Quarterly Reviews of Biophysics **36**, 3 (2003), pp. 257–306. © 2003 Cambridge University Press
DOI: 10.1017/S0033583503003895 Printed in the United Kingdom


# Advances in biomolecular simulations: methodology and recent applications

Jan Norberg* and Lennart Nilsson*

Center for Structural Biochemistry, Department of Biosciences at Novum, Karolinska Institutet, SE-141 57 Huddinge, Sweden

**Abstract.** Molecular dynamics simulations are widely used today to tackle problems in biochemistry and molecular biology. In the 25 years since the first simulation of a protein computers have become faster by many orders of magnitude, algorithms and force fields have been improved, and simulations can now be applied to very large systems, such as protein–nucleic acid complexes and multimeric proteins in aqueous solution. In this review we give a general background about molecular dynamics simulations, and then focus on some recent technical advances, with applications to biologically relevant problems.



* Correspondence may be addressed to either author.
J. Norberg: Tel.: +46-8-6089263; Fax: +46-8-6089290; E-mail: Jan.Norberg@biosci.ki.se
L. Nilsson: Tel.: +46-8-6089228; Fax: +46-8-6089290; E-mail: Lennart.Nilsson@biosci.ki.se

## 1. Introduction

In the five decades since molecular dynamics (MD) simulations were first introduced and applied to elastic hard spheres (Alder & Wainwright, 1957). The method has proved to be very versatile with a wide range of applications, and today MD simulations are routinely used in studies of large and complex biological systems; an analysis of citation frequencies for some of the commonly used programs indicates that after a rapid initial increase the number of reports using these programs now has leveled off (Fig. 1). This evolution is in part due to an impressive development in computer power, but also to the continued development of new parameters for the energy function and new techniques for running a MD simulation. At this stage it is possible to investigate and elucidate microscopic properties of systems involving tens of thousands of atoms, of complex chemical reactions, on timescales of nanoseconds and beyond. On the horizon one may find MD simulations covering the dynamics of viruses, the protein-folding process, and various large protein complexes making up the molecular machines that we now begin to perceive in the biological cell. In this review we focus on what we believe are interesting technical advances in MD simulations of biological macromolecules over the last few years. We wish to stress that the review is not intended to fully cover all technical improvements. Here for instance technical advances for Brownian dynamics, Car–Parrinello dynamics, ligand docking, Poisson–Boltzmann calculations and pure quantum mechanics are not included.

*Abbreviations*: FM, fast multipole; MD, molecular dynamics; PME, particle–mesh Ewald; $P^3M$, particle–particle particle–mesh; r-RESPA, reversible reference system propagator algorithm; r.m.s., root mean square.

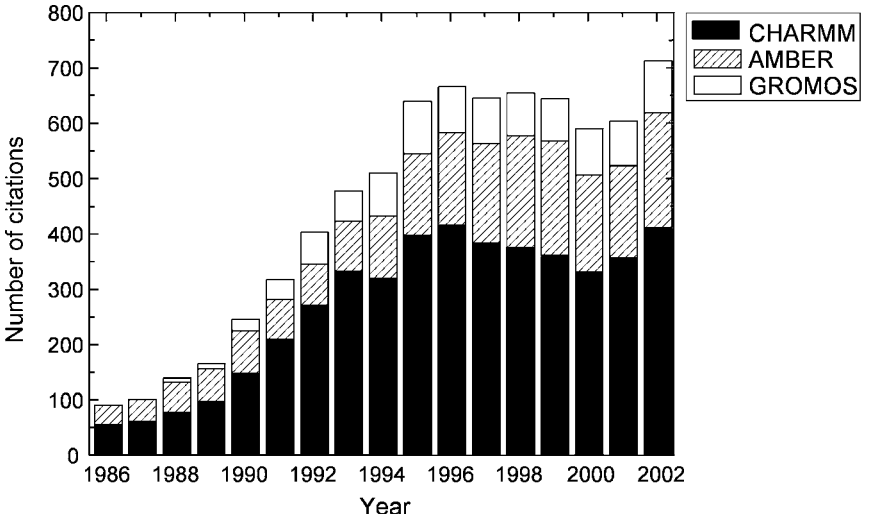


**Fig. 1.** Number of citations per year to the AMBER (▨), CHARMM (■), and GROMOS (□) programs 1986–2002. Analyzed from ISI Web of Science using the following references or search patterns: AMBER (Weiner & Kollman, 1981; Pearlman *et al.* 1995; 'Pearlman, AMBER, 1991|1995'; 'Case, AMBER, 1999'); CHARMM (Brooks *et al.* 1983); GROMOS (van Gunsteren & Berendsen, 1987; Scott *et al.* 1999; 'van Gunsteren, GROMOS|Groningen, 1996').

The first hard spheres simulations were soon followed by more realistic simulations of simple systems, using a Lennard–Jones potential for liquid argon (Rahman, 1964), or studies of complex liquids; the first simulation of a polyatomic liquid focused on water (Rahman & Stillinger, 1971). These studies are the background for the development of MD simulations of biological macromolecules, which were first presented with the simulation of the bovine pancreatic trypsin inhibitor (BPTI) *in vacuo* (McCammon, 1977; McCammon *et al.* 1977). Several years later the first nucleic-acid simulation was published (Levitt, 1983).

The early years of MD simulations of proteins and nucleic acids have been reviewed (McCammon & Harvey, 1987; Brooks *et al.* 1988; Karplus & Petsko, 1990), and this early period is also the subject of recent short reviews (Levitt, 2001; Karplus & McCammon, 2002; Karplus, 2003). The 25th anniversary of the first biomolecular MD simulation is marked by dedicated issues of the *Journal of Computational Chemistry* [**23** (1), 2002] and *Accounts of Chemical Research* [**35** (6), 2002].

Several reviews have covered the progress in MD simulations of nucleic acids (Auffinger & Westhof, 1998; Cheatham & Brooks, 1998; Cheatham & Kollman, 2000; Beveridge & McConnell, 2000; Norberg & Nilsson, 2002). Reviews concerning MD simulations of membranes and proteins are usually more specific, focusing on, for instance, membrane dynamics (Forrest & Sansom, 2000), protein hydration (Pettitt *et al.* 1998), protein folding (Brooks, 1998), viral capsid proteins (Phelps *et al.* 2000), collective protein motions (Berendsen & Hayward, 2000), and protein reaction dynamics (Karplus, 2000). Different aspects of MD simulation methodology are covered in a number of reviews (Kollman, 1996; Berne & Straub, 1997; Cheatham & Brooks, 1998; Levy & Gallicchio, 1998; Schlick, 1999; Schlick *et al.* 1999; Tuckerman & Martyna, 2000; Wang *et al.* 2001).

Modern MD simulations of proteins or nucleic acids are usually carried out with the solvent explicitly treated, typically allowing simulations in the nanosecond range, but recently a 1-μs

simulation of the villin headpiece subdomain was conducted (Duan *et al.* 1998; Duan & Kollman, 1998). Nowadays MD simulations have the power to gain insight into different kinds of chemical reactions at atomic level. Therefore the MD simulation technique has become quite powerful in elucidating features of biological macromolecules.

A number of different MD programs and force fields for simulating the dynamic behavior of biological molecules have been developed. Although most programs are closely associated with a specific force field, often developed together with the program, programs and force fields are in general separate entities and might thus, in principle, be used in a mixed fashion. If the functional forms used by two different force fields are the same such a mixed use could be accomplished with little more than a conversion of the format of the files, which store the force field.

Examples of commonly used programs for MD simulations biomolecules are AMBER (Cornell *et al.* 1995), CHARMM (Brooks *et al.* 1983), ENCAD (Levitt *et al.* 1995), GROMOS (van Gunsteren & Berendsen, 1987; Scott *et al.* 1999), NAMD (Nelson *et al.* 1996). The major packages have similar capabilities in terms of what systems can be studied and the kinds of simulations that they allow the user to perform: you need to be able to define and manipulate your molecular system, the actual simlation has to be performed and finally the results have to be analyzed. The way this is achieved varies. CHARMM is a single program which carries out all the different tasks involved as specified in a keyword-based English-like command language, whereas AMBER is a suite of programs that perform specific tasks, while GROMOS is a set of FORTRAN subroutines and a suite of programs, which may be combined in different ways.

## 2. Set-up of MD simulations

One important aspect of modern MD simulations is that the practitioner may choose other statistical mechanical ensembles than the microcanonical (Tuckerman & Martyna, 2000). For instance the Nosé–Hoover dynamics has been modified to include a chain of thermostating extended variables (Martyna *et al.* 1992). These dynamic equations produce a canonical ensemble. Below we consider the developments of techniques for constant-pressure dynamics, grand-canonical dynamics and boundary conditions.

### 2.1 Constant-pressure dynamics

Several methods have been developed for running MD simulations at constant pressure. Among the first methods developed for constant pressure are the extended system algorithm (Andersen, 1980; Nosé & Klein, 1983), the constraint algorithm (Evans & Morris, 1983, 1984), the weak coupling method (Berendsen *et al.* 1984), the hybrid method (Martyna *et al.* 1994), and the Langevin piston method (Feller *et al.* 1995). An integration scheme for MD simulations in an exact isobaric–isothermal ensemble has been suggested (Melchionna *et al.* 1993) based on the equations of motion formulated by Hoover (1985). Originally from the reversible reference system propagator algorithm (r-RESPA) another integration scheme has been developed for constant-pressure MD simulations (Procacci & Berne, 1994). The statistical mechanics ensemble of the MD simulations which use the weak coupling thermostat (Berendsen *et al.* 1984) has been investigated (Morishita, 2000), and found to approach the canonical ensemble for small coupling constants, but for large coupling constants the energy exchange was suppressed and therefore the ensemble approaches the microcanonical ensemble. A generalization of the Langevin piston

method to allow for stochastic motion has been presented and shown to produce an isobaric–isothermal ensemble (Kolb & Dünweg, 1999).

The isobaric–isothermal MD method of a system with holonomic constraints was analyzed (Ciccotti *et al.* 2001) using non-Hamiltonian statistical mechanics (Tuckerman *et al.* 1999). The equations by Kneller & Mülders (1996) were found to generate an isobaric–isothermal ensemble only in systems with non-zero total force (Ciccotti *et al.* 2001). Further a set of equations of motion for a proper isobaric–isothermal ensemble was described.

### 2.2 Grand-canonical dynamics

A general description of the grand-canonical Monte Carlo method has been presented (Valleau & Cohen, 1980). It was demonstrated that by fixing the chemical potential the average number or concentration can be determined. A decade ago the grand adiabatic ensemble method for MD simulations was introduced (Çağin & Pettitt, 1991a). In this method the number of particles is used as a continuous variable and therefore fractional molecules can be analyzed. A simple fluid, which had variable number of molecules, has been examined using this method (Çağin & Pettitt, 1991b). An extension of the grand-canonical method has been reported ( Ji *et al.* 1992). In this study the density dependence of the chemical potential of a water model was investigated. For grand-canonical ensemble simulations inclusion of umbrella sampling with a biasing in the extended Hamiltonian technique improves the efficiency of atomic annihilation and creation processes (Shroll & Smith, 1999). A comment concerning this report has compared the fluctuations of particle numbers between that MD simulation and the cavity-biased Monte Carlo method (Mezei, 2000). It was found that the Monte Carlo approach performed better than the MD simulation.

### 2.3 Boundary conditions

When a biological molecule is solvated in a cubic or rectangular box of water periodic boundary conditions are usually applied (Born & von Karmen, 1912; Allen & Tildesley, 1987; van Gunsteren & Berendsen, 1990). These conditions guarantee that all atoms are surrounded by neighbors, which are real atoms if they are inside the original box, and images of the atoms if they are outside. Periodic boundary conditions can also be applied for other geometries like the rhombic dodecahedron (Wang & Krumhansl, 1972) or the truncated octahedron (Adams, 1979). These can substantially reduce the size of the system and therefore decrease the required computational time. In the longest simulation (1 μs) of a protein, thus far, the truncated octahedron box was used (Duan & Kollman, 1998). Alternative boundary conditions, which allow non-periodic geometries, like the stochastic boundary conditions can also be applied (Kratky, 1980; Berkowitz & McCammon, 1982). This procedure has been applied to energy minimizations of a protein (Brooks & Karplus, 1983) and to simulations of protein dynamics (Brooks *et al.* 1985a; Brünger *et al.* 1985). Periodic boundary conditions and stochastic boundary conditions for DNA fragments have recently been compared (Norberg & Nilsson, 1995b). The accuracy and stability of the simulations was similar, but the stochastic boundary simulations were found to be significantly faster than the MD simulations with periodic boundary conditions. MD simulations using only a thin layer of water molecules around the protein have been performed and shown promising results (Guenot & Kollman, 1992; Steinbach & Brooks, 1993; Sen & Nilsson, 1999b). A different method uses the rotational symmetry boundary condition, which is based on the

icosahedral symmetry and reduces the number of solvent molecules in the system (Çağin & Pettitt, 1991c). Another technique named the dynamic surface-boundary-condition method consists of adding a monolayer of special water molecules on the surface of the protein ( Juffer & Berendsen, 1993). Recently the glide-plane boundary-condition technique for interfacial systems has been developed (Wong & Pettitt, 2000). In comparison with periodic boundary conditions, the number of atoms was reduced by half and therefore required less CPU time. The results were very similar, however, the existing artifact in the glide-plane boundary-condition method requires further exploration (Wong & Pettitt, 2000). A simple approach, which used weakly harmonic restraining of a 5-Å shell of water, has been presented (Sankararamakrishnan *et al.* 2000). This procedure worked for MD simulations with temperatures in the 300–500 K range.

