D-14195 Berlin, Germany (reich@zib.de)

Abstract. The numerical integration of highly oscillatory Hamiltonian systems, such as those arising in molecular dynamics or Hamiltonian partial differential equations, is a challenging task. Various methods have been suggested to overcome the step-size restrictions of explicit methods such as the Verlet method. Among these are multiple-time-stepping, constrained dynamics, and implicit methods. In this paper, we investigate the suitability of time-reversible, semi-implicit methods. Here semi-implicit means that only the highly oscillatory part is integrated by an implicit method such as the midpoint method or an energy-conserving variant of it. The hope is that such methods will allow one to use a step-size k which is much larger than the period ε of the fast oscillations.

However, our results are not encouraging. Even in the absence of resonance-type instabilities, we show that in general one must require that k^2/ε be small enough. Otherwise the method might become unstable and/or it might lead to a wrong approximation of the slowly varying solution components. The latter situation might, in some cases, even require that k/ε be small in order to avoid this danger. While certain (semi-implicit) energy conserving methods prove to be robust for some model problems, they may also yield deceptively-looking, wrong solutions for other simple model problems, in circumstances where the corresponding constrained dynamics formulation may not be easily derived and used.

1 Introduction

In this paper, we discuss semi-implicit/implicit integration methods for highly oscillatory Hamiltonian systems. Such systems arise, for example, in molecular dynamics [1] and in the finite dimensional truncation of Hamiltonian partial differential equations. Classical discretization methods, such as the Verlet method [19], require step-sizes k smaller than the period ε of the fast oscillations. Then these methods find pointwise accurate approximate solutions. But the time-step restriction implies an enormous computational burden. Furthermore, in many cases the high-frequency responses are of little or no interest. Consequently, various researchers have considered the use of semi-implicit/implicit methods, e.g. [6, 11, 9, 16, 18, 12, 13, 8, 17, 3].

* The work of this author was partially supported under NSERC Canada Grant OGP0004306.

A popular implicit discretization is the (implicit) midpoint method [7] which, applied to a system of the type

$$\frac{d}{dt}\boldsymbol{q} = \boldsymbol{p}\,, \tag{1a}$$

$$\frac{d}{dt}\boldsymbol{p} = -\nabla U(\boldsymbol{q})\,, \tag{1b}$$

yields the discretization

$$\boldsymbol{q}_{n+1} = \boldsymbol{q}_n + k\,\boldsymbol{p}_{n+1/2}\,, \tag{2a}$$

$$\boldsymbol{p}_{n+1} = \boldsymbol{p}_n - k\,\nabla U(\boldsymbol{q}_{n+1/2}) \tag{2b}$$

with $\boldsymbol{p}_{n+1/2} = [\boldsymbol{p}_{n+1} + \boldsymbol{p}_n]/2$, etc. The behavior of the midpoint method when applied to highly-oscillatory systems

$$\frac{d}{dt}\boldsymbol{q} = \boldsymbol{p}\,, \tag{3a}$$

$$\frac{d}{dt}\boldsymbol{p} = -\nabla W(\boldsymbol{q}) - \kappa g(\boldsymbol{q})\nabla g(\boldsymbol{q})\,, \tag{3b}$$

W, g potential functions, $\kappa \gg 1$, has been discussed in various papers (see, for example, [6, 11, 9, 16, 3]). It has been pointed out that, for step-sizes $k \gg \varepsilon := 1/\sqrt{\kappa}$, the midpoint method can become unstable due to resonances [9, 16], i.e., for specific values of k. However, generic instabilities arise if the step-size k is chosen such that k^2/ε is not small [3, 6, 18]. For systems with a rotational symmetry this has been shown rigorously in [6]. This effect is generic for highly oscillatory Hamiltonian systems, as argued for in [3] in terms of decoupling transformations and proved for a linear time varying system without symmetry.

Even further complications are to be expected for general systems of the type (3). These are related to the approximation of the slowly varying solution components and other related quantities of (3) for $\kappa \to \infty$ by the corresponding solution of the constrained system DAE

$$\begin{aligned} \frac{d}{dt}\boldsymbol{q} &= \boldsymbol{p}\,, \\ \frac{d}{dt}\boldsymbol{p} &= -\nabla W(\boldsymbol{q}) - \lambda\nabla g(\boldsymbol{q})\,, \\ 0 &= g(\boldsymbol{q})\,. \end{aligned} \tag{4}$$

In general, the solution components of the DAE (4) are the correct limits (as $\kappa \to \infty$) of the corresponding slowly varying solution components of the free dynamics only if an additional (conservative) force term is introduced in the constrained system [14, 5]. It turns out [3] that the midpoint method may falsely approximate this correcting force term to zero unless $k = \mathcal{O}(\varepsilon)$, which leads to a step-size restriction of the same order of magnitude as explicit

methods such as the Verlet method! In cases where the additional force term is zero this further complication does not apply, of course. The specific form of the additional force term depends on the initial conditions and the potential function g in (3a)-(3b).

Since fully implicit methods are very expensive when long-range forces are present, we consider semi-implicit methods that are based on the splitting of the force field into weak forces and strong forces, the latter causing the high-frequency oscillations [12]. The limitations of the methods considered here are not related to this splitting, i.e., corresponding fully implicit methods would not remove these limitations.

The purpose of this paper is twofold: (i) We summarize possible difficulties with the midpoint method (other than resonance instability, which has been treated extensively elsewhere) by looking at a simple (molecular) model problem. (ii) We investigate the suitability of some energy conserving methods.

Regarding the second objective (ii), it has been argued [6] that energy conserving methods might be better suited for the numerical integration of highly oscillatory problems such as (3). We show that such a statement must be made with extreme caution. The main concern is that energy conserving methods might wrongly approximate slowly varying solution quantities. Even worse, unlike the midpoint method where a blatant non-conservation of energy indicates potential trouble, energy conserving methods, by definition, do not provide such a warning flag. Thus, energy conserving methods are potentially more misleading on one hand, and they do not seem to provide a significant gain on the other hand. Consequently, we generally discourage their use for highly oscillatory Hamiltonian systems, unless special circumstances warrant it.

2 Some Energy Conserving Methods

Given a general autonomous, separable Hamiltonian system (1), the Hamiltonian

$$H(\mathbf{q}, \mathbf{p}) = \frac{1}{2}\mathbf{p}^T\mathbf{p} + U(\mathbf{q})$$

remains constant, of course, along exact solution trajectories. Here we wish an approximate, numerical solution to preserve this invariant as well, i.e.,

$$H(\mathbf{q}_{n+1}, \mathbf{p}_{n+1}) = H(\mathbf{q}_n, \mathbf{p}_n), \quad n = 0, 1, \ldots . \tag{5}$$

Moreover, we want the resulting numerical method to remain time-reversible, which precludes the most obvious projection schemes from further consideration.

Time-reversible energy conserving methods can be obtained by appropriate modifications to the (time-reversible) midpoint method. Two such modifications are: (i) scaling of the force field by a scalar such that total energy

is conserved [6, 12], and (ii) using a symmetric projection technique onto the hypersurface of constant energy. Here are the details.

2.1 The Simo-Gonzales Method

The method considered in [6] can be derived as follows. Let us assume that $U(\mathbf{q}) = V(r)$, where $r = \|\mathbf{q}\|$. Then (1b) reads

$$\frac{d}{dt}\boldsymbol{p} = -\frac{V'(r)}{r}\boldsymbol{q}\,.$$

Consider for the nth time step the linear harmonic oscillator

$$\bar{H}(\mathbf{q},\mathbf{p}) = \frac{1}{2}\mathbf{p}^T\mathbf{p} + \frac{\sigma}{2}\mathbf{q}^T\mathbf{q}$$

where $\sigma = \sigma_n$ is constant. Discretizing the equations of motion using the midpoint method

$$\mathbf{q}_{n+1} = \mathbf{q}_n + k\mathbf{p}_{n+1/2} \tag{6a}$$

$$\mathbf{p}_{n+1} = \mathbf{p}_n - k\sigma\mathbf{q}_{n+1/2} \tag{6b}$$

yields $\bar{H}(\mathbf{q}_{n+1},\mathbf{p}_{n+1}) = \bar{H}(\mathbf{q}_n,\mathbf{p}_n)$, because $\bar{H}$ is quadratic [2]. Comparing this to (5) we see that (5) is satisfied if we choose

$$\sigma = 2\frac{V(r_{n+1}) - V(r_n)}{r_{n+1}^2 - r_n^2} = \frac{2}{r_{n+1}+r_n}\,\frac{V(r_{n+1}) - V(r_n)}{r_{n+1} - r_n}\,, \tag{6c}$$

where $r_n = \|\mathbf{q}_n\|$. The method (6) is then energy conserving. It also conserves linear and angular momentum.

The energy conserving method (6) is a close variation of the method

$$\mathbf{q}_{n+1} = \mathbf{q}_n + k\mathbf{p}_{n+1/2} \tag{7a}$$

$$\mathbf{p}_{n+1} = \mathbf{p}_n - k\frac{V'(r_{n+1/2})}{r_{n+1/2}}\mathbf{q}_{n+1/2} \tag{7b}$$

(because $\frac{V(r_{n+1})-V(r_n)}{r_{n+1}-r_n}$ can be viewed as a difference approximation of $V'(r_{n+1/2})$) where

$$r_{n+1/2} = \frac{1}{2}(r_{n+1} + r_n) = \frac{1}{2}(\|\mathbf{q}_{n+1}\| + \|\mathbf{q}_n\|)\,. \tag{7c}$$

Note that this latter method differs from the midpoint method, where one would use $r(\mathbf{q}_{n+1/2}) = \|(\mathbf{q}_{n+1} + \mathbf{q}_n)/2\|$ instead of (7c) for $r_{n+1/2}$ in (7b). For highly oscillatory systems with $k \gg \varepsilon$, this can be a significant difference, because r is discretized directly in (7). An example in §4 below shows that the midpoint method can become unstable while (7) and (6) remain stable.

For a general Hamiltonian system (1), a straightforward generalization of (6) reads

$$\mathbf{q}_{n+1} = \mathbf{q}_n + k\mathbf{p}_{n+1/2} \tag{8a}$$

$$\mathbf{p}_{n+1} = \mathbf{p}_n - k\sigma\nabla U(\mathbf{q}_{n+1/2}) \tag{8b}$$

where

$$\sigma = \frac{U(\mathbf{q}_{n+1}) - U(\mathbf{q}_n)}{\nabla U(\mathbf{q}_{n+1/2})^T(\mathbf{q}_{n+1} - \mathbf{q}_n)}\,. \tag{8c}$$

For refined variants of this method in the context of many-particle systems, see [15, 12].

2.2 A Projection Method

Here a symmetric projection step is used to enforce conservation of energy. Let $\boldsymbol{a}(\boldsymbol{q},\boldsymbol{p})$ and $\boldsymbol{b}(\boldsymbol{q},\boldsymbol{p})$ be two vector-valued functions such that $(\boldsymbol{p}^T\boldsymbol{a}(\boldsymbol{q},\boldsymbol{p}) + \nabla U(\boldsymbol{q})^T\boldsymbol{b}(\boldsymbol{q},\boldsymbol{p}))$ is bounded away from zero. Then we propose the following modified midpoint method,

$$\begin{aligned}
\boldsymbol{q}_{n+1} &= \boldsymbol{q}_n + k\,\boldsymbol{p}_{n+1/2} + \lambda\,\boldsymbol{b}(\boldsymbol{q}_{n+1/2}, \boldsymbol{p}_{n+1/2})\,,\\
\boldsymbol{p}_{n+1} &= \boldsymbol{p}_n - k\,\nabla U(\boldsymbol{q}_{n+1/2}) + \lambda\,\boldsymbol{a}(\boldsymbol{p}_{n+1/2}, \boldsymbol{p}_{n+1/2})\,,\\
0 &= H(\boldsymbol{q}_{n+1}, \boldsymbol{p}_{n+1}) - H(\boldsymbol{q}_n, \boldsymbol{p}_n)\,.
\end{aligned}$$

The parameter λ is determined by the requirement that the total energy $H = \boldsymbol{p}^T\boldsymbol{p}/2 + U(\boldsymbol{q})$ is conserved.

The particular choice $\boldsymbol{a}(\boldsymbol{q},\boldsymbol{p}) = \nabla U(\boldsymbol{q})$ and $\boldsymbol{b}(\boldsymbol{q},\boldsymbol{p}) = 0$ leads to the method (8). This is further discussed in [12].

Another option is $\boldsymbol{a}(\boldsymbol{q},\boldsymbol{p}) = \boldsymbol{p}$ and $\boldsymbol{b}(\boldsymbol{q},\boldsymbol{p}) = \nabla U(\boldsymbol{q})$. This guarantees that we are discretizing a pure index-2 DAE for which λ is well-defined. But for this choice we observed severe difficulties with Newton's method, where a step-size smaller even than what is required by explicit methods is needed to obtain convergence. In fact, it can be shown that when the linear harmonic oscillator is cast into such a projected DAE, the linearized problem can easily become unstable for $k > \varepsilon$. Another way is to check the conditions of the Newton-Kantorovich Theorem, which guarantees convergence of the Newton method. These conditions are also found to be satisfied only for a very small step size k, if ε is small.

This latter modified midpoint method does work well, however, for the long time integration of Hamiltonian systems which are not highly oscillatory. Note that conservation of any other first integral can be enforced in a similar manner. To our knowledge, this method has not been considered in the literature before in the context of Hamiltonian systems, although it is standard among methods for incompressible Navier-Stokes (where its time-reversibility is not an issue, however).

For highly oscillatory Hamiltonian systems, the best energy conserving midpoint variant that we are aware of is (6). In the sequel we therefore examine only its performance.

3 A Simple Model Problem

As our first model problem, we take the motion of a diatomic molecule under an external force field. For simplicity, it is assumed that (i) the motion is planar, (ii) the two atoms have equal mass $m = 1$, and (iii) the chemical bond is modeled by a stiff harmonic spring with equilibrium length $r_0 = 1$. Denoting the positions of the two atoms by $\boldsymbol{q}_i \in \mathcal{R}^2$, $i = 1, 2$, the corresponding Hamiltonian function is of type

$$H = \frac{\boldsymbol{p}_1^T \boldsymbol{p}_1}{2} + \frac{\boldsymbol{p}_2^T \boldsymbol{p}_2}{2} + \frac{\kappa}{2}(\|\boldsymbol{q}_1 - \boldsymbol{q}_2\| - 1)^2 + V_1(\|\boldsymbol{q}_1 - \boldsymbol{q}_1^0\|) + V_2(\|\boldsymbol{q}_2 - \boldsymbol{q}_2^0\|)\,.$$

Here $\kappa \gg 1$ is the force constant of the harmonic spring, $\|\cdot\|$ denotes the Euclidian norm of a vector in $\mathcal{R}^2$, the functions $V_i : \mathcal{R} \to \mathcal{R}$, $i = 1, 2$, are assumed to be smooth but arbitrary otherwise, and $\boldsymbol{q}_i^0$, $i = 1, 2$, are two fixed reference vectors.

Let us introduce the following abbreviations:

$$\begin{aligned} \hat{r}_1 &:= \|\boldsymbol{q}_1 - \boldsymbol{q}_1^0\|\,, \\ \hat{r}_2 &:= \|\boldsymbol{q}_2 - \boldsymbol{q}_2^0\|\,, \\ r_{12} &:= \|\boldsymbol{q}_1 - \boldsymbol{q}_2\|\,. \end{aligned}$$

Then the equations of motion are

$$\frac{d}{dt}\boldsymbol{q}_1 = \boldsymbol{p}_1\,, \tag{9a}$$

$$\frac{d}{dt}\boldsymbol{q}_2 = \boldsymbol{p}_2\,, \tag{9b}$$

$$\frac{d}{dt}\boldsymbol{p}_1 = -\frac{V_1'(\hat{r}_1)}{\hat{r}_1}(\boldsymbol{q}_1 - \boldsymbol{q}_1^0) - \frac{\kappa}{r_{12}}(r_{12} - 1)(\boldsymbol{q}_1 - \boldsymbol{q}_2)\,, \tag{9c}$$

$$\frac{d}{dt}\boldsymbol{p}_2 = -\frac{V_2'(\hat{r}_2)}{\hat{r}_2}(\boldsymbol{q}_2 - \boldsymbol{q}_2^0) - \frac{\kappa}{r_{12}}(r_{12} - 1)(\boldsymbol{q}_2 - \boldsymbol{q}_1)\,. \tag{9d}$$

The qualitative solution behavior becomes more apparent when going to local coordinates, i.e., we rewrite the equations of motion in terms of the center of mass

$$\boldsymbol{q}_c = \frac{\boldsymbol{q}_1 + \boldsymbol{q}_2}{2}\,,$$

the internal bond stretching $r = r_{12}$, and the angle of rotation ϕ determined by

$$\boldsymbol{\psi}(\phi) = \begin{pmatrix} \cos\phi \\ \sin\phi \end{pmatrix} = \frac{\boldsymbol{q}_1 - \boldsymbol{q}_2}{r_{12}}\,.$$

This coordinate transformation gives rise to a corresponding transformation of the momenta via the canonical lift transformation [10]. Thus the corresponding conjugate momenta are $\boldsymbol{p}_c \in \mathcal{R}^2$, defined by

$$\boldsymbol{p}_c = \boldsymbol{p}_1 + \boldsymbol{p}_2 \,,$$

and $p_r, p_\phi \in \mathcal{R}$, defined by[1]

$$\frac{\boldsymbol{p}_1 - \boldsymbol{p}_2}{2} = \frac{p_r}{r}(\boldsymbol{q}_1 - \boldsymbol{q}_2) + \frac{p_\phi}{r^2} J(\boldsymbol{q}_1 - \boldsymbol{q}_2) \tag{10}$$

where

$$J = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} .$$

To transform the Hamiltonian H into these new coordinates, we use the identities

$$\begin{aligned} \frac{1}{2}\boldsymbol{p}_1^T\boldsymbol{p}_1 + \frac{1}{2}\boldsymbol{p}_2^T\boldsymbol{p}_2 &= \frac{1}{4}(\boldsymbol{p}_1 + \boldsymbol{p}_2)^T(\boldsymbol{p}_1 + \boldsymbol{p}_2) + \frac{1}{4}(\boldsymbol{p}_1 - \boldsymbol{p}_2)^T(\boldsymbol{p}_1 - \boldsymbol{p}_2) \,, \\ &= \frac{1}{4}\boldsymbol{p}_c^T\boldsymbol{p}_c + p_r^2 + r^{-2}p_\phi^2 \end{aligned}$$

and

$$\begin{aligned} \boldsymbol{q}_1 &= \boldsymbol{q}_c + \frac{r}{2}\boldsymbol{\psi}(\phi) \,, \\ \boldsymbol{q}_2 &= \boldsymbol{q}_c - \frac{r}{2}\boldsymbol{\psi}(\phi) \,. \end{aligned}$$

Thus we obtain

$$H = \frac{\boldsymbol{p}_c^T\boldsymbol{p}_c}{4} + p_r^2 + r^{-2}p_\phi^2 + \frac{\kappa}{2}(r-1)^2 + V_1(\boldsymbol{q}_c, r, \phi, \boldsymbol{q}_1^0) + V_2(\boldsymbol{q}_c, r, \phi, \boldsymbol{q}_2^0) \,. \tag{11}$$

The corresponding equations of motion are

$$\frac{d}{dt}\boldsymbol{q}_c = \frac{1}{2}\boldsymbol{p}_c \,, \tag{12a}$$

$$\frac{d}{dt}\boldsymbol{p}_c = -\frac{V_1'(\hat{r}_1)}{\hat{r}_1}(\boldsymbol{q}_1 - \boldsymbol{q}_1^0) - \frac{V_2'(\hat{r}_2)}{\hat{r}_2}(\boldsymbol{q}_2 - \boldsymbol{q}_2^0) \tag{12b}$$

for the center of mass,

$$\frac{d}{dt}r = 2p_r \,, \tag{12c}$$

$$\begin{aligned} \frac{d}{dt}p_r = -\kappa(r-1) + 2p_\phi^2 r^{-3} - \frac{V_1'(\hat{r}_1)}{2\hat{r}_1}\left(\frac{r}{2} + \boldsymbol{\psi}^T[\boldsymbol{q}_c - \boldsymbol{q}_1^0]\right) + \\ - \frac{V_2'(\hat{r}_2)}{2\hat{r}_2}\left(\frac{r}{2} - \boldsymbol{\psi}^T[\boldsymbol{q}_c - \boldsymbol{q}_2^0]\right) \end{aligned} \tag{12d}$$

[1] Explicit expressions for p_r and p_ϕ are obtained by premultiplying (10) by $(\boldsymbol{q}_1 - \boldsymbol{q}_2)^T$, $(\boldsymbol{q}_1 - \boldsymbol{q}_2)^T J$ respectively.

for the internal bond vibrations, and

$$\frac{d}{dt}\phi = 2r^{-2}p_\phi\,, \tag{12e}$$

$$\frac{d}{dt}p_\phi = r\frac{V_1'(\hat{r}_1)}{2\hat{r}_1}\boldsymbol{\psi}^T J(\boldsymbol{q}_c - \boldsymbol{q}_1^0) - r\frac{V_2'(\hat{r}_2)}{2\hat{r}_2}\boldsymbol{\psi}^T J(\boldsymbol{q}_c - \boldsymbol{q}_2^0) \tag{12f}$$

for the rotation of the molecule. Here we have used

$$\nabla_x \hat{r}_i = \frac{1}{2}\hat{r}_i^{-1}\nabla_x \hat{r}_i^2, \qquad x = r, \phi, \boldsymbol{q}_c\,,$$

$$\hat{r}_i^2 = (\boldsymbol{q}_c - \boldsymbol{q}_i^0)^T(\boldsymbol{q}_c - \boldsymbol{q}_i^0) + \frac{1}{4}r^2 - (-1)^i r\,\boldsymbol{\psi}^T(\boldsymbol{q}_c - \boldsymbol{q}_i^0), \qquad i = 1, 2\,.$$

In the context of our semi-implicit methods, we typically consider the special case $V_1 = V_2 = 0$ which leads to the simplified equations of motion

$$\frac{d}{dt}\boldsymbol{q}_c = \frac{1}{2}\boldsymbol{p}_c\,, \tag{13a}$$

$$\frac{d}{dt}\boldsymbol{p}_c = \boldsymbol{0}\,, \tag{13b}$$

$$\frac{d}{dt}r = 2p_r\,, \tag{13c}$$

$$\frac{d}{dt}p_r = -\kappa(r-1) + 2p_\phi^2 r^{-3}\,, \tag{13d}$$

$$\frac{d}{dt}\phi = 2r^{-2}p_\phi\,, \tag{13e}$$

$$\frac{d}{dt}p_\phi = 0\,. \tag{13f}$$

4 Numerical Approximation

The standard discretization for the equations (9) in molecular dynamics is the (explicit) Verlet method. Stability considerations imply that the Verlet method must be applied with a step-size restriction $k < \varepsilon := \sqrt{2/\kappa}$. Various methods have been suggested to avoid this step-size barrier. The most popular is to replace the stiff spring by a holonomic constraint, as in (4). For our first model problem, this leads to the equations

$$\begin{aligned}
\frac{d}{dt}\boldsymbol{q}_1 &= \boldsymbol{p}_1\,,\\
\frac{d}{dt}\boldsymbol{q}_2 &= \boldsymbol{p}_2\,,\\
\frac{d}{dt}\boldsymbol{p}_1 &= -\frac{V_1'(\hat{r}_1)}{\hat{r}_1}(\boldsymbol{q}_1 - \boldsymbol{q}_1^0) - \frac{\lambda}{r_{12}}(\boldsymbol{q}_1 - \boldsymbol{q}_2)\,,\\
\frac{d}{dt}\boldsymbol{p}_2 &= -\frac{V_2'(\hat{r}_2)}{\hat{r}_2}(\boldsymbol{q}_2 - \boldsymbol{q}_2^0) - \frac{\lambda}{r_{12}}(\boldsymbol{q}_2 - \boldsymbol{q}_1)\,,\\
0 &= r_{12} - 1\,.
\end{aligned}$$

In local coordinates, the constraint is $r \equiv 1$, so equations (12) simplify directly into

$$\begin{aligned}
\frac{d}{dt}\boldsymbol{q}_c &= \frac{1}{2}\boldsymbol{p}_c\,,\\
\frac{d}{dt}\boldsymbol{p}_c &= -\frac{V_1'(\hat{r}_1)}{\hat{r}_1}(\boldsymbol{q}_1 - \boldsymbol{q}_1^0) - \frac{V_2'(\hat{r}_2)}{\hat{r}_2}(\boldsymbol{q}_2 - \boldsymbol{q}_2^0)\,,\\
\frac{d}{dt}\phi &= 2p_\phi\,,\\
\frac{d}{dt}p_\phi &= \frac{V_1'(\hat{r}_1)}{2\hat{r}_1}\boldsymbol{\psi}^T J(\boldsymbol{q}_c - \boldsymbol{q}_1^0) - \frac{V_2'(\hat{r}_2)}{2\hat{r}_2}\boldsymbol{\psi}^T J(\boldsymbol{q}_c - \boldsymbol{q}_2^0)\,.
\end{aligned}$$

The constrained equations of motion in cartesian coordinates can be solved by the SHAKE or (the essentially equivalent) RATTLE method (see [8]) which requires the solution of a non-linear system of equations in the Lagrange multiplier function λ. The equivalent formulation in local coordinates can still be integrated by using the explicit Verlet method.

The main disadvantage of this approach arises when the limit constrained system is different from (4), as mentioned in the introduction and demonstrated in §5 for our second model problem.

Another way to overcome the step-size restriction $k < \varepsilon$ is to use multiple-time-stepping methods [4] or implicit methods [17, 18, 12, 3]. In this paper, we examine the latter possibility. But for large molecular systems, fully implicit methods are very expensive. For that reason, we focus on the general class of semi-implicit methods depicted in Fig. 1 [12]. In this scheme, Step 3 of the nth time step can be combined with Step 1 of the $(n+1)$st time step. This then is a staggered two-step splitting method. We refer to [12] for further justification.

Note that, in local coordinates, Step 2 is equivalent to integrating the equations (13). Thus, Step 2 can either be performed in local or in cartesian coordinates. We consider two different implicit methods for this purpose, namely, the midpoint method and the energy conserving method (6) which, in this example, coincides with the method (7) (because the V term appearing in (6) and (7) for $\mathbf{q} = \mathbf{q}_1 - \mathbf{q}_2$ is quadratic here). These methods are applied to the formulation in cartesian and in local coordinates and the properties of the resulting propagation maps are discussed next.

SEMI-IMPLICIT INTEGRATOR

Step 1.

$$\bar{\boldsymbol{p}}_{1,n} = \boldsymbol{p}_{1,n} - k\frac{V_1'(\hat{r}_{1,n})}{2\hat{r}_{1,n}}(\boldsymbol{q}_{1,n} - \boldsymbol{q}_1^0)\,,$$

$$\bar{\boldsymbol{p}}_{2,n} = \boldsymbol{p}_{2,n} - k\frac{V_2'(\hat{r}_{2,n})}{2\hat{r}_{2,n}}(\boldsymbol{q}_{2,n} - \boldsymbol{q}_2^0)\,.$$

Step 2.

Apply one step of size k to approximately integrate the fast system

$$\begin{aligned}
\frac{d}{dt}\boldsymbol{q}_1 &= \boldsymbol{p}_1\,,\\
\frac{d}{dt}\boldsymbol{q}_2 &= \boldsymbol{p}_2\,,\\
\frac{d}{dt}\boldsymbol{p}_1 &= -\frac{\kappa}{r_{12}}(r_{12}-1)(\boldsymbol{q}_1 - \boldsymbol{q}_2)\,,\\
\frac{d}{dt}\boldsymbol{p}_2 &= -\frac{\kappa}{r_{12}}(r_{12}-1)(\boldsymbol{q}_2 - \boldsymbol{q}_1)
\end{aligned}$$

using an implicit method with initial conditions $(\boldsymbol{q}_{1,n}, \boldsymbol{q}_{2,n}, \bar{\boldsymbol{p}}_{1,n}, \bar{\boldsymbol{p}}_{2,n})$. Denote the result by $(\boldsymbol{q}_{1,n+1}, \boldsymbol{q}_{2,n+1}, \tilde{\boldsymbol{p}}_{1,n+1}, \tilde{\boldsymbol{p}}_{2,n+1})$.

Step 3.

$$\boldsymbol{p}_{1,n+1} = \tilde{\boldsymbol{p}}_{1,n+1} - k\frac{V_1'(\hat{r}_{1,n+1})}{2\hat{r}_{1,n+1}}(\boldsymbol{q}_{1,n+1} - \boldsymbol{q}_1^0)\,,$$

$$\boldsymbol{p}_{2,n+1} = \tilde{\boldsymbol{p}}_{2,n+1} - k\frac{V_2'(\hat{r}_{2,n+1})}{2\hat{r}_{2,n+1}}(\boldsymbol{q}_{2,n+1} - \boldsymbol{q}_2^0)\,.$$

Fig. 1. A semi-implicit integrator: the implicit scheme is applied only to the fast system.

4.1 Local Coordinates

The midpoint discretization of (13) yields

$$\boldsymbol{q}_{c,n+1} = \boldsymbol{q}_{c,n} + k\boldsymbol{p}_{c,n}/2\,, \tag{14a}$$

$$\boldsymbol{p}_{c,n+1} = \boldsymbol{p}_{c,n}\,, \tag{14b}$$

$$r_{n+1} = r_n + 2k\,p_{r,n+1/2}\,, \tag{14c}$$

$$p_{r,n+1} = p_{r,n} - k\kappa(r_{n+1/2} - 1) + kp_{\phi,n+1/2}^2 r_{n+1/2}^{-3}\,, \tag{14d}$$

$$\phi_{n+1} = \phi_n + 2kr_{n+1/2}^{-2}p_{\phi,n}\,, \tag{14e}$$

$$p_{\phi,n+1} = p_{\phi,n}\,. \tag{14f}$$

As remarked earlier, we are interested in the behavior of this approximation for step-sizes k much larger than the period of the fast bond vibrations,

which is of order ε. But then, the Hamiltonian (11) is almost quadratic. The midpoint approximation almost reproduces this Hamiltonian [3, 2], which provides an energy norm estimate for the approximate solution. A simple perturbation argument yields from this that the midpoint method applied to the local coordinates formulation of the fast system (13) is unconditionally stable and that, in the limit $\kappa \to \infty$, it conserves energy exactly. Note that, in local coordinates, the midpoint method and its energy conserving variants are basically equivalent.

4.2 Cartesian Coordinates

The interesting question is now what happens if the midpoint method is applied to the cartesian formulation (9) with $V_1 = V_2 = 0$. The equations are

$$\boldsymbol{q}_{1,n+1} = \boldsymbol{q}_{1,n} + k\boldsymbol{p}_{1,n+1/2}\,, \tag{15a}$$

$$\boldsymbol{q}_{2,n+1} = \boldsymbol{q}_{2,n} + k\boldsymbol{p}_{2,n+1/2}\,, \tag{15b}$$

$$\boldsymbol{p}_{1,n+1} = \boldsymbol{p}_{1,n} - \frac{k\,\kappa}{\tilde{r}_{n+1/2}}(\tilde{r}_{n+1/2} - 1)(\boldsymbol{q}_{1,n+1/2} - \boldsymbol{q}_{2,n+1/2})\,, \tag{15c}$$

$$\boldsymbol{p}_{2,n+1} = \boldsymbol{p}_{2,n} - \frac{k\,\kappa}{\tilde{r}_{n+1/2}}(\tilde{r}_{n+1/2} - 1)(\boldsymbol{q}_{2,n+1/2} - \boldsymbol{q}_{1,n+1/2})\,, \tag{15d}$$

where $\tilde{r}_{n+1/2} := \|\boldsymbol{q}_{1,n+1/2} - \boldsymbol{q}_{2,n+1/2}\|$. This discretization becomes unstable if the step-size k becomes too large [6]. As discussed in [3] by means of a simple linear time-varying problem, stability requires that

$$\alpha := k^2\sqrt{\kappa}$$

be sufficiently small. Since the midpoint discretization of the equations of motion in local coordinates (14) is stable for any step size k, the present instability might come as a surprise. The explanation can be given as follows [6, 3]: In cartesian coordinates the fast vibrations and the slow translational and rotational degrees of freedom are not decoupled as they are in the local coordinate formulation. For large step sizes as compared to ε, this coupling leads to the destabilization of the midpoint method.

To obtain the unconditional stability of the midpoint method in local coordinates, one would have to consider the decoupling transformation from cartesian to local coordinates for the discrete variables $\boldsymbol{q}_{1,n}$ etc. But this transformation, which for the continuous variables is not constant, necessarily is in error which depends on k, not ε. The stability properties of the discrete dynamical systems obtained by the midpoint discretization in the different sets of coordinates may therefore be significantly different when $k \gg \varepsilon$ [3].

This instability is avoided when applying the method (6), as proved in [6]. The latter method yields here

$$\boldsymbol{q}_{1,n+1} = \boldsymbol{q}_{1,n} + k\boldsymbol{p}_{1,n+1/2}\,, \tag{16a}$$

$$\boldsymbol{q}_{2,n+1} = \boldsymbol{q}_{2,n} + k\boldsymbol{p}_{2,n+1/2}\,, \tag{16b}$$

$$\boldsymbol{p}_{1,n+1} = \boldsymbol{p}_{1,n} - \frac{k\,\kappa}{r_{n+1/2}}(r_{n+1/2}-1)(\boldsymbol{q}_{1,n+1/2} - \boldsymbol{q}_{2,n+1/2})\,, \tag{16c}$$

$$\boldsymbol{p}_{2,n+1} = \boldsymbol{p}_{2,n} - \frac{k\,\kappa}{r_{n+1/2}}(r_{n+1/2}-1)(\boldsymbol{q}_{2,n+1/2} - \boldsymbol{q}_{1,n+1/2})\,, \tag{16d}$$

where

$$r_{n+1/2} := \frac{1}{2}(\|\boldsymbol{q}_{1,n+1} - \boldsymbol{q}_{2,n+1}\| + \|\boldsymbol{q}_{1,n} - \boldsymbol{q}_{2,n}\|)\,.$$

The gain in stability can now be interpreted as resulting from the direct midpoint discretization of the rapidly vibrating, local variable r, thus avoiding the potentially damaging discrete decoupling transformation.

k	$\sqrt{\kappa}$	α	ΔH
0.05	0.5e+4	12.5	0.71e-3
0.05	1.0e+4	25	0.83e-3
0.1	0.5e+4	50	0.39e-2
0.1	1.0e+4	100	0.33e-2
0.2	0.5e+4	200	0.11e-1
0.2	1.0e+4	400	0.12e-1

Table 1. Maximum error in the energy using the semi-implicit method with the energy conserving method (6) for the strong forces.

Numerical Experiment We now present numerical results for non-zero V_1 and V_2. In particular, we take

$$V_1(\hat{r}_1) := \frac{1}{2}(\hat{r}_1 - 1)^2\,, \qquad \hat{r}_1 = \|\boldsymbol{q}_1 - [1.5, 0]^T\|\,,$$

$$V_2(\hat{r}_2) := \frac{1}{2}(\hat{r}_2 - 1)^2\,, \qquad \hat{r}_2 = \|\boldsymbol{q}_2 + [1.5, 0]^T\|\,.$$

As initial conditions we chose $\boldsymbol{q}_1 = [0.6, 0]^T$, $\boldsymbol{q}_2 = [-0.4, 0]^T$, $\boldsymbol{p}_1 = [1, 1]^T$, and $\boldsymbol{p}_2 = [-1, -1]^T$.

We apply the semi-implicit algorithm to handle the weak potentials V_i, and the energy conserving method (16) for the stiff forces. The maximal error in the total energy, i.e.

$$\Delta H = \max_{t\in[0,500]} |H(t) - H(0)|\,,$$

can be found in Table 1 for various values of k, κ, and α. Even for large κ and α, the value of ΔH depends quadratically on the step size k.

A difficulty with the energy conserving method (6), in general, is the solution of the corresponding nonlinear equations [6]. Here, however, using the initial iterate $(\mathbf{q}_n + k\mathbf{p}_n, \mathbf{p}_n)$ for $(\mathbf{q}_{n+1}, \mathbf{p}_{n+1})$, even for large values of α we did not observe any difficulties with the convergence of Newton's method.

On the other hand, our computations using (15) indicate that the midpoint method becomes unstable for $\alpha > 1$.

Next, we replace the stiff spring potential $\kappa(r-1)^2/2$ by the Morse potential

$$\kappa(1 - e^{-(r-1)/2})^2,$$

where $r = \|\mathbf{q}_1 - \mathbf{q}_2\|$ as before. The methods (6) and (7) are no longer the same, and (7) does not conserve the energy exactly, even in the absence of the slow potentials. However, these two methods remain close. Repeating the experiments of Table 1 produces qualitatively similar results for the Morse potential. The computations with either (6) or (7) are stable. The energy error, as well as the error in fast energy ΔE_f defined in the next section, appear to converge quadratically in k and be independent of κ, when $k \gg \sqrt{\kappa}$. Similar numerical experiments with the midpoint method run into difficulties.

5 Modified Model Problem

For molecules with more than two atoms the frequencies of the bond stretching modes and/or bond angle bending modes are, in general, no longer constant along the slowly varying solution components. This implies additional complications not present for our simple diatomic model problem. In particular, in the limit $\kappa \to \infty$, replacing the stiff spring terms by holonomic constraints as in (4) leads to a qualitatively wrong dynamics which has to be corrected by introducing an additional force term [14, 5]. The size of this additional force term depends on the limiting initial conditions.

Unfortunately, discretization methods with large step sizes applied to such problems tend to miss this additional force term [3]. Furthermore, even if the implicit midpoint method is applied to a formulation in local coordinates, similar problems occur [3]. Since the midpoint scheme and its variants (6) and (7) are basically identical in local coordinates, the same problem can be expected for the energy conserving method (6). To demonstrate this, let us consider the following modified model problem [13]:

In the model problem described earlier, replace the stiff spring potential $V_s = \kappa(r-1)^2/2$ by

$$V_s = \frac{\kappa}{2}(\|A(\boldsymbol{q}_1 - \boldsymbol{q}_2)\| - 1)^2$$

with

$$A = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & 1 \end{bmatrix} .$$

This implies that the two "atoms" oscillate about a mutual distance $d = r(\phi)$ that depends on the angle ϕ and is given by an ellipse. Let

$$d = \|A(\boldsymbol{q}_1 - \boldsymbol{q}_2)\| .$$

The equations of motion are an obvious modification of (9) where $\boldsymbol{q}_1 - \boldsymbol{q}_2$ is replaced by $A^2(\boldsymbol{q}_1 - \boldsymbol{q}_2)$ and r_{12} is replaced by d. The total energy H of the system decomposes into the vibrational energy E_f of the stiff "spring" and the energy in the slowly varying degrees of freedom. The time-evolution of $E_f(t)$ is crucial for the limiting behavior of (3a)-(3b) as $\kappa \to \infty$. Let us derive an explicit expression for the energy E_f. The conjugate momentum corresponding to d is

$$p_d := \frac{d}{2\|A^2(\boldsymbol{q}_1 - \boldsymbol{q}_2)\|^2} (\boldsymbol{p}_1 - \boldsymbol{p}_2)^T A^2 (\boldsymbol{q}_1 - \boldsymbol{q}_2) .$$

The vibrational energy E_f is now given by

$$E_f = \omega^2(\boldsymbol{q}_1 - \boldsymbol{q}_2)\, p_d^2 + \frac{\kappa}{2}(d-1)^2$$

with

$$\omega(\boldsymbol{q}_1 - \boldsymbol{q}_2) := \frac{\|A^2(\boldsymbol{q}_1 - \boldsymbol{q}_2)\|}{d} .$$

Note that E_f corresponds to a high-frequency harmonic oscillator with a slowly varying frequency ω. This implies that the vibrational energy E_f is not constant along solution curves. Instead, we have the adiabatic invariant $J_f := E_f/\omega$ which is preserved up to terms of order $\varepsilon = \sqrt{2/\kappa}$ over a time-interval of order at least one. In the limit $\kappa \to \infty$, the slow motion is no longer given by simply enforcing $d-1 = 0$ as a holonomic constraint. Instead, a force term corresponding to the additional potential energy

$$V_c = J_f \omega(\boldsymbol{q}_1 - \boldsymbol{q}_2)$$

has to be added in (4).

Numerical Experiment We consider the case $V_1 = V_2 = 0$ and apply the energy conserving method (6). (Note that the methods (6) and (7) are still equivalent, with the corresponding notation $r \leftarrow d$, $V(d) \leftarrow \frac{\kappa}{2}(d-1)^2$, and $\mathbf{q} \leftarrow A^2(\mathbf{q}_1 - \mathbf{q}_2)$.) As initial values we took $\boldsymbol{q}_1 = [0, 0.5]^T$, $\boldsymbol{q}_2 = [0, -0.5]^T$, $\boldsymbol{p}_1 = [1, 1]^T$, and $\boldsymbol{p}_2 = [-1, -1]^T$. The maximal variation in the vibrational energy, i.e.

$$\Delta E_f = \max_{t \in [0,10]} |E_f(t) - E_f(0)| ,$$

k	κ	$k\sqrt{\kappa}$	ΔE_f	ΔJ_f
0.1e-2	1.0e+4	0.1	0.42	0.57e-1
0.1e-1	1.0e+4	1	0.24	0.14
0.5e-1	1.0e+4	5	0.59e-1	0.29
0.1e-3	1.0e+6	0.1	0.41	0.57e-2
0.1e-2	1.0e+6	1	0.23	0.13
0.1e-1	1.0e+6	10	0.62e-2	0.29
0.2e-1	1.0e+6	20	0.60e-2	0.29

Table 2. Maximum variation in the vibrational energy E_f and the adiabatic invariant J_f using the energy conserving method (6).

and in the adiabatic invariant, i.e.

$$\Delta J_f = \max_{t\in[0,10]} |J_f(t) - J_f(0)|,$$

can be found in Table 2. Note that, for $k\sqrt{\kappa} \gg 1$, the method enforces $\Delta E_f \approx 0$ (and thus $V_c = 0$) instead of (approximately) conserving the adiabatic invariant J_f. This implies that the energy conserving method introduces an error of order one in the slow solution quantities, unless $k\sqrt{\kappa}$ is bounded by a constant of magnitude order 1.[2] This is a step-size restriction which is comparable to the explicit Verlet method stability restriction, up to a moderate factor. In fact, the energy conserving method leads to the same wrong solution behavior as a naive enforcement of the holonomic constraint $d = 1$ by a method such as SHAKE. Note that, due to the energy conserving nature of the method, no indication of this wrong solution behavior will be given unless the adiabatic invariant is explicitly computed. Again we did not observe any difficulties in solving the nonlinear equations by Newton's method, so no alarm bells of any kind rang while computing a wrong solution.

References

1. M.P. Allen and T.J. Tildesley. *Computer Simulations of Liquids.* Clarendon Press, Oxford, 1987.
2. U. Ascher. Stabilization of invariants of discretized differential systems. *Numerical Algorithms*, 14:1–23, 1997.
3. U. Ascher and S. Reich. The midpoint scheme and variants for Hamiltonian systems: advantages and pitfalls. *SIAM J. Sci. Comput.*, to appear.
4. J.J. Biesiadecki and R.D. Skeel. Danger of multiple-time-step methods. *J. Comput. Phys.*, 109:318–328, 1993.

[2] To be more precise, this error occurs in the limit $\kappa \to \infty$ with $E_f = \mathcal{O}(1)$ and step-size k such that $k\sqrt{\kappa} = const. \gg 1$. This error does not occur if $E_f = 0$ for the analytic problem, i.e., in case there is no vibrational energy in the stiff "spring" which implies $V_c = 0$.

5. F.A. Bornemann and Ch. Schütte. Homogenization of Hamiltonian systems with a strong constraining potential. *Physica D*, 102:57–77, 1997.
6. O. Gonzalez and J. Simo. On the stability of symplectic and energy-momentum algorithms for nonlinear Hamiltonian systems with symmetry. *Comp. Meth. in Appl. Mech. Eng.*, 134:197–222, 1996.
7. E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems.* Springer-Verlag, 1991.
8. B. Leimkuhler and R. Skeel. Symplectic integrators in constrained Hamiltonian systems. *J. Comp. Phys.*, 112:117–125, 1994.
9. M. Mandziuk and T. Schlick. Resonance in the dynamics of chemical systems simulated by the implicit midpoint scheme. *Chem. Phys. Lett.*, 237:525–535, 1995.
10. J.E. Marsden and T.S. Ratiu. *Introduction to Mechanics and Symmetry.* Springer-Verlag, New York, 1994.
11. L.R. Petzold, L.O. Jay, and J. Yen. Numerical solution of highly oscillatory ordinary differential equations. *Acta Numerica*, pages 437–484, 1997.
12. S. Reich. Enhanced energy conserving methods. *BIT*, 36:122–134, 1996.
13. S. Reich. Preservation of adiabatic invariants under symplectic discretization. *App. Numer. Math.*, to appear, 1998.
14. H. Rubin and P. Ungar. Motion under a strong constraining force. *Comm. Pure Appl. Math.*, 10:65–87, 1957.
15. J.C. Simo and O. Gonzales. Assessment of energy-monentum and symplectic schemes for stiff dynamical systems. *The American Society of Mechanical Engineering*, 93-WA/PVP-4, 1993.
16. T. Schlick, M. Mandziuk, R.D. Skeel, and K. Srinivas. Nonlinear resonance artifacts in molecular dynamics simulations. *J. Comput. Phys.*, 139:1–29, 1998.
17. R.D. Skeel, G. Zhang, and T. Schlick. A family of symplectic integrators: stability, accuracy, and molecular dynamics applications. *SIAM J. Sci. Comput.*, 18:203–222, 1997.
18. R.D. Skeel and M. Zhang. Cheap implicit symplectic integrators. *Appl. Numer. Math.*, 25:297–302, 1997.
19. L. Verlet. Computer experiments on classical fluids. *Phys. Rev.*, 159:98–103, 1967.

Molecular Dynamics in Systems with Multiple Time Scales: Reference System Propagator Algorithms

Bruce J. Berne

Department of Chemistry, Columbia University,
New York, NY 10027

Abstract. Systems with multiple time scales, and with forces which can be subdivided into long and short range components are frequently encountered in computational chemistry. In recent years, new, powerful and efficient methods have been developed to reduce the computational overhead in treating these problems in molecular dynamics simulations. Numerical reversible integrators for dealing with these problems called r-RESPA (Reversible Reference System Propagator Algorithms) are reviewed in this article. r-RESPA leads to considerable speedups in generating molecular dynamics trajectories with no loss of accuracy. When combined with the Hybrid Monte Carlo (HMC) method and used in the Jump-Walking and the Smart-Walking algorithms, r-RESPA is very useful for the enhanced sampling of rough energy landscapes in biomolecules.

1 Introduction

Molecular Dynamics (MD) is one of the major tools in the arsenal of computational chemistry and physics. It grew out of attempts to understand the static and dynamical properties of hard sphere fluids and its first appearance[1, 2] was in a form applicable to impulsive forces (1957). Several years later (1964), Rahman extended MD to monoatomic liquids in which the atoms interact pairwise through the Lennard-Jones pair potential.[3] This major development was followed soon after (1968) by the first application of MD to fluids containing diatomic molecules[4, 5] interacting through continuous potentials and then to triatomic molecules[6] (1971). It was these applications of MD to molecules interacting through continuous force fields that set the stage for all subsequent applications of MD in computational chemistry. Several excellent monographs exist which treat the methodology in detail.[7, 8, 9, 10]

One of the problems encountered in applying molecular dynamics to the simulation of complex systems is the presence of both fast and slow degrees of freedom. One must choose a small time step to achieve stable integration of the equations of motion for the fast motion and must then generate a very large number of time steps to achieve sufficient sampling of the slow degrees of freedom. Another major bottleneck is the calculation of the long range electrostatic forces. These are called "intrinsic" multiple time scale problems.

Even with fast methods for generating molecular dynamics (MD) or Monte Carlo (MC) trajectories, the problems of sampling conformational states separated by large energy barriers remains an obstacle to progress. This problem raises another and more serious kind of multiple time scale problem – one due to the presence of a rugged energy landscape with the attendant separation of time scales arising from activated barrier crossing. This latter problem will be refered to as the "extrinsic" multiple time scale problem.

In recent years a variety of powerful new molecular dynamics and Monte Carlo methods have been developed to address the "intrinsic" and "extrinsic" multiple time scale respectively. Accurate numerical integrators are required for questions involving real dynamical problems such as transport and energy relaxation. Thus in Sec. 2 we discuss accurate numerical integrators for "intrinsic" multiple time scale problem. On the other hand, in the simulation of biomolecular systems, one is often interested in computing equilibrium averages and thermodynamic quantities. For this purpose, the exact time dependence is not required, since all that is needed is the correct and efficient sampling of the thermally accessible configurations of the system, a problem made difficult by the "extrinsic" multiple time scales connected with the omnipresent energy barriers in systems with rough energy landscapes. A variety of techniques, such as stochastic dynamics, Monte Carlo, Hybrid Monte Carlo, J-Walking etc can be used, some of which are discussed in Sec. 3. First new methods for generating accurate dynamical trajectories are described and then methods based on inaccurate dynamics for sampling state space in systems with rough energy landscapes are treated.

2 Methods for Dealing with the Intrinsic Multiple Time Scale Problem in Molecular Dynamics

In complex systems the set of fast degrees of freedom arises both from vibrations of stiff bonds or particles with small mass. An example of the latter is the fast vibrational motions of the C-H and O-H bonds in biomolecules and the O-H bonds of water. In systems with multiple time scales it is necessary to choose a time step much smaller than the periods of the fastest motions and to recalculate the forces after each small time step. It then requires very long runs to sample the conformational space of the slower degrees of freedom. To bypass this problem some fast degrees of freedom can be eliminated by constraining the length of the stiff bonds.[11] Constrained molecular dynamics suffers from several problems: (a) bond constraints introduce additional angular correlations is torsion angle distribution functions that are not found in the flexible systems in nature; (b) constraints cannot be used to eliminate problems like the fast librational motion of water; (c) the integrators often used in constrained MD are neither reversible nor symplectic. This latter problem means that constrained dynamics cannot be wed to Monte Carlo

methods like Hybrid Monte Carlo[12] which require reversible integrators to insure detailed balance.

In conventional MD the forces are recomputed after each time step. The force calculations account for as much as 95% of the CPU time in an MD simulation. In systems with long range forces, the force computation becomes the major bottleneck to the computation. When using the direct pairwise evaluation, the computational effort required to compute the long-range Coulomb forces on N interacting particles is of order N^2. A variety of strategies, such as the fast multipole method and the particle-particle-mesh Ewald method, have been introduced to reduce the computational effort in calculating the forces. Building on earlier reference system propagator algorithm (RESPA) based integrators,[13, 14, 15, 16] a class of new reversible and symplectic integrators have been invented that greatly reduces the "intrinsic" multiple time scale problem. By using a reversible Trotter factorization of the classical propagator[17] one can generate simple, accurate, reversible and symplectic integrators that allow one to integrate the fast motions using small time steps and the slow degrees of freedom using large time steps.[17] This approach allows one to split the propagator up into a fast part, due to the high frequency vibrations, and slow parts, due to short range, intermediate range, and long range forces, in a variety of ways. These new integrators, called reversible reference system algorithms (r-RESPA), require for the treatment of all-atom force fields no more CPU time than constrained dynamics and often lead to even larger improvements in speed. Although r-RESPA is quite simple to implement, there are many ways to factorize the propagator. A recent paper shows how to avoid bad strategies.[18] Applications of these methods to Car-Parrinello *ab initio* molecular dynamics has resulted in speedups by a factor of approximately five in semiconductor materials.[19, 20, 21] There has been significant progress in recent years to apply these methods to systems of biological relevance.[22, 23, 24, 25, 26]

2.1 Background

As is well known, Molecular Dynamics is used to simulate the motions in many-body systems. In a typical MD simulation one first starts with an initial state of an N particle system $\Gamma = (x_1, \ldots, x_f, p_1, \ldots, p_f)$ where $f = 3N$ is the number of degrees of freedom in the system. After sampling the initial state one numerically solves Hamilton's equations of motion:

$$\begin{aligned} \dot{x}_i &= \frac{\partial H}{\partial p_i} \\ \dot{p}_i &= -\frac{\partial H}{\partial x_i} \end{aligned} \tag{1}$$

subject to the initial conditions. Although there are many possible finite difference approximations or integrators[7] to solve the equations of motion,

we will focus on the Verlet integrator.[27] Over the years this integrator has undergone various extensions and modifications. For example, Andersen *et al.* have introduced the velocity Verlet integrator.[28] In this integrator, the positions $x_i(\Delta t)$ and velocities $\dot{x}_i(\Delta t)$ after one time step Δt are related to the positions $x_i(0)$ and velocities $v_i(0) = \dot{x}_i(0) = p_i(0)/m$ at the beginning of the time step by:

$$x_i(\Delta t) = x_i(0) + \Delta t \dot{x}_i(0) + \frac{(\Delta t)^2}{2} \frac{F(\{x_i(0)\})}{m}$$
$$\dot{x}_i(\Delta t) = \dot{x}_i(0) + \frac{\Delta t}{2m}[F_i(\{x_i(0)\}) + F_i(\{x_i(\Delta t)\})] \tag{2}$$

for $i = 1, \ldots, 3N$, where $F_i = -\partial U(\{x_i\})/\partial x_i$ is the force on the coordinate x_i. The forces at any time can be computed from the potential function $U(\{x_i\})$ and are functions of all of the position coordinates at that time.

One property of the exact trajectory for a conservative system is that the total energy is a constant of the motion.[12] Finite difference integrators provide approximate solutions to the equations of motion and for trajectories generated numerically the total energy is not strictly conserved. The exact trajectory will move on a constant energy surface in the $6N$ dimensional phase space of the system defined by,

$$H(\Gamma) = E. \tag{3}$$

The numerical trajectory will wander off this energy surface. If the trajectory is stable it will wander on an energy shell

$$|E - H(\Gamma)| \leq \Delta E. \tag{4}$$

The smaller the time step Δt used in the integrator, the more accurate will be the trajectory and the smaller will be the thickness of the energy shell on which the trajectory wanders. If the time step is too large the integrator will generate an unstable trajectory and the energy will diverge after a small number of time steps. This will happen if during a time step the errors in the new positions give rise to very large changes in the forces between particles. Then on the next time step the particles will speed up giving rise to still larger errors in the next positions and to even larger changes in the forces. This situation eventually results in disaster. In general, one must choose time steps sufficiently small that the forces do not change significantly. This means that the time step must be small enough for the fastest motions in the system.

One of the advantages of the Verlet integrator is that it is time reversible and symplectic[30, 31, 32]. *Reversibility* means that in the absence of numerical round off error, if the trajectory is run for many time steps, say $n\Delta t$, and the velocities are then reversed, the trajectory will retrace its path and after $n\Delta t$ more time steps it will land back where it started. An integrator can be viewed as a mapping from one point in phase space to another. If this mapping is applied to a measurable point set of states at one time, it will

map these states to another measurable point set. If the mapping is *symplectic*, the measure of the initial point set will be equal to the measure of the final point set. The mapping then satisfies Liouville's Theorem[12] and conserves the measure in phase space. Thus, like the exact solution of Hamilton's equations of motion, symplectic integrators, such as the Verlet integrator (see Eq. 2) , are reversible and measure conserving. In recent years it has been understood that symplectic integrators are more stable than non-symplectic integrators.[31, 32] It can be shown that dynamics generated by a symplectic integrator will conserve not the true Hamiltonian, but rather a modified, time step dependent Hamiltonian, $\tilde{H}(\Delta t)$ in one dimension and is postulated to do so in many dimension [33]. This theorem guarantees that Eq. 4 will hold for all time, t and that the integrator will be stable.

2.2 Integrators Generated from Factorizing the Classical Propagator

Before discussing the method for handling the problem of multiple time step molecular dynamics, it is useful to show how simple operator algebra can be used to generate reversible integrators.[17] The starting point for this is the definition of the classical Liouvillian, a Hermitian operator on functions of the state variables. The Liouvillian is defined in terms of the Poisson Bracket, $\{, H\}$, of whatever it operates on with the Hamiltonian H of the system. In Cartesian coordinates it has the form,

$$iL = \{, H\} = \dot{x}\frac{\partial}{\partial x} + F\frac{\partial}{\partial p_x}, \tag{5}$$

where F is the force ($F = -\partial V/\partial x$), $V(x)$ is the potential function, $\dot{x} = p_x/m$ is the velocity and p is the momentum. For simplicity of notation we treat only a one dimensional system (one position coordinate and one conjugate momentum); nevertheless it should be recognized that for general systems the Liouvillian involves a sum over all degrees of freedom.

The operator,

$$G(t) \equiv e^{iLt}, \tag{6}$$

is the propagator of the classical motion. Thus the state of the system after one time step Δt is found by applying the propagator to the initial state so that

$$\begin{pmatrix} x(\Delta t) \\ p(\Delta t) \end{pmatrix} = e^{iL\Delta t} \begin{pmatrix} x(0) \\ p(0) \end{pmatrix}. \tag{7}$$

Now assuming any decomposition of the Liouvillian into two parts,

$$iL = iL_1 + iL_2, \tag{8}$$

one can use the reversible Trotter factorization of the propagator to approximate the true propagator,

$$e^{(iL_1+iL_2)\Delta t} = e^{iL_1\Delta t/2}e^{iL_2\Delta t}e^{iL_1\Delta t/2} + O(\Delta t^3). \tag{9}$$

Applying this to an initial state of the system represented by the column vector,

$$\begin{pmatrix} x(0) \\ p(0) \end{pmatrix}, \tag{10}$$

gives the state after one time step Δt,

$$\begin{pmatrix} x(\Delta t) \\ p(\Delta t) \end{pmatrix} = e^{iL_1\frac{\Delta t}{2}}e^{iL_2\Delta t}e^{iL_1\frac{\Delta t}{2}}\begin{pmatrix} x(0) \\ p(0) \end{pmatrix}. \tag{11}$$

If we subdivide the Liouvillian into the two parts by separating the force and velocity terms,

$$iL_1 = F\frac{\partial}{\partial p_x} \qquad \text{and} \qquad iL_2 = \dot{x}\frac{\partial}{\partial x}, \tag{12}$$

and apply this factorization to the propagator, we obtain:

$$\begin{pmatrix} x(\Delta t) \\ p(\Delta t) \end{pmatrix} = e^{\frac{\Delta t}{2}F\frac{\partial}{\partial p_x}}e^{\Delta t\dot{x}\frac{\partial}{\partial x}}e^{\frac{\Delta t}{2}F\frac{\partial}{\partial p_x}}\begin{pmatrix} x(0) \\ p(0) \end{pmatrix}. \tag{13}$$

Each of the factorized operators are displacement operators and can thus be applied seriatim to the initial state vector to give the final solution,

$$\begin{pmatrix} x(\Delta t) \\ p(\Delta t) \end{pmatrix} = \begin{pmatrix} x(0) + \Delta t\dot{x}(0) + \frac{\Delta t^2}{2m}F(x(0)) \\ p(0) + \frac{\Delta t}{2}\left[F(x(\Delta t)) + F(x(0))\right] \end{pmatrix}. \tag{14}$$

This procedure is then repeated after each time step. Comparison with Eq. (2) shows that the result is the velocity Verlet integrator and we have thus derived it from a split-operator technique; which is not the way that it was originally derived. A simple interchange of the L_1 and L_2 operators yields an entirely equivalent integrator,

$$\begin{pmatrix} x(\Delta t) \\ p(\Delta t) \end{pmatrix} = \begin{pmatrix} x(0) + \frac{\Delta t}{2m}\left[p(0) + p(\Delta t)\right] \\ p(0) + \Delta tF\left(\left[x(0) + \frac{\Delta t}{2}p(0)\right]\right) \end{pmatrix}. \tag{15}$$

which by symmetry we call the position Verlet integrator, an integrator of the same accuracy as the velocity Verlet integrator.

An interesting property of these integrators is that the Jacobian of the transformation from the state at time 0 to the time Δt is

$$\left|\frac{\partial(x(\Delta t), p_x(\Delta t))}{\partial(x(0), p_x(0))}\right| = 1. \tag{16}$$

Thus these integrators are measure preserving and give trajectories that satisfy the Liouville theorem.[12] This is an important property of symplectic integrators, and, as mentioned before, it is this property that makes these integrators more stable than non-symplectic integrators.[30, 33]

By now it should be clear that this kind of operator algebra can be a useful method for generating integrators. We show, in the following, how it can be applied to generate a wide variety of methods for treating the multiple time scale problem.

2.3 Reference System Propagator Algorithms

The aforementioned factorizations of the classical propagator can be used to generate efficient reversible and symplectic integrators for systems with long and short range forces and for systems in which the degrees of freedom can be subdivided into fast and slow subsets. All of the methods described below are called Reference System Propagator Algorithms (RESPA); a name that we gave to our initial attempts to use an underlying reference system propagator for the fast motion. This early effort resulted in non-reversible integrators.[13, 14, 15, 16] If the Liouville operator of the system is decomposed into a "reference system" part, iL_{ref}, and a "correction part", $i\delta L$, as

$$iL = iL_{ref} + i\delta L. \tag{17}$$

Trotter factorization of the propagator then leads to

$$e^{iL\Delta t} = e^{i\delta L\Delta t/2} e^{iL_{ref}\Delta t} e^{i\delta L\Delta t/2}. \tag{18}$$

In this context the velocity Verlet integrator is equivalent to taking the reference system to be the dynamical system with all of the forces turned off; that is, the ideal gas system. In some cases the reference system can be solved analytically, and we refer to these methods as the Numerical Analytical Propagator Algorithm (NAPA). The development of symplectic, reversible RESPA (r-RESPA) integration methods grew out of our earlier attempts to devise multiple time scale integrators based on the generation of the dynamics of a reference system and, in principle, exact correction to it.[13, 14, 15, 16] The latter, being non-reversible, guided us in the direction of analyzing the structure of the classical propagator and the use of the symmetric Trotter factorization. In fact in the development of r-RESPA and r-NAPA we have adopted many of many of the strategies used in our earlier non-reversible RESPA (nr-RESPA).[17] All of these r-RESPA integrators are also symplectic. First we treat the problem where there are fast and slow degrees of freedom (or light and heavy particles). Then we treat the case where the forces can be subdivided into short and long range components. Finally, we show how the long and short range force factorizations can be combined with the fast and slow factorization yielding a speedup which is approximately the product of the speedups achieved when these factorizations are used separately. There are many variations on the theme introduced here.

2.4 Fast and Slow Processes

In many cases the dynamical system consists of fast degrees of freedom, labeled x, and slow degrees of freedom, labeled y. An example is that of a fluid containing polyatomic molecules. The internal vibrations of the molecules are often very fast compared to their translational and orientational motions. Although this and other systems, like proteins, have already been treated using RESPA,[17, 34, 22, 23, 24, 25, 26] another example, and the one we focus on here, is that of a system of very light particles (of mass m) dissolved in a bath of very heavy particles (mass M).[14] The positions of the heavy particles are denoted y and the positions of the light particles are denoted by x. In this case the total Liouvillian of the system is:

$$iL = iL_x + iL_y, \tag{19}$$

where

$$\begin{aligned} iL_x &= \dot{x}\frac{\partial}{\partial x} + F_x\frac{\partial}{\partial p_x} \\ iL_y &= \dot{y}\frac{\partial}{\partial y} + F_y\frac{\partial}{\partial p_y}. \end{aligned} \tag{20}$$

With this break up the reversible Trotter factorization of the propagator is

$$G_{xyx}(t) = e^{(iL_x+iL_y)\Delta t} = e^{iL_x\Delta t/2}e^{iL_y\Delta t}e^{iL_x\Delta t/2} + O(\Delta t^3). \tag{21}$$

For the slow (y) motion the time step Δt may be chosen large whereas for the fast motion this time will be too large. Thus this propagator can be expressed as;

$$G_{xyx}(\Delta t) = G_x(\frac{\Delta t}{2})G_y(\Delta t)G_x(\frac{\Delta t}{2}), \tag{22}$$

where

$$G_y(\Delta t) = e^{\frac{\Delta t}{2}F_y\frac{\partial}{\partial p_y}}e^{\Delta t\dot{y}\frac{\partial}{\partial y}}e^{\frac{\Delta t}{2}F_y\frac{\partial}{\partial p_y}}, \tag{23}$$

$$G_x(\Delta t/2) = \left[e^{\frac{\delta t}{4}F_x\frac{\partial}{\partial p_x}}e^{\frac{\delta t}{2}\dot{x}\frac{\partial}{\partial x}}e^{\frac{\delta t}{4}F_x\frac{\partial}{\partial p_x}}\right]^n, \tag{24}$$

and $\Delta t = n\delta t$, where n is a whole number. The r-RESPA integrator involves the following: the heavy particles are integrated for one half of one large time step $\Delta t/2$, the light particles are then integrated for n small time steps δt, such that $\Delta t = n\delta t$, and the heavy particles are integrated for one half of one

large time step. A simple example of fortran pseudocode for this is:

$$
\begin{aligned}
&v_h \leftarrow v_h + F_h \tfrac{\Delta t}{4m} \\
&x_h \leftarrow x_h + v_h \tfrac{\Delta t}{2} \\
&v_h \leftarrow v_h + F_h \tfrac{\Delta t}{4m} \\
&\text{do } i = 1, n \\
&\qquad v_l \leftarrow v_l + F_l \tfrac{\delta t}{2m} \\
&\qquad x_l \leftarrow x_l + v_l \delta t \\
&\qquad v_l \leftarrow v_l + F_l \tfrac{\delta t}{2m} \\
&\text{end do} \\
&v_h \leftarrow v_h + F_h \tfrac{\Delta t}{4m} \\
&x_h \leftarrow x_h + v_h \tfrac{\Delta t}{2} \\
&v_h \leftarrow v_h + F_h \tfrac{\Delta t}{4m}
\end{aligned}
$$

where we designate the fast motion by subscript l (standing for light particles) and the slow motion by subscript h (standing for heavy particles).

This procedure is very cost efficient when the fast (or light) particles are the dilute component because then one only has to update the forces on the heavy particles (the expensive part of the computation) every large time step instead of every small time step as would be the case in the straightforward application of the Verlet integrator. For example when applied to a system containing 64 particles of mass 1 dissolved in 800 solvent atoms of mass 100, the CPU time for the full simulation took only slightly longer than it would if the complete system was made up of heavy particles.[14] In contrast, application of the usual Verlet integrator using the small time step required for the light particles but evaluating all the forces after each one of these small time steps required approximately ten times the CPU time used in the RESPA integrator. The same accuracy was achieved in these two different treatments.

Another important application of this strategy was to the vibrational relaxation of a stiff diatomic molecule dissolved in a Lennard-Jones solvent. As is typical of such problems, the frequency of the oscillator can be an order of magnitude or more larger than the typical frequencies found in the spectral density of the solvent. Thus very small time steps are required to to integrate the equations of motion, but because there are very few accepting solvent modes at the frequency of the oscillator, its vibrational relaxation time will be very long, largely occurring by a multiphonon mechanism. In the past it was not practicable to simulate these processes directly. Using a form of r-RESPA modified for the specific case of an oscillator dissolved in a slow solvent, we have been able to reduce the CPU time required for these calculations by factors of ten in many cases making possible the direct simulation of such energy transfer problems.[34] When this strategy has been applied to the calculation of the IR and Raman spectrum of crystalline buckminster-

fullerene, speedups of as much as a factor of forty have been obtained.[35] It is important to note that the strategy outlined here is a direct generalization of the strategy we introduced in our original RESPA papers,[13] and is distinct from other attempts to deal with multiple time scales.

2.5 Long and Short Range Forces

Another immediate application of r-RESPA is to the case when the force can be subdivided into a short range part and a long range part. One way for effectuating this break up is to introduce a switching function, $s(x)$ that is unity at short inter-particle separations and 0 at large inter-particle separations. We introduced this strategy in our earlier non-reversible RESPA paper[15] where we expressed the total force as,

$$F(x) = s(x)F(x) + (1 - s(x))F(x) = F_s(x) + F_l(x). \tag{25}$$

The switching function $s(x)$ was taken to be a sigmoidal function (usually a cubic spline) whose inflection point (switching point) and skin-depth can be optimized. The short range force $F_s(x) = s(x)F(x)$ defines the time step to be used in a molecular dynamics calculation. In the velocity Verlet integrator one must compute the full force after each time step. If only the short range force were present, the CPU cost would be small because each particle would only interact with its nearest neighbors. It is the long range force $F_l(x) = (1 - s(x))F(x)$ which is costly to calculate. We introduced this strategy into the r-RESPA propagator factorization,[17] and as with the non-reversible RESPA, we showed that this can significantly reduce the CPU cost of the simulation.