## 3. Force fields

The accuracy of the energy function directly affects the accuracy and stability of MD simulations of biological macromolecules. For reasons of computational efficiency when dealing with large macromolecular systems the energy function has to be simple, and since force calculations are needed in MD simulations, the function should also be analytically differentiable. Even though the simulations are Newtonian, the energy function is an approximation of the Born–Oppenheimer quantum mechanical energy surface of the system, and *ab initio* quantum chemical calculations on model compounds today are indispensable in the development of these empirical energy functions.

Several research groups have developed program packages for MD simulations of biomolecules, often in conjunction with an energy function. Most of these MD simulation programs apply the same form of the energy function (Levitt, 2001), with harmonic terms describing bonds and angles, Fourier series for torsions, and pairwise van der Waals and Coulombic interactions between atoms that are separated by three or more bonds. The major differences between the force fields are in the way the parameters of the energy function are derived. This, together with the interdependencies among the parameters, results in parameter sets where individual parameters cannot, and should not, be directly compared. Instead comparisons have to be made based on the ability of the energy functions to reproduce relevant data for systems of interest. The issue is further convoluted by different treatments of long-rang electrostatic interactions, but overall recent energy functions have similar performance.

With the availability of increasingly powerful computers macromolecular simulations are now regularly being performed in solution, and in the development of the newer energy functions more attention has been paid to the balance between solute–solute, solute–solvent and solvent–solvent interaction energies. Energy functions commonly used for the major classes of biomolecules are briefly characterized below in sections 3.1–3.4, and finally recent efforts towards inclusion of polarization effects are described in Section 3.5.

### 3.1 Proteins

Highly accurate force-field parameters have been presented for the AMBER energy function (Cornell *et al.* 1995). In a comparison between the AMBER95 and the modified AMBER C96 force fields (Ono *et al.* 2000), free-energy profiles were calculated for small peptides in explicit water and *in vacuo*. Extended conformations of the peptide were found with the C96 force field and more turn-like conformations with the AMBER95 force field. A modified version of the

AMBER force field, with refined carbon–carbon torsional parameters, which is suitable for cyclic $\beta$-amino-acid derivatives, has been described (Christianson *et al.* 2000). This parameter set should be helpful in predicting qualitatively the helical secondary structures of $\beta$-amino-acid oligomers.

Partial charges for the AMBER force field have been determined for biological and organic molecules using the restrained electrostatic potential (RESP) methodology (Wang *et al.* 2000). The study considered 55 molecules for which experimental data existed and an average absolute error of 0·28 kcal mol$^{-1}$ was found. This was somewhat better than what was obtained for the CHARMm/MSI, MM3, and MMFF force fields. Further the torsional parameters in peptides were optimized and this gave better energies than those found for the AMBER95 force field (Cornell *et al.* 1995). In a recent investigation RESP charges have been derived for molecular models that take into account 'lone pairs' on lone-pair donor sites and atom-centered polarizabilities (Cieplak *et al.* 2001). In this methodology the partial charges are determined self-consistently from both the partial charges and the induced dipoles to accurately reproduce the quantum mechanical electrostatic potential. A genetic algorithm and a systematic search method have been applied to optimize the force-field parameters (Wang & Kollman, 2001). In both methods the r.m.s. errror of 0·51 kcal mol$^{-1}$ was smaller than the r.m.s. error of 0·78 kcal mol$^{-1}$ found with the AMBER force field (Cornell *et al.* 1995).

A few years ago protein parameters (version 22) for the all-atom CHARMM force field were developed (MacKerell *et al.* 1998). The atomic charges were derived from *ab initio* calculations of interactions between water molecules and model compounds. A number of different peptides and proteins *in vacuo* and in crystals were used for testing the parameters.

The van der Waals parameters for the aliphatic $CH_n$ united atoms have been reparameterized for the GROMOS96 energy function (Daura *et al.* 1998b). In this derivation the parameters were adjusted to produce accurate room-temperature enthalpies of vaporization and vapor pressures or densities of a set of solvated alkanes. For this reparameterization the cut-off radius was increased from 8 to 16 Å. The investigation found free energies of hydration with a deviation of 0·2 kcal mol$^{-1}$ from the experimental values. The newer GROMOS96 force-field version has been compared to the old GROMOS87 version (Stocker & van Gunsteren, 2000). For the comparison two 2 ns MD simulations of hen egg-white lysozyme were used in conjunction with NMR and X-ray data. The GROMOS96 force field showed better agreement with experimental data than the GROMOS87 force field. A number of biochemical systems have been used for benchmarking the GROMOS96 program (Bonvin *et al.* 2000). The performance was measured for a number of different computers. For the non-bonded interaction pairs and long-range interactions a new pair-list technique was applied and found to yield a speed-up of 1·5–3 times more compared to standard procedures.

The all-atom version of the OPLS force field has been reported (OPLS-AA; Jorgensen *et al.* 1996). Most of the bond-stretching and angle-bending parameters were taken from the AMBER force field (Cornell *et al.* 1995), but the torsional and non-bonded energetic parameters were derived using *ab initio* molecular orbital calculations and Monte Carlo simulations (Jorgensen *et al.* 1996). The free energy of hydration showed an average error of less than 0·5 kcal mol$^{-1}$ for some model compounds. By using a classical approach force-field parameters were developed for amines (Rizzo & Jorgensen, 1999). The experimental features of pure liquids and gas-phase *ab initio*-calculated hydrogen-bond strengths were reproduced. Transferability of the parameters to other solvents was demonstrated. The study concluded that there was no need to include polarizability. In a recent study the OPLS-AA force field was improved by refitting the Fourier

torsional coefficients (Kaminski *et al.* 2001). The average energy r.m.s. deviation was significantly reduced and transferability of the parameters was demonstrated. Further a protocol for torsional fitting was devised. For the OPLS-AA force-field torsion parameters have been derived using quantum mechanical calculations (Damm & van Gunsteren, 2000). Three different parameter sets were determined for both high- and low-energy conformations. These parameter sets were then applied in the MD simulations of a $\beta$-heptapeptide and compared to previous data, which were obtained by using the GROMOS96 force field. The results showed a melting temperature of approximately 360 K for the OPLS-AA force field. This was an increase of the melting temperature in comparison with the GROMOS96 force field, for which 340 K was found.

Energetic parameters for the ENCAD program have been reported (Levitt *et al.* 1995). Nanosecond trajectories for proteins in solution have been found to be stable with the ENCAD force-field parameters. An extensive comparison of a number of force fields to *ab initio* and experimental data has been conducted (Halgren, 1999). The MMFF94 and MMFF94s force fields demonstrated the best results. The ECEPP force-field parameters have been updated (Némethy *et al.* 1992). An improvement of the energy minimum for peptides containing Pro-Pro or Ala-Pro sequences was obtained.

A comparison between 20 different force fields using high-level *ab initio* quantum chemical calculations has been conducted (Beachy *et al.* 1997). This testing used an alanine tetrapeptide as a model compound. The study found discrepancies in all of the force fields, and it is not clear how relevant the alanine tetrapeptide is as a model system for proteins (MacKerell *et al.* 1998), Further it was suggested that for obtaining accurate electrostatics the force fields needed to incorporate off-atom partial charges. Force-field parameters, which included fluctuating dipoles, have been developed for alanine, serine and phenylalanine (Stern *et al.* 1999). The intramolecular parameters, torsional parameters, and Lennard–Jones parameters were adopted from the OPLS force field (Jorgensen *et al.* 1996). These parameters accurately reproduced energetics from quantum mechanical calculations (Stern *et al.* 1999). For an alanine peptide a polarizable force field named OPLS-FQ has been described (Banks *et al.* 1999). This force field used the non-electrostatic parameters from the all-atom OPLS-AA force field (Jorgensen *et al.* 1996) and the polarizability was accounted for through the fluctuating charge model (Rick *et al.* 1994). The accuracy was on the same level as has been found for other force fields.

A recent report has described a new force field, which includes fluctuating atomic charges (Möllhoff & Sternberg, 2001). In this force field the charges were determined using the bond polarization theory. The study presented a test, which showed r.m.s. deviation of interaction energies, hydrogen-bond distances, and hydrogen-bond angles comparable to other force fields.

Effective charges have been determined for different conformational states of an alanine dipeptide (Ridard & Lévy, 1999). These calculations were done with a specific fitting algorithm to fit the *ab initio* electrostatic potential from the SCF wave function as previously described (Lévy & Enescu, 1998).

Atomic charges have been determined by using a least-squares fit to the quantum mechanical electrostatic potential with a restraint (REPD; Henchman & Essex, 1999a). This method has an advantage over the RESP approach (Bayly *et al.* 1993) because the restraint was optimized to produce charges close to the OPLS charges (Jorgensen *et al.* 1996). The accuracy of the REPD method was tested by calculating the free energies of hydration of 22 organic molecules (Henchman & Essex, 1999b). An absolute error of 0·7 kcal $mol^{-1}$ was obtained.

*Ab initio*, density functional theory, SIBFA molecular mechanics (MM) calculations have been performed for formamide and *cis-N*-methylacetamide (*cis*-NMA) (Gresh *et al.* 1999). The SIBFA formalism showed larger dimerization energies of formamide and *cis*-NMA than of alanine and glycine dipeptides. From the analysis of SIBFA calculations the monopole–dipole and monopole–quadrupole energy terms were found to be very important for the better formamide dimerization compared to the glycyl dipeptide dimerization.

### 3.2 Nucleic acids

In the second-generation AMBER force field, nucleic-acid parameters were consistently developed (Cornell *et al.* 1995). The charges were obtained using a 6-31G* basis set and RESP fitting. In contrast the van der Waals parameters were derived from liquid MD simulations.

Nucleic-acid parameters for the CHARMM version 27 force field have been derived (Foloppe & MacKerell, 2000). The parameter refinement protocol was based on small molecule data from experiments and *ab initio* calculations and nucleic-acid features in the condensed phase. These parameters were determined to be compatible with the protein and lipid parameters in CHARMM. A number of DNA and RNA structures have been simulated to verify the quality of the CHARMM version 27 force-field parameters (MacKerell & Banavali, 2000). The study analyzed dynamic, hydration and structural features in these simulations. In aqueous solution B-DNA was observed to be stable in the B form, contrary to what has been found for the CHARMM version 22 parameters (MacKerell *et al.* 1995). A stabilized A-DNA form was obtained in 75% ethanol. In comparison with other force fields the CHARMM27 showed less B form than the BMS force field (Langley, 1998), but more B form than the AMBER force field (Cornell *et al.* 1995).

The intrinsic energetics of the torsion angles in nucleic acids has been calculated for both the AMBER and CHARMM force fields (Bosch *et al.* 2001). Several improvements were included in the CHARMM version 27 parameters compared to version 22 parameters. The nucleic-acid structure with the lowest energy was well balanced in both AMBER and CHARMM. This study concluded that both the force fields could still be further refined, especially considering conformational transitions.

A large number of MD simulations have been performed to attest the performance of the BMS force field for nucleic acids (Langley, 1998). In this force field the backbone dihedral and phosphate parameters were derived and the other parameters were adopted from various force fields. From the MD simulations both A-DNA/B-DNA and B-DNA/A-DNA transitions were observed. In 75% ethanol solution the B-DNA/A-DNA transition was seen and also the reverse. These 41 ns MD simulations using the BMS force field demonstrated both correct sequence-dependent and environmental effects of DNA structures.

Force-field parameters for phosphorothioate DNA have been determined (Bertrand *et al.* 1999). These parameters took into account the sulfur effect on structure and were incorporated into the FLEX force field in JUMNA (Lavery *et al.* 1995). Good agreement with experimental results was obtained. Stable nanosecond MD simulations of DNA oligonucleotides have also been obtained using the ENCAD force-field parameters (Levitt *et al.* 1995). The W99 force field for organic molecules has been extended to include nitrogen atoms (Williams, 2001). This parameter set was tested on nucleoside and peptide molecular crystals and showed good agreement with determined crystal structures.

### 3.3 Carbohydrates

Maltoheptaose was analyzed using small-angle X-ray scattering profiles, Monte Carlo calculations and MD simulations with different force fields (Shimada *et al.* 2000). The Cff91 force field yielded the best agreement with experimental data. Further it was observed that Glycam99 gave better results than the older version Glycam93.

The carbohydrate parameters in GROMOS have been improved using MD and molecular mechanics (Spieser *et al.* 1999). In this study the force constant of the glycopyranose rings was increased to stabilize the ring inversion and the bond-angle potential was changed to account for the exo-anomeric effect.

A MD study of the oligosaccharide $\alpha$-D-maltose demonstrated the importance of including explicit water in the simulation (Ott & Meyer, 1996). The study observed only short-lived hydrogen bonds. Further the study incorporated the exo-anomeric effect by parameterization of the GROMOS force field.

Seven test cases have been used to evaluate 20 different force fields (Pérez *et al.* 1998). Principal-component analysis was applied to give a global view of carbohydrate parameters. The relationship between the force fields was determined and this will be useful in selecting parameters.

An atomic charge set for carbohydrate molecules has been derived (Woods & Chappelle, 2000) using the RESP methodology (Bayly *et al.* 1993). In the investigation the potential energy was observed to be inversely proportional to the restraint weight used in the RESP procedure (Woods & Chappelle, 2000). The charges were implemented into the GLYCAM force field (Woods *et al.* 1995). Further the study suggested a refinement of the van der Waals parameters or an incorporation of the polarizable charge model (Woods & Chappelle, 2000). Partial charges for carbohydrates were determined by averaging charges, which have been calculated for different structural conformations taken from a MD simulation (Basma *et al.* 2001). Interestingly the investigation provided standard deviations for the partial charges. The hydroxyl group conformations significantly affected the internal energies.

The CHEAT force field was developed for carbohydrates based on the CHARMM force field (Grootenhuis & Haasnoot, 1993). In this force field the carbohydrate hydroxyl groups are represented with united atoms. To obtain better agreement with NMR data a reparameterization has been carried out resulting in the CHEAT95 force field (Kouwijzer & Grootenhuis, 1995). This force field allowed for carbohydrate simulations without explicit water molecules. The study further pointed out that the force field should not be used for protein–carbohydrate interactions.

### 3.4 Phospholipids

The united atom AMBER force-field parameters for phospholipid membranes have been refined (Smondyrev & Berkowitz, 1999). In a test case a dipalmitoyl phosphatidylcholine bilayer was simulated for 1 ns at 323 K. The area per head group and order parameters for the Sn-2 hydrocarbon tail were in good agreement with experimental data. In contrast the conformations of the head groups in the simulations disagreed with what has been seen in the crystal structure.

For the CHARMM program the phospholipid parameters have been described (Schlenkrich *et al.* 1996). Thereafter parameters for the unsaturated hydrocarbons were added (Feller *et al.* 1997). In a recently updated version of the phospholipid parameters (Feller & MacKerell, 2000) major changes were made for the Lennard–Jones hydrocarbon parameters and the torsional

parameters, and the phosphate moiety was reparameterized to achieve more accurate atomic charges and torsional parameters.

In the GROMOS force field the parameters for the dipalmitoyl phosphatidylcholine bilayer have been published (Egberts *et al.* 1994). More recently these parameters and other phospholipid parameters have been thoroughly discussed (Berendsen & Tieleman, 1998). Recently the lipid force-field parameters of GROMOS have been reparameterized (Schuler *et al.* 2001). For alkanes the free energy of hydration was in very good agreement with experiments. The parameters for *n*-alkanes, cyclo-, iso-, and neoalkanes were significantly improved. This will enable MD simulations of membranes and micelles and other mixed apolar biochemical systems.

Another set of force-field parameters for lipids has been developed (Stouch *et al.* 1991). These parameters and the CHARMM version 22 parameters have been assessed in MD simulations of glyceryl phosphorylcholine crystals (Tu *et al.* 1995). The study found that these parameters outperformed the CHARMM parameters in comparison with experimental values. In an investigation of a bilayer of 128 fully hydrated phospholipids different boundary conditions and force-field parameters were compared (Tieleman & Berendsen, 1996). The techniques and force fields for MD simulations of lipid membranes have been reviewed (Tobias *et al.* 1997; Berendsen & Tieleman, 1998; Tobias, 2001).

### 3.5 Polarization

To include the effects of polarization a method, in which the atomic-centered induced dipole was described as induced charges on the atom and its neighboring atoms, has been reported (Winn *et al.* 1999). This study is an extension of the previously published multipole analysis method for deriving charges (Winn *et al.* 1997). The method takes advantage of using induced charges instead of induced dipoles (Winn *et al.* 1999). Further it was found that a correction term was needed for the electrostatics.

Recent improvements in polarizable force fields have been reviewed (Halgren & Damm, 2001 and references therein). The initial attempts several years ago to develop a polarizable force field focused on the water molecule. An electrostatic model, which includes atomic charges, atomic dipoles, and polarization, has been developed (Mannfors *et al.* 2000). Charges and dipoles have been determined by fitting to *ab initio* electric potentials of isolated molecules and the polarizability parameters were obtained from potentials by applying different electric fields to the molecules. The study found an increased accuracy by including polarization compared to condensed phase models.

## 4. Electrostatics

Long-range electrostatic interactions determine to a large degree the conformation of a biological macromolecule and are also very important for the initial encounter in many association processes. A number of reviews have focused on the importance of long-range electrostatic interactions in biochemical systems (Harvey, 1989; Berendsen, 1993). A recent review concerning electrostatic interactions has focused on the methodology and applications to membranes (Tobias, 2001).

The long-range nature of electrostatic interactions makes them computationally the most expensive part of a MD simulation; for a finite system with $N$ charges, $\sim N^2$ interactions have to be computed, which can be prohibitively expensive for large systems. The traditional way of

handling this problem is to truncate the range, and only include interactions between charges separated by a distance shorter than some cut-off, typically around 10 Å. The Ewald summation technique (Ewald, 1921), which in principle is the correct way to treat electrostatic interactions in infinite periodic systems, has recently gained popularity in the biomolecular simulation community due to the introduction of computationally efficient variants (Toukmaji & Board, 1996; Sagui & Darden, 1999), which reduce the computational complexity to $N \log N$. Several studies, (see below) have looked for artifacts due to the strict periodicity imposed by these methods. For large non-periodic systems multipole expansion and multi-grid methods, which are $O(N)$, offer substantial speed-ups, and there are also combinations of multipole and Ewald summation methods.

Different truncation methods for long-range electrostatic interactions have been compared in MD simulations of both proteins and highly charged nucleic acids (Norberg & Nilsson, 2000 and references therein), and it was shown that spherical cut-off methods, with an appropriate tapering function such as the atom-based force-shift truncation method (Brooks *et al.* 1985b; Steinbach & Brooks, 1994), provide accurate and stable nanosecond trajectories that behave essentially in the same way as those obtained with Ewald methods.

### 4.1 Spherical truncation methods

The first developed truncation methods are the spherical truncation methods, which are used together with a Verlet neighbor list that is updated regularly. These methods neglect all Coulomb or van der Waals interactions where the distance between the two atoms is greater than a certain radius, the cut-off distance. There are a number of different ways to implement the spherical truncation methods (Brooks *et al.* 1983, 1985b; Steinbach & Brooks, 1994). A simple way is to truncate the energy or the force abruptly at the cut-off distance, but usually some kind of smoothing function $S(r)$ is applied:

$$E(r_{ij}) = E^{\mathrm{Coul}}(r_{ij}) \cdot (S(r_{ij};\, r_{\mathrm{on}};\, r_{\mathrm{off}})).$$

Spherical truncation comes in a variety of flavors, but the methods can be broadly classified according to whether the neighbor list is atom–atom, or group–group based, if the smoothing function modifies the energy over the whole range $r < r_{\mathrm{off}}$ (shifting) or over a narrow range $r_{\mathrm{on}} < r < r_{\mathrm{off}}$ (switching), and if $S$ makes the energy and/or the force go to zero at $r_{\mathrm{off}}$. With group-based truncation it is known that energy is not conserved (Steinbach & Brooks, 1994), which is easily seen for the simple case of two dipoles with an interaction energy $E_{\mathrm{d}}$ when they are just within the cut-off distance. If one of the dipoles now moves just outside the cut-off it can rotate 180° and bounce back as a result of bumping into something, essentially at no energetic cost. When it re-appears within the cut-off distance of the first dipole, their relative orientation has changed and the interaction energy is $-E_{\mathrm{d}}$, without a corresponding change elsewhere in the system. If the atom groups are not neutral the effects can be severe, and we believe that reported cases of unstable simulations of nucleic acids using spherical truncation are probably caused by the use of group-based truncation schemes. In a recent study of a number of different truncation methods a detailed description of the methods is given and the application of the methods to proteins and nucleic acids is discussed (Norberg & Nilsson, 2000 and references therein). The study demonstrated that the atom-based force-shifting function and the particle–mesh Ewald (PME) technique generated stable and very similar nanosecond trajectories for double-stranded DNA. This issue has also been discussed in a recent review (Cheatham & Kollman, 2000).

The twin-range cut-off method (van Gunsteren & Berendsen, 1990) has been examined for an aqueous sodium chloride solution (Auffinger & Beveridge, 1995). This spherical cut-off method determines the short-range interactions at every step of the MD simulation, but the long-range interactions are only calculated when the non-bonded list is also calculated. Perturbations were obtained at the cut-off limit, but these were probably due to the use of a charge-group-based truncation and not a result of the twin range itself.

In MD simulations of a sodium ion in water boxes with different edges another approach to truncate the electrostatic interactions was applied (Ledauphin & Vergoten, 2000). For the solute–solute and solute–solvent interactions three cut-offs along the $x$-, $y$- and $z$-axes were used, whereas the solvent–solvent interactions were treated with a classical spherical cut-off. Good agreement with experimental results was found for the ion–oxygen radial distribution functions.

## 4.2 Ewald summation methods

The infinite sum needed to compute the Coulomb energy of a charged particle in an inifinite periodic system converges slowly, but this can be speeded up if a cancelling charge distribution is added and the sum is rewritten as the following Ewald sum:

$$4\pi\varepsilon_0 E_{\text{ele}} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} q_i q_j \sum_{l=0}^{\infty}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \frac{\operatorname{erfc}(\kappa|r_{ij}+a|)}{|r_{ij}+a|} - \frac{\kappa}{2\sqrt{\pi}}\sum_{i=1}^{N}\sum_{j=1}^{N} q_i q_j \delta_{ij}$$
$$+ \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \frac{q_i q_j}{\pi L_x L_y L_z} \sum_{l=0}^{\infty}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \frac{\exp\left(-\frac{|k|^2}{4\kappa^2}\right)\cos(k\cdot r_{ij})}{|k|^2} + J(\mathbf{D},\varepsilon'),$$

here erfc is the complementary error function, the first term is the real-space summation, the second term is a correction for the self-energy of the cancelling charge, the third term is the reciprocal space summation, which is truncated at an ellipsoidal boundary, and the fourth term is a surface correction term, which involves the unit cell dipole moment, **D**, and the dielectric constant, $\varepsilon'$, of the surrounding medium. $k = 2\pi(l/L_x, m/L_y, n/L_z)$ is the reciprocal space lattice vector. $\kappa$ determines the width of the cancelling charge distribution, here a gaussian ($\rho_i(r) = q_i\kappa^3 \exp\{-\kappa^2 r^2\}/\pi^{\frac{3}{2}}$) centered at each real charge but other functional forms are also possible. With a wide charge distribution (small $\kappa$) the exponential in the reciprocal space sum decays rapidly to zero, and only a few terms have to be included. A narrow distribution, on the other hand, makes the erfc in the real-space sum fall to zero very quickly, hence an optimum value of $\kappa$ has to be choosen, such that total time is minimized for a given required accuracy. For the surface correction term, $J(\mathbf{D}, \varepsilon')$, it has been suggested that vacuum boundary conditions are preferable to tinfoil boundary conditions in simulations of biomolecules (Vorobjev & Hermans, 1999).

Several different methods based on the Ewald summation have been developed. An efficient Ewald method for large condensed phase systems (Smith & Pettitt, 1995), in which the real space was partitioned into two parts, one long-range and one short range, was tested on an ionic system and found to be slower than the traditional spherical truncation method, but with less artifacts. Possible artifacts of the Ewald summation method in liquid solutions and biomolecules have been investigated (Smith & Pettitt, 1996; Smith *et al.* 1997; Hünenberger & McCammon, 1999;

Weber *et al.* 2000), with the general conclusion that the artifacts are small, in particular for systems with large unit cells and solvents with a high dielectric permittivity.

The reciprocal sum of the Ewald sum can be turned into an $O(N \cdot \log N)$ method (PME) by using fast Fourier transform of the charge distribution distributed onto a grid ('mesh') (Darden *et al.* 1993; Sagui & Darden, 1999). This method has been further extended using B-spline interpolation of the reciprocal space structure factor (Essmann *et al.* 1995). The importance of the real-space term in the Ewald summation approach has been extensively examined (Wolf *et al.* 1999). Evaluation of the real-space term can also be made more efficient using a tree code, with multipole expansions to approximate the interactions between particles, which reduces the complexity from $O(N^2)$ to $O(N \log N)$ (Duan & Krasny, 2000). Several recent reviews discuss applications of the PME method in MD simulations of proteins and nucleic acids (Cheatham & Brooks, 1998; Darden *et al.* 1999; Cheatham & Kollman, 2000; Norberg & Nilsson, 2002). In a comparison of the PME method with other commonly used truncation methods (Norberg & Nilsson, 2000) it was observed that the PME method produced accurate and stable trajectory of a DNA oligonucleotide, but very similar results were also obtained with appropriate spherical truncation of the electrostatic interactions.