Introducing the above force breakup into the Liouvillian gives,

$$iL = \dot{x}\frac{\partial}{\partial x} + F_s\frac{\partial}{\partial p_x} + F_l\frac{\partial}{\partial p_x} = iL_s + F_l\frac{\partial}{\partial p_x}. \tag{26}$$

The system defined by the Liouvillian L_s is called the reference system. Now applying the Trotter factorization to the propagator $\exp(iL_s + F_1\frac{\partial}{\partial p_x})\Delta t$ arising from this subdivision gives the new propagator,[17]

$$G_{lsl}(\Delta t) = e^{\frac{\Delta t}{2}F_l\frac{\partial}{\partial p_x}}e^{iL_s\Delta t}e^{\frac{\Delta t}{2}F_l\frac{\partial}{\partial p_x}}, \tag{27}$$

where with $\Delta t = n\delta t$

$$e^{iL_s\Delta t} = \left[e^{\frac{\delta t}{2}F_s\frac{\partial}{\partial p_x}}e^{\delta t\dot{x}\frac{\partial}{\partial x}}e^{\frac{\delta t}{2}F_s\frac{\partial}{\partial p_x}}\right]^n. \tag{28}$$

Thus the propagator in Eq. (27) produces the following dynamics algorithm:

$$
\begin{aligned}
& v \leftarrow v + F_l \tfrac{\Delta t}{2m} \\
& \text{do } i = 1, n \\
& \qquad v \leftarrow v + F_s \tfrac{\delta t}{2m} \\
& \qquad x \leftarrow x + v\,\delta t \\
& \qquad v \leftarrow v + F_s \tfrac{\delta t}{2m} \\
& \text{end do} \\
& v \leftarrow v + F_l \tfrac{\Delta t}{2m}.
\end{aligned}
$$

Note that while the velocities will be updated on two different time-scales, the positions will be updated using only the smallest time-step. This procedure[17] allows one to update the expensive long range force much less frequently than updating the cheap short range forces and thus saves CPU time without sacrificing accuracy. Even for simple systems like a liquid consisting of atoms interacting through a Lennard-Jones potential this procedure leads to a speedup of as much as 400%. It is important to note that if one takes the switching function to be a Heaviside function, an approximation not recommended, the factorization of the propagator introduced reduces to the so called Verlet I integrator introduced by Grubmuller *et al.*.[36] However, factorizations like the one in Sections 2.4 and 2.6 are distinct from the Verlet I integrator and are not treated in ref. [36]. This should dispell some confusion with respect to these issues.

It is worth calling to attention one difference between the force subdivision used in r-RESPA[17] and the one used in the original non-reversible RESPA.[15] In the non-reversible RESPA paper we included the value of the long range force at the beginning of the time interval into the reference system equation of motion which was then integrated for n small time steps. We then solved the correction equation involving the difference between the true force and the reference system force for one large time step. This was shown to lead to a more stable integration scheme with much smaller long time drift than when the long range force was not introduced into the reference set of equations. Unfortunately, in the r-RESPA factorization there is no way to introduce the long range force at the beginning of the interval into the reference system propagator because that would remove reversibility. Strategies are being developed to implement such effects in new reversible integrators [37].

2.6 Combining Force Subdivision and Dynamic Subdivision

The preceding breakup for light and heavy particles can be combined with breaking the forces up into short and long range forces in r-RESPA[17] in a similar manner to what was done in non-reversible RESPA.[16] We can then further factorize the three propagators appearing in Eq. (22) by using the factorization used to generate the velocity Verlet integrator with the forces

divided into short and long range parts. This yields the propagator for one large time step:[17]

$$G_{xyx}(\Delta t) = G_{lsl}^{(x)}\left(\frac{\Delta t}{2}\right) G_{lsl}^{(y)}(\Delta t) G_{lsl}^{(x)}\left(\frac{\Delta t}{2}\right), \tag{29}$$

where

$$G_{lsl}^{(x)}(\Delta t) = e^{\frac{\Delta t}{2} F_{xl}(x,y)\frac{\partial}{\partial p_x}} e^{iL_{xs}\Delta t} e^{\frac{\Delta t}{2} F_{xl}(x,y)\frac{\partial}{\partial p_x}}, \tag{30}$$

$$G_{lsl}^{(y)}(\Delta t) = e^{\frac{\Delta t}{2} F_{yl}(x,y)\frac{\partial}{\partial p_y}} e^{iL_{ys}\Delta t} e^{\frac{\Delta t}{2} F_{yl}(x,y)\frac{\partial}{\partial p_y}}, \tag{31}$$

and the middle propagator for x is integrated as:

$$e^{iL_{xs}\Delta t} = \left[e^{\frac{\delta t}{2} F_{xs}\frac{\partial}{\partial p_x}} e^{\delta t \dot{x}\frac{\partial}{\partial x}} e^{\frac{\delta t}{2} F_{xs}\frac{\partial}{\partial p_x}} \right]^{n_1}. \tag{32}$$

Likewise, the middle propagator for y is integrated n_2 times with a time step of $\Delta t_1 = n_1 \delta t$

$$e^{iL_{ys}\Delta t} = \left[e^{\frac{n_1\delta t}{2} F_{xs}\frac{\partial}{\partial p_y}} e^{n_1 \delta t \dot{y}\frac{\partial}{\partial y}} e^{\frac{n_1\delta t}{2} F_{ys}\frac{\partial}{\partial p_y}} \right]^{n_2}. \tag{33}$$

Thus $\Delta t = n_2 \Delta t_1 = n_1 n_2 \delta t$. It is simple matter to write down the Fortran pseudocode for this breakup.

2.7 The Applications of RESPA to Proteins and Chemical Systems

In order to apply the techniques discussed above to the MD simulation of biomolecules, one takes the Liouville operator for a macromolecule in *vacuo* containing N atoms to be

$$iL = \sum_{i}^{3N} [\dot{x}_i \frac{\partial}{\partial x_i} + F_i(x)\frac{\partial}{\partial p_i}\ ], \tag{34}$$

where

$$F(x) = F_{\text{stret}}(x) + F_{\text{bend}}(x) + F_{\text{tors}}(x) + F_{\text{Hbond}}(x) + F_{\text{vdW}}(x) + F_{\text{elec}}(x)\ , \tag{35}$$

$F_{\text{stret}}, F_{\text{bend}}, F_{\text{tors}}, F_{\text{Hbond}}, F_{\text{vdW}}$, and F_{elec} represent the forces for stretch, bending, torsion (including improper torsion), hydrogen-bonding, van der Waals, and electrostatic interactions, respectively. Their functional forms can be found elsewhere [38, 39]. The databases of parameters for these functional forms are generally called force fields. There are several force fields available for biomolecular simulations, such as AMBER [39] , OPLS [40] and CHARMM [41], etc.

In an atomic level simulation, the bond stretch vibrations are usually the fastest motions in the molecular dynamics of biomolecules, so the evolution of the stretch vibration is taken as the "reference" propagator with the smallest time step. The nonbonded interactions, including van der Waals and electrostatic forces, are the slowest varying interactions, and a much larger time-step may be used. The bending, torsion and hydrogen-bonding forces are treated as intermediate time-scale interactions.

In addition, the non-bonded forces can be divided into several regions according to pair distances. The near region is normally more important than the distant region because the non-bonded forces decay with distance. Since most of the CPU time in a MD simulation is spent in the calculation of these non-bonded interactions, the separation in pair distance results in valuable speedups. Using a 3-fold distance split, the non-bonded forces are separated in 3 regions: near, medium, and far distance zones. Thus, the Liouville operator can be express as a sum of five terms

$$L = L_1 + L_2 + L_3 + L_4 + L_5 \ , \tag{36}$$

where

$$iL_1 \equiv \dot{x}\frac{\partial}{\partial x} \ + F_1(x)\frac{\partial}{\partial p} \tag{37}$$

$$iL_i \equiv F_i(x)\frac{\partial}{\partial p}, \qquad i = 2,3,4,5 \tag{38}$$

and

$$F_1(x) \equiv F_{\text{stret}}(x) \tag{39}$$

$$F_2(x) \equiv F_{\text{bend}}(x) + F_{\text{tors}}(x) + F_{\text{Hbond}}(x) \tag{40}$$

$$F_3(x) \equiv F_{\text{vdW}}^{\text{near}}(x) + F_{\text{elec}}^{\text{near}}(x) \tag{41}$$

$$F_4(x) \equiv F_{\text{vdW}}^{\text{med}}(x) + F_{\text{elec}}^{\text{med}} \tag{42}$$

$$F_5(x) \equiv F_{\text{vdW}}^{\text{far}}(x) + F_{\text{elec}}^{\text{far}} \ . \tag{43}$$

To separate the non-bonded forces into near, medium, and far zones, pair distance separations are used for the van der Waals forces, and box separations are used for the electrostatic forces in the Fast Multipole Method,[24] since the box separation is a more convenient breakup in the Fast Multipole Method (FMM). Using these subdivisions of the force, the propagator can be factorized according to the different intrinsic time scales of the various components of the force. This approach can be used for other complex systems involving long range forces.

2.8 Efficient Integrators for Systems with Coulomb Potentials

One of the most expensive parts of a MD or MC simulations is the computation of long range interactions. Since the CPU time required for the

calculation of these forces scales as $O(N^2)$, where N is the number of force centers in the system, direct calculation of these forces in large systems makes molecular dynamics (or Monte Carlo) infeasible for large protein-water systems. The standard approach has been to truncate the long range forces so that their calculation scales as $O(N)$ for large enough systems. Unfortunately, truncation introduces significant non physical effects. To eliminate surface effects and to avoid the errors caused by truncation it is now becoming common to use periodic boundary conditions and to invoke Ewald summation. Optimal application of Ewald summation also scales as $O(N^{3/2})$ and thus becomes prohibitively expensive for large systems. Procacci and Marchi have combined Ewald with RESPA for protein solutions by including the total Fourier sum in the intermediate time loop.[42] A better strategy for applying r-RESPA to Ewald boundary conditions involves subdividing the Fourier space sum in such a way that the short time contribution is placed in the inner short time loop of RESPA and the "true" long range and slow part of the sum is put in the outer loop.[18]

There are three different algorithms for the calculation of the electrostatic forces in systems with periodic boundary conditions: (a) the (optimized) Ewald method, which scales like $O(N^{3/2})$; (b) the Particle Mesh Ewald (PME) method, which scales like $O(N \log N)$; and (c) the periodic Fast Multipole Method (PFMM), which scales like $O(N)$. For very large systems ($N \gtrsim 10^5$) it is expected that the PFMM will be the best choice, given its linear algorithmic complexity. It is of interest to determine the break-even point for these two methods. Because PME scales as $O(N \ln N)$ and periodic-FMM scales as $O(N)$, PFMM will be faster than PME for N greater than some N_0. The break-even point for these two methods combined with r-RESPA will be different because the implementation of r-RESPA will be different in these two cases. This break-even point has not yet been determined systematically. Figueirido *et al.* estimated that the break-even point for protein-water systems is $N_0 \approx 20,000$. Despite the significant progress in this field the optimal strategy has yet to be found.

Fast Multipole Methods To manage the calculation of all of the electrostatic interactions, several groups have experimented with approximate schemes, of which the most widely used is the Fast Multipole Method (FMM) of Greengard and Rokhlin[43, 44] and its variants.[45, 46, 47, 48, 49] This algorithm decreases the computational burden to $O(N)$ by cleverly exploiting a hierarchy of clusters and using multipolar expansions to approximate the potential produced by these clusters. The basic principle of FMM is rather elegant. It interpolates the potential and force on a particular charge due to distant charges not by direct calculation, but by using the local expansion of fields produced by the multipoles generated from those distant charges. It first organizes multipole representations of charge distributions in hierarchically structured boxes, then transforms these multipoles into local field

expansions. Each particle then interacts with the local field of distant multipoles. Meanwhile, the near range interactions are calculated directly by pairwise evaluation. Thus, the potential (and force) consists of two parts:

$$\Phi(\boldsymbol{x}) = \Phi^{\text{near}}_{\text{direct}}(\boldsymbol{x}) + \Phi^{\text{far}}_{\text{multipole}}(\boldsymbol{x}) \tag{44}$$

where $\Phi^{\text{near}}_{\text{direct}}$ contains the near range inter-particle interactions, and $\Phi^{\text{far}}_{\text{multipole}}$ contains the contribution from distant particles. A top-down FMM recursive method was proposed for multipole generation by Zhou and Berne[24]. Their method is based on White and Head-Gordon's simplified derivation [45]. Zhou and Berne have incorporated r-RESPA in this top-down FMM algorithm and applied it to isolated all-atom proteins.[24] They were able to achieve speed-ups on the order of fifteen-fold for the photo-synthetic reaction center over the direct UN-truncated calculation of the forces using the standard velocity verlet integrator. Fig 1 shows a comparison between the cpu times required by ordinary velocity verlet and r-RESPA for different size proteins. The figure also shows improvements that can be achieved by combining efficient algorithms such as the fast-multipole method (FMM) with r-RESPA.

The fast multipole method was first extended to periodic systems by Schmidt and Lee.[50] Figueirido *et al.* also designed a periodic FMM with a full derivation of the local field expansion which scales as $O(N)$.[26] These authors combined PFMM with r-RESPA producing in a very powerful algorithm (r-RESPA/PFMM) that is expected to be the optimum strategy for dealing with very large systems (see below).

Particle Mesh Ewald Methods Recently the particle mesh Ewald method (PME), and a smooth variant of it (SPME), developed by Darden *et al.*., have been described in the literature [51, 52, 53, 46]. These algorithms are based on Hockney and Eastwood's [54] idea of assigning charges to a mesh according to their real space positions; the CPU time savings come from applying the Fast Fourier Transform (FFT) to the particle mesh to accelerate the reciprocal-space calculations of the Ewald sum and to use a small cutoff in real space. The algorithms are found to be of order $O(N \log N)$. This method has been combined with r-RESPA by Procacci, Darden and Marchi.[55]

3 New Sampling Methods for the Extrinsic Multiple Time Scale Problem

Biomolecular systems often have rough energy landscapes. The sampling of rugged energy landscape poses special problems for molecular dynamics and Monte Carlo. As the system moves from one potential energy basin to another it must cross barriers that are large compared to kT. The crossing of such barriers are rare events and thus very long runs are required to sample the configuration space. In such systems the barriers are due to at least two

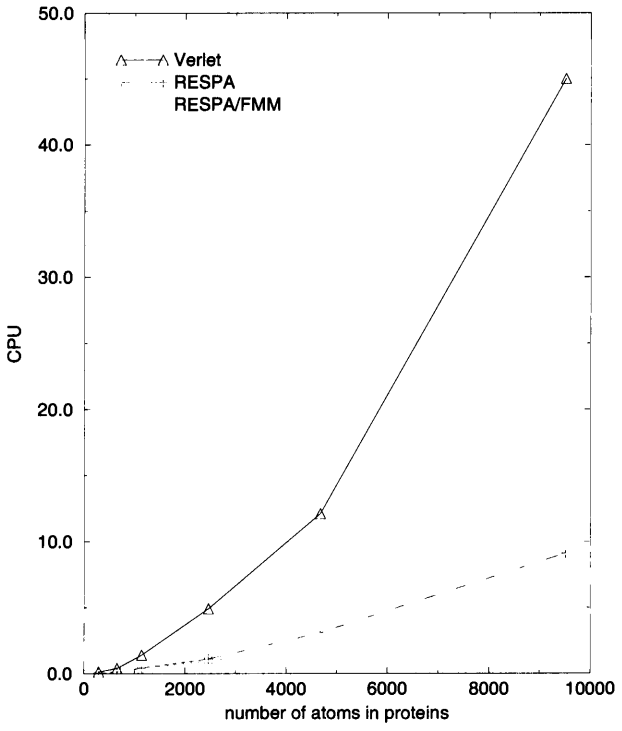


Fig. 1. CPU times (in hours) for 1 ps MD runs for various proteins using three different methods, direct velocity Verlet with a time-step 0.5 fs, r-RESPA with direct evaluation of electrostatic forces and an overall time-step of 4.0 fs, and r-RESPA/TFMM with an overall time-step 4.0 fs (combination of (2,2,2,2) in force breakup).The energy conservation parameter $\log \Delta E$ for the three methods are comparable. The CPU time (hours) is for RISC6000 /MODEL 590 computer.

classes of interactions. First there are the local barriers that separate stable states of the torsion angles. Then there are barriers arising from close encounters of atoms on side chains as well as on the primary chain which result from very repulsive (r^{-12}) non-bonded interactions. There is a long history of using fictitious dynamics to sample the configuration space of complex systems. These schemes fall into three classes: Brownian or Langevin dynamics; BGK (Bhatnanger, Gross, Krook) dynamics; and Monte Carlo methods. One can accelerate all of these methods by a clever break-up of the forces, as we have done for molecular dynamics with RESPA and r-RESPA, but this will not solve the problem of sampling the rare barrier crossing events frequently enough to determine long time averages needed for determining the thermodynamic averages or for gaining insights into reaction paths between initial

and final conformational states of large molecules. New methods are required to deal with rugged energy surfaces.

New Monte Carlo methods have been devised to deal with the intrinsic multiple time scale problem.[56, 10] The Hybrid Monte Carlo (HMC) method, combined with r-RESPA, as outlined in Sec. 2, can be used to speedup the intrinsic multiple time scale problem in MC sampling. In addition, new methods for speeding up barrier crossings in systems with rough energy landscapes like the Jump-Walking (J-Walking) method[57] and the Smart-Walking (S-Walking) method[58] can be combined with HMC and thereby r-RESPA. These methods are particularly useful for sampling the conformation space of many-body systems, such as proteins.[24, 26]Lastly, it is worth mentioning methods that allow the Lennard-Jones diameters σ to fluctuate or that allow the barriers in the torsion angle potential to fluctuate. These methods very rapidly explore the configuration space.[59]

Hybrid Monte Carlo In standard MC only single particle moves are tried and accepted or rejected. Attempts to make many particle moves of the system before applying the Metropolis acceptance criterion leads to such small acceptance probabilities that this method is not efficient. Moreover it requires the recalculation of the whole potential after each attempted move, a costly computation especially when the move is likely to be rejected. One efficient method for generating collective moves is the Hybrid Monte Carlo method invented by Duane and Kennedy.[12] In this method one starts with a configuration of the system and samples momenta of the particles from a Maxwell distribution. Molecular dynamics is used to move the whole system for a time Δt and, because this time may be sufficiently large as to cause a reasonable energy change due the lack of strict energy conservation, one then accepts or rejects the move using the Metropolis criterion based on $\exp(-\beta H)$ where H is the hamiltonian of the system. This step is repeated over and over. In HMC, bad MD is used to generate efficient MC. It is important that the integrator used for generating the solution to the equations of motion be reversible because only then will this method satisfy detailed balance and only then will the method generate the canonical distribution and the Boltzmann distribution. A number of authors have further elaborated the HMC method.[60, 61, 62, 63]

Since many systems of interest in chemistry have intrinsic multiple time scales it is important to use integrators that deal efficiently with the multiple time scale problem. Since our multiple time step algorithm, the so-called reversible Reference System Propagator Algorithm (r-RESPA) [17, 24, 18, 26] is time reversible and symplectic, they are very useful in combination with HMC for constant temperature simulations of large protein systems.

In HMC the momenta are constantly being refreshed with the consequence that the accompanying dynamics will generate a spatial diffusion process superposed on the inertial dynamics, as in BGK or Smoluchowski dynamics. It is well known from the theory of barrier crossing that this added spatial

diffusion can lead to smaller rates for barrier crossing. Thus the HMC or BGK methods may suppress barrier crossing. Parenthetically, it is important to note that stochastic methods such as Langevin dynamics or BGK dynamics will behave similarly. We have found[64] that in some systems the Nose' thermostat[65] may have a more beneficial sampling of different basins. Reversible integration schemes for these methods have been developed.[66] One way to improve these methods is to couple them to new methods for accelerating the dynamics on rugged energy landscapes such as the J-Walking method[57, 67, 68] or the S-Walking method.[58]

Jump Walking In the J-walking method, the MC or HMC sampling at the desired low temperature is infrequently punctuated by sampling from a higher temperature distribution for the same system. Since a higher temperature MC simulation can involve larger attempted moves and more frequent barrier crossings, this allows the system to access more conformational states according to the high temperature Boltzmann distribution. Then, the lower temperature walker attempts occasional jumps to the conformation states of the high temperature walker, thus enhancing the barrier crossing. The trial sampling distribution for these occasional jumps is the Boltzmann distribution at the higher temperature. The method is so constructed that one generates the correct low temperature Boltzmann distribution. Since the energy landscape of biomolecules contains very high barriers, it is often necessary to use many high temperature walks spaced at intervals of approximately 50 K and the CPU time required by this method will scale as the number of high temperature walks.

Smart Walking Jumping directly into a high temperature structure is not the only way to use the conformational space information from the J-Walker. Instead, the structure can be first relaxed before being jumped into.[58] Approximate minimization with a steepest descent method (or conjugate gradient method) will generate structures close to the local minimum. These relaxed configurations will significantly decrease the potential energy, and thus increase the jump success ratio dramatically. Each minimized structure is then regarded as one of the possible trial moves at low temperature and are accepted or rejected with acceptance probability function, that generates a Boltzmann distribution at the low temperature. Unlike the J-Walking acceptance probability, this scheme, which is called Smart Walking[58] (or S-Walking), will dramatically increase the jump success ratio from one basin to another. It also enables the system to explore more phase space and undergo more efficient barrier-crossings. This S-walking method avoids the linear increase of CPU time and memory usage required by the multiple-stage J-Walking method, because it is not necessary to use multiple stages for most systems, even though it would be very easy to implement a multi-stage S-Walking procedure. S-Walking preserves detailed balance approximately

provided the time between S-jumps is much longer than the time required by the low temperature walker to explore its local basin effectively (more discussion follows in the section on results). This new S-Walking algorithm only requires a simple modification of the J-Walking algorithm.

4 Summary

Integrators based on r-RESPA, when combined with enhanced methods for calculating long-range electrostatic forces, such as the FMM or SPME schemes have led to a considerable speed-up in the CPU time for large scale simulations of biomacromolecular solutions. Since r-RESPA is symplectic such integrators are very stable. Moreover since r-RESPA is time reversible it can be used in Hybrid Monte Carlo and satisfies the condition of detailed balance. This HMC method can be used in enhanced sampling methods such as J-Walking and S-Walking methods which lead to a more rapid exploration of rugged energy landscapes and thus to enhanced conformational searches.

Acknowledgements

This work was supported by grants from the National Science Foundation and the National Institutes of Health.

References

1. B. J. Alder and T. E. Wainwright. *J. Chem. Phys.*, 27:1208–9, 1957.
2. B. J. Alder and T. E. Wainwright. *J. Chem. Phys.*, 31:459–56, 1959.
3. A. Rahman. *Phys. Rev.*, 136A:405–11, 1964.
4. G. D. Harp and B. J. Berne. *J. Chem. Phys.*, 49:1249–54, 1968.
5. G. D. Harp and B. J. Berne. *Phys. Rev.*, A2:975–96, 1970.
6. A. Rahman and F. H. Stillinger. *J. Chem. Phys.*, 55:3336–59, 1971.
7. M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids.* Oxford University Press, Oxford, 1987.
8. J. A. McCammon and S. C. Harvey. *Dynamics of Proteins and Nucleic Acids.* Cambridge University Press, Cambridge, 1987.
9. J. M. Haille. *Molecular Dynamics Simulation: Elementary Methods.* John Wiley & Sons, USA, 1992.
10. D. Frenkel and B. Smit. *Understanding Molecular Simulation.* Academic Press, 1996.
11. J. P. Ryckaert and G. Ciccotti and H. J. C. Berendsen. *J. Comp. Phys.*, 23:327, 1977.
12. S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth. *Phys. Rev. Lett. B*, 195:216–222, 1987.
13. M. E. Tuckerman, G. J. Martyna, and B. J. Berne. *J. Chem. Phys.*, 93:1287–1291, (1990).

14. M. Tuckerman, B. J. Berne, and A. Rossi. *J. Chem. Phys.*, 94:1465–1469, (1991).
15. B. J. Berne M. Tuckerman and G. Martyna. *J. Chem. Phys.*, 94:6811–6815, (1991).
16. M. Tuckerman and B. J. Berne. *J. Chem. Phys.*, 95:8362–8364, (1991).
17. M. E. Tuckerman, B. J. Berne, and G.J. Martyna. *J. Chem. Phys.*, 97:1990–2001, (1992).
18. S. J. Stuart, R. Zhou, and B. J. Berne. *J. Chem. Phys.*, 105:1426–1436, (1996).
19. M.E. Tuckerman and M. Parrinello. *J. Chem. Phys.*, 101:1302–1315, (1994).
20. M. E. Tuckerman and G. J. Martyna and B. J. Berne. *J. Chem. Phys.*, 93:1287, 1990.
21. J. Hutter, M.E. Tuckerman, and M. Parrinello. *J. Chem. Phys.*, (1995).
22. D. Humphreys, R. A. Friesner, and B.J. Berne. *J. Phys. Chem.*, 98:6885–6892, (1994).
23. D. Humphreys, R. A. Friesner, and B. J. Berne. *J. Phys. Chem.*, 99:10674–10685, (1995).
24. R. Zhou and B.J. Berne. *J. Chem. Phys.*, 103:9444–9458, (1995).
25. M. Watanabe and M. Karplus. *J. Chem. Phys.*, 99:8063–8074, 1993.
26. F. Figueirido, R. Zhou, B. J. Berne, and R. M. Levy. *J. Chem. Phys.*, 106:9835–9849, (1997).
27. L. Verlet. *Phys. Rev.*, 159:98–103, 1967.
28. T. A. Andrea, W. C. Swope, and H. C. Andersen. *J. Chem. Phys.*, 79:4576–4584, 1983.
29. H. Goldstein. *Classical Mechanics.* Addison Wesley Publishing Company, Inc., 1980.
30. Haruo Yoshida. *Phys. Lett. A.*, 150:262–268, (1990).
31. J. J. Biesiadecki and R. D. Skeel. *J. Comp. Phys.*, 109:318–328, 1993.
32. S. K. Gray, D. W. Noid, and B. G. Sumpter. *J. Chem. Phys.*, 101:4062–4072, 1994.
33. S. Auerbach and A. Friedman. *J. Comp. Phys.*, 93:189, 1991.
34. M.E. Tuckerman and B. J. Berne. *J. Chem. Phys.*, 98:7301–7318, (1993).
35. P. Procacci and B.J. Berne. *J. Chem. Phys.*, 101:2421, (1994).
36. H. Grubmuller, H. Heller, A. Windemuth, and K. Schulten. *Molecular Simulation*, 6:121–142, (1991).
37. G.J. Martyna and M.E. Tuckerman. *J. Chem. Phys.*, 102:8071–8077, (1996).
38. F. Mohamadi and N. G. J. Richards and W. C. Guida and R. Liskamp and M. Lipton and C. Caufield and G. Chang and T. Hendrickson and W. C. Still. *J. Comp. Chem.*, 11:440, 1990.
39. S. J. Weiner and P. A. Kollman and D. T. Nguyen and D. A. Case. *J. Comp. Chem.*, 7:230, 1986.
40. W. L. Jorgensen and J. Tirado-Rives. *J. Am. Chem. Soc.*, 110:1657, 1988.
41. B. R. Brooks and R. E. Bruccoeri and B. D. Olafson and D. J. States and S. Swaminathan and M. Karplus. *J. Comp. Chem.*, 4:187–217, 1983.
42. P. Procacci and M. Marchi. *J. Chem. Phys.*, 104:3003–3012, 1996.
43. L. Greengard. *The Rapid Evaluation of Potential Fields in Particle Systems.* The MIT Press, Cambridge, Massachusetts, 1988.
44. L. Greengard and V. Rokhlin. *J. Comp. Phys.*, 73:325, 1987.
45. C. A. White and M. Head-Gordon. *J. Chem. Phys.*, 101:6593–6605, 1994.
46. J. Shimada and H. Kaneko and T. Takada. *J. Comp. Chem.*, 15:28, 1994.

47. H.-Q. Ding and N. Karasawa and W. A. Goddard III. *J. Chem. Phys.*, 97:4309, 1992.
48. M. Saito. *Molecular Simulations*, 8:321–333, 1992.
49. C. Niedermeier and P. Tavan. *Molecular Simulation.*, 17:57–66, (1996).
50. K. E. Schmidt and M. A. Lee. *J. Stat. Phys.*, 63:1223–1235, 1991.
51. T. Darden, D. M. York, and L. G. Pedersen. Particle mesh Ewald: An $N\log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.*, 98:10089–10092, 1993.
52. H. G. Petersen. *J. Chem. Phys.*, 103:3668–3679, 1995.
53. U. Essman, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. *J. Chem. Phys.*, 103:8577–8593, 1995.
54. R. W. Hockney and J. W. Eastwood. *Computer Simulation Using Particles.* Adam Hilger, Bristol-New York, 1989.
55. P. Procacci and T. Darden and M. Marchi. *J. Phys. Chem.*, 100:10464–10468, 1996.
56. B. J. Berne and John E. Straub. *Current Topics in Structural Biology*, 7 No 2:181–189, (1997).
57. D. D. Frantz, D. L. Freeman, and J. D. Doll. *J. Chem. Phys.*, 93:2769–2784, 1990.
58. R. Zhou and B. J. Berne. *J. Chem. Phys.*, 107:9185–9196, (1997).
59. Z. Liu and B. J. Berne. *J. Chem. Phys.*, 99:6071, (1993).
60. U. H. E. Hansmann, Y. Okamoto, and F. Eisenmenger. *Chem. Phys. Lett.*, 259:321–330, 1996.
61. R. M. Neal. *J. Comp. Phys.*, 111:194–203, 1994.
62. S. Gupta, A. Irback, F. Karsch, and B. Petersson. *Phys. Lett. B*, 242:437–443, 1990.
63. P. B. Markenzie. *Phys. Lett. B*, 226:369–371, 1989.
64. M. Tuckerman, B. J. Berne, G. Martyna, and M. Klein. *J. Chem. Phys.*, 99:2796–2784, (1993).
65. S. Nosé. *J. Chem. Phys.*, 81:511–519, 1984.
66. G.J. Martyna, M.E. Tuckerman, D.J. Tobias, and M.L. Klein. *Mol. Phys.*, 87:1117–1157, (1996).
67. D. L. Freeman, D. D. Frantz, and J. D. Doll. *J. Chem. Phys.*, 97:5713, 1992.
68. A. Matro, D. L. Freeman, and R. Q. Topper. *J. Chem. Phys.*, 104:8690, 1996.

The Five Femtosecond Time Step Barrier*

Robert D. Skeel and Jesús A. Izaguirre

Beckman Institute and Department of Computer Science
University of Illinois at Urbana-Champaign

Abstract. Simulation of the dynamics of biomolecules requires the use of a time step in the range 0.5–1 fs to obtain acceptable accuracy. Nevertheless, the bulk of the CPU time is spent computing interactions, such as those due to long-range electrostatics, which vary hardly at all from one time step to the next. This unnecessary computation is dramatically reduced with the use of multiple time stepping methods, such as the Verlet-I/r-RESPA method, which is based on approximating "slow" forces as widely separated impulses. Indeed, numerical experiments show that time steps of 4 fs are possible for these slow forces but unfortunately also show that a long time step of 5 fs results in a dramatic energy drift. Moreover, this is less pronounced if one uses a yet larger long time step! The cause of the problem can be explained by exact analysis of a simple two degree-of-freedom linear problem, which predicts numerical instability if the time step is just less than half the period of the fastest normal mode. To overcome this, a modification of the impulsive Verlet-I/r-RESPA method is proposed, called the mollified impulse method. The idea is that one modifies the slow part of the potential energy so that it is evaluated at "time averaged" values of the positions, and one uses the gradient of this modified potential for the slow part of the force. Various versions of the algorithm are implemented for water and numerical results are presented.

1 Introduction

Answers to questions in structural biology are often sought by means of long time biomolecular simulations using empirical classical mechanical force fields. Nevertheless, the shortest time scales present in the simulation have limited the integration step to 0.5–1 fs, which is many orders of magnitude smaller than the desired simulation time interval. Fairly recently the introduction of the Verlet-I [8, 9]/r-RESPA [23] "impulse" multiple time step (MTS) method has permitted an increase to 4 fs in the length of the longest time step. This article describes this development as well as empirical [4] and analytical [7] evidence indicating that timesteps of 5 fs or greater are not possible with the impulse MTS method (for unconstrained molecular dynamics). Nevertheless, a modification to this method – the "mollified" impulse method [7] – yields stable dynamics for longest timesteps as great as 7 fs.

MTS methods exploit the existence of different time scales arising from the many interactions present in the force field. For expository purposes assume

* This work was supported in part by NIH Grant P41RR05969 and NSF Grants BIR-9318159 and DMS-9600088.

that the potential energy function is expressed as the sum of just two parts $U^{\text{fast}}(x) + U^{\text{slow}}(x)$, where x denotes the collection of all atomic positions. The equations of motion for unconstrained constant NVE dynamics are

$$M\frac{d^2}{dt^2}X = F^{\text{fast}}(X) + F^{\text{slow}}(X) \tag{1}$$

where M is a diagonal matrix of atomic masses, $x = X(t)$ are the atomic trajectories, $F^{\text{fast}} = -U_x^{\text{fast}}$, and $F^{\text{slow}} = -U_x^{\text{slow}}$. The partitioning is chosen so that an appropriate time step Δt for the slow part is significantly larger than an appropriate time step δt for the fast part. To be specific about the length of an appropriate time step, there is a recipe in [17], which suggests that Δt^2 be chosen to be proportional to the reciprocal of the largest eigenvalue of the mass weighted Hessian $M^{-1/2}U_{xx}^{\text{slow}}(x)M^{-1/2}$, and analogously for δt^2.

Although unconstrained dynamics is being considered here, the ideas extend to the case where bond lengths (and bond angles) are constrained. Also, the ideas are applicable to other than constant NVE simulations.

This article is organized as follows: Sect. 2 explains why it seems important to use symplectic integrators, Sect. 3 describes the Verlet-I/r-RESPA impulse MTS method, Sect. 4 presents the 5 femtosecond time step barrier, Sect. 5 introduce a possible solution termed the "mollified" impulse method (MOLLY), and Sect. 6 gives the results of preliminary numerical tests with MOLLY.

2 Importance of Symplectic Integration

It is appropriate to consider first the question of what kind of accuracy is expected from a simulation. In molecular dynamics (MD) very small perturbations to initial conditions grow exponentially in time until they completely overwhelm the trajectory itself. Hence, it is inappropriate to expect that accurate trajectories be computed for more than a short time interval. Rather it is expected only that the trajectories have the correct statistical properties, which is sensible if, for example, the initial velocities are randomly generated from a Maxwell distribution.

The use of a numerical integrator to approximate the exact propagator of a system of ordinary differential equations (ODEs) yields a numerical solution which can be interpreted as the exact solution of a slightly different system of ODEs.[1] If the given system is a Hamiltonian system (as it is for constant-energy MD), then the slightly different system is Hamiltonian if and only if the integrator is symplectic [21]. In particular, this implies that any given energy surface in phase space is changed only slightly by the use of symplectic

[1] This statement is not exactly true – the slightly different system of ODEs is defined by an asymptotic expansion in powers of Δt which is generally divergent.

numerical integration, and it suggests that statistical properties of long-time dynamics are retained.

A transformation in phase space from positions x and momenta p

$$\bar{x} = \xi(x,p), \quad \bar{p} = \pi(x,p) \tag{2}$$

is said to be *symplectic* if its Jacobian matrix (of partial derivatives) satisfies

$$\begin{bmatrix} \xi_x & \xi_p \\ \pi_x & \pi_p \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix} \begin{bmatrix} \xi_x & \xi_p \\ \pi_x & \pi_p \end{bmatrix} = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}. \tag{3}$$

The exact propagator for a Hamiltonian system for any given time increment Δt is symplectic. As a consequence it possesses the Liouville property of preserving volume in phase space.

Additional evidence favoring the use of symplectic integrators is an observation [18] concerning hybrid Monte Carlo methods: to get the property of detailed balance needed for valid sampling, it is enough to use a numerical integrator which is volume preserving and reversible. As was just mentioned, symplectic integrators preserve volume in phase space.

On the basis of these considerations it is suggested that the design of numerical integration techniques might employ the following two hypotheses:

1. desired computables are insensitive to small changes in the Hamiltonian $\frac{1}{2}p^{\mathrm{T}}M^{-1}p + U(x)$, and
2. symplectic numerical integration produces small changes to the Hamiltonian.

3 The Verlet-I/r-RESPA Impulse Method

For MD the reduced system

$$M\frac{d^2}{dt^2}X = F^{\text{fast}}(X) \tag{4}$$

is cheap to integrate compared to the cost of integrating the full system. For example, for the MD program NAMD [19] on an ATM-connected cluster of eight HP 9000-735/125 workstations applied to a system of 36,000 atoms, the CPU time required to compute the "fast" short-range forces is 7 seconds whereas that required for the "slow" long-range forces is 29 seconds. Short-range forces consisted of all forces except electrostatic forces beyond 8 Å. The latter were calculated using a parallel enhanced fast multipole method [15]. An efficient numerical integrator would evaluate F^{slow} sparingly and incorporate its values into an integration of the reduced system in the most advantageous way. The question is how to do this.

A good place to start the discussion is the "impulse method"

$$M\frac{d^2}{dt^2}X = F^{\text{fast}}(X) + \sum_{m=-\infty}^{\infty} \Delta t\, \delta(t - m\Delta t) F^{\text{slow}}(X) \tag{5}$$

proposed by [24], in which the slow force is approximated by a sequence of appropriately weighted impulses. If the reduced system is analytically solvable, then the above approximation to the full system is solvable by inserting jumps into the analytical integration of $(d/dt)X(t)$ at integer multiples of the time step Δt. Let $P = M(d/dt)X$, and suppose values X^{n-1}, P^{n-1}, and $F^{\text{slow},n-1} = F^{\text{slow}}(X^{n-1})$ are given. Then one step of the impulse method can be expressed as follows:

half a kick

$$P^{n-1+\epsilon} = P^{n-1} + \frac{\Delta t}{2} F^{\text{slow},n-1}. \tag{6}$$

a vibration Propagate X^{n-1}, $P^{n-1+\epsilon}$ by integrating

$$\frac{d}{dt}X = M^{-1}P, \quad \frac{d}{dt}P = F^{\text{fast}}(X) \tag{7}$$

for an interval Δt to get X^n, $P^{n-\epsilon}$.

half a kick

$$F^{\text{slow},n} = F^{\text{slow}}(X^n), \tag{8}$$

$$P^n = P^{n-\epsilon} + \frac{\Delta t}{2} F^{\text{slow},n}. \tag{9}$$

The symbols $P^{n-1+\epsilon}$ and $P^{n-\epsilon}$ represent momenta just after the $(n-1)$st kick, and just before the nth kick, respectively.

The application in [24] is to celestial mechanics, in which the reduced problem for U^{fast} consists of the Keplerian motion of planets around the sun and in which the U^{slow} impulses account for interplanetary interactions. Application to MD is explored in [14]. It is not easy to find a reduced problem that can be integrated analytically however. The choice $U^{\text{fast}} \equiv 0$ is always possible and this yields the simple but effective leapfrog/Störmer/Verlet method, whose use according to [22] dates back to at least 1793 [5]. This connection should allay fears concerning the quality of an approximation using Dirac delta functions.

The need to integrate the fast forces analytically can be avoided by approximating it by a sequence of suitably weighted impulses but more closely spaced in time than the slow force impulses:

$$M\frac{d^2}{dt^2}X = \sum_{m=-\infty}^{\infty} \delta t\, \delta(t - m\delta t) F^{\text{fast}}(X) + \sum_{n=-\infty}^{\infty} \Delta t\, \delta(t - n\Delta t) F^{\text{slow}}(X). \tag{10}$$

The idea is illustrated by Fig. 1. These equations constitute a readily understandable and concise representation of the widely used Verlet-I/r-RESPA impulse MTS method. The method was described first in [8, 9] but tested

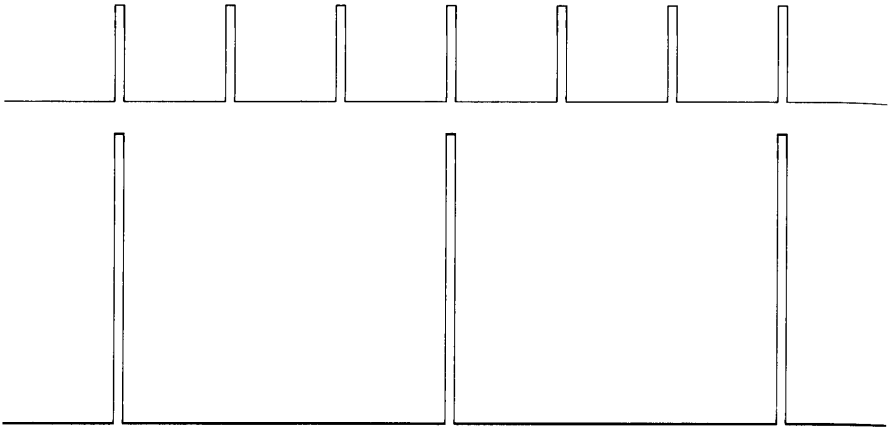


Fig. 1. Schematic for the impulse multiple time stepping method.

first in [23]. This latter reference describes the method in terms of an operator splitting and proposes the splitting of nonbonded interactions using switching functions. The symplecticness of this method is first noted in [3]. Of course, the idea generalizes to more than two different time steps. Also, it is most practical to choose time steps with ratios that are integers.

4 Time Step Barriers

Based on accuracy considerations alone, it seems that time steps of the order $\Delta t = 16\,\text{fs}$ should be possible for the slowest forces for MD [6]. Contrast this with the 0.5–1 fs value needed for the fastest forces.

Nevertheless, it was recognized already in the first paper on the impulse method [9] that resonance might be expected if the timestep Δt is approximately equal to the period of the fastest normal mode, which is 9 or 10 fs for biomolecules with flexible bonds to hydrogen. (Flexible bonds to hydrogen give modes with the highest frequency because bond stretching interactions have the largest force constants and hydrogen atoms have the smallest mass.) This resonance can be most readily explained if one assumes δt is infinitesimal and neglects the dependence of the slow force on X. Then, the impulse MTS method simplifies to

$$M\frac{d^2}{dt^2}X = F^{\text{fast}}(X) + \sum_n \Delta t\ \delta(t - n\Delta t)F^{\text{slow}}, \tag{11}$$

which is a system of ODEs with a periodic forcing function. If the frequency of the forcing function coincides with a natural frequency of the unforced system, then resonance produces an oscillation in the trajectory whose amplitude increases with time. This artifact of the impulse MTS method was first

demonstrated, for simple problems, in [3]. A more recent analytical study [7] shows that the trajectory error in the impulse method is proportional to Δt rather than Δt^2 if F^{fast} is fast enough to produce resonance.

More surprisingly, there is also a problem for time steps Δt just smaller than *half* the period of the fastest normal mode, which is approximately 5 fs. There is in [4] a revealing experiment for the Verlet-I/r-RESPA impulse MTS method. In a 1000-fs simulation on a 36000-atom water–protein–DNA system it was found that $\Delta t = 4$ fs produces very little energy drift, that $\Delta t = 5$ fs yields dramatic energy growth, and that $\Delta t = 6$ fs gives only half as much energy growth as $\Delta t = 5$ fs. In all cases the small time step δt was 1 fs. This experiment suggests a *stability* problem associated with timesteps Δt approximately equal to half the period of the fastest normal mode. As a check a power spectrum was computed for 20 ps of a 180 K simulation of a 20 Å sphere of the flexible TIP3P water used in that simulation. This is shown by Fig. 2. Peaks are discernible at frequencies corresponding to periods of 9.3 and 12.1 fs, evidently the periods of the two normal modes for the stretching of the O–H bonds.

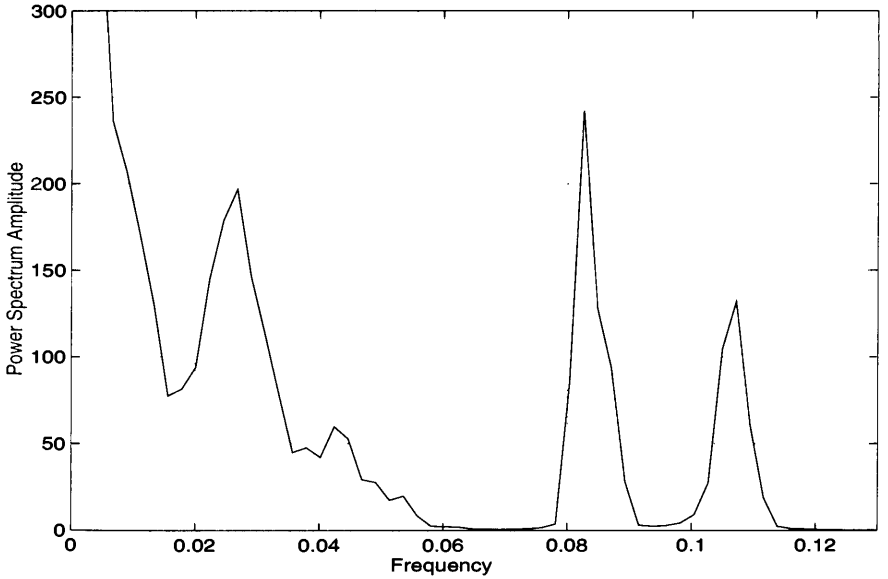


Fig. 2. Power spectrum of water dynamics with frequency in units of fs^{-1}.

An upper limit of 4 fs on the longest time step is also the experience of other researchers. For example, Fig. 2 in [10] shows that energy conservation is good for long time steps in the range 0.5 to 4 fs but dramatically worsens

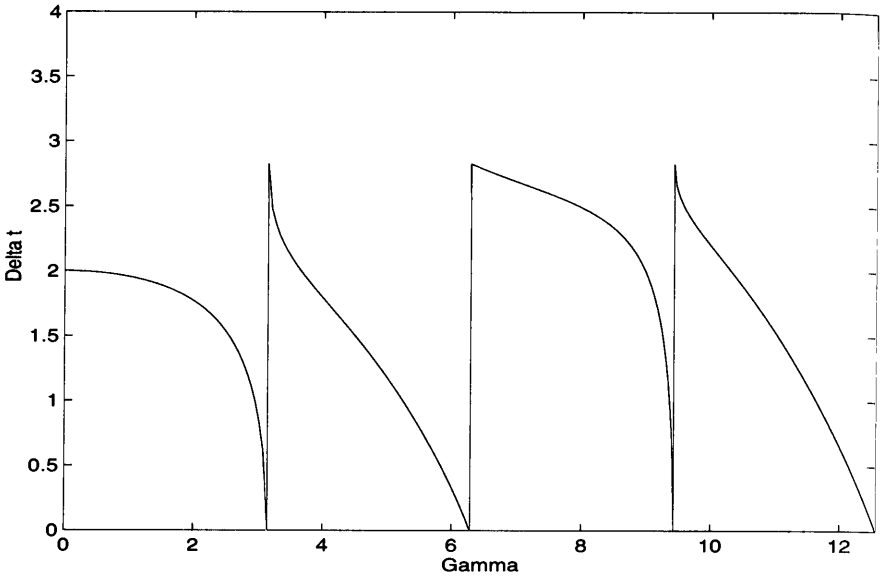


Fig. 3. Stability boundary for impulse method applied to the 2-spring problem. Gamma is $\omega_1 \Delta t$ and Delta t is Δt.

for a 5 fs time step. See also [12, 13] for evidence of a time step barrier at half the period of the fastest normal mode.

An instability of the impulse MTS method for Δt slightly less than half the period of a normal mode is confirmed by an analytical study of a linear model problem [7]. For another analysis, see [2]. A special case of this model problem, which gives a more transparent description of the phenomenon, is as follows: Consider a two-degree-of-freedom system with Hamiltonian $\frac{1}{2}p_1^2 + \frac{1}{2}p_2^2 + \frac{1}{2}\omega_1^2 x_1^2 + \frac{1}{4}(x_2 - x_1)^2$ This models a system of two springs connecting two balls of unit mass where one spring of stiffness ω_1^2 connects an immovable object to the first ball and a second spring of stiffness $\frac{1}{2}$ connects the balls together. The x_1, x_2 are displacements from equilibrium. The instability for the impulse MTS method is studied by choosing a timestep Δt for the second spring (the term $\frac{1}{4}(x_2 - x_1)^2$ in the Hamiltonian) and an infinitesimal timestep for the first spring (which in effect is method (5)). Analysis shows that the dynamics depends on only the two parameters $\gamma = \omega_1 \Delta t$ and Δt. Instability occurs for certain combinations of γ and Δt. This is shown by Fig. 3, where the region above the curve constitutes the region of instability. In this diagram, a change in Δt with ω_1 held fixed corresponds to a move in the radial direction. Details concerning the construction of this diagram are given in Appendix A.

It is to be expected that this barrier on the long time step would be shifted upward if the highest frequencies were eliminated. These frequencies can be removed with only modest loss of accuracy by constraining selected bond lengths and bond angles. The resulting constrained system of ODEs can be numerically integrated by the popular SHAKE [20] extension of the leapfrog/Verlet method. Although SHAKE is not symplectic, there is a cosmetic modification to it known as RATTLE [1] which is symplectic [16].

The method presented in the next section is an attempt to overcome the barrier due to the highest frequencies whatever their origin. Although it has been implemented and tested for unconstrained dynamics only, there is no fundamental reason why it cannot be applied to overcome the less restrictive time step barrier arising in constrained dynamics.

5 The Mollified Impulse Method

In an effort to counteract the accuracy reduction of the impulse method in a resonance situation, a modification to the impulse method is proposed in [7]. There, the term $U^{\text{slow}}(x)$ is replaced by $U^{\text{slow}}(\mathcal{A}(x))$, where $\mathcal{A}(x)$ represents a time averaging of positions x due to motion producible by the fast forces F^{fast} in the neighborhood of x. The purpose of this change to the potential [2] is to change the slow forces from $F^{\text{slow}}(x)$ to $\mathcal{A}_x(x)^{\mathrm{T}} F^{\text{slow}}(\mathcal{A}(x))$. Modifying the potential instead of the force ensures that the force is conservative and that the integrator is symplectic. Therefore, this generalization of the impulse method preserves symplecticness and at the same time improves the impulse method's accuracy and stability, as explained below.

The original motivation [7] for using $\mathcal{A}_x(x)^{\mathrm{T}} F^{\text{slow}}(\mathcal{A}(x))$ was to compensate for the inaccuracies arising from evaluating $F^{\text{slow}}(x)$ at point values of $x = X(t)$ sampled at large time increments Δt. Changing the point of evaluation of F^{slow} from x to $\mathcal{A}(x)$ brings about some *accuracy* gains, because $F^{\text{slow}}(\mathcal{A}(x))$ is a better description of the quickly varying $F^{\text{slow}}(X(t))$ than values of F^{slow} at step points. A more important benefit, however, is gained by multiplying the force $F^{\text{slow}}(\mathcal{A}(x))$ by the matrix $\mathcal{A}_x(x)^{\mathrm{T}}$. The latter also tremendously benefits the *stability* of the integrator by filtering out those components of the slow force that would excite components of the fast force susceptible to resonance. The F^{slow} impulse is *mollified.*

An example of a time averaging function $\mathcal{A}(x)$ is the formula termed *LongAverage* in [7]:

$$\mathcal{A}(x) = \frac{1}{\Delta t} \int_0^{\Delta t} \tilde{X}(t) dt \tag{12}$$

[2] This is not inconsistent with the problem being solved, because in the limit as $\Delta t \to 0$, one would have $\mathcal{A}(x) = x$.

where $\tilde{X}(t)$ solves an *auxiliary* problem

$$M\frac{d^2}{dt^2}\tilde{X} = F^{\text{fast}}(\tilde{X}), \quad \tilde{X}(0) = x, \quad \frac{d}{dt}\tilde{X}(0) = 0. \tag{13}$$

The derivation of the mollified impulse method in [7] suggests that the same integrator be used for the auxiliary problem as that used for integrating the *reduced primary* problem $M(d^2/dt^2)X = F^{\text{fast}}(X)$ between impulses. Of course, $\mathcal{A}_x(x)^{\text{T}}$ is also needed. For the partitionings $U^{\text{fast}} + U^{\text{slow}}$ typically used in MD, this would lead unfortunately to a matrix $\mathcal{A}_x(x)^{\text{T}}$ with a great many nonzeros. However, it is probably important to take into account only the fastest components of U^{fast} [7]. Hence, it would seem sufficient to use only the fastest forces $F^{\text{fastest}}(x)$ in the time averaging calculation.

Suppose values X^{n-1}, P^{n-1}, and $F^{\text{slow},n-1} = F^{\text{slow}}(X^{n-1})$ are given. One step of the efficient version of MOLLY just described can be expressed as follows:

half a mollified kick

$$P^{n-1+\epsilon} = P^{n-1} + \frac{\Delta t}{2}F^{\text{slow},n-1}. \tag{14}$$

a vibration Propagate X^{n-1}, $P^{n-1+\epsilon}$ by integrating

$$\frac{d}{dt}X = M^{-1}P, \quad \frac{d}{dt}P = F^{\text{fast}}(X) \tag{15}$$

for an interval Δt to get X^n, $P^{n-\epsilon}$.

a time averaging Calculate a temporary vector of time-averaged positions $\bar{X}^n = \mathcal{A}(X^n)$ and a Jacobian matrix $J^n = \mathcal{A}_x(X^n)^{\text{T}}$. The time averaging function $\mathcal{A}(x)$ uses only $F^{\text{fastest}}(x)$.

half a mollified kick

$$F^{\text{slow},n} = J^n F^{\text{slow}}(\bar{X}^n), \tag{16}$$

$$P^n = P^{n-\epsilon} + \frac{\Delta t}{2}F^{\text{slow},n}. \tag{17}$$

The symbols $P^{n-1+\epsilon}$ and $P^{n-\epsilon}$ represent momenta just after the $(n-1)$st kick, and just before the nth kick, respectively. Note that $\bar{X}^n$ is used only for the purpose of evaluating F^{slow}; it does not replace the value of X^n.

The calculation of $\mathcal{A}(x)$ and $\mathcal{A}_x(x)$ can be done in a systematic manner. First the calculation of $\mathcal{A}(x)$ is coded, and then this is differentiated with respect to each of the components of x to yield code for $\mathcal{A}_x(x)$. An example of this procedure for the leapfrog method is given in Appendix B.

Different time averagings are, of course, possible. Various time averagings can be defined by

$$\mathcal{A}(x) = \frac{1}{\Delta t}\int_0^\infty \phi\left(\frac{t}{\Delta t}\right)\tilde{X}(t)dt \tag{18}$$

where $\phi(s)$ is a weight function. Some interesting choices are

LongAverage

$$\phi(s) = \begin{cases} 1, & |s| < 1, \\ \frac{1}{2}, & |s| = 1, \\ 0, & |s| > 1. \end{cases} \tag{19}$$

LongLinearAverage

$$\phi(s) = \begin{cases} 1 - \frac{1}{2}s, & 0 \le s \le 2, \\ 0, & s \ge 2. \end{cases} \tag{20}$$

LongQuadraticAverage

$$\phi(s) = \begin{cases} \frac{1}{4}(3 - s^2), & 0 \le s \le 1, \\ \frac{1}{8}(3 - s)^2, & 1 \le s \le 3, \\ 0, & s \ge 3. \end{cases} \tag{21}$$

If the fastest forces F^{fastest} consist only of bond stretching and angle bending, then it is possible to define a time averaging $\mathcal{A}$ as a projection onto the equilibrium value of the bond lengths and bond angles. This technique appears to have better stabilizing properties. Further details are to be provided elsewhere [11]. Here this time averaging is called *Equilibrium.*

6 Numerical Tests for MOLLY

The mollified impulse method was implemented in NAMD and tested on a 20 Å diameter sphere of flexible TIP3P water at 370 K. The switching function in [10] was used with a transition distance of 4 Å to 6.5 Å to separate the slow long-range electrostatic interactions from the remaining forces. The cutoff was chosen to be short to ensure that slow forces are of significant magnitude. Experiments such at those in [4] suggest that flexible water models are particularly sensitive to destabilizing artifacts in numerical integrators.

Shown in Fig. 4 are plots of total pseudoenergy vs. simulation time for two time averaging methods mentioned in this article, plus the *Equilibrium* and impulse method. The pseudoenergy is the true total energy with $U^{\text{slow}}(x)$ replaced by $U^{\text{slow}}(\mathcal{A}(x))$. This is the quantity which the integrator is trying to conserve, so it is a more sensitive indicator than the true energy of instability. In all cases the long time step Δt is 8 fs and the short time step δt is 1 fs. Only the *Equilibrium* version of MOLLY does not lead to energy increase in this experiment. The *LongQuadraticAverage* version gives a slight energy increase while the *LongAverage* gives a large energy increase. Ranking of the two time

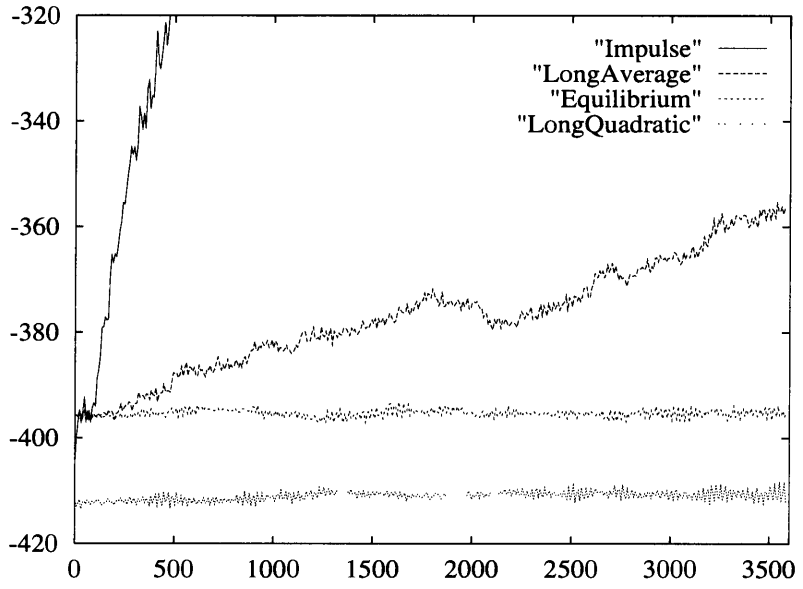


Fig. 4. Total pseudoenergy (in kcal/mol) vs. simulation time (in fs) for time averaging, *Equilibrium*, and impulse methods. (Δt for all methods equals 8 fs.)

averagings is in direct relation to the extensiveness of their time averaging and is consistent with the stability analysis of [7]. Notice the dramatic rise in energy for the impulse method.

The *Equilibrium* version was tested also on a 40 Å diameter sphere of water with a switching function transition distance of 8 Å to 13 Å. There was in this case a slight rise in energy when Δt was 8 fs.

Fig. 5 shows that all three time averaging methods succeed for a long timestep Δt of 5 fs.

Acknowledgment

We are grateful to Dorina Kosztin for reviewing a draft of the manuscript.

References

1. H. Andersen. Rattle: a 'velocity' version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys.*, 52:24–34, 1983.
2. E. Barth and T. Schlick. Extrapolation versus impulse in multiple-timestepping schemes: Linear analysis and applications to Newtonian and Langevin dynamics. *J. Chem. Phys.*, 1997. In press.

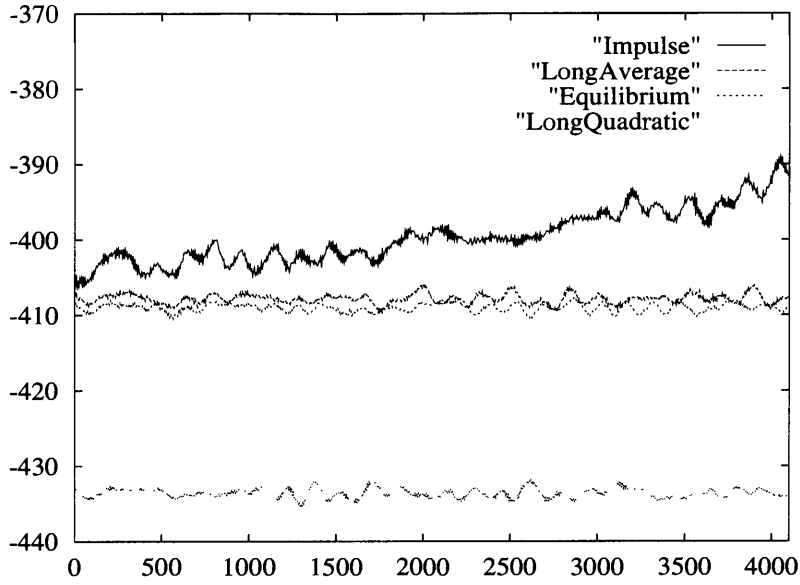


Fig. 5. Total pseudoenergy (in kcal/mol) vs. simulation time (in fs) for time averaging, *Equilibrium*, and impulse methods. (Δt for all methods equals 5 fs.)

3. J. J. Biesiadecki and R. D. Skeel. Dangers of multiple-time-step methods. *J. Comput. Phys.*, 109:318–328, 1993.
4. T. Bishop, R. D. Skeel, and K. Schulten. Difficulties with multiple timestepping and the fast multipole algorithm in molecular dynamics. *J. Comput. Chem.*, 18:1785–1791, 1997.
5. J. Delambre. *Mem. Acad. Turin*, 5:143, 1790–1793.
6. T. Forester and W. Smith. On multiple time-step algorithms and the Ewald sum. *Mol. Sim.*, 13:195–204, 1994.
7. B. García-Archilla, J. M. Sanz-Serna, and R. D. Skeel. Long-time-step methods for oscillatory differential equations. *SIAM J. Sci. Comput.* To appear. [Also Tech. Rept. 1996/7, Dep. Math. Aplic. Comput., Univ. Valladolid, Valladolid, Spain].
8. H. Grubmüller. Dynamiksimulation sehr großer Makromoleküle auf einem Parallelrechner. Master's thesis, Physik-Dept. der Tech. Univ. München, Munich, 1989.
9. H. Grubmüller, H. Heller, A. Windemuth, and K. Schulten. Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Molecular Simulation*, 6:121–142, 1991.
10. D. D. Humphreys, R. A. Friesner, and B. J. Berne. A multiple-time-step molecular dynamics algorithm for macromolecules. *J. Phys. Chem.*, 98:6885–6892, 1994.
11. J. Izaguirre, S. Reich, and R. D. Skeel. Longer time steps for molecular dynamics. Submitted.

12. D. Janežič and F. Merzel. An efficient symplectic integration algorithm for molecular dynamics simulations. *J. Chem. Inf. Comput. Sci.*, 35:321–326, 1995.
13. D. Janežič and F. Merzel. An efficient split integration symplectic method for molecular dynamics simulations of complex systems. In A. Sydow, editor, *15th IMACS World Congress on Scientific Computation, Modelling and Applied Mathematics*, volume 1, pages 493–498, Berlin, 1997. Wissenschaft & Technik Verlag.
14. D. Janežič and F. Merzel. Split integration symplectic method for molecular dynamics integration. *J. Chem. Inf. Comput. Sci.*, 37:1048–1054, 1997.
15. J. A. Board, Jr., Z. S. Hakura, W. D. Elliot, D. C. Gray, W. J. Blanke, and J. F. Leathrum, Jr. Scalable implementations of multipole-accelerated algorithms for molecular dynamics. In *Proceedings of the Scalable High-Performance Computing Conference*, pages 87–94, Los Alamitos, Calif., 1994. IEEE Computer Society Press.
16. B. Leimkuhler and R. D. Skeel. Symplectic numerical integrators in constrained Hamiltonian systems. *J. Comput. Phys.*, 112:117–125, 1994.
17. T. R. Littell, R. D. Skeel, and M. Zhang. Error analysis of symplectic multiple time stepping. *SIAM J. Numer. Anal.*, 34:1792–1807, 1997.
18. B. Mehlig, D. W. Heermann, and B. M. Forrest. Hybrid Monte Carlo method for condensed-matter systems. *Phys. Rev. B*, 45:679–685, 1992.
19. M. Nelson, W. Humphrey, A. Gursoy, A. Dalke, L. Kalé, R. D. Skeel, and K. Schulten. NAMD—a parallel, object-oriented molecular dynamics program. *Intl. J. Supercomput. Applics. High Performance Computing*, 10:251–268, 1996.
20. J. Ryckaert, G. Ciccotti, and H. Berendsen. Numerical integration of the cartesian equation of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.*, 23:327–341, 1977.
21. J. Sanz-Serna and M. Calvo. *Numerical Hamiltonian Problems.* Chapman and Hall, London, 1994.
22. S. Toxvaerd. Comment on: Reversible multiple time scale molecular dynamics. *J. Chem. Phys*, 99:2277, 1993.
23. M. Tuckerman, B. J. Berne, and G. J. Martyna. Reversible multiple time scale molecular dynamics. *J. Chem. Phys*, 97:1990–2001, 1992.
24. J. Wisdom. The origin of the Kirkwood gaps: A mapping for asteroidal motion near the 3/1 commensurability. *Astr. J.*, 87:577–593, 1982.

Appendix A Stability Conditions for the Impulse Method

Here are given details indicating how Fig. 3 was obtained from the analytical study in [7]. The problem considered there is a system with Hamiltonian $\frac{1}{2}p_1^2 + \frac{1}{2}p_2^2 + \frac{1}{2}\omega_1^2 x_1^2 + \frac{1}{4}(x_2 - x_1)^2 + \frac{1}{2}\omega_2^2 x_2^2$, which models a system of *three* springs connecting two balls of unit mass where the third spring is of stiffness ω_2^2 and connects the second ball to an immovable object. Instability is studied by choosing a timestep Δt for the second spring and an infinitesimal timestep for the other two springs. Presented here is the special case $\omega_2 = 0$, which, of course, simplifies the stability conditions given in [7]. These simplified stability conditions depend on the two parameters $\gamma = \omega_1 \Delta t$ and Δt. Instability occurs when

$$\left(\frac{\Delta t}{2}\right)^2 > 2(1+\cos\gamma)(1+\cos\gamma+2\gamma^{-1}\sin\gamma)^{-1}$$
$$\text{for } 0 \le \gamma < \pi \text{ or } 2\pi \le \gamma < 3\pi \text{ or } \ldots$$

and

$$\left(\frac{\Delta t}{2}\right)^2 > (1-\cos\gamma)(1+|\gamma^{-1}\sin\gamma|^{1/2})^{-2}$$
$$\text{for } \pi \le \gamma < 2\pi \text{ or } 3\pi \le \gamma < 4\pi \text{ or } \ldots,$$

which is shown in Fig. 3 with the region above the curve constituting the region of instability.

Appendix B Calculation of the Jacobian Matrix of an Averaging Function

As an example suppose that the leapfrog method with time step δt is coded for the calculation of $\mathcal{A}(x)$. This is then differentiated to obtain $\mathcal{A}_x(x)$. The result is the following code for calculating $\mathcal{A}(x)$ and $\mathcal{A}_x(x)$: Initialization is given by

$$\begin{aligned} X &:= x, & X_x &:= I, \\ P &:= 0, & P_x &= 0, \\ B &:= 0, & B_x &= 0, \end{aligned} \tag{22}$$

and step by step integration by

$$\begin{aligned} P &:= P + \tfrac{1}{2}\delta t F^{\text{fastest}}(X), & P_x &:= P_x + \tfrac{1}{2}\delta t F_x^{\text{fastest}}(X)X_x, \\ B &:= B + \tfrac{1}{2}\delta t X, & B_x &:= B_x + \tfrac{1}{2}\delta t X_x, \\ X &:= X + \delta t M^{-1} P, & X_x &:= X_x + \delta t M^{-1} P_x, \\ B &:= B + \tfrac{1}{2}\delta t X, & B_x &:= B_x + \tfrac{1}{2}\delta t X_x, \\ P &:= P + \tfrac{1}{2}\delta t F^{\text{fastest}}(X), & P_x &:= P_x + \tfrac{1}{2}\delta t F_x^{\text{fastest}}(X)X_x. \end{aligned} \tag{23}$$

The value $(1/\Delta t)B$ is used for $\mathcal{A}(x)$ and $(1/\Delta t)B_x$ for $\mathcal{A}_x(x)$.