A method has been proposed for systems that are periodic in two dimensions and have a finite length in the third dimension, and was also shown to yield a significant reduction in computational time (Yeh & Berkowitz, 1999). In a comparison of the standard Ewald summation method and the PME method with 2D periodicity a cubic box of water was simulated with both methods and the accuracy was found to be on the same level for both methods, but for 2D periodicity the PME method was approximately 5 times faster than the Ewald summation method (Kawata & Nagashima, 2001).

The particle–particle particle–mesh ($P^3M$) method uses a charge-shaping function to split the potential into a real part and a reciprocal space one (Hockney & Eastwood, 1981). In this method particle–particle summation is used for the short-range non-bonded interactions and the long-range non-bonded interactions are determined from the Poisson equation with periodic boundary conditions (Sagui & Darden, 1999). It has been shown that using polynomials truncated to a finite spatial range, instead of a Gaussian, as charge-shaping function leads to an increase in computational speed, user friendliness and improved accuracy (Hünenberger, 2000).

The relative r.m.s. force errors for the reciprocal sums in the $P^3M$, PME and smooth PME methods are very similar (Darden *et al.* 1997), but less accurate than the standard Ewald summation (Deserno & Holm, 1998a). The latter study recommended the $P^3M$ method, which uses an analytically derived optimal-influence function to minimize the force errors, and in a following paper the accuracy of the $P^3M$ method was calculated (Deserno & Holm, 1998b).

The reversible reference system propagator algorithm (r-RESPA) (Tuckerman *et al.* 1992) with multiple time-steps can be used in combination with the PME (Kawata & Mikami, 1999, 2000) or $P^3M$ Ewald method (Zhou *et al.* 2001a) to further accelerate the computations. Not only the Coulombic interactions but also the van der Waals interactions may be evaluated using the PME method (Kawata & Mikami, 1999, 2000). A symmetric Trotter factorization of the Liouville operator together with the PME method were used to construct a multiple time-step force-splitting formalism for biological molecules (Batcho *et al.* 2001). The study presented a position-Verlet scheme for dynamic integration and optimized parameters for the Ewald method. A discussion about the speed-up of these techniques in relation to other commonly applied procedures was reported.

### 4.3 Fast multipole (FM) methods

The FM method (Greengard & Rokhlin, 1987) is based on conventional multipole expansions for the electrostatic potential, and tree algorithms, in which distant particles recursively grouped into multipoles are used to calculate the energies and forces. A recent review covers to some extent the basics of the FM method (Sagui & Darden, 1999).

For MD simulations of large proteins a method, which combines the FM technique and the r-RESPA, has been developed (Figueirido *et al.* 1997). The study found that the energy was conserved for time-steps up to 12 fs for molecules of 40 000 atoms. A considerable speed-up was seen over the Verlet integration combined with the Ewald summation. In comparison with the PME method the presented technique showed similar efficiency up to 20 000 atoms, while for even larger molecules the FM method with r-RESPA was more efficient. A multiple time-step method in conjunction with the Nosé–Hoover chain method for controlling the temperature and pressure has been implemented (Cheng *et al.* 1999). By using these methods a time-step of 4 fs could be allowed to obtain stable MD simulations. The study concluded that these methods are clearly better than the SHAKE method concerning speed-up and temperature control.

With the computational complexity of the Ewald ($N^2$), PME ($N \log N$), $P^3M$ ($N \log N$) and FM ($N$) methods, it is clear that FM will be the fastest for very large systems. Benchmarks of the speed of the methods are very dependent on details of code implementation/optimization, and thus the break-even points are difficult to define accurately (Pollock & Glosli, 1996; Toukmaji & Board, 1996).

### 4.4 Reaction-field methods

Instead of trying to include explicitly all interactions in an infinite system, as in the Ewald or FM methods, or truncating the sum at some cut-off distance, one may include effects of the surroundings using a reaction field, which takes care of the response of the dielectric medium beyond the cut-off (Allen & Tildesley, 1987; Tironi *et al.* 1995; Im *et al.* 2001; Walser *et al.* 2001; Banavali *et al.* 2002). Recent implementations of generalized reaction-field methods allow not only dipolar but also ionic solutions to be treated. The reaction-field method is an approximation, but it seems to perform quite well (Tironi *et al.* 1995; Boresch & Steinhauser, 1999; Rozanska & Chipot, 2000; Walser *et al.* 2001).

## 5. Implicit solvation models

The aqueous environment found *in vivo* is essential for most biomolecular functions, even though many proteins retain their structure under very dry conditions or *in vacuo*. Early protein simulations benefited from this by being performed without any solvent, thus saving a substantial amount of computing time since in a simulation of a protein in water, 90% or more of the time may be spent on computing water–water interactions. For highly charged systems, such as nucleic acids, vacuum simulations were less successful, and today most simulations incorporate solvent effects, either with an explicit solvent representation or implicitly. In addition to providing dielectric screening, an explicit solvent contributes direct and specific interactions that may be important, particularly close to the solute. The importance of explicit solvents in MD simulations for the thermodynamic decomposition of free energies and for the handling of the electrostatic effects has been reviewed (Levy & Gallicchio, 1998). On the other hand, implicit

solvation models allow more direct estimation of free energies of solvation than can be obtained with explicit solvation (Lazaridis & Karplus, 2003). The statistical mechanics behind the implicit solvation models has been thoroughly described (Roux & Simonson, 1999). This review further outlined the non-polar and electrostatic contributions of the solvation free energy. In another review MD simulations, which include the generalized Born model, have been compared to explicit solvent MD simulations (Simonson, 2001). Implicit models using the generalized Born model have been compared to explict solvation (Simonson, 2001), solvent effects on biological molecules and different methods (MD, QM, continuum models) for treating solvent features have been reviewed (Cramer & Truhlar, 1999; Orozco & Luque, 2000).

The generalized Born model has been investigated regarding the performance and vectorization as described in the AMBER program (Sosa *et al.* 2001). MD simulations, which used the implicit Born model, were significantly faster than fully hydrated protein MD simulations.

A recent approach used a surface of active polarons to model the main part of the bulk solvent (Kimura *et al.* 2000). This method applied a thin layer of explicit water molecules with external forces and polarons, which are variable charges that correct for the long-range electrostatics, to solvate a biomolecule. In several test cases good agreement with experiments were obtained.

In a hybrid MD approach the solvent–solvent interactions were handled by a coarse-graining model (Malevanets & Kapral, 2000). Here all the other interactions were treated with full MD. The solvent interactions followed the right hydrodynamic equations and a test showed that this hybrid scheme captured both hydrodynamic and microscopic effects.

The solvent effects are included in MD simulations through the solvent-accessible surface area in the generalized Langevin dynamics method (Oliva *et al.* 2000). This method was applied to the carboxypeptidase A inhibitor protein and found to be promising for simulating large biomolecular systems.

In the effective energy function EEF1 the CHARMM united-atom energy function was combined with a Gaussian solvent-exclusion model for the free energy of solvation (Lazaridis & Karplus, 1999). This solvation model was developed from small model compounds. Further the ionic side-chains were neutralized and for the charge–charge interactions a distant-dependent dielectric constant was applied. The report also demonstrated the applicability of the energy function in MD simulations. In a recent review a number of newly developed effective energy functions have been discussed (Lazaridis & Karplus, 2000).

A recent investigation has demonstrated stable and cost-effective MD simulations of biomolecules using a continuum solvent approach (Marchi *et al.* 2001). The method is based on a polarization-density free-energy functional with a minimum at the electrostatic equilibrium. Further, applying a pseudopotential technique expanded the solvent polarization and after a discretization of the functional the equations of motion were generated.

Another way to simulate a specifically hydrated region of a large protein while the distant solvent molecules are represented with a continuum dielectric has been reported (Im *et al.* 2001). In this methodology the solvent-shielded static field from the biomolecule was determined once from the finite-difference Poisson–Boltzmann method and the solvent reaction field was generated from a basis-set expansion, in which the coefficients correspond to the electrostatic multipoles. This method was tested on the active site area of the aspartyl–tRNA synthetase protein and compared to the finite-difference Poisson–Boltzmann method.

A continuum solvation model for the Merck molecular force field has been parameterized (Cheng *et al.* 2000). This generalized Born/surface area (GB/SA) solvation model showed free energies of solvation with good accuracy.

A fast method named the buried atom elimination (BAE) technique, for finding inaccessible atoms, has been developed (Weiser *et al.* 1999). This method speeds up the surface area computation and has been incorporated into the MacroModel program. From the testing it was found that the method increased the efficiency in the energy minimizations and MD simulations while keeping the same accuracy in computation.

In a recent investigation a smooth solvation potential was developed (York & Karplus, 1999). The basics for this potential are the conductor-like screening model (Klamt & Schüürmann, 1993). A simple solvent-accessible surface with a surface discretization based on angular quadrature rules was used (York & Karplus, 1999). The model was based on the solution of a number of linear equations, which produce a smooth potential and have no singularities.

In a protein-folding study MD simulations with explicit solvent and generalized Born model were compared (Bursulaya & Brooks, 2000). The thermodynamics of folding was determined for the $\beta$-sheet protein Betanova. Both models gave similar results for the location of the global minimum, the absence of folding barrier, and the unimportance of hydrogen bonds in folding. Discrepancies were observed regarding the protein stability and the folding transition temperature. The investigation stated that the generalized Born model was good enough for analyzing the folding of small polypeptides.

The binding of an octapeptide to the murine MHC class I protein has been compared using both explicit and implicit solvent models (Zhang *et al.* 2001). The surface-generalized Born implicit solvent model was significantly faster than the free-energy perturbation simulations in explicit water, but it was concluded that implicit models are not yet accurate enough to treat the conformational freedom of the protein.

## 6. Speeding-up the simulation

For a given computational effort the total time of a simulation is determined by the number of interactions that have to be evaluated at each time-step, the number of degrees of freedom that have to be propagated, and the step size. To enhance the speed in MD simulations one can try to reduce the number of interactions, or the number of degrees of freedom; the efficiency of interaction evaluations can be improved. If the basic time-step can be increased there is a potential for speeding up the simulation, and a number of algorithms concerning the integration time-step and the factors that limit the step size have been reviewed (Schlick *et al.* 1997). The number of interactions may be reduced by using implicit solvation as discussed in Section 5, or by using a force field with only heavy atoms represented (Section 3); other coarse-graining methods, using rigid bodies have also been suggested. Constraints are commonly used both to enforce a rigid geometry of molecules or fragments of molecules, and to allow a longer time-step.

### 6.1 SHAKE and its relatives

The fundamental step size in the numeric integration is determined by the period of the fastest motions in the system, and the SHAKE algorithm (Ryckaert *et al.* 1977) for holonomic constraints [of the form $f(\mathbf{R}, t) = 0$, where $\mathbf{R}$ denotes the coordinates all particles in the system] was one of the first devices to address this problem by removing the fastest degrees of freedom from the system, The SHAKE algorithm is an easily implemented and efficient method which has been widely used in MD simulations for some time. It can be applied to bonds or bond angles,

but due to the coupling between dihedral and bond-angle motions it is usually only applied to bonds, and often only to X–H bonds, which exhibit the fastest vibrations. This still allows 2 fs time-steps with the Verlet or leap-frog integrators, compared to 0·5–1 fs without SHAKE. To remove the fast vibrations of hydrogens in MD simulations a holonomic constraint variant was constructed (Ryckaert *et al.* 2001).

SHAKE may run into convergence problems for large coordinate displacements, and it does not parallelize well because of its iterative nature. The constraint method LINCS (Hess *et al.* 1997), which resets the constraints rather than their derivatives, was designed to remedy these issues, and it achieves a speed-up of 3–4 times compared to the original SHAKE algorithm (Ryckaert *et al.* 1977). In SHAKE the linearized constraint equations are treated as independent and solved iteratively. Several partly analytical variants have been introduced, often with an exact matrix inversion instead of the approximation with independent equations (Mertz *et al.* 1991; Kräutler *et al.* 2001). These algorithms are still iterative, due to the linearization of the constraint equations, but fewer iterations are required. The speed-up depends on the trade-off between fewer iterations and the cost of inverting the matrix, which becomes expensive for large systems of connected constraints. For systems with a small number of inter-dependent constraints, such as water, or when high accuracy is desired these methods are advantageous.

RATTLE (Andersen, 1983), and the analytical variant SETTLE (Miyamoto & Kollman, 1992) also allow velocities to be constrained, and they can thus be implemented in integrators that use velocities.

A generalization of SHAKE (GSHAKE) that allows for general non-holonomic constraints along with holonomic constraints, using a similar iterative procedure as SHAKE has recently been developed (Kutteh, 1999) and shown to be efficient, with no numerical drift for large numbers of constraints.

In the QSHAKE method, which is also an iterative constraint method, holonomic SHAKE-like constraints are combined with quaternion dynamics for rigid fragments (Forester & Smith, 1998). The study found that the QSHAKE algorithm used fewer holonomic constraints, obtained convergence by fewer iterations, and was more stable for large time-steps than the SHAKE method.

### 6.2 Multiple time-step algorithms

In multiple time-step methods fast and slow motions are separated, often based on distance classes, assuming that interactions involving larger distances vary more slowly, however, other schemes have also been devised. This results in a speed-up if the main computational burden is associated with the longer timescale (Tuckerman & Martyna, 2000; Tuckerman *et al.* 2000). Multiple time-steps can be combined with various Ewald techniques (see Section 5) and algorithms such as the Nosé–Hoover chain algorithm for pressure and temperature control (Marchi & Procacci, 1998; Cheng & Merz, 1999). Several multiple time-step algorithms have been investigated and stable MD trajectories have been reported with the long time-step as large as 8 fs (García-Archilla *et al.* 1998; Grubmüller & Tavan, 1998; Izaguirre *et al.* 1999; Leimkuhler & Reich, 2001).

The MBO(N)D (multi-body order (N) dynamics) method combines rigid-body dynamics with multiple time-steps. In this approach the high-frequency harmonic motions are removed while allowing the low-frequency anharmonic motion (Chun *et al.* 2000). This is managed by using different levels of detail to model different parts of the molecule. In tests the MBO(N)D method

has reproduced the essential dynamical global properties of proteins and nucleic acids (Chun *et al.* 2000).

### 6.3 Other algorithms

Digital signal processing techniques have been used to remove high-frequency motions in MD simulations (Sessions *et al.* 1989). In the digitally filtered MD approach (Phillips *et al.* 2000) the atomic velocities were modified such that the vibrational motion was controlled based on the frequency, allowing high selectivity and control.

The highest frequency motions are connected with the lightest atoms in the system, the hydrogens. Building hydrogens as dummy atoms or increasing the hydrogen mass could extend the time-step from 2 to 7 fs and 4 fs, respectively (Feenstra *et al.* 1999).

Two integration algorithms, which conserve the energy and are reversible in molecular simulations, have been developed (Zhou *et al.* 2000). The energy is evaluated at each iteration and, therefore, these methods are computationally more expensive, but longer time-steps can be utilized compared to SHAKE.

## 7. Conformational space sampling

Methods to speed-up the simulation also increase the sampling of available conformations, and since this is very important in many applications several techniques have been developed to further increase conformational sampling in MD or Monte Carlo simulations (Berne & Straub, 1997). There are two basic approaches, either to flatten the energy surface so that it becomes easier to pass barriers between conformations, or to use several copies of a region of interest, which increases sampling of this region at a small cost increase if the major computational effort lies in handling the remainder of the system. Flattening of the energy surface can either be done in a general way, to increase overall sampling, or it can be done along a defined reaction coordinate to facilitate a specific transformation.

### 7.1 Multiple-copy simultaneous search (MCSS) and locally enhanced sampling (LES)

The MCSS procedure was developed for finding energetically favorable positions of functional groups on a protein surface (Miranker & Karplus, 1991). A huge number of copies of a functional group were placed at a specific site and thereafter simultaneously minimized or simulated. In rational drug design the functionality maps from this method can be of importance. The method has also been extended to include target flexibility (Stultz & Karplus, 1999). Favorable positions of functional groups in a binding site can be found and even novel minima, which are not present in the fixed structure. To enhance the sampling the LES approximation was applied. In the study the viral protein HIV-1 protease with methanol and methyl ammonium was used for testing.

LES, a related method in which a fragment of the system is present in several copies has been used to calculate diffusion pathways for carbon monoxide through lupine leghemoglobin protein (Czerminski & Elber, 1991). The diffusion rates in leghemoglobin was faster than that in myoglobin due to the lack of helix D and to the more flexible CE loop in leghemoglobin as has been shown experimentally. LES has also been applied in studies of nucleic-acid conformations (Simmerling *et al.* 1998).

Sampling may be enhanced if the barriers between local minima are reduced, which can be achieved through conformational flooding, a method which destabilizes the initial state of the system by adding a potential centered at this initial state (Grubmüller, 1995). Potential smoothing can also be handled by the diffusion equation method. This method and its extensions have been thoroughly investigated (Pappu *et al.* 1998). Two methods for energy optimization of biomolecules have been analyzed (Hart *et al.* 2000). The quantitative and qualitative correlations between potential smoothing and simulated annealing were described.

The success rate of a ligand trying to escape from a binding pocket in a receptor is low in standard MD simulations if the pocket is not quite open. In this case the rate may be increased by the addition of a randomly directed force on the ligand, and changing the direction of the force if the ligand does not move a certain distance over a period of time (Lüdemann *et al.* 2000). The exit path(s) finally found may then be used in more quantitative analyses such as a PMF calculation.

An interesting development is the scheme for running combinations of MD simulations and Monte Carlo simulations (LaBerge & Tully, 2000). This mixed method was shown to converge to the correct equilibrium probability distribution, and it may enhance sampling in cases where part of the system, for example the solvent, is more efficiently handled by Monte Carlo, while the dynamics of the solute is treated by MD.

In replica-exchange dynamics several copies (replicas) of the system are simulated simultaneously, at different temperatures, and at certain intervals replicas are exchanged (i.e. their temperatures are switched), usually with a probability based on the energy and temperature difference between the two replicas (Sugita & Okamoto, 1999). The method has been successfully applied in protein folding and peptide conformation analysis (Zhou *et al.* 2001b; Feig *et al.* 2003).

### 7.2 Steered or targeted MD

In the steered MD approach external forces are used to maneuver the molecule, for example in binding and unbinding processes. This procedure makes it possible to speed-up biomolecular processes, which are usually too slow to handle, and to probe mechanical functions. Recently the steered MD technique and the atomic force experiments have been reviewed (Israelewitz *et al.* 2001). A couple of studies have built potentials of mean force from steered MD trajectories (Balsera *et al.* 1997; Evans & Ritchie, 1997; Gullingsrud *et al.* 1999).

In targeted MD the simulation is driven towards a specific conformation of the solute, usually by means of a force that depends on the conformational difference between the current structure and the target structure. A targeted MD method for determining the transition pathway between different configurations of a protein has been proposed (Schlitter *et al.* 1993), and applied to the T $\leftrightarrow$ R transition in insulin, as well as to the molecular switch in ras p21 (Ma & Karplus, 1997), the GroEL chaperone complex (Ma *et al.* 2000), and to protein folding (Ferrara *et al.* 2000).

### 7.3 Self-guided MD

To increase the efficiency of searching the conformational space a self-guided MD simulation technique has been developed (Wu & Wang, 1998). In this method a guiding force was incorporated into the equations of motion. The guiding force was determined as a time average of the force from the same MD simulation. An advantage of this self-guided technique over the conventional MD simulation was the increased search efficiency of the conformational space and

the possibility to overcome energy barriers. A recent extension of the self-guided MD simulation approach has been described (Shinoda & Mikami, 2001). The study demonstrated that the algorithm was stable at different pressures and temperatures. Lahiri *et al.* (2001) found that for a simple model system with a test particle in a double-well potential sampling rates could actually be reduced by the self-guiding procedure, unless the correlation between the physical and self-guiding forces was rather small.

An approach for efficient sampling of the conformational space of proteins is the leap-dynamics method (Kleinjung *et al.* 2000). Here a combination of the essential dynamics method (de Groot *et al.* 1997) and MD simulations were used. This scheme was applied to an alanine dipeptide and the conformations obtained were observed with all but one of the methods. Further the BPTI and the Y35G mutant were studied. The combined method correctly predicted the mobility of the mutant compared to the native structure as judged from NMR measurements.

## 7.4 Leaving the standard 3D Cartesian coordinate system: 4D MD and internal coordinate MD

A MD method with four spatial dimensions has been presented (van Schaik *et al.* 1993). The procedure with distance restraints from NMR data has been applied to structure refinement of cyclosporin A and of *lac*-repressor headpiece proteins. For these two cases the standard refinement protocol failed. The study found that the proteins could be refined and concluded that the method might be important for loop modeling. Excess chemical potentials may also be computed using PMF calculations along the fourth dimension, as was shown for the hydration free energy of camphor (Pomès *et al.* 1999).

An internal-coordinate MD approach devised for flexible biochemical systems has been developed (Lee *et al.* 2001). This formalism is based on the spectroscopic B-matrix in non-redundant generalized coordinates.

## 7.5 Temperature variations

Another way to sample the conformational space of a biological macromolecule using MD simulations is to use a number of different temperatures. This has turned out to be a very efficient and powerful tool for investigating various chemical problems. For instance high-temperature MD simulations have been used to study protein folding (see Section 10.1). This procedure has also been applied to nucleic acids for analyzing the glass transition (Norberg & Nilsson, 1996). This report found a glass transition around 223–234 K for a hexamer duplex (Fig. 2).

A computationally efficient procedure called the repeated-annealing sampling method has been proposed (Kamiya & Higo, 2001). The method was applied to tri-*N*-acetyl-D-glucosamine *in vacuo* and was found to give comparable results to conventional multi-canonical MD simulations.

To describe a rough energy landscape the random-energy approximation has been used and the density of states is usually assumed to be Gaussian distributed. Another distribution, in which the $N$ independent rotors build up the biomolecule, is based on the rotational isomeric approximation (Pérez *et al.* 1998). The sampling efficiency of the conformational space has been explored using a random search and the iterative simulated annealing approach (Corcho *et al.* 2000). For the random search method the density of states of the molecule displayed a bimodal

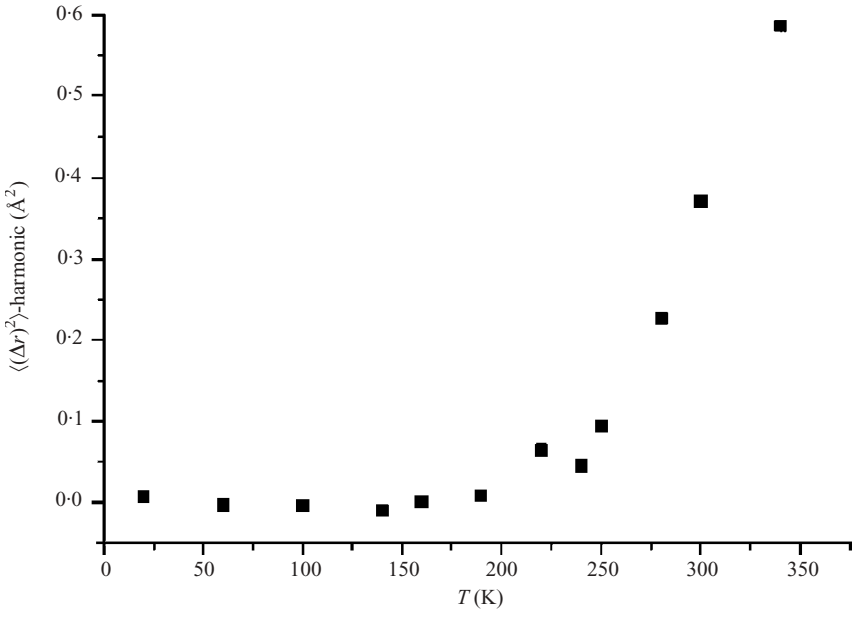


**Fig. 2.** Temperature dependence of the anharmonic contribution obtained by subtraction of the harmonic contribution from the mean-square fluctuations averaged over all atoms of the $d(CGCGCG)_2$ duplex (Norberg & Nilsson, 1996). [Figure reprinted with permission from the *Proceedings of the National Academy of Sciences USA* **93** (1996), 10173–10176.] (Copyright 1996 National Academy of Sciences, USA.)

distribution, whereas for the simulated annealing approach structures were found only in the lower peak of the distribution.

## 8. Thermodynamic calculations

The basic description for the calculation of free energies was formulated many years ago (Zwanzig, 1954), and applications in biochemistry have flourished for two decades, as described in the number of review articles focusing on free-energy calculations, with detailed descriptions of the methodology for a large number of approaches (Beveridge & DiCapua, 1989; Straatsma & McCammon, 1992; Kollman, 1993, 1996; Gilson *et al.* 1997; Lamb & Jorgensen, 1997; Kollman *et al.* 2000, 2001).

### 8.1 Lambda ($\lambda$) dynamics

The efficiency of a free-energy perturbation calculation usually depends on the pathway chosen for the transformation. By using a set of coupling parameters, $\lambda_i$, that are treated as dynamical variables, which are allowed to change during the simulation, it is possible to find alternative, optimal, pathways in an automated fashion. Free-energy calculations can thus be transformed into potentials of mean-force calculations in the $\lambda$ variable space, with umbrella sampling on the $\lambda$ variables (Kong & Brooks, 1996), a methodology that could be important for molecular design problems like relative binding-affinity calculations. With a set of $\lambda_i$ several perturbations can be performed simultaneously, and the largest $\lambda_i$ at the end of the simulation is the one connected to the optimal change, for example allowing a series of related ligands to be ranked according to

their affinity for a target (Banba & Brooks, 2000). The essential conformations of 5-membered-ring heterocycle derivatives bound to an artificial binding site were successfully sampled using the $\lambda$-dynamics technique in a study of the sampling efficiency of ligand orientations and conformations (Banba *et al.* 2000).