Long Time Step MD Simulations Using Split Integration Symplectic Method*

Dušanka Janežič and Franci Merzel

National Institute of Chemistry, Hajdrihova 19, 1000 Ljubljana, Slovenia
e-mail: dusa@kihp5.cmm.ki.si

Abstract. The design and analysis of an explicit Split Integration Symplectic Method (SISM) for molecular dynamics (MD) simulations is described. SISM uses an analytical treatment of high frequency motions within a second order generalized leapfrog scheme. SISM is up to an order of magnitude faster than the commonly used leapfrog-Verlet (LFV) algorithm which is of the same order and complexity as SISM. The main restriction on time step in the general MD simulations, which stems from the high-frequency motion is, to a large extent, overcome in this approach. The simulation results for selected examples show that SISM posses long term stability and the ability to use long time steps. This should significantly extend the scope of the presently used algorithms and thus contribute to the general applicability of MD algorithms.

1 Introduction

Computer simulations in general, and MD in particular, represent a new scientific methodology. Theoretical breakthroughs involve both new concepts and the mathematical tools of development.

Among the main theoretical methods of investigation of the dynamic properties of macromolecules are molecular dynamics (MD) simulations and harmonic analysis. MD simulation is a technique in which the classical equation of motion for all atoms of a molecule is integrated over a finite period of time. Harmonic analysis is a direct way of analyzing vibrational motions. Harmonicity of the potential function is a basic assumption in the normal mode approximation used in harmonic analysis. This is known to be inadequate in the case of biological macromolecules, such as proteins, because anharmonic effects, which MD has shown to be important in protein motion, are neglected [1, 2, 3].

The problem which arises in performing MD simulations of complex systems is that only short simulations can be performed due to the high frequency bond stretching and angle bending motion which requires the use of small time step in MD simulations. Therefore the efficiency and the scope of the MD method will be increased by algorithms allowing for a long time step without loss of stability [4]. Enlargement of the scope of the existing methods

* This work was supported by the Ministry of Science and Technology of Slovenia under grant No. J1-7346-104-97.

of molecular dynamics (MD) of complex molecular systems may be achieved by the introduction of the symplectic integration schemes for Hamiltonian systems [5, 5, 15, 16] and their parallel implementations [9, 10, 11].

We recently introduced an efficient symplectic algorithm for MD simulations of a class of isolated linear molecules for which it can be assumed that bond stretching satisfactorily describes all vibrational motions [6]. The method allowed the use of a time step at least a factor of ten larger than the standard leapfrog-Verlet method [7]. We also proposed the extension of the method devised in ref. [6] to a system of linear molecules by introducing the Split Integration Symplectic Method (SISM) for MD simulations [8]. This approach uses the idea of composition methods, which are particularly useful for numerical integration of differential equations when the equations have some special structure that is advantageous to preserve [9]. Ideally one would like to know the fastest method for a given problem with a given accuracy.

The derivation of SISM was inspired by the work of Wisdom and Holman [16], who used symplectic mapping method based on a separation of the Hamiltonian, for the N body problem, to compute the time evolution of the planet Pluto over a billion-year period. They split the total Hamiltonian into a Kepler part representing the low-frequency terms of the Hamiltonian and into an interaction part representing the high-frequency terms. Then they applied a second order symplectic integrator, a generalized leap-frog integration scheme. Introducing the averaging principle [17] to the interaction part enabled them to use a step size an order of magnitude larger than can be used by traditional numerical methods. The SISM is, owing to the analytical treatment of the high frequency motions, quite distinct from other approaches using fractional step methods. These approaches which separate motion into different time scales and use multiple time-steps [18, 19, 10] proved to be stable and to accelerate the integration of the equations of motion. The idea for treating the reference system (harmonic oscillator) analytically and the rest of the propagator numerically was also introduced in ref. [21].

In this article we describe an extension of SISM to a system of molecules for which it can be assumed that both bond stretching and angle bending describe satisfactorily all vibrational motions of the molecule. The SISM presented here allows the use of an integration time step up to an order of magnitude larger than possible with other methods of the same order and complexity.

2 Molecular Dynamics

For a given potential energy function, one may take a variety of approaches to study the dynamics of macromolecules. The most exact and detailed information is provided by MD simulations in which one solves the equations of motion for the atoms constituting the macromolecule and any surrounding environment. With currently available techniques and methods it is possible

to simulate the dynamics of macromolecules for up to a few nanoseconds. To study slower and more complex processes in macromolecules, it is generally necessary to use methods other than straightforward MD simulations or, alternatively, possible improvements on efficiency of MD simulation techniques should be considered.

A good MD algorithm permits a large time step to be used while preserving acceptable energy conservation. Major methodological progress has been made in the directions of longer time scales and larger systems size [11, 12, 13].

3 Harmonic Analysis

Harmonic analysis is an alternative approach to MD. The basic assumption is that the potential energy can be approximated by a sum of quadratic terms in displacements.

To determine the vibrational motions of the system, the eigenvalues and eigenvectors of a mass-weighted matrix of the second derivatives of potential function has to be calculated. Using the standard normal mode procedure, the secular equation

$$|\mathbf{F} - \lambda \mathbf{M}| = \mathbf{0} \tag{1}$$

is solved, where $\mathbf{F} = \nabla^2 E$ is the Hessian and $\mathbf{M} = \mathrm{diag}\,(m_i)$ is a diagonal mass matrix. In the root-masss-weighted cartesian displacement coordinate system, which provides a symmetric secular equation, this becomes

$$|\mathbf{M}^{-1/2}\mathbf{F}\mathbf{M}^{-1/2} - \lambda| = \mathbf{0} \tag{2}$$

For a nonlinear molecule composed of N atoms, $3N-6$ eigenvalues provide the normal or fundamental vibrational frequencies of the vibration and and the associated eigenvectors, called normal modes give the directions and relative amplitudes of the atomic displacements in each mode.

Harmonic analysis (normal modes) at given temperature and curvature gives complete time behavior of the system in the harmonic limit [1, 2, 3]. Although the harmonic model may be incomplete because of the contribution of anharmonic terms to the potential energy, it is nevertheless of considerable importance because it serves as a first approximation for which the theory is highly developed. This model is also useful in SISM which uses harmonic analysis.

4 Hamilton Equations of Motion

Symplecticness is a characterization of Hamiltonian systems in terms of their solution. The solution operator $\Phi_H\,(t, t_0)$ defined by

$$(\boldsymbol{p}, \boldsymbol{q}) = \Phi_H\,(t, t_0)\,(\boldsymbol{p}^0, \boldsymbol{q}^0) \tag{3}$$

is the value at time t of the solution of Hamiltonian system that at $t = t_0$ has the initial condition $(\boldsymbol{p}^0, \mathbf{q}^0)$ where t, t_0 are fixed parameters and $(\mathbf{p}^0, \mathbf{q}^0)$ is a variable [5].

For autonomous Hamiltonians $\frac{dH}{dt} = 0$ the solution in phase space depends only on $t - t_0$ where the t-flow $\phi_{t,H}$ is defined as

$$(\boldsymbol{p}, \boldsymbol{q}) = \phi_{t,H}(\boldsymbol{p}_0, \boldsymbol{q}_0) \tag{4}$$

where $(\boldsymbol{p}, \boldsymbol{q})$ is the solution and takes at t = 0 the value $(\boldsymbol{p}_0, \boldsymbol{q}_0)$.

A mapping is said to be symplectic or canonical if it preserves the differential form $d\boldsymbol{p} \wedge d\boldsymbol{q}$ which defines the symplectic structure in the phase space. Differential forms provide a geometric interpretation of symplecticness in terms of conservation of areas which follows from Liouville's theorem [14]. In one-degree-of-freedom example symplecticness is the preservation of oriented area. An example is the harmonic oscillator where the t-flow is just a rigid rotation and the area is preserved. The area-preserving character of the solution operator holds only for Hamiltonian systems. In more then one-degree-of-freedom examples the preservation of area is symplecticness rather than preservation of volume [5].

Symplectic integration methods replace the t-flow $\phi_{t,H}$ by the symplectic transformation $\psi_{t,H}$, which retains "Hamiltonian" features of $\phi_{t,H}$. They poses a backward error interpretation property which means that the computed solutions are solving exactly or, at worst, approximately a nearby Hamiltonian problem which means that the points computed by means of symplectic integration, lay either exactly or at worst, approximately on the true trajectories [5].

To perform MD simulation of a system with a finite number of degrees of freedom the Hamilton equations of motion

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}, \quad \frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}, \quad i = 1 \ldots d \tag{5}$$

where H is the Hamiltonian, q_i and p_i are the coordinate and momentum, respectively, and d is the number of degrees of freedom are to be solved.

5 Leapfrog-Verlet Algorithm

The simplest of the numerical techniques for the integration of equations of motion is leapfrog-Verlet algorithm (LFV), which is known to be symplectic and of second order. The name leapfrog steams from the fact that coordinates and velocities are calculated at different times.

The typical MD Hamiltonian H of the system is the sum of kinetic and potential energy

$$H = T + U \tag{6}$$

$$T = \sum_i \frac{\mathbf{p}_i^2}{2m_i} \tag{7}$$

$$\begin{aligned} U = &\textstyle\sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\varphi(\varphi - \varphi_0)^2 + \\ &+ \textstyle\sum_{dihed.} k_\vartheta\,(1 + cos(n\vartheta - \delta)) + \\ &+ \textstyle\sum_{i>j} \frac{e_i e_j}{r_{ij}} + \sum_{i>j} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \end{aligned} \tag{8}$$

where m_i is the mass of the i-atom, b_0, φ_0 and δ are reference values for bond lengths, angles and dihedral angles, respectively, and k_b, k_φ and k_ϑ are corresponding force constants; i and j run over all atoms, e_i denotes the charge on the i-atom and r_{ij} is the distance between atoms i and j, ε_{ij} and σ_{ij} are the corresponding constants of Lennard-Jones potential.

The LFV integration method propagates coordinates and momenta on the basis of the equation of motion (5) by the following relations

$$q_i' = q_i + \frac{p_i}{m} \frac{\Delta\tau}{2} \tag{9}$$

$$p_{i+1} = p_i - \Delta\tau \left(\frac{\partial U}{\partial q} \right)_{q=q_i'} \tag{10}$$

$$q_{i+1} = q_i' + \frac{p_{i+1}}{m} \frac{\Delta\tau}{2} \tag{11}$$

where q_i is the coordinate, p_i is the momentum, $\dim(p_i,\, q_i) = d$, d is the number of degrees of freedom, $\Delta\tau$ is the time step, and m is the mass of the corresponding atom.

6 Split Integration Symplectic Method

The explicit symplectic integrator can be derived in terms of free Lie algebra in which Hamilton equations (5) are written in the form

$$\frac{d\mathbf{x}}{dt} = \{\mathbf{x}, H\} = \widehat{L}_H \mathbf{x} \tag{12}$$

where $\{\mathbf{x}, H\}$ denotes the Poisson bracket, $\widehat{L}_H$ is the Poisson bracket operator, and $\mathbf{x} = (\mathbf{q}, \mathbf{p})$ is a vector in the phase space composed of the coordinates and momenta of all particles. The formula

$$\mathbf{x}|_{\tau_0 + \Delta\tau} = \exp(\Delta\tau \widehat{L}_H)\mathbf{x}|_{\tau_0} \tag{13}$$

provides a way for integrating the Hamiltonian system in terms of Lie operators [17]. It is the formal solution of Hamilton equations or the exact time evolution of trajectories in phase space from τ_0 to $\tau_0 + \Delta\tau$, and $\Delta\tau$ is a time step. The trouble with it lays in the impossibility of evaluating $\exp(\Delta\tau \widehat{L}_H)$.

The Lie formalism used is the key in the development of symplectic integration. Symplectic integration consists in replacing $\exp(\Delta\tau\widehat{L}_H)$ by a product of symplectic maps which approximate $\exp(\Delta\tau\widehat{L}_H)$ to a given order [18].

The construction of an efficient algorithm rests on the ability to separate the Hamiltonian into parts which are themselves integrable and also efficiently computable. Suppose that the MD Hamiltonian H defined by (6) is split into two parts as

$$H = H_0 + H_r \tag{14}$$

where

$$H_0 = H^{anal} \tag{15}$$

$$H_r = H - H^{anal} \tag{16}$$

and H^{anal} denotes the part of the Hamiltonian which can be treated analytically. Then the following approximation for (13) can be used

$$\mathbf{x}|_{\tau_0+\Delta\tau} \approx \exp(\frac{\Delta\tau}{2}\widehat{L}_{H_0})\exp(\Delta\tau\widehat{L}_{H_r})\exp(\frac{\Delta\tau}{2}\widehat{L}_{H_0})\mathbf{x}|_{\tau_0} \tag{17}$$

which prescribes how to propagate from one point in phase space to another. First, the system is propagated a half step evolution with H_0, then a whole step with H_r, and finally another half step with H_0. This scheme is called the generalized leapfrog [19] and was widely used as a basis for development of multiple time-step MD integration algorithms [18, 19]. This integration scheme was also used in the development of SISM, a second order symplectic integration algorithm for MD integration.

We choose the Hamiltonian H_0 as describing the internal vibrational motion of the molecules and also the translation and rotation of molecules. It represents the dynamically leading contribution whose potential depends only on constant parameters of the simulation. This separation of the potential function allows the calculation of normal modes only once, at the beginning of the calculation. This term includes, in general, all bonding, angle bending and torsional angle interactions within the harmonic approximation. The dynamics driven by the vibrational part of H_0 is resolved analytically, by means of internal normal coordinates that rotate in phase space with the corresponding vibrational frequencies which are obtained by solving the secular equation at the outset of the calculation [1, 2, 3]. SISM was derived step-by-step with regard to which terms, describing the internal motions of the system studied, were included into the model Hamiltonian.

6.1 SISM Treatment of Bond Stretching and Angle Bending Terms

For the model Hamiltonian used in this study it was assumed that the bond stretching and angle bending satisfactorily describe all vibrational motions

for a system of planar molecules

$$H_0 = h.p. \left(\sum_i \frac{\mathbf{p}_i^2}{2m_i} + \sum_{bonds} k_b \left(b - b_0\right)^2 + \sum_{angles} k_\theta \left(\theta - \theta_0\right)^2 \right) + H_{tran/rot} \tag{18}$$

$$H_r = \sum_{i>j} \frac{e_i e_j}{r_{ij}} + \sum_{i>j} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \text{corrections to } H_0 \tag{19}$$

where h.p. denotes the harmonic part; m_i is the mass of the i-atom, b_0 and θ_0 are reference values for bond lengths and angles respectively, and k_b and k_θ are corresponding force constants, and $H_{tran/rot}$ corresponds to the translational and rotational part of kinetic energy; i and j run over all atoms, e_i denotes the charge on the i-atom and r_{ij} is the distance between atoms i and j, ε_{ij} and σ_{ij} are the corresponding constants of Lennard-Jones potential. By corrections to H_0 we mean the anharmonic and coupling terms to H_0. Figure 1 displays the SISM solution procedure for MD integration.

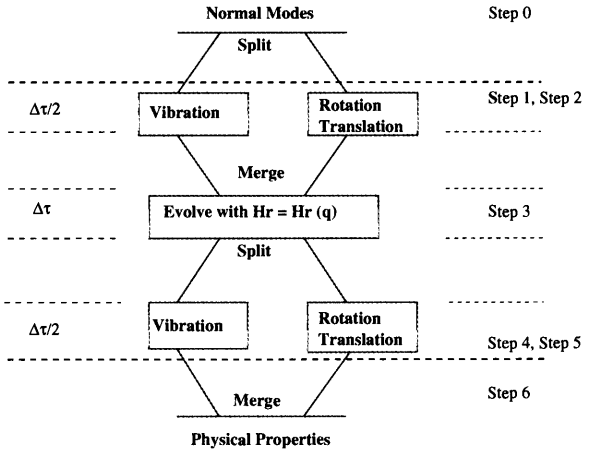


Fig. 1. The Split Integration Symplectic Method (SISM) solution procedure.

Following the procedure defined by (17) SISM can be written explicitly as follows

Step 0: Perform the normal modes of the harmonic part of H_0 to get the vibrational frequencies $\boldsymbol{\nu}$ and normal mode vectors which compose the transformational matrix $\mathbf{V}$.

$$\textbf{Split: } \mathbf{p},\ \mathbf{q} \Rightarrow \boldsymbol{\pi},\ \boldsymbol{\varrho},\ \mathbf{r}^*, \mathbf{v}^*, \boldsymbol{\omega}^*, \mathbf{e}^*$$

The procedure **Split** selects the internal displacement coordinates, $\boldsymbol{\varrho}$, and momenta, $\boldsymbol{\pi}$, (describing vibrations), the coordinates, $\mathbf{r}^*$, and velocities, $\mathbf{v}^*$, of the centers of molecular masses, angular velocities, $\boldsymbol{\omega}^*$, and directional unit vectors, $\mathbf{e}^*$, of the molecules from the initial Cartesian coordinates, $\mathbf{q}$, and from momenta, $\mathbf{p}$. Thus, the staring values for algorithm loop are prepared.

Step 1: Vibration

Propagate by harmonic part of H_0 for the time $\Delta\tau/2$. This corresponds to the rotation of internal normal coordinates, P_i' and Q_i', in the phase space by the corresponding vibrational frequency ν_i

$$P_i^0 = \sum_k \frac{1}{\sqrt{m_k}} \mathbf{V}_{ik}^T \pi_k$$

$$Q_i^0 = \sum_k \sqrt{m_k} \mathbf{V}_{ik}^T \varrho_k$$

$$\begin{bmatrix} P_i' \\ Q_i' \end{bmatrix} = \begin{bmatrix} \cos(\nu_i \frac{\Delta\tau}{2}) & -\nu_i \sin(\nu_i \frac{\Delta\tau}{2}) \\ (1/\nu_i)\sin(\nu_i \frac{\Delta\tau}{2}) & \cos(\nu_i \frac{\Delta\tau}{2}) \end{bmatrix} \begin{bmatrix} P_i^0 \\ Q_i^0 \end{bmatrix}$$

Step 2: Rotation/Translation

Translation of the center of mass

$$\mathbf{r}'^* = \mathbf{r}^* + \mathbf{v}^* \frac{\Delta\tau}{2}$$

$\mathbf{r}^*$ is the center of mass, $\Delta\tau$ is the time step.

Rotation of the internal coordinate system $\mathbf{e}^*$ for an angle $\omega^* \frac{\Delta\tau}{2}$ about the vector $\boldsymbol{\omega}^*/\omega^*$

$$\mathbf{e}'^* = \mathbf{R}\left(\frac{\Delta\tau}{2}\boldsymbol{\omega}^*\right)\mathbf{e}^*$$

It is assumed that $\boldsymbol{\omega}^*$ and $\mathbf{v}^*$ remain unchanged at $\frac{\Delta\tau}{2}$ time step since there is no external forces.

Step 3:

$$\textbf{Merge} : \boldsymbol{\pi}', \boldsymbol{\varrho}', \mathbf{r}'^*, \mathbf{v}^*, \boldsymbol{\omega}^*, \mathbf{e}'^* \Rightarrow \mathbf{p}, \mathbf{q}$$

The procedure **Merge** transforms the internal displacement coordinates and momenta, the coordinates and velocities of centers of masses, and directional unit vectors of the molecules back to the Cartesian coordinates and momenta.

Evolve with $H_r = H_r(\mathbf{q})$ means only a shift of all momenta for a corresponding impulse of force (SISM requires only one force evaluation per integration step).

$$p_i' = p_i - \Delta\tau \left(\frac{\partial H_r}{\partial q}\right)_{q=q_i}$$

$$q_i' = q_i$$

$$\textbf{Split: p', q'} \Rightarrow \boldsymbol{\pi}, \boldsymbol{\varrho}, \mathbf{r}^*, \mathbf{v}^*, \boldsymbol{\omega}^*, \mathbf{e}^*$$

Step 4: Rotation/Translation

Translation of the center of mass and rotation of the internal coordinate system as in **Step2**

$$\mathbf{r}'^* = \mathbf{r}^* + \mathbf{v}^* \frac{\Delta\tau}{2} \qquad \mathbf{e}'^* = \mathbf{R}\left(\frac{\Delta\tau}{2}\omega^*\right)\mathbf{e}^*$$

Step 5: Vibration

Again propagate by harmonic part of H_0 for the time $\Delta\tau/2$ in the phase space.

$$P_i^0 = \sum_k \frac{1}{\sqrt{m_k}} \mathbf{V}_{ik}^T \pi_k$$

$$Q_i^0 = \sum_k \sqrt{m_k} \mathbf{V}_{ik}^T \varrho_k$$

$$\begin{bmatrix} P_i' Q_i' \end{bmatrix} = \begin{bmatrix} \cos(\nu_i \frac{\Delta\tau}{2}) & -\nu_i \sin(\nu_i \frac{\Delta\tau}{2}) \\ (1/\nu_i)\sin(\nu_i \frac{\Delta\tau}{2}) & \cos(\nu_i \frac{\Delta\tau}{2}) \end{bmatrix} \begin{bmatrix} P_i^0 \\ Q_i^0 \end{bmatrix}$$

$$\pi_k' = \sqrt{m_i} \sum_i \mathbf{V}_{ik} P_i'$$

$$\varrho_k' = \frac{1}{\sqrt{m_i}} \sum_i \mathbf{V}_{ik} Q_i'$$

This concludes one whole integration step. At this point,

$$\textbf{Merge} : \boldsymbol{\pi}', \boldsymbol{\varrho}', \mathbf{r}'^*, \mathbf{v}^*, \boldsymbol{\omega}^*, e'^* \Rightarrow \mathbf{p}, \mathbf{q}$$

has to be performed (to get new momenta and coordinates **p, q**) whether the physical properties of the system, e.g., energy and displacements, are to be derived.

Step 6: Return to **Step 1** until the desired number of calculation steps is completed.

The analytical treatment of high frequency terms in the Hamiltonian proposed here allows to use SISM significantly longer integration time step than can be used by other methods of the same order and complexity.

Although the whole scheme of SISM seems to be computationally more demanding then LFV owing to several transformations used (e.g., from Cartesian to internal coordinates, from internal to normal coordinates etc.), it is not so, especially for the systems of small molecules. The computational cost of transformations in SISM scales as $N.n^2$ while the calculation of long-range forces and the energy of the system (common for both SISM and LFV) requires approximately $N^2.n^2$ operations, where N is the number of molecules in the system, and n is the number of atoms in the molecule. Therefore, all extra work in SISM used in coordinate transformations is prevailed by a long-range force and energy calculations [20].

6.2 SISM Treatment of only Bond Stretching Term

For the model Hamiltonian used in this study it was assumed that bond stretching satisfactorily describes all internal vibrational motions for a system of linear molecules and the split parts of the Hamiltonian were of the form

$$H_0 = i.p. \left(\sum_i \frac{\mathbf{p}_i^2}{2m_i} \right) + \sum_{bonds} k_b (b - b_0)^2 + H_{tran/rot} \tag{20}$$

$$H_r = \frac{\sum_{i>j} e_i e_j}{r_{ij}} + \sum_{i>j} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \tag{21}$$

and $i.p.$ refer to the internal part of kinetic energy which corresponds to molecular vibrations. The difference in the potential function compared to that used in the previous section is that here is no angle bending term in the Hamiltonian. Again, the analytical treatment of high frequency terms in Hamiltonian permits the SISM to use a much longer integration time step than the standard methods of the same order and complexity [8].

SISM for an Isolated Linear Molecule An efficient symplectic algorithm of second order for an isolated molecule was studied in details in ref. [6]. Assuming that bond stretching satisfactorily describes all vibrational motions for linear molecule, the partitioned parts of the Hamiltonian are

$$\begin{aligned} H_0 &= \textstyle\sum_i \frac{\mathbf{p}_i^2}{2m_i} + \sum_{bonds} k_b (b - b_0)^2 \\ H_r &= \textstyle\sum_{i>j} \frac{e_i e_j}{r_{ij}} + \sum_{i>j} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \end{aligned} \tag{22}$$

Again, the algorithm allows at least ten times larger time step to be used than LFV for the same accuracy.

The fourth order approximation for (13)

$$\begin{aligned} \boldsymbol{x}|_{t_0+\Delta t} \approx \exp(\tfrac{1}{2(1+\alpha)} \Delta t \widehat{L}_{H_0}) \exp(\tfrac{1}{1+\alpha} \Delta t \widehat{L}_{H_r}) \exp(\tfrac{\alpha}{2(1+\alpha)} \Delta t \widehat{L}_{H_0}) \\ \exp(\tfrac{\alpha-1}{\alpha+1} \Delta t \widehat{L}_{H_r}) \\ \exp(\tfrac{\alpha}{2(1+\alpha)} \Delta t \widehat{L}_{H_0}) \exp(\tfrac{1}{1+\alpha} \Delta t \widehat{L}_{H_r}) \exp(\tfrac{1}{2(1+\alpha)} \Delta t \widehat{L}_{H_0}) \boldsymbol{x}|_{t_0} \end{aligned} \tag{23}$$

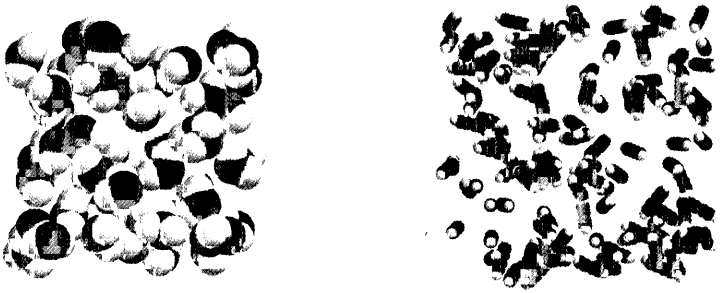
where $\alpha = 1 - \sqrt[3]{2}$ gives the fourth order SISM.

Following the procedure defined by (23) the fourth order SISM for MD simulations written explicitly can be found in ref. [22]. In the fourth order SISM additional steps in the algorithm occur due to additional force evaluations.

The fourth order SISM allows the same enlargement of the time step as the second order algorithm. However, it is computationally more demanding since three force evaluations per time step occur. In case of MD integration the greater order gives no advantage in calculation except greater accuracy which is known not to be very important in MD simulations.

7 Numerical Experiments

We shall illustrate the SISM described with two examples. The model system of a box of water molecules and the system of a box of linear molecules which are depicted in Figure 2.



(a) (b)

Fig. 2. (a) The model system of box of 50 H_2O (water) molecules for T = 300 K, and L = 15 Å(dark are O atoms, white are H atoms). (b) The model system of 128 H-(-C≡C-)$_2$-H (butadiyne) molecules for $\rho = 0.1$ g/cm^3, T = 300 K, and L = 47 Å(dark are C atoms, white are H atoms).

In order to compare the efficiency of the SISM with the standard LFV method, we compared computational performance for the same level of accuracy. To study the error accumulation and numerical stability we monitored the error in total energy, ΔE, defined as

$$\Delta E = \frac{1}{M}\sum_{i=1}^{M}\left|\frac{E_0 - E_i}{E_0}\right| \tag{24}$$

where E_i is the total energy at step i, E_0 is the initial energy, and M is the total number of time steps.

7.1 System of a Box of Water Molecules

The algorithm was applied to the MD simulations of a box of water molecules. The three-center water model was used [23]. The initial positions were at the equilibrium therefore all displacements were zero. The initial velocities were

defined by the equipartition theorem. The periodic boundary conditions were imposed. In this way it is possible to model the system that is effectively bounded but is nevertheless spatially homogeneous as far as boundaries are concerned.

The primary metric used to quantify the accuracy of the results of a simulation is the conservation of total energy. The long term stability is evident by the results shown in Figure 3, which depicts the time evolution of the total energy for the system of a box of water molecules for two different methods (LFV and SISM) for time step equal to 1 fs and trajectories of length of 0.5 ps. In comparison with LFV the total energy is even better conserved using SISM.

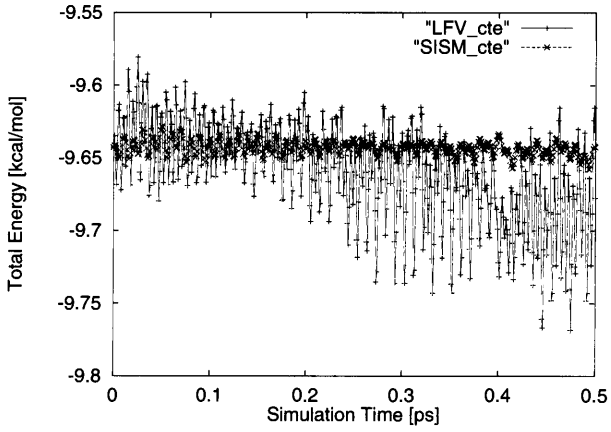


Fig. 3. Conservation of total energy for LFV and SISM for a system of 50 H_2O molecules, box L = 15 Å, and time step for both methods is 1fs

Figure 4 shows the error in total energy for a different time steps for a system of water molecules. The solid line represents LFV and the dashed line represents SISM. It can be observed that the integration time step using SISM can be enlarged many fold in comparison with LFV. The largest time step used is 5 fs since using longer time steps excise growth in the total energy occurs and the results no longer represent physical reality. This result agree well with those obtained by other authors when similar methods are used [24, 25].

In Table 1 the CPU time required by the two methods (LFV and SISM) for 1000 MD integration steps computed on an HP 735 workstation are compared for the same model system, a box of 50 water molecules, respectively. The computation cost per integration step is approximately the same for both methods so that the speed up of the SISM over the LFV algorithm is deter-

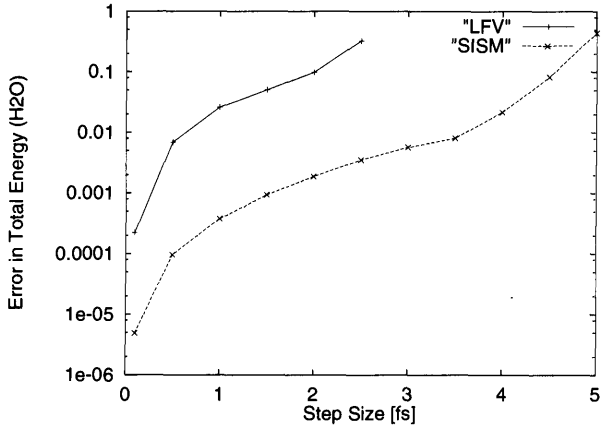


Fig. 4. Error in total energy for LFV and SISM for water.

mined mainly by the difference in step size which is significant. Qualitatively the same results were obtained in our previous works [6, 8].

Table 1. CPU Time for 1000 MD steps of 50 H_2O molecules in a box with L = 15 Åusing the LFV and the SISM for equal time step of 1 fs computed on an HP 735 workstation

MD Method	CPU Time (s)
LFV	50.80
SISM	53.80

7.2 System of Linear Molecules

Figure 2(b) displays the model system of 128 H-(-C≡C-)$_2$-H (butadiyne) molecules; ρ = 0.1 g/cm^3, L = 47Å, and T = 300 K. The initial conditions for coordinates and velocities of the system and the system parameters were the same as in the previous study [8].

The results of the error in total energy for test molecules, a system of linear butadiyne molecules, (ρ = 0.1 g/cm^3, corresponding to the box size L = 47 Å) using two different methods (LFV and SISM) are presented on Figure 5. It can be seen that for the same level of accuracy, the time step in SISM can be up to an order of magnitude larger than in LFV. Also, LFV is stable for only very short time steps, up to 4 fs, while SISM is stable for much longer time steps, even for time step larger than 25 fs. It should

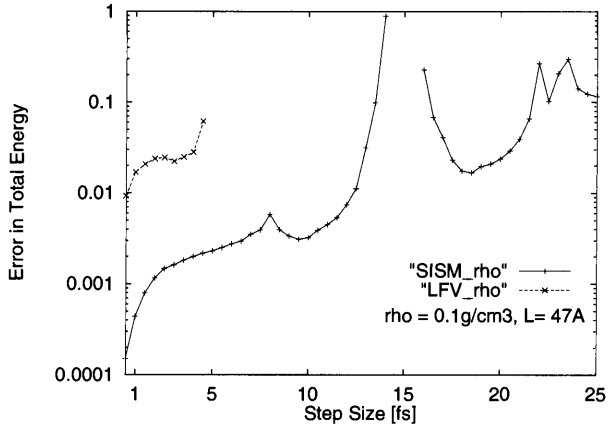


Fig. 5. Error in total energy for the LFV algorithm and the SISM of a system of 128 H-(-C≡C-)$_2$-H (butadiyne) molecules. Results are plotted for two different algorithms (LFV and SISM) and for density $\rho = 0.1$ g/cm^3, corresponding to the box size L = 47 Å.

be noted that such large time steps no longer represent physical reality and are identified with linear molecules possessing no angle bending or torsional interactions. There are large variations in total energy which might be due to the so called "step size resonances" [26, 27]. This phenomenon is due to the symplectic methods which seem to introduce artificial coupling among the motions associated with various frequencies leading to instability [28].

Isolated Linear Molecule Figure 6 shows the error in total energy for an isolated linear molecule H-(-C≡C-)$_5$-H. It is obvious that for the same level of accuracy, the time step in the SISM can be ten times or more larger as in the LFV. Furthermore, the LFV method is stable for only very short time steps, up to 5 fs, while the SISM is stable even for a time step up to 200 fs. However, such large time steps no longer represent physical reality and are a particular property identified with linear molecules without bending or torsional intramolecular interactions.

Note that there are also variations in total energy which might be due to the so called "step size resonance" [26, 27]. Shown are also results for fourth order algorithm which gives qualitatively the same results as the second order SISM. This show that the "step size resonances" are not due to the low order integration method but rather to the symplectic methods [28].

Notes and Comments. Further improvements in efficiency were achieved by implementing the method on computers with highly parallel architecture. SISM performs in parallel as LFV which means the speed up is gained due to longer time step which can be used by SISM [20].

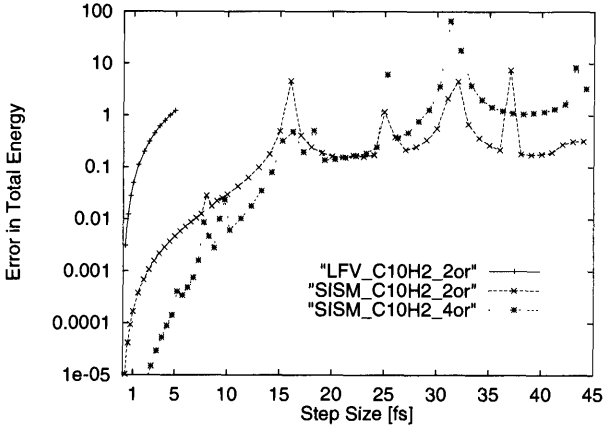


Fig. 6. Error in total energy for LFV, and the second and the fourth order SISM for H-(-C≡C-)$_5$-H. Results are plotted for two different algorithms (-|-, LFV, $-x-$, the second order SISM, and $-*-$, the fourth order SISM).

Much work remains to be done in the development of this approach to explore the advantages and limitations of the method. The method will be extended to force fields that include torsional terms; large systems such as biological macromolecules will also be treated.

References

1. Brooks, B. R., Janežič, D., Karplus, M.: Harmonic Analysis of Large Systems: I. Methodology. J. Comput. Chem. **16** (1995) 1522-1542
2. Janežič, D., Brooks, B. R.: Harmonic Analysis of Large Systems: II. Comparison of Different Protein Models. J. Comput. Chem. **16** (1995) 1543-1553
3. Janežič, D., Venable, R. M., Brooks, B. R.: Harmonic Analysis of Large Systems. III. Comparison with Molecular Dynamics. J. Comput. Chem. **16** (1995) 1554-1566
4. Allen, M. P., Tildesley, D. J.: Computer Simulation of Liquids. Clarendon Press: Oxford (1987)
5. Sanz-Serna, J. M. Symplectic Integrators for Hamiltonian Problems: An Overview. Acta Numerica (1991) 243-286
6. Sanz-Serna, J. M., Calvo, M. P.: Numerical Hamiltonian Problems. Chapman and Hall: London (1994)
7. Janežič, D., Orel, B.: Implicit Runge-Kutta Method for Molecular Dynamics Integration. J. Chem. Inf. Comput. Sci. **33** (1993) 252-257
8. Janežič, D., Orel, B.: Improvement of Methods for Molecular Dynamics Integration. Int. J. Quant. Chem. **51** (1994) 407-415
9. Trobec, R.; Jerebic, I.; D.Janežič, D. Parallel Algorithm for Molecular Dynamics Integration. Parallel Computing **19** (1993) 1029-1039

10. Janežič, D.; Trobec, R. Parallelization of an Implicit Runge-Kutta Method for Molecular Dynamics Integration. J. Chem. Inf. Comput. Sci. **34** (1994) 641-646
11. R. Trobec and D. Janežič; Comparison of Parallel Verlet and Implicit Runge-Kutta Method for Molecular Dynamics Integration. J. Chem. Inf. Comp. Sci., **35** (1995) 100-105
12. Janežič, D., Merzel, F.: An Efficient Symplectic Integration Algorithm for Molecular Dynamics Simulations. J. Chem. Inf. Comput. Sci. **35** (1995) 321-326
13. Verlet, L.: Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. Physical Review **159** (1967) 98-103
14. Janežič, D., Merzel, F.: Split Integration Symplectic Method for Molecular Dynamics Integration. J. Chem. Inf. Comput. Sci. **37** (1997) 1048-1054
15. McLachlan, R. I.: On the Numerical Integration of Ordinary Differential Equations by Symplectic Composition Methods. SIAM J. Sci. Comput. **16** (1995) 151-168
16. Wisdom, J.; Holman, M. Symplectic Maps for the N-body Problem. Astron. J. **102** (1991) 1528-1538
17. Wisdom, J. The Origin of the Kirkwood Gaps: A Mapping for Asteroidal Motion Near the 3/1 Commensurability. Astron. J. **87** (1982) 577-593
18. Tuckerman, M., Martyna, G. J., Berne, J.: Reversible Multiple Time Scale Molecular Dynamics. J. Chem. Phys. **97** (1992) 1990-2001
19. Tuckerman, M., Berne, J.: Vibrational Relaxation in Simple Fluids: Comparison of Theory and Simulation. J. Chem. Phys. **98** (1993) 7301-7318
20. Humphreys, D. D., Friesner, R. A., Berne, B. J.: A Multiple-Time Step Molecular Dynamics Algorithm for Macromolecules. J. Chem. Phys. **98** (1994) 6885-6892
21. Tuckerman, M., Martyna, G. J., Berne, J.: Molecular Dynamics Algorithm for Condensed Systems with Multiple Time Scales. J. Chem. Phys. **93** (1990) 1287-1291
22. Watanabe, M., Karplus, M.: Dynamics of Molecules with Internal Degrees of Freedom by Multiple Time-Step Methods. J. Chem. Phys. **99** (1995) 8063-8074
23. Figueirido, F., Levy, R. M., Zhou, R., Berne, B. J.: Large Scale Simulation of Macromolecules in Solution: Combining the Periodic Fast Multiple Method with Multiple Time Step Integrators. J. Chem. Phys. **106** (1997) 9835-9849
24. Derreumaux, P., Zhang, G., Schlick, T, Brooks, B.R.: A Truncated Newton Minimizer Adapted for CHARMM and Biomolecular Applications. J. Comp. Chem. **15** (1994) 532-555
25. Arnold, V. I.: Mathematical Methods of Classical Mechanics. Springer: New York (1978)
26. Goldstein, H.: Classical Mechanics. Addison-Wesley: Reading (1965)
27. Forest, E., Ruth, R. D.: Fourth-Order Symplectic Integration. Phys. D **43** (1990) 105-117
28. Yoshida, H.: Recent Progress in the Theory and Application of Symplectic Integrators. Celestial Mechanics and Dynamical Astronomy **56** (1993) 27-43
29. Trobec, R., Merzel, F., Janežič, D.: On the Complexity of Parallel Symplectic Molecular Dynamics Algorithms. J. Chem. Inf. Comput. Sci. **37** (1997) 1055-1062
30. Janežič, D., Merzel, F.: An Efficient Split Integration Symplectic Method for Molecular Dynamics Simulations of Complex Systems. In: Proceedings of the

15th IMACS World Congress on Scientific Computation, Modelling and Applied Mathematics, Ed. Sydow, A., Wissenschaft und Technik Verlag, Berlin 1 (1997) 493-498
31. Jorgensen, W.L., Chandrasekar, J., Madura, J.D., Impey, W., Klein, M.L.: J. Chem. Phys. **79** (1983) 926
32. Grubmüller, H., Heller, H., Windemuth, A., Schulten, K.: Generalized Verlet Algorithm for Efficient Molecular Dynamics Simulations with Long-range Interactions. Molecular Simulation **6** (1991) 121-142
33. Garcia-Archilla, B., Sanz-Serna, J.M., Skeel, R.D.: Long-Time-Steps Methods for Oscillatory Differential Equations. SIAM J. Sci. Comput. (to appear)
34. Leimkuhler, B. J., Reich, S., Skeel, R. D.: Integration Methods for Molecular Dynamics. In: IMA Volumes in Mathematics and its Applications. Eds. Mesirov, J., Schulten, K., Springer-Verlag, Berlin **82** (1995)
35. Wisdom, J., Holman, M.: Symplectic Maps for the n-Body Problem: Stability Analysis. Astron. J. **104** (1992) 2022-2029
36. Schlick, T., Barth, E., Mandziuk M.: Biomolecular Dynamics at Long Timesteps: Bridging the Timescale Gap Between Simulation and Experimentation. Ann. Rev. Biophy. Biomol. Struct. **26** (1997) 181-222

Comparison of Geometric Integrators for Rigid Body Simulation

Benedict J. Leimkuhler *

Department of Mathematics
University of Kansas
Lawrence, KS 66045

Abstract. Geometric integrators are numerical timestepping schemes which preserve invariant structures associated to physical dynamical systems. For example, a symplectic integrator is one which preserves a strong differential invariant of the flows of Hamiltonian systems (the 2-form $dq \wedge dp$ associated to canonical variables q, p). For constrained systems such as the rigid body, preservation of geometric phase-flow structure is complicated by the choice of coordinates and the need for efficiency. Nowhere are these issues more critical than in the simulation of rigid body systems. In recent work, several alternative geometric approaches to rigid body systems integrators have been proposed and applied in molecular simulation. In this article, these methods are introduced and compared with a simple model problem.

1 Introduction

Rigid body dynamics play a fundamental role in molecular simulation. The replacement of small bonded atomic groups by rigid units can eliminate the most stringent timestep (stability) restrictions due to interatom bond stretches. Moreover, the use of rigid bodies can be helpful in formulating reduced variable 'macroscopic' models for large polymers. For these reasons it is important to develop accurate, stable and reliable integration methods for systems of interacting rigid bodies. This article discusses recent work on constructing *geometrically sensitive* integrators for rigid body systems which typically exhibit improved qualitative properties compared to traditional methods.

Until lately, it was widely believed that the most important criteria for the design of a discretization method were the order of accuracy (the degree of the leading term in an error expansion in powers of the stepsize), error constants (the magnitude of the coefficient of the leading monomial in the error expansion), and the stability interval (which defines the largest stable integration stepsize for a given linear oscillator frequency).

In the context of molecular simulation, particularly biomolecular modelling, a critical aspect for numerical simulation is the presence of long-range Coulombic forces which render the force computations much more costly

* The author was supported by NSF Grant No. DMS-9627330.

than the other parts of the computation. For this reason, usually only methods which are *explicit in the forces* can usually be considered for molecular applications.

Recent mathematical work suggests that—especially for nonlinear phenomena—certain geometric properties can be as important as accuracy and (linear) stability. It has long been known that the flows of Hamiltonian systems posess invariants and symmetries which describe the behavior of groups of nearby trajectories. Consider, for example, a two-dimensional Hamiltonian system such as the planar pendulum ($H = \frac{1}{2}p^2 - \cos(q)$) or the polar Kepler problem ($H = \frac{1}{2}p^2 - 1/q + \frac{l^2}{2q^2}$). Such a Hamiltonian system defines a one-parameter family of maps $\{\Phi_t : t \in \mathbf{R}\}$ of $\mathbf{R}^2$ which take points (p, q) to their evolution through t units of time. If such a map is applied point-wise to a region R of the pq-plane, the area is found to remain invariant. This is the simplest manifestation of the hierarchy of *symplectic* (or *Poincaré*) *invariants* which exist for Hamiltonian flow maps. The fact that many physical Hamiltonians are invariant under the substitution $p \to -p$ implies that the corresponding flow maps are *time-reversible*: starting from an initial point (p, q), integrating forward in time τ units, then changing the sign of the momentum, is the same as first changing the sign of the momentum, then integrating *backward* in time. Since flow maps of Hamiltonian systems are always symplectic, and often time-reversible, it is natural to impose similar restrictions on the numerical integration method used for their approximation. Symplectic integration algorithms (and, to a lesser extent, reversible methods) have been shown to possess excellent long-term energy stability, often far superior to 'traditional' (nonsymplectic) methods (even those with higher-order local error). For a recent example showing the improved long-term stability of integration methods incorporating geometric structure, see [15]. In particular, one typically observes drift in the energy error over long time intervals when a nonsymplectic method is used, whereas no such systematic drift is seen for symplectic integrations. An explanation for the promising behavior of symplectic methods is that, up to a very small error of magnitude $O(\exp(-K/\Delta t))$, such a method can be viewed as the flow map of a nearby Hamiltonian systems, constructed through an asymptotic expansion in powers of the stepsize $\tilde{H} = H + \Delta t H^{(1)} + \Delta t^2 H^{(2)} + \ldots$. For more detailed discussions of these theoretical issues, the reader is referred to the recent literature [7, 18, 28].

Methods for simulating rigid bodies typically rely on introduction of some set of generalized coordinates describing the position and orientation of each body. There are several popular parameterizations in use, including (1) quaternions, (2) cartesian (particle) models, and (3) orientation matrix description. Once a set of variables is chosen, the equations of motion are recast in those variables, and the resulting differential equations are then discretized in time using some numerical timestepping method.

The choice of parameterization and the design of a discretization method are not independent: Some choices of parameters will facilitate symplectic/reversible discretization while others may make this task very difficult or render the resulting scheme practically useless because of the computational expense involved.

Parameterization by the use of Hamilton's quaternions has been popular for rigid body molecular computations for many years (see [12, 2]). Unfortunately, the use of these coordinates impedes symplectic/reversible discretization since it 'mixes' the position and momentum variables in a Hamiltonian description. For practical calculations on large systems, one is essentially forced to use nonsymplectic methods such as Gear predictor-corrector methods, Runge-Kutta schemes or extrapolation. These nonsymplectic methods can introduce artificial asymptotically stable (i.e. dissipative) equilibria in what is supposed to be a conservative system (see [11] and the experiments of this article). Moreover, this effect can be viewed as a direct consequence of the nonsymplectic character of the map, since it is well known that iterated symplectic maps do not possess such behaviors. Although one can easily develop "projected" energy-conserving variants of the quaternionic scheme, their long term behavior is little better than the unprojected variant (an example is included in this article).

Particle models offer a simple means for easily and efficiently incorporating the symplectic structure. In some sense, the particle description is exceedingly natural: the standard definition of a rigid body is a relatively rigid collection of massive point particles (see e.g. [1]). The particles need not have direct physical significance: given any rigid body whose inertial tensor, center of mass, and total mass are provided, one can develop an equivalent representation in terms of point masses subject to rigid rod constraints. An important benefit of this choice of integration variables is that the equations of holonomically constrained particle motion are easily solved by use of the SHAKE (or RATTLE) discretization [30, 3], a generalization of Verlet which has been shown to be symplectic [22]. This approach was used by Ciccotti et al [9] to treat small rigid polyatoms, albeit without recognition of the symplectic character of the algorithm. More recently [5], the technique was generalized and applied to treat chains of rigid bodies, with the assistance of special SHAKE-SOR and sparse Newton methods for treating the nonlinear equations arising at each step of integration.

There are two applications where the particle approach is at a disadvantage. First, if the interactions between bodies are naturally expressed not as "site-to-site" potentials but, for example, in terms of the dipolar alignment of the rigid bodies, then the translation of these interaction forces to forces acting on the individual particles can be computationally expensive. Second, whenever the rigid bodies are not linked by constraints but interact only through "soft" forces such as Lennard-Jones or Coulombic potentials, then it turns out that an *explicit* symplectic integration method, not requiring the solution of any nonlinear algebraic system, is available.

The key to these more efficient treatments is a natural canonical formulation of the rigid body dynamics in terms of "rotation matrices." The orientational term of the Lagrangian in these variables can be written simply as

$$\mathcal{L}_{\text{rot}} = \frac{1}{2}\text{trace}(\dot{Q}J\dot{Q}^T), \tag{1}$$

where Q is the 3×3 rotation matrix (subject to an orthogonality constraint $Q^TQ = E$, E the 3×3 identity), and J is a constant 3×3 related to the inertial tensor.[1] In these variables, the orientational interactions of two bodies are trivial to write down, and the work involved in computing the interaction potential and torque is minimal. Moreover, for a free rigid body expressed in the rotation matrix formulation, Reich [27] and, independently, MacLachlan [24] have given a simple reduction procedure that leads to an explicit second order symplectic and reversible discretization; this approach is easily adapted to free rigid bodies interacting in soft forces. The Reich-MacLachlan scheme does not apply to systems linked by constraints; for these problems we have two reasonable options: (1) a particle model or (2) a method which uses rotational matrices but which enforces the orthogonality of the rotation matrix as a set of holonomic constraints, using SHAKE to discretize the resulting system [25]. A discussion of these two alternatives to rigid body dynamics may be found in [26].

These various techniques were recently applied to molecular simulations [11, 20]. Both of these articles used the rotation matrix formulation, together with either the explicit reduction-based integrator or the SHAKE method to preserve orthogonality directly. In numerical experiments with realistic model problems, both of these symplectic schemes were shown to exhibit vastly superior long term stability and accuracy (measured in terms of energy error) compared to quaternionic schemes.

In this article, we briefly describe these symplectic methods, citing recent articles for most of the details of derivation and implementation. We compare the various algorithms in terms of theoretical and implementation aspects, as well as in simple numerical experiments.

2 Background: Symplectic and Reversible Discretization

An example of a symplectic/time-reversible method is the Verlet (leap-frog) scheme. This method is applicable to *separated*[2] Hamiltonian systems of the

[1] If $I = \text{diag}(I_1, I_2, I_3)$ is the inertial tensor, then J is also a diagonal matrix with $J_i = I_k + I_l$, for (i, k, l) a cyclic permutation of $(1, 2, 3)$.

[2] We will use the term "separated" for this class of Hamiltonians. Usually the term "separable" is used in numerical analysis to describe this class, but this usage conflicts with an established meaning of the same term in the literature of quantum mechanics.

form

$$H(p,q) = T(p) + V(q).$$

We will assume in this article that the system is time-reversible, so $T(p) = T(-p)$. Dichotomic Hamiltonians arise from elementary particle models, the simplest nontrivial class of conservative systems. Moreover, even seemingly more complex systems can usually be written in the dichotomic form through change of variables or introduction of additional degrees of freedom.

A timestep of size Δt with the Verlet method ("velocity Verlet") takes (q_0, p_0) to (q_1, p_1) and can be divided into three steps: (1) a 'kick'

$$p_{\frac{1}{2}} := p_0 - \frac{1}{2}\Delta t \nabla_q V(q_0)$$

(2) 'drift'

$$q_1 := q_0 + \Delta t \nabla_p T(p_{\frac{1}{2}})$$

and (3) another 'kick':

$$p_1 := p_{\frac{1}{2}} - \frac{1}{2}\Delta t \nabla_q V(q_1)$$

The symmetry $T(p) = T(-p)$ implies that reversing the order of these three steps and changing the sign of τ and p results in exactly the same method. In other words, Verlet is time-reversible. (In practice, the equations are usually reduced to equations for the positions at time-steps and the momenta at half-steps, only, but for consideration of time-reversibility or symplecticness, the method should be formulated as a mapping of phase space.)

We will introduce the following notation to describe the flow map of a Hamiltonian system with Hamiltonian H:

$$\Phi_{\Delta t} = \exp(\Delta t L_H)$$

where L_H is the operator

$$L_H = \sum_{i=1}^{N} \frac{\partial H}{\partial q_i}\frac{\partial}{\partial p_i} - \frac{\partial H}{\partial p_i}\frac{\partial}{\partial q_i}$$

and we define the exponential by its formal series ($\exp(a) = E + a + a^2/2 + \ldots$). With these definitions, we can view the Verlet method as an approximation to $\exp(\Delta t L_H)$ generated by concatenating the flows of T and V, in other words

$$\exp(\Delta t L_H) \approx \exp(\frac{1}{2}\Delta t L_V)\exp(\Delta t L_T)\exp(\frac{1}{2}\Delta t L_V)$$

This approximation is a special case of the Baker-Campbell-Hausdorff lemma; for additional discussion and extensions to more general classes of methods,

the reader is referred to the literature (see, e.g., [32]). Because the Verlet method is a concatenation of Hamiltonian flow maps, each of which is a symplectic map, we see that the overall map is also symplectic. Such a splitting framework was used effectively by Berne, Martyna and Tuckerman in their work on "multiple timestepping" [8].

The concept of a symplectic method is easily extended to systems subject to holonomic constraints [22]. For example the RATTLE discretization is found to be a symplectic discretization. Since SHAKE is algebraically equivalent to RATTLE, it, too, has the long-term stability of a symplectic method.

A key feature required of a Hamiltonian system that leads to an efficient method based on splitting is the ability to separate the Hamiltonian into p-dependent and q-dependent terms.

3 Rigid Body Formulation and Discretization

We now consider the formulation of the equations of motion for a rigid body pinned at its center of mass and acted on by a (possibly nonlinear) potential field. The Lagrangian in this case is

$$\mathcal{L} = \mathcal{L}_{\text{rot}} - V_{\text{ext}}$$

where V_{ext} is a function of position (i.e. the orientation matrix Q) of the body.

We consider each of three different choices of formulation and the subsequent effect that these have on discretization of the problem.

3.1 Quaternions

Introduce parameters $\sigma = (\sigma_1, \sigma_2, \sigma_3, \sigma_4)$ subject to the constraint $\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \sigma_4^2 = 1$. The rotation matrix is defined in terms of these four parameters by

$$Q = \begin{bmatrix} \sigma_1^2 + \sigma_2^2 - \sigma_3^2 - \sigma_4^2 & 2(\sigma_2\sigma_3 - \sigma_1\sigma_4) & 2(\sigma_2\sigma_4 + \sigma_1\sigma_3) \\ 2(\sigma_2\sigma_3 + \sigma_1\sigma_4) & \sigma_1^2 - \sigma_2^2 + \sigma_3^2 - \sigma_4^2 & 2(\sigma_3\sigma_4 - \sigma_1\sigma_2) \\ 2(\sigma_2\sigma_4 - \sigma_1\sigma_3) & 2(\sigma_3\sigma_4 + \sigma_1\sigma_2) & \sigma_1^2 - \sigma_2^2 - \sigma_3^2 + \sigma_4^2 \end{bmatrix}$$

We will assume that the reference (initial) axis of the top is μ_0. Then $\mu(t) := Q(t)\mu_0$ gives the orientation of the top at time t.

The quaternions obey coupled differential equations involving the angular velocities $\omega_1, \omega_2, \omega_3$ expressed in the body frame (i.e. ω_1 represents the angular velocity about the first axis of inertia, etc.). These differential equations take the form

$$\frac{d\sigma}{dt} = B(\sigma)\omega; \qquad I\frac{d\omega}{dt} = \tau + \omega \times I\omega$$

where $I = \mathrm{diag}(I_1, I_2, I_3)$ is the (diagonalized) inertial tensor,

$$B(\sigma) = \begin{bmatrix} \sigma_1 & -\sigma_2 & -\sigma_3 & -\sigma_4 \\ \sigma_2 & \sigma_1 & -\sigma_4 & -\sigma_3 \\ \sigma_3 & \sigma_4 & \sigma_1 & -\sigma_2 \\ \sigma_4 & -\sigma_3 & \sigma_2 & \sigma_1 \end{bmatrix},$$

and τ is the applied torque due to the potential field.

A "Hamiltonian" version of the quaternionic description is also possible by viewing the quaternions as a set of generalized coordinates, introducing those variables into the rigid body Lagrangian (1), and finally determining the canonical momenta through the formula

$$p_\sigma = \frac{\partial L}{\partial \dot{\sigma}},$$

subject to appropriate constraints ($|\sigma|^2 = 1$, $\sigma \cdot \dot{\sigma} = 0$). The Hamiltonian, when expressed in terms of these generalized positions and momenta, would have a nonseparated character (i.e. it would not be expressible as $T(p)+V(q)$).

The form of the Hamiltonian impedes efficient symplectic discretization. While symplectic discretization of the general constrained Hamiltonian system is possible using, e.g., the methods of Jay [19], these methods will require the solution of a nontrivial nonlinear system of equations at each step which can be quite costly. An alternative approach is described in [10] ("impetus-striction") which essentially converts the Lagrange multiplier for the constraint to a differential equation before solving the entire system with implicit midpoint; this method also appears to be quite costly on a per-step basis.

Thus we find that the choice of quaternion variables introduces barriers to efficient symplectic-reversible discretization, typically forcing us to use some off-the-shelf explicit numerical integrator for general systems such as a Runge-Kutta or predictor-corrector method.

3.2 Particles

A particle description of the rigid body can be obtained by introducing four point masses at positions q_i, $i = 1, \ldots 4$, with corresponding masses m_i. One then determines the positions and masses so that the point masses determine the prescribed inertial tensor, total mass, and center of mass, with rod constraints introduced as needed to rigidify the body (see [5] for discussion). The resulting equations of motion, those of a system of particles subject to quadratic constraints, may be treated using either SHAKE or RATTLE discretization, which has been shown to be symplectic [22]. The particle approach requires the integration of twelve position variables and twelve momenta in the general case, plus six algebraic constraints (for planar rigid bodies, this can be reduced slightly). SHAKE or RATTLE discretization reduces to solving the six nonlinear equations for the multipliers, and

requires one external force/torque evaluation per timestep. An inefficiency arises if the forces/torques are not given in terms of the point masses, since those forces must then be translated at each step into the new variables.

3.3 Rotation Matrices

Rotation matrices may be viewed as an alternative to particles. This approach is based directly on the orientational Lagrangian (1). Viewing the elements of the rotation matrix as the coordinates of the body, we directly enforce the constraint $Q^TQ = E$. Introducing the canonical momenta P in the usual manner, there results a constrained Hamiltonian formulation which is again treatable by SHAKE/RATTLE [25, 27, 20]. For a single rigid body we arrive at equations for the orientation of the form[25, 27]

$$\begin{aligned}\frac{d}{dt}Q &= PJ^{-1},\\ \frac{d}{dt}P &= \hat{\tau} - 2Q\Lambda,\\ Q^TQ &= E,\end{aligned}$$

where $\hat{\tau} = -\frac{\partial}{\partial Q}V(Q)$ is a matricial representation of the torque acting on the body due to a potential function V. After SHAKE discretization, we have[25]

$$\begin{aligned}Q^{n+1} &= Q^n + \Delta t P^{n+1/2}J^{-1}\\ P^{n+\frac{1}{2}} &= P^{n-\frac{1}{2}} + \Delta t\hat{\tau}^n - 2Q^n\Lambda^n\\ (Q^{n+1})^TQ^{n+1} &= E\end{aligned}$$

These equations reduce to a 3×3 matrix Ricatti equation in this case. In the appendix of [20], the efficient iterative solution of this nonlinear system is considered, as is the specialization of the method for linear and planar molecules. In the special case of linear molecules, the SHAKE-based method reduces to a scheme previously suggested by Fincham[14].

It has been observed by [27, 24] that the equations of motion of a free rigid body are subject to reduction. (For a detailed discussion of this interesting topic, see [23].) This leads to an unconstrained Lie-Poisson system which is directly solvable by splitting, i.e. the Euler equations in the angular momenta:

$$\dot{\pi} = \pi \times I^{-1}\pi, \qquad \pi = (\pi_1, \pi_2, \pi_3).$$

Setting $H = \frac{1}{2}\pi \cdot I^{-1}\pi = \frac{1}{2}(\pi_1^2/I_1 + \pi_2^2/I_2 + \pi_3^2/I_3)$, and introducing the symplectic structure matrix $\mathcal{J} = \text{skew}(\pi)$, we observe that the system is of the form $\frac{d}{dt}\pi = \mathcal{J}\nabla H$. Then writing $H = H_1 + H_2 + H_3$, with $H_i = \frac{1}{2I_i}\pi_i^2$, we can obtain a second order "symplectic" (symplectic in the sense of the J symplectic structure) explicit algorithm by solving in sequence, each for a timestep of size Δt,

$$\dot{\pi} = \frac{1}{2}\mathcal{J}\nabla H_1, \quad \dot{\pi} = \frac{1}{2}\mathcal{J}\nabla H_2, \quad \dot{\pi} = \mathcal{J}\nabla H_3, \quad \dot{\pi} = \frac{1}{2}\mathcal{J}\nabla H_2, \quad \dot{\pi} = \frac{1}{2}\mathcal{J}\nabla H_1.$$

Each of the Hamiltonians H_i is easily integrated in terms of rotations about one or the other of the Eulerian axes.

This approach has recently been applied to treat N rigid body molecular systems[11]. At each step of the integration, the orientational degrees of freedom for each rigid body are solved by implementing this reduction, realized as operations directly on the rotation matrix, Q, together with an associated update of the canonical momentum variable P; this is combined with a leapfrog integration of the translational degrees of freedom. Notes are also provided in [11] indicating how these methods are implemented in conjunction with site-site or dipole-type potentials, and some special tricks for handling symmetric molecules more efficiently are discussed.

Observe that, in principle, it is possible to introduce quaternions in the solution of the free rotational part of a Hamiltonian splitting, although there is no compelling reason to do so, since the rotation matrix is usually a more natural coordinatization in which to describe interbody force laws.

3.4 Other Methods

One can write down any number of alternative schemes for treating constrained rigid body dynamics. One frequently suggested idea is to differentiate the constraints several times to eliminate the Lagrange multipliers, then solve the resulting ordinary differential equations for the positions and velocities using a generalized leapfrog or other integration method. This approach introduces additional computational complexity, as well as a drift from the orthogonality constraint if the constraint is not enforced directly by some sort of projection. Moreover, this method is not symplectic. A symplectic approach similar to this can be developed, as in [21], by employing Dirac's concept of weak invariants, but the resulting method is again inefficient compared to the above-mentioned schemes. Various other methods have been suggested [1, 13], but none appear to be serious competitors to the symplectic approaches for long term integrations.

One may attempt to correct an integrator by an energy projection, i.e., in the general case of integrating some Hamiltonian $H(q,p)$, after a specified number l of steps, we would solve the equation

$$H(q, \alpha p) = H_0$$

for a scale factor α; the momenta are then adjusted accordingly. This is a common technique and results in a method which better conserves energy. Such a method will sometimes show slight improvement in its dynamical behavior, but experience indicates that the potential improvement diminishes with the complexity of the system and the system dimension. Even in our relatively simple experiment (see below) little improvement is obtained.

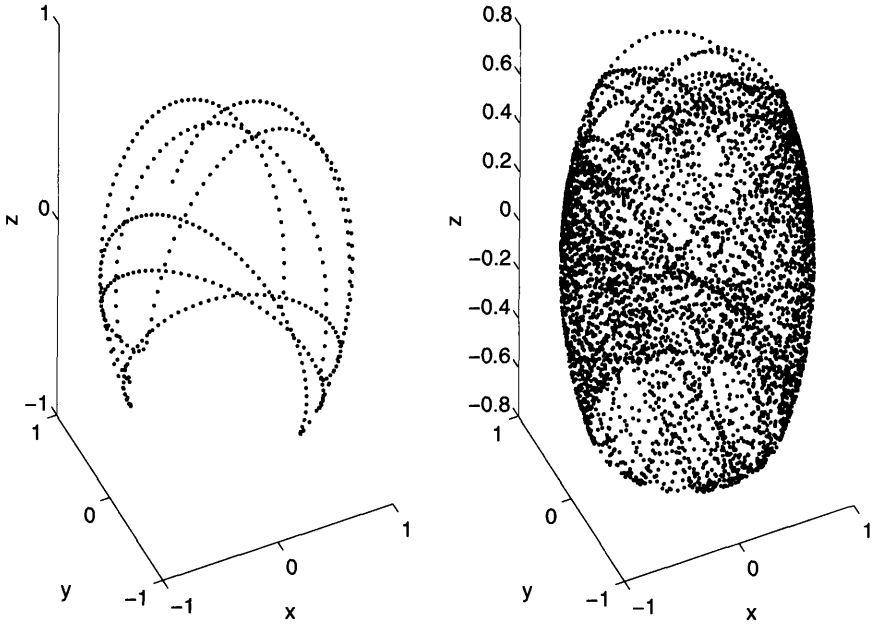


Fig. 1. Motion of a material point on the body over time (left, short time interval; right, long interval). The rigid body swings repeatedly toward the plane where it is repelled by the strong short-range force.

4 Comparisons

Some comparisons among the various types of methods for molecular systems may be found in [11] (rotational SHAKE vs. quaternions/Gear multistep methods) and [20] (reduction/splitting vs. quaternions/extrapolation methods). Qualitatively speaking, the stories are similar. In a single step, the higher-order quaternionic schemes are seen to have much smaller local truncation error, however, the quaternionic schemes are found to exhibit a clear drift in energy, whereas the symplectic schemes show no such problems. In long term simulations, the symplectic methods clearly win out over the nonsymplectic schemes, and are, from this point of view, much more efficient as well.

As an illustration, we consider a simple example of a top with a fixed point at the center of mass moving in an applied field not dissimilar from those encountered in molecular simulations. Specifically, we used

$$V_{\text{ext}} = \phi((\beta + Q_{33})); \qquad \phi(d) = -d^{-1} + \epsilon d^{-10}; \beta = 1.1, \quad \epsilon = 0.001.$$

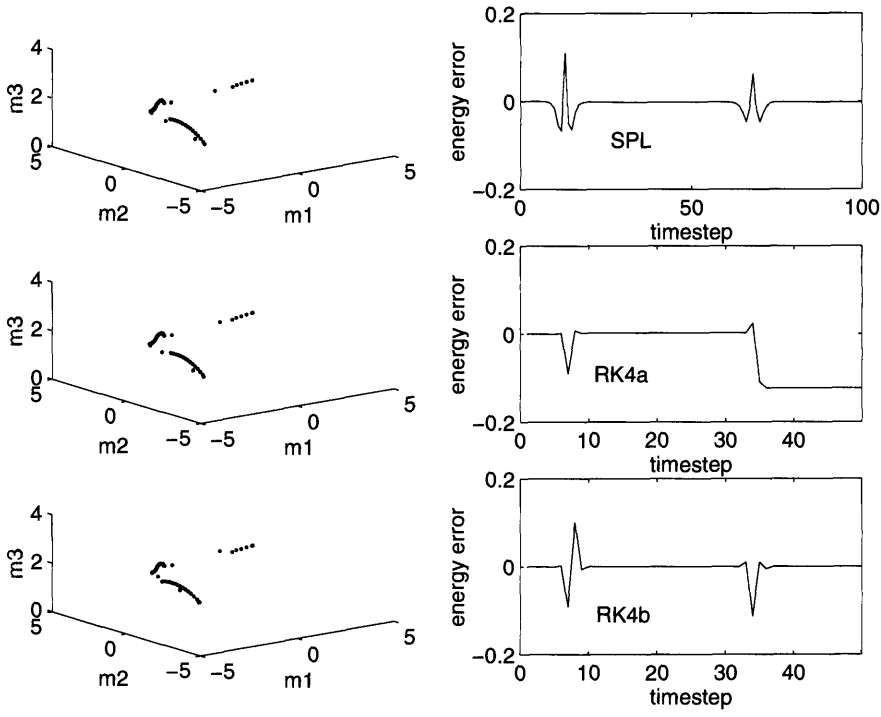


Fig. 2. Top left: short time interval results for the SPL method with stepsize $\Delta t = 0.05$ (100 steps); top right: corresponding energy variation. Center and Bottom: equivalent diagrams for RK4a ($\Delta t = 0.1$) and RK4b ($\Delta t = 0.1$).