## 8.2 Extracting thermodynamic information from simulations

The analysis of free-energy pertubation or umbrella-sampling simulations often requires that data from several simulations are joined in a smooth fashion, and a formalism called the weighted histogram analysis method (WHAM) based on the multiple histogram technique has been developed (Ferrenberg & Swendsen, 1989; Kumar *et al.* 1992; Boczko & Brooks, 1993). This approach allows for multiple dimensions and produces smoother potential surfaces than conventional methods to splice together data sets. A generalized way to estimate thermodynamic properties of biochemical systems based on the maximum-likelihood method has been derived (Bartels, 2000), and shown to be equal to WHAM.

The methodology has also been applied to nucleic acids (Norberg & Nilsson, 2002). For instance the conformational free-energy landscape of a trinucleotide has been determined (Norberg & Nilsson, 1995a). This potential of a mean-force profile displayed a global minimum corresponding to the crystal structure (Fig. 3). The *cis/trans*-imide isomerization of a blocked peptide dimer has been investigated using the multi-canonical ensemble MD simulations with WHAM (Ono *et al.* 1999). In the study both the *cis*- and *trans*-isomer states were observed. To calculate an adiabatic potential energy map the automatic refinement procedure (AMRP) was implemented (Crouzy *et al.* 1999). This method was applied to the isomerization of two bonds in the retinal of bacteriorhodopsin and good experimental agreement was found. In a recent approach WHAM was used for free-energy perturbations together with the umbrella sampling (Souaille & Roux, 2001). One advantage was that a specific region of the conformational space could be chosen by window potential.

A way to estimate thermodynamic properties of biochemical systems based on the maximum-likelihood method has been derived (Bartels, 2000). This approach was found to be equal to WHAM for the potential of mean-force calculations and to the composite reference state method for free-energy difference calculations.

## 8.3 Non-Boltzmann thermodynamic integration (NBTI)

A recently developed method for free-energy calculations is the NBTI method (Ota & Brünger, 1997). This method has shown good results for free-energy calculations of the difference in solvation of butane and propanol. The binding free energies between two amine inhibitors and trypsin have been determined using this method (Ota *et al.* 1999). The NBTI simulation sampled the conformational space better and produced binding free energies closer to experimental values than the thermodynamic integration method.

## 8.4 Other methods

By using MD simulations together with soft-core potential scaling the relative hydration free energies of organic molecules were determined (Mordasini & McCammon, 2000). The method is based on an initially generated MD simulation using an unphysical reference state. Thereafter this

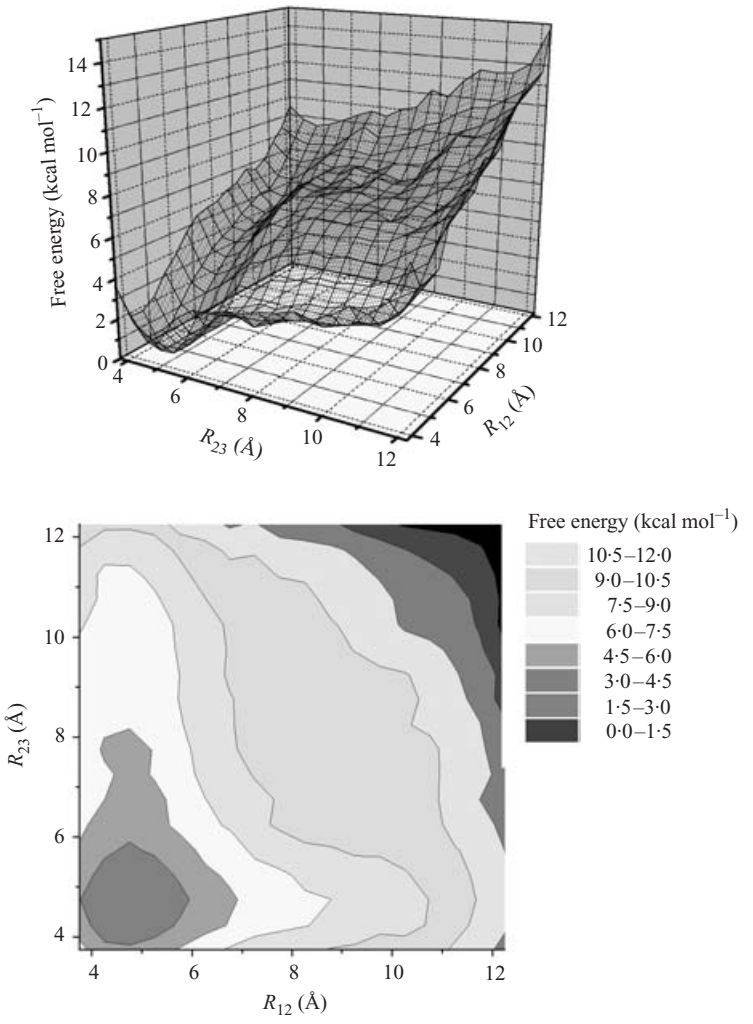


**Fig. 3.** Conformational free-energy landscape of the ApApA trimer using 2D PMF calculations (Norberg & Nilsson, 1995). [Reprinted with permission from the *Journal of Physical Chemistry* **100** (1995), 2550–2554]. (Copyright 1995 American Chemical Society.)

trajectory was used for calculating the free-energy differences between molecules. The study found that for small mutations the method was good, but for large mutations severe discrepancy was obtained.

In the linear interaction energy method only the end states of of the thermodynamic cycle are simulated, and the free energy of binding, $\Delta G_{\text{bind}}$, in a protein–ligand complex is obtained from (Åqvist *et al.* 2002)

$$\Delta G_{\text{bind}} = \alpha(\langle E_{\text{vdW}}\rangle_{\text{bound}} - \langle E_{\text{vdW}}\rangle_{\text{free}}) + \beta(\langle E_{\text{Coul}}\rangle_{\text{bound}} - \langle E_{\text{Coul}}\rangle_{\text{free}}) + \gamma,$$

where $E$ is the interaction energy between the ligand and its surroundings (solvent or protein in the free and bound simulations, respectively), and $\alpha$, $\beta$ and $\gamma$ are empirical parameters, which play an important role in obtaining accurate absolute-binding free energies. These parameters have been investigated for several protein-ligand interactions (Wang *et al.* 1999). A correlation

between the empirical $\beta$-parameter and the weighted non-polar desolvation ratio was observed. Another test study focused on 15 inhibitors of the enzyme neuraminidase (Wall *et al.* 1999). The linear interaction method was evaluated and a r.m.s. error of 1·51 kcal $mol^{-1}$ was found for the binding free energies.

In a recent investigation several approaches for ligand–protein binding free-energy calculations were evaluated (Sham *et al.* 2000). The study focused on the linear response approximation method, the linear interaction method, the protein–dipoles Langevin–dipoles (PDLD) method, and a semi-microscopic version of the PDLD method. Similar results were obtained for all the methods, but the semi-microscopic PDLD method was significantly faster than the other methods.

The hydration free energy was determined using the developed generalized linear response method (Chen & Tropsha, 1999). A test of the technique was carried out for a number of neutral organic compounds and good agreement with experiments were seen. One advantage of the method was the reduced computational cost.

To obtain accurate free energies the molecular mechanics and continuum calculations have been combined (Kollman *et al.* 2000). This approach is the most recent development in calculating free energies of biomolecules. The review covers applications concerning both proteins and nucleic acids. This approach was applied to the villin headpiece subdomain protein to differentiate between the native, random coil and intermediate states (Lee *et al.* 2000b).

An analytic gradient method for optimization of the transition state in solution chemical reactions has been presented (Okuyama-Yoshida *et al.* 1998). This method utilized the force and Hessian on the free-energy surface and can be determined from MD simulations. Further the study derived intrinsic reaction coordinate equations on the free-energy surface.

The force on the free-energy surface from MD techniques was used in the free-energy gradient algorithm (Okuyama-Yoshida *et al.* 2000). This method was applied to geometry optimization of the glycine zwitterion in aqueous solution. The glycine zwitterion was observed to be stable in the aqueous solution and this has also been found in experimental studies.

Free-energy differences of amino-acid substitutions in a protein have been evaluated (Shobana *et al.* 2000). In the study the focus was on the effects from the equilibration time, number of windows, and choice of transformation path. The intrinsic problem with dummy atoms in free-energy calculations was investigated. Stable free-energy differences with good reversibility and minimal variability were found.

Another method has been described for estimating the electrostatic free-energy differences using MD simulations (King & Barford, 1993). The results were within approximately 1–2 kcal $mol^{-1}$ of data from thermodynamic integration calculations. A speed-up of half the computational time was found. Since the method is based on the fact that the initial and final state are not too different the integrand of the thermodynamic integration equation was approximated with a linear function of the coupling parameter.

In a recent study the thermodynamic integration procedure based on Jarzynski's identity (Jarzynski, 1997) has been used to obtain equations for free-energy differences (Hummer, 2001). Free energies from this fast-growth thermodynamic integration showed very good agreement with results from umbrella sampling together with WHAM. It was further pointed out in the study that the fast-growth thermodynamic integration is very suitable for massively parallel computation.

One way of calculating the free-energy difference is to generate a long trajectory using a soft-core potential scaling for the mutated functional groups (Liu *et al.* 1996). This involves sampling

of the essential conformations of both the initial and final states and thereafter these conformations are used to determine the free-energy difference using the perturbation formula. This extrapolation method has been tested on a series of organic molecules and for molecules of similar size the results were reliable, but the inaccuracy increased for large functional groups.

Two issues, which are essential in free-energy calculations, have been analyzed (Sen & Nilsson, 1999a). First the dependence of the free-energy difference on the extent of perturbation for a specific transformation was investigated. Second the sampling error in free-energy difference calculations of mutagenesis in the free state and in complex was checked. In the first case no dependence was seen and in the second case the sampling errors were similar for the two different approaches.

Four different techniques, the thermodynamic integration method, free-energy perturbation method, the acceptance ratio method of Bennett, and a WHAM-based method, for determining the free energy have been compared (Radmer & Kollman, 1997). The two last methods showed the smallest errors when no intermediate states were taken into account. In the report an extension of the last method was also presented.

The ability to use higher derivatives of the free energy with respect to the coupling constant parameter, $\lambda$, has been analyzed (Smith & van Gunsteren, 1994). In the study it was observed that the derivatives up to the third had converged in a 750 ps simulation. The study also discussed the advantages of using higher derivatives in different applications.

By using a single MD simulation the hydration free energy of several small molecules were determined (Schäfer *et al.* 1999). In this investigation a previously published method was employed (Liu *et al.* 1996). Here the solute molecule was modeled as a dummy with soft-core interactions (Schäfer *et al.* 1999). The hydration free energies were very accurate and the computational efficiency was increased by 2–3 orders.

A second dynamic pharmacophore model based on multiple crystal structures has been developed (Carlson *et al.* 1999). This dynamic model was an improvement over the static models and can be used for searching databases to identify new inhibitors.

The free energy between two states of a biochemical system in equilibrium has been determined from a number of independent irreversible short simulations (Hendrix & Jarzynski, 2001). This fast-growth free-energy algorithm was applied to a Lennard–Jones fluid for calculating the excess chemical potential.

The accuracy of free-energy perturbation calculations has been assessed (Lu & Kofke, 2001). Here it is argued that the free-energy perturbation calculation should be carried out in a specific direction, i.e. that in which the entropy of the target is smaller than the entropy of the reference.

## 9. QM/MM calculations

The fundamentals of the QM/MM methodology were first described in a study of the catalytic mechanism of lysozyme (Warshel & Levitt, 1976). Several years later the QM/MM method was applied to exchange reactions in solution and the advantages and weaknesses were examined in detail (Singh & Kollman, 1986; Field *et al.* 1990). The total hamiltonian, $\boldsymbol{H}_{\mathrm{tot}}$, for the system is split into three parts

$$\boldsymbol{H}_{\mathrm{tot}} = \boldsymbol{H}_{\mathrm{MM}} + \boldsymbol{H}_{\mathrm{QM}} + \boldsymbol{H}_{\mathrm{QM/MM}},$$

where the two first terms describe the interactions within the MM and QM regions respectively. Once a certain description is chosen for each of these, the remaining issues, in the third term,

contain the interactions between the MM and QM parts. In these calculations the QM part is the most time consuming, nevertheless both *ab initio* and density functional methods, as well as semi-empirical QM codes, are used for $\boldsymbol{H}_{\mathrm{QM}}$ today.

Several recent reviews of QM/MM focus on the basics of the QM/MM method (Monard & Merz, 1999), the problem of how to treat the boundary between the QM and MM regions (Monard & Merz, 1999), combination with free-energy calculations of chemical reactions in enzymes (Monard & Merz, 1999; Kollman *et al.* 2001; Villà & Warshel, 2001), or on the validity of the approximations in mixed quantum-classical systems involving dynamical processes (Egorov *et al.* 1999).