This represents an attractive (Coulombic) potential coupled with a repulsive "soft" wall, relative to a plane situated just below the rigid body. The rigid body is repeated drawn toward the plane, then repelled sharply from the wall.

We integrated the system with four different methods: (I) the method based on a Hamiltonian splitting followed by reducing the free rigid body to the Euler equations which we will label "SPL", (II) rotational SHAKE ("ROT"), (III) quaternions/Runge-Kutta-4 ("RK4a"), and (IV) The same RK-4 method except that the angular velocities were scaled every 20 timesteps so that the energy was preserved ("RK4b"). We used an asymmetric rigid body with inertial axes $I_1 = 2$, $I_2 = 3$, $I_3 = 4.5$, started from an initially horizontal configuration with initial angular momenta $(2, 2, 2)$. A 3D plot of the motion of a material point on the body is shown in Fig. 1. It turns out to be advantageous to visualize the motion in terms of the locus of momenta $\Omega = \{m(t) | t \in \mathbf{R}\}$.

In experiments, the two symplectic methods ROT and SPL performed very similarly in terms of error propagation and long term stability. The ex-

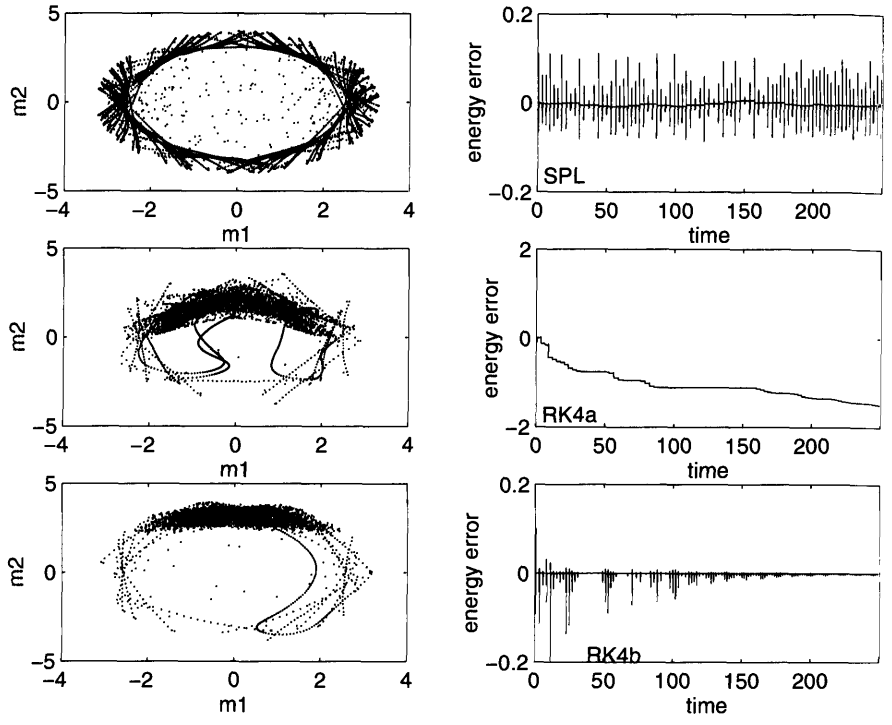


Fig. 3. Top left: long time interval results for the SPL method with stepsize $\Delta t = 0.05$ (to T=250); top right: corresponding energy variation. Center and Bottom: equivalent diagrams for RK4a ($\Delta t = 0.1$) and RK4b ($\Delta t = 0.1$).

planation is that these two methods can be viewed as based on an identical splitting into kinetic and potential energy terms and differ essentially in the means of propagating the free rigid body. In general, in cases such as this one where either symplectic method can be used, it seems that the reduction method is to be preferred on the basis of robustness and also computational efficiency. In other applications, for example for chains of rigid bodies connected by constraints, the reduction method is not available, so the SHAKE approach (either with particles or rotation matrices) is the natural choice.

The implementations were tested by ensuring that they agreed in a short integration with a very small stepsize and that they exhibited the correct order of accuracy (two for the symplectic methods, and four for the methods based on RK-4). The locus of momenta from a short interval calculation (100 timesteps) with splitting method and a stepsize $\Delta t = 0.05$ is shown in Fig. 2(top). The Runge-Kutta method uses four times as many evaluations of the applied torque per timestep as the splitting or shake schemes; however, it proved impossible to obtain reasonable results using a fourfold larger timestep

in the RK schemes. Therefore, we compare those methods using $\Delta t = 0.1$. On the short interval, the results obtained with both the unprojected (Fig. 2(center) and projected RK (Fig. 2(bottom)) methods appear to be quite similar to those of the symplectic method.

Note the curious "return" property exhibited by the energy in the symplectic method; this is a manifestation of the "nearby Hamiltonian" mentioned in the introduction (see [7, 18, 28]).

Looking on a longer time scale reveals quite a different story. Comparative results are shown in Fig. 3 for the same three methods. The splitting method (top) obtains a regular, symmetric momentum portrait that is not qualitatively different from the true dynamics; energy oscillates but does not drift substantially. The Runge-Kutta method (center), on the other hand, rapidly obtains a wholly nonphysical dissipative equilibrium. This loss of structure is associated to a breakdown in energy conservation, but projecting the method so that energy is conserved (bottom) does not help.

Although smaller timesteps would eventually allow the RK-4 method to produce more accurate results, the method must pay a large price in terms of function evaluations. Qualitatively similar results can be expected from other popular nonsymplectic schemes.

These experiments confirm observations in the recent articles [20] and [11]: symplectic methods easily outperform more traditional quaternionic integration methods in long term rigid body simulations.

References

1. R. Ahlrichs and S. Brode, Comput. Phys. Commun. **42**, 59 (1986).
2. M.A. Allen and D.J. Tildesley, *Computer Simulation of Liquids*, Oxford Science Press, Oxford (1987).
3. Andersen, H.C., J. Comput. Phys.,**52**, 24 (1983).
4. V.I. Arnold, *Mathematical Methods of Classical Mechanics*, Springer-Verlag, New York (1978).
5. E. Barth and B. Leimkuhler, Commun. Fields Inst., **10**, 25 (1995).
6. E. Barth, B. Leimkuhler and S. R Reich, SIAM J. Sci. Comput, to appear.
7. G. Benettin and A. Giorgilli, *J. Stat. Phys.*, **74**, 1117-1143 (1994).
8. B. Berne, M. Tuckerman, G. Martyna, J. Chem. Phys., 97, 1990 (1993).
9. Ciccotti, G. Ferrario, M., and Ryckaert, J.P., Mol. Phys., **47**, 6:1253 (1982).
10. D.J. Dichmann and J.H. Maddocks, J. Nonl. Sci., **6**, 271 (1992).
11. A. Dullweber, B. Leimkuhler, R. McLachlan, J. Chem. Phys., **107**, 5840 (1997).
12. D.J. Evans, Mol. Phys. **34**, 317 (1977).
13. D. Fincham, Mol. Simul. **8**, 165 (1992).
14. D. Fincham, Mol. Simul. **11**, 79 (1993).
15. J. Frank, W. Huang, and B. Leimkuhler, J. Comput. Phys., **133**, 160 (1997).
16. H. Goldstein, *Classical Mechanics*, Addison-Wesley, Reading, Massachussetts (1980).
17. E. Hairer, Appl. Num. Math., to appear (1997).
18. E. Hairer and C. Lubich, *Numer. Math.*, **76**, 111-162 (1997).

19. L. Jay, SIAM J. Num. Anal., **33**, 368-387 (1996).
20. A. Kol, B. Laird, and B. Leimkuhler, J. Chem. Phys., **107**, 2580 (1997).
21. B. Leimkuhler and S. R Reich, Math. Comp., **63**, 589 (1994).
22. B. Leimkuhler and R. D Skeel, J. Comput. Phys., **112**, 117 (1994).
23. J. Marsden and T. Ratiu, *Mechanics and Symmetry*, Springer-Verlag, New York, 1996.
24. R.I. McLachlan, Phys. Rev. Lett., **71**, 3043 (1993).
25. R.I. McLachlan and C. Scovel, J. Nonlinear Sci. **5**, 233 (1995).
26. S. Reich, Symplectic integrators for systems of rigid bodies, Fields Institute Communications, **10**, 181-191 (1996).
27. S. Reich, Physica D, **76**, 375-383 (1994).
28. S. Reich, *Dynamical Systems, Numerical Integration, and Exponentially Small Estimates*, Habilitationsthesis, Konrad Zuse Center, Free University, Berlin (1997).
29. R. D. Ruth, IEEE Trans. Nucl. Sci., **30**, 2669 (1983).
30. J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, J. Comput. Phys. **23**, 327 (1977).
31. J.M. Sanz-Serna, BIT, **28**, 877 (1988).
32. J.M. Sanz-Serna and M.P. Calvo, *Numerical Hamiltonian Problems*, Chapman and Hall, New York (1995).
33. D. Wei and G.N. Patey, Phys. Rev. A **46**, 2043 (1992).

Part IV

Quantum-Classical Simulations

New Methods in Quantum Molecular Dynamics of Large Polyatomic Systems

Pavel Jungwirth[1] and R. Benny Gerber[2,3]

[1] J. Heyrovsky Institute of Physical Chemistry, Academy of Sciences of the Czech Republic, Dolejskova 3, 18223 Prague 8, Czech Republic
[2] Department of Physical Chemistry and The Fritz Haber Research Center, The Hebrew University, Jerusalem 91904, Israel
[3] Department of Chemistry, University of California, Irvine, CA 92697-2025, U.S.A.

Abstract. We present new methods for time-dependent quantum mechanical simulations of large polyatomic systems and their applications to photochemical processes in clusters. Two related approaches are discussed: The Classical Separable Potential (CSP) approach, and its extension towards Configuration Interaction (CI-CSP). The former scheme assumes separability of the vibrational modes of the system, and describes each mode as moving in a mean field due to the other modes. The basic idea, which allows for quantum simulations of hitherto unaccesibly large systems, is that the effective single-mode potentials are obtained from a classical MD simulation that precedes the quantum calculation. The second approach represents an improvement that corrects for correlations between different modes, resulting in a scheme of good accuracy. Applications of the methods are presented for dynamics following photodetachment in a small $I^-(Ar)_2$ cluster (where comparison with numerically exact calculation is possible) and for photoexcitation dynamics and spectroscopy of atomic and molecular impurities in large clusters, such as $I_2(Ar)_{17}$ and $I_2(Ar)_{47}$. Future directions of method development are suggested in the light of the algorithmic aspects and the applications.

1 Introduction

The preferable theoretical tools for the description of dynamical processes in systems of a few atoms are certainly quantum mechanical calculations. There is a large arsenal of powerful, well established methods for quantum mechanical computations of processes such as photoexcitation, photodissociation, inelastic scattering and reactive collisions for systems having, in the present state-of-the-art, up to three or four atoms, typically.[1-9] Both time-dependent and time-independent "numerically exact" algorithms are available for many of the processes, so in cases where potential surfaces of good accuracy are available, excellent quantitative agreement with experiment is generally obtained. In addition to the full quantum-mechanical methods, sophisticated semiclassical approximations have been developed that for many cases are essentially of near-quantitative accuracy and certainly at a level sufficient for the interpretation of most experiments.[10-13] These methods also are com-

putationally applicable in the present state-of-the-art, to systems of several atoms only.

As far as methodology is concerned, the situation for large polyatomic systems and for condensed phases is very different. Quantum mechanical simulation methods of quantitative accuracy have not been available for dynamical processes in systems of many coupled degrees of freedom. Classical Molecular Dynamics (MD) simulations have been the prime theoretical tool for interpretation, analysis and comparison with experiment. They are computationally feasible in the present state-of-the-art for systems having up to tens of thousands of atoms, and have been extensively applied to processes in biomolecules, large molecular aggregates and condensed phases. However, many physical properties cannot be treated classically, at least not quantitatively. Resonance Raman spectroscopic intensities in condensed phases, and state-to-state transitions in atom scattering from large molecules are only two examples in a virtually endless list of observable properties for which classical calculations are not adequate.[14] To this one must add the need to describe quantum effects such as zero-point motions, nonadiabatic transitions and tunneling events in the dynamics of many large systems. The need for a quantitative quantum-mechanical description of large polyatomic systems is thus well recognized, but hitherto suitable methods for this purpose have not been at hand. The situation in this respect for some time-independent properties is much better: For instance, methods such as the Diffusion Quantum Monte Carlo (DQMC) and related techniques have been used very effectively to compute ground-state energies and structural properties of large quantum clusters and solids.[15–18] Likewise, Feynman Path Integral methods were very successfully applied to the calculation of thermodynamic and structural properties of such systems at thermal equilibrium.[19–21] Another example are the ground and the low-lying vibrational eigenfunctions of large polyatomic molecules, that were recently calculated for several such systems by the Vibrational Self-Consistent Field (VSCF) method and its extensions.[22–24] Thus, calculations of the ground and of the fundamental excited vib-rotational states were reported for the protein BPTI.[24]

Within the time-dependent framework, the Time-Dependent Self-Consistent Field (TDSCF) approximation is known, in the context of electronic structure theory, from the early years of quantum theory,[29] but its applications to problems of molecular dynamics are much more recent.[30–42,23] Methods of much improved accuracy were developed on the basis of the TDSCF approximation.[34–42] Also, some technical hurdles in applying the method for larger systems were overcome in the last few years at least for some types of interaction potentials, so recently calculations have been reported for problems of significantly increased number of degrees of freedom.[16,37] Different methods for time-dependent quantum simulations of large systems were pursued vigorously by several research groups, and much progress was made. Some of the novel methods proposed are still confined to models or special systems, while for a few of the others realistic applications are already at

hand. Examples of such recent methods are the Path Integral approach developed by Makri and coworkers[25,26] and the semiclassical centroid density approximation of Voth *et al.*[27,28]

The paper discusses two related recent methods for time-dependent quantum simulations of many-atom systems in which the present authors have been involved, the Classical Separable Potential (CSP) method and its Configuration Interaction extension (CI-CSP). While by no means developed to completion or general in applicability yet, these methods provide already practical working algorithms for a wide range of applications. Indeed, in the very short time since the introduction of the CSP method, many applications to large realistic systems were already made, showing that the method is already a "production tool" for extensive simulations of systems with tens to hundreds of degrees of freedom.

The CSP method involves approximations that separate the different modes, and treats each mode as moving in the average field of all the other modes. The effective single-mode potentials are determined from independent classical Molecular Dynamics simulations carried out at the outset, and then used in the quantum calculation. The CI-CSP extension goes beyond the separability approximation by expanding the multidimensional time-dependent wavepacket into a sum of separable terms with variationally determined coefficients. The choice of the CI terms is guided by the individual classical trajectories. In this way, the number of CI terms is grossly reduced and due to a favorable (non-exponential) scaling with the number o degrees of freedom the method is applicable to large polyatomic systems. The applications that will be mentioned in the present paper are all for cluster systems. This seems to us a natural choice: In trying to develop simulation methods ultimately applicable to condensed phase problems, large clusters offer physical properties similar to these phases, while still retaining the simplicity of finite systems.

Sec. 2 presents the CSP method, while an improved approach built on it (CI-CSP) is described in Sec. 3. Sec. 4 gives some examples of applications of the CSP and CI-CSP methods to the photochemical ultrafast dynamics in clusters. Directions for future progress and improvements are discussed in Sec. 5.

2 The Classical Separable Potential (CSP) Method

The principal idea behind the CSP approach is to use input from Classical Molecular Dynamics simulations, carried out for the process of interest as a first preliminary step, in order to simplify a quantum mechanical calculation, implemented in a subsequent, second step. This takes advantage of the fact that classical dynamics offers a reasonable description of many properties of molecular systems, in particular of average quantities. More specifically, the method uses classical MD simulations in order to determine effective

potentials such that by the use of the latter a subsequent quantum calculation becomes far simpler[43)-(48)].

Here, we give here a brief outline of the methods as introduced in Refs. 43, 44, and 47. Suppose that the initial state of the system is $\psi_0(q_1, ..., q_N)$. From ψ_0, the Wigner phase-space distribution $D(q_1, ..., q_N; p_1, ..., p_N)$ is computed. This distribution is used to sample initial positions and momenta $q_1^{(\alpha)}$, ... , $q_N^{(\alpha)}; p_1^{(\alpha)}, ..., p_N^{(\alpha)}$, for a classical trajectory simulation of the process of interest. The set of trajectories $q_1^{(\alpha)}(t), ..., q_N^{(\alpha)}(t)$, for $\alpha = 1, ..., n_T$, can be used to generate in the classical framework any quantity of interest at each time point t. From the trajectories, we compute the average potential acting on each mode j:

$$\bar{V}_j(q_j, t) = \sum_{\alpha=1}^{n_T} V(q_1^{(\alpha)}(t), .., q_{j-1}^{(\alpha)}(t), q_j, q_{j+1}^{(\alpha)}(t), .., q_N^{(\alpha)}(t))\omega_\alpha + \frac{1-N}{N}\bar{V}(t) \quad (1)$$

where $V(q_1, ..., q_N)$ is the full potential function of the system, and the summation in (24) extends over all MD trajectories. ω_α is the weight of the α trajectory in the initial state distribution (as determined from the Wigner distribution). Note that in (24), all the coordinates q_l are evaluated at trajectory points, except the q_j for which the mean potential $\bar{V}_j(q_j, t)$ is evaluated. The j-th coordinate is kept as in the full potential function. $\bar{V}(t)$ in (24) is a coordinate-independent function of time given by:

$$\bar{V}(t) = \sum_{\alpha=1}^{n_T} V(q_1^{(\alpha)}(t), ..., q_j^{(\alpha)}(t), ..., q_N^{(\alpha)}(t))\omega_\alpha \quad (2)$$

The inclusion of the second term on the right-hand-side of (24) is convenient, because in this way the average (over the trajectories) of the sum of the single-mode potentials, $\sum_{j=1}^{n} \bar{V}(q_j, t)$, equals to the average of the full potential function $V(q_1, ..., q_N)$. Using the single-mode effective potentials $\bar{V}_j(q_j, t)$, one can solve a time-dependent Schrödinger equation for each mode:

$$i\hbar \frac{\partial \phi_j(q_j, t)}{\partial t} = [T_j + \bar{V}_j(q_j, t)]\phi_j(q_j, t) \quad (3)$$

where T_j is the kinetic energy operator for mode j. A separable approximation for the total wavepacket of the system is given by

$$\psi(q_1, ..., q_N, t) = \prod_{j=1}^{N} \phi_j(q_j, t) \quad (4)$$

With the above definitions, there is no additional overall phase factor to be included in (27). Eqs. (24)-(27) are the CSP approximation.[43-47] Like TDSCF, CSP is a separable approximation, using a time-dependent mean potential for each degree of freedom. However, the effective potentials in CSP

are obtained from classical MD, and their evaluation is computationally very easy, not involving multidimensional integrals as in the case of TDSCF. This gives CSP a computational advantage that for large systems is enormous, especially for systems where the potential function cannot be simplified to reduce the dimensionality of the TDSCF integrals.

The accuracy of the CSP approximation is, as test calculations for model systems show, typically very similar to that of the TDSCF.[43] The reason for this is that for atomic scale masses, the classical mean potentials are very similar to the quantum mechanical ones. CSP may deviate significantly from TDSCF in cases where, e.g., the dynamics is strongly influenced by classically forbidden regions of phase space. However, for simple tunneling cases it seems not hard to "fix" CSP, by running the classical trajectories slightly above the barrier. In any case, for typical systems the classical estimate for the mean potential functions works extremely well.

It should be noted that due to the use of classical trajectories to evaluate the time-dependent single-mode potentials, the quantum CSP energy is not strictly conserved (TDSCF conserves energy, since the single-mode potentials are obtained self-consistently with the single-mode wavefunctions). The extent of the non-conservation is, however, small and it does not seem to create significant difficulties. The smaller the difference between the TDSCF and the CSP single-mode potentials, the smaller is the CSP energy non-conservation. The more near-classical a system is, the smaller is the non-conservation effect. In general, CSP is expected to be less accurate than TDSCF, although only slightly so, because the quantum effects are not included in the evaluation of the single-mode potentials. In special cases, however, CSP may give slightly better results than TDSCF, because the CSP mean-potentials are calculated from an approach that includes correlation effects between the modes (classical MD). In summary: TDSCF and CSP give very similar results in most cases, but for systems with appreciable quantum effects, TDSCF is more accurate than CSP, while the latter may prove somewhat accurate when correlation effects are quite important.

It is obvious that CSP depends, as does TDSCF, on the choice of coordinates. As pointed out in Sec. 2.2, numerical convenience often limits the choice of the coordinates. CSP may, however, offer practical prospects for the choice of physically optimal modes. The deviation of the true potential from CSP separability is given by:

$$\Delta V_{corr}(q_1, ..., q_N, t) \equiv V(q_1, ..., q_N) - \sum_{j=1}^{N} \bar{V}_j(q_j, t) \tag{5}$$

ΔV_{corr} can be evaluated readily from the classical MD simulation for any choice of coordinate system, and it may be possible to determine the modes that give the smallest ΔV_{corr}. These should be optimal CSP modes. Work along these lines is in progress in our group. So far, however, the coordi-

nate choices in the CSP applications were all straightforward and based on qualitative physical consideration.

3 Configuration Interaction Extension of the CSP method

The main extension hitherto introduced for the CSP approximation is referred to as the Configuration Interaction CSP (CI-CSP) method.[48] This approach represents the full non-separable wavepacket as a linear combination, with time-dependent coefficients, of terms that are each separable. The applications of this method are very encouraging. The method can provide major improvements over CSP, and can probably be pursued at a level of high accuracy, while still being computationally feasible for large systems.[48] The method is flexible, and allows for the use of physical considerations to optimize it. The most important advantage is that within CI-CSP, the classical MD simulations greatly simplify the determination of the terms or configurations, that play the most important role in the total wavefunction. The approach is based on the following: The CSP wavepacket is obtained from mean single-mode potentials, computed as an average over all the trajectories. To include correlation effects, it is necessary to include also wavefunction terms computed from potentials that represent "fluctuations" from the average. Such potentials can be computed along a limited set of trajectories, even individual trajectories (rather than as an average over all trajectories). In the version of CI-CSP discussed here, we start with propagating separable wavepackets using effective potentials obtained from individual trajectories:

$$\bar{V}_j^{(\alpha)}(q_j,t) = V(q_1^{(\alpha)}(t),..,q_{j-1}^{(\alpha)}(t),q_j,q_{j+1}^{(\alpha)}(t),..,q_N^{(\alpha)}(t)) + \frac{1-N}{N}\bar{V}^{(\alpha)}(t) \quad (6)$$

The coordinate-independent quantities $\bar{V}^{(\alpha)}(t)$ are given by:

$$\bar{V}^{(\alpha)}(t) = V(q_1^{(\alpha)},...,q_j^{(\alpha)}(t),...,q_N^{(\alpha)}(t)) \quad (7)$$

In addition to wavepackets propagated along effective potentials corresponding to individual trajectories, we also propagate in CI-CSP the CSP wavefunctions, governed by the average potentials $\bar{V}_j(q_j,t)$ of Eq. (24). The CI-CSP *ansatz* for the total wavepacket of the system is then as follows:[48]

$$\Psi(q_1,...,q_N,t) = c_o(t)\prod_{j=1}^{N}\phi_j(q_j,t) + \sum_{j,\alpha}^{N,M_j} s_j^{(\alpha)}(t)\phi_1...\phi_j^{(\alpha)}...\phi_N +$$

$$+ \sum_{j<j',\alpha,\beta}^{N,M_j,N,M} d_{j\alpha j'\beta}(t)\phi_1...\phi_j^{(\alpha)}...\phi_{j'}^{(\beta)}...\phi_N \quad (8)$$

Here $\phi_j(q_j,t)$ are the CSP functions, $\phi_j^{(\alpha)}(q_j,t)$ denotes the q_j-mode wavefunction propagated on the effective potential generated from the α trajectory. These wavefunctions are orthogonalized and included in the CI expansion when the overlap between ϕ_j and $\phi_j^{(\alpha)}$ drops below a certain threshold (typically 99%). In this way, the CI space grows in the course of the dynamical evolution, corresponding to the increase of correlation between the modes. In the above expansion, M_j is the number of trajectories actually included in the CI expansion (with respect to functions of the mode j), $c_o(t)$ is the CSP coefficient, $s_j^\alpha(t)$ are single mode excitation coefficients, while $d_{j\alpha j'\beta}(t)$ are coefficients corresponding to two-mode excitations. The language of excitation is borrowed from time-independent CI in electronic structure theory,[80] with which the present formulation has strong analogies. Within this analogy, the *ansatz* (31) corresponds to configuration interaction with single and double excitations (neglecting higher-order excitations). From a physical point of view one should rather talk about mode-mode correlations instead of double excitations and mode 'polarization' terms instead of single excitations but we stay with the usual notation. To obtain equations of motion for the coefficients $c_o(t), s_j^\alpha(t), d_{j\alpha j'\beta}(t)$ of the wavefunction expression (31), we substitute the *ansatz* into the time-dependent Schrödinger equation of the system, Eq. (1), with the fully coupled potential. To get practical working equations we make several approximations:

(i) Consider for simplicity the case when the initial state is separable (extension to a non-separable initial state is straightforward, though technically more complicated). Then at $t = 0, c_o = 1$ and all the other CI coefficients are zero. In a short-time approximation we then take into account only coupling between the CSP state, and doubly excited states (it can be shown that coupling between CSP and singly excited states is always zero - this is a time-dependent analogue of the Brillouin theorem, familiar from electronic structure theory[80]). In this way we neglect very small (at least initially) single-single, single-double, and double-double excitation couplings.

(ii) Consider an integral $\langle \Phi | V | \hat{\Phi} \rangle$ where Φ and $\hat{\Phi}$ differ in two modes j and j' only. We approximate this integral by $\langle \phi_j \phi_{j'} | V_{jj'}^{CSP2} | \hat{\phi}_j \hat{\phi}_{j'} \rangle$, where

$$V_{jj'}^{CSP2}(q_j, q_{j'}, t) = \sum_{\alpha=1}^{n_T} V(q_1^{(\alpha)}(t), ..., q_j, ..., q_{j'}, ..., q_N^{(\alpha)}(t))\omega_\alpha \tag{9}$$

where n_T is the total number of trajectories.

Substitution of the *ansatz* (31) into the Schrödinger equation (1) for the full system, together with the above approximations, yields the following equations for the coefficients $c_o(t), d_{j\alpha j'\beta}(t)$ of the CI expansion (31):

$$i\frac{\partial c_o}{\partial t} = \sum_{j,\alpha,j',\beta}^{N,M,N_j,M_{j'}} d_{j\alpha j'\beta}(t) \langle \phi_j \phi_{j'} | V_{jj'}^{CSP2} | \hat{\phi}_j^{(\alpha)} \hat{\phi}_{j'}^{(\beta)} \rangle \tag{10}$$

$$i\frac{\partial d_{j\alpha j'\beta}(t)}{\partial t} = c_o(t)\langle\hat{\phi}_j^{(\alpha)}\hat{\phi}_{j'}^{(\beta)}|V_{jj'}^{CSP2}|\phi_j\phi_{j'}\rangle \tag{11}$$

For separable initial states the "single excitation" terms can be set to zero at all times at this level of approximation. Eqs. (32),(33),(34) together with the CSP equations and with the *ansatz* (31) for the total wavefunction are the working equations for the approach. This form, without further extension, is valid only for short time-domains (typically, a few picoseconds at most). For large times, higher correlations, i.e. interactions between different singly and doubly excited states must be included.

An important computational advantage of CSP and CI-CSP for large systems is that the methods can be applied to any potential function without any need to simplify the latter by expansion, assumption of pairwise interactions, etc., as is the case for TDSCF. Applications of CSP for systems having up to $\sim 10^3$ degrees of freedom, and of CI-CSP for systems of up to ~ 100 modes are not computationally extremely demanding, and were indeed carried out on quite modest scalar workstations, of the type of SGI Indigo, HP Appollo, etc. Finally, implementation of CSP on a parallel computer is straightforward and almost ideal in efficiency. Recently, there has been extensive work on application of the CSP method on parallel computer systems with very encourageing results.[46,47,83]

4 Applications of CSP and of CI-CSP

Almost from the introduction of the CSP and CI-CSP methods, applications to large systems with realistic potential functions became possible. A general CSP code was written and is available.[47] The applications of CSP so far include a study of electron photodetachment dynamics for $I^-(Ar)_n, n = 2, ..., 12$,[44,83] a study of the dynamics following electronic excitation of $Ba(Ar)_n$ for $n = 10$, $n = 20$,[45] and a study of electronic excitation, cage effects and vibrational dephasing and relaxation dynamis for I_2 in $I_2(Rg)_n$, for $Rg = Ar$, Xe and $n = 17, 47$ (corresponding respectively to complete first and complete first two solvation layers around the iodine).[47] The dynamics of $Ba(Ar)_n$ following the excitation of the Ba in this cluster involves nonadiabatic transitions, since there are three quasidegenerate p-states of the Ba atom, and the system is then governed by three potential energy surfaces corresponding to these state, with possible non-adiabatic transitions between them. The simulations of Ref. 45 thus went beyond simple CSP, and were in fact a three-configuration CI-CSP, with a separable term in the CI wavefunction for each adiabatic electronic state. This approach, while not a converged quantum treatment, is expected already to be more reliable than semiclassical "surface hopping" methods that treat nuclear motions classically.[81] Comparison of the CI-CSP with semiclassical "surface

hopping" simulations for $Ba(Ar)_{10}$ and $Ba(Ar)_{20}$, suggests that the "surface hopping" results for properties such as population ratios of the electronic states at, say, $t \approx 1$ ps after excitation, can differ from quantum results by a significant factor, of the order of ~ 2 in this case.[45]

The small $I^-(Ar)_2$ complex serves as a benchmark system since numerically exact calculation is still possible here. By numerically exact calculation we mean here a 3-dimensional wavepacket propagation using the time-dependent vibrational Schrödinger equation with the fully coupled interaction potential. The experimentally observed vibrationally resolved photoelectron spectrum can be modeled as a Fourier transform of the calculated autocorrelation function, i. e. the overlap of the initial total wavepacket with the wavepacket at time t. At the same time the complex autocorrelation function is a very sensitive quantity for testing the quality of approximate approaches since it depends not only on the amplitude but also on the phase of the wavepacket. Fig. 1 depicts the short time CSP and CI-CSP autocorrelation functions, compared to exact and TDSCF results. First, we note that there is qualitative agreement among all four approaches indicating that mean-field methods represent a reasonable approximation even though there is no significant separation in mode frequencies in this system. The excellent agreement between CSP and TDSCF demonstrates that only minor errors are introduced by replacing the quantum mean-field integrals by averages over classical trajectories. Finally, inclusion of two-mode correlations significantly improves the autocorrelation funcion and brings it closer to the exact one.

For $I_2(Ar)_{17}$, an extensive CI-CSP simulation was carried out,[48] and the results were compared with those of the simple CSP approximation.[46] Both calculations are for the ultrafast dynamics following excitation of the I_2 into the B state. We found that the CI-CSP calculation, including "doubly excited configurations", is close to converged for times up to $t \sim 500$ fs, when 1500 configurations are included. Fig. 2 shows $|c_o(t)|^2$, the coefficient of the CSP term and the doubly excited terms $|d_{j\alpha j'\beta}|^2$ in the full CI-CSP wavefunction, versus time.

It turns out that the CSP approximation dominates the full wavefunction, and is therefore almost exact till $t \approx 80$ fs. This timescale is already very useful: The first ≈ 20 fs are sufficient to determine the photoadsorption lineshape and, as turns out, the first ≈ 80 fs are sufficient to determine the Resonance Raman spectrum of the system. Simple CSP is almost exact for these properties. As Fig. 3 shows, for later times the accuracy of the CSP decays quickly: for $t \approx 500$ fs in this system, the contribution of the CSP approximation to the full CI wavefunction is almost negligible. In addition, this wavefunction is dominated not by a few specific terms of the CI expansion, but by a whole host of configurations. The "decay" of the CSP approximation was found to be due to "hard collisions" between the iodine atoms and the "surrounding wall" of argons. Already the first hard collision brings a major deterioration of the CSP approximation, but also the role of the second collision can be clearly identified. As was mentioned, for $t < 80$ fs, the CSP

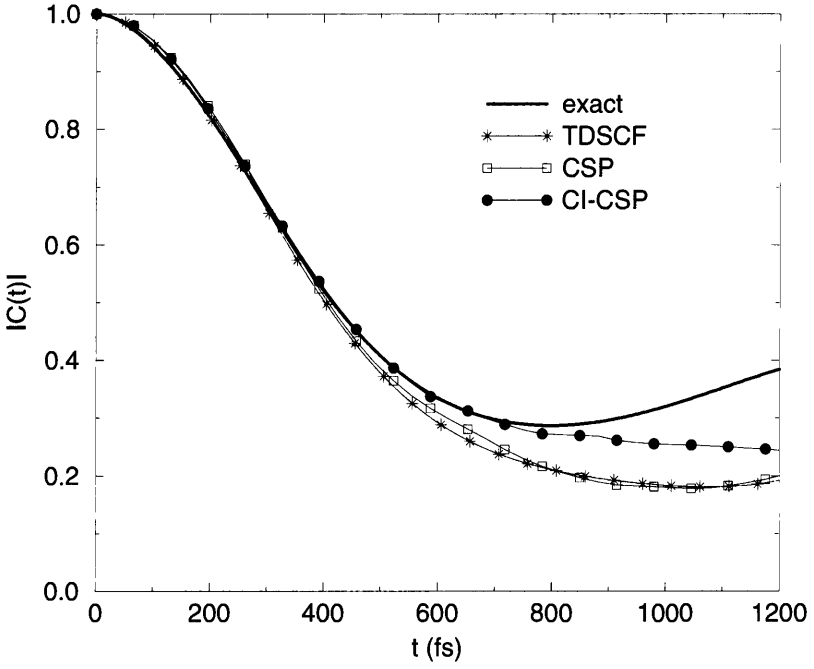


Fig. 1. Comparison between the CID-CSP, CSP, TDSCF, and the numerically exact autocorrelation functions.

is essentially of quantitative accuracy. Fig. 3 shows the Resonance Raman spectrum for $I_2(Ar)_{17}$.[46]

The intensities are plotted vs. v, the final vibrational quantum number of the transition. The CSP results (which for this property are almost identical with CI-CSP) are compared with experimental results for I_2 in a low-temperature Ar matrix.[82] The agreement is excellent. Also shown is the comparison with gas-phase, isolated I_2. The "solvent effect" on the Raman intensities is clearly very large and qualitative. These show that CSP calculations for short timescales can be extremely useful, although for later times the method breaks down, and CI-CSP should be used.

5 Concluding Remarks

Two methods for time-dependent quantum simulations of many-atom systems are examined in this article: the CSP-based and the CI-CSP-based algorithms. The CSP method begins with a separable approximation for

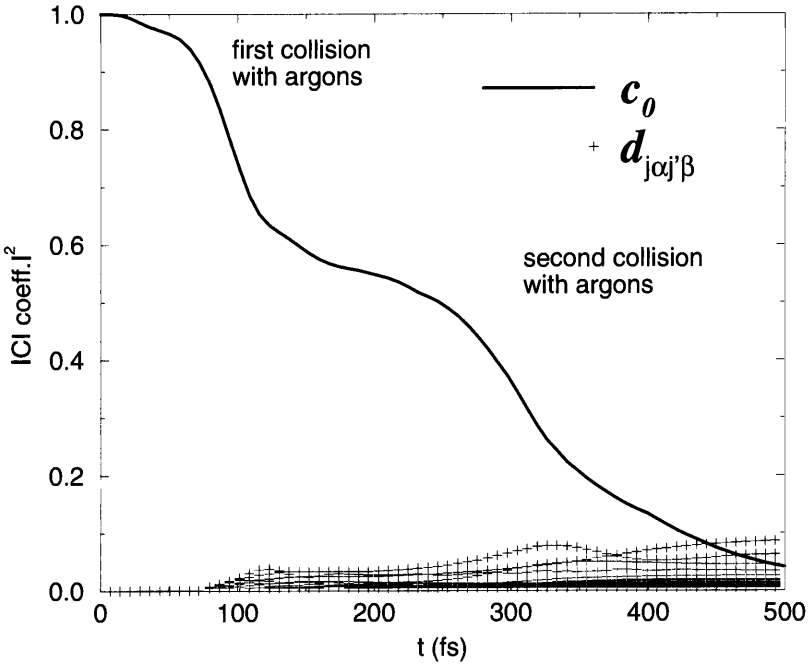


Fig. 2. The contribution c_o of the CSP approximation to the CI wavefunction and the correlation coefficients $d_{j\alpha j'\beta}$ versus time.

the wavefunction, and describes each mode as governed by a mean time-dependent potential due to the other modes, the effective potentials being determined from classical MD simulations. Correlations are included within the CI-CSP approach by expanding the total wavepacket into a sum of separable terms. Classical MD calculation assist in the selection of important terms and in the evaluation of CI coupling terms. Both methods can and have already been applied to systems of substantial sizes, and are thus already practical simulation tools. Applications of CSP and CI-CSP to systems having of the order of $\sim$ 100 degrees of freedom are already at hand, and CSP calculations for biomolecules of 10^4 degrees of freedom are in advanced stages of progress. Clearly, the introduction of correlation effects between modes is computationally the most demanding aspect of the approach. In our belief, a search for simplified methods for correlation corrections should be one of the most useful directions to be followed in the future. An efficient treatment of important correlation effects will benefit in particular processes involving, e.g., chemical (substitution) reactions and nonadiabatic processes, in which

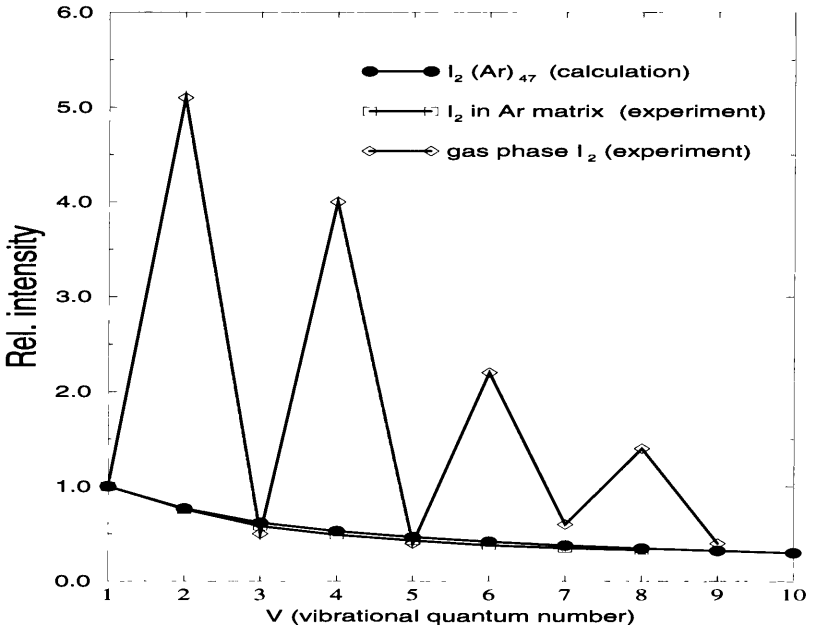


Fig. 3. Resonance Raman Intensities for I_2 $(Ar)_{47}$

the role of dynamical correlations is both extremely important and hard to handle computationally.

The computational efficiency is a major advantage of CSP and CI-CSP, and we expect that in the forthcoming few years CSP-based methods will be extensively used as practical tools for the study of an increased range of dynamical processes in large systems.

Acknowledgements. We thank Dr. B. Schmidt, E. Fredj, Prof. M. A. Ratner, and Dr. A. Marks for helpful discussions and suggestions. The research described here was supported by the Chemistry Division of the NSF (grant CHE-9503258 to R.B.G.) and by the Granting Agency of the Academy of Sciences of the Czech Republic (grant A4040706 to P.J.).

6 References

1. G.C. Schatz, *J. Phys. Chem.* **100**, 12839 (1996).
2. R. Kosloff, *Ann. Rev. Phys. Chem.* **45**, 145 (1994).

3. J.Z.H. Zang and W.H. Miller, *J. Chem. Phys.* **92**, 1811 (1990).
4. U. Manthe, T. Seideman and W.H. Miller, *J. Chem. Phys.* **92**, 10078 (1990).
5. D. Neuhauser, *J. Chem. Phys.* **9272** (1994).
6. U. Peskin and N. Moiseyev, *J. Chem. Phys.* **99**, 4590 (1993).
7. D.J. Kouri, Y. Houng, W. Zhu and D.K. Hoffman, *J. Chem. Phys.* **100**, 3362 (1994).
8. D.H. Zang and J.Z.K. Zang, *J. Chem. Phys.* **100**, 2679 (1994).
9. R. Schinke, *Photodissociation Dynamics* (Cambridge University Press, Cambridge, 1993).
10. W.H. Miller, *Adv. Chem. Phys.* **30**, 77 (1975).
11. E.J. Heller, *J. Chem. Phys.* **94**, 2723 (1994).
12. B.W. Spath and W.H. Miller, *J. Chem. Phys.* **104**, 95 (1996).
13. J.G. Kay and W.H. Miller, *J. Chem. Phys.* **104**, 95 (1996).
14. G.C. Schatz and M.A. Ratner, *Quantum Mechanics in Chemistry* (Prentice Hall, New Jersey, 1993).
15. J.B. Anderson, *J. Chem. Phys.* **63**, 1499 (1975).
16. M.A. Suhm and R.O. Watts, *Phys. Rep.* **204**, 293 (1991).
17. M.A. McMahan, R.N. Barnet and K.B. Whaley, *J. Chem. Phys.* **99**, 8816 (1993).
18. S. Broude and R.B. Gerber, *Chem. Phys. Lett.* **258**, 416 (1996).
19. D. Scharf, G.J. Martyna and M.L. Klein, *J. Chem. Phys.* **99**, 8997 (1993).
20. V. Buch, *J. Chem. Phys.* **100**, 7610 (1994).
21. D. Scharf, G.J. Martyna, D. Li, G.A. Voth and M.L. Klein, *J. Chem. Phys.* **99**, 9013 (1993).
22. J.M. Bowman, *Acc. Chem. Res.* **19**, 202 (1986).
23. R.B. Gerber and M.A. Ratner, *Adv. Chem. Phys.* **70**, 97 (1988).
24. A.E. Roitberg, R.B. Gerber, R. Elber and M.A. Ratner, *Science* **268**, 1319 (1995).
25. N. Makri, *Chem. Phys. Lett.* **193**, 435 (1992).25. N. Makri, *Chem. Phys. Lett.* **193**, 435 (1992).
26. M. Topaler and N. Makri, *Chem. Phys. Lett.* **210**, 285 (1993).
27. J. Cao and G.A. Voth, *J. Chem. Phys.* **100**, 5106 (1994).
28. J. Cao and G.A. Voth, *J. Chem. Phys.* **101**, 6168 (1994).
29. P.A.M. Dirac, *Proc. Camb. Phil. Soc.* **26**, 376 (1930).
30. E.J. Heller, *J. Chem. Phys.* **64**, 63 (1976).
31. R.B. Gerber, V. Buch and M.A. Ratner, *J. Chem. Phys.* **77**, 3022 (1982).
32. Z. Li and R.B. Gerber, *J. Chem. Phys.* **99**, 8637 (1993).
33. E. Fredj, R.B. Gerber and M.A. Ratner, *J. Chem. Phys.* **105**, 1121 (1996).
34. N. Makri and W.H. Miller, *J. Chem. Phys.* **87**, 5781 (1987).
35. H.-D. Meyer, U. Manthe and L.S. Cederbaum, *Chem. Phys. Lett.* **165**, 73 (1990).
36. U. Manthe, H.-D. Mayer and L.S. Cederbaum, *J. Chem. Phys.* **97**, 3199 (1992).
37. G.A. Worth, H.-D. Mayer and L.S. Cederbaum, *J. Chem. Phys.* (in press).

38. J.-Y. Fang and H. Guo, *J. Chem. Phys.* **101**, 5831 (1994).
39. J.-Y. Fang and H. Guo, *J. Chem. Phys.* **102**, 2404 (1995).
40. Z. Kotler, E. Neria and A. Nitzan, *Comput. Phys. Commun.* **63**, 243 (1991).
41. J. Campos-Martinez and R.D. Coalson, *J. Chem. Phys.* **93**, 4740 (1990).
42. A.D. Hammerich, R. Kosloff and M.A. Ranter, *Chem. Phys. Lett.* **171**, 97 (1990).
43. P. Jungwirth and R.B. Gerber, *J. Chem. Phys.* **102**, 6046 (1995).
44. P. Jungwirth and R.B. Gerber, *J. Chem. Phys.* **102**, 8855 (1995).
45. P. Jungwirth and R.B. Gerber, *J. Chem. Phys.* **104**, 5803 (1996).
46. P. Jungwirth, E. Fredj and R.B. Gerber, *J. Chem. Phys.* **104**, 9932 (1996).
47. P. Jungwirth, E. Fredj, P. Zdanska and R.B. Gerber, *Computers&Chemistry* **21**, 419 (1997).
48. P. Jungwirth, E. Fredj and R.B. Gerber, *J. Chem. Phys.* **107**, 8963 (1997).
49. D.J. Thouless, *The Quantum Mechanics of Many-Body Systems* (Academic Press, New York, 1961) pp. 88-93.
50. P. Bonche, S.E. Koonin and J.W. Negele, *Phys. Rev.* **C13**, 1226 (1976).
51. S. Lent, *Phys. Rev.* **C21**, 1594 (1980).
52. J.W. Negele, *Physics Today* **38**, 24 (1985).
53. K. Goeke and P.-G. Reinhard, *Time-Dependent Hartree-Fock and Beyond*, Lecture Notes in Physics, Vol. 171 (Springer-Verlag, Berlin, 1982).
54. V. Buch, M.A. Ratner and R.B. Gerber, *Chem. Phys. Lett.* **101**, 44 (1983).
55. R.B. Gerber, R. Kosloff and M. Berman, *Comput. Phys. Reports* **5**, 59 (1986).
56. M.D. Feit, J.A. Fleck and A. Steiger, *J. Comp. Phys.* **47**, 412 (1982).
57. U. Peskin, R. Kosloff and N. Moiseyev, *J. Chem. Phys.* **99**, 4590 (1993).
58. T.R. Horn, R.B. Gerber and M.A. Ratner, *J. Chem. Phys.* **91**, 1813 (1989).
59. T.R. Horn, R.B. Gerber, J.J. Valentini and M.A. Ratner, *J. Chem. Phys.* **94**, 6728 (1991).
60. R.H. Bisseling, R. Kosloff, R.B. Gerber, M.A. Ratner, L. Gibson and C. Cerjan, *J. Chem. Phys.* **87**, 2760 (1987).
61. A. Garcia-Vela, *J. Chem. Phys.* **104**, 1047 (1996).
62. R. Alimi, R.B. Gerber, A.D. Hammerich, R. Kosloff and M.A. Ratner, *J. Chem. Phys.* **93**, 6484 (1990).
63. A. Garcia-Vela and R.B. Gerber, *J. Chem. Phys.* **103**, 3463 (1995).
64. A.B. McCoy, R.B. Gerber and M.A. Ratner, *J. Chem. Phys.* **101**, 1975 (1994).
65. A.B. McCoy, *Mol. Phys.* **85**, 965 (1995).
66. Z. Li and R.B. Gerber, *Chem. Phys. Lett.* **104**, 5803 (1996).
67. J.O. Jung and R.B. Gerber, *J. Chem. Phys.* (in press).
68. A.Y. Rom and R.B. Gerber, *J. Chem. Phys.* (submitted).
69. A. Jäckle and H.-D. Meyer, *J. Chem. Phys.* **104**, 7974 (1996).
70. A. Jäckle and H.-D. Meyer, *J. Chem. Phys.* **102**, 5605 (1995).
71. G.D. Billing and G. Jolicard, *Chem. Phys. Lett.* **221**, 75 (1994).

72. J.R. Waldeck, J. Campos-Martinez and R.D. Coalson, *J. Chem. Phys.* **94**, 2773 (1991).
73. N.P. Blake and H. Metiu, in *Femtosecond Chemistry*, ed. J. Manz and L. Wöste (VCH, Weinheim, 1995) p. 532.
74. J. Kucar, H.-D. Meyer and L.S. Cederbaum, *Chem. Phys. Lett.* **140**, 525 (1987).
75. R.B. Gerber and A. Alimi, *Chem. Phys. Lett.* **184**, 69 (1991).
76. J. Campos-Martinez and R.D. Coalson, *J. Chem. Phys.* **99**, 9629 (1993).
77. E. Fredj, R.B. Gerber and M.A. Ratner (to be published).
78. E.J. Heller, *J. Chem. Phys.* **62**, 1544 (1975).
79. M.P. Allen and D.J. Tildesley, *Computer Simulation of Liquids*, (Clarendon, Oxford, 1987).
80. R. McWeeny, *Methods of Molecular Quantum Mechanics*, 2nd edition (Academic Press, San Diego, 1992).
81. A.I. Krylov, R.B. Gerber, M.A. Gaveau, J.M. Mestdagh, B. Schilling and J.P. Visticot, *J. Chem. Phys.* **104**, 3651 (1996).
82. J.F. Grzykowski and L. Andrews, *J. Raman Spectros.* **4**, 99 (1975).
83. P. Jungwirth and B. Schmidt, *Chem. Phys. Lett.*, 275 (1997) 127.
83. R. B. Gerber, P. Jungwirth, E. Fredj, and A. Y. Rom *Modern methods for multidimensional dynamics computations in chemistry*, ed. D. L. Thompson (World Scientific, River Edge, NJ, 1998).

Approximation Properties and Limits of the Quantum-Classical Molecular Dynamics Model

Christof Schütte[1,2] and Folkmar A. Bornemann[1]

[1] Konrad-Zuse-Zentrum, Takustr. 7, 14195 Berlin, Germany
[2] Freie Universität Berlin, Fachbereich Mathematik, Arnimallee 2–6, 14195 Berlin, Germany

Abstract. In molecular dynamics applications there is a growing interest in including quantum effects for simulations of larger molecules. This paper is concerned with *mixed quantum-classical* models which are currently discussed: the so-called QCMD model with variants and the time-dependent Born-Oppenheimer approximation. All these models are known to approximate the full quantum dynamical evolution—under different assumptions, however. We review the meaning of these assumptions and the scope of the approximation. In particular, we characterize those typical problematic situations where a mixed model might largely deviate from the full quantum evolution. One such situation of specific interest, a non-adiabatic excitation at certain energy level crossings, can promisingly be dealt with by a modification of the QCMD model that we suggest.

1 Introduction

In molecular dynamics applications there is a growing interest in *mixed quantum-classical* models various kinds of which have been proposed in the current literature. We will concentrate on two of these models: the *adiabatic* or time-dependent Born-Oppenheimer (BO) model, [8, 13], and the so-called QCMD model.[1] Both models describe most atoms of the molecular system by the means of classical mechanics but an important, small portion of the system by the means of a wavefunction. In the BO model this wavefunction is *adiabatically* coupled to the classical motion while the QCMD model consists of a *singularly perturbed* Schrödinger equation nonlinearly coupled to classical Newtonian equations, §2.2.

This paper is meant as a contribution to *systematize* the quantum-classical modeling of molecular dynamics. Hence, we are interested in an extended theoretical understanding of the models rather than to further contribute to the bunch of numerical experiments which have been performed on certain models by applying them to particular molecular systems. Thus, we will carefully review the assumptions under which our models are known to approximate the full quantum dynamical (QD) evolution of the system. This knowledge

[1] The number of articles applying this model is so large that we only mention four articles, [2][3][9][17], as the starting points to different lines of discussion.

allows for a characterization of the typical problematic situations where the mixed models might largely deviate from the QD evolution.

The present paper is organized as follows: In a first step, the derivation of QCMD and related models is reviewed in the framework of the semiclassical approach, §2. This approach, however, does not reveal the close connection between the QCMD and BO models. For establishing this connection, the BO model is shown to be the adiabatic limit of both, QD and QCMD, §3. Since the BO model is well-known to fail at *energy level crossings*, we have to discuss the influence of such crossings on QCMD-like models, too. This is done by the means of a relatively simple test system for a specific type of such a crossing where non-adiabatic excitations take place, §4. Here, all models so far discussed fail. Finally, we suggest a modification of the QCMD system to overcome this failure.

To simplify we restrict our study to the case of a system with just two "particles" of significantly different masses, m and M, having coordinates $x \in \mathbb{R}^m$ and $q \in \mathbb{R}^d$. Thus, the time-dependent Schrödinger equation becomes

$$i\hbar\,\partial_t\Psi \;=\; \left(-\tfrac{\hbar^2}{2M}\mathcal{T}_q \;-\; \tfrac{\hbar^2}{2m}\mathcal{T}_x \;+\; V(x,q)\right)\,\Psi.$$

Here, the kinetic operators are typically given by the corresponding Laplacians $\mathcal{T}_q = \Delta_q$ and $\mathcal{T}_x = \Delta_x$ or similar selfadjoint differential operators. The corresponding solution $\Psi = \Psi(x,q,t)$ describes what we call the full QD evolution of the system.

By assumption, the mass ratio $\epsilon^2 = m/M$ is a small parameter. Thus, rescaling the Schrödinger equation properly in time and potential transforms[2] it into the singularly perturbed equation

$$i\epsilon\,\partial_t\Psi \;=\; \left(-\tfrac{\epsilon^2}{2}\mathcal{T}_q \;-\; \tfrac{1}{2}\mathcal{T}_x \;+\; V(x,q)\right)\,\Psi. \tag{1}$$

In many applications, x and q will not necessarily be coordinates of "particles" but other degrees of freedom of the system under consideration. Typically however, a proper choice of the coordinate system allows the initial quantum state to be approximated by a product state (cf. [11], §IIb):

$$\Psi(x,q,t=0) \;=\; \phi_*(q)\,\cdot\,\psi_*(x). \tag{2}$$

We will throughout assume this initial condition to be given.

2 Semiclassical Approach to QCMD

The semiclassical approach to QCMD, as introduced in [10], derives the QCMD equations within two steps. First, a *separation* step makes a tensor ansatz for the full wavefunction separating the coordinates x and q:

$$\Psi(x,q,t) \;\approx\; \Psi_\otimes \;=\; \phi(q,t)\,\cdot\,\psi(x,t). \tag{3}$$

[2] Time is scaled according to $\hbar t/\sqrt{mM} \to t$, implying a new potential $(m/\hbar^2)V$.

Second, a *semiclassical*, or *WKB*, ansatz approximates the "classical" wavefunction ϕ by

$$\phi(q,t) \approx \phi_{\mathrm{QC}} = a(q,t) \exp\left(\tfrac{i}{\epsilon} S(q,t)\right). \tag{4}$$

We will study the equations of motion that result from inserting all this in the full Schrödinger equation, Eq. (1). However, we would like to remind the reader that *not* the derivation of these equations of motion is the main topic here but the question of the quality of the underlying approximations.

2.1 Separation and TDSCF

Inserting the separation ansatz, i.e., $\Psi_{\otimes}$, results in two nonlinearly coupled single particle Schrödinger equations, the so-called *time dependent self-consistent field* (TDSCF) equations:[3]

$$i\epsilon\partial_t\phi = \left(-\tfrac{\epsilon^2}{2}\mathcal{T}_q + \langle\psi, V\psi\rangle\right)\phi, \qquad i\epsilon\partial_t\psi = \left(-\tfrac{1}{2}\mathcal{T}_x + \langle\phi, V\phi\rangle\right)\psi. \tag{5}$$

Here, $\langle\psi, V\psi\rangle = U_\phi$ denotes the ψ-averaged potential as seen by ϕ, still depending on the coordinate q. Likewise, $U_\psi = \langle\phi, V\phi\rangle$ includes integration with respect to q and depends on x. In the following, $\langle\cdot,\cdot\rangle$ will similarly denote integration with respect to x, q, or x and q, yielding expressions that depend on the other coordinate.

Approximation Property We assume that the "classical" wavefunction ϕ is an approximate δ-function, i.e., for all times $t \in [0,T]$ the probability density $|\phi(t)|^2 = |\phi(q,t)|^2$ is concentrated near a location $q(t)$ with "width," i.e., position uncertainty, $\delta(t)$. Then, the quality of the TDSCF approximation can be characterized as follows:

Theorem 1 (Thm. 4.1. in [6]). *For all $t \in [0,T]$, let ϕ have compact support[4] of width $\delta(t) < \delta$. Then, the TDSCF wavefunction $\Psi_{\otimes}$ approximates the full QD solution Ψ of Eq. (1) up to an error of order δ, i.e.,*

$$\Psi_{\otimes} = \Psi + \mathcal{O}(\delta) \quad \text{in} \quad [0,T].$$

Thus, TDSCF is the better an approximation of full QD the sharper located the probability density $|\phi|^2$ remains in the course of the evolution.

[3] More precisely, Eq. (5) is only valid up to additional phase terms, cf. [6], §IVa, or [11], §IIIa, for details.

[4] In this case, let δ be the diameter of the support: $\delta(t) = \operatorname{diam}\operatorname{supp}|\phi(t)|^2$.

2.2 Semiclassical Ansatz and QCMD

Inserting Ψ_{QC} into Eq. (1), or equivalently the WKB-ansatz for ϕ into the TDSCF system Eq. (5), results in equations of motion for a and S (for details cf. [6, 10]) and an one-particle Schrödinger equation,

$$i\epsilon\partial_t\psi_{\mathrm{ens}} = \left(-\tfrac{1}{2}\mathcal{T}_x + \int V(x,q)\,a^2(q,t)\,dq\right)\psi_{\mathrm{ens}}. \tag{6}$$

Notice that the solution ψ_{ens} is not identical to ψ but an approximation of it. The evolution of a and S in time may conveniently be described via the following classical Newtonian equations of motion: Given the initial values

$$a(q,0) = a_*(q) \qquad \text{and} \qquad S(q,0) = S_*(q), \tag{7}$$

we denote by $q(t) = q(t;q_0,\dot{q}_0)$ and $\dot{q}(t) = \dot{q}(t;q_0,\dot{q}_0)$ the solutions of the initial value problem

$$\ddot{q} = -\mathrm{grad}_q\,\langle\psi_{\mathrm{ens}}, V\psi_{\mathrm{ens}}\rangle(q), \qquad q(0) = q_0,\ \dot{q}(0) = \dot{q}_0 = \nabla S_*(q_0). \tag{8}$$

The *probability density* a^2 at a point $q = q(t) = q(t;q_0,\dot{q}_0)$ is obtained by transport of the initial probability, i.e.,

$$a^2(q(t;q_0,\dot{q}_0),t) = a_*^2(q_0,\dot{q}_0)\ J, \tag{9}$$

with J denoting the Jacobian of $q = q(t;q_0,\dot{q}_0)$ with respect to q_0. In addition, the *action* or *phase* S at $q = q(t)$ is given by integrating the corresponding Lagrangian along this trajectory, [1]:

$$S(q(t;q_0,\dot{q}_0),t) = S_*(q_0) + \int_0^t \left(\tfrac{1}{2}\dot{q}(s) - \langle\psi_{\mathrm{ens}}, V\psi_{\mathrm{ens}}\rangle(q(s))\right)ds. \tag{10}$$

Since Eq. (6) depends on the probability density $|\phi(q,t)|^2 = a^2(q,t)$ only, we may put the solution for S aside. Thus, we get a system that couples the classical equation Eq. (8) for computing a^2 to the one-particle Schrödinger equation, Eq. (6). A numerical simulation of the evolution as described by Eqs. (6), (8), and (9) has to compute a *bundle* of classical trajectories that sample the probability distribution a^2 and are nonlinearly coupled via Eq. (6).

We assume now, that the initial probability distribution $|\phi|^2\big|_{t=0} = a_*^2$ is an approximate δ-function at q_0. In this case, Eq. (9) makes it obvious, a^2 *remains* to be an approximate δ-function as long as the approximation is valid. Thus, the single trajectory $q(t) = q(t;q_0,\dot{q}_0)$ is an appropriate sampling of the probability density and $a^2(q,t) = \delta(q - q(t))$ simplifies the integral in Eq. (6) so that the final *QCMD equations* of motion read

$$\begin{aligned} i\epsilon\partial_t\psi_{\mathrm{QC}} &= \left(-\tfrac{1}{2}\mathcal{T}_x + V(x,q(t))\right)\psi_{\mathrm{QC}}, \\ \ddot{q} &= -\mathrm{grad}_q\langle\psi_{\mathrm{QC}}, V\psi_{\mathrm{QC}}\rangle(q). \end{aligned} \tag{11}$$

Caustics The above formulae can only be valid as long as Eq. (9) describes a unique map in position space. Indeed, the underlying Hamilton-Jacobi theory is only valid for the time interval [0,T] if at all instances $t \in [0,T]$ the map $(q_0, \dot{q}_0) \to q(t; q_0, \dot{q}_0)$ is one-to-one, [6, 19, 1], i.e., as long as trajectories with different initial data do *not* cross each other in position space (cf. Fig. 1). Consequently, the *detection* of any caustics in a numerical simulation is only possible if we propagate a trajectory bundle with different initial values. Thus, in pure QCMD, Eq. (11), caustics *cannot* be detected.

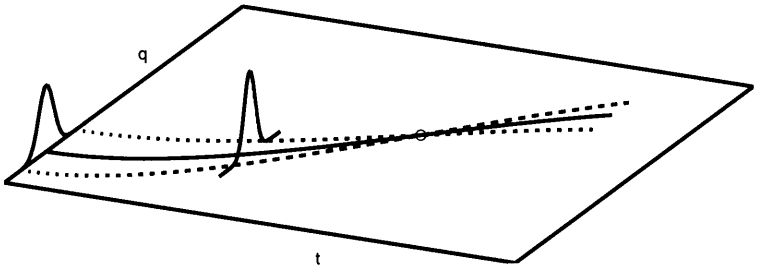


Fig. 1. Illustration of a caustic. Different trajectories sample the probability distribution. If they cross each other in position space, the transport or probability density is not longer unique and the approximation might break down.

Approximation Property Excluding caustics we can exploit the results of semiclassical approximation theory [19]. This leads to the following statement:

Theorem 2 (Thm. 4.2. in [6]). *Let ϕ initially*[5] *have width $\delta(0) < \delta$ and let ϵ be small enough. Moreover, assume that caustics do not appear in time interval $[0,T]$. Then, the semiclassical wavefunctions ψ_{ens} and ψ_{QC} approximate the TDSCF wavefunction ψ up to an error of order $\delta^2 + \epsilon$, i.e.,*

$$\psi_{\text{ens}} = \psi + \mathcal{O}(\delta^2 + \epsilon) \quad \text{and} \quad \psi_{\text{QC}} = \psi + \mathcal{O}(\delta^2 + \epsilon) \qquad \text{in } [0,T].$$

The QCMD solution q approximates expectation value of the classical position, $\langle q \rangle_{\text{QD}} = \langle \Psi_{\text{QD}}, q\Psi_{\text{QD}} \rangle$, of the full QD solution Ψ_{QD} as:

$$q(t) = \langle q \rangle_{\text{QD}} + \mathcal{O}(\delta^2 + \epsilon) \qquad \text{in } [0,T].$$

Referring to Thm. 1 we can conclude that—excluding caustics—QCMD (and QCMD bundles) approximates full QD up to an error of order $\mathcal{O}(\delta+\epsilon)$. These

[5] Because of Eq, (9), the condition from Thm. 1 concerning the small width can herein be restricted to the initial condition.

approximation result extends to cases in which certain types of caustics (focal points) are present by including phase shifts; cf. [19]. However, this cannot fully explain the bunch of numerical observations in which the presence of caustics does *not* influence the quality of the approximation at all. Thus, we might be interested in a justification of QCMD which avoids the problem of caustics. We will achieve this via studying the adiabatic limit of QCMD in §3.

2.3 Density Formulation of Semiclassical QCMD Bundles

A particularly convenient notation for trajectory bundle system can be introduced by using the classical Liouville equation which describes an ensemble of Hamiltonian trajectories by a phase space density $f = f(q, \dot{q}, t)$. In textbooks of classical mechanics, e.g. [12], it is shown that Liouville's equation

$$\partial_t f = \mathcal{L}[V]f, \qquad \mathcal{L}[V] = \left(\nabla_q V(q)^T \cdot \nabla_{\dot{q}} + \dot{q}^T \cdot \nabla_q\right), \tag{12}$$

describes the transport of an initial probability density $f(q, \dot{q}, 0) = f_*$ along the trajectories of the classical equation of motion $\ddot{q} = -\mathrm{grad}_q V$ in the sense that

$$f(q(t), \dot{q}(t), t) = f_*(q_0, \dot{q}_0). \tag{13}$$

Here, we denote by $(q(t), \dot{q}(t))$ the trajectory starting at $(q_0, \dot{q}_0)$. Thus, the transport of the semiclassical probability density a^2 according to Eq. (9) is just given by the Liouville equation with the potential $\langle \psi_{\mathrm{ens}}, V\psi_{\mathrm{ens}} \rangle$ and $a^2(q,t) = \int f(q, \dot{q}, t) d\dot{q}$:

$$\begin{aligned} i\epsilon\partial_t \psi_{\mathrm{ens}} &= \left(-\tfrac{1}{2}\mathcal{T}_x + \int f(q, \dot{q}, t) V(x, q)\, dq\, d\dot{q}\right) \psi_{\mathrm{ens}}, \\ \partial_t f &= \mathcal{L}[\langle \psi_{\mathrm{ens}}, V\psi_{\mathrm{ens}} \rangle]\, f. \end{aligned} \tag{14}$$

We will refer to this model as to the *semiclassical QCMD bundle*. Eqs. (7) and (8) would suggest certain initial conditions for f_*. However, those would *not* include any momentum uncertainty, resulting in a wrong disintegration of the probability distribution in q as compared to the full QD. For including an initial momentum uncertainty, a Gaussian distribution in position space is used

$$f(q, \dot{q}, 0) = f_*(q, \dot{q}) = \tfrac{1}{\alpha} \exp\left(-\tfrac{1}{\delta^2}(q - q_*)^2\right) \exp(-\tfrac{2\delta^2}{\epsilon^2}(\dot{q} - \dot{q}_*)^2), \tag{15}$$

with a normalizing constant α and an initial momentum expectation $\dot{q}_*$.

3 Adiabatic Limit Approach to QCMD

If the parameter ϵ is very small, we are in the case of M being much larger than m. Thus, the limit $\epsilon \to 0$ is the limit of infinite mass M, i.e., the *adiabatic*

limit of fast quantum motion of small particles around (infinitely) slowly changing positions of increasingly heavy "nuclei." We will study the *limit equations* governing the QCMD solutions for this adiabatic limit. Therefore, we rewrite the QCMD system, Eq. (11), by explicitly denoting the dependence of its solution $(q_\epsilon, \dot{q}_\epsilon, \psi_\epsilon)$ on the parameter ϵ:[6]

$$\begin{aligned} \ddot{q}_\epsilon &= -\mathrm{grad}_q \langle \psi_\epsilon, H(q_\epsilon)\psi_\epsilon \rangle, & q_\epsilon(0) = q_*,\ \dot{q}_\epsilon(0) = \dot{q}_* \\ i\epsilon \partial_t \psi_\epsilon &= H(q_\epsilon)\,\psi_\epsilon, & \psi_\epsilon|_{t=0} = \psi_*. \end{aligned} \tag{16}$$

where $H = H(q)$ is the q-parametrized one-particle Hamiltonian

$$H(q) \;=\; -\tfrac{1}{2}\mathcal{T}_x \;+\; V(x,q). \tag{17}$$

We restrict ourselves to finite-dimensional Hilbert spaces,[7] making H a Hermitian matrix. We denote the eigenvalues of $H(q)$ by $E_k(q)$ and consider the spectral decomposition

$$H(q) \;=\; \textstyle\sum_k E_k(q)\, P_k(q), \tag{18}$$

where P_k is the orthogonal projection onto the eigenspace associated with E_k. With respect to a quantum state ψ, the number $\theta_k = \langle \psi, P_k \psi \rangle$ is the *population* of the *energy level* E_k.[8] The surfaces $E_k = E_k(q)$ are called *energy levels.* Those positions q_c at which *energy level crossings* occur, i.e.,

$$E_k(q_c) \;=\; E_l(q_c) \quad \text{for some } k \neq l,$$

will be the points of special interest in this section.