If the QM part of the system is covalently linked to the MM part, the treatment of the boundary between the two has to be defined. This may be done through the introduction of additional link atoms, usually hydrogen-like, between the MM and QM parts (Singh & Kollman, 1986; Field *et al.* 1990, 2000). The link atoms complete the valency of the QM region while interacting through a MM bond with the MM region. Other approaches to this problem have also been proposed: Using a connection atom that interacts with the semi-empirical QM atoms as a differently parameterized QM atom and with the MM atoms as a conventional MM atom (Antes & Thiel, 1999) results in smooth potential surfaces because no extra centers are used. In the generalized hybrid orbital method (Gao *et al.* 1998; Amara *et al.* 2000) a number of hybrid orbitals are introduced on the boundary atom between the QM and MM fragments, allowing new biochemical systems to be handled without reparameterization. The pseudobond formalism (Zhang *et al.* 1999, 2000) in which the pseudobond mimicking the original bond is only treated in the QM part and, is based on a 1 free-valence atom with an effective core potential that replaces the boundary atom of the MM part and forms a boundary atom of the QM part. In a comparison of the link atom formalism and the local self-consistent field approach (Reuter *et al.* 2000) it is shown that the selection of the QM/MM electronic interactions in the frontier region plays an essential role in determining the overall energy and the QM electron distribution. In general the two methods show similar accuracy when care is taken concerning the frontier between the QM and MM regions.

The methodology is now at a point where it is of interest to improve the accuracy of the QM description at a reasonable computational cost. One example of this is a study of models of triosephosphate isomerase-catalyzed reactions, where the self-consistent-charge density-functional tight-binding (SCC-DFTB) method (Cui *et al.* 2001) was used in the QM part and found to give reliable energies.

In a spirit similar to the QM/MM approach QM calculations can be performed on a set of classically computed configurations (Wood *et al.* 1999; Sakane *et al.* 2000), molecular electrostatic potentials have been calculated using classical methods far from the molecule, and QM methods at a closer range (Hernández *et al.* 2000). Solvent effects have also been determined from QM together with MD simulations based on a mean-field theory (Sánchez *et al.* 2000), including mutual polarization of the solute and solvent molecules.

## 10. MD simulations of protein folding and unfolding

Proteins fold to their native structure through physicochemical processes as determined by their amino-acid sequences (Anfinsen, 1973). The most widely accepted view of how the protein-folding process occurs is based on the energy landscape view (Onuchic *et al.* 1997). Here the folding to the native structure is regarded as a process with multiple pathways, on an energy

landscape resembling a funnel with local minima, which could trap the protein in misfolded states. Recently MD simulations have been used in conjunction with experiments to characterize the denatured and transition-state conformations of proteins (reviewed in Brooks, 1998; Shea & Brooks, 2001). The folding pathways of the 36-residue villin headpiece subdomain were examined using a 1-μs MD simulation (Duan & Kollman, 1998), which currently is the longest MD simulation of a protein conducted in aqueous solution. In the simulation starting from an unfolded conformation a hydrophobic collapse and helix formation were observed followed by conformational rearrangements. Two folding pathways to a native-like structure were found.

Simulations generate large amounts of data, and to allow meaningful comparisons of a folding/unfolding simulation with experimental observations, and with other simulations, pertinent characteristics of the process have to be defined, and tools to analyze these properties have to be developed. Commonly used characteristics are: the number of amino acid–amino acid contacts (native or overall); radius of gyration ($R_g$): r.m.s. deviation from native structure; the structure index, which is a product of secondary structure and local packing interactions at a given residue (Daggett, 2002); the $\phi$-value, which for a given amino acid substitution measures the ratio of (de)stabilization of the transition state (TS) and folded state (S) relative to the unfolded state (U), $\phi = \Delta\Delta G_{TS-U}/\Delta\Delta G_{F-U}$ (Fersht, 1995). Several analysis methods, based on geometry or more general properties, for the protein unfolding process have been applied to a set of proteins (Kazmirski *et al.* 1999). In these unfolding simulations a variety of pathways to the denatured state were found, and it was concluded that a folding/unfolding pathway characterized by some physical property may involve structurally diverse conformations.

In the following sections the focus is on the folding of proteins and a few recent studies of shorter peptides. Folding of peptides has been reviewed elsewhere (Brooks & Case, 1993; Brooks, 1998; Shea & Brooks, 2001).

## 10.1 High-temperature effects

In an early high-temperature simulation of the S-peptide analog in aqueous solution at 358 K (Tirado-Rives & Jorgensen, 1991), the most interesting findings were that the breaking and formation of main-chain hydrogen bonds occurred through the following sequence $\alpha \leftrightarrow 3_{10} \leftrightarrow$no hydrogen bond. In the following decade several high-temperature unfolding simulations on small proteins were reported.

Upon unfolding of the BPTI (Daggett & Levitt, 1992, 1993) a compact, partially denatured state, which was 10–25% larger than the native state, was observed, and the volume increase was not due to penetrating water molecules. The hydrophobic core was well preserved and the major conformational changes were found in the loop regions. The simulations proposed a model for the molten globule state. In a more recent study the unfolding of BPTI was examined by eleven high-temperature simulations in water (Kazmirski & Daggett, 1998a). These unfolding structures were found to agree with experimental results from NMR, fluorescence energy transfer and ANS binding. In a neutral pH environment a slight increase of the $R_g$ for the unfolded conformations was observed, and protonation of acidic residues was found to result in an expansion of the protein.

Similarly expanded and partially unfolded structures of apomyoglobin showed secondary-structure content and tertiary-structure contacts in agreement with experimental findings. The most stable $\alpha$-helical regions were found to be in contact with the other helices (Tirado-Rives & Jorgensen, 1993).

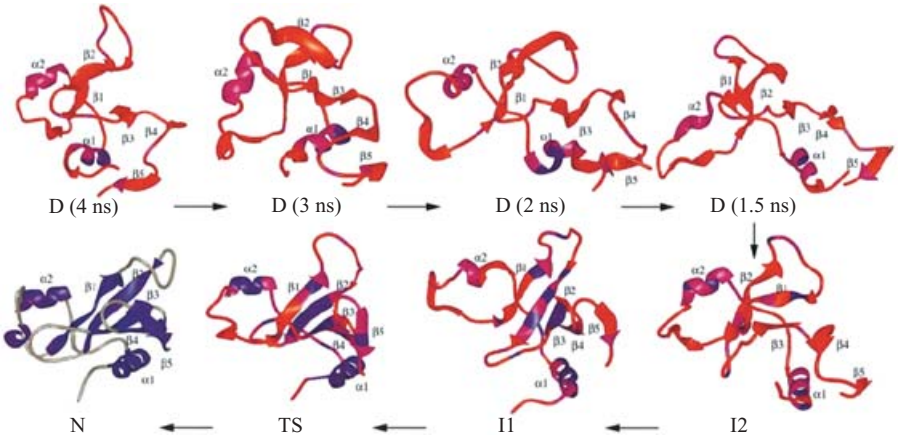


**Fig. 4.** Structures from a simulation of barnase colored by structure index values $S$ (light red $S<0{\cdot}3$, red $S\geqslant 0{\cdot}3$ and $S<0{\cdot}7$, blue $S\geqslant 0{\cdot}7$). Here I1 is the major intermediate state, I2 is the second more disordered intermediate state, TS is the major transition state, N is the native NMR structure and D are the denatured states (Wong *et al.* 2000). [Reprinted from the *Journal of Molecular Biology* **296** (2000), 1257–1282, with permission from the publisher, Academic Press.]

In unfolding simulations of lysozyme, two-state behavior with no distinct folding domains (Mark & van Gunsteren, 1992) has been observed, indicating that an initial collapse was followed by formation of a secondary structure. Similar results were obtained in a study in which a nearly complete loss of secondary structure and a divided hydrophobic core were found (Hünenberg *et al.* 1995), supporting a folding model with a hydrophobic collapse preceding the secondary and tertiary structure formation. Williams *et al.* (1997) found that the lysozyme $\beta$-sheet was still intact in the transition state, whereas in simulations by Kazmirski & Daggett (1998b), the $\beta$-domain unfolded first, leaving the $\alpha$-domain structured. The observed conformational substates suggested that the major kinetic intermediate might be made up of partially folded states with differences in helix packing (Kazmirski & Daggett, 1998b).

Barnase and CI2 have been the subject of a large number of unfolding simulations by several groups. In barnase the core unfolds concurrently with (Caflisch & Karplus, 1994, 1995), or followed by (Bond *et al.* 1997; Li & Daggett, 1998) fraying secondary structure. The unfolding of the core has been observed to be accompanied by water penetration after disruption of parts of the secondary structure (Caflisch & Karplus, 1994). In a number of high-temperature barnase simulations two intermediates on the unfolding pathway were found, and the calculated structure index, which takes into account both secondary and tertiary interactions, agreed quite well with the experimentally measured $\phi$-values from protein engineering and NMR experiments (Bond *et al.* 1997; Li & Daggett, 1998; Daggett *et al.* 1998). A heterogeneous denatured ensemble was obtained in which fluid hydrophobic clusters were found, together with residual structure in the $\alpha$1 and $\alpha$2 helices and tertiary contacts in the $\beta$(3–4) region (Wong *et al.* 2000) also in agreement with NMR data (Fig. 4). Engineered disulfide bonds were found to constrain the denatured state (Clarke *et al.* 2000).

For CI2 the secondary structure was partially unchanged, but in the hydrophobic core tertiary contacts were broken (Li & Daggett, 1994, 1996). $\phi$-Values determined from the unfolding simulations were in quantitative agreement with experimental values. The loss of secondary and tertiary contacts occurred concurrently in these simulations, and the transition-state structure

was thus closer to the native state than to the unfolded state. The structure index for the transition-state conformation was calculated and found to be in good overall agreement with experimental $\phi$-values, which suggests that in the transition state the $\alpha$-helix is partially intact and the $\beta$-sheet disrupted. The data further indicated that the folding proceeds according to the nucleation-collapse mechanism. From a simulated transition-state ensemble of the CI2 protein, mutants with an accelerated folding rate were designed (Ladurner *et al.* 1998). The folding rate was increased 40-fold by elimination of unfavorable interactions at the C terminus and in the reactive loop region. In unfolding simulations with an implicit solvent model the initial unfolding event occurred between the helix and the $\beta$3–$\beta$4 strands, and in the transition state only 25% of the native contacts were formed (Lazaridis & Karplus, 1997). A wide diversity in the unfolding pathways was found, suggesting a reconciliation of the the classical, single path, view of folding with the new view in which the energy hypersurface has a bias towards the native state, thus allowing folding to occur without following a specific path. In an investigation of the 46-residue helical fragment B of protein A the expansion of the transition-state volume was similar to what has been found experimentally for CI2 (Guo *et al.* 1997), and a native-like topology of the transition state was seen with approximately 30% of the native tertiary structure and 50–70% of the helical structure.

Folding via the nucleation-condensation mechanism has been observed for the $\alpha+\beta$ protein suc1, which is involved in cell cycle regulation (Alonso *et al.* 2000; Schymkowitz *et al.* 2000), and for the FKBP12 protein (Fulton *et al.* 1999). The suc1 study used $\phi$-value analysis to verify the conformations of the rate-limiting transition state for the folding–unfolding and for the intermediate state in the refolding. In the protein the area around the $\beta$2- and $\beta$4-strands builds up the folding nucleus as suggested by the nucleation-condensation mechanism. Native secondary structure was obtained at the $\alpha$1-helix and at the $\beta$-sheet. The suc1 protein was observed to fold by a two-state process and the overall results from the MD simulations agreed very well with experimental data. In FKBP12 all the interactions in the transition states for folding were much weaker than in the native state. No secondary structure was fully formed in the transition state, in which the $\beta$2-, $\beta$4- and $\beta$5-strands were the most structured and comprised the folding nucleus together with the C terminus of the $\alpha$-helix (Fulton *et al.* 1999).

In the engrailed homeodomain protein secondary and tertiary structure folding are uncoupled, leading to a very rapid folding process. This allows extrapolation of experimental refolding and unfolding data to 100 °C, giving an estimated half-life for unfolding of 7·5 ns at 100 °C (Mayor *et al.* 2000), a timescale which is accessible to MD simulations. This unfolding was seen in the MD simulation. Furthermore the transition states from MD simulations at 100 °C and at 225 °C were quite similar and in agreement with the experimentally determined position of the transition state along the folding pathway.