3.1 Adiabatic Limit of QCMD

The limit equation governing $\lim_{\epsilon\to 0} q_\epsilon$ can be motivated by referring to the *quantum adiabatic theorem* which originates from work of Born and Fock [4, 20]: The classical position q influences the Hamiltonian very slowly compared to the time scale of oscillations of ψ_ϵ, in fact, "infinitely slowly" in the limit $\epsilon \to 0$. Thus, in analogy to the quantum adiabatic theorem, one would expect that the population of the energy levels remain *invariant* during the evolution:

$$\lim_{\epsilon\to 0} \theta_k^\epsilon(t) = \lim_{\epsilon\to 0} \langle \psi_\epsilon, P_k(q_\epsilon)\psi_\epsilon \rangle \;=\; \theta_k^0 = \langle \psi_*, P_k(q_*)\psi_* \rangle. \tag{19}$$

The *constant* θ_k^0 is the initial population of level E_k and thus computable from the initial data, Eq. (16). All this turns out to be true if the following assumption on the eigenspaces and eigenenergies of $H(q)$ is fulfilled:

[6] We will often add an index ϵ in order to refer to a family of solutions.

[7] The reader may think of a finite dimensional subspace of the original state space. This subspace may, e.g., be associated with a suitable discretization in space. For a generalization of Thm. 3 to the infinitely dimensional case, see [5].

[8] If the eigenspace to E_k is one-dimensional and Φ_k is a corresponding normalized eigenvector, then we have $P_k = \Phi_k^* \otimes \Phi_k$ and the population is $\theta_k = |\langle \Phi_k, \psi \rangle|^2$.

(A) The spectral decomposition Eq. (18) of H depends smoothly on q.

This assumption allows to prove that the limit solution $q_{\mathrm{BO}} = \lim_{\epsilon \to 0} q_\epsilon$ is given by:

$$\ddot{q}_{\mathrm{BO}} = -\mathrm{grad}_q \sum_k \theta_k^0 E_k(q_{\mathrm{BO}}), \qquad q_{\mathrm{BO}}(0) = q_*, \; \dot{q}_{\mathrm{BO}}(0) = \dot{q}_*. \tag{20}$$

We refer to this equation as to the *time-dependent Born-Oppenheimer* (BO) model of adiabatic motion. Notice that Assumption (A) does *not* exclude energy level crossings along the limit solution q_{BO}. Using a density matrix formulation of QCMD and the technique of weak convergence one can prove the following theorem about the connection between the QCMD and the BO model:

Theorem 3 (Thm. III.1 in [5], Thm. 2.1 in [7]). *Let $q_{\mathrm{BO}} = q_{\mathrm{BO}}(t)$ be the solution of the BO equation, Eq. (20), and assumption (A) be given. Any energy level crossing at $q_{\mathrm{BO}}(t_c)$ with $t_c \in [0, T]$ fulfills the transversality condition*

$$\tfrac{d}{dt}(E_k(q_{\mathrm{BO}}(t)) - E_l(q_{\mathrm{BO}}(t)))\big|_{t=t_c} \neq 0.$$

Then, the adiabatic invariance Eq. (19) holds and the limit of the sequence q_ϵ of QCMD solutions is q_{BO}.

3.2 Adiabatic Limit of QD

Thus, the time-dependent BO model describes the adiabatic limit of QCMD. If QCMD is a valid approximation of full QD for sufficiently small ϵ, the BO model has to be the adiabatic limit of QD itself. Exactly this question has been addressed in different mathematical approaches, [8], [13], and [18]. We will follow HAGEDORN [13] whose results are based on the product state assumption Eq. (2) for the initial state with a special choice concerning the dependence of ϕ_* on ϵ:

$$\phi_*(q) = \tfrac{1}{A_\epsilon} \exp\left(-\tfrac{1}{4\epsilon}(q - q_*)^2\right) \exp\left(\tfrac{i}{\epsilon}\dot{q}_* q\right), \tag{21}$$

with the initial momentum expectation $\dot{q}_*$ and a normalization constant A_ϵ. This scaling guarantees that the wavefunction ϕ behaves uniformly classically.[9] Using this initial condition and the BO solution q_{BO} a wavefunction Ψ_{BO} is constructed which approximates the full QD solution Ψ_ϵ up to an error $\mathcal{O}(\sqrt{\epsilon})$, [13].[10] For simplicity, let us discuss the position expectation

$$\langle q \rangle_\epsilon^{\mathrm{QD}} = \langle \Psi_\epsilon, q\Psi_\epsilon \rangle(t).$$

[9] Let be $V = 0$. Then, ϕ describes a free particle. With Eq. (21), the disintegration of the wavepaket makes its width increase like $\sqrt{\epsilon}\sqrt{1 + t^2/4}$ in the limit $\epsilon \to 0$. Thus, the velocity of its disintegration is classical and independent from ϵ.

[10] Ψ_ϵ is the family of solutions of Eq. (1) with initial states due to Eq. (2) and Eq. (21). The initial quantum state ψ_* is assumed to be independent from ϵ with only finitely many energy levels E_k, $k = 1, \ldots, n$ being initially excited.

instead of the wavefunctions. Here, the statement of HAGEDORN is:

Theorem 4 (Thm. 2.1 in [13]). *Assume $q_{\mathrm{BO}} = q_{\mathrm{BO}}(t)$ to be the solution of the BO equation, Eq. (20), in a finite time interval $[0,T]$. Moreover, let there be no energy level crossings along q_{BO}. Then, for ϵ small enough, we have*

$$\langle q \rangle_{\epsilon}^{\mathrm{QD}} = q_{\mathrm{BO}} + \mathcal{O}(\sqrt{\epsilon}) \quad \textit{in } [0,T].$$

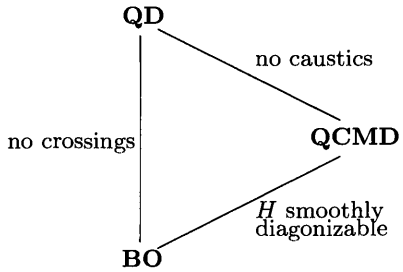


Fig. 2. The BO model is the adiabatic limit of full QD if energy level crossings do not appear. QCMD is connected to QD by the semiclassical approach if no caustics are present. Its adiabatic limit is again the BO solution, this time if the Hamiltonian H is smoothly diagonalizable. Thus, QCMD may be justified indirectly by the adiabatic limit excluding energy level crossings and other discontinuities of the spectral decomposition.

Altogether, the three different models discussed so far are interconnected as sketched in Fig. 2. Now, we can by-pass the problems connected to caustics: For ϵ being small enough QCMD is justified as an approximation of QD if we exclude energy level crossings and discontinuities of the spectral decomposition.

However, there remains one major question:

Can QCMD describe *non-adiabatic processes*; is there any situation in which BO fails but QCMD or its bundle variants are still useful?

By what we have seen before such a situation can only occur if there is an energy level crossing where Assumption (A) of Thm. 3 is hurt. In the next section, we will present a test example of this situation.

3.3 Energy Level Crossings with Non-Adiabatic Excitations

In his book [16], HAGEDORN classifies all energy level crossings that can occur generically with an electronic Hamiltonian according to the associated symmetries. Each symmetry yields a finite number of typical, generic energy level

crossings. These generic situations are mathematically described by a *normal form* which reduces the general problem Eq. (1) to a simple low dimensional test problem. For time-reversible Hamiltonians there is just one normal form of an energy level crossing which moreover hurts the Assumption (A).

This normal form reduces the Schödinger equation, Eq. (1), to a specific form where $q \in \mathbb{R}^2$ remains a particle's position but x becomes a spin-like coordinate:

$$i\epsilon\,\dot{\Psi} = \left(-\tfrac{\epsilon^2}{2}\mathcal{T}_q + H(q)\right)\Psi. \tag{22}$$

Herein, $H = H(q)$ and $\mathcal{T}_q$ denote 2×2 Hermitian matrices, the entries of H being potential operators and $\mathcal{T}_q$ being diagonal

$$\mathcal{T}_q = \begin{pmatrix} \Delta_q & 0 \\ 0 & \Delta_q \end{pmatrix}. \tag{23}$$

Thus, $\Psi \in L^2(\mathbb{R}^2) \times L^2(\mathbb{R}^2)$ consists of two components $\Psi = (\Psi_1, \Psi_2)^T$, each of which a function in the usual Hilbert space. The Hamiltonian is specified by the particular matrix

$$H(q) = \begin{pmatrix} q^1 & q^2 \\ q^2 & -q^1 \end{pmatrix}. \tag{24}$$

The eigenvalues of H are $E_1(q) = -|q|$ and $E_2(q) = |q|$. Excluding the crossing at the origin $q = 0$ and using polar coordinates $q^1 = r\cos\varphi$ and $q^2 = r\sin\varphi$, yields the corresponding eigenvectors in the form

$$\Phi_1 = \begin{pmatrix} -\sin(\varphi/2) \\ \cos(\varphi/2) \end{pmatrix}, \qquad \Phi_2 = \begin{pmatrix} \cos(\varphi/2) \\ \sin(\varphi/2) \end{pmatrix}.$$

The occurrence of the argument $\varphi/2$ shows that these eigenvectors are defined up to a sign only. For a unique representation we have to cut the plane along a half-axis. By this, Φ_1 and Φ_2 become smooth vector fields uniquely defined on the cut plane. They cannot, however, be continued over the cut, but change their roles there instead. Thus, we have the situation of a crossing at which the eigenvector field is discontinuous and Assumption (A) of Thm. 3 is hurt.

In the pure BO model, this discontinuity will be ignored. Let the initial values be given by

$$q_* = (q_*^1, 0), \qquad \dot{q}_* = (\dot{q}_*^1, 0), \qquad \psi_* = (1, 0)^T, \tag{25}$$

with $q_*^1 > 0$ and $\dot{q}_*^1 < 0$ so that the initial motion is towards the crossing. In this case, the pure BO equations read $\ddot{q}^1 = -1$ and $\ddot{q}^2 = 0$, i.e., the solution is

$$q_{\mathrm{BO}}^1(t) = -\tfrac{1}{2}t^2 + \dot{q}_*^1\,t + q_*^1, \qquad q_{\mathrm{BO}}^2(t) = 0,$$

moving through the crossing at $t_c = \sqrt{(\dot{q}^1_*)^2 + 2q^1_*} + \dot{q}^1_*$.

As long as we have not passed the crossing, i.e., for $t < t_c$, Thm. 4 describes the limit $\epsilon \to 0$. Thus, the populations will be constant in $[0, t_c)$ in the limit $\epsilon \to 0$: $\theta_1 = 1, \theta_2 = 0$. The crossing itself induces a true excitation of the second energy level, [14, 16]:[11]

$$\theta^\epsilon_2(t) = \theta^+_2 + o(1), \text{ for } t = t_c + \delta, \text{ with } \theta^+_2 = \left[1 + \frac{\pi}{\dot{q}^1_{\mathrm{BO}}(t_c)}\right]^{-1}. \tag{26}$$

Thus, passing the crossing induces a deeply *non-adiabatic* process. Directly behind the crossing Thm. 4 applies again, so that the information concerning the redistribution of population at the crossing is sufficient to denote the limit solution q_{Ha} for $\epsilon \to 0$: While the second component remains zero ($q^2_{\mathrm{Ha}} \equiv 0$) we now have

$$q^1_{\mathrm{Ha}}(t) = \begin{cases} q^1_{\mathrm{BO}}(t) & : \ t < t_c \\ (2\theta^+_2 - 1)(t - t_c)^2 + \dot{q}^1_{\mathrm{BO}}(t_c)(t - t_c) & : \ t \geq t_c \end{cases}. \tag{27}$$

With initial conditions Eq. (25), the QCMD solution can be determined explicitly. Surprisingly, there is for all ϵ:

$$q_\epsilon \equiv q_{\mathrm{BO}}, \quad \psi^1_\epsilon(t) = \exp\left(-\tfrac{i}{\epsilon}\int_0^t q^1_\epsilon(\tau)\, d\tau\right), \quad \psi^2_\epsilon \equiv 0. \tag{28}$$

Thus, neither BO nor QCMD can describe the non-adiabatic excitation at the crossing. However, as studied in [7], there is yet another feature of the QCMD model that could turn out to be useful here and might help to include the non-adiabatic process. After the crossing the adiabatic limit of QCMD is, in a sense, *not uniquely determined*:

Theorem 5 (§4 in [7]). *Let $q_{\mu,\epsilon}$ be the QCMD solution to the initial conditions*

$$q_* = (q^1_*, 0), \qquad \dot{q}_* = (\dot{q}^1_*, \mu), \qquad \psi_* = (1, 0)^T$$

with $\mu > 0$. Then, the limit process $\epsilon, \mu \to 0$ is not unique, specifically

$\lim_{\epsilon \to 0} q_{\mu=0,\epsilon} \equiv q_{\mathrm{BO}}$ *and* $\lim_{\mu \downarrow 0} \lim_{\epsilon \to 0} q_{\mu,\epsilon}(t) = q_{\mathrm{BO}}(t) + (t^2, 0) = \tilde{q}_{\mathrm{BO}}(t)$.

Actually all points $\bar{q}$ between the two curves q^1_{BO} and $\tilde{q}^1_{\mathrm{BO}}$ can be obtained as a limit solutions belonging to a particular pair of sequences $\epsilon, \mu \to 0$.

In a way, the limit set is thus the entire *funnel* between the two extreme cases q^1_{BO} and $\tilde{q}^1_{\mathrm{BO}}$, Fig. 5. This effect is called *Takens-chaos*, [21, 5, 7]. As a consequence of this theorem each momentum uncertainty effects a kind of "disintegration" process at the crossing. Thus, one can reasonably expect to reproduce the true excitation process by using QCMD trajectory bundles for "sampling the funnel." To realize this idea, we have to study the full quantum solution and compare it to suitable QCMD trajectory bundles.

[11] For the connection of this result to the well-known Landau-Zener formula [23] see [15].

4 Energy Level Crossings and QCMD Bundles

To illustrate the effect of the crossing on a QD solution Ψ_ϵ, Fig. 3 shows the projection of the probability density $|\Psi_\epsilon|^2$ onto the q^1-plane for $\epsilon = 1/100$. We observe that the density disintegrates after passing the crossing. Its two main arms propagate along the curves q_{BO} and $\tilde{q}_{\mathrm{BO}}$ described above, i.e., the funnel of Thm. 5 reappears in the QD solution, however, this time, with an internal statistical structure. Fig. 4 shows the corresponding picture for a suitable QCMD trajectory bundle computation that clearly reflects the properties of the full QD solution. The following paragraphs explain how this QCMD bundle has been constructed.

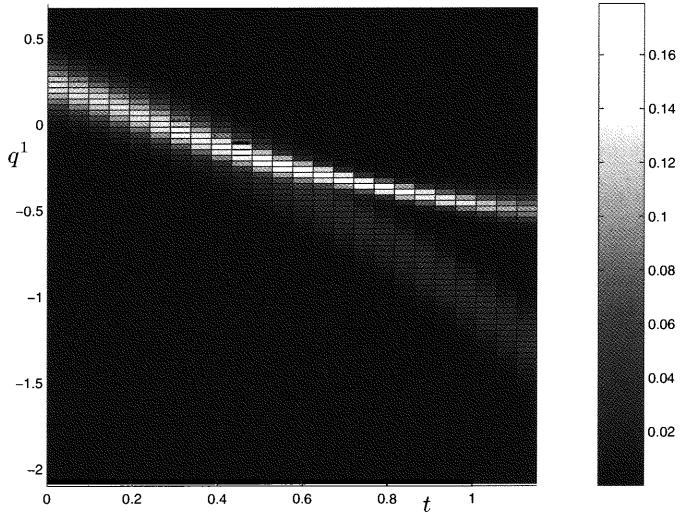


Fig. 3. Quantum solution Ψ_ϵ of the test system of §3.3 for $\epsilon = 1/100$. Ψ_ϵ computed numerically using Fourier pseudospectral methods in space and a symplectic discretization in time. Reduced q^1-density $\int |\Psi_\epsilon(q^1, q^2, t)|^2 dq^2$ versus t and q^1. Initial data due to (21) and (25) with $q^1_* = 0.2$ and $\dot{q}^1_* = -0.8$.

4.1 Different Trajectory Bundles

Unfortunately, the semiclassical QCMD bundles, Eq. (14) in §2.3, are only of limited use here. To understand this, let us consider the q-expectation $[q]$ of its solution (ψ_{ens}, f). Recall that the expectation of a classical observable $A = A(q, \dot{q})$ with respect to the phase-space density f is given by

$$[A](t) \;=\; \int A(q, \dot{q})\, f(q, \dot{q}, t)\, d\dot{q}\, dq.$$

Integration of Eq. (14) directly yields

$$\tfrac{d^2}{dt^2}[q] \;=\; -[\langle \psi_{\mathrm{ens}}, \mathrm{grad}_q V(q) \psi_{\mathrm{ens}} \rangle].$$

Because the Hamiltonian $H = V$ depends *linearly* on q, the last expectation value is actually independent of f:

$$[\langle \psi_{\text{ens}}, \text{grad}_q V(q)\psi_{\text{ens}}\rangle] = \int \langle \psi_{\text{ens}}, \text{grad}_q V(q)\psi_{\text{ens}}\rangle \, f(q,\dot{q},t)\, dqd\dot{q}$$
$$= \underbrace{\langle \psi_{\text{ens}}, \text{grad}_q V(q)\psi_{\text{ens}}\rangle}_{\text{independent from } f} \cdot \underbrace{\int f(q,\dot{q},t) dqd\dot{q}}_{=1}.$$

Thus, $[q]$ and ψ_{ens} obey the following single trajectory QCMD system

$$i\epsilon\dot{\psi}_{\text{ens}} = V([q])\psi_{\text{ens}}, \qquad \frac{d^2}{dt^2}[q] = -\langle \psi_{\text{ens}}, \text{grad}_q V([q])\psi_{\text{ens}}\rangle.$$

Hence, we just have proven the following proposition:

Proposition 6. *Let (ψ_{ens}, f) be the solution of the bundle equation, Eq. (14), and $[q]$ be the corresponding q-expectation. If the potential V depends linearly on q, $[q]$ is identical with the solution q of the single trajectory QCMD model with initial values $q(0) = [q](0)$, $\dot{q}(0) = [\dot{q}](0)$, and $\psi(0) = \psi_{\text{ens}}(0)$.*

Since QCMD reproduces the BO solution, we again have $[q] = q_{\text{BO}}$ ignoring the non-adiabatic excitation process at the crossing. Consequently, we have to modify the very concept of QCMD bundles.

Remark: The statement of Prop. 6 is also valid for the q-expectation $\langle q\rangle = \langle \Psi_\otimes, q\Psi_\otimes\rangle$ of the TDSCF solution. Consequently, *TDSCF fails near the crossing*, a fact, which emphasizes that the reason for this failure is connected to the separation step.
Actually, Fig. 4 has been obtained using the following modification of the QCMD bundle: one propagates an *ensemble* of *independent*, single QCMD trajectories $(q_k, \dot{q}_k, \psi_k)$, $k = 1, \ldots, N$:

$$i\epsilon\partial_t\psi_k = \left(-\tfrac{1}{2}\mathcal{T}_x + V(x, q_k(t))\right)\psi_k, \qquad \ddot{q}_k = -\text{grad}_q\langle \psi_k, V\psi_k\rangle(q_k). \quad (29)$$

Initially, all the ψ_k are identical: $\psi_k(t=0) = \psi_*$ and the classical states $(q_k, \dot{q}_k)$ sample the density $|\phi_*(t=0)|^2$ according Eq. (15),[12] i.e., there is a weight factor w_k for each trajectory k. Consequently, for each time t the probability distribution $p(q^1)$ in q^1 can be approximated on any sufficiently large interval $[q^1, q^1 + \Delta q^1]$ by adding the weights of all trajectories passing this interval at time t:

$$p(t)|_{[q^1, q^1+\Delta q^1]} = \sum_{q_k^1(t)\in[q^1,q^1+\Delta q^1]} w_k. \quad (30)$$

A comparison of Fig. 4 and Fig. 3 shows that this *uncoupled QCMD bundle* reproduces the disintegration of the full QD solution. However, there are minor quantitative differences of the statistical distribution. Fig. 5 depicts

[12] For Hagedorn's initial condition Eq. (21), we have to choose $\beta = 2/\epsilon$ in Eq. (15).

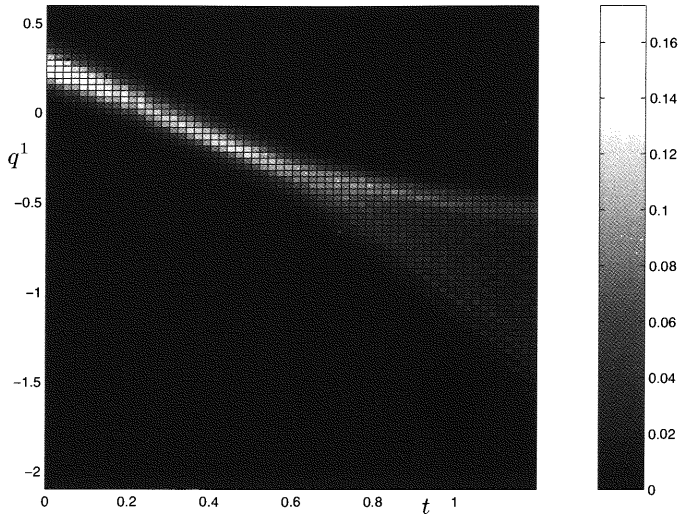


Fig. 4. q^1-density p, Eq. (30), of a simulation for $\epsilon = 1/100$ using the uncoupled QCMD bundle. Same situation as in Fig. 3.

the corresponding q^1-expectation values together with HAGEDORN's limit q^1-expectation q_{Ha} of QD for $\epsilon \to 0$. We observe that for $\epsilon = 1/100$ and $\epsilon = 1/500$ the q^1-expectation of the uncoupled QCMD bundle approximates the q^1-expectation of the corresponding QD solution which lie close to q_{Ha}.

Acknowledgments. Folkmar Bornemann greatfully acknowledges the hospitality of the Courant Institute of Mathematical Sciences, New York University, where he spent the academic year 96/97. His work was supported in part by the U.S. Department of Energy under contract DE-FG02-92ER25127.

References

1. Arnold, V. I.: Mathematical Methods of Classical Mechanics. 2nd edition. Springer Verlag, Berlin, Heidelberg, New York, Tokyo (1989)
2. P. Bala, P. Grochowski, B. Lesyng, and J. A. McCammon: Quantum-classical molecular dynamics. Models and applications. In: Quantum Mechanical Simulation Methods for Studying Biological Systems (M. Fields, ed.). Les Houches, France (1995)
3. Billing, G. D.: Quantum-Classical Methods. In: Numerical Grid Methods and Their Application to Schrödinger's equation (C. Cerjan, eds.). Kluwer Academics Publishers (1993)
4. Born, M., Fock, V.: Beweis des Adiabatensatzes. Z. Phys. **51** (1928) 165–180
5. Bornemann, F. A.: Homogenization in Time of Singularly Perturbed Conservative Mechanical Systems. Manuscript (1997) 146pp

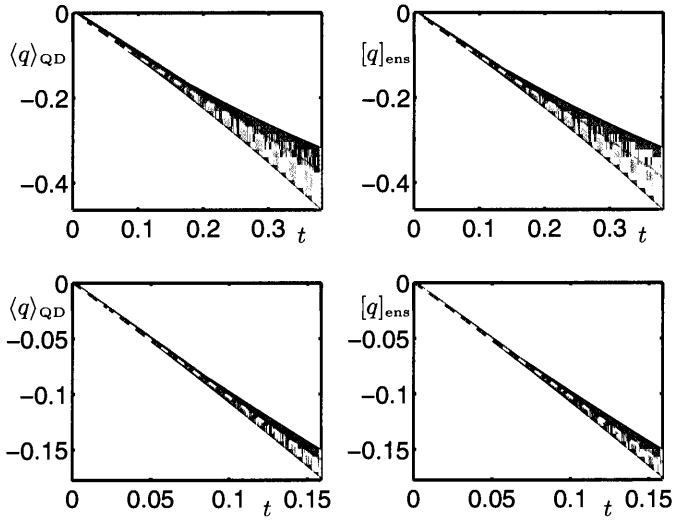


Fig. 5. Comparison of the q_1 expectation value of the uncoupled QCMD bundle ($[q]_{ens}$) and full QD ($\langle q \rangle_{QD}$) for the test system for $\epsilon = 1/100$ (pictures on top) and $\epsilon = 1/500$ (below). Initial data as in Fig. 3. The shaded domain indicates the funnel between the two curves q_{BO} and $\tilde{q}_{BO}$ (cf. Thm. 5). The light dashed line shows Hagedorn's limit solution q_{Ha} and the dense lines $\langle q \rangle_{QD}$ (left hand pictures) and $[q]_{ens}$ (right hand pictures).

6. Bornemann, F. A., Nettesheim, P., Schütte, Ch.: Quantum-Classical Molecular Dynamics as an Approximation to Full Quantum Dynamics. J. Chem. Phys., **105** (1996) 1074-1083
7. Bornemann, F. A., Schütte, Ch.: On the Singular Limit of the Quantum-Classical Molecular Dynamics Model. Preprint SC 97-07 (1997) Konrad-Zuse-Zentrum Berlin. SIAM J. Appl. Math. (submitted)
8. Combes, J. M.: The Born-Oppenheimer approximation. Acta Phys. Austriaca **17** (1977) Suppl. 139–159
9. Garcia-Vela, A., Gerber, R. B.: Hybrid quantum-semiclassical wave packet method for molecular dynamics: Application to photolysis of Ar...HCl. J. Chem. Phys. **98** (1993) 427–43
10. Gerber, R. B., Buch, and V., Ratner, M. A.: Time–dependent self–consistent field approximation for intramolecular energy transfer. J. Chem. Phys. **66** (1982) 3022–3030
11. Gerber, R. B., Ratner, M. A.: Self-consistent field methods for vibrational excitations in polyatomic systems. Adv. Chem. Phys. **70** (1988) 97–132
12. Goldstein, H.: Classical Mechanics. 2nd edition. Addison-Wesley, Reading, MA, (1980)
13. Hagedorn, G. A.: A time dependent Born-Oppenheimer approximation. Comm. Math. Phys. **77** (1980) 1–19
14. Hagedorn, G. A.: Electron energy level crossing in the time-dependent Born-Oppenheimer approximation. Theor. Chim. Acta **67** (1990) 163–190

15. Hagedorn, G. A.: Proof of the Landau-Zener formula in an adiabatic limit with small eigenvalue gaps. Commun. Math. Phys. **136** (1991) 433–449
16. Hagedorn, G. A.: Molecular propagation through electron energy level crossings. Mem. Amer. Math. Soc. **536** (1994) 130pp
17. Haug, K., Metiu, H.: A test of the possibility of calculating absorption spectra by mixed quantum-classical methods. J. Chem. Phys. **97** (1992) 4781–4791
18. Maslov, V.P.: The Complex WKB Method for Nonlinear Equations I. Birkhäuser Basel, Boston, Berlin (1994)
19. Maslov, V. P., Fedoriuk, M. V.: Semi-Classical Approximation in Quantum Mechanics. D. Reidel Publishing Company, Dordrecht, Boston, London (1981)
20. Messiah, A.: Quantum Mechanics, Vol. II. 2nd edition. Wiley, New York (1967)
21. Takens, F.: Motion under the influence of a strong constraining force. In: Global Theory of Dynamical Systems, Evanston 1979 (Z. Nitecki and C. Robinson, eds.). Springer-Verlag, Berlin, Heidelberg, New York (1980)
22. Tully, J. C.: Nonadiabatic Processes in Molecular Collisions. In: Dynamics of Molecular Collisions, Part B (W.H. Miller, ed.). Plenum, New York (1976)
23. Zener, C.: Non-adiabatic crossing of energy levels, Proc. R. Soc. London, Ser. A **137** (1932) 696–702

Numerical Integrators for Quantum-Classical Molecular Dynamics

Peter Nettesheim[1] and Christof Schütte[1,2]

[1] Konrad Zuse Zentrum Berlin, Takustr. 7, 14195 Berlin, Germany
[2] Freie Universität Berlin, Fachbereich Mathematik, Arnimallee 2–6, 14195 Berlin, Germany

Abstract. It was revealed that the QCMD model is of canonical Hamiltonian form with symplectic structure, which implies the conservation of energy. An efficient and reliable integrator for transfering these properties to the discrete solution is the symplectic and explicit PICKABACK algorithm. The only drawback of this kind of integrator is the small stepsize in time induced by the splitting techniques used to discretize the quantum evolution operator. Recent investigations concerning Krylov iteration techniques result in alternative approaches which overcome this difficulty for a wide range of problems. By using iterative methods in the evaluation of the quantum time propagator, these techniques allow for the stepsize to adapt to the classical motion and the coupling between the classical and the quantum mechanical subsystem. This yields a drastic reduction of the numerical effort. The pros and cons of both approaches as well as the suitable applications are discussed in the last part.

1 Introduction

Various kinds of *mixed quantum-classical* models have been introduced in the literature. We will concentrate on the so-called quantum-classical molecular dynamics (QCMD) model, which consists of a Schrödinger equation coupled to classical Newtonian equations (cf. Sec. 2).

In this paper, we focus on numerical techniques for integrating the QCMD equations of motion. The aim of the paper is to systematize the discussion concerning numerical integrators for QCMD by:

- giving a derivation of the different techniques based on a common construction principle,
- classifying the application problems in order to link together the properties of the integrators and the structure of the problem under consideration.

For this purpose, a short overview will be given concerning some *theoretical* properties of the QCMD model (Sec. 2). This will allow for a suitable classification of the application problems. In the course of the following discussion, we will introduce two different classes of integration techniques:

In Sec. 3, some recent developments of "structure conserving" integrators will be reviewed. Such *symplectic* or *symmetric* integrators are build to preserve certain geometric properties of the exact QCMD solution like energy

conservation or reversibility. They are preferable for applications to long-term simulations, but turn out to have some crucial disadvantages when short-term simulations up to a given precision are wanted.

Hence, as the second class of techniques, we discuss *adaptive* methods for accurate short-term integration (Sec. 4). For this class, it is the major requirement that the discretization allows for the stepsize to adapt to the classical motion and the coupling between the classical and the quantum mechanical subsystem. This means, that we are interested in discretization schemes which avoid stepsize restrictions due to the fast oscillations in the quantum part. We can meet this requirement by applying techniques recently developed for evaluating matrix exponentials iteratively [12]. This approach yields an adaptive *Verlet-based exponential integrator* for QCMD.

Finally, in Sec. 5, the theoretical results are illustrated by applying two adaptive schemes to the collinear photo dissociation of ArHCl.

2 The QCMD Model

There are various approaches to the problem of coupling quantum degrees of freedom to classical degrees of freedom. The QCMD model is given by the following equations of motion:

$$\begin{aligned} i\hbar\dot{\psi} &= \Big(\underbrace{T + V(x,q)}_{H}\Big)\,\psi, \qquad T = -\frac{\hbar^2}{2m}\Delta_x \qquad \psi|_{t=0} = \psi_0 \\ M\dot{q} &= p, \qquad\qquad\qquad\qquad\qquad\quad q(0) = q_0, \\ \dot{p} &= -\langle\psi, \nabla_q V\psi\rangle, \qquad\qquad\qquad\quad p(0) = p_0 \end{aligned} \tag{1}$$

The quantum degrees of freedom are described by a wave function $\psi = \psi(x,t)$. It obeys Schrödinger's equation with a parameterized coupling potential V which depends on the location $q = q(t)$ of the classical particles. This location $q(t)$ is the solution of a classical Hamiltonian equation of motion in which the time–dependent potential arises from the expectation value of V with regard to ψ. For simplicity of notation, we herein restrict the discussion to the case of only two interacting particles. Nevertheless, all the following considerations can be extended to arbitrary many particles or degrees of freedom.

2.1 Conservation Properties of the QCMD Model

In a first discretization step, we apply a suitable spatial discretization to Schrödinger's equation, e.g., based on pseudospectral collocation [15] or finite element schemes. From now on, we consider ψ, T, V and H as denoting the corresponding vector and matrix representations, respectively. The total

energy expectation value of the system

$$\mathcal{H} = \psi^* H(q)\psi + \frac{1}{2M}|p|^2, \tag{2}$$

to which we will simply refer as "energy", is a constant of motion [2]. Another conserved quantity is the norm of the wave function, due to the unitary propagator in the quantum part. We are interested in constructing numerical integrators which reproduce these conservation properties. To this end, it is enormously helpful to observe that the QCMD equations are of canonical Hamiltonian form with respect to $\mathcal{H}$. In order to illustrate this fact, we decompose the wave function into a scaled real and imaginary part $\psi = (q_\psi + ip_\psi)/\sqrt{2\hbar}$ and introduce generalized positions $Q = (q_\psi, q)^T$ and momenta $P = (p_\psi, p)^T$. This allows for denoting the whole system (1) in canonical Hamiltonian form:

$$\dot{Q} = \frac{\partial}{\partial P}\mathcal{H}, \qquad \dot{P} = -\frac{\partial}{\partial Q}\mathcal{H}, \tag{3}$$

with the usual symplectic structure (cf. [16]).

2.2 Adiabatic Limit of QCMD

QCMD describes a coupling of the "fast" motions of a quantum particle to the "slow" motions of a classical particle. In order to classify the types of coupled motion we eventually have to deal with, we first analyze the case of an extremely heavy classical particle, i.e., the limit $M \to \infty$ or, better, $m/M \to 0$. In this "adiabatic limit", the classical motion is so slow in comparison with the quantal motion that it cannot induce an excitation of the quantum system. That means, that the populations $\theta_k(t) = |\langle\psi(t), \Phi_k(q(t))\rangle|$ of the eigenstates $\Phi_k(q)$ of the Hamiltonian $H(q)$ remain *constant* along the classical path $q = q_{\text{BO}}$. Hence, the limit populations $\theta_k(t) = \theta_k(0)$ may be computed from the initial conditions. The classical limit path is given by the time-dependent Born-Oppenheimer model:

$$M\,\ddot{q}_{\text{BO}} = -\sum_k \theta_k(0)^2\,(\text{grad}_q E_k)(q_{\text{BO}}), \quad q_{\text{BO}}(0) = q_0,\ \dot{q}_{\text{BO}}(0) = p_0/M,$$

where the $E_k(q)$ are the eigenenergies of $H(q)$ (for details concerning the adiabatic limit see [21] in this collection, or [3, 1]). The associated asymptotic expression for the wave function

$$\psi(t) = \sum_k \theta_k(0)\,\exp\left(-\frac{i}{\hbar}\int_0^t E_k(q_{\text{BO}}(s))ds\right)\Phi_k(q_{\text{BO}}(t)) + \underbrace{\mathcal{O}(\sqrt{m/M})}_{\text{nonadiabaticity}} \tag{4}$$

deserves our attention because it uncovers some essential features of QCMD motion:

1. The quantal motion is highly oscillatory with frequencies given by the eigenvalues of the Hamiltonian H.
2. For m/M small enough, the populations of the eigenstates Φ_k are nearly constant and the quantal motion is given in terms of the evolution of the eigenstates and eigenenergies E_k along q_{BO}.
3. For larger values of m/M, we have to expect *nonadiabatic* redistribution of the populations induced by the classical motion.

2.3 Classification of Application Problems

It is the aim of this paper to take into account a wide range of systems to which QCMD is applied. For a precise understanding of the situation, it is necessary to recognize the differences between these applications, because these differences demand for specific features of the numerical integrator. In the following, we will describe a suitable classification of the application problems.

1. For *long term simulations*, it turns out that the reproduction of the conservation properties is most important in order to ensure reliable results.
2. For *short term simulations*, accuracy requirements on the discrete solution make sense and we advocate error controlling adaptive integrators. Moreover, we have to further subdivide our classification due to the observations in Sec. 2.2:
   (a) Problems with (nearly) adiabatic motion.
   (b) Problems with essentially nonadiabatic motion.

In most real life applications, the evaluation of the forces acting on the classical particles (i.e., the evaluation of the gradient of the interaction potential) is by far the most expensive operation due to the large number of classical degrees of freedom. Therefore we will concentrate on numerical techniques which try to minimize the number of force evaluations.

3 Structure Conserving Integration Schemes

Since we have discovered the underlying Hamiltonian structure of the QCMD model we are able to apply methods commonly used to construct suitable numerical integrators for Hamiltonian systems. Therefore we transform the QCMD equations (1) into the Liouville formalism. To this end, we introduce a new state z in the phase space, $z = (\mathcal{Q}_{\mathcal{N}}, \mathcal{P}_{\mathcal{N}})^T$, and define the nonlinear Liouville operator $L_{\mathcal{H}} z_i = \{z_i, \mathcal{H}\}$, using the common Poisson brackets $\{\ ,\ \}$. This permits us to denote the QCMD equations (1) in the form $\dot{z} = L_{\mathcal{H}} z$. The formal solution can now be written as

$$z(\tau) = e^{\tau L_{\mathcal{H}}} z(0). \tag{5}$$

At this point we may apply well-known approximation techniques. For each decomposition of $\mathcal{H}$, i.e., $\mathcal{H} = \mathcal{H}_1 + \mathcal{H}_2 + \ldots$, the corresponding Lie-generator decomposes accordingly

$$L_{\mathcal{H}} = L_{\mathcal{H}_1} + L_{\mathcal{H}_2} + \ldots .$$

Using splitting schemes of the exponential function allows for a generation of numerical integrators. For example [24, 22]:

$$e^{\tau(L_{\mathcal{H}_1}+L_{\mathcal{H}_2})} = e^{\tau L_{\mathcal{H}_1}} e^{\tau L_{\mathcal{H}_2}} + \mathcal{O}\left(\tau^2\right) \tag{6}$$

$$e^{\tau(L_{\mathcal{H}_1}+L_{\mathcal{H}_2})} = e^{\frac{\tau}{2} L_{\mathcal{H}_1}} e^{\tau L_{\mathcal{H}_2}} e^{\frac{\tau}{2} L_{\mathcal{H}_1}} + \mathcal{O}\left(\tau^3\right), \tag{7}$$

which can easily be extended to higher orders [25]. It should be pointed out that the given approximation orders, here and in the following, essentially depend on the smoothness of the solution, i.e., the error term may be affected by the highest quantum frequency excited.

Note, that the choice of the $\mathcal{H}_k$ crucially influences the properties of the resulting integrator.

3.1 Symplectic Integrators

A well-known property of symplectic integrators is the conservation of the total energy within a very accurate deviation range even for long term simulations. It can be shown that symplectic integrators in application to Hamiltonian systems solve a system corresponding to a modified Hamiltonian with a small stepsize-dependent perturbation [8]. This leads to a "quasi conservation" of some first integrals, so that, for example, the total energy of the discrete solution oscillates around its initial value with a small amplitude that decreases with the stepsize used (cf. Fig. 1). This "structural stability" makes symplectic integrators superior for long term simulations.

A convenient and constructive approach to attain symplectic maps is given by the composition of symplectic maps, which yields again a symplectic map. For appropriate $\mathcal{H}_k$, the splittings (6) and (7) are exactly of this form: If the $\mathcal{H}_k$ are Hamiltonians with respect to the whole system, then the $\exp(\tau L_{\mathcal{H}_k})$ define the phase flow generated by these $\mathcal{H}_k$. Thus, the $\exp(\tau L_{\mathcal{H}_k})$ are symplectic maps on the whole phase space and the compositions in (6) and (7) are symplectic maps, too. Moreover, in order to allow for a direct numerical realization, we have to find some $\mathcal{H}_k$ for which either $\exp(\tau L_{\mathcal{H}_k})$ has an analytic solution or a given symplectic integrator.

Pickaback We decompose $\mathcal{H}$ into a kinetic and a potential term:

$$\mathcal{H} = \mathcal{H}_1 + \mathcal{H}_2 \quad \text{with} \quad \mathcal{H}_1 = \frac{p^2}{2\,M} + \psi^* T\,\psi \quad \text{and} \quad \mathcal{H}_2 = \psi^* V(q)\,\psi.$$

As shown in [16], the two corresponding flow maps, $\exp(\tau L_{\mathcal{H}_1})$ and $\exp(\tau L_{\mathcal{H}_2})$, can be represented analytically. Using the second order Strang

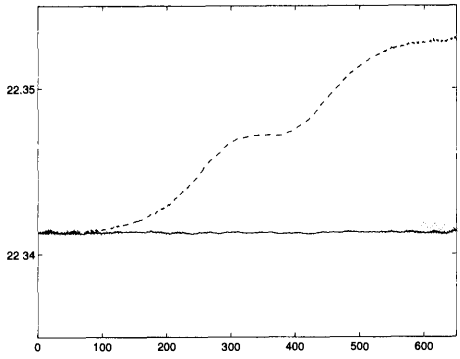


Fig. 1. Total energy (in kJ/mol) versus time (in fs) for different integrators for a collinear collision of a classical particle with a harmonic quantum oscillator (for details see [2]). Dashed line: Nonsymplectic scheme. Dotted: Symplectic integrator of first order. Solid: PICKABACK (symplectic, second order).

splitting (7), we derive an integration scheme which is explicit, symplectic and symmetric. This scheme was denoted PICKABACK emphasizing the interwoven structure of the partial steps.

$$\begin{aligned} q_{1/2} &= q_0 + \frac{\tau}{2}\frac{p_0}{M} \\ \psi_{1/2} &= \exp\Big(-i\frac{\tau}{2}T_N\Big)\,\psi_0 \\ p_1 &= p_0 - \tau\,\psi^*_{1/2}\,D_q V(q_{1/2})\psi_{1/2} \\ \psi_1 &= \exp\Big(-i\frac{\tau}{2}T_N\Big)\,\exp\Big(-i\tau V_N(q_{1/2})\Big)\,\psi_{1/2} \\ q_1 &= q_{1/2} + \frac{\tau}{2}\frac{p_1}{M}. \end{aligned} \tag{8}$$

A main advantage of PICKABACK is its reliability. But the reader might notice, that the splitting of the quantum propagator $\exp(-i\frac{\tau}{2}H)$ restricts the stepsize to the order of the inverse of the largest eigenvalue of H. Thus, the overall time steps are connected to the shortest significant period of phase oscillation in the quantum subsystem – demanding more evaluations of the pure classical forces than required by the classical motion itself. In order to circumvent the problem we switch to symmetric but no longer symplectic methods.

3.2 Symmetric Integration Schemes

Beneath the conservation properties of QCMD its equations of motion possess another important geometric structure by being time reversible. As shown in [10], the application of symmetric integrators to reversible problems yields

the solution of a perturbed but again reversible problem. Hence, all the characteristics which are connected to reversibility are structurally inherited if the discretization scheme is symmetric.

The splitting technique, introduced above for the construction of symplectic schemes, is also adequate for symmetric ones. Now, the only condition is that we have to split $e^{\tau L_{\mathcal{H}}}$ symmetrically. To this end, let us consider the Liouville generator for the Hamiltonian $\mathcal{H}$ from above:

$$L_{\mathcal{H}} = \underbrace{(\nabla_q \mathcal{H})^T \nabla_p - (\nabla_p \mathcal{H})^T \nabla_q}_{L_{\mathcal{H}}^{cl}} + \underbrace{(\nabla_{q_\psi} \mathcal{H})^T \nabla_{p_\psi} - (\nabla_{p_\psi} \mathcal{H})^T \nabla_{q_\psi}}_{L_{\mathcal{H}}^{qm}}$$

decomposing as $L_{\mathcal{H}} = L_{\mathcal{H}}^{cl} + L_{\mathcal{H}}^{qm}$, with $L_{\mathcal{H}}^{cl}$ acting on the classical coordinates and $L_{\mathcal{H}}^{qm}$ acting on the quantum subsystem only. This permits to produce symmetric schemes via, for example, the second order Strang splitting:

$$e^{\tau L_{\mathcal{H}}} = e^{\frac{\tau}{2} L_{\mathcal{H}}^{qm}} e^{\tau L_{\mathcal{H}}^{cl}} e^{\frac{\tau}{2} L_{\mathcal{H}}^{qm}} + \mathcal{O}\left(\tau^3\right)$$

Using the symmetric Verlet algorithm for integrating $\exp(\tau L_{\mathcal{H}}^{cl})$ yields:

$$\begin{aligned} \psi_{1/2} &= \exp\Big(-i\frac{\tau}{2\hbar} H(q_0)\Big)\,\psi_0 \\ \text{Leapfrog} \begin{cases} q_{1/2} &= q_0 + \frac{\tau}{2}\frac{p_0}{M} \\ p_1 &= p_0 - \tau\,\psi_{1/2}^{*}\, D_q V(q_{1/2})\psi_{1/2} \\ q_1 &= q_{1/2} + \frac{\tau}{2}\frac{p_1}{M} \end{cases} \\ \psi_1 &= \exp\Big(-i\frac{\tau}{2\hbar} H(q_1)\Big)\,\psi_{1/2}. \end{aligned} \tag{9}$$

The question remains how to evaluate $\exp(-i\tau H(q_0)/(2\hbar))\psi_i$ while retaining the symmetric structure. In Sec. 4.2 we will introduce some iterative techniques for evaluating the matrix exponential but the approximative character of these techniques will in principle destroy the symmetry.

Symmetric multiple time stepping An intriguingly simple idea for realizing a symmetric approximation of the matrix exponential is presented in [20] (and is extended to symplectic splittings in [17]). It copes with the different time scales of classical and quantum degrees of freedom by splitting the quantum propagation in some small "substeps". The resulting scheme is a variant of (9) with its quantum steps replaced by

$$\begin{aligned} \psi_{1/2} &= \left(e^{-i\frac{\tau}{2\hbar n} H(q_0)}\right)^n \psi_0 \\ &= \left(e^{-i\frac{\tau}{4\hbar n} T} e^{-i\frac{\tau}{2\hbar n} V(q_0)}\, e^{-i\frac{\tau}{4\hbar n} T}\right)^n \psi_0 + \mathcal{O}\left(\frac{\tau^3}{n^2}\right) \end{aligned}$$

The splitting of the quantum propagator negatively effects the efficiency of the scheme especially if m/M is small, i.e., if the quantum oscillation are much faster than the classical motion and the number n of substeps is becoming inefficiently large.

4 Adaptive Methods

4.1 Adaptive Stepsize Control

We have to pay a price for the advantages of symplectic and symmetric methods: The stepsize τ has to be constant during the simulation, because, up to now, there is no appropriate strategy for efficiently controlling the stepsize without destroying the "structural stability". This means, that the overall stepsize has to be reduced until it fulfills the accuracy requirements during the whole integration period. In many real life applications of QCMD, the dynamical behavior of the solution can change dramatically during the course of the simulation (collisions, excitation processes). In principle, one would like to make large time steps where "nothing important happens" and small ones where it is necessary to resolve important processes, i.e., the stepsize should be adapted to the accuracy wanted. In Numerical Analysis, such stepsize control strategies have extensively been discussed. In the following, we will give a brief overview on the usual strategy (for details cf. [4, 9]). The conceptual framework requires the control of the approximation error *in each time step* via choosing the stepsize with respect to a given accuracy requirement. That is, the stepsize is controlled in a way which bounds the local approximation error by a given tolerance TOL.

The local error in the step from time t to $t+\tau$, i.e., the error, which is produced by calculating a discrete solution in this step instead of exactly solving the QCMD equations, is given as follows:

$$\epsilon_\tau(t+\tau) \;=\; \Phi_p^\tau\, z(t) - \exp(\tau L_{\mathcal{H}})\, z(t),$$

where $\exp(\tau L_{\mathcal{H}})\, z(t)$ denotes the exact solution of the QCMD model and Φ_p^τ the discrete evolution of order p and with stepsize τ, for example the map given by (9).

Unfortunately, this local error ϵ_τ cannot be calculated, since we do not know the exact solution to the QCMD equations. The clue to this problem is given by the introduction of an approximation to ϵ_τ. Let us consider another discrete evolution Φ_q^τ with an order $q > p$ and define an error estimation $\hat{\epsilon}_\tau$ via $\hat{\epsilon}_\tau(t+\tau) = \Phi_p^\tau\, z(t) - \Phi_q^\tau\, z(t)$.

The control scheme tries to choose the stepsize τ so that $\|\hat{\epsilon}_\tau\| = \mathrm{TOL}$ in some adequate norm. In case of a tolerance exceeding error, i.e., for $\|\hat{\epsilon}_\tau\| > \mathrm{TOL}$, one reduces the stepsize according to

$$\tau_{new} \;=\; \sqrt[p+1]{\rho\ \mathrm{TOL}/\|\hat{\epsilon}_\tau\|}\ \tau_{old}. \tag{10}$$

with an additional safety factor $\rho < 1$. The same formula is used in order to predict a proper stepsize for the next step. Problems can arise, when the error approaches zero. We cope with them by restricting the allowed increase of the stepsize.

For realizing (10), we need an adequate norm for measuring the error. It obviously makes no sense to use an Euclidian norm of z indiscriminately of

quantum and classical parts. We advocate the use of a scaled norm in the classical subsystem and the usual 2-Norm for the quantum part:

$$\|\hat{\epsilon}_\tau(t)\| = \sqrt{\|\psi(t) - \hat{\psi}(t)\|_2^2 + \left|\frac{q(t) - \hat{q}(t)}{\max(q(t), s_{min})}\right|^2 + \left|\frac{p(t) - \hat{p}(t)}{\max(p(t), s_{min})}\right|^2},$$

where ψ, q and p denote the results of Φ_q^τ and $\hat{\psi}, \hat{q}$ and $\hat{p}$ that of Φ_q^τ. A threshold value $s_{min} > 0$ avoids an exploding error for locations or momenta close to zero.

The error estimate approximates the error of the propagation with the less accurate method Φ_p^τ. Nonetheless, the next step is started with the more precise result of Φ_q^τ.

We are now concerned with the selection of two integration methods of different order. A first idea – which we are not advocating – is to use the PICKABACK integrator (8) as Φ_q^τ together with a first order scheme based on the Trotter formula (6) replacing Φ_p^τ. Recalling that the stepsize of these methods are dominated by the splitting of $\exp(-i\tau H/\hbar)$, we actually foresee the effect of such an adaptive method. The scheme correctly resolves the dynamical behavior but forces the stepsize to remain restricted to the order of the inverse of the largest eigenvalue of the Hamiltonian. An illustrative example of these drawbacks is given in Sec. 5.

4.2 A Verlet-Based Adaptive Integrator

A more convincing approach leads to an adaptive method based on the symmetric second order scheme (9). As a first step, we have to introduce a first order scheme substituting Φ_p^τ of the previous section. In what follows, we use the following pair of schemes:

2nd order symmetric scheme as Φ_q^τ

$$\begin{cases} \psi_{1/2} &= \exp\left(-i\frac{\tau}{2\hbar}H(q_0)\right)\,\psi_0 \\ q_{1/2} &= q_0 + \frac{\tau}{2}\frac{p_0}{M} \\ p_1 &= p_0 - \tau\,\psi_{1/2}^*\,D_qV(q_{1/2})\psi_{1/2} \\ q_1 &= q_{1/2} + \frac{\tau}{2}\frac{p_1}{M} \\ \psi_1 &= \exp\left(-i\frac{\tau}{2\hbar}H(q_1)\right)\,\psi_{1/2} \end{cases} \tag{11}$$

For comparison: 1st order Euler scheme as Φ_p^τ

$$\begin{cases} \hat{\psi}_1 &= \exp\left(-i\frac{\tau}{\hbar}H(q_0)\right)\,\psi_0 \\ \hat{q}_1 &= q_0 + \tau\frac{p_0}{M} \\ \hat{p}_1 &= p_0 - \tau\,\psi_{1/2}^*\,D_qV(\hat{q}_1)\psi_{1/2} \end{cases}$$

When considering the construction of *exactly* symmetric schemes, we are obstructed by the requirement to find *exactly* symmetric approximations to $\exp(-i\tau H/(2\hbar))$. But it is known [10], that the usual stepsize control mechanism destroys the reversibility of the discrete solution. Since we are applying this mechanism, we now may use approximations to $\exp(-i\tau H/(2\hbar))$ which are not precisely symmetric, i.e., we are free to take advantage of the superior efficiency of iterative methods for evaluating the matrix exponential. In the following, we will compare three different approaches.

Chebyshev Approximation The well known expansion of $\exp(-i\tau H/\hbar)$ into Chebyshev polynomials T_k [23] is one of the most frequently used integration technique in numerical quantum dynamics:

$$\exp(-\tfrac{i}{\hbar}\, H\, \tau)\psi(t) \approx \textstyle\sum_{k=1}^{N} \alpha_k(\rho\tau)\, T_k(-\tfrac{i}{\hbar} H)\psi$$

with appropriately chosen coefficients α_k and an estimate ρ for the spectral radius of the Hamiltonian H. This technique allows for large stepsizes if the truncation index N is chosen large enough. The N necessary for achieving a specific accuary depends linearly on the stepsize τ and the spectral radius of H. We use an adaptive stopping criterion for the iteration based on the decay of the coefficients α_k [14].

Krylov Approximation of the Matrix Exponential The iterative approximation of the matrix exponential based on Krylov subspaces (via the Lanczos method) has been studied in different contexts [12, 19, 7]. After the iterative construction of the Krylov basis $\{v_1, \ldots, v_n\}$, the matrix exponential is approximated by using the representation A of $H(q)$ in this basis:

$$\exp\left(-\tfrac{i\tau}{\hbar} H(q)\right)\psi \;\approx\; V\, \exp\left(-\tfrac{i\tau}{\hbar} A\right)\, V^*\psi, \quad \text{with } V = [v_1, \ldots, v_n].$$

The evaluation of $\exp(-i\tau A/\hbar)$ is cheap since A is tridiagonal.

In [13], an efficient residual error estimation scheme has been introduced for controlling the quality of the approximation. This gives us a stopping criterion for the iteration guaranteeing that the quality of the approximation fits to the accuracy requirements of the stepsize control.

In most cases, this Lanczos–based technique proves to be superior to the Chebyshev method introduced above. It is the method of choice for the application problems of class 2b of Sec. 2. The Chebyshev method is superior only in the case that nearly all eigenstates of the Hamiltonian are substantially occupied.

However, using the Lanczos iteration for evaluating the matrix exponential produces two eventual drawbacks. Firstly, the iteration does not use any of the information gathered in the last step. But if the eigenvectors undergo only minor changes from step to step, some approximate eigenvectors of the last step may be used as good initial choices for the next iteration. This

idea can be realized by using Block-Lanczos iteration instead of the pure Lanczos scheme. The second drawback is important if the motion under consideration is nearly adiabatic and only a few, let us say m, eigenstates are occupied. By approximating these eigenstates in a Krylov basis with typically $d > m$ basis vectors, the Lanczos scheme necessarily introduces (small) artificial populations of other than the m states occupied. From time step to time step, this will lead to an artificial and unwanted blow-up of the dimension of the occupied subspace.

Subspace-Controlling Iteration Methods Out of this observation we also studied some subspace-controlling algorithms. In these approaches, we do *not* try to construct an (eventually large) basis set for transforming the Hamiltonian into a form appropriate for an efficient evaluation of the matrix exponential. Instead of this, we directly approximate a (small) basis set for the relevant (small) subspace. Only then, the matrix exponential is computed using this basis. In the course of the iteration, appropriate error estimates control whether the subspace dimension has to be increased or may be reduced. Mainly two techniques were tested in order to evaluate the basis set: a simultaneous minimization of the Rayleigh quotient in the subspace via an appropriately preconditioned conjugate gradient iteration [6] and a multi grid approach to the eigenvalue problem as introduced in [5]. Both techniques prove to be superior to the Lanczos approach for nearly adiabatic problems with very few eigenstates occupied (class 2a). But they quickly get inefficient if a nonadiabatic excitation of previously unimportant states is essential.

5 An Illustrative Example

In this section, the theoretical results are checked and illustrated by numerical simulations. Therefore we consider a well-known test problem which is of class 2b in our classification from page 399: a photo dissociation process of a collinear ArHCl molecule (see Fig. 2). The photo dissociation is modeled via a transition of the bounding Hydrogen-Chlorine ground state into a repulsive excited state. The Hydrogen starts oscillating between Argon and Chlorine transferring more and more kinetic energy to the Argon atom.

Using Jacobi coordinates and reduced masses, the Hydrogen-Chlorine interaction is modeled quantum mechanically whereas the Ar–HCl interaction classically. The potentials used, initial data and additional computational parameters are listed in detail in [16].

Obviously, one test example is not enough to illuminate all the effects pointed out previously. Thus, we have to concentrate herein on some main ideas. An extensively example-based comparison is in preparation [18].

The stepsize controlling adaptive QCMD integrators presented in the previous section differ only with respect to the approximation of the quantum propagation. We herein compare three of these integrators, all of them

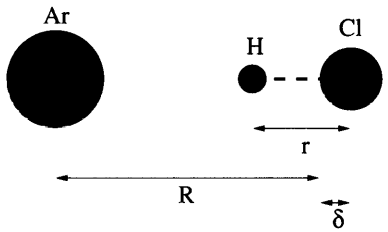


Fig. 2. Collinear ArHCl-system with the Jacobi-coordinates used.

equipped with the stepsize control mechanism (10): two integrators based on the pair of discretization schemes from page 404, with on one hand a Chebyshev approximation and on the other hand a Krylov approximation of the matrix exponential, and – just to show the stepsize restriction due to the splitting of the quantum propagator – a stepsize controlled Pickaback scheme.

To begin with, we compare the stepsizes used in the simulations (Fig. 3). As pointed out before, it seems to be unreasonable to equip the Pickaback scheme with a stepsize control, because, as we indeed observe in Fig. 3, the stepsize never increases above a given level. This level depends solely on the eigenvalues of the quantum Hamiltonian. When analyzing the other integrators, we observe that the stepsize control just adapts to the dynamical behavior of the classical subsystem. The internal (quantal) dynamics of the Hydrogen-Chlorine subsystem does not lead to stepsize reductions.

Large stepsizes result in a strong reduction of the number of force field evaluations per unit time (see left hand side of Fig. 4). This represents the major advantage of the adaptive schemes in comparison to structure conserving methods. On the right hand side of Fig. 4 we see the number of FFTs (i.e., matrix-vector multiplication) per unit time. As expected, we observe that the Chebyshev iteration requires about double as much FFTs than the Krylov techniques. This is due to the fact that only about half of the eigenstates of the Hamiltonian are essentially occupied during the process. This effect occurs even more drastically in cases with less states occupied.

Everything seems to be in favor of the stepsize controlling schemes, but the reader might notice, that there are —up to now— some drawbacks. When computing the autocorrelation function, e.g., the complex valued function $\psi(0)^*\psi(t)$, used for instance in the calculation of absorption spectra, we find a substantial phase shift in the discrete solution. This results from the fact that the error control mechanism is not adapted to detect phase errors because both schemes from page 404 depend on *pointwise* updates of the Hamiltonian only. It should be possible to overcome this problem by applying alternative Verlet-based schemes using *averaged* updates and forces. Promising candidates for these schemes were recently introduced by HOCHBRUCK and LUBICH (see [11] in this collection). Extensive numerical experiments using

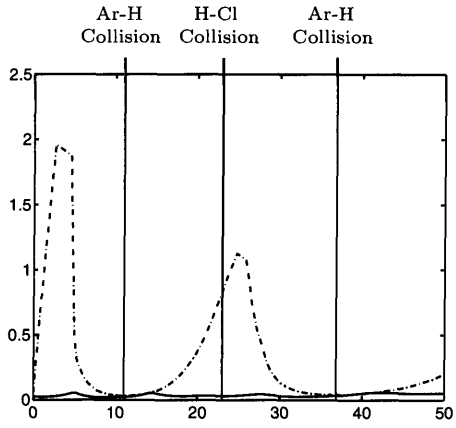


Fig. 3. Stepsize τ used in the simulation of the collinear photo dissociation of ArHCl: the adaptive Verlet-based exponential integrator using the Lanczos iteration (dash-dotted line) for the quantum propagation, and a stepsize controlling scheme based on PICKABACK (solid line). For a better understanding we have added horizontal lines marking the collisions (same tolerance TOL). We observe that the quantal H-Cl collision does not lead to any significant stepsize restrictions.

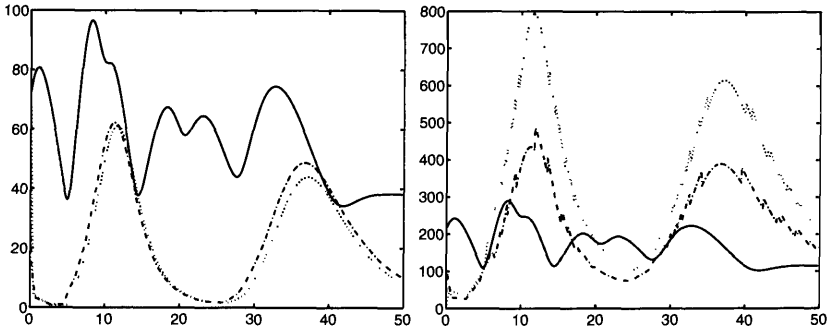


Fig. 4. Photo dissociation of ArHCl. Left hand side: the number of force field evaluations per unit time. Right hand side: the number of Fast-Fourier-transforms per unit time. Dotted line: adaptive Verlet with the Chebyshev approximation for the quantum propagation. Dash-dotted line: with the Lanczos iteration. Solid line: stepsize controlling scheme based on PICKABACK. If the FFTs are the most expensive operations, PICKABACK-like schemes are competitive, and the Lanczos iteration is significantly cheaper than the Chebyshev approximation.

such schemes in the adaptive context introduced herein will be presented in a forthcoming paper [18].

6 Conclusions

We discussed numerical integration techniques for different classes of applications of QCMD. We mainly distinguished between *long* and *short term* simulations. Short term simulations are characterized by the fact that specific (global) accuracy requirements for the numerical solution make sense. For long term simulations one is more interested in certain stability and conservation properties of the solution, despite the fact that its global accuracy might be spoiled by the amplification of numerical error. Consequently, the advocated numerical techniques should be divided into two categories:

Long term simulations require "structurally stable" integrators. *Symplectic* and *symmetric* methods nearly perfectly reproduce structural properties of the QCMD equations, as, for example, the conservation of the total energy. We introduced an explicit symplectic method for the QCMD model — the PICKABACK scheme— and a symmetric method based on multiple time stepping.

For short-term simulations we advocate the use of *stepsize controlling* integrators which gain efficiency by adapting the stepsize to the dynamics of the system. We presented an adaptive Verlet-based exponential integrator for QCMD with iterative evaluation of the quantum propagation. It permits us to use stepsizes which are not restricted by the fast phase oscillations in the quantum part. For the iterative realization of the quantum propagation steps, we analyzed three different approaches: the Chebyshev approximation, the Lanczos iteration and a subspace controlling method. For the application problems with a (nearly) adiabatic behavior (class 2a from page 399), the subspace controlling method appeared to be best suited, because it does not artificially blow up the excited subspace. For all other cases of short-term simulation (class 2b), we advocate the Lanczos iteration scheme owing to its efficient adaption of the basis set to the dynamical behavior.

Acknowledgement. It is a pleasure to thank M. Hochbruck and Ch. Lubich for their helpful comments on the topics of this paper.

References

1. Bornemann, F. A.: Homogenization in time of singularly perturbed conservative mechanical systems. Manuscript (1997) 146pp
2. Bornemann, F. A., Nettesheim, P., Schütte, Ch.: Quantum-classical molecular dynamics as an approximation to full quantum dynamics. J. Chem. Phys., **105** (1996) 1074-1083

3. Bornemann, F. A., Schütte, Ch.: On the singular limit of the quantum-classical molecular dynamics model. Preprint SC 97-07 (1997) Konrad-Zuse-Zentrum Berlin. SIAM J. Appl. Math. (submitted)
4. P. Deuflhard and F. Bornemann: Numerische Mathematik II — Integration gewöhnlicher Differentialgleichungen. Walter de Gruyter, Berlin, New York (1994)
5. P. Deuflhard, T. Friese, F. Schmidt, R. März, and H.-P. Nolting: Effiziente Eigenmodenberechnung für den Entwurf integriert–optischer Chips. In W. Jäger, Th. Lohmann, and H. Schunck, editors, Mathematik – Schlüsseltechnologie für die Zukunft. Springer Verlag.
6. B. Dietrich: Numerische Behandlung der zeitabhängigen Schrödingergleichung mit iterativen Unterraummethoden. Master's thesis, Freie Universität Berlin (1997)
7. V.L. Druskin and L.A. Knizhnerman: Krylov subspace approximation of eigenpairs and matrix functions in exact and computer arithmetics. *Num. Lin. Alg. Appl.*, **2** (1995) 205-217
8. E. Hairer and Ch. Lubich: The life-span of backward error analysis for numerical integrators. Numer. Math. **76** (1997) 441–462
9. E. Hairer, S.P. Nørsett, and G. Wanner: Solving Ordinary Differential Equations I, Nonstiff Problems. Springer Verlag, Berlin, Heidelberg, New York, Tokyo, 2nd edition (1993)
10. E. Hairer and D. Stoffer: Reversible long–term integration with variable step sizes. Report (1995)
11. M. Hochbruck and Ch. Lubich: A bunch of time integrators for quantum/classical molecular dynamics. (1998) (this volume)
12. M. Hochbruck and Ch. Lubich: On Krylov subspace approximations to the matrix exponential operator. SIAM J. Numer. Anal. **34** (1997) (to appear)
13. M. Hochbruck, Ch. Lubich, and H. Selhofer: Exponential integrators for large systems of differential equations. SIAM J. Sci. Comp. (1998) (to appear)
14. W. Huisinga: Faber-, Newton- und Krylov-Approximation zur Integration großer Differentialgleichungssysteme aus der Quantendynamik. Master's thesis, Freie Universität Berlin (1997)
15. Ronnie Kosloff: Quantum molecular dynamics on grids. In R. E. Wyatt and J. Z. Zhang, editors, Dynamics of Molecules and Chemical Reactions, pages 185–230. Marcel Dekker, New York (1996)
16. P. Nettesheim, F. A. Bornemann, B. Schmidt, and Ch. Schütte: An explicit and symplectic integrator for quantum-classical molecular dynamics. Chem. Phys. Lett. **256** (1996) 581–588
17. P. Nettesheim, and S. Reich: Symplectic multiple-time-stepping integrators for quantum-classical molecular dynamics. (1998) (this volume)
18. P. Nettesheim, Ch. Schütte, M. Hochbruck, and Ch. Lubich. (work in preparation)
19. T.J. Park and J.C. Light: Unitary quantum time evolution by iterative Lanczos reduction. J. Chem. Phys. **85** (1986) 5870–5876
20. U. Schmitt and J. Brickmann: Chem. Phys. **208(45)** (1996)
21. Schütte, Ch., Bornemann, F. A.: Approximation Properties and Limits of the Quantum-Classical Molecular Dynamics Model. Preprint SC 97-41 (1997) Konrad-Zuse-Zentrum Berlin (submitted to proceedings of MacroMM97)
22. G. Strang: On the construction and comparison of difference schemes. SIAM J. Numer. Anal. **5** (1968) 506–517

23. H. Tal-Ezer and R. Kosloff: An accurate and efficient scheme for propagating the time dependent Schrödinger equation. J. Chem. Phys. **81(9)** (1984) 3967–3971
24. H.F. Trotter: On the product of semi-groups of operators. Proc. Am. Math. Soc. **10** (1959) 545–551
25. H. Yoshida: Construction of higher order symplectic integrators. Physics Letters A **150** (1990) 262–268

Symplectic Multiple-Time-Stepping Integrators for Quantum-Classical Molecular Dynamics

Peter Nettesheim and Sebastian Reich

Konrad-Zuse-Zentrum, Takustr. 7, D-14195 Berlin, Germany
E-mail: nettesheim@zib.de, reich@zib.de

Abstract. The overall Hamiltonian structure of the Quantum-Classical Molecular Dynamics model makes – analogously to classical molecular dynamics – symplectic integration schemes the methods of choice for long-term simulations. This has already been demonstrated by the symplectic PICKABACK method [19]. However, this method requires a relatively small step-size due to the high-frequency quantum modes. Therefore, following related ideas from classical molecular dynamics, we investigate symplectic multiple-time-stepping methods and indicate various possibilities to overcome the step-size limitation of PICKABACK.

1 Introduction

In this paper, we consider the symplectic integration of the so-called Quantum-Classical Molecular Dynamics (QCMD) model. In the QCMD model (see [11, 9, 2, 3, 6] and references therein), most atoms are described by classical mechanics, but an important small portion of the system by quantum mechanics. This leads to a coupled system of Newtonian and Schrödinger equations.

We focus on so-called symplectic methods [18] for the following reason: It has been shown that the preservation of the symplectic structure of phase space under a numerical integration scheme implies a number of very desirable properties. Namely,

- symplectic methods preserve the total energy over very (exponentially) long periods of time up to small fluctuations [2, 11, 14] and
- symplectic methods also conserve the adiabatic invariants of the problem under consideration [15].

Note that the same results have not been shown for symmetric (time-reversible) integration methods, although symmetric methods seem to perform quite well in practice. For a discussion of symmetric methods in the context of the QCMD model see [16, 17, 13].

For ease of presentation, we consider the case of just one quantum degree of freedom with spatial coordinate x and mass m and N classical particles with coordinates $\boldsymbol{q} \in \mathcal{R}^{3N}$ and diagonal mass matrix $\boldsymbol{M} \in \mathcal{R}^{3N \times 3N}$. Upon

denoting the interaction potential by $V(x, \boldsymbol{q})$, we obtain the following equations of motion for the QCMD model:

$$\begin{aligned} i\hbar \frac{\partial}{\partial t}\psi &= H(\boldsymbol{q})\,\psi\,, \\ \frac{\partial}{\partial t}\boldsymbol{q} &= \boldsymbol{M}^{-1}\boldsymbol{p}\,, \\ \frac{\partial}{\partial t}\boldsymbol{p} &= -\langle \psi, \nabla_{\boldsymbol{q}} V(\boldsymbol{q})\psi\rangle - \nabla_{\boldsymbol{q}} U_{cl}(\boldsymbol{q}) \end{aligned}$$

with U_{cl} a purely classical potential energy function and with $H(\boldsymbol{q})$ the quantum Hamiltonian operator given by

$$H(\boldsymbol{q}) = T + V(x, \boldsymbol{q}), \qquad T = -\frac{\hbar^2}{2m}\Delta_x.$$

In the sequel, we assume that the quantum subsystem has been truncated to a finite-dimensional system by an appropriate spatial discretization and a corresponding representation of the wave function ψ by a complex-valued vector $\boldsymbol{\psi} \in \mathcal{C}^d$. The discretized quantum operators T, V and H are denoted by $\boldsymbol{T} \in \mathcal{C}^{d\times d}$, $\boldsymbol{V}(\boldsymbol{q}) \in \mathcal{C}^{d\times d}$ and $\boldsymbol{H}(\boldsymbol{q}) \in \mathcal{C}^{d\times d}$, respectively. In the following construction of the time-propagators, we will exploit special matrix structures of some spatial discretizations:

a) $\boldsymbol{V}(\boldsymbol{q})$ is diagonal,
b) $\boldsymbol{H}(\boldsymbol{q})$ is real-valued, and
c) all other cases.

2 Conservation Properties of the QCMD Model

For long-term simulations, it generally proves advantageous to consider numerical integrators which pass the structural properties of the model onto the calculated solutions. Hence, a careful analysis of the conservation properties of QCMD model is required. A particularly relevant constant of motion of the QCMD model is the total energy of the system

$$\mathcal{H} = \frac{\boldsymbol{p}^T \boldsymbol{M}^{-1}\boldsymbol{p}}{2} + \boldsymbol{\psi}^* \boldsymbol{H}(\boldsymbol{q})\boldsymbol{\psi} + U_{cl}(\boldsymbol{q}). \tag{1}$$

Here $\boldsymbol{\psi}^*$ denotes the conjugate transpose of $\boldsymbol{\psi}$. Another conserved quantity is the norm of the vector $\boldsymbol{\psi}$, i.e., $\boldsymbol{\psi}^*\boldsymbol{\psi} = const.$ due to the unitary propagation of the quantum part.

In the context of this paper, the most important conservation property of QCMD is related to its canonical Hamiltonian structure which implies the symplecticness of the solution operator [1]. There are different ways to

consider the QCMD model as a canonical Hamiltonian system with Hamiltonian (1). Here we follow the presentation given in [3, 16]: We decompose the complex-valued vector ψ into its real and imaginary part, i.e.,

$$\psi = \frac{1}{\sqrt{2\hbar}}(q_\psi + ip_\psi)\,.$$

Then, after introducing generalized positions $\boldsymbol{Q} = (\boldsymbol{q}_\psi^T, \boldsymbol{q}^T)^T \in \mathcal{R}^{d+3N}$ and generalized momenta $\boldsymbol{P} = (\boldsymbol{p}_\psi^T, \boldsymbol{p}^T)^T \in \mathcal{R}^{d+3N}$, the equations of motion can be written as

$$\frac{d}{dt}\boldsymbol{Q} = +\nabla_{\boldsymbol{P}}\mathcal{H}(\boldsymbol{Q},\boldsymbol{P})\,,$$
$$\frac{d}{dt}\boldsymbol{P} = -\nabla_{\boldsymbol{Q}}\mathcal{H}(\boldsymbol{Q},\boldsymbol{P})\,.$$

These equations of motion are also time-reversible [13].

Finally, we like to mention that the QCMD model reduces to the Born-Oppenheimer approximation in case the ratio of the mass m of the quantum particles to the masses of the classical particles vanishes [6]. This implies that the populations $|\theta_i(t)|^2$, $i = 1, \ldots, k$, corresponding to the eigenvalues $E_i(\boldsymbol{q}(t)))$ of the operator $\boldsymbol{H}(\boldsymbol{q})$ become adiabatic invariants.

Note that the conservation of total energy and the conservation of the adiabatic invariants associated to the Born-Oppenheimer limit of the QCMD model provide a simple test for the behavior of a numerical integrator.

3 Construction of Symplectic Integrators

Our aim is the construction of numerical integrators which reproduce the conserved quantities in long-term simulations. To this end, we focus on symplectic methods, i.e., methods that conserve the canonical structure of phase space [18]. A convenient way to derive symplectic methods for general Hamiltonian systems is based on an appropriate splitting of the Hamiltonian $\mathcal{H}$ into a sum of sub-Hamiltonians, e.g., the two-term decomposition $\mathcal{H}_1 + \mathcal{H}_2$, each of which corresponds either to an explicitly solvable system or has a given symplectic integrator [18]. This procedure can be illustrated using a phase space representation of the Hamiltonian flow. The time-evolution over Δt units of time is then given by $\exp(\Delta t L_{\mathcal{H}})$ where $L_{\mathcal{H}}$ is the Liouville operator of the whole system [18, 19]. The Liouville operator $\exp(\Delta t L_{\mathcal{H}})$ can be approximated via the second order Strang splitting [18]:

$$\exp(\Delta t L_{\mathcal{H}}) = \exp(\frac{\Delta t}{2} L_{\mathcal{H}_1}) \exp(\Delta t L_{\mathcal{H}_2}) \exp(\frac{\Delta t}{2} L_{\mathcal{H}_1}) + \mathcal{O}(\Delta t^3). \quad (2)$$

The resulting numerical method is obviously symplectic since $\exp(\frac{\Delta t}{2} L_{\mathcal{H}_1})$ and $\exp(\Delta t L_{\mathcal{H}_2})$ are symplectic maps and the composition of symplectic maps yields a symplectic map.

The symplectic PICKABACK method [19], for instance, uses the following selection:

$$\mathcal{H}_1 = \frac{\boldsymbol{p}^T \boldsymbol{M}^{-1} \boldsymbol{p}}{2} + \boldsymbol{\psi}^* \boldsymbol{T} \boldsymbol{\psi} \qquad \text{and} \qquad \mathcal{H}_2 = \boldsymbol{\psi}^* \boldsymbol{V}(\boldsymbol{q}) \boldsymbol{\psi} + U_{cl}(\boldsymbol{q}) \,.$$

The corresponding differential equations can be solved explicitly provided the operator $\boldsymbol{V}(\boldsymbol{q})$ is diagonal.

PICKABACK conserves total energy up to small fluctuations and the norm of the vector $\boldsymbol{\psi}$ exactly. Its main drawback is the step-size restriction which is of the order of the inverse of the largest eigenvalue of the quantum operator $\boldsymbol{H}(\boldsymbol{q})$. Thus, if the evaluation of the operator $\boldsymbol{V}(\boldsymbol{q})$ and the gradients $\nabla_{\boldsymbol{q}} \boldsymbol{V}(\boldsymbol{q})$ and $\nabla_{\boldsymbol{q}} U_{cl}(\boldsymbol{q})$ are expensive due to long-range interactions, then the PICKABACK scheme can become inefficient, i.e., the permitted step-size might be much smaller than required by the pure classical dynamics. To overcome this problem, symmetric integration schemes are considered in [16, 17] and [13].

4 Symplectic Multiple-Time-Stepping Methods

Here we suggest a different approach that propagates the system using multiple step-sizes, i.e., few steps with step-size Δt are taken in the "slow" classical part whereas many smaller steps with step-size δt are taken in the highly oscillatory quantum subsystem (see, for example, [19, 4] for symplectic multiple-time-stepping methods in the context of classical molecular dynamics). Therefore, we consider a splitting of the Hamiltonian $\mathcal{H} = \mathcal{H}_1 + \mathcal{H}_2$ in the following way:

$$\mathcal{H}_1 = \frac{\boldsymbol{p}^T \boldsymbol{M}^{-1} \boldsymbol{p}}{2} \qquad \text{and} \qquad \mathcal{H}_2 = \boldsymbol{\psi}^* \boldsymbol{H}(\boldsymbol{q}) \boldsymbol{\psi} + U_{cl}(\boldsymbol{q}) \,.$$

Let us write down the corresponding differential equations. First for $\mathcal{H}_1$:

$$\begin{aligned} i\hbar \frac{d}{dt} \boldsymbol{\psi} &= \boldsymbol{0} \,, \\ \frac{d}{dt} \boldsymbol{q} &= \boldsymbol{M}^{-1} \boldsymbol{p} \,, \\ \frac{d}{dt} \boldsymbol{p} &= \boldsymbol{0} \,, \end{aligned}$$

next for $\mathcal{H}_2$:

$$i\hbar \frac{d}{dt} \boldsymbol{\psi} = \boldsymbol{H}(\boldsymbol{q}) \boldsymbol{\psi} \,, \tag{3}$$

$$\frac{d}{dt} \boldsymbol{q} = \boldsymbol{0} \,, \tag{4}$$

$$\frac{d}{dt} \boldsymbol{p} = -\boldsymbol{\psi}^* \nabla_{\boldsymbol{q}} \boldsymbol{V}(\boldsymbol{q}) \boldsymbol{\psi} - \nabla_{\boldsymbol{q}} U_{cl}(\boldsymbol{q}) \,. \tag{5}$$

The solution to $\mathcal{H}_1$ is just a translation of classical particles with constant momentum $\boldsymbol{p}$.

The intriguing point about the second set of equations is that $\boldsymbol{q}$ is now kept constant. Thus the vector $\boldsymbol{\psi}$ evolves according to a time-dependent Schrödinger equation with time-independent Hamilton operator $\boldsymbol{H}(\boldsymbol{q})$ and the update of the classical momentum $\boldsymbol{p}$ is obtained by integrating the Hellmann-Feynman forces [3] acting on the classical particles along the computed $\boldsymbol{\psi}(t)$ (plus a constant update due to the purely classical force field).

Upon computing the eigenvalues of the operator $\boldsymbol{H}(\boldsymbol{q})$, the equations (3)-(5) can be solved exactly. However, this is, in general, an expensive undertaken. Therefore we proceed with the following multiple-time-stepping approach: The first step is to consider the identity

$$\exp(\Delta t L_{\mathcal{H}_2}) = \underbrace{\exp(\delta t L_{\hat{\mathcal{H}}_2}) \cdots \exp(\delta t L_{\hat{\mathcal{H}}_2})}_{j \text{ times}} \exp(\Delta t L_{U_{cl}}),$$

where $\delta t = \Delta t/j$, $j \gg 1$, and

$$\hat{\mathcal{H}}_2 = \boldsymbol{\psi}^* \boldsymbol{H}(\boldsymbol{q}) \boldsymbol{\psi}.$$

The second step is to use this identity in (2) which yields

$$\exp(\Delta t L_{\mathcal{H}}) = \exp(\frac{\Delta t}{2} L_{\mathcal{H}_1}) \underbrace{\exp(\delta t L_{\hat{\mathcal{H}}_2}) \cdots \exp(\delta t L_{\mathcal{H}_2})}_{j \text{ times}} \cdot \cdot \exp(\Delta t L_{U_{cl}}) \exp(\frac{\Delta t}{2} L_{\mathcal{H}_1}) + \mathcal{O}(\Delta t^3).$$

The last step is to find a symplectic, second order approximation $\Phi_{\delta t}$ to $\exp(\delta t L_{\hat{\mathcal{H}}_2})$. In principle, we can use any symplectic integrator suitable for time-dependent Schrödinger equations (see, for example, [9]). Here we focus on the following three different possibilities corresponding to special properties of the spatially truncated operators $\boldsymbol{H}(\boldsymbol{q})$ and $\boldsymbol{V}(\boldsymbol{q})$.

a) Provided that $\boldsymbol{V}(\boldsymbol{q})$ *is diagonal*, an efficient method $\Phi_{\delta t}$ is obtained by exploiting the natural splitting of the quantum operator $\boldsymbol{H}(\boldsymbol{q}) = \boldsymbol{T} + \boldsymbol{V}(\boldsymbol{q})$ in a procedure similar to the one used in PICKABACK. This yields two exactly solvable subsystems [19]

$$\hat{\mathcal{H}}_{2,1} = \boldsymbol{\psi}^* \boldsymbol{T} \boldsymbol{\psi} \qquad \text{and} \qquad \hat{\mathcal{H}}_{2,2} = \boldsymbol{\psi}^* \boldsymbol{V}(\boldsymbol{q}) \boldsymbol{\psi}.$$

Again, we use (2) to construct a symplectic, second order approximation $\Phi_{\delta t}$ to $\exp(\delta t\, L_{\hat{\mathcal{H}}_2})$. The resulting integrator for QCMD is of second order,

explicit, symplectic, and conserves the norm of the wave-function:

$$
\begin{aligned}
\boldsymbol{q}_{1/2} &= \boldsymbol{q}_0 + \frac{\Delta t}{2}\boldsymbol{M}^{-1}\boldsymbol{p}_0\,, \\
\hat{\boldsymbol{p}}_0 &= \boldsymbol{p}_0 - \Delta t \nabla_{\boldsymbol{q}} U_{cl}(\boldsymbol{q}_{1/2})\,, \\
j \text{ times the} \\
\text{application of } \Phi_{\delta t}; \\
k = 1 \ldots j
\end{aligned}
\left\{
\begin{aligned}
\hat{\boldsymbol{\psi}}_{k/j} &= \exp\left(-\tfrac{i}{\hbar}\tfrac{\delta t}{2}\boldsymbol{T}\right)\boldsymbol{\psi}_{(k-1)/j} \\
\hat{\boldsymbol{p}}_{k/j} &= \hat{\boldsymbol{p}}_{(k-1)/j} - \delta t\, \hat{\boldsymbol{\psi}}^*_{k/j} \nabla_{\boldsymbol{q}} \boldsymbol{V}(\boldsymbol{q}_{1/2}) \hat{\boldsymbol{\psi}}_{k/j} \\
\boldsymbol{\psi}_{k/j} &= \exp\left(-\tfrac{i}{\hbar}\tfrac{\delta t}{2}\boldsymbol{T}\right)\exp\left(-\tfrac{i}{\hbar}\delta t\, \boldsymbol{V}(\boldsymbol{q}_{1/2})\right)\hat{\boldsymbol{\psi}}_{k/j}
\end{aligned}
\right.
$$

$$
\begin{aligned}
\boldsymbol{p}_1 &= \hat{\boldsymbol{p}}_1\,, \\
\boldsymbol{q}_1 &= \boldsymbol{q}_{1/2} + \frac{\Delta t}{2}\boldsymbol{M}^{-1}\boldsymbol{p}_1\,.
\end{aligned}
$$

b) If the spatially discretized quantum Hamiltonian operator $\boldsymbol{H}(\boldsymbol{q})$ is *real-valued*, i.e.,

$$
\mathcal{H} = \frac{1}{2\hbar}\boldsymbol{q}_\psi^T H(\boldsymbol{q})\boldsymbol{q}_\psi + \frac{1}{2\hbar}\boldsymbol{p}_\psi^T \boldsymbol{H}(\boldsymbol{q})\boldsymbol{p}_\psi + \frac{1}{2}\boldsymbol{p}^T\boldsymbol{M}^{-1}\boldsymbol{p} + U_{cl}(\boldsymbol{q})\,,
$$

then the Hamiltonian $\hat{\mathcal{H}}_2$ can be written as

$$
\hat{\mathcal{H}}_2 = \underbrace{\frac{1}{2\hbar}\boldsymbol{q}_\psi^T\boldsymbol{H}(\boldsymbol{q})\boldsymbol{q}_\psi}_{\hat{\mathcal{H}}_{2,1}} + \underbrace{\frac{1}{2\hbar}\boldsymbol{p}_\psi^T\boldsymbol{H}(\boldsymbol{q})\boldsymbol{p}_\psi}_{\hat{\mathcal{H}}_{2,2}}
$$

and the equations of motion corresponding to each of the two terms in the Hamiltonian $\hat{\mathcal{H}}_2$, namely

$$
\begin{aligned}
\hbar\frac{d}{dt}\boldsymbol{q}_\psi &= \boldsymbol{0}\,, \\
\hbar\frac{d}{dt}\boldsymbol{p}_\psi &= -\boldsymbol{H}(\boldsymbol{q})\boldsymbol{q}_\psi\,, \\
\frac{d}{dt}\boldsymbol{q} &= \boldsymbol{0}\,, \\
\frac{d}{dt}\boldsymbol{p} &= -\frac{1}{2\hbar}\boldsymbol{q}_\psi^T\nabla_{\boldsymbol{q}}\boldsymbol{V}(\boldsymbol{q})\boldsymbol{q}_\psi\,,
\end{aligned}
$$

and

$$
\begin{aligned}
\hbar\frac{d}{dt}\boldsymbol{q}_\psi &= \boldsymbol{H}(\boldsymbol{q})\boldsymbol{p}_\psi\,, \\
\hbar\frac{d}{dt}\boldsymbol{p}_\psi &= \boldsymbol{0}\,, \\
\frac{d}{dt}\boldsymbol{q} &= \boldsymbol{0}\,, \\
\frac{d}{dt}\boldsymbol{p} &= -\frac{1}{2\hbar}\boldsymbol{p}_\psi^T\nabla_{\boldsymbol{q}}\boldsymbol{V}(\boldsymbol{q})\boldsymbol{p}_\psi\,,
\end{aligned}
$$

can be solved analytically. Thus we define

$$\Phi_{\delta t} = \exp\left(\frac{\delta t}{2} L_{\hat{\mathcal{H}}_{2,1}}\right) \exp(\delta t L_{\hat{\mathcal{H}}_{2,2}}) \exp\left(\frac{\delta t}{2} L_{\hat{\mathcal{H}}_{2,1}}\right) .$$

For stability reasons, the micro-step-size δt has to be chosen smaller than the inverse of the largest eigenvalue of the (scaled) truncated quantum operator $\hbar^{-1}\boldsymbol{H}(\boldsymbol{q})$. This can imply a very small value of δt compared to the macro-step-size Δt.

c) The most straightforward but also an expensive $\Phi_{\delta t}$ is obtained by discretizing the equations of motion corresponding to $\hat{\mathcal{H}}_2$ by the (symplectic) implicit midpoint rule which results in

$$\begin{aligned} \psi_{k/j} &= \psi_{(k-1)/j} + \delta t\, \boldsymbol{H}(\boldsymbol{q_{1/2}})\psi_{(k-1/2)/j} , \\ \hat{\boldsymbol{p}}_{k/j} &= \hat{\boldsymbol{p}}_{(k-1)/j} - \delta t\, \psi^*_{(k-1/2)/j} \nabla_{\boldsymbol{q}} \boldsymbol{V}(\boldsymbol{q_{1/2}})\psi_{(k-1/2)/j} , \end{aligned}$$

$k = 1, \ldots, j$, with $\psi_{(k-1/2)/j} = (\psi_{k/j} + \psi_{(k-1)/j})/2$. Note that each integration step requires the solution of a d-dimensional linear system of equations in the unknown $\psi_{k/j}$.

Our multiple-time-stepping methods are close to methods suggested in [16, 17]. The method considered in [16] is time-reversible but not symplectic. More importantly, the method updates the momenta $\boldsymbol{p}$ of all classical particles only once per macro-time-step Δt. As indicated in [10, 13], this might lead to a substantial phase drift in the discrete solution. In [10], an averaging procedure of the quantum-classical Hellmann-Feynman force field along $\psi(t)$ is suggested to overcome this problem. Note that, for the multiple-time-stepping schemes suggested here, this averaging is carried out automatically and is a direct consequence of proposed splitting of the Hamiltonian equations of motion. We finally like to mention that symplectic methods are also discussed in [17]. In particular, the suggested methods are symplectic in the quantum part and the classical part if considered separately. However, this does not imply that the overall method is symplectic.

5 Conclusions

We have derived time-reversible, symplectic, and second-order multiple-time-stepping methods for the finite-dimensional QCMD model. Theoretical results for general symplectic methods imply that the methods conserve energy over exponentially long periods of time up to small fluctuations. Furthermore, in the limit $m \to 0$, the adiabatic invariants corresponding to the underlying Born-Oppenheimer approximation will be preserved as well. Finally, the phase shift observed for symmetric methods with a single update of the classical momenta $\boldsymbol{p}$ per macro-time-step Δt should be avoided by

the suggested methods. The additional costs for this frequent update per micro-time-step δt are relatively low. Note that the update only requires taking the inner product $\psi^* \nabla_q V(q) \psi$ with respect to a constant gradient $\nabla_q V(q)$ and only with respect to those classical particles that interact with the quantum degree of motion.

Acknowledgement. It is a pleasure to thank Christof Schütte for discussions on the subject of this paper.

References

1. A.I. Arnold. *Mathematical methods of classical mechanics.* Springer-Verlag, 1978.
2. G. Benettin and A. Giorgilli. On the Hamiltonian interpolation of near to the identity symplectic mappings with application to symplectic integration algorithms. *J. Statist. Phys.*, 74:1117–1143, 1994.
3. H.J.C. Berendsen and J. Mavri. Quantum simulation of reaction dynamics by density matrix evolution. *J. Phys. Chem.*, 97:13464–13468, 1993.
4. J.J. Biesiadecki and R.D. Skeel. Dangers of multiple-time-step methods. *J. Comput. Phys.*, 109:318–328, 1993.
5. F.A. Bornemann, P. Nettesheim, and Ch. Schütte. Quantum-classical molecular dynamics as an approximation for full quantum dynamics. *J. Chem. Phys.*, 105(3):1074–1083, 1996.
6. F.A. Bornemann and Ch. Schütte. On the singular limit of the quantum-classical molecular dynamics model. Preprint SC 97-07, ZIB Berlin, 1997. Submitted to SIAM J. Appl. Math.
7. A. García-Vela, R.B. Gerber, and D.G. Imre. Mixed quantum wave packet/classical trajectory treatment of the photodissociation process ArHCl → Ar+H+Cl. *J. Chem. Phys.*, 97:7242–7250, 1992.
8. R.B. Gerber, V. Buch, and M.A. Ratner. Time-dependent self-consistent field approximation for intramolecular energy transfer. *J. Chem. Phys.*, 66:3022–3030, 1982.
9. S.K. Gray and D.E. Manolopoulos. Symplectic integrators tailored to the time-dependent Schrödinger equation. *J. Chem. Phys.*, 104:7099–7112, 1996.
10. M. Hochbruck and Ch. Lubich. A bunch of time integrators for quantum/classical molecular dynamics. this volume.
11. E. Hairer and Ch. Lubich. The life-span of backward error analysis for numerical integrators. *Numer. Math.*, 76:441–462, 1997.
12. P. Nettesheim, F.A. Bornemann, B. Schmidt, and Ch. Schütte. An explicit and symplectic integrator for quantum-classical molecular dynamics. *Chemical Physics Letters*, 256:581–588, 1996.
13. P. Nettesheim and Ch. Schütte. Numerical integrators for quantum-classical molecular dynamics. this volume.
14. S. Reich. Backward error analysis for numerical integrators. *SIAM J. Numer. Anal.*, to appear, 1999.
15. S. Reich. Preservation of adiabatic invariants under symplectic discretization. *Applied Numerical Mathematics*, to appear, 1998.

16. U. Schmitt and J. Brinkmann. Discrete time-reversible propagation scheme for mixed quantum classical dynamics. *Chem. Phys.*, 208:45–56, 1996.
17. U. Schmitt. Gemischt klassisch-quantenmechanische Molekulardynamik im Liouville-Formalismus. Ph.D. thesis (in german), Darmstadt, 1997.
18. J.M. Sanz-Serna and M.P. Calvo. Numerical Hamiltonian Systems. Chapman and Hall, London, 1994.
19. M. Tuckerman, B.J. Berne, and G. Martyna. Reversible Multiple Time Scale Molecular Dynamics. *J. Chem. Phys.*, 97:1990–2001, 1992.

A Bunch of Time Integrators for Quantum/Classical Molecular Dynamics

Marlis Hochbruck and Christian Lubich

Mathematisches Institut, Universität Tübingen, Auf der Morgenstelle 10,
D–72076 Tübingen, Germany.
E-mail: marlis@@na.uni-tuebingen.de, lubich@@na.uni-tuebingen.de

Abstract. We present novel time integration schemes for Newtonian dynamics whose fastest oscillations are nearly harmonic, for constrained Newtonian dynamics including the Car-Parrinello equations of *ab initio* molecular dynamics, and for mixed quantum-classical molecular dynamics. The methods attain favorable properties by using matrix-function vector products which are computed via Lanczos' method. This permits to take longer time steps than in standard integrators.

1 Introduction

In this paper we present a number of time integrators for various problems ranging from classical to quantum molecular dynamics. These integrators share some common features: they are new, they are second-order accurate and time-reversible, they improve substantially over standard schemes in well-defined model situations — and none of them has been tested on real applications at the time of this writing. This last feature will hopefully change in the near future [20].

On a more technical level, a further common feature of all the schemes proposed here is that they require the computation of the product of an analytic function of a symmetric matrix with a vector. Integration schemes employing matrix functions apparently have not hitherto been used in practice, except in a few special cases where direct diagonalization is possible. However, since the mid-eighties, starting with a paper by Park and Light [21] on quantum propagators, Lanczos' method has been put to good use in approximating matrix-function vector products. More recently, the excellent convergence properties of this approach have been clarified in [7, 12]. This motivated the development of the general-purpose ODE integrator `exp4` in [15]. Employing matrix functions in an integrator enables us to obtain favorable properties, such as exact integration of linear differential systems with constant coefficients. This is very advantageous for problems where the fastest oscillations are nearly harmonic. Matrix functions add a welcome element to the construction of integration schemes. The freedom thus gained can be used for designing new integrators tailored to specific applications, as we try to demonstrate with the methods in this paper. A theoretical error analysis of these schemes is given in [13, 14]. We hope that the proposed integrators, or closely related ones, will be found useful in molecular dynamics.

2 Newtonian Equations of Motion

In this section we consider the classical equations of motion of particles in cases where the highest-frequency oscillations are nearly harmonic: The positions $\mathbf{y}(t) = \{y_i(t)\}$ evolve according to the second-order system of differential equations

$$\mathbf{M}\ddot{\mathbf{y}} = -\mathbf{A}\mathbf{y} + \mathbf{f}(\mathbf{y}). \tag{1}$$

Here, $\mathbf{M}$ is a constant, symmetric positive definite mass matrix. We assume without loss of generality that $\mathbf{M}$ is simply the identity matrix $\mathbf{I}$. Otherwise, this is achieved by the familiar transformation

$$\mathbf{y} \mapsto \mathbf{M}^{1/2}\mathbf{y}, \quad \mathbf{A} \mapsto \mathbf{M}^{-1/2}\mathbf{A}\mathbf{M}^{-1/2}, \quad \mathbf{f}(\mathbf{y}) \mapsto \mathbf{M}^{-1/2}\mathbf{f}(\mathbf{M}^{-1/2}\mathbf{y}).$$

We assume that $\mathbf{A}$ is a symmetric and positive semi-definite matrix. The case of interest is when the largest eigenvalue of $\mathbf{A}$ is significantly larger than the norm of the derivative of the nonlinear force $\mathbf{f}$. $\mathbf{A}$ may be a constant matrix, or else $\mathbf{A} = \mathbf{A}(\mathbf{y})$ is assumed to be slowly changing along solution trajectories, in which case $\mathbf{A}$ will be evaluated at the current averaged position in the numerical schemes below. In the standard Verlet scheme, which yields approximations $\mathbf{y}^n$ to $\mathbf{y}(n\Delta t)$ via

$$\mathbf{y}^{n+1} - 2\mathbf{y}^n + \mathbf{y}^{n-1} = \Delta t^2(-\mathbf{A}\mathbf{y}^n + \mathbf{f}(\mathbf{y}^n)) \,, \tag{2}$$

the time step Δt is then limited by the inverse of the largest eigenvalue of $\mathbf{A}$. We are aiming at methods whose step size is restricted only by the nonlinearity $\mathbf{f}$. To motivate the derivation of such a scheme, we start from the linear equation with constant force vector

$$\ddot{\mathbf{y}} = -\mathbf{A}\mathbf{y} + \mathbf{f} \qquad (\mathbf{A}, \mathbf{f} \text{ constant}) \tag{3}$$

whose exact solution satisfies

$$\mathbf{y}(t + \Delta t) - 2\mathbf{y}(t) + \mathbf{y}(t - \Delta t) = \Delta t^2 \, \sigma(\Delta t^2 \mathbf{A}) \, (-\mathbf{A}\mathbf{y}(t) + \mathbf{f})$$

with the complex function

$$\sigma(z) = \left(\frac{\sin(\sqrt{z}/2)}{\sqrt{z}/2} \right)^2 .$$

Approximating the nonlinear force $\mathbf{f}(\mathbf{y})$ over a time step by a suitable constant vector leads to a scheme whose origins for scalar equations can be traced back to [10]:

$$\mathbf{y}^{n+1} - 2\mathbf{y}^n + \mathbf{y}^{n-1} = \Delta t^2 \, \sigma(\Delta t^2 \mathbf{A}) \, (-\mathbf{A}\mathbf{y}^n + \mathbf{f}^n) \tag{4}$$

with

$$\mathbf{f}^n = \mathbf{f}(\bar{\mathbf{y}}^n) \ . \tag{5}$$

Here, the obvious choice would be $\bar{\mathbf{y}}^n = \mathbf{y}^n$. However, as was first proposed in [8] for a different method described at the end of this section, it turns out to be favorable to take an averaged position value

$$\bar{\mathbf{y}}^n = \phi(\Delta t^2 \mathbf{A})\, \mathbf{y}^n \tag{6}$$

with a suitably chosen filter function $\phi(z)$, or eventually a translated version of (6), with some reference position $\mathbf{y}^*$,

$$\bar{\mathbf{y}}^n = \mathbf{y}^* + \phi(\Delta t^2 \mathbf{A})\, (\mathbf{y}^n - \mathbf{y}^*) \ .$$

The choice of ϕ strongly influences the mixing of frequencies by the nonlinearity. Two possible choices are

$$\phi(z) = \phi_0(z) = \frac{\sin \sqrt{z}}{\sqrt{z}} \ , \tag{7}$$

which is a filter function suggested in [8], or preferably

$$\phi(z) = \phi_1(z) = \left(1 + \tfrac{1}{6}(1 - \cos \sqrt{z})\right) \phi_0(z) \ . \tag{8}$$

Of course, to make the scheme (4) practical, we must be able to compute the products of matrix functions $\sigma(\Delta t^2 \mathbf{A})$ and $\phi(\Delta t^2 \mathbf{A})$ with vectors efficiently. This will be discussed in Section 5.

From the derivation of the method (4) it is obvious that the scheme is exact for constant-coefficient linear problems (3). Like the Verlet scheme, it is also time-reversible. For the special case $\mathbf{A} = \mathbf{0}$ it reduces to the Verlet scheme. It is shown in [13] that the method has an $O(\Delta t^2)$ error bound over finite time intervals for systems with bounded energy. In contrast to the Verlet scheme, this error bound is independent of the size of the eigenvalues λ_k of $\mathbf{A}$.

It turns out that the error is essentially determined by a two-dimensional scalar function $\epsilon(x, y)$ evaluated at $x = \omega_k \Delta t$, $y = \omega_l \Delta t$, where $\omega_k = \sqrt{\lambda_k}$ are the frequencies of the linear system (3). This error function depends strongly on the choice of ϕ. The most obvious choice $\phi \equiv 1$ suffers from resonances, which express themselves in singularities of ϵ at arguments that are integer multiples of π. This leads to a loss of accuracy in resonance situations. The same is true for the choice $\phi = \sigma$. For ϕ given by (7) and (8) no such problems occur. The choice (8) has the advantage that it drastically reduces the error function $|\epsilon(x, y)|$ for small x, y and therefore gives better accuracy. See [13] for more details on the error analysis.

A widely used variant of the Verlet scheme is its velocity version:

$$\begin{aligned} \mathbf{v}^{n+1/2} &= \mathbf{v}^n + \frac{\Delta t}{2}(-\mathbf{A}\mathbf{y}^n + \mathbf{f}^n) \\ \mathbf{y}^{n+1} &= \mathbf{y}^n + \Delta t\, \mathbf{v}^{n+1/2} \\ \mathbf{v}^{n+1} &= \mathbf{v}^{n+1/2} + \frac{\Delta t}{2}(-\mathbf{A}\mathbf{y}^{n+1} + \mathbf{f}^{n+1}) \end{aligned} \tag{9}$$

with $\mathbf{f}^n = \mathbf{f}(\mathbf{y}^n)$. Similarly, also the scheme (4) admits a velocity version, which this time works with averaged velocities

$$\mathbf{v}^n \approx \phi_0(\Delta t^2 \mathbf{A})\, \dot{\mathbf{y}}(n\Delta t) \approx \frac{1}{2\Delta t} \int_{-\Delta t}^{\Delta t} \dot{\mathbf{y}}(n\Delta t + \tau)\, d\tau.$$

This one-step version of (4) reads (for $\mathbf{f}^n$ of (5) with $\bar{\mathbf{y}}^n$ of (6))

$$\begin{aligned} \mathbf{v}^{n+1/2} &= \mathbf{v}^n + \frac{\Delta t}{2}\sigma(\Delta t^2 \mathbf{A})\,(-\mathbf{A}\mathbf{y}^n + \mathbf{f}^n) \\ \mathbf{y}^{n+1} &= \mathbf{y}^n + \Delta t\, \mathbf{v}^{n+1/2} \\ \mathbf{v}^{n+1} &= \mathbf{v}^{n+1/2} + \frac{\Delta t}{2}\sigma(\Delta t^2 \mathbf{A})\,(-\mathbf{A}\mathbf{y}^{n+1} + \mathbf{f}^{n+1})\ . \end{aligned} \tag{10}$$

A different "long-time-step method" was previously proposed by García-Archilla, Sanz-Serna, and Skeel [8]. Their *mollified impulse method*, which is based on the concept of operator splitting and also reduces to the Verlet scheme for $\mathbf{A} = \mathbf{0}$ and admits second-order error estimates independently of the frequencies of $\mathbf{A}$, reads as follows when applied to (1):

$$\begin{aligned} \mathbf{v}_+^n &= \mathbf{v}^n + \frac{\Delta t}{2}\phi(\Delta t^2 \mathbf{A})\, \mathbf{f}^n \\ \begin{pmatrix} \mathbf{y}^{n+1} \\ \mathbf{v}_-^{n+1} \end{pmatrix} &= \exp\left(\Delta t \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{A} & \mathbf{0} \end{bmatrix} \right) \begin{pmatrix} \mathbf{y}^n \\ \mathbf{v}_+^n \end{pmatrix} \\ \mathbf{v}^{n+1} &= \mathbf{v}_-^{n+1} + \frac{\Delta t}{2}\phi(\Delta t^2 \mathbf{A})\, \mathbf{f}^{n+1} \end{aligned} \tag{11}$$

where $\mathbf{f}^n$ is again given by (5) and (6). Eliminating the (non-averaged) velocities, this scheme can be shown to become

$$\mathbf{y}^{n+1} - 2\mathbf{y}^n + \mathbf{y}^{n-1} = \Delta t^2 \sigma(\Delta t^2 \mathbf{A})\,(-\mathbf{A}\mathbf{y}^n + \mathbf{f}^n) + \Delta t^2 \delta(\Delta t^2 \mathbf{A})\, \mathbf{f}^n\ ,$$

where $\delta(z) = \phi(z)\,\phi_0(z) - \sigma(z)$. A comparison of theoretical properties of (10) and (11) is given in [13].

3 Car-Parrinello Equations of *Ab Initio* Molecular Dynamics, Constrained Newtonian Dynamics

In the Car-Parrinello method [6] (and see, e.g., [24, 25, 16, 4]), the adiabatic time-dependent Born-Oppenheimer model is approximated by a fictitious Newtonian dynamics in which the electrons, represented by a set of

wave functions $\{|\psi_j\rangle\}$, follow the motion of the ions, represented by a set of positions $\{\mathbf{R}_J\}$. The model involves a small fictitious-mass parameter μ which keeps the fictitious kinetic energy of the electrons small compared to the kinetic energy of the ions. The dynamics is determined by the Kohn-Sham energy functional $E(\{\mathbf{R}_J\}, \{|\psi_j\rangle\})$, and by the orthonormality constraint for the orbitals. The Euler-Lagrange equations of motion then read, with Lagrange multipliers Λ_{jk},

$$\begin{aligned} M_J \ddot{\mathbf{R}}_J &= -\frac{\partial E}{\partial \mathbf{R}_J} \\ \mu |\ddot{\psi}_j\rangle &= -\frac{\partial E}{\partial |\psi_j\rangle} + \sum_k \Lambda_{jk} |\psi_k\rangle \\ \langle \psi_j | \psi_k \rangle &= \delta_{jk} \ . \end{aligned}$$

After spatial (spectral) discretization, this becomes a large finite-dimensional system in the evolution variables $\mathbf{y} = [\{\mathbf{R}_J\}, \{\psi_j\}]$ of the form

$$\begin{aligned} \ddot{\mathbf{y}} &= -\mathbf{A}\mathbf{y} + \mathbf{f}(\mathbf{y}) + \mathbf{G}^T(\mathbf{y})\,\boldsymbol{\lambda} \\ \mathbf{g}(\mathbf{y}) &= \mathbf{0} \end{aligned} \tag{12}$$

where $\mathbf{G} = \partial \mathbf{g}/\partial \mathbf{y}$. The splitting of the forces might be such that $\mathbf{A}$ represents the discretized Laplacian contained in the Kohn-Sham Hamiltonian, or, e.g., $\mathbf{A} = \mathbf{A}(\mathbf{y})$ contains in addition the Jacobian of the local interaction forces. The choice of the most effective splitting may depend on physical insight into the specific problem at hand, and quite probably on extensive numerical experiments.

The standard numerical integrators for the constrained system (12) are the SHAKE scheme [23], which extends the Verlet method (2),

$$\begin{aligned} \mathbf{y}^{n+1} - 2\mathbf{y}^n + \mathbf{y}^{n-1} &= \Delta t^2 \left(-\mathbf{A}\mathbf{y}^n + \mathbf{f}(\mathbf{y}^n) + \mathbf{G}^T(\mathbf{y}^n)\,\boldsymbol{\lambda}^n\right) \\ \mathbf{g}(\mathbf{y}^{n+1}) &= \mathbf{0} \end{aligned}$$

and the RATTLE scheme [1], which extends the velocity Verlet method (9). There, one first solves for $\mathbf{y}^{n+1}$ in

$$\begin{aligned} \mathbf{v}^{n+1/2} &= \mathbf{v}^n + \frac{\Delta t}{2}\left(-\mathbf{A}\mathbf{y}^n + \mathbf{f}(\mathbf{y}^n) + \mathbf{G}^T(\mathbf{y}^n)\,\boldsymbol{\lambda}^n\right) \\ \mathbf{y}^{n+1} &= \mathbf{y}^n + \Delta t\, \mathbf{v}^{n+1/2} \\ \mathbf{g}(\mathbf{y}^{n+1}) &= \mathbf{0} \end{aligned}$$

and subsequently for $\mathbf{v}^{n+1}$ in

$$\begin{aligned} \mathbf{v}^{n+1} &= \mathbf{v}^{n+1/2} + \frac{\Delta t}{2}\left(-\mathbf{A}\mathbf{y}^{n+1} + \mathbf{f}(\mathbf{y}^{n+1}) + \mathbf{G}^T(\mathbf{y}^{n+1})\,\widehat{\boldsymbol{\lambda}}^{n+1}\right) \\ \mathbf{G}(\mathbf{y}^{n+1})\mathbf{v}^{n+1} &= \mathbf{0} \ . \end{aligned}$$

In the same ways, the scheme (4) extends to

$$\begin{aligned} \mathbf{y}^{n+1} - 2\mathbf{y}^n + \mathbf{y}^{n-1} &= \Delta t^2\, \sigma(\Delta t^2 \mathbf{A})\, \left(-\mathbf{A}\mathbf{y}^n + \mathbf{f}(\bar{\mathbf{y}}^n) + \mathbf{G}^T(\bar{\mathbf{y}}^n)\, \boldsymbol{\lambda}^n\right) \\ \mathbf{g}(\mathbf{y}^{n+1}) &= \mathbf{0}\,, \end{aligned} \tag{13}$$

and its averaged-velocity version (10) extends to

$$\begin{aligned} \mathbf{v}^{n+1/2} &= \mathbf{v}^n + \frac{\Delta t}{2}\, \sigma(\Delta t^2 \mathbf{A})\, \left(-\mathbf{A}\mathbf{y}^n + \mathbf{f}(\bar{\mathbf{y}}^n) + \mathbf{G}^T(\bar{\mathbf{y}}^n)\, \boldsymbol{\lambda}^n\right) \\ \mathbf{y}^{n+1} &= \mathbf{y}^n + \Delta t\, \mathbf{v}^{n+1/2} \\ \mathbf{g}(\mathbf{y}^{n+1}) &= \mathbf{0} \end{aligned} \tag{14}$$

in the first half-step, and

$$\begin{aligned} \mathbf{v}^{n+1} &= \mathbf{v}^{n+1/2} + \frac{\Delta t}{2}\, \sigma(\Delta t^2 \mathbf{A})\, \left(-\mathbf{A}\mathbf{y}^{n+1} + \mathbf{f}(\bar{\mathbf{y}}^{n+1}) + \mathbf{G}^T(\bar{\mathbf{y}}^{n+1})\, \widehat{\boldsymbol{\lambda}}^{n+1}\right) \\ \mathbf{G}(\bar{\mathbf{y}}^{n+1})\mathbf{v}^{n+1} &= \mathbf{0} \end{aligned} \tag{15}$$

in the second half-step. Here $\bar{\mathbf{y}}^n$ is again defined as in (6).

Both methods are time-reversible. For $\mathbf{A} = \mathbf{0}$, they reduce to SHAKE and RATTLE. In contrast to SHAKE and RATTLE, the time step is not restricted by the largest eigenvalue of $\mathbf{A}$.

4 Quantum-Classical Molecular Dynamics

In the mixed quantum-classical molecular dynamics (QCMD) model (see [11, 9, 2, 3, 5] and references therein), most atoms are described by classical mechanics, but an important small portion of the system by quantum mechanics. The full quantum system is first separated via a tensor product ansatz. The evolution of each part is then modeled either classically or quantally. This leads to a coupled system of Newtonian and Schrödinger equations.

For ease of presentation only, we here consider the case of two particles having spatial coordinates $\mathbf{x}$ and $\mathbf{y}$, and masses m and M, with $m \ll M$. With the interaction potential $V(\mathbf{x}, \mathbf{y})$, the quantum Hamiltonian H is given by

$$H(\mathbf{y}) = -\frac{\hbar^2}{2m} \Delta_{\mathbf{x}} + V(\mathbf{x}, \mathbf{y})\,.$$

The equations of motion of the QCMD model read

$$\begin{aligned} M\ddot{\mathbf{y}} &= -\langle \psi | H'(\mathbf{y}) | \psi \rangle \\ i\hbar\, |\dot{\psi}\rangle &= H(\mathbf{y})\, |\psi\rangle \end{aligned}$$

where $H' = \nabla_{\mathbf{y}} H = \nabla_{\mathbf{y}} V$. After spatial discretization in $\mathbf{x}$, we obtain a large finite-dimensional system

$$\begin{aligned} M\ddot{\mathbf{y}} &= -\boldsymbol{\psi}^*\, \mathbf{H}'(\mathbf{y})\, \boldsymbol{\psi} \\ i\hbar\, \dot{\boldsymbol{\psi}} &= \mathbf{H}(\mathbf{y}) \boldsymbol{\psi} \end{aligned} \tag{16}$$

with the matrix $\mathbf{H}(\mathbf{y}) = \mathbf{T} + \mathbf{V}(\mathbf{y})$, where both $\mathbf{T}$ and $\mathbf{V}(\mathbf{y})$ are real and symmetric, and $\mathbf{T}$ is positive semi-definite. $\mathbf{T}$ is a constant matrix (the discretized Laplacian) of which we assume no bounds. On the other hand, $\mathbf{V}(\mathbf{y})$ and its first two derivatives are assumed to be moderately bounded. Typically, $\mathbf{T}$ is transformed to diagonal form by discrete Fourier transforms, and $\mathbf{V}(\mathbf{y})$ is a diagonal matrix.

A quite successful integrator for (16), which is based on the idea of operator splitting, is the PICKABACK scheme of Nettesheim et al. [19]. With the shorthand notation

$$\mathrm{i} = i/\hbar \ ,$$

their scheme reads

$$\begin{aligned}
\mathbf{y}^{n+1/2} &= \mathbf{y}^n + \frac{\Delta t}{2}\mathbf{v}^n \\
\boldsymbol{\psi}^{n+1/2} &= \exp\left(-\mathrm{i}\,\frac{\Delta t}{2}\,\mathbf{T}\right)\boldsymbol{\psi}^n \\
\mathbf{v}^{n+1} &= \mathbf{v}^n - \frac{\Delta t}{M}\,(\boldsymbol{\psi}^{n+1/2})^*\,\mathbf{H}'(\mathbf{y}^{n+1/2})\,\boldsymbol{\psi}^{n+1/2} \\
\boldsymbol{\psi}^{n+1} &= \exp\left(-\mathrm{i}\,\frac{\Delta t}{2}\,\mathbf{T}\right)\exp\left(-\mathrm{i}\,\Delta t\,\mathbf{V}(\mathbf{y}^{n+1/2})\right)\boldsymbol{\psi}^{n+1/2} \\
\mathbf{y}^{n+1} &= \mathbf{y}^{n+1/2} + \frac{\Delta t}{2}\mathbf{v}^{n+1} \ .
\end{aligned} \tag{17}$$

This scheme requires the exponential only of matrices that are diagonal or transformed to diagonal form by fast Fourier transforms. Unfortunately, this matrix splitting leads to time step restrictions of the order of the inverse of the largest eigenvalue of $\mathbf{T}/\hbar$. A simple, Verlet-like scheme that uses no matrix splitting, is the following:

$$\mathbf{y}^{n+1} - 2\mathbf{y}^n + \mathbf{y}^{n-1} = -\frac{\Delta t^2}{M}\,(\boldsymbol{\psi}^n)^*\,\mathbf{H}'(\mathbf{y}^n)\,\boldsymbol{\psi}^n \tag{18}$$

$$\boldsymbol{\psi}^n = \exp\left(-\mathrm{i}\,\frac{\Delta t}{2}\,\mathbf{H}^n\right)\boldsymbol{\psi}^{n-1/2} \tag{19}$$

$$\boldsymbol{\psi}^{n+1/2} = \exp\left(-\mathrm{i}\,\frac{\Delta t}{2}\,\mathbf{H}^n\right)\boldsymbol{\psi}^n = \exp\left(-\mathrm{i}\,\Delta t\,\mathbf{H}^n\right)\boldsymbol{\psi}^{n-1/2} \tag{20}$$

with $\mathbf{H}^n = \mathbf{H}(\mathbf{y}^n)$. We note that here (and in the schemes to follow) the $\mathbf{y}$-recursion could be rewritten in the velocity form (9). The action of the exponential is now approximated by the Lanczos method described in Section 5, with a Krylov subspace corresponding to the matrix $\mathbf{H}^n$ and the vector $\boldsymbol{\psi}^{n-1/2}$. Like (17), the scheme (18)–(20) is time-reversible, is unitary in the quantum part, evaluates the Hamiltonian and its gradient at the same position, and has formal order of accuracy 2. However, second-order error

bounds independently of the highest frequencies can be obtained only under unreasonable smoothness assumptions about the solution.

In the following we devise, following [14], an efficiently implementable scheme which leads to favorable error bounds independently of the highest frequencies under the mere assumption that the system has bounded energy. The scheme will be time-reversible, and robust in the singular limit of the mass ratio m/M tending to 0.

We first deal with the $\mathbf{y}$-equation. We start from the identity

$$\mathbf{y}(t+\Delta t) - 2\mathbf{y}(t) + \mathbf{y}(t-\Delta t) = \int_0^{\Delta t} (\Delta t - \tau)\,(\ddot{\mathbf{y}}(t+\tau) + \ddot{\mathbf{y}}(t-\tau))\,d\tau \ ,$$

which leads us to replace the pointwise force evaluation of (18) by

$$\begin{aligned} \mathbf{y}^{n+1} - 2\mathbf{y}^n + \mathbf{y}^{n-1} &= \frac{1}{M}\int_0^{\Delta t} (\Delta t - \tau)\,(\mathbf{f}^n(\tau) + \mathbf{f}^n(-\tau))\,d\tau \\ \mathbf{f}^n(\tau) &= -\boldsymbol{\phi}^n(\tau)^*\,\mathbf{H}'(\mathbf{y}^n)\,\boldsymbol{\phi}^n(\tau) \\ \boldsymbol{\phi}^n(\tau) &= \exp\left(-\mathrm{i}\tau\mathbf{H}^n\right)\boldsymbol{\psi}^n \ , \end{aligned} \tag{21}$$

where $\boldsymbol{\psi}^n$ can be determined from (19)–(20). Section 5 shows that the scheme (21) can still be implemented efficiently using Lanczos' method, at the expense of computing a number of inner products in addition to (18). The accuracy of this scheme is limited by the recursion for $\boldsymbol{\psi}$.

With $\dot{\mathbf{H}}^n = \mathbf{H}'(\mathbf{y}^n)\mathbf{v}^n$ and $\mathbf{v}^n = (\mathbf{y}^{n+1} - \mathbf{y}^{n-1})/(2\Delta t)$, we define the Hermitian matrix

$$\mathbf{J}^n = \int_0^{1/2} \exp(-\mathrm{i}\theta\Delta t\mathbf{H}^n)\,\theta\dot{\mathbf{H}}^n\,\exp(\mathrm{i}\theta\Delta t\mathbf{H}^n)\,d\theta \tag{22}$$

and consider the following symmetric and norm-preserving scheme:

$$\begin{aligned} \boldsymbol{\psi}^n &= \exp\left(\mathrm{i}\Delta t^2\mathbf{J}^n\right)\exp\left(-\mathrm{i}\,\frac{\Delta t}{2}\,\mathbf{H}^n\right)\boldsymbol{\psi}^{n-1/2} \ , \\ \boldsymbol{\psi}^{n+1/2} &= \exp\left(-\mathrm{i}\,\frac{\Delta t}{2}\,\mathbf{H}^n\right)\exp\left(-\mathrm{i}\Delta t^2\overline{\mathbf{J}}^n\right)\boldsymbol{\psi}^n \ . \end{aligned} \tag{23}$$

This formula can be motivated by applying the variation of constants formula to $i\hbar\dot{\boldsymbol{\psi}} = \mathbf{H}^n\boldsymbol{\psi} + (\mathbf{H}(\mathbf{y}) - \mathbf{H}^n)\boldsymbol{\psi}$. The method can be implemented using Chebyshev approximations to the exponential, see [14].

For the combined scheme (21), (23), second-order error bounds are derived in [14]. These bounds hold independently of the size of the eigenvalues of $\mathbf{T}$, and without assumptions about the smoothness of the solution, which in general is highly oscillatory.

For the system (16) it is known [5] under non-resonance assumptions that in the limit $m/M \to 0$ the motion of the classical particle is governed

by the Born-Oppenheimer potential (more precisely, by its approximation corresponding to the space discretization (16)),

$$M\ddot{\mathbf{y}} = -\nabla_{\mathbf{y}} U_{BO}(\mathbf{y}) \ .$$

Letting $m/M \to 0$ in the numerical method, it can be shown that the solution given by (21) tends to a small perturbation of the Verlet method formally applied to that equation:

$$\mathbf{y}^{n+1} - 2\mathbf{y}^n + \mathbf{y}^{n-1} = \frac{\Delta t^2}{M} \left(-\nabla_{\mathbf{y}} U_{BO}(\mathbf{y}^n) + O(\Delta t^2) \right) \ .$$

This is apparently not true for (18) and similar schemes with pointwise force evaluation.

Recently, there has been an effort to correct the tensor-product ansatz underlying the QCMD and many other models. In the Configuration Interaction extension of the Classical Separable Potential approach [17], Jungwirth and Gerber use first classical trajectories to guide the selection of important correlation terms and to simplify multidimensional integral evaluations for a subsequent quantum propagation. That approach requires the numerical solution of a large number of 1-dimensional Schrödinger equations with time-dependent Hamiltonian,

$$i\hbar\,\dot{\psi} = \mathbf{H}(t)\psi \ .$$

These can again be solved by the scheme (23), interpreting now $\mathbf{H}^n = \mathbf{H}(t^n)$ and $\dot{\mathbf{H}}^n = \dot{\mathbf{H}}(t^n)$ in (23) and (22).

5 Implementation Using Lanczos' Method

A common ingredient of all the schemes proposed in this article is the computation of the product of an entire function of a real symmetric matrix with a vector, $\varphi(\mathbf{S})\mathbf{b}$. In the various schemes, this is needed for $\varphi(z) = \sigma(\Delta t^2 z)$, $\phi(\Delta t^2 z)$, or $e^{-\mathrm{i}\Delta t z}$, and for $\mathbf{S} = \mathbf{A}$ or $\mathbf{S} = \mathbf{H}$. Computing $\varphi(\mathbf{S})\mathbf{b}$ is an easy task when the eigendecomposition of $\mathbf{S}$ is available, e. g., when $\mathbf{S}$ is a spectrally discretized Laplacian. However, computing the eigendecomposition of a general symmetric matrix of large dimension is prohibitive because of memory requirements and computational effort, in particular so when a different matrix occurs in every time step.

We here describe the alternative of approximating $\varphi(\mathbf{S})\mathbf{b}$ via Lanczos' method. The Lanczos process [18, 22] recursively generates an orthonormal basis $\mathbf{Q}_m = [\mathbf{q}_1, \ldots, \mathbf{q}_m]$ of the mth Krylov subspace $\mathrm{span}\{\mathbf{b}, \mathbf{S}\mathbf{b}, \ldots, \mathbf{S}^{m-1}\mathbf{b}\}$ such that

$$\mathbf{S}\mathbf{Q}_m = \mathbf{Q}_m \mathbf{L}_m + \beta_m [\mathbf{0} \ \cdots \ \mathbf{0} \ \mathbf{q}_{m+1}]$$

with the $m \times m$ tridiagonal Lanczos matrix

$$\mathbf{L}_m = \mathbf{Q}_m^T \mathbf{S} \mathbf{Q}_m.$$

This construction requires one matrix-vector multiplication with $\mathbf{S}$ and two inner products in each recursive step. Therefore, it is not necessary to store $\mathbf{S}$ explicitly as a matrix. The Lanczos process yields the approximation [21, 7, 12]

$$\varphi(\mathbf{S})\mathbf{b} \approx \mathbf{Q}_m \varphi(\mathbf{L}_m) \mathbf{Q}_m^T \mathbf{b},$$

where we note that $\mathbf{Q}_m^T \mathbf{b} = \|\mathbf{b}\| \cdot [1\ 0\ \cdots\ 0]^T$. A robust and inexpensive stopping criterion for the Lanczos iteration based on a generalized residual is described in [15]. Convergence properties are studied in [7, 12]. Since m is typically very small compared to the dimension of $\mathbf{S}$, $\varphi(\mathbf{L}_m)$ can easily be computed by diagonalization of $\mathbf{L}_m$,

$$\mathbf{L}_m = \mathbf{U}_m \mathbf{D}_m \mathbf{U}_m^T$$

with an $m \times m$ orthogonal matrix $\mathbf{U}_m$ and diagonal matrix $\mathbf{D}_m$. This makes the algorithms of Sections 2–4 practical. The required number m of Lanczos steps is at worst of the magnitude of $\|\Delta t^2 \mathbf{A}\|$ or $\|\Delta t \mathbf{H}\|$, and often considerably smaller. This may however lead to (relatively mild) time step restrictions.

Finally we discuss the implementation of (21). We approximate (omitting the time superscript n)

$$\begin{aligned} \boldsymbol{\phi}(\tau) &= \exp(-\mathrm{i}\tau \mathbf{H})\boldsymbol{\psi} \\ &\approx \mathbf{Q}_m \exp(-\mathrm{i}\tau \mathbf{L}_m) \mathbf{Q}_m^T \boldsymbol{\psi} = \mathbf{Q}_m \mathbf{U}_m \exp(-\mathrm{i}\tau \mathbf{D}_m)\mathbf{c} \end{aligned}$$

with $\mathbf{c} = \mathbf{U}_m^T \mathbf{Q}_m^T \boldsymbol{\psi}$, so that we have

$$\begin{aligned} \mathbf{f}(\tau) &= -\boldsymbol{\phi}(\tau)^* \mathbf{H}' \boldsymbol{\phi}(\tau) \\ &\approx -\mathbf{c}^* \exp(\mathrm{i}\tau \mathbf{D}_m) \mathbf{U}_m^T \mathbf{K}_m \mathbf{U}_m \exp(-\mathrm{i}\tau \mathbf{D}_m)\mathbf{c}\ . \end{aligned}$$

Here,

$$\mathbf{K}_m = \mathbf{Q}_m^T \mathbf{H}' \mathbf{Q}_m$$

requires the computation of $m^2/2$ long inner products for each component of $\mathbf{y}$. With this approximation, the integral in (21) is easily computed analytically.

Acknowledgement. We thank Ch. Schütte and P. Nettesheim for stimulating discussions on the topics of this paper.

References

1. H. C. Andersen. Rattle: A "velocity" version of the Shake algorithm for molecular dynamics calculations. *J. Comp. Phys.*, 52:24–34, 1983.

2. H. J. C. Berendsen and J. Mavri. Quantum simulation of reaction dynamics by density matrix evolution. *J. Phys. Chem.*, 97:13464–13468, 1993.
3. F. A. Bornemann, P. Nettesheim, and Ch. Schütte. Quantum-classical molecular dynamics as an approximation for full quantum dynamics. *J. Chem. Phys.*, 105(3):1074–1083, 1996.
4. F. A. Bornemann and Ch. Schütte. A mathematical investigation of the Car-Parrinello method. Preprint SC 96-19, ZIB Berlin, 1996. To appear in Numer. Math.
5. F. A. Bornemann and Ch. Schütte. On the singular limit of the quantum-classical molecular dynamics model. Preprint SC 96-07, ZIB Berlin, 1996. Submitted to SIAM J. Appl. Math.
6. R. Car and M. Parrinello. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Letter*, 55:2471–2474, 1985.
7. V. L. Druskin and L. A. Knizhnerman. Krylov subspace approximations of eigenpairs and matrix functions in exact and computer arithmetic. *Numer. Lin. Alg. Appl.*, 2:205–217, 1995.
8. B. García-Archilla, J. M. Sanz-Serna, and R. Skeel. Long-time-step methods for oscillatory differential equations. Applied Mathematics and Computation Reports 1996/7, Universidad de Valladolid, 1996.
9. A. García-Vela, R. B. Gerber, and D. G. Imre. Mixed quantum wave packet/classical trajectory treatment of the photodissociation process ArHCl → Ar+H+Cl. *J. Chem. Phys.*, 97:7242–7250, 1992.
10. W. Gautschi. Numerical integration of ordinary differential equations based on trigonometric polynomials. *Numer. Math.*, 3:381–397, 1961.
11. R. B. Gerber, V. Buch, and M. A. Ratner. Time-dependent self-consistent field approximation for intramolecular energy transfer. *J. Chem. Phys.*, 66:3022–3030, 1982.
12. M. Hochbruck and Ch. Lubich. On Krylov subspace approximations to the matrix exponential operator. *SIAM J. Numer. Anal.*, 34:1911–1925, 1997.
13. M. Hochbruck and Ch. Lubich. A Gautschi-type method for oscillatory second-order differential equations. Tech. Rep., Universität Tübingen, 1998.
14. M. Hochbruck and Ch. Lubich. Exponential integrators for quantum-classical molecular dynamics. Tech. Rep., Universität Tübingen, 1998. In preparation.
15. M. Hochbruck, Ch. Lubich, and H. Selhofer. Exponential integrators for large systems of differential equations. *SIAM J. Sci. Comput.*, 1998. To appear.
16. J. Hutter, M. E. Tuckerman, and M. Parrinello. Integrating the Car-Parrinello equations. III. Techniques for ultrasoft pseudopotentials. *J. Chem. Phys.*, 102(2):859–871, 1995.
17. P. Jungwirth and R. B. Gerber. Quantum dynamics of large polyatomic systems using a classically based separable potential method. *J. Chem. Phys.*, 102:6046–6056, 1995.
18. C. Lanczos. Solution of systems of linear equations by minimized iterations. *J. Res. Nat. Bureau Standards*, 49:33–53, 1952.
19. P. Nettesheim, F. A. Bornemann, B. Schmidt, and Ch. Schütte. An explicit and symplectic integrator for quantum-classical molecular dynamics. *Chemical Physics Letters*, 256:581–588, 1996.
20. P. Nettesheim, Ch. Schütte, M. Hochbruck, and Ch. Lubich. Work in preparation.
21. T. J. Park and J. C. Light. Unitary quantum time evolution by iterative Lanczos reduction. *J. Chem. Phys.*, 85:5870–5876, 1986.

22. B. N. Parlett. *The Symmetric Eigenvalue Problem.* Prentice-Hall, Englewood Cliffs, N.J., 1980.
23. J. P. Ryckaert, G. Cicotti, and H. J. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comp. Phys.*, 23:327–341, 1977.
24. M. E. Tuckerman and M. Parrinello. Integrating the Car-Parrinello equations. I. Basic integration techniques. *J. Chem. Phys.*, 101(2):1302–1315, 1994.
25. M. E. Tuckerman and M. Parrinello. Integrating the Car-Parrinello equations. II. Multiple time scale techniques. *J. Chem. Phys.*, 101(2):1316–1329, 1994.

Applications of Ab-Initio Molecular Dynamics Simulations in Chemistry and Polymer Science

Robert J. Meier

DSM Research, P.O. Box 18, 6160 MD Geleen, The Netherlands

Abstract. This paper presents results from quantum molecular dynamics simulations applied to catalytic reactions, focusing on ethylene polymerization by metallocene catalysts. The entire reaction path could be monitored, showing the full molecular dynamics of the reaction. Detailed information on, e.g., the importance of the so-called agostic interaction could be obtained. Also presented are results of static simulations of the Car-Parrinello type, applied to orthorhombic crystalline polyethylene. These simulations for the first time led to a first principles value for the ultimate Young's modulus of a synthetic polymer with demonstrated basis set convergence, taking into account the full three-dimensional structure of the crystal.

1 Introduction

The Car-Parrinello quantum molecular dynamics technique, introduced by Car and Parrinello in 1985 [1], has been applied to a variety of problems, mainly in physics. The apparent efficiency of the technique, and the fact that it combines a description at the quantum mechanical level with explicit molecular dynamics, suggests that this technique might be ideally suited to study chemical reactions. The bond breaking and formation phenomena characteristic of chemical reactions require a quantum mechanical description, and these phenomena inherently involve molecular dynamics. In 1994 it was shown for the first time that this technique may indeed be applied efficiently to the study of, in that particular application catalytic, chemical reactions [2]. We will discuss the results from this and related studies we have performed.

Secondly, the ultimate properties of polymers are of continuous interest. Ultimate properties are the properties of ideal, defect free, structures. So far, for polymer crystals the ultimate elastic modulus and the ultimate tensile strength have not been calculated at an appropriate level. In particular, convergence as a function of basis set size has not been demonstrated, and most calculations have been applied to a single isolated chain rather than a three-dimensional polymer crystal. Using the Car-Parrinello method, we have been able to achieve basis set convergence for the elastic modulus of a three-dimensional infinite polyethylene crystal. These results will also be discussed.

2 Metallocene-Based Olefin Polymerization

The fact that metallocenes employed as homogeneous catalysts are well-defined organometallic species make them ideally suitable for a theoretical molecular modelling study. In the past, studies on both classical Ziegler-Natta catalysts as well as on metallocenes have generally been performed with semi-empirical quantum mechanical methods or low level ab-initio methods. In more recent years, several groups have reported high level ab initio calculations on metallocene complexes, see, e.g., [3–5]. In particular the paper by Ahlrichs et al. [4] showed that this high level was necessary to retrieve qualitatively correct energetic data. All these calculations concerned static molecular structures, i.e. calculation of structures and energies of reactant states, transition states and products. Moreover, only the energetics were usually considered and entropic considerations were not taken into account. These limitations may be overcome by applying a quantum molecular dynamics approach. We will summarize results from quantum molecular dynamics simulations we have performed on ethylene insertion in various metallocenes and related species, which has produced a full record of the chemical reaction including dynamics and bond formation as well as bond breaking phenomena.

The insertion of ethylene into the bridged di(cyclopentadienyl) methylzirconocene $(SiH_2Cp_2)ZrCH_3^+$, previously studied by Morokuma et al. [3], was reinvestigated using quantum molecular dynamics simulation. The structure of the reactant complex including the inserting ethylene is depicted in Fig. 1. Since the Car-Parrinello method also allows for energy minimisations (referring to $T = 0\,\mathrm{K}$), these calculations were carried out first, and dynamics simulations were performed starting from energy minimised reactant complexes. In addition, we studied the corresponding titanocene complex (the Zirconium atom is replaced by a Titanium atom).

The structure of the metallocene cation energy minimised with the Car-Parrinello method agrees well with the experimentally obtained crystal structures of related complexes. Typical features of the structure as obtained from X-ray diffraction on crystals of very similar neutral complexes (e.g., the dichlorides), such as small differences in distances between C atoms within a cyclopentadienyl (Cp) ring, as well as differences in distances between the C atoms of the Cp ring and the Zr atom, were revealed from the simulations.

The full ab-initio molecular dynamics simulation revealed the insertion of ethylene into the Zr-C bond, leading to propyl formation. The dynamics simulations showed that this first step in ethylene polymerisation is extremely fast. Figure 2 shows the distance between the carbon atoms in ethylene and between an ethylene carbon and the methyl carbon, from which it follows that the insertion time is only about 170 fs. This observation suggests the absence of any significant barrier of activation at this stage of the polymerisation process, and for this catalyst. The absence or very small value of a barrier for insertion of ethylene into a bis-cyclopentadienyl titanocene or zirconocene has also been confirmed by static quantum simulations reported independently

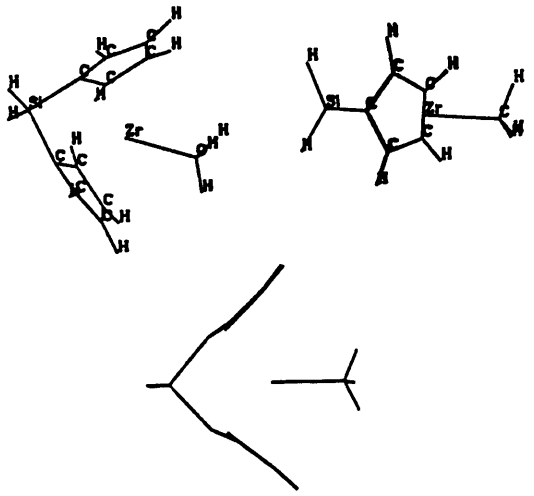


Fig. 1. The structure of the ethylene-zirconocene complex $(SiH_2Cp_2)ZrCH_3^+$-C_2H_4. The corresponding titanocene has basically the same structure, except that the Ti-C distances are obviously different from the Zr-C distances.

by other groups. A similar result was obtained for the titanocene complex, cf. Fig. 2.