The folding mechanism and thermodynamics of segment B1 of streptococcal protein G has been investigated using high-temperature MD simulations (Sheinerman & Brooks, 1998a, b). First a database of initial conformations was generated using MD simulations at 400 K and thereafter followed by equilibration and sampling at 298 K. In the investigation an initial collapse was observed prior to the secondary-structure formation. A free-energy landscape of the folding (Fig. 5) along the reaction coordinate, which was defined by the $R_g$, was constructed using WHAM (Ferrenberg & Swendsen, 1989; Boczko & Brooks, 1993). The energy landscape of segment B1 of protein G is displayed in Figure 5. The protein structure formation was found to arise from two regions, the N terminus and the turn connecting the third and fourth $\beta$-strands. In the folding process the $\alpha$-helix was formed early and the $\beta$-sheet followed the overall topology

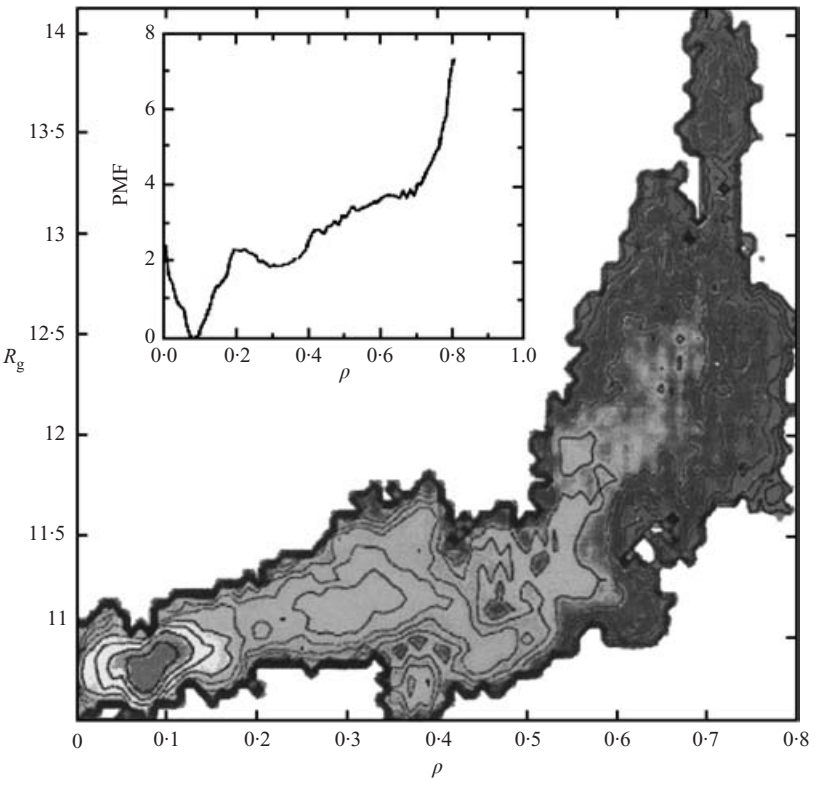


**Fig. 5.** Free-energy surfaces along the reaction coordinate $\rho$ (defined by the number of native contacts) (inset) and as a function of $\rho$ and $R_g$ (radius of gyration). Contours are drawn every 0·5 kcal mol$^{-1}$ (Sheinerman & Brooks, 1998b). [Figure reprinted with permission from the *Proceedings of the National Academy of Sciences USA* **95** (1998), 1562–1567.] (Copyright 1998 National Academy of Sciences, USA.)

(Sheinerman & Brooks, 1998a, b). Water molecules were observed to be present in the protein core until the late stage of the folding process, when 80% of the native contacts were formed. In the N terminus of the $\alpha$-helix an experimentally determined non-native contact, which was important in the early stages of the folding, was observed.

The folding free-energy landscape for the three-stranded $\beta$-sheet protein Betanova constructed using the WHAM method applied to MD simulations starting from a database obtained from four high-temperature simulations (Bursulaya & Brooks, 1999). This demonstrated that the folding process took place in two collapse stages. The driving force for folding was the native side-chain contacts, with approximately 60% of the native contacts being formed in the transition state. The thermodynamic stability of Betanova was estimated to approximately 1 kcal mol$^{-1}$. The free-energy folding surface for Betanova was found to be similar for the implicit generalized Born model and the explicit water model at 275 K (Bursulaya & Brooks, 2000). Another synthetic 20-residue $\beta$-sheet peptide has been investigated using high-temperature implicit solvation MD simulations (Ferrara & Caflisch, 2000). In the major folding pathway the contacts were formed between the second and third strands and thereafter the N-terminal strand docked onto these strands. This folding picture agreed with the free-energy surface, which was observed for the Betanova protein (Bursulaya & Brooks, 1999).

Potato carboxypeptidase inhibitor was folded from partially unfolded conformations using MD simulations with disulfide bond constraints (Martí-Renom *et al.* 1998). The partially unfolded structures were obtained from high-temperature MD simulations and seven conformations were chosen for the nine refolding MD simulations, in which the temperature was decreased to room temperature. In all except one case the refolding simulations re-created the native simulation conformation within 1·8–3·2 Å for the main-chain of the core. Just a few native hydrogen bonds were re-created and the structures obtained from more unfolded initial conformations were closer to the native structure.

### 10.2 Co-solvent and polarization effects

Barnase has been studied in both water and 8 M urea solution at 298 K for 0·5 ns and at 358 K for 2·0 ns (Tirado-Rives *et al.* 1997). This study examined the effect of urea in comparison with aqueous solution on the folding process. The room-temperature water simulation was in close agreement with experimental data. In the high-temperature simulation the unfolding pathway closely mimicked the energetic trends seen by protein engineering. The observations from both the urea simulations suggested that urea promotes protein unfolding by stabilization of the unfolded conformations instead of destabilizing the native form. Another investigation has also studied the unfolding of barnase by MD simulations with lengths of 555–870 ps in water and urea solution at two different temperatures, 300 and 360 K (Caflisch & Karplus, 1999). For the simulations with and without urea the solvent accessible surface area of polar and non-polar groups were very similar. This observation disagreed with the first study (Tirado-Rives *et al.* 1997), that found the solvent-accessible surface area of hydrophobic regions to increase when urea was present. The study further suggested that effects of both polar and non-polar groups of proteins were involved in urea denaturation (Caflisch & Karplus, 1999).

The folding reactions of a $\beta$-heptapeptide and a $\beta$-hexapeptide in methanol have been thoroughly studied using MD simulations at different temperatures (Daura *et al.* 1997, 1998a, 1999b). The free energy of folding was approximated to approximately 0 and 1 kcal mol$^{-1}$ for the $\beta$-heptapeptide and $\beta$-hexapeptide, respectively. At room temperature the $3_1$-helix, which is the experimentally observed structure of the $\beta$-heptapeptide, was not stable and at 350 K the peptide unfolded and refolded in a time period of approximately 200 ps. Similar conformations of the peptide were obtained both without and with NOE distance restraints from NMR experiments. The importance of explicit solvent was also demonstrated (Daura *et al.* 1999a).

Non-classical and many-body treatment of the interactions in the folding of the Betanova protein have been examined (van der Vaart *et al.* 2000) in a study of the effects from charge transfer and polarization, which were found to be nearly constant during the folding process, and therefore the driving force for folding was mainly the electrostatic interactions. These observations demonstrated that effective two-body potentials could be used for an accurate picture of the folding of Betanova.

### 10.3 External force effects

A 46-residue 3-helix bundle from fragment B of staphylococcal protein A was examined using MD simulations, in which the unfolding process was driven along the $R_g$ as a reaction coordinate (Boczko & Brooks, 1995). The study indicated that the interaction between helices I and II was important in the initial phase of folding and thereafter helix III associated with these helices. The folded state had 2·6 kcal mol$^{-1}$ lower free energy than the unfolded state.

Hen egg-white lysozyme unfolding has been investigated using a penalty function, which increased the mean radius of the protein (Hünenberger *et al.* 1995). Both the force constant in the penalty function and location of the center of mass of the protein clearly affected the unfolding process. This approach to model unfolding will always bias the unfolding. A folding mechanism with docking of preformed $\alpha$ and $\beta$ secondary structures was suggested. The unfolding of hen egg-white lysozyme has been examined using the path exploration with distance constraints approach (Gilquin *et al.* 2000). An important role in the unfolding was the broken contacts between the $\alpha$-domain and the $\beta$-domain. This $\beta$-sheet unfolding was due to the release of the Ile55-Leu56 turn from the hydrophobic patch of the $\alpha$-domain. The MD simulation study suggested that the Ile55 residue was critical for correct folding and this has also been supported by experimental results.

The unfolding of two titin subunits, I27 domain and the homology-built I28 domain, was examined using steered MD simulations, in which one terminus was fixed and a force was applied to the other terminus (Lu *et al.* 1998). In the study six MD simulations with different pulling speeds were performed. To initiate unfolding of the Ig domains the backbone hydrogen bonds between the parallel $\beta$-strands A′ and G and antiparallel $\beta$-strands A and B were broken. After the separation of the $\beta$-strands the hydrophobic core became exposed and then the domains could unravel. Due to lack of computer resources the MD simulations had a speed 6–8 orders of magnitude larger than in the experiments. This yielded a 4 times larger peak force in the MD simulations compared to experiments. The force-induced unfolding pathway was unlikely to be the same as that observed for spontaneous unfolding. To investigate the extension that precedes the main unfolding event MD simulations were carried out with forces from 20 to 500 pN being applied to the termini of the I27 domain (Marszalek *et al.* 1999). An extension of approximately 2 Å was observed for forces up to 50 pN, and for forces larger than 100 pN the I27 domain showed extensions of 6–9 Å, in good agreement with AFM experiments, which showed an extension of 6·6 Å. If the hydrogen bonds between the $\beta$-strands broke then the forces were larger than 100 pN and unfolding of the domain occurred. The breaking of these water-mediated hydrogen bonds was found to be a key event in the force-induced unfolding process and took place before the unfolding could proceed (Lu & Schulten, 2000; Lu *et al.* 2000). From the theory of mean first-passage time the energy barrier between the folded and unfolded states was estimated to have a width of around 3 Å and a height of approximately 20 kcal $mol^{-1}$ (Lu & Schulten, 1999a), which agreed well with the experimentally determined 22 kcal $mol^{-1}$. After the barrier was passed only weak forces were needed to fully unfold the protein and break the remaining inter-strand hydrogen bonds.

Stretching and unfolding have been observed in ten different protein domains which were investigated using forced MD simulations (Lu & Schulten, 1999b). The domains were divided in two classses: 'I' for proteins, which exhibit a potential barrier at short extension; 'II' for proteins, which do not show a potential barrier against stretching. This barrier arises due to the breaking of several inter-strand backbone hydrogen bonds. A force larger than 1500 pN was needed to unfold the class I $\beta$-sandwich domains. Thus the study suggested that this simulation technique might be used to design proteins with barriers of desired height.

Unfolding of the fibronectin type-3 domains $^{9}$Fn3 and $^{10}$Fn3 and the heterodimer of these domains has been studied using MD simulations with applied external forces (Paci & Karplus, 1999). The two domains showed similar unfolding pathways, but in the heterodimer the $^{9}$Fn3 domain was completely unfolded before the initiation of unfolding of the $^{10}$Fn3 domain. The van der Waals interactions were found to be the dominating forces in stabilizing the $\beta$-sandwich

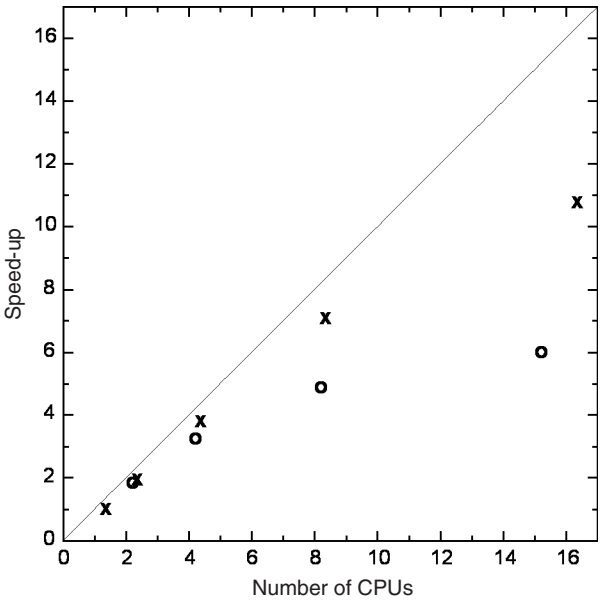


**Fig. 6.** Efficiency in parallel computing using CHARMM for different numbers of nodes. The 14 000 atom system run on 866 MHz Pentium III processors, using Fast Ethernet (○) or SCI communication hardware (×) (Scali AS, Norway).

and also the separate $\beta$-sheets. Thus the electrostatic interactions only played a minor role in stabilizing the domains. Two $\beta$-sandwich proteins, a fibronectin type-III (Fn3) domain and an immunoglobulin (Ig) domain, and two $\alpha$-helical proteins, an acyl–coenzyme A-binding domain and an $\alpha$-spectrin domain, have been unfolded using force-induced MD simulations (Paci & Karplus, 2000). In the unfolding process of the Fn3 domain two barriers were observed, but for the Ig domain just one barrier. The unfolding of the Ig domain required a larger force than was needed for the Fn3 domain. It was also found that larger forces were required for the $\beta$-sandwich domains compared to the $\alpha$-helical domains