The speed of the reaction as indicated by the dynamics simulations, i.e. the time span within which the C_{ethyl}-C_{methyl} bond shortens from about 3.3 Å to a single C-C bond length of 1.6 Å, is an interesting result, because this implies that the reaction time is on the same time scale as the slow motions within the molecule. This result places question-marks on results of static energy minimisation of the transition state of such a reaction. Furthermore, the trajectory file produced by the simulation showed that starting from the ethylene/zirconocene π-complex, the ethylene shifts towards the methyl group that represents the polymer chain. This results in a four-centre transition state, loosely described as a metallacyclobutane ring-type complex. However, given the aforementioned time-scale of the reaction, it might be questioned whether it still makes sense to refer to previously proposed reaction mechanisms, and talk in terms of a four-centre transition state, one of the primary details of the Cossee mechanism [6]. The time-scale of the reaction is so short that it does not seem useful to adopt terminology related to the kind of quasi-equilibrium structures involved in some traditionally proposed schemes. In fact, we have not found a true (quasi-stable) transition state, and the results of the simulation provide evidence for an alternative mechanism due to Brookhart and Green [7]. In that mechanism the formation of a so-called α-agostic interaction is assumed to be necessary for the reaction to proceed. The term agostic interaction is used for situations in which a

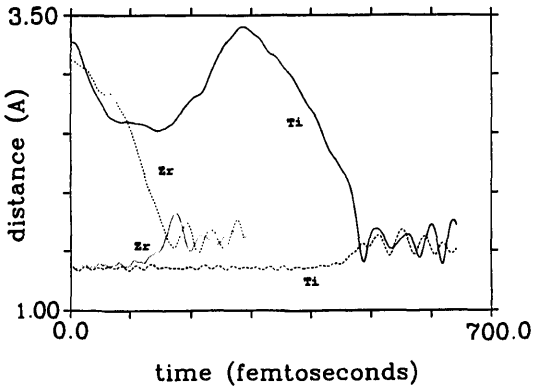


Fig. 2. Time-evolution of the methyl/ethyl C-C distances for both the zirconocene and the corresponding titanocene catalyst. The two curves starting at around 3.2 Å represent the distance between the methyl carbon atom and the nearest-by ethylene carbon atom in the zirconocene-ethylene and the titanocene-ethylene complex, respectively. The two curves starting at around 1.35 Å reflect the ethylene internal C-C bond lengths in the two complexes.

hydrogen atom is covalently bonded simultaneously to both a carbon atom and to a transition metal atom. In this mechanism the π-electron cloud of the inserting ethylene does not start to coordinate to the metal ion, as in the Cossee mechanism, but the ethylene shifts to a position in which one of its carbon atoms comes in direct contact with the methyl carbon attached to the metal centre (see Fig. 1). The formation of the agostic interaction (methyl hydrogen with the metal centre) causes the other two methyl hydrogens to lie practically in a plane with the methyl carbon and the metal ion. This geometry facilitates the approach of the ethylene and the formation of a carbon-carbon bond between it and the methyl group. This is exactly what was observed in the quantum dynamics simulation, for both the titanocene and the zirconocene. The formation of an agostic interaction is illustrated by Figure 3. Moreover, whereas the optimised starting structure had no α-H agostic interaction, during the molecular dynamics run an α-H agostic interaction is established in the early stages of the insertion process. The final complex seems to be stabilised by a γ-H agostic interaction, although severely influenced by the dynamics.

Further simulations have been performed. In contrast to what was observed for bis-cyclopentadienyl metallocenes, mono-cyclopentadienyl systems did reveal a significant barrier to insertion [10]. However, for all these systems it turned out that insertion only proceeded after the formation of a relatively stable agostic interaction, an observation that clearly supports the Brookhart-Green mechanism.

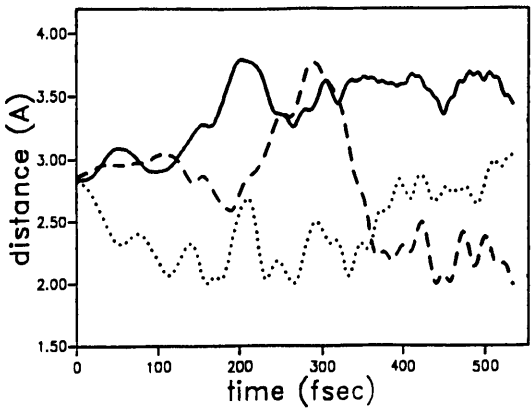


Fig. 3. Time evolution of the distance between the Zr atom and each of the three hydrogen atoms belonging to the methyl group (the original methyl group bonded to the Zr) in the zirconocene-ethylene complex. The time-evolution of one of the hydrogen atoms depicted by the dotted curve shows the development of an α-agostic interaction. Later on in the simulation (after about 450 fs) one of the other protons (broken curve) takes over the agostic interaction (which is then a γ-agostic interaction).

Finally, from the dynamics simulations it was found that the Cp rings are very flexible, and during the simulation they are not very closely attached to the metal centre as often anticipated. For further details regarding the metallocene and mono-cyclopentadienyl based olefin polymerisation studied by quantum molecular dynamics simulation we refer to [2] and [8–10].

3 The Ultimate Young's Modulus for Crystalline Polyethylene

Ultra-high molecular weight polyethylene can be ultra-drawn, thus forming ultra-oriented fibres with high elastic modulus and high tensile strength. These good mechanical properties are the consequence of the high molecular weight and the high degree of orientation of the polymer chains [11]. For such high performance materials it is relevant to evaluate the difference between the highest experimental values obtained for the mechanical properties and the ultimate values calculated by theory. The latter values refer to the ideal, defect-free, material. Because the defect-free material can not be made experimentally (and if it could, one would not be able to determine experimentally that it was indeed defect-free), it is necessary to calculate the ultimate properties. It will be obvious that a correctly calculated ultimate value must be higher than the highest reported experimental value.

A number of papers have reported the calculation of Young's modulus of polyethylene [12–16], whereas papers reporting a calculated tensile strength are scarce [12, 13]. According to Crist and Hereña [14], the calibrated semi-empirical values reported by Horn et al. and Meier probably establish the most reliable theoretical ultimate value for the Young's modulus. The reported ab initio studies [13, 14] show a significant dependence of the modulus on the basis set, which is not unexpected because the modulus depends on the second derivative of the energy with displacement, and is thus as sensitive to the basis set as a (calculated) harmonic vibrational frequency. For the latter, density functional theory usually does a somewhat better job than the Hartree-Fock method employed in modulus studies up till now. More importantly, the reported ultimate Young's modulus extrapolated from the results of the calculations [13] is below the highest reported experimental result [17], and this invalidates the calculated value. Obvious possible causes for failure of theoretical calculations to calculate the ultimate modulus for polyethylene on an ab initio basis are, in arbitrary order, (i) the absence of demonstrated basis set convergence, (ii) the simulation of a single chain rather than a crystal, and (iii) inadequacy of the Hartree-Fock method, even at the MP2 level.

Starting from this situation, we have applied total energy calculations, using the first principles Car-Parrinello method, to a crystal unit cell of orthorhombic polyethylene (see Fig. 4). Because of the application of periodic boundary conditions the system may be considered as an infinite, perfect, crystal. The Car-Parrinello method employs the density functional formalism. Basis set convergence may be demonstrated by increasing the plane-wave cut-off energy, as we will show below. The fact that the calculation of the Young's modulus involves elongation of the unit cell box in one direction, one may anticipate problems with the use of a finite plane wave basis at the interfaces between the boxes. Indeed, when not accounting for this problem a significant scatter of the energy is found as a function of box length; a correction due to Francis and Payne [18] resolves this problem. The resulting total energy curve as function of elongation of the box length is shown in Fig. 5, and from the curvature at the minimum the Young's modulus can be obtained [19, 20]. Basis set convergence was tested and the calculated Young's moduli for a crystal structure slightly different from that displayed in Fig. 4 are shown in Fig. 6. It is observed that the value for the modulus initially "fluctuates" as a function of basis set size, which might be the reason that some Hartree-Fock based values were found to be low compared to the highest experimental value reported. For the correct experimental geometry of orthorhombic polyethylene at $T = 4\,\mathrm{K}$ we evaluated a value for the Young's modulus of 334 GPa, whereas optimization of the crystal structure with the computational tools led to a value of 366 GPa. The difference is acceptable and, more importantly, the values are above the available experimental values, and in the range usually found when attempting to extrapolate experimental values to the ideal, perfectly drawn, material.

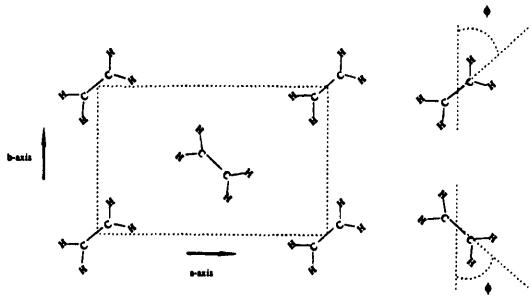


Fig. 4. Structure of orthorhombic polyethylene unit cell projected on the *ab* plane. The right-hand pictures define the setting angle ϕ.

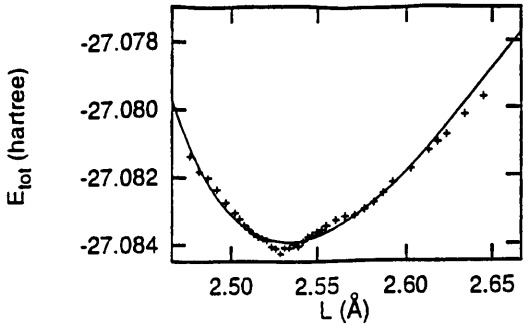


Fig. 5. Total energy curve with finite basis correction (due to Francis and Payne [18]). (1 hartree = 627 kcal/mol or 2624 kJ/mol).

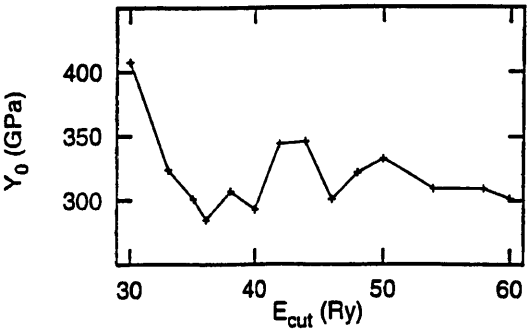


Fig. 6. The calculated Young's modulus as a function of cut-off energy (basis set size). Convergence is basically reached for a cut-off of 54 Ry.

Finally, the modelling of the setting angle (see Fig. 4 for definition) is problematic in empirical force field methods [21]. As far as we know, no adequate data from quantum simulations are available. The shallowness of the potential energy as a function of the setting angle ϕ is a likely reason why the setting angle is so sensitive on the method of calculation employed. Evaluation of the setting angle by structure optimization using simulation methods then serves as a critical test of the calculational method, and we have therefore calculated the dependence of total energy on setting angle, with results displayed in Fig. 7. The angle which gives minimal energy is 42.4°, in very good agreement with the available experimental value of 42° [22] measured at room temperature.

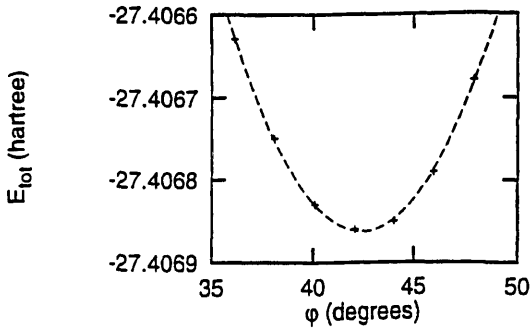


Fig. 7. Total energy as a function of setting angle ϕ. The minimal energy value corresponds to a setting angle of 42.4°.

4 Conclusions

The results presented demonstrate that it has become feasible to study the full dynamics of a (catalytic) reaction from the reactant complex to the product state. This type of theoretical modelling provides previously inaccessible insight in the polymerisation reaction. This is a very significant improvement in methods for further investigating and elucidating (fast) chemical reaction mechanisms. Several groups have now adopted this strategy and papers arising from several research groups have recently been published.

Regarding mechanical properties of polymers, the efficiency of the Car-Parrinello approach has enabled us to evaluate the ultimate Young's modulus of orthorhombic polyethylene, and demonstrate basis set convergence for that property.

5 Acknowledgements

The work presented in this paper was the result of the joint effort with the following colleagues. For the simulations on catalysis Simonetta Iarlori (IBM Italy), Franco Buda (Scuola Normale Superiore, Pisa, Italy) and Gerard van Doremaele (DSM Research) were involved. For the work on polyethylene the simulations were performed by Joost Hageman, with help from Martina Heinemann and Rob de Groot (all University of Nijmegen, The Netherlands).

References

1. R. Car and M. Parrinello, *Phys. Rev. Letters* **55**, 2471 (1985).
2. R.J. Meier, G.H.J. van Doremaele, S. Iarlori and F. Buda, *J. Am. Chem. Soc.* **116** (1994) 7274.
3. H. Kawamura-Kuribayashi, N. Koga and K. Morokuma, *J. Am. Chem. Soc.* **114** (1992) 8687.
4. H. Weiss, M. Ehrig and R. Ahlrichs, *J. Am. Chem. Soc.* **116** (1994) 4919.
5. T.K. Woo, L. Fan and T. Ziegler, *Organometallics* **13** (1994) 432.
6. P. Cossee, *J. Catal.* **3** (1964) 80.
7. M. Brookhart and M.L.H. Green, *J. Organometall. Chem.* **250** (1983) 395.
8. S. Iarlori, F. Buda, G. van Doremaele and R.J. Meier, *Makromoleculare Chemie - Makromolekulare Symposia* **89** (1995) 369.
9. G.H.J. van Doremaele, R.J. Meier, S. Iarlori and F. Buda, *J. Mol. Struct. (Theochem)* **363** (1996) 269.
10. S. Iarlori, F. Buda, R.J. Meier and G.H.J. van Doremaele, Molecular Physics 87(1996)801.
11. B. Crist, *Annu. Rev. Mater. Sci.* **25** (1995) 295.
12. D.S. Boudreaux, *J. Polym. Sci. Polym. Phys. Ed.* **11** (1973) 1285.
13. S. Suhai, *J. Polym. Sci. Polym. Phys. Ed.* **84** (1983) 1341.
14. B. Crist and P.G. Hereña, *J. Polym. Sc., Part B: Polym. Phys.* **34** (1996) 449.
15. T. Horn, W.W. Adams, R. Pachter and D. Haaland, *Polymer* **34** (1993) 2481.
16. R.J. Meier, *Macromolecules* **26** (1993) 4376.
17. P.J. Barham and A. Keller, *J. Polym. Sci. Polym. Lett. Ed.* **17** (1979) 591.
18. G.P. Francis and M.C. Payne, *J. Phys.: Condensed Matter* **2** (1990) 4395.
19. J. Hageman, R.A. de Groot, M. Heinemann and R.J. Meier, *Macromolecules* **30** (1997) 5953.
20. J.C.L. Hageman, R.A. de Groot and R.J. Meier, *Computational Materials Science* **20** (1998), 180.
21. R.L. McCullough and P.H. Lindenmeyer, *Kolloid-Z. u. Z. Polymere* **250** (1972).
22. C.W. Bunn, *Trans. Faraday Soc.* **35** (1939) 482.

Polarons of Molecular Crystal Model by Nonlocal Dynamical Coherent Potential Method

Sergiy V. Izvekov

Institute for Materials Research and Engineering, National University of Singapore, Singapore 119260; FAX: (65) 8720785; e-mail: s-izvekov@imre.org.sg

Abstract. The nonlocal dynamical coherent potential approximation (NDCPA) is formulated to calculate a single-electron(exciton) Green's function of polaron due to the interaction of an electron(exciton) with phonons with dispersion. This approximation is an extension of the dynamical CPA. The NDCPA provides an efficient means calculating of an approximate Green's function for a dynamical model of electrons(excitons) strongly coupled to optical or acoustical phonons, in the entire ranges of the electron(exciton)-phonon coupling strengths and electron (exciton) transfer. The electron(exciton)-phonon coupling in the Hamiltonian may involve terms of any order with respect to the phonon operators. A set of recurrent equations is derived in the case of a system at zero temperatures, from which the coherent potential can be obtained as a function of energy E and momentum k. A simple algorithm for the polaron spectra calculations is obtained for a linear electron(exciton)-phonon coupling in the antiadiabatic limit. The algorithm is applied to calculate absorption spectra of excitons linear and locally coupled to phonon without dispersion.

1 Introduction

An electron(exciton) with a lattice or molecular phonon cloud round itself can be regarded as a new single particle state - a lattice or molecular (exciton) polaron. It is impossible in a small space like this to present a comprehensive review of all the relevant work done in the field of polaron states up to now. Being one of the simplest, most important and directly applicable examples of quasiparticle description of properties of many-particle systems, the polaron problem has often served as a model for both purely theoretical speculations and direct explanations of experimental data. The variety of aims corresponds to a variety of approaches as well as tools: from the field-theoretical expansions in powers of electron(exciton)-phonon coupling constant [1], through different kinds of interpolation (variational) approaches [2] to the small polaron theories invoking so-called polaron canonical transformation [3]. Each of the approaches is tailored to a special purpose and is good in its own (broader or narrower) region of system parameters [3]. For instance, variational interpolation approaches concentrate on the problem of the polaron dispersion law, leaving polaron quasiparticle damping over the

whole range of parameters mostly aside. In the present work, we attempt to partly fill this gap by presenting theory which is in principal able to provide such information.

To determine a single-particle Green's function, which provides information about physical properties (energy spectrum, spectral functions) of polaron states, a self-energy function should be calculated. Approaches based on the coherent potential approximation provide effective tools for the calculation of a polaron self-energy function for a wide class of models and allow consequently (nonperturbatively) to account multiphonon scatterings. Attempts to construct such an algorithm in the case of linear electron(exciton)--phonon coupling have been made before. In [4] the self-energy function has been expanded into an integral chain fraction with the use of diagram techniques. In [5] the expansion of Green's function into a continuous fraction has been performed by Haydock method. It is known, that the coherent potential approximation (CPA) is the most sophisticated virtual-crystal approximation scheme for disordered electronic systems [6-8]. The first attempt to use the notion of the coherent potential for periodic solids in which the role of disorder is played by other elemental excitations which interact with the electron, appeared probably in the $s-d$ model of magnetic semiconductors [9,10]. Almost simultaneously with the coherent potential method for magnetic semiconductors, the so-called (local) Dynamical Coherent Potential Approximation (DCPA) was suggested by Sumi [11-13]. This approximation can be applied to a system in which only an Einstein phonon spectrum and site diagonal electron(exciton)-phonon coupling are assumed. For a system with non-diagonal electron(exciton)-phonon coupling the DCPA is completely inadequate because of the important role of correlated scatterings among different sites. Another difficulty in using the DCPA is that the resulting condition for the coherent potential in [11-13] requires a numerical solution of a set of functional equations. Lately several attempts have been made to improve the (local) DCPA [14-18]. In [14] the dynamical CPA has been reformulated in terms of so-called locators that resulted in simplification of equations that define a coherent potential. S. Abe has extended the (local) DCPA to two-particle Green's functions [15,16], and reformulated this approximation for a model of Frenkel exciton coupled to dispersive phonons [17,18]. The (local) DCPA has been employed to calculate energy spectra [11], absorption and emission spectra [12,17], resonance Raman spectra and time-resolved spectra [16] in exciton-phonon systems. It has been successfully used to explain experiments of absorption spectra in molecular crystals [19,20].

The present paper is devoted to the theoretical formulation and numerical implementation of the NDCPA. The dynamical CPA is a one-site approximation in which variation of a site local environment (due to the presence, for example, of phonons with dispersion) is ignored. It is known from the coherent potential theory for disordered solids [21], that one can account in some extension the variation of a site local environment through an introduction of a nonlocal coherent potential which depends on the difference between site

coordinates. The Fourier transform of the coherent potential is a function of a quasimomentum vector. It is clear, that introduction the nonlocal coherent potential in this way is a different level of approximation because the system is implied to have translational symmetry. For a dynamical disorder which is brought about by phonons the use of the nonlocal coherent potential is natural and does not require any additional approximation. The proposed nonlocal dynamical coherent potential approximation (NDCPA) follows this approach to improve the dynamical CPA. In principal, NDCPA can be applied to calculate a single-particle Green's function of the Hamiltonian with arbitrary electron(exciton)-phonon coupling. This approach maintains also technical preferences: at zero temperature the algorithm is reduced to recurrent linear equations from which a coherent potential can be determined as a function of polaron momentum and energy. The algorithm is greatly simplified if only inelastic scatterings of the electron(exciton) by phonons is taken into account. That is possible if the polaron spectrum is shifted into the energy region where the processes of elastic scatterings can be neglected. In particular, this situation may be realized in the case of an optical phonon spectrum and a narrow electron(exciton) bandwidth (so-called antiadiabatic limit). This problem will be discussed in more details in Sec. 3. In the antiadiabatic limit and assuming an Einstein phonon spectrum a Green's function can be obtained analytically for the case of a model with linear electron(exciton) site diagonal electron(exciton)-phonon interaction. If the energy gain of exciton localization is much larger than the electron(exciton) bandwidth the algorithm can be easily employed to calculate the lowest (localized) electron(exciton) state for the arbitrary ratio of electron(exciton) and phonon bandwidths. The antiadiabatic limit combined with purely local electron(exciton)-phonon interaction also allows to construct a simple recurrent algorithm in which the elastic scattering of electron(exciton) by a phonon field is accounted for.

The paper is organized as follows: in the next section the model is introduced. In the section 3 the nonlocal dynamical CPA at zero temperature is presented. Analytical expressions for the coherent potential are obtained in the antiadiabatic limit for two cases. The first case relates to the polaron energies where the elastic scatterings can be neglected. In the second one, the elastic n-phonon short-range scattering correction is obtained. In the section 4 the algorithm is used to calculate numerically absorption spectra of a system of excitons strongly coupled to Einstein phonons in a cubic crystal with one molecule per unit cell.

2 Model Hamiltonian

In order to demonstrate the NDCPA a model of a system of excitons strongly coupled to phonons in a crystal with one molecule per unit cell is chosen. This model is called here the molecular crystal model. The Hamiltonian of

this system is written as

$$H = \sum_k E_k a_k^+ a_k + \sum_q \omega(q)(b_q^+ b_q + \frac{1}{2}) + \frac{1}{N^{1/2}} \sum_{n,q} \chi(q) e^{iqn} a_n^+ a_n (b_q^+ + b_{-q}). \tag{1}$$

Here $a_k(a_k^+)$ is the annihilation (creation) operator of an exciton with the momentum k and energy E_k, operator $a_n(a_n^+)$ annihilates (creates) an exciton at the n-th site, $b_q(b_{-q}^+)$ is the annihilation (creation) operator of a phonon with the momentum q and energy $\omega(q)$, $\chi(q)$ is the exciton-phonon coupling function, N is the total number of crystal molecules. The exciton energy is $E_k = \epsilon_o + t_k$, where ϵ_0 is the change of the energy of a crystal molecule with excitation, and t_k is the Fourier transform of the energy transfer matrix elements.

This model can be employed to describe an exciton interacting with molecular or lattice vibrations. The coupling function χ is $q-$independent if the exciton-phonon interaction is taken to be site-diagonal and purely local in the exciton and phonon coordinates. This case is realized, for example, for the exciton interacting with molecular vibrations. The $q-$dependent coupling function represents the interaction of the exciton with lattice phonons that arise due to the modulation of van der Waals and Coulomb forces by lattice vibrations. Accounting for the resonant interaction results in $k-$dependent terms in the coupling function. Generally, the magnitude of $k-$dependent and $k-$independent terms in the coupling function varies from one material to another. For a material in which the interaction of the exciton with the local environment is strong, the exciton-phonon interaction operator can be taken as in Eq.(1) with an appropriately chosen coupling function $\chi(q)$.

Hamiltonians equivalent to (1) have been used by many authors for the consideration of a wide variety of problems which relate to the interaction of electrons or excitons with the local environment in solids [22-25]. The model with a Hamiltonian containing the terms describing the interaction between excitons or electrons also allows for the use of NDCPA. For example, the Hamiltonian (1) in which the electron-electron interaction terms are taken into account becomes equivalent to the Hamiltonians (for instance, of Holstein type) of some theories of superconductivity [26-28].

3 Nonlocal Dynamical CPA

Let us introduce the coherent potential $v_k(E)$ which is thought to be dependent on energy E and exciton momentum k. The coherent potential is translational invariant in the site representation. The Hamiltonian (1) is transformed with the coherent potential taken into account as

$$H = H^{eff} + H', \tag{2}$$

with

$$H^{eff} = \sum_k (E_k + v_k) a_k^+ a_k + \sum_q \omega(q)(b_q^+ b_q + \frac{1}{2}),$$
$$H' = \sum_n a_n^+ a_n \varphi_n - \sum_{n,m} v_{n,m} a_n^+ a_m.$$

In H' the site representation is used and

$$\varphi_n = \sum_m g_{n,m}(b_m^+ + b_m), \tag{3}$$

where

$$b_m = \frac{1}{N^{1/2}} \sum_q b_q e^{iqm}$$

is the annihilation operator of a phonon associated with the m-th lattice site, and

$$g_{n,m} = \frac{1}{N} \sum_q \chi(q) e^{iq(n-m)},$$
$$v_{n,m} = \frac{1}{N} \sum_k v_k e^{ik(n-m)}.$$

In the further manipulations the site representation will be used for convenience. The Fourier transform with respect to time of a single-exciton retarded Green's function $G^{eff}(t)$ of a system under the Hamiltonian H^{eff} in the site representation for exciton coordinates and Fock's representation for phonon coordinates is written as

$$G_{ml}^{eff}(z) =$$
$$\frac{1}{N} \sum_k \sum_{s=0}^{\infty} \frac{1}{s!} \frac{1}{N^s} \sum_{\substack{q_1,\dots,q_s \\ i_1,\dots,i_s \\ j_1,\dots,j_s}} \frac{e^{ik(m-l)} e^{iq_1(i_1-j_1)+\cdots+iq_s(i_s-j_s)}}{z - E_k - v_k(z) - \omega(q_1) - \cdots - \omega(q_s)} \times$$
$$| i_1, \dots, i_s >< j_1, \dots, j_s | . \tag{4}$$

Here

$$| i_1, \dots, i_m >= b_{i_1}^+ \cdots b_{i_m}^+ \; | 0 >,$$

and z denotes the complex energy $z = E + i\gamma$ with an exciton decay rate γ which is assumed to be independent of energy and momentum. It is assumed in the expressions below that the summations are performed over indexes

$s, s_i, i = 1, 2, \ldots$. Expanding the Green's function $G_{ml}(z)$ of the exciton with Hamiltonian H in powers of propagator $G^{eff}_{ml}(z)$ we get

$$G_{ml}(z) = [z - H_{eff} - (\hat{\varphi} - \hat{v}(z))]^{-1}_{ml} = \\ = G^{eff}_{ml}(z) + G^{eff}_{ms}(z)\ (\hat{\varphi} - \hat{v}(z))_{ss_1}\ G^{eff}_{s_1 l}(z) + \cdots . \tag{5}$$

Here $\hat{\varphi}$ and $\hat{v}(z)$ are operator matrixes with elements

$$\varphi_{mn} = \delta_{mn}\varphi_n \tag{6}$$

and $v_{mn}(z)$, respectively.

We restrict ourselves to the case of the system at zero temperature. This is not relevant from the point of view of methodology. The case of nonzero temperatures may be considered in the completely analogous fashion. At zero temperature the coherent potential is chosen so that the $G_{ml}(z)$ and $G^{eff}_{ml}(z)$ averaged over phonon degrees of freedom in phonon vacuum are equal

$$< 0 \mid G_{ml}(z) \mid 0 > = < 0 \mid G^{eff}_{ml}(z) \mid 0 > . \tag{7}$$

For the case of nonzero temperatures the vacuum averages in Eq.(7) should be replaced by thermal averages over phonon populations. Using (7) and (5) we obtain that the scattering of an exciton in the effective medium by the perturbation $(\hat{\varphi} - \hat{v}(z))$ is described by the following self-consistent condition

$$< 0 \mid (\hat{\varphi} - \hat{v}(z))_{ml} + (\hat{\varphi} - \hat{v}(z))_{ms}\ G^{eff}_{ss_1}(z) \times \\ (\hat{\varphi} - \hat{v}(z))_{s_1 l} + \cdots \mid 0 > = < 0 \mid T_{ml} \mid 0 > = 0. \tag{8}$$

Performing summation in (8) we obtain the $t-$matrix operator $\hat{T}$. Once $\hat{T}$ is found Eq.(8) is written as

$$< 0 \mid \left[\hat{I} - (\hat{\varphi} - \hat{v}(z))\hat{G}^{eff}(z)\right]^{-1} (\hat{\varphi} - \hat{v}(z)) \mid 0 > = 0. \tag{9}$$

$\hat{I}$ is the unity matrix. The Green's function $\hat{F}(z)$ of the exciton under the Hamiltonian (1) in which the exciton-phonon coupling terms are excluded is given by the expression

$$\hat{F}^{-1}(z) = \hat{G}^{eff\,-1}(z) + \hat{v}(z). \tag{10}$$

With the use of the operator $\hat{F}(z)$ Eq.(9) is rewritten as

$$< 0 \mid [\hat{I} - \hat{\varphi}\hat{F}(z)]^{-1}\hat{\varphi} \mid 0 > = < 0 \mid [\hat{I} - \hat{\varphi}\hat{F}(z)]^{-1} \mid 0 > \hat{v}(z). \tag{11}$$

This is the operator (or matrix) equation for the coherent potential $\hat{v}(z)$. The formal expression for $\hat{v}(z)$ can be easily written. The exciton polaron is

determined by the Green's function for the exciton in the effective medium defined by the coherent potential

$$< 0 \mid G^{eff}_{ml}(z) \mid 0 >= \frac{1}{N} \sum_{k} \frac{e^{ik(m-l)}}{z - E_k - v_k(z)}. \tag{12}$$

The question how to calculate the phonon vacuum averages in Eq.(11) remains. Let us derive equations which determine them.

Let $\tilde{i}_n$, $\tilde{i}_n + i_k$, and $\tilde{i}_n - i_k$ represent sets of sites $(i_1, \ldots, i_n)$, $(i_1, \ldots, i_n, i_k)$, and $(i_1, \ldots, i_{k-1}, i_{k+1}, \ldots, i_n)$, respectively. The set (0) and (i_1) represent, respectively, an empty and an one-site set. In the expressions below we will not always indicate the dependence of the operators $\hat{F}(z)$ and $\hat{v}(z)$ on the energy z. All the calculations are performed with $\hat{F}(z)$, $\hat{v}(z)$ taken at the same energy. Defining

$$a^{\tilde{i}_n}_{ml} =< 0 \mid [\hat{I} - \hat{\varphi}\hat{F}]^{-1}_{ml} \mid i_1, \ldots, i_n > \tag{13}$$

and using Eqs.(11), (6), and (3) the coherent potential v_{ml} is expressed as

$$v_{ml} = [a^{(0)}]^{-1}_{ms} a^{(s_1)}_{sl} g_{s_1 l}. \tag{14}$$

Employing the obvious operator identity

$$[\hat{I} - \hat{\varphi}\hat{F}]^{-1} = \hat{I} + [\hat{I} - \hat{\varphi}\hat{F}]^{-1}\hat{\varphi}\hat{F}, \tag{15}$$

and definition (13) we obtain that the matrix elements $a^{\tilde{i}_n}_{ml}$ are defined by the following set of equations

$$\begin{aligned} a^{(0)}_{ml} &= \delta_{ml} + a^{(s_1)}_{ms} g_{ss_1} F^{(0),(0)}_{sl} \\ &\vdots \\ a^{\tilde{i}_n}_{ml} &= a^{\tilde{s}_n + s_1}_{ms} (n_{\tilde{s}_n, s_1} + 1)^{1/2} g_{ss_1} F^{\tilde{s}_n, \tilde{i}_n}_{sl} + a^{\tilde{s}_n - s_1}_{ms} (n_{\tilde{s}_n, s_1})^{1/2} g_{ss_1} F^{\tilde{s}_n, \tilde{i}_n}_{sl} \\ &\vdots \end{aligned} \tag{16}$$

Here

$$F^{\tilde{i}_n, \tilde{j}_n}_{ml}(z) = \frac{1}{N^{n+1}} \sum_{\substack{k \\ q_1, \cdots, q_n}} \frac{e^{ik(m-l)} e^{iq_1(i_1 - j_1) + \cdots + iq_n(i_n - j_n)}}{z - E_k - \omega(q_1) - \cdots - \omega(q_n)} \tag{17}$$

is a $n-$phonon matrix element of the operator $\hat{F}$ and $n_{\tilde{i}_n, i_k} = (\delta_{i_1, i_k} + \cdots + \delta_{i_n, i_k})$. Eqs.(16) are a set of recurrent linear equations from which $a^{(0)}_{ml}$ and $a^{(i_1)}_{ml}$ can be determined. Inserting then $a^{(0)}_{ml}$ and $a^{(i_1)}_{ml}$ in Eq.(14) finally we find the coherent potential. With the use of Eqs.(11), (15) we can easily write down equations similar to (16) for exciton-phonon coupling operator $\hat{\varphi}$ of more complex (for instance, quadratic) phonon operator structure .

The set of equations in (16) are greatly simplified if we make some assumptions which allow the exciton and phonon site nondiagonal matrix elements $F_{ml}^{\tilde{i}_n,\tilde{j}_n}$ to be neglected. In the case of an Einstein phonon spectrum these matrix elements are exactly zero and Eqs.(16) contain only $F_{ml}^{\tilde{i}_n,\tilde{i}_n}$. For latter reference the matrix elements of $\hat{F}$ which are site diagonal in exciton coordinates, and both in exciton and phonon coordinates are denoted as

$$F_{ml}^{(n)}(z) = \frac{1}{N^{n+1}} \sum_{\substack{k \\ q_1,\cdots,q_n}} \frac{e^{ik(m-l)}}{z - E_k - \omega(q_1) - \cdots - \omega(q_n)}, \tag{18}$$

$$F^{(n)}(z) = \frac{1}{N^{n+1}} \sum_{\substack{k \\ q_1,\cdots,q_n}} \frac{1}{z - E_k - \omega(q_1) - \cdots - \omega(q_n)}, \tag{19}$$

respectively. The nondiagonal $F_{ml}^{\tilde{i}_n,\tilde{j}_n}$ can be eliminated if the energies E satisfy the following condition

$$|E - \epsilon_0 - n\omega(0)| \gg B + n\Omega, \tag{20}$$

where B and Ω are the exciton and phonon bandwidths, respectively. The fact that this condition is unfulfilled means that for a polaron state with energy E elastic (with impulse transfer) $n-$phonon scatterings of the exciton are relevant. We have to account $F^{(n)}$ in Eqs.(16) only if we assume that polaron spectrum is located in the energy region where elastic scatterings are irrelevant. Such situation may be realized in the case of a low dispersion optical phonon spectrum ($\Omega \ll \omega(0)$), and a narrow exciton band ($B \ll \omega(0)$). As it will be seen, in this case the polaron spectrum is shifted by the energy gain of exciton localization. Due to this shifting the energies in the polaron bands may obey the condition (20). If the condition (20) is not satisfied for polaron band n_0 then the n_0-phonon elastic scatterings of the exciton should be taken into account and, therefore, $F_{ml}^{\tilde{i}_{n_0},\tilde{j}_{n_0}}$ should be kept in Eqs.(16). In other words we have to make elastic scattering corrections to calculate polaron band for this case. It is obvious that if the phonons are acoustical the elastic $n-$phonon scattering becomes relevant beginning with a sufficiently large n for an arbitrary value of the polaron energy E (see (20)).

Let us first consider a case of the polaron spectrum when condition (20) is fulfilled so that the nondiagonal matrix elements of the operator $\hat{F}$ can be neglected. we denote as v_{inel} the coherent potential in which only the inelastic scatterings are accounted. It is easy show by explicit calculations that

$$\frac{a_{ml}^{\tilde{i}_n+p}}{a_{ml}^{\tilde{i}_n+p'}} = \frac{g_{lp}}{g_{lp'}} \frac{(n_{\tilde{i}_n,p'})^{1/2}}{(n_{\tilde{i}_n,p})^{1/2}}. \tag{21}$$

Therefore, Eqs.(16) contain only $a_{ml}^{\tilde{p}_n}$. Here $\tilde{p}_n = (\underbrace{p,\ldots,p}_{n})$. It is easily seen that $a_{ml}^{\tilde{p}_n} = 0$ when $m,l \neq p$. Finally we conclude that Eqs.(16) actually form a three-diagonal set for the values $a_{pp}^{\tilde{p}_n}$. This enables us to express the coherent potential in the form of a continuous fraction. In order to show this let us introduce $t^{(n)}$ as

$$t^{(n)} = \frac{a_{pp}^{\tilde{p}_n}}{a_{pp}^{\tilde{p}_{n-1}}} \frac{1}{n^{1/2}}. \tag{22}$$

Then Eqs.(16) combined with Eqs. (14) and (21) yield

$$\begin{aligned} v_{inel}(z) &= t^{(1)} \frac{G}{g_{00}} \\ &\vdots \\ t^{(n)} &= \frac{g_{00}}{\dfrac{1}{F^{(n)}(z)} - (n+1)t^{(n+1)} \dfrac{G}{g_{00}}} \\ &\vdots \end{aligned} \tag{23}$$

where $g_{00} = g_{pp}$ and

$$G = \sum_s g(s)^2 = \frac{1}{N} \sum_q \mid \chi(q) \mid^2 .$$

From Eqs.(23) we get the coherent potential v_{inel} in the form

$$v_{inel}(z) = \cfrac{G}{\dfrac{1}{F^{(1)}(z)} - \cfrac{2G}{\dfrac{1}{F^{(2)}(z)} - \cfrac{3G}{\ddots}}}. \tag{24}$$

The energy spectrum of the exciton polaron as a function of k is obtained by the solution of an equation

$$z - E_k - v(z) = 0. \tag{25}$$

Eq.(25) with $v(z) = v_{inel}(z)$ can be further simplified if we assume an Einstein phonon spectrum ($\omega(q) = \omega_0$) and neglect the energy transfer matrix elements in $F^{(n)}$ (in this case $F^{(n)-1}(z) = z - \epsilon_0 - n\omega_0$). Then with the use of the formulae [29]

$$\sum_{n=0}^{\infty} \frac{a^n/n!}{z+a-n} e^{-a} = \cfrac{1}{z - \cfrac{a}{z-1 - \cfrac{2a}{z-2 - \cfrac{3a}{\ddots}}}},$$

for complex z with a nonvanishing imaginary part Eq.(25) is written as

$$t_k - \frac{1}{\sum_{n=0}^{\infty} \left(\frac{1}{z - \epsilon_0 + S\omega_0 - n\omega_0} \right) \frac{S^n}{n!} e^{-S}} = 0, \tag{26}$$

where $S = G/\omega_0^2$. For a limit case $t_k = 0 (B = 0)$ when the exciton energy spectrum converges to ϵ_0 which may correspond to a highly localized exciton or an exciton trapped on an impurity we obtain from (12), (25), and (26) the exciton Green's function in the form

$$G(z) = \sum_{n=0}^{\infty} \left(\frac{1}{z - \epsilon_0 + S\omega_0 - n\omega_0} \right) \frac{S^n}{n!} e^{-S}. \tag{27}$$

This result can be easily obtained with the usual methods [30]. Thus we see that the present approximation becomes valid for $B = 0$. On the whole, formulae (26) accounts the configuration mixing among polaron states with different number of phonons. The polaron spectrum calculated with the use of Eq.(26) consists of equally spaced bands which are shifted to lower energies by $S\omega_0$ and have a renormalized dispersion law as compared to a dispersion law of a free exciton. To show the latter let us suppose that for energies in the n_0−th polaron band we have only one leading term of the sum in (26) which corresponds to n_0−phonon scatterings. Then the dispersion law in the n_0−th polaron band may be roughly described by the expression

$$E = \epsilon_0 - S\omega_0 + n_0\omega_0 + \frac{S^{n_0}}{n_0!} e^{-S} t_k. \tag{28}$$

Such renormalization can be obtained in the framework of the small polaron theory [3]. $S\omega_0$ is the energy gain of exciton localization. Let us note that the condition (20) and, therefore, Eq.(26) is correct for $S \gg B/\omega_0$ and *arbitrary* B/ω_0 for the lowest energy of the exciton polaron. So Eq.(26) can be used to evaluate the energy of a self-trapped exciton when the energy of the vibrational or lattice relaxation is much larger then the exciton bandwidth.

It is possible to make elastic scattering corrections to the algorithm (24) in the case of an Einstein phonon spectrum and purely local exciton-phonon coupling. If we calculate the energy of the polaron state at the value $E \approx n\omega_0$ only the matrix elements $F_{ml}^{(n)}$ should be considered in Eqs.(16). In this case

Eqs.(16) are rewritten as

$$
\begin{gathered}
a^{(0)}_{mp} = \delta_{mp} + a^{(p)}_{mp} g_{00} F^{(1)} \\
\vdots \\
a^{\tilde{p}_{n-1}}_{mp} = a^{\tilde{p}_n}_{mp} F^{(n-1)}(n)^{1/2} g_{00} + a^{\tilde{p}_{n-2}}_{mp} F^{(n-1)}(n-1)^{1/2} g_{00} \\
a^{\tilde{p}_n}_{ml} = \sum_{s \neq p} a^{\tilde{p}_n + s}_{ms} F^{(n)}_{sl} g_{00} + a^{\tilde{p}_{n+1}}_{mp} F^{(n)}_{pl} (n+1)^{1/2} g_{00} + a^{\tilde{p}_{n-1}}_{mp} F^{(n)}_{pl} n^{1/2} g_{00} \\
a^{\tilde{p}_n + l}_{ml} = a^{\tilde{p}_n + \tilde{l}_2}_{ml} F^{(n\delta_{pl}+1)}(n\delta_{pl}+2)^{1/2} g_{00} + a^{\tilde{p}_n}_{ml} F^{(n\delta_{pl}+1)}(n\delta_{pl}+1)^{1/2} g_{00} \\
\vdots
\end{gathered}
\tag{29}
$$

Here $\tilde{p}_n + \tilde{l}_k = (\underbrace{p, \ldots, p}_{n}, \underbrace{l, \ldots, l}_{k})$. It is easily seen that in Eqs.(29) $a^{\tilde{p}_n + \tilde{l}_k}_{ml} = 0$ and $a^{\tilde{p}_r}_{ml} = 0$ when $m \neq p$. The Eqs.(29) with $a^{\tilde{p}_n + \tilde{l}_k}_{pl}, p \neq l$ in left-hand side form a three-diagonal set of equations for $a^{\tilde{p}_n + \tilde{l}_k}_{pl}$ quite similar to Eqs.(23). Therefore, solving this system we obtain

$$
\frac{a^{\tilde{p}_n + l}_{pl}}{a^{\tilde{p}_n}_{pl}} = v_{inel} \frac{1}{g_{00}}. \tag{30}
$$

Let us now introduce a matrix with elements

$$
t_{pl} = \frac{a^{\tilde{p}_n}_{pl}}{a^{\tilde{p}_n}_{pp}}. \tag{31}
$$

Using Eqs. (29) and (30) we obtain that the coherent potential v_{el} with elastic scattering correction is determined by the following equations

$$
\begin{gathered}
v_{el}(z) = t^{(1)} g_{00} \\
\vdots \\
t^{(n-1)} = \frac{g_{00}}{\dfrac{1}{F^{(n-1)}(z)} - n t^{(n)} g_{00}} \\
t_{pl} = t_{ps} F^{(n)}_{sl}(z) v_{inel}(z) - \\
F^{(n)}_{pl}(z) v_{inel}(z) + t^{(n+1)} F^{(n)}_{pl}(z)(n+1) g_{00} + \frac{1}{t^{(n)}} F^{(n)}_{pl}(z) g_{00} \\
t^{(n+1)} = \frac{g_{00}}{\dfrac{1}{F^{(n+1)}(z)} - (n+2) t^{(n+2)} g_{00}} \\
\vdots
\end{gathered}
\tag{32}
$$

From Eqs.(32) for t_{pl} when $p \neq l$ we can find the Fourier transform $t(k)$ of t_{pl} through $t^{(n)}$ and $t^{(n+1)}$ only. Then, inserting $t(k)$ in Eqs. (32) with t_{pp} in

the left-hand side we get equation which contains $t^{(n)}$ and $t^{(n+1)}$ only from which $t^{(n)}$ can be expressed through $t^{(n+1)}$

$$t^{(n)} = \frac{g_{00}}{\left(\frac{1}{N}\sum_k \frac{1}{z-E_k-n\omega_0-v_{inel}(z)}\right)^{-1} + v_{inel}(z) - (n+1)t^{(n+1)}g_{00}}. \quad (33)$$

Eqs.(32) where $t^{(n)}$ is calculated by the formulae (33) represent the recurrent algorithm for determination of v_{el}.

Above mathematics shows that the changes in the model Hamiltonian (1) that do not involve the exciton-phonon coupling terms, - for instance inclusion the exciton-exciton (electron-electron) interaction, lead only to the respective change of $F_{ml}^{\tilde{i}_n,\tilde{j}_n}$ in Eqs.(16).

4 Calculations and Discussions

To verify effectiveness of NDCPA we carried out the calculations of absorption spectra for a system of excitons locally and linearly coupled to Einstein phonons at zero temperature in cubic crystal with one molecule per unit cell (probably the simplest model of exciton-phonon system of organic crystals). Absorption spectrum is defined as an imaginary part of one-exciton Green's function taken at zero value of exciton momentum vector

$$I_{abs}(E) = -\frac{1}{\pi} Im \frac{1}{z - v_0(z)}, \quad (34)$$

where $z = E + i\gamma$, and γ is an exciton decay rate which defines the width of absorption spectrum lines. To find coherent potential $v_0(z)$ Eqs.(16) were solved by iterations. We kept only five first equations (i.e. multiphonon scatterings were accounted up to the fifth order). The number of sites in the periodic region that is the interval of change of phonon indexes in the matrix elements a in Eqs.(16) was assumed to be equal to seven. Let note that this region actually is the area in which we account the exciton-phonon interaction.

To achieve faster convergence and avoid overflow during iterations, equations (16) were rewritten in terms of $t_{pl}^{(n)} = a_{pl}^{\tilde{p}_n}/a_{pp}^{\tilde{p}_n}$ (see(31) and (32)). Diagonal elements $t_{pp}^{(n)}$ were calculated by expanding into the continuous fraction on the each step of iterations.

In Fig. 1 the absorption spectra for a number of values of excitonic bandwidth B are depicted. The phonon energy ω_0 is chosen as energy unit there. The presented pictures correspond to three cases of relation between values of phonon and excitonic bandwidths - $B < \omega_0$, $B = \omega_0$, $B > \omega_0$. The first picture ($B = 0.3$) corresponds to the antiadiabatic limit ($B \ll \omega_0$), which can be handled with the small polaron theories [3]. The last picture($B = 10$) represents the adiabatic limit ($B \gg \omega_0$), that fitted for the use of variation approaches [2]. The intermediate cases B=0.8 and B=1 can't be treated with these techniques. The overall behavior of spectra seems to be reasonable and

is in good agreement with results obtained by other authors with different methods [5], [12], [15].

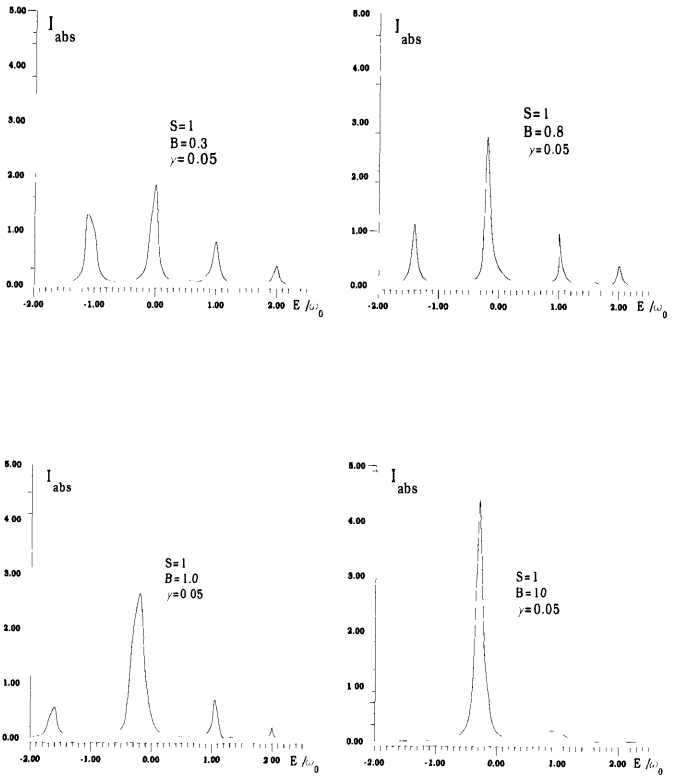


Figure 1: Absorption spectra for system of excitons linear and locally coupled with strength S to nondispersive phonons with energy ω_0 calculated in NDCPA. B is excitonic bandwidth, and γ is excitonic decay rate.

The NDCPA seems to be a very reasonable way to treat the properties of both electrons and excitons interacting with phonons with dispersion. In principal, the NDCPA can be applied to a system of the Hamiltonian with the electron(exciton)-phonon coupling terms of arbitrary structure. The NDCPA results in an algorithm which can be effectively treated numerically (for example, iteratively). The application of the NDCPA is not restricted to the

case of zero temperature and can be invoked to deal with the systems with different distributions (thermal or nonequilibrium) of phonon populations. For such systems the averages in Eq.(7) should be performed over the respective distributions. The algorithm is greatly simplified in the antiadiabatic limit and for polaron energies where the elastic exciton-phonon scatterings are irrelevant. In this case if the electron(exciton)-phonon coupling term is linear the NDCPA algorithm is reduced to a three-diagonal linear set of equations and, therefore, the coherent potential can be obtained analytically. It should be noted that for electron(exciton)-phonon terms including the phonon operator of nth-order the coherent potential is represented by a $2n+1$-diagonal linear set of equations and can be easily calculated numerically up to high orders of electron(exciton)-phonons scatterings. It is also possible to obtain the $n-$phonon short-range scattering correction when the role of this scattering is not negligible in the formation of a polaron band.

References

1. A.A.Abrikosov,L.P.Gor'kov, and I.E.Dzyaloshinskii Methods of Quantum Field Theory in Statistical Mechanics, (Dover, New York,1975).
2. T.D.Lee, D.Pines Phys Rev. **88**, 960(1952).
3. I.G.Lang and Yu.A.Firsov Zh. Eksp. Teor. Phys. **43**, 1843 (1962) [Sov. Phys. -JETP **16**, 1301 (1963)].
4. N.V. Tkach Teor. Mat. Phys.(USSR) **61**, 400(1984) (in Russian).
5. A.V. Sherman Phys. Stat. Sol. (b) **131**, 225(1985).
6. P. Soven Phys. Rev. **156**, 809(1967).
7. D.W. Taylor Phys. Rev. **156**, 1017(1967).
8. B. Velicky, S. Kirkpatrick, and H. Ehrenreich Phys. Rev. **175**, 747(1968).
9. A. Rangette, A. Yanase , and J. Kübler Sol. State Comm. **12**, 171(1973).
10. K. Kubo J. Phys. Soc. Jpn. **36**, 32(1974).
11. H. Sumi J. Phys. Soc. Jpn. **36**, 770(1974).
12. H. Sumi J. Phys. Soc. Jpn. **38**, 825(1975).
13. H. Sumi J. Chem. Phys. **67**, 2943(1977).
14. V. Čhápek, V. Špička Czech. J. Phys. B **34**, 115(1984).
15. S. Abe J. Phys. Soc. Jpn. **57**, 4029(1988).
16. S. Abe J. Phys. Soc. Jpn. **57**, 4036(1988).
17. S. Abe J. Lumin. **45**, 272(1990).
18. S. Abe J. Phys. Soc. Jpn. **59**, 1496(1990).
19. Y. Tokura and T. Koda Solid State Commun. **40**, 299(1981).
20. Y. Wada, Y. Tokura, and T. Koda J. Chem. Phys. **86**, 3009(1987).
21. N.F. Berk, D.J. Shazeer, and R.A. Tahir-Kheli Phys. Rev. **B8**, 2496(1973).
22. E.I. Rashba Sov. Phys. -JETP **27**, 292(1968).
23. H. Sumi and Y. Toyozawa J. Phys. Soc. Jpn. **31**, 342(1972).
24. G. Venzl and S.F. Fischer Phys. Rev. **B32**, 6437(1985).
25. J. Singh and A. Matsui Phys. Rev. **B36**, 6094(1987).
26. T. Holstein Ann. Phys (New York) **8**, 325(1959).
27. J. Ranninger Phys. Rev. **B48**, 13166(1993).
28. F. Marsiglio Phys. Rev. **B42**, 2416(1990).

29. W.Gautschi, in Handbook of Mathematical Functions, edited by Abramowitz M. and Stegun I.A. (Dover, New York,1964), p.295.
30. M. Lax J. Chem. Phys. **20**, 1753(1952).

Part V

Parallel Force Field Evaluation

Ewald and Multipole Methods for Periodic N-Body Problems *

John A. Board, Jr.[1], Christopher W. Humphres[1], Christophe G. Lambert[2], William T. Rankin[1], and Abdulnour Y. Toukmaji[1]

[1] Dept. of Electrical and Computer Engineering, Duke University, Durham, NC USA
[2] Dept. of Computer Science, Duke University, Durham, NC USA

Abstract. Many realistic biomolecular simulations require use of periodic boundary conditions to create a surface-free environment for the molecule of interest and associated solvent molecules to interact. Electrostatic interactions are the principal computational cost of such simulations. We have implemented two codes: a parallel variant of an Ewald summation method which computes the effect of infinite periodic boundary conditions, and a parallel variant of a multipole algorithm which explicitly computes the interactions within a large but finite periodic system. Each has a regime of applicability, with Ewald favoring smaller systems and fewer processors, and the multipole methods favoring larger systems and more processors. Simulations can now include a full treatment of periodic electrostatics to three or four significant figures of accuracy for a computational cost equivalent to that of a 12Å cutoff simulation.

1 Motivation

Several groups have previously reported parallel implementations of multipole based algorithms for evaluating the electrostatic n-body problem and the related gravitational n-body problem [1, 2]. These methods permit the evaluation of the mutual interaction between n particles in serial time proportional to $n \log n$ or even n under certain conditions, with further reductions in computation time from parallel processing.

Our work is targeted to biomolecular simulation applications, where the objective is to illuminate the structure and function of biological molecules (proteins, enzymes, etc) ranging in size from dozens of atoms to tens of thousands of atoms today, with the desire to increase this limit to millions of atoms in the near future. Such molecular dynamics (MD) simulations simply apply Newton's law to each atom in the system, with the force on each atom being determined by evaluating the gradient of the potential field at each atom's position. The potential includes contributions from bonding forces,

* Supported by NSF ASC-9318159, NSF CDA-9422065, NIH Research Resource RR08102, and computer time from the North Carolina Supercomputing Center. An earlier version of this paper was presented at the Eighth SIAM Conference on Parallel Processing for Scientific Computing.

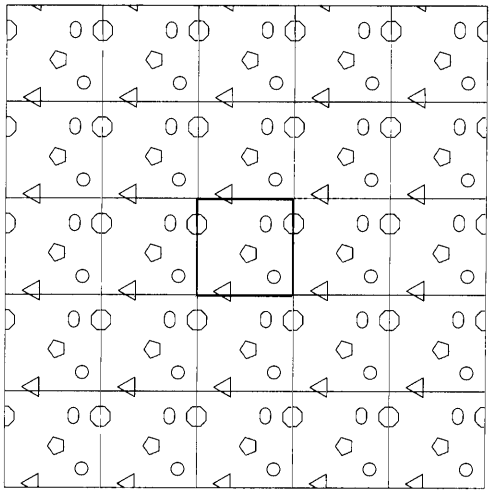


Fig. 1. Periodic boundary conditions protect the inner simulation cell from disturbing effects of having all its particles close to the surface. With PBCs in force, as a particle moves out of the box on one side, one of its images will move back into the box on the opposite side.

van der Waals interactions, and other sources, but the computationally dominant contribution is from electrostatics, as all the other terms are short ranged, requiring computation time that grows only linearly in the size of the system.

For simulations of the motions of the atomic constituents of these molecules to be meaningful, the molecules must be placed in a natural environment, such as in a water bath or inside a membrane wall; this increases the total number of atoms in the simulation by a significant factor (typically between 2 and 10). Even a "large" water bath by these standards is still extremely tiny, being only a few water molecules deep. Such simulations are adequate in some cases, but many properties of interest are distorted by surface effects and orientational correlations imposed by the small water bath and the finite system boundary. Periodic Boundary Conditions (PBC) have long been used to overcome the effects of a tiny simulation region; by replicating the original simulation region a finite or infinite number of times in all directions (Fig. 1), the system's "boundary" is pushed out much further; to infinity in the case of infinite PBCs. Particles in the replicated cell simply mimic the motions of the particles in the original unit cell; each particle in the original cell feels the force induced by all other particles and all periodic images of all particles (including itself).

The simplest form of PBCs simply provides one layer of replicated cells about the original simulation region. Infinite PBCs are the other extreme, with an infinite number of layers of surrounding cells. We will describe two

different methods for evaluating forces in periodic assemblies of particles. We will present results from each method in isolation, i.e. simply computing the electrostatic interactions among a group of particles, and we will close the paper with results from inserting each solver into a complete molecular dynamics program.

2 Macroscopic Multipole Method

Our first method for treating periodic systems is an extension to fast multipole methods for rapidly evaluating the electrostatic interactions within a unit cell. The serial version of the macroscopic multipole algorithm is described in detail in [5]. Briefly, the multipole algorithms in their several variants all subdivide the original simulation cell into a hierarchy of increasingly fine regions (typically via an oct-tree); the different algorithms then impose some criterion to decide whether two regions are sufficiently distant from each other to interact via the multipole approximation rather than directly. The macroscopic method extends this hierarchy of interacting regions upward by creating new regions consisting of increasingly large numbers of copies of the original unit cell, as in Fig. 2. As these aggregates get farther from the initial unit cell, they become larger, so that they continue to satisfy the multipole acceptance criterion which typically involves the ratio of the size of the distant box to the separation between that box and the target cell (in this case, the entire simulation cell). The complete description of this method in [5] describes two alternative schemes for this replication and provides the relevant error bounds in each case. A degenerate case of the method provides the equivalent of a single layer of surrounding cells, so that particles which move out of the simulation box are immediately replaced by their image on the opposite side.

This hierarchical macroscopic replication of the original unit cell is carried out k times, where k is an arbitrary parameter. The cells in each new layer are 27 times larger than the cells at the previous layer (in three dimensions); eight levels of this replication convert a 100,000 atom unit cell into a region of approximately 28 quadrillion particles. Even at this level of replication, the cost of evaluating the forces on all the particles in the original unit cell is negligibly greater than the cost of simply evaluating the forces due to the particles in the unit cell alone, i.e. the standard multipole algorithms. Unlike the Ewald methods discussed below, strict charge neutrality in the unit cell is not a necessary condition for these methods.

The speed of the method comes from two sources. First, all of the "macroscopic" cells of the same size have exactly the same internal structure, as they are simply formed of tessellated copies of the original cell, thus each has exactly the same multipole expansion. We need compute a new multipole expansion only once for each level of macroscopic agglomeration. Second, the structure of the periodic copies is fixed; we can precompute a single transfer

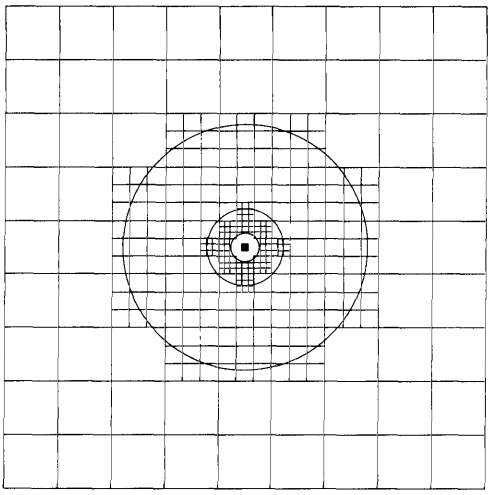


Fig. 2. The Macroscopic multipole algorithm creates exponentially larger aggregates of the original unit cell (small solid box in center) to rapidly build up a large but finite periodic system.

function to describe the effect of all of the copies on the original unit cell. In the case of constant pressure simulations where the box size fluctuates, the transfer function needs to be scaled. Only in anisotropic constant pressure simulations, where the size and shape of the unit cell both fluctuate, does the function need to be recalculated periodically.

Parallelizing this method was not difficult, given that we already had parallel versions of several multipole algorithms to start from. The entire macroscopic assembly, given its precomputed transfer function, is handled by a single processor which has to perform k extra multipole expansions, one for each level of the macroscopic tree. Each processor is already typically performing many hundreds or thousands of such expansions, so the extra work is minimal.

Our multipole code D-PMTA, the Distributed Parallel Multipole Tree Algorithm, is a message passing code which runs both on workstation clusters and on tightly coupled machines such as the Cray T3D/T3E [11]. Figure 3 shows the parallel performance of D-PMTA on a moderately large simulation on the Cray T3E; the scalability is not affected by adding the macroscopic option.

3 Ewald Summation

Ewald summation was invented in 1921 [7] to permit the efficient computation of lattice sums arising in solid state physics. PBCs applied to the unit cell of a crystal yield an infinite crystal of the appropriate symmetry; performing

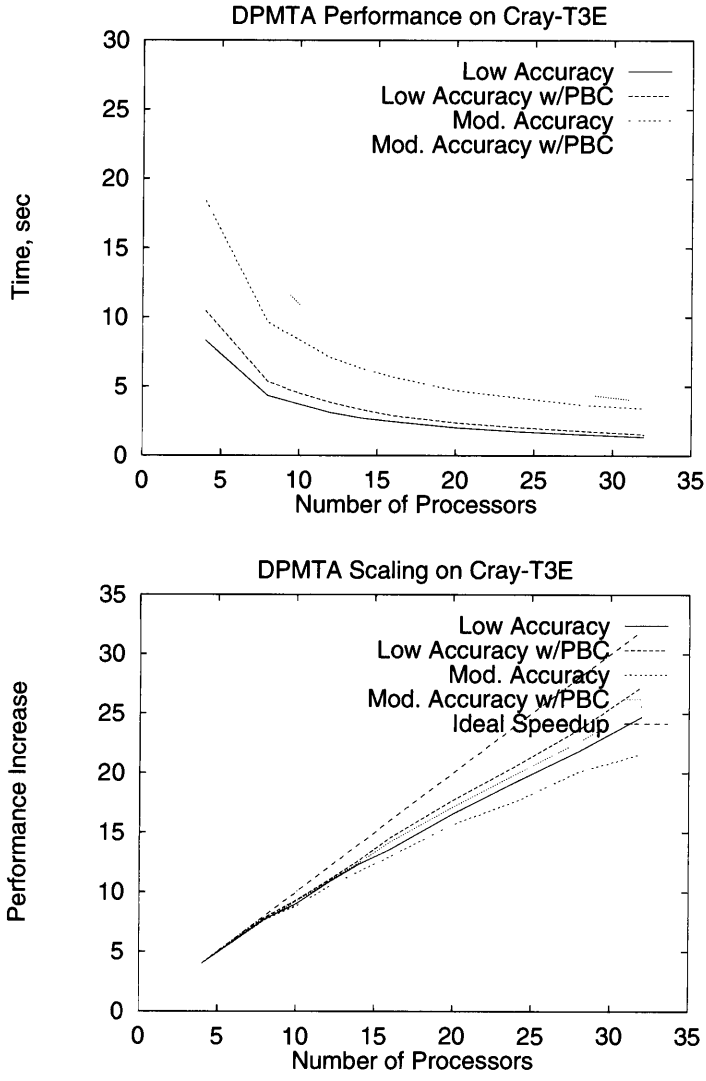


Fig. 3. Performance (top) and scaling behavior (bottom) of D-PMTA on the Cray T3E when simulating 70,000 particles.

sums over this infinite crystal lattice allows calculation of many properties of the material. The potential on particle i due to all other particles j and all periodic images of all particles can be written as

$$\Phi_i = \sum_{\substack{j=1 \\ j \neq i \text{ if } \mathbf{n}=(0,0,0)}}^{N} \sum_{\mathbf{n}} \frac{Z_i Z_j}{r_{i,j_\mathbf{n}}} \tag{1}$$

where $\mathbf{n}$ is an integer triple identifying a particular image cell: $\mathbf{n} = (0,0,0)$ indicates the "real" simulation cell, and all other cells are identified by their displacement from the central cell, i.e., $\mathbf{n} = (1,0,0)$ identifies the image cell immediately adjacent on the x axis, and $\mathbf{n} = (i,j,k)$ identifies the image displaced i unit cells in x, j in y, and k in z. $r_{i,j_\mathbf{n}}$, then, is the distance between particle i in the original cell and particle j in the $\mathbf{n}$th image cell. Ewald recognized that this slowly and indeed conditionally convergent sum can be recast as two rapidly converging sums, one in real space and one in reciprocal space. One physical explanation of Ewald's observation is that an auxiliary Gaussian charge distribution can be both added to and subtracted from the original charge distribution. The real space sum is now in terms of the rapidly converging complementary error function erfc(r) rather than the slowly converging $\frac{1}{r}$ thanks to the Gaussian screening of the charges. The other sum over the Gaussian counter-charges can be Fourier transformed into a rapidly converging form in reciprocal space.

Ewald's formalism reduces the infinite lattice sum to a serial complexity of n^2 in the number of particles n, which has been reduced to $n \log n$ in more recent formulations. A review of variants on Ewald summation methods which includes a more complete derivation of the basic method is in [3].

For biomolecular applications, as with the macroscopic method above, the "unit cell" is the entire original simulation volume; it is replicated throughout space resulting in an infinite simulation region. Convergence of the Ewald sum does require charge neutrality in the simulation volume, which can complicate some biomolecular simulations by requiring the introduction of potentially unphysical counter charges.

3.1 Particle-Mesh Ewald

One of the most efficient algorithms known for evaluating the Ewald sum is the Particle-mesh Ewald (PME) method of Darden et al. [8, 9]. The use of Ewald's "trick" of splitting the Coulomb sum into real space and Fourier space parts yields two distinct computational problems. The relative amount of work performed in real space *vs* Fourier space can be adjusted within certain limits via a free parameter in the method, but one is still left with two distinct calculations. PME performs the real-space calculation in the conventional manner, evaluating the complementary error function within a cutoff

radius. To speed up the Fourier space calculation, PME borrows from the Particle-Particle, Particle-Mesh (P^3M) method of Hockney and Eastwood [10] to interpolate all the randomly spaced particles onto a regular mesh. The Fourier sum is then computed via FFT techniques and the results interpolated back to the actual particle sites. PME originally employed Lagrange interpolation [8], but the revised PME which we have implemented uses B-spline interpolation functions [9]. The smoothness of B-splines allows the force expressions to be evaluated analytically, with high accuracy, by differentiating the real and reciprocal energy equations rather than using finite differencing techniques.

Our interest has been to parallelize PME; some of our efforts are described in [4]. The three dimensional FFT needed by PME is notoriously difficult to parallelize, while the real-space contribution is quite easy to parallelize, so we first bias the work as much as possible to favor the real-space term. We can do this by adjusting the free parameter in the method which controls the Gaussian width of the fictitious charge and counter-charge distributions added to the system. A larger value of this width allows a coarser grid in Fourier space (and thus less work to evaluate the FFT) but a larger real-space cut-off radius for the complementary error function (and thus more work in the real-space sum). The ability to trade off work between the real and reciprocal space sums is further limited by the requirement that the real-space cut-off radius cannot exceed half the side length of the original simulation cell.

The real-space sum is particularly easy to parallelize as it simply involves a spatial decomposition with appropriate attention to the overlap between adjacent regions due to the cut-off radius. Performance for the real-space contribution alone is given in Fig. 4.

The Fourier sum, involving the three dimensional FFT, does not currently run efficiently on more than perhaps eight processors in a network-of-workstations environment. On a more tightly coupled machine such as the Cray T3D/T3E, we obtain reasonable efficiency on 16 processors, as shown in Fig. 5. Our initial production implementation was targeted for a small workstation cluster, so we only parallelized the real-space part, relegating the Fourier component to serial evaluation on the "master" processor. By Amdahl's principle, the 16% of the work attributable to the serially computed Fourier sum limits our potential speedup on 8 processors to 6.25, a number we are able to approach quite closely.

4 Performance Comparison

The results in the prior two sections were for the Macroscopic multipole and PME solvers in isolation. A complete MD simulation involves much more than these routines. In addition to computing the short range interactions from bonding forces, etc., the particle positions and velocities need to be updated each timestep. Additionally, efficient MD programs recognize that the

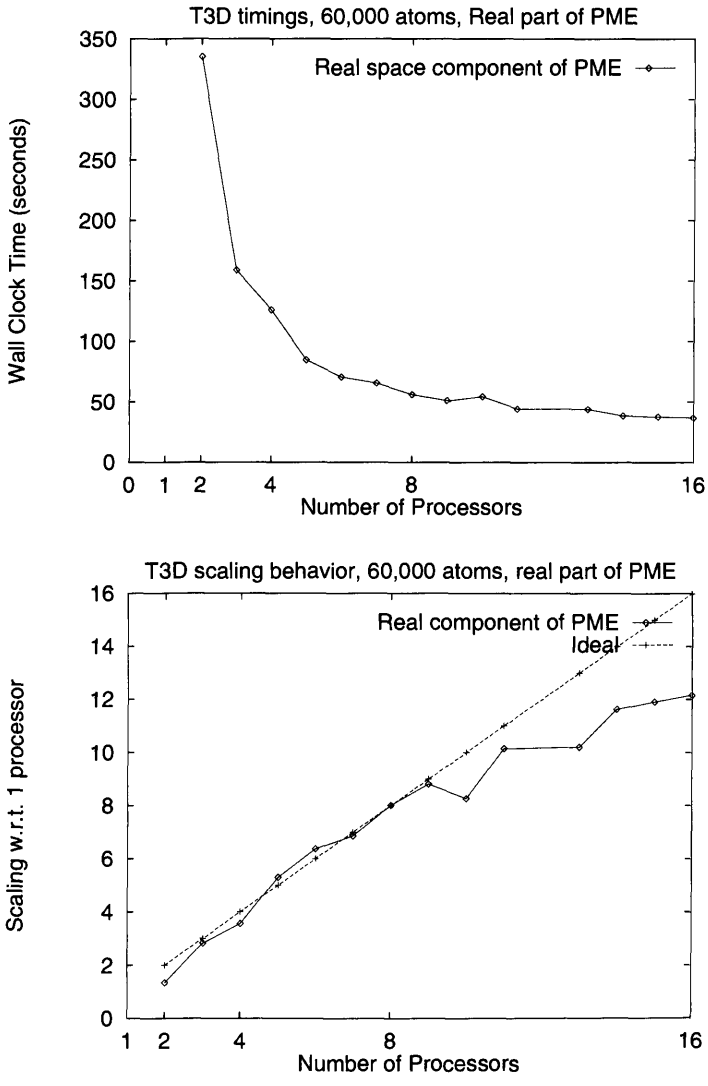


Fig. 4. Performance (top) and scaling behavior (bottom) of the real space part of PME on the Cray T3D.

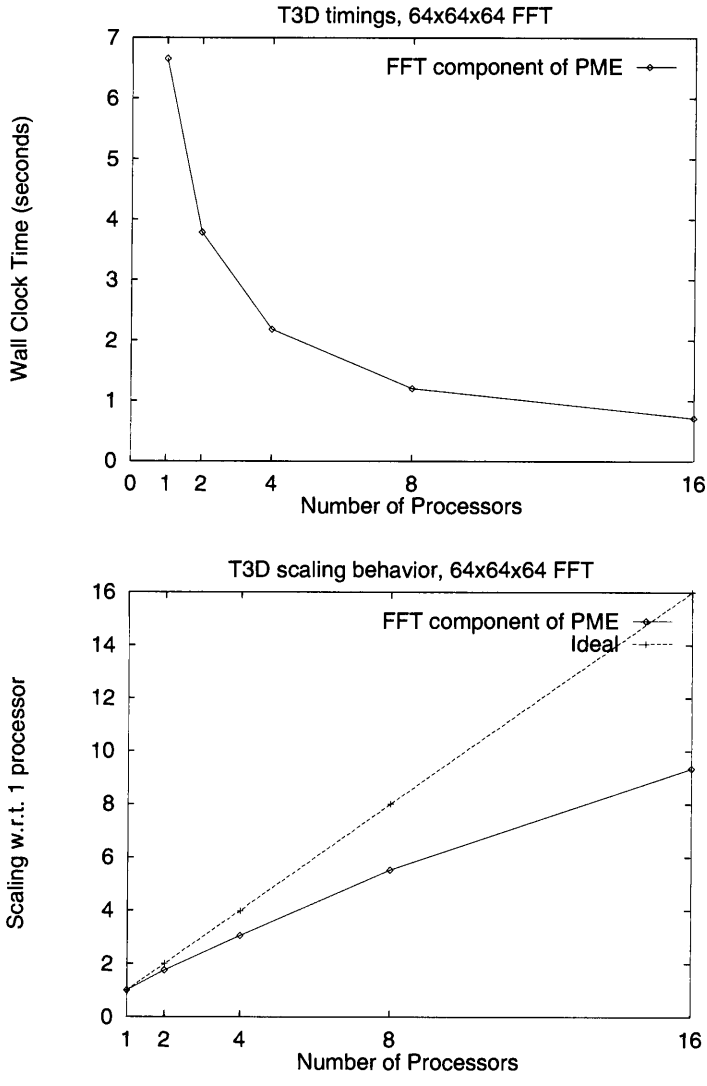


Fig. 5. Performance (top) and scaling behavior (bottom) of 64x64x64 3D FFT on the Cray T3D.

force on a given particle due to long-range (distant Coulomb) interactions varies slowly, so it is not necessary to do a full Coulomb (and thus multipole or Ewald) solve each timestep. Instead, major and minor timesteps are established, with a full Coulomb solve performed each major timestep, which is perhaps 4-10 minor timesteps long. The forces due to long-range interactions are interpolated between major timesteps from estimates derived from the last few major timesteps.

Despite all the effort to reduce both the frequency of Coulomb solves (periodic or not) and the computational complexity of each call when required, the long-range force evaluation remains the dominant computational cost of MD simulations.

These comparisons are for the serial versions of the algorithms only. Figure 6 compares the running time of complete MD simulations (in this case, a box containing 23,832 water molecules, or equivalently 71,496 atoms) performed with the different periodic solvers, as well as traditional methods utilizing cutoff radii. The MD program used is SigmaXplus, a research version of SigmaX from Dr. Jan Hermans at the University of North Carolina at Chapel Hill, modified to accept both PME and our multipole codes. The solvers used are summarized in Table 1.

C-#	cutoff radius method – all non-bonded forces between particles within # angstroms of each other are computed explicitly
P3-UNC	rewritten PME code, tightly integrated to SigmaX, low accuracy
P5-UNC	rewritten PME code, tightly integrated to SigmaX, moderate accuracy
P3-Dk	original Duke PME code, loosely coupled to SigmaX, low accuracy
P5-Dk	original Duke PME code, loosely coupled to SigmaX, moderate accuracy
DP-4	Multipole code, 4 levels of macroscopic expansion, 4 terms in the multipole expansions, low accuracy
DP-8	Multipole code, 4 levels of macroscopic expansion, 8 terms in the multipole expansions, moderate accuracy
DP-12	Multipole code, 4 levels of macroscopic expansion, 12 terms in the multipole expansions, high accuracy

Table 1. Parameters for SigmaX test runs.

The salient comparisons are between the bars marked "P3-Dk," our initial parallel PME implementation, and "DP-4," the macroscopic multipole method with four levels of macroscopic boxes. Though it is difficult to create a completely fair comparison in terms of the relative accuracy of the potentials and forces as computed by the two methods, the parameters for these simulations were tuned to give comparable overall accuracy[1]. PME is clearly

[1] Accuracy is defined here as relative to an infinitely periodic reference system. The values of the energies and forces in the reference system can be determined to

faster, running in only two thirds the time of the multipole methods. The DP-8 and DP-12 bars are higher accuracy simulations with more accurate multipole expansions. The very fast P3-UNC bar resulted from rewriting our PME code to share data structures with SigmaX to avoid the overhead involved in calling the original PME; this version also used a table lookup for erfc, which reduced accuracy somewhat but significantly improved speed. The multipole codes would also benefit somewhat from a tight integration with SigmaX, but it would be more difficult to implement this for the multipole codes than for the relatively simpler PME codes.

Since no real physical system is in fact infinitely periodic, there has always been some concern that calculations with Ewald and related methods may not reflect reality. The macroscopic multipole algorithm gives us a mechanism to study the effect of finite vs infinite periodicity. Referring back to Section 2 and Figure 2, if the macroscopic replication process is carried out k times (forming k "shells" of increasingly larger numbers of copies of the original unit cell around the unit cell), the force and energy values computed converge very rapidly to the infinite Ewald value. For $k = 3$, forces are already within one part in 10^{-3} of the infinite system value; by $k = 5$ or 6 we have exceeded single precision floating point accuracy in agreement. Details of this comparison can be found in [6].

5 Conclusions

So which solver to use? In regions of low to moderate accuracy, and on small numbers of processors, PME is a clear winner. Recall the results above from SigmaX were serial results, but both PME and the PMTA derivatives work well on small workstation clusters (4-16 processors). On larger machines and/or when higher accuracy is required, the multipole codes have an advantage. We expect that PME can be made to scale well to a larger number of processors on a tightly coupled machine such as the Cray T3E, but the intrinsic communication requirements of the 3D FFT probably limit PME's performance on a loosely coupled workstation cluster to current levels. A more thorough examination of the speed/accuracy/scalability trade-off is underway now; this should result in stronger guidelines for our biochemist collaborators on which solver to use when.

6 Acknowledgments

We happily acknowledge Dr. Tom Darden from the National Institute for Environmental Health Sciences for access to and help in parallelizing PME.

machine precision via a highly accurate (and slow) conventional Ewald calculation; the results of such a computation are the reference values for these relative accuracy determinations.

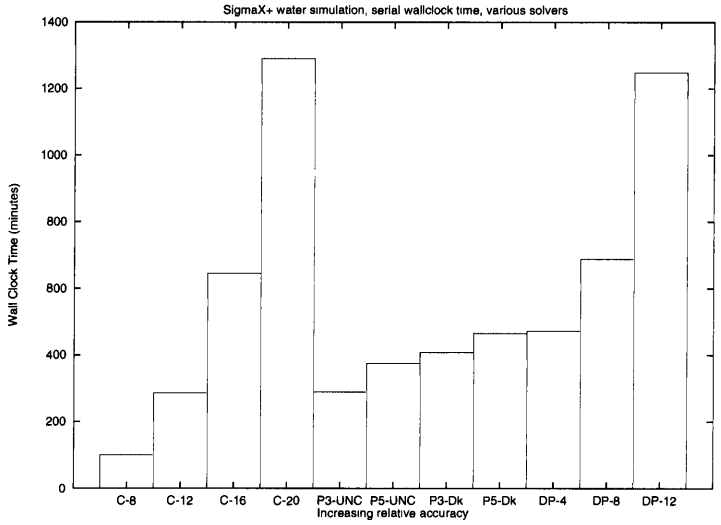


Fig. 6. Wallclock time for a complete run of SigmaX with various solvers. In each case, 23,832 water molecules are simulated for 1000 minor timesteps of 2 fs; the Coulomb solver (PME or PMTA) was called every 6 minor timesteps. Details of the solvers used are in Table 1.

We also are grateful to Drs. Jan Hermans and Yun Ru Huai at the University of North Carolina at Chapel Hill for access to SigmaX and help in integrating our codes with SigmaX. We thank the reviewers for helpful comments in revising the manuscript.

References

1. J. A. Board, Jr. et al., *Scalable variants of Multipole-Accelerated Algorithms for Molecular Dynamics Applications*, Proceedings, Seventh SIAM Conference on Parallel Processing for Scientific Computing, SIAM, Philadelphia (1995), pp. 295–300.
2. M. S. Warren and J. K. Salmon, *A Parallel, Portable and Versatile Treecode*, Proceedings, Seventh SIAM Conference on Parallel Processing for Scientific Computing, SIAM, Philadelphia (1995), pp. 319–324.
3. A. Toukmaji and J. A. Board, Jr., *Ewald Sum Techniques in Perspective: A Survey*, Comput. Phys. Comm., 95 (1996), pp. 73–92.
4. A. Toukmaji and D. Paul and J. A. Board, Jr., *Distributed Particle-Mesh Ewald: A Parallel Ewald Summation Method*, Proceedings, International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA'96), CSREA Press (1996), pp. 33–43.
5. C. G. Lambert and T. A. Darden, and J. A. Board, Jr., *A Multipole-Based Algorithm for Efficient Calculation of Forces and Potentials in Macroscopic Periodic Assemblies of Particles*, J. Comp. Phys. 126 (1996), pp. 274–285.

6. C. G. Lambert, *Multipole-based Algorithms in Molecular Biophysics and Nonparametric Statistics*, Ph.D. Dissertation, Duke University Department of Computer Science, 1997.
7. P. Ewald, Ann. Phys. 64 (1921), pp. $253ff$.
8. T. Darden and D. York and L. Pedersen, J. Chem. Phys. 98 (1993), pp. $10089ff$.
9. T. Darden and U. Essmann and H. Lee and L. Perera and M. Berkowitz and L. Pedersen, J. Chem. Phys. 103 (1995), pp. $8577ff$.
10. R. Hockney and J. Eastwood, *Computer Simulation Using Particles*, McGraw-Hill, New York (1981).
11. W. T. Rankin and J. A. Board, Jr., *A Portable Distributed Implementation of the Parallel Multipole Tree Algorithm*, Proceedings, Fourth IEEE International Symposium on High Performance Distributed Computing, IEEE Computer Society Press (1995), pp. 17–22.

Avoiding Algorithmic Obfuscation in a Message-Driven Parallel MD Code

James C. Phillips, Robert Brunner, Aritomo Shinozaki, Milind Bhandarkar, Neal Krawetz, Attila Gursoy, Laxmikant Kalé, Robert D. Skeel, and Klaus Schulten

Theoretical Biophysics Group, University of Illinois and Beckman Institute, 405 North Mathews Avenue, Urbana, IL 61801, USA

Abstract. Parallel molecular dynamics programs employing shared memory or replicated data architectures encounter problems scaling to large numbers of processors. Spatial decomposition schemes offer better performance in theory, but often suffer from complexity of implementation and difficulty in load balancing. In the program NAMD 2, we have addressed these issues with a hybrid decomposition scheme in which atoms are distributed among processors in regularly sized patches while the work involved in computing interactions between patches is decomposed into independently assignable compute objects. When needed, patches are represented on remote processors by proxies. The execution of compute objects takes place in a prioritized message-driven manner, allowing maximum overlap of work and communication without significant programmer effort. In order to avoid obfuscation of the simulation algorithm by the parallel framework, the algorithm associated with a patch is encapsulated by a single function executing in a separate thread. Output and calculations requiring globally reduced quantities are similarly isolated in a single thread executing on the master node. This combination of features allows us to make efficient use of large parallel machines and clusters of multiprocessor workstations while presenting minimal barriers to method development and implementation.

Introduction

This paper describes the design history of the program NAMD, developed by members of the Theoretical Biophysics Group at the University of Illinois starting in 1994. The intent is to give the reader a better understanding of the conflicting forces which shape the design of a parallel molecular dynamics code and to demonstrate the need for advanced features such as multiple threads and message-driven execution.

From a software design perspective, a molecular dynamics program carries out a very simple algorithm. The gradient of a potential energy function is calculated for all atoms in a system, yielding a force; this force is then employed by an integration algorithm to update the positions of the atoms for the next force evaluation. Aside from issues of reading data, generating output, and the actual integration algorithm there is only this basic cycle of force evaluation and integration which is carried out every timestep.

Molecular dynamics simulations run for millions of timesteps consuming months of computer time. It is the length of simulations which has led to the use of parallel computing in this field. It is the iterative nature of the molecular dynamics algorithm which produces the challenge, for although efficiently parallelizing independent force evaluations is trivial, the force evaluations for a sequence of timesteps must be individually parallel to realize a speedup. Also, even if force evaluation consumes the vast majority of computer time, it may be advantageous to perform the integration in parallel as well, increasing scalability according to Amdahl's law [1].

Parallelism increases programming complexity and with it the need for sound software engineering practices. This is especially true for programs designed for public use in an academic environment since the primary developers of such codes are often graduate students who tend to move on after obtaining their degrees. In addition, molecular dynamics is not a static field and the users of such software often propose new algorithms and techniques to be added to a working code. Thus, a complex program such as a parallel molecular dynamics code must be sufficiently well designed and documented that it can be maintained and enhanced by future generations of programmers. Those portions of the code which are most likely to be modified, such as the integration algorithm, must therefore be especially clear and modularly separated from the remaining code with well-documented interfaces.

The following sections cover the design goals, decisions, and outcomes of the first two major versions of NAMD and present directions for future development. It is assumed that the reader has been exposed to the basics of molecular dynamics [2, 3, 4] and parallel computing [5]. Additional information on NAMD is available electronically [6].

NAMD 1

NAMD [7] was born of frustration with the maintainability of previous locally developed parallel molecular dynamics codes. The primary goal of being able to hand the program down to the next generation of developers is reflected in the acronym NAMD: Not (just) Another Molecular Dynamics code. Specific design requirements for NAMD were to run in parallel on the group's then recently purchased workstation cluster [8] and to use the fast multipole algorithm [9] for efficient full electrostatics evaluation as implemented in DPMTA [10].

Two implementation decisions could be made immediately. First, DPMTA is based on the PVM message-passing library [11] and therefore it was necessary to base NAMD on PVM as well. All communication done by NAMD, however, would use an intermediate interface to allow communications to be easily retargeted to MPI [12] or other standards, and to simplify later implementation of communication optimizations such as combining messages destined for the same processor. Second, after much debate C++ was selected

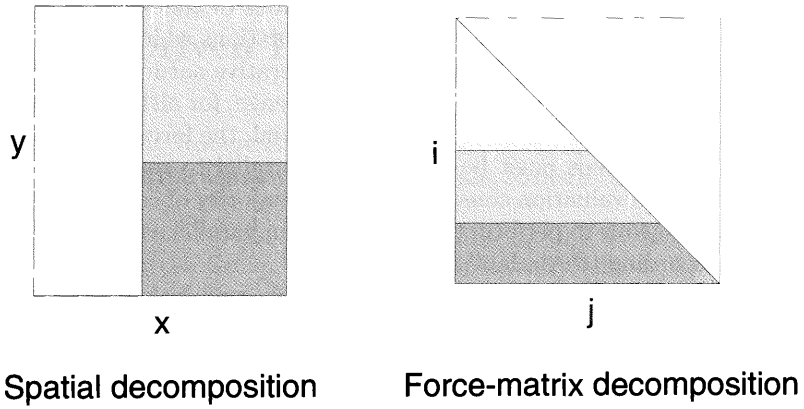


Fig. 1. Nonbonded force evaluation may be distributed among processors according to atomic coordinates, as in spatial decomposition (left), or according to the indices of the interacting atoms, as in force-matrix decomposition (right). Shades of gray indicate processors to which interactions are assigned.

as the development language. This was based on the desire to use an object-oriented design and on prior good experiences in developing the visualization program VMD [13]. There was concern that existing C++ compilers were not uniformly mature and hence to ensure portability across platforms exotic features (at the time) such as templates would be avoided in NAMD. In order to avoid possible performance problems [14] time-critical sections of code like force evaluation were reduced to plain C, many functions were inlined, and virtual functions were avoided.

Parallel molecular dynamics codes are distinguished by their methods of dividing the force evaluation workload among the processors (or *nodes*). The force evaluation is naturally divided into bonded terms, approximating the effects of covalent bonds and involving up to four nearby atoms, and pairwise nonbonded terms, which account for the electrostatic, dispersive, and electronic repulsion interactions between atoms that are not covalently bonded. The nonbonded forces involve interactions between all pairs of particles in the system and hence require time proportional to the square of the number of atoms. Even when neglected outside of a cutoff, nonbonded force evaluations represent the vast majority of work involved in a molecular dynamics simulation.

Methods of decomposing the nonbonded force evaluation fall into two classes, *spatial decomposition* [15] in which atoms and their interactions are divided among processors based on their coordinates, and *force-matrix decomposition* [16] in which the calculation of the interaction between a pair of atoms is assigned to a processor without considering the location of either atom (Fig. 1). Spatial decomposition scales better to large numbers of

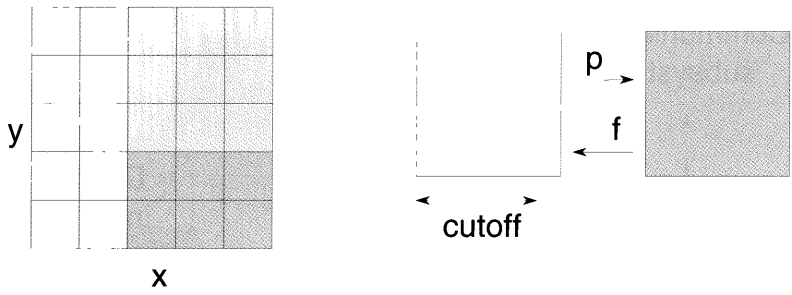


Fig. 2. Patches divide the simulation space into a regular grid of cubes, each larger than the nonbonded cutoff. Interactions between atoms belonging to neighboring patches are calculated by one of the patches which receives a positions message (p) and returns a force message (f). Shades of gray indicate processors to which patches are assigned.

processors because it takes locality of communication into account, while force-matrix decomposition is easier to implement and load-balance.

NAMD implemented spatial decomposition and addressed the load balancing issue by dividing the simulation space into a large number of cubes called *patches* (Fig. 2). A patch serves three purposes. First, it is a region of space larger than the cutoff distance for nonbonded force evaluation, and can therefore function in a cell list or linked-list method [17] to accelerate distance checking for nonbonded interactions. Second, a patch is a unit of parallelizable work which can be reassigned to balance load among processors—each node possessing several patches. Finally, a patch is a *message-driven object* that receives atomic coordinates from some of its neighboring patches, calculates interactions, and returns forces while sending coordinates to and receiving forces from its other neighbors.

Message-driven execution [18] is a parallel processing technique in which communication latency is hidden by overlapping computation and communication. This is achieved by executing computations specified by the messages as they arrive instead of in a fixed serial order. Every coordinate message that arrives contains data that allows some subset of the force evaluation to be carried out, primarily nonbonded interactions between atoms of the patch which sent the message and those of the patch which receives it. Messages are prioritized such that those which generate off-node communication (such as position messages from off-node patches) are processed before messages between patches on the same node. (Actually, the main message loop in NAMD attempted to receive each of the several types of messages in order of priority, providing only roughly prioritized message execution.)

NAMD was implemented in an object-oriented fashion (Fig. 3). Patches, the encapsulated communication subsystem, the molecular structure, and various output methods were objects. Every patch owned specialized objects

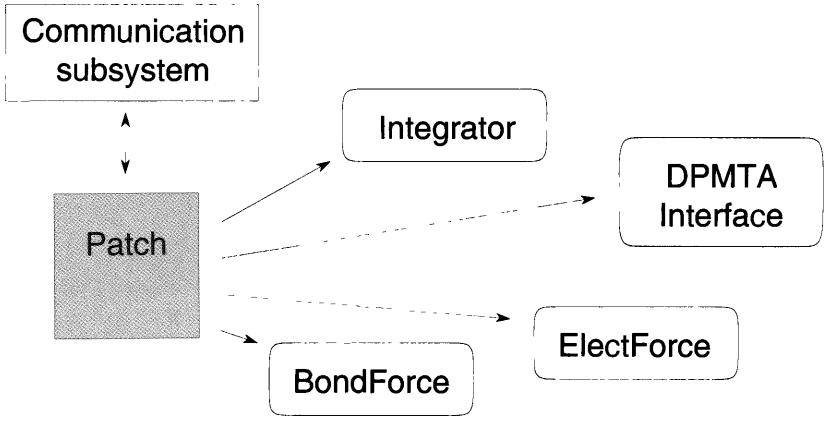


Fig. 3. NAMD 1 employs a modular, object-oriented design in which patches communicate via an encapsulated communication subsystem. Every patch owns an integrator and a complete set of force objects for bonded (BondForce), nonbonded (ElectForce), and full electrostatic (DPMTA) calculations.

responsible for integration, the several types of force calculations, and the interface to the DPMTA full electrostatics package. This made the system modular in that new forces or integration methods could be added with minimal modification of existing code.

Once it entered production mode, the strengths and weaknesses of the NAMD design could be determined. C++, message-driven execution, and the concept of patches had each proven their utility and the program performed well on small numbers of processors. There were also some problems. Load balancing was hampered because most of the work was concentrated in a few patches near the center of the system (simulations lacked periodic boundary conditions). A patch with multiple neighbors on the same node would send several identical messages to that node; the workaround for this unnecessarily complicated the communication system. Finally, it was found that a patch-centric flow of control created a mixing of the essentially serial simulation algorithm with the parallel logic for responding to incoming messages, obfuscating both and requiring an understanding of the message structure in order to make trivial modifications to the iterative loop. For these reasons, it was decided that a major redesign was necessary and work began on NAMD 2.

NAMD 2

NAMD 2 added several new design goals. First, parallel performance needed to be increased through more parallelism and better load balancing. Second, communication efficiency needed to be improved without adding application-

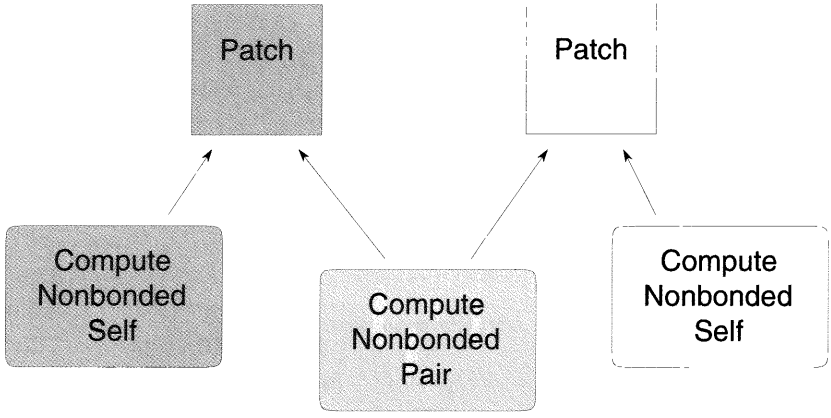


Fig. 4. In NAMD 2 forces are calculated not by force objects owned by individual patches, but rather by independent compute objects which depend on one or more patches for atomic coordinates. As suggested by shading in this illustration, a compute object need not reside on the same node as the patches upon which it depends.

specific code to the communication subsystem. Third, the simulation algorithm's outer loop should be made explicit and parallel logic in this section of code eliminated. Finally, the design needed to be able to take advantage of the eventual availability of kernel-level threads on a newly-acquired cluster of symmetric multiprocessor shared-memory workstations; a node would be able to control several processors in a common memory space.

NAMD 2 did not use PVM as its parallel communication protocol, switching instead to the Charm++/Converse system developed locally by the group of L. V. Kalé. While NAMD 1 simulated message-driven execution in PVM, Charm++ [19] provides direct support for NAMD's message-driven object paradigm and provides tools for analyzing the performance of parallel programs. (A Charm++ version of NAMD 1 was also implemented but maintaining both versions required too much manpower.) Converse [20] is an underlying communications layer which is portable to most parallel machines and features the ability to let multiple parallel languages coexist in a single code. This later feature allowed us to continue using the PVM-based DPTMA package [21]. Converse also incorporates multiple threads into its messaging system, the utility of which is described below. NAMD 2 also made aggressive use of C++ templates in order to provide efficient yet safe and convenient container classes and employed a more thoroughly object-oriented design.

In order to improve parallelism and load balancing, a hybrid force-spatial decomposition scheme was adopted in NAMD 2. Rather than decomposing the nonbonded computation into regions of space or pairwise atomic interactions, the basic unit of work was chosen to be interactions between atoms

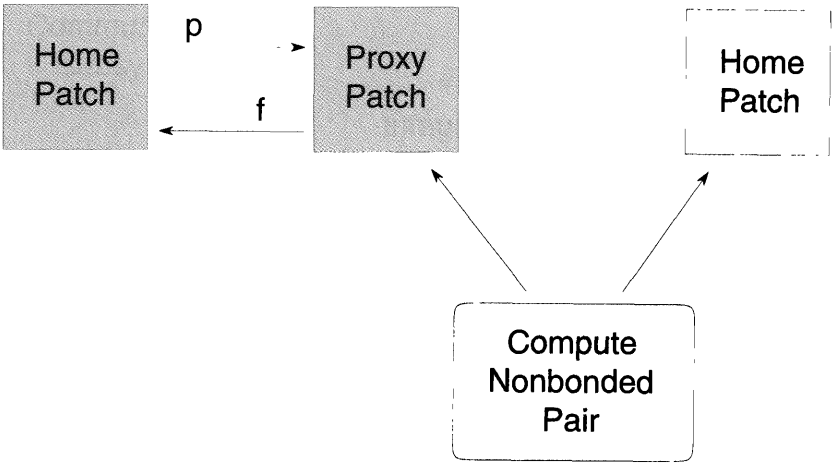


Fig. 5. Compute objects requiring off-node patches do not engage in off-node communication but rather interact with local proxy patches. When force evaluations are required the home patch sends positions messages (p) to its proxies and receives force messages (f) containing the results of off-node calculations. The proxy patch in this illustration exists on the same node as the compute object but represents the off-node home patch with which it communicates.

in regions of space. This was represented in the object-oriented design of NAMD 2 by moving responsibility for calculating forces from objects owned by a patch to more general *compute objects* that were responsible only for nonbonded interactions between atoms in a pair of patches, or within a single patch (Fig. 4).

Moving responsibility for the force computation away from the patches required a move away from pure message-driven execution to *dependency-driven execution* in which patches control the data (atomic coordinates) needed for compute objects to execute. A compute object, upon creation, registers this dependency with those patches from which it needs data. The patch then triggers force calculation by notifying its dependent compute objects when the next timestep's data is available. Once a compute object has received notification from all of the patches it depends on, it is placed in a prioritized queue for eventual execution.

Load balancing can then be achieved in NAMD 2 by moving compute objects and patches between nodes. But what if a compute object and a patch it depends on are on different nodes? Compute objects individually communicating with off-node patches would generate a huge amount of redundant communication. Therefore, patches are represented on other nodes by *proxy patches*, which implement the same interface as *home patches* for dealing with compute objects and handling dependencies but receive coordinates from and

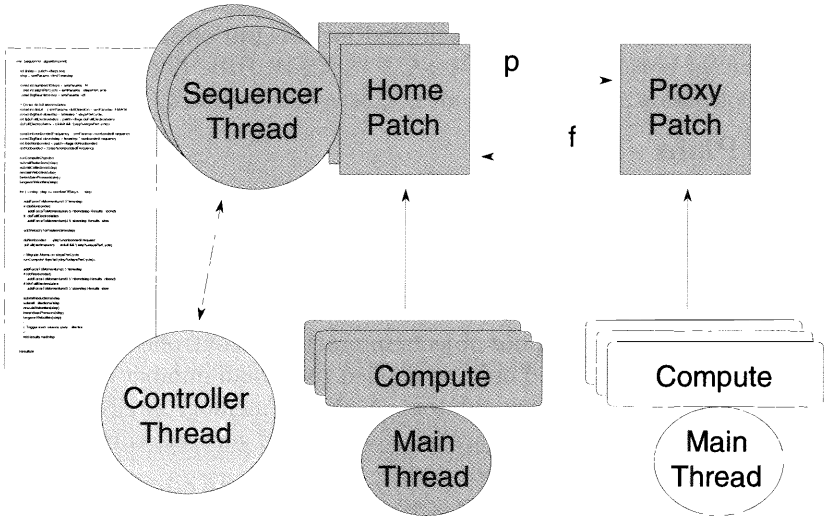


Fig. 6. Multiple threads in NAMD 2 allow the integration algorithm to be expressed sequentially as a single function. This function, shown illegibly at left, runs in sequencer threads associated with home patches. A similar function running in a controller thread on the master node communicates with the sequencers to deal with output and global calculations. Compute objects execute in the larger stack space of each node's main thread.

send forces to their respective home patches rather than performing integration themselves (Fig. 5). Thus data is replicated on those nodes where it is needed with a minimum of communication while no off-node communication is done by compute objects.

The logic associated with the patch has been greatly simplified by separating compute objects and limiting communication to patches and proxies, but one additional step is needed to fully separate sequential molecular dynamics algorithm from the complex logic of a message-driven parallel code. A *sequencer thread* is associated with every patch. This thread runs a single function which contains an explicit loop over all of the timesteps in the simulation (Fig. 6). In this way, the integration algorithm can be inspected in a single section of code closely resembling the outer loop of a serial molecular dynamics program. All of the parallel logic is hidden inside of a force evaluation function called by the sequencer that simply propagates coordinates to proxies and notifies all registered dependent compute objects that coordinates are available for calculating forces before suspending the sequencer thread. The thread is later awakened when all dependent compute objects and proxies have deposited their forces. A similar *controller thread* on the master node coordinates energy output and global aspects of the integration algorithm such as calculating velocity rescaling factors. Thread suspension is

also used to wait for unavailable data such as energies needed for output in the case of the controller or forces needed for integration in the case of the sequencer.

Future Plans

As noted above, one of the goals of NAMD 2 is to take advantage of clusters of symmetric multiprocessor workstations and other non-uniform memory access platforms. This can be achieved in the current design by allowing multiple compute objects to run concurrently on different processors via kernel-level threads. Because compute objects interact in a controlled manner with patches, access controls need only be applied to a small number of structures such as force and energy accumulators. A shared memory environment will therefore contribute almost no parallel overhead and generate communication equal to that of a single-processor node.

Although the current multithreaded implementation of sequencers works well and provides a clearly visible algorithm, threads have several drawbacks. Extra memory is required for multiple stacks, there is overhead from context-switching between threads, and a running sequencer cannot migrate between processors along with its patch. These problems will be addressed by using the Structured Dagger coordination language [22], which enables programmers to specify partial order between entry methods of an object. Using constructs such as overlap, forall, and when-blocks, one can easily express dependencies between entry methods of an object while letting the system do the buffering, bookkeeping, etc. required for the specified flow of control.

Finally, the ultimate in algorithmic flexibility can be achieved by the addition of a scripting language interface to NAMD. Such an interface, most likely based on Tcl [23], will allow the end user to modify the simulation algorithm without recompiling and to implement multi-stage simulation protocols in a single script. By adopting an existing scripting and extension language such as Tcl, Perl or Python [24], the end user will avoid learning a special-purpose language and enjoy the benefits of a well-designed and fully featured programming environment. The success of the Tcl interface in VMD [13], the Theoretical Biophysics Group's biomolecular visualization package, makes this line of development almost inevitable.

Acknowledgements

The primary developers of NAMD 1 were M. Nelson, W. Humphrey, A. Gursoy, A. Dalke and R. Brunner. The primary developers of NAMD 2 were J. Phillips, A. Shinozaki, R. Brunner, N. Krawetz, M. Bhandarkar and A. Gursoy. NAMD development was performed at the National Institutes of Health Resource for Concurrent Biological Computing under the supervision of principal investigators L.V. Kalé, R. Skeel, and K. Schulten. This work was

supported by the National Institutes of Health (NIH PHS 5 P41 RR05969-04 and NIH HL 16059) and the National Science Foundation (NSF/GCAG BIR 93-18159 and NSF BIR 94-23827 EQ). JCP was supported by a Computational Science Graduate Fellowship from the United States Department of Energy.

References

1. Amdahl, G. M.: Validity of the single processor approach to achieve large scale computing capabilities. *In* Proc. AFIPS spring computer conf. vol. 30. AFIPS Press, Reston, Virginia, 1967.
2. Allen, M. P., Tildesley, D. J.: Computer Simulation of Liquids. Oxford University Press, New York, 1987.
3. Brooks III, C. L., Karplus, M., Pettitt, B. M.: Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics. Advances in Chemical Physics, vol. LXXI. John Wiley & Sons, New York, 1988.
4. McCammon, J. A., Harvey, S. C.: Dynamics of Proteins and Nucleic Acids. Cambridge University Press, Cambridge, 1987.
5. Almasi, G. S., Gottlieb, A.: Highly Parallel Computing. 2nd edn. Benjamin/Cummings, Redwood City, California, 1994.
6. Theoretical Biophysics Group. http://www.ks.uiuc.edu/.
7. Nelson, M., Humphrey, W., Gursoy, A., Dalke, A., Kalé, L., Skeel, R. D., Schulten, K.: NAMD— A parallel, object-oriented molecular dynamics program. J. Supercomputing App. **10** (1996) 251–268.
8. Lin, M., Hsieh, J., Du, D. H. C., Thomas, J. P., MacDonald, J. A.: Distributed network computing over local ATM networks. *In* Proceedings of Supercomputing '94. IEEE Computer Society Press, Los Alamitos, California, 1994.
9. Greengard, L., Rokhlin, V.: A fast algorithm for particle simulation. J. Comp. Phys. **73** (1987) 325–348.
10. Rankin, W., Board, J.: A portable distributed implementation of the parallel multipole tree algorithm. IEEE Symposium on High Performance Distributed Computing. Duke University Technical Report 95-002.
11. Geist, A., Beguelin, A., Dongarra, J., Jiang, W., Manchek, R., Sunderam, V.: PVM: Parallel Virtual Machine: A Users' Guide and Tutorial for Networked Parallel Computing. MIT Press, Cambridge, Massachusetts, 1994.
12. Snir, M., Otto, S., Huss-Lederman, S., Walker, D., Dongarra, J.: MPI: The Complete Reference. MIT Press, Cambridge, Massachusetts, 1995.
13. Humphrey, W. F., Dalke, A., Schulten, K.: VMD – Visual molecular dynamics. J. Mol. Graphics. **14** (1996) 33–38.
14. Haney, S. W.: Is C++ fast enough for scientific computing? Computers in Physics. **8** (1994) 690–694.
15. Clark, T., Hanxleden, R., McCammon, J., Scott, L.: Parallelizing molecular dynamics using spatial decomposition. *In* Proceedings of the scalable high performance computing conference, May 23-25, 1994, Knoxville, Tennessee. IEEE Computer Society Press, Los Alamitos, California, 1994.
16. Plimpton, S., Hendrickson, B.: A New Parallel Method for Molecular Dynamics Simulation of Macromolecular Systems. 1994. Technical Report SAND94-1862. Sandia National Laboratories.

17. Hockney, R. W., Eastwood, J. W.: Computer Simulation Using Particles. McGraw-Hill, New York, 1981.
18. Kalé, L. V.: The Chare Kernel parallel programming language and system. *In* Proceedings of the international conference on parallel processing vol. II. CRC Press, Boca Raton, Florida, 1990.
19. Kalé, L., Krishnan, S.: Charm++: A Portable Concurrent Object Oriented System Based on C++. *In* Proceedings of the Conference on Object Oriented Programming Systems, Languages and Applications. A. Paepcke, editor. ACM Press, New York, N.Y., 1993.
20. Kalé, L. V., Bhandarkar, M., Jagathesan, N., Krishnan, S., Yelon, J.: Converse: An interoperable framework for parallel programming. *In* Proceedings of the 10th international parallel processing symposium. IEEE Computer Society Press, Los Alamitos, California, 1996.
21. Kalé, L. V., Bhandarkar, M., Brunner, R., Krawetz, N., Phillips, J., Shinozaki, A.: NAMD: A case study in multilingual parallel programming. *In* Proceedings of the 10th international workshop on languages and compilers for parallel computing. Springer-Verlag, Berlin, 1998.
22. Kalé, L. V., Bhandarkar, M.: Structured Dagger: A coordination language for message-driven programming. *In* Proceedings of the second international europar conference. Lecture Notes in Computer Science, vol. 1123–1124. Springer-Verlag, Berlin, 1996.
23. Ousterhout, J.: Tcl and the Tk Toolkit. Addison-Wesley, Reading, Massachusetts, 1994.
24. Watters, A., Rossum, G. V., Ahlstrom, J. C.: Internet Programming With Python. M & T Books, Sebastopol, California, 1996.

Parallel Molecular Dynamics Using Force Decomposition

Daniel Okunbor and Ravi Murty

Department of Computer Science
University of Missouri-Rolla
Rolla, MO 65401

Abstract. Research interests in molecular dynamics (MD) and its applications have increased significantly over the past few decades. This is due partly to the advances in software and hardware components of computer technology. The main computational goal of recent research work in molecular dynamics has been to reduce the computational cost of the force calculations which evidently accounts for approximately ninety percent of the total CPU time for most MD simulations. This paper describes parallel algorithms for force calculations using the force decomposition approach. These parallel algorithms have been tested and found to be highly portable and scalable. Numerical experiments on IBM SP/2 indicate that these algorithms have improved speedups and efficiencies.

1 Introduction

Molecular dynamics (MD) studies the time evolution of N interacting particles via the solution of classical Newton's equations of motion.

$$m_i \frac{d^2 r_i(t)}{dt^2} = f_i, \ i = 1, 2, \ldots, N, \tag{1}$$

where $r_i = (x_i(t), y_i(t), z_i(t))^{\mathrm{T}}$ and $v_i = (\dot{x}_i(t), \dot{y}_i(t), \dot{z}_i(t))^{\mathrm{T}}$ are respectively, the position and velocity vectors of the i-th particle at time t and m_i is the mass. The ultimate goal being to evaluate the dynamical structure of particles in order to reveal the chaotic characteristics of the system as in solar systems or the conformational analyses of the system as in biomolecules. During the time of the simulation, different measures are employed and these measures are quantitative in nature and apparently, it is the desire of computational scientists that these quantities be reasonably accurate so that inferences drawn on them are close to being realistic as much as possible.

Although, the notion of molecular dynamics was known in the early turn of the century, the first conscious effort in the use of computer for molecular dynamics simulation was made by Alder and Wainright, who in their paper [1] reported the application of molecular dynamics to realistic particle systems. Using hard spheres potential and fastest computers at the time, they were able to simulate systems of 32 to 108 atoms in 10 to 30 hours. Since the work of Alder and Wainright, interests in MD have increased tremendously, see

[2]-[37]. This rapid increase in research interests is due largely to the growth in computer technology, particularly, in software and hardware advances in multiprocessor technology.

Molecular dynamics conceptually involves two phases, namely, the force calculations and the numerical integration of the equations of motion. In the first phase, force interactions among particles based on the negative gradient of the potential energy function U,

$$f_i = -\nabla_{r_i} U(r_1, r_2, \ldots, r_N),$$

of the particles are computed. The numerical integrator to update particle positions and momenta, that is commonly used because of its simplicity, second-order of accuracy, reasonable stability and symplectic property (see [26, 27] for details) is the Störmer-Verlet method

$$\begin{aligned} v_{i,n+\frac{1}{2}} &= v_{i,n} + \frac{\Delta t}{2m_i} f_{i,n}, \\ r_{i,n+1} &= r_{i,n} + \Delta t v_{i,n+\frac{1}{2}}, \\ v_{i,n+1} &= p_{i,n+\frac{1}{2}} + \frac{\Delta t}{2m_i} f_{i,n+1}, \end{aligned}$$

where $r_{i,n} \approx r_i(t_n)$, $v_{i,n} \approx v_i(t_n)$ and $f_{i,n} \approx f_i(t_n)$ are the approximate position, velocity and force vectors, respectively, of the i-th particle at time $t_n = n\Delta t$ and Δt is the time step. In biomolecular modeling, the potential energy function is composed of the bonded and non-bonded interaction terms,

$$U = U_{\text{bonded}} + U_{\text{non-bonded}}.$$

For a detailed discussion of the force components, see [23]. The bonded interactions act only with nearest neighbor particles and therefore have linear time complexity. The nonbonded interactions comprise of the short-range interactions as in van der Waals potential and long-range interactions as in electrostatic (or Columbic) potential. While it is possible to obtain linear time complexity for the short-range interactions using the distance cut-off strategy, the long-range interaction have quadratic time complexity. This is the bottleneck present in nearly all realistic MD simulations, particularly, when large number of particles are involved.

Computational issues that are pertinent in MD simulations are time complexity of the force calculations and the accuracy of the particle trajectories including other necessary quantitative measures. These two issues overwhelm computational scientists in several ways. MD simulations are done for long time periods and since numerical integration techniques involve discretization errors and stability restrictions which when not put in check, may corrupt the numerical solutions in such a way that they do not have any meaning and therefore, no useful inferences can be drawn from them. Different strategies such as globally stable numerical integrators and multiple time steps implementations have been used in this respect (see [27, 31]).

The computational complexity of most MD as pointed out in the preceding paragraphs is dominated by the calculations of $O(\frac{N^2}{2})$ pairwise interactions in the force. The rapid development in parallel computer architecture and fast parallel algorithms have provided ways to reduce this complexity significantly allowing for the simulation of systems of large number of particles. The simulation of system with one billion particles have been reported in the literature. These simulations are academic in nature, practical simulations of more realistic system consisting of one billion particles is still underway. Perhaps the impediment one faces in this are the complexity and ever changing nature of parallel architecture and the fast parallel algorithms. It is evidence from the amount of research in parallel molecular dynamics, that distributed memory multiprocessor systems (MIMD) is an acceptable platform for doing molecular dynamics. This is further harness with the improvement in message-passing programming software, making code portability possible across different platforms. Fast parallel algorithms using different techniques to distribute problem tasks to processors have been developed. Classifications that are currently being used for parallel molecular dynamics simulations are *particle/atom*, *force/interaction* and *domain/spartial* decomposition. The particle decomposition algorithm distributes particles to processors irrespective of their physical positions in the computational domain. This approach generally uses systolic topology for interprocessor communication when replicated data is not possible in each processor. In the force decomposition technique, the interaction terms in the skew-symmetric force matrix

$$f = \begin{bmatrix} 0 & f_{12} & f_{13} & f_{14} \cdots & f_{1N} \\ -f_{12} & 0 & f_{23} & f_{24} \cdots & f_{2N} \\ -f_{13} & f_{23} & 0 & f_{34} \cdots & f_{3N} \\ -f_{14} & -f_{24} & -f_{34} & 0 \cdots & f_{4N} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -f_{1N} & -f_{2N} & -f_{3N} & -f_{4N} \cdots & 0 \end{bmatrix}$$

are assigned to processors in *fine grain* (interaction terms are sent to processors one at a time-a processor must complete the assigned interaction term before being assigned another interaction term), *medium grain* (rows/columns of the force matrix are are assigned to processors one at a time) or *coarse grain* (blocks of rows/columns or subblocks of the force matrix are sent to processors to compute one at a time). The domain decomposition partitions the computational domain into subdomains and assigning particles within a subdomain to a processor. The approach used in fast multipole algorithm (FMA) of Greengard and Rokhlin [11], the particle-particle, particle-mesh algorithm (PPPM) of Hockney and Eastwood [17] and hierarchical-tree algorithm of Barnes and Hut [3]. The fast multipole algorithm which supports generally the electrostatic force and uses Taylor series expansions to approximate far-field of the force function and which, when systematically combined

with hierarchical-tree structure produces an $O(N)$ algorithm with a large constant term.

In this paper, we would describe parallel algorithms for force calculations using the force decomposition technique. These parallel algorithms (which are described in Section 2) are the **checkerboard partitioning** developed by Taylor *et al.* [12] following the concept suggested by Plimpton [28] the **force-row interleaving** and **force-stripped row** methods developed by Murty and Okunbor [22]. All three parallel algorithms have been tested and found to be highly portable and scalable. Numerical experiments on IBM SP/2 described in Section 4, indicate that these algorithms have improved speedups and efficiencies.

2 Parallel Force Decomposition

The force decomposition algorithm maps all possible interactions to processors and does not require inter-processor communication during the force calculation phase of MD simulation. However, to obtain the net force on each particle for the update phase would need global communication. In this section, we will present parallel algorithms based on force decomposition.

2.1 Checkerboard Partitioning Method

This approach is based on the scheme suggested by Taylor *et al.* [12] and Plimpton [28]. The force matrix is divided into $\sqrt{P} \times \sqrt{P}$ blocks, where P is number of processors. The processors are conceptually thought of as having a two-dimensional mesh topology, that is, $\sqrt{P} \times \sqrt{P}$-mesh. Note that this is not the physical architecture of the parallel system. This arrangement is used only to describe the algorithm. Since the force matrix has $N(N-1)$ pairs to be computed, each processor is assigned $\frac{N}{\sqrt{P}} \times \frac{N-1}{\sqrt{P}}$ interactions, contained within a single (i,j)-block of the force matrix. Let us use P_{ij} to denote the processor that is assigned the (i,j)-block, and P_{ji} to denote its *transpose* processor. As indicated in the the preceding sections, the force matrix is skew symmetric, therefore, only the upper (or lower) triangular part of the force matrix must be used in order to remove unnecessary calculations. In light of this, the interactions in a given block of the force matrix are distributed to the processor and its transpose processor. Inter-processor communications are done only among processors in the same row and transpose processors. Note that the processors on the diagonal are responsible for necessary interactions in the diagonal blocks, since they do not have transpose processors.

Let us illustrate this with the mapping of a system of 16 particles on 16 processors. The diagonal processor P_{11} computes interactions in the $(1,1)$-block, which are interactions among particles $(1,2,3,4)$, processor P_{22} computes interactions among particles $(5,6,7,8)$, processor P_{33} computes interactions among particles $(9,10,11,12)$ and processor P_{44} computes interactions

among particles $(13, 14, 15, 16)$. For the off-diagonal blocks, processors P_{12} and P_{21} share computations of interactions between particles in $(1, 2, 3, 4)$ and $(5, 6, 7, 8)$, processors P_{13} and P_{31} share interactions between particles in $(1, 2, 3, 4)$ and $(9, 10, 11, 12)$, processors P_{14} and P_{41} share interactions between particles in $(1, 2, 3, 4)$ and $(13, 14, 15, 16)$, processors P_{23} and P_{32} share interactions between particles in $(5, 6, 7, 8)$ and $(9, 10, 11, 12)$, and so on. At the end of all interaction calculations, processors communicate with transpose processors and processors in the same row.

For this algorithm, each processor is assigned $\frac{3}{2}\frac{N}{\sqrt{P}}$ atoms, so the force calculation time is $O(\frac{N^2}{P})$. Using the communication scheme mentioned above, each processor communicates with $(\sqrt{P}-1)$ processors in each row and column. Thus the total number of terms being communicated per step is $(\sqrt{P}-1)(\frac{3}{2}\frac{N}{\sqrt{P}})$. Therefore, $O(N)$ CPU time is required in communicating the net force per step. Therefore,

$$T_p = \underbrace{O(\frac{N^2}{P})}_{\text{Computation}} + \underbrace{O(\frac{3}{2}N - \frac{3}{2}\frac{N}{\sqrt{P}})}_{\text{Communication}}. \tag{2}$$

2.2 Force-Row Interleaving Method

One of the important factors in the design of algorithms for parallel systems is the issue of load balance. This can be emphasized by considering the following simple case. Since most of the algorithms execute synchronously, if one processor finishes 10% earlier than the rest of the group, there will be no major effect on the overall efficiency of the parallel application. However, if one processor finishes 10% later than the rest of the processors, the efficiency of the application will drop drastically. The next two algorithms presented in this section achieve load balance by dividing the computational load equally among processors. Details of these algorithms are found in the paper by Murty and Okunbor [22].

The first algorithm is called the force-row interleaving method. The algorithm consists of a sequence of row assignments. In the *first assignment*, processor k is assigned row k of the force matrix, $0 \leq k \leq (P-1)$. Let P_k denote a processor with rank k. This means that in the first assignment processor P_k will compute $(N-k-1)$ interaction between atoms. If all processors start their first assignment and work in parallel, then intuitively, processor $P-1$ is expected to finish computing all interaction assigned to it, since it computes one interaction less than processor $P_{(p-2)}$, two interactions less than $P_{(p-3)}$ and so on. Processors that complete their *first assignment* start computing interactions for the next available row - their *second assignment*. The sequence of row assignments continues until all rows of the force matrix have been computed.

It can be observed from the above discussion that processor P_k, $0 \leq k \leq (P-1)$ in general will be responsible for computing interactions for rows k, $2P-K-1$, $2P+K$, $4P-K-1, \ldots$

The complexity of the force computation function described here is computed by counting the number of force terms evaluated by each processor. The cost of force computation for processor i, which is denoted by C_i is,

$$C_i = \frac{11N^2}{16P} - \frac{N}{2P} - \frac{3N}{4}. \tag{3}$$

Since every processor computes the same number of computations, the above equation gives an the time spent in computing forces $\approx O(\frac{N^2}{P})$. At the end of each force step, a global all-to-all reduction operation updates the entire force vector. Since the size of the force vector is N, this requires $O(N)$ time. Therefore the overall time complexity of this method is,

$$T_p = \underbrace{O(\frac{N^2}{P})}_{\text{Computation}} + \underbrace{O(N)}_{\text{Communication}} . \tag{4}$$

The complexity analysis shows that the load is evenly balanced among processors and therefore we should expect speedup close to P and efficiency close to 100%. There are however few extra terms in the expression of the time complexity (first order terms in N), that exist because of the need to compute the next available row in the force matrix. These row allocations can be computed ahead of time and this overhead can be minimized. This is done in the next algorithm. Note that, the communication complexity is the worst case for all interconnection topologies, since simple broadcast and gather on distributed memory parallel systems are assumed.

2.3 Force-Stripped Row Method

This algorithm is an improvement over the algorithms described in the previous subsections. The idea behind this algorithm is fairly simple. To ensure load balance, the rows of the force matrix will be allocated in such a way that the load on all processor is equal.

As before, let C_i, denote the cost of force computation on processor i, $0 \leq i \leq (P-1)$. Processor i is assigned l_i rows of the force matrix and for load balance $l_1, l_2, \ldots, l_P$ will satisfy $l_1 \leq l_2 \leq l_3 \leq \ldots \leq l_P$. This algorithm computes *a priori* the row assignment so that the load sent to processors is balanced. Some typical values are listed in Table 6.

The total number of force terms calculated by processor i is given by

$$C_i = (N - \sum_{j=1}^{i-1} l_j - \frac{l_i + 1}{2}) l_i. \tag{5}$$

For complete load balance,

$$C_1 = C_2 = C_3 = \ldots = C_p,$$
$$l_1 + l_2 + l_3 + \ldots + l_p = N. \tag{6}$$

These quadratic equations are solved for l_i to give,

$$l_i = \frac{t_i - \sqrt{(t_i)^2 - \frac{4}{P}(N^2 - N)}}{2},$$

where

$$t_i = (2N - 1 - 2\sum_{j=1}^{i-1} l_j). \tag{7}$$

Generally, the computed values of $l_1, l_2, \ldots, l_P$ are not integers and are therefore rounded to nearest integers. In particular, $l_1, l_2, \ldots, l_{P-1}$ are rounded to the nearest integers $\lceil l_1 \rceil, \lceil l_2 \rceil, \ldots \lceil l_{P-1} \rceil$ and $l_P = (N - \sum_{j=1}^{j=P-1} \lceil l_j \rceil)$ rows are assigned to the last processor ($P_{(P-1)}$).

The time complexity of this approach can be calculated easily by substituting equation(7) in equation(5). The result is,

$$C_k = \frac{(N^2 - N)}{2P} \approx O(\frac{N^2}{P}).$$

Since this approach maps all possible interactions to processors, no communication is required during force calculation. Moreover, the row assignments are completed before the first step of the simulation. The computation of the bounds for each processor require $O(P^2)$ time, which is very negligible compared to N (for $N \gg P$). The communication required at the end of each step to update the position and velocity vectors is done by reducing force vectors of length N, and therefore scales as $O(N)$ per node per time step. Thus the overall complexity of this algorithm is,

$$T_p = \underbrace{O(\frac{N^2}{P})}_{Computation} + \underbrace{O(N)}_{Communication} \tag{8}$$

The time complexity of this algorithm shows that the force computation does not involve any extra overheads and therefore, the speedup should be equal to P and efficiency 100% in theory.

3 Benchmark System

We consider a Lennard-Jones fluid consisting of atoms interacting with a Lennard-Jones potential given by

$$U(r) = 4\epsilon[(\frac{\sigma}{r})^{12} - (\frac{\sigma}{r})^6],$$

where ϵ is the depth of the potential well at the minimum in $U(r)$ and σ is the collision diameter. For liquid argon,

$$\sigma = 3.405\text{Å}, \quad \epsilon = 120^{\circ}\text{K} \times k_{\text{B}} \text{ and } m = 39.95\,\text{a.m.u.}$$

where $1\,\text{a.m.u} = 1.66057 \times 10^{-24}\text{g}$ is the atomic mass unit, $k_{\text{B}} = 1.38064 \times 10^{-16}\text{erg/mole/}^{\circ}\text{K}$ is the Boltzmann constant and m is mass of an atom. We simulate atoms in a cubic box and select the number of atoms N so that periodic boundary conditions permit a perfect lattice appropriate for the physical system under investigation. Liquid argon crystalizes in a face-centered cubic (fcc) structure, that is, an atom is present at each of the eight corners and one in each middle of the six sides of the cubic box. It is therefore natural to choose $N = 4k^3$, where k is integer. This way, the cubic box can be divided into k^3 smaller cubic boxes and 4 atoms assigned to each smaller cubic box with one atom in each middle of the three visible faces and one in the corner.

4 Results on the IBM SP/2

This section presents some of the simulation results obtained by simulating systems of sizes 4000, 6912, 10976, 16384 and 32000 atoms on the IBM-SP/2. The simulations were performed on 4, 8 and 16 processors, respectively. Although, the simulated system size and the number of processors can be scaled easily, this section does not show all results.

Table 1 describes the timing results (in seconds) for a system of 4000 atoms on 4, 8 and 16 nodes. The average CPU seconds for 10 time steps per processor is calculated. In the case of the force-stripped row and force-row interleaving algorithms the CPU time is reduced by half each time the number of processors is doubled. This indicates a perfect speedup and efficiency as described in Table 2. Tables 3, refibm:table3 and 5 describe the timing results, speedups and efficiencies for larger systems. In particular, Table 4 shows the scaling in the CPU time with increase in the system size. These results are very close to predicted theoretical results.

Lastly, Table 6 describes the assignment of rows to processors for some typical cases, and the load in each case (indicating the number of force interactions computed by each processors in the corresponding case). These are based on equations in Section 3. Several important points can be noted from the results shown in the table. Firstly, it can be observed that in the 4 processor case, processor P_3 computes half the maximum number of rows in the force matrix which leads to a load balanced assignment. This would not be the case if processors were assigned equal number of rows. Moreover, when the number of processors is increased from 4 to 16, the load on each processor reduces by a factor of 4, but is still equal on every processor.

Table 1. CPU Timings (in seconds) for a 4000 atom system, on 4, 8 and 16 IBM-SP/2

	P = 4		P = 8		P = 16	
Algorithm	$\mathbf{T_{force}}$	T_{Commun}	T_{force}	T_{Commun}	T_{force}	T_{Commun}
Force-Row Interleaving	7.622	0.0097	3.8090	0.0125	1.908	0.0182
Checkerboard	7.7063	0.0900	-	-	1.999	0.0763
Force-Stripped Row	7.617	0.0078	3.800	0.0105	1.909	0.01723

Table 2. Speedup and Efficiency results for a system of 4000 atoms on 4, 8 and 16 processors

	P = 4		P = 8		P = 16	
Algorithm	S_4	$E_4(\%)$	S_8	$E_8(\%)$	S_{16}	$E_{16}(\%)$
Force-Row Interleaving	3.897	99.68	7.899	98.74	15.494	96.83
Checkerboard	3.937	99.30	-	-	15.27	95.43
Force-Stripped Row	3.9960	99.90	7.949	99.30	15.51	96.88

Table 3. CPU Timings (in seconds) for a 10976 atom system, on 4, 8 and 16 IBM-SP/2

	P = 4		P = 8		P = 16	
Algorithm	$\mathbf{T_{force}}$	T_{Commun}	T_{force}	T_{Commun}	T_{force}	T_{Commun}
Force-Row Interleaving	57.55	.0.0202	28.80	0.0282	14.40	0.0436
Checkerboard	57.82	0.4936	-	-	14.78	0.4233
Force-Stripped Row	57.50	0.0199	28.77	0.0269	14.33	0.0918

Table 4. CPU Timings (in seconds) for a 16384 and 32000 atom systems 16 IBM-SP/2 nodes

	N= 16384		N = 32000	
Algorithm	$\mathbf{T_{force}}$	T_{Commun}	T_{force}	T_{Commun}
Force-Row Interleaving	32.105	0.1168	122.45	0.1137
Checkerboard	32.30	0.8841	123.05	2.286
Force-Stripped Row	32.005	0.0916	122.22	0.1026

References

1. B. J. Alder and T. E. Wainright, "Studies in molecular dynamics II: Behavior of a small number of elastic spheres", *J. Chem. Phys.*, vol 33, 1439-1447, 1960.
2. M. P. Allen and D. J. Tildesley, Computer Simulation Of Liquids, Oxford science publication 1987.
3. J. Barnes and P. Hut, "A hierarchical O($N \log N$) force calculation algorithm", *Nature*, Vol 324, 446-49, 1986.

Table 5. Speedup and Efficiency results for a system of 16384 and 32000 atoms 16 processors

	N = 16384		N = 32000	
Algorithm	S_{16}	$E_{16}(\%)$	S_{16}	$E_{16}(\%)$
Force-Row Interleaving	15.89	99.30	15.89	99.30
Checkerboard	15.648	97.80	15.649	97.81
Force-Stripped Row	15.91	99.44	15.95	99.72

Table 6. Table showing the assignment of rows of the force matrix for 4 processors.

Row assignment	**N = 4000**	**N = 10976**	**N = 32000**
l_0	536	1471	4287
l_1	636	1744	5085
l_2	829	2273	6628
l_3	1999	5488	16000
Total	4000	10976	32000
Load per processor	2.000500e+06	1.506044e+07	1.280040e+08

4. D. Brown, J. H. R. Clarke, M. Okuda and T. Yamazaki, "A domain decomposition strategy for molecular dynamics simulations on distributed memory machines", *Comp. Phys. Comm.*, Vol 74, 67-80, 1993.
5. T. W. Clark, J. A. McCammon, L. R. Scott, "Parallel molecular dynamics", *Proc. of the fifth SIAM conference on Parallel Processing for Scientific Computing*, 338-44, 1992.
6. R. Duncan, "A survey of parallel computer architectures", *Computer*, Vol 23, no 2, 5-16, 1990.
7. D. Fincham, "Choice of timestep in molecular dynamics simulation", *Comp. Phys. Comm.*, Vol 40, no 2&3, 263-9, 1986.
8. D. Fincham and B. J. Ralston, "Molecular dynamics simulation using the Cray-1 vector processing computer", *Comp. Phys. Comm.*, Vol 23, no 2, 127-34, 1981.
9. G. C. Fox, M. A. Johnson, G. A. Lyzenga, S. W. Otto, J. K. Salmon and D. W. Walker, Solving Problems On Concurrent Processors: Volume 1, Prentice Hall, Englewood Cliffs, 1988.
10. S. K. Gray, D. W. Noid and B. G. Sumpter, "Symplectic integrators for large scale molecular dynamics simulations: A comparison of several explicit methods", *J. Chem. Phys.*, Vol 101, no 5, 4062-72, 1994.
11. L. Greengard and V. Rokhlin, "A fast algorithm for particle simulations", *J. Comput. Phys.*, Vol 73, no 2, 325-48, 1987.
12. D. L. Greenwell, R. K. Kalia, J. C. Patterson and P. Vashishta, "Molecular dynamics algorithm on the connection machine", *Int. J. High Speed Computing*, Vol 1, no 2, 321-8, 1989.
13. W. Gropp, E. Lusk and A. Skjellum, Using MPI Portable Parallel Programming with Message-Passing Interface, Scientific and Engineering Computation series, 1994.
14. S. Gupta, "Computing aspects of molecular dynamics simulation", *Comp. Phys. Comm.*, Vol 70, no 2, 243-70, 1992.

15. J. M. Haile, Molecular Dynamics Simulation, Elementary methods, John Wiley and sons, 1992.
16. P. A. J Hilbers and K. Esselink, "Parallel computing and molecular dynamics simulations", *Computer Simulations in Chemical Physics, Proc. of the NATO advanced study institute on new perspectives in computer simulations in chemical physics*, 473-95, 1993.
17. R. W. Hockney and J. W. Eastwood, Computer Simulations Using Particles, Institute of Physics Publishing, Bristol, 1988.
18. Y. Hwang, R. Das, F. H. Saltz, M. Hadošček and B. R. Brooks, "Parallelizing molecular dynamics programs for distributed-memory machines", *IEEE Computational Science and Engineering* , Vol 2, no 2, 18-29, 1995.
19. D. Janežič and F. Merzel, "An efficient symplectic integration algorithm for molecular dynamics simulations" , *J. Chem. Info. Comp. Sci.*, Vol 35, no 2, 321-6, 1995.
20. D. Janežič and R. Trobec, "Parallelization of an implicit Runge-Kutta method for molecular dynamics integration", *J. Chem. Info. Comp. Sci.* , Vol 34, no 3, 641-6, 1994.
21. T. G. Mattson and G. R. Shanker, "Portable molecular dynamics software for parallel computing", *ACS Symposium Series 592*, 133-50.
22. R. Murty and D. Okunbor, "Efficient parallel algorithms for molecular dynamics simulations", submitted to *Parallel Computing.*
23. D. W. Noid, B. G. Sumpter, B. Wunderlich and G. A. Pfeffer, "Molecular dynamics simulations of polymers: Methods for optimal Fortran programming", *J. Comput. Chem.*, 11(2), 236-241, 1990.
24. D. Okunbor, "Integration methods for N-body problems", *Proc. of the second International Conference On Dynamic Systems*, 1996.
25. D. Okunbor, "Parallel molecular dynamics on connection machines", *Wuhan J. Natural Sci.*, Vol 11, no 3&4, 337-43, 1996.
26. D. Okunbor, "Canonical methods for Hamiltonian systems: Numerical experiments", *Physica D*, 60, 314-322, 1992.
27. D. Okunbor and R. Skeel, "Canonical numerical methods for molecular dynamics simulations", *J. Comput. Chem.*, 15(1), 72-79, 1994.
28. S. Plimpton, "Fast parallel algorithms for short-range molecular dynamics", *J. Comput. Phys.*, Vol 117, no 1, 1-19, 1995.
29. S. Plimpton and B. Hendrickson, "A new parallel method for molecular dynamics simulation of macromolecular systems", *J. Comput. Chem.*, Vol 17, no 3, 326-37, 1996.
30. H. Schreiber, O. Steinhauser and P. Schuster, "Parallel molecular dynamics of biomolecules", *Parallel Computing*, Vol 18, no 5, 557-73, 1992.
31. R. D. Skeel and J. J. Biesiadecki, "Symplectic integrations with variable stepsize", *Annals Numer. Math.*, 191-198, 1994.
32. W. Smith, "A replicated data molecular dynamics strategy for the parallel Ewald sum", *Comp. Phys. Comm.*, Vol 62, no 3, 392-406, 1992.
33. W. Smith, "Molecular dynamics on hypercube parallel computers", *Comp. Phys. Comm.*, Vol 62, no 2&3, 229-48, 1991.
34. W. Smith and T. R. Forester, "Parallel macromolecular simulations and the replicated data strategy", *Comp. Phys. Comm.*, Vol 79, no 1, 52-62, 1994.
35. V. E. Taylor, R. L. Stevens and K. E. Arnold, "Parallel molecular dynamics: Communication requirements for massively parallel machines", *Proc. Frontiers*

'95, the fifth symposium on the frontiers of Massively Parallel Computation, 156-63, 1994.

36. R. Trobec, I. Jerebic and D. Janežič, "Parallel algorithms for molecular dynamics integration", *Parallel Computing*, Vol 19, no 9, 1029-39, 1993.
37. A. Windemuth, "Advanced algorithms for molecular dynamics simulations: The program PMD", *ACS Symposium Series 592*, 151-69, 1995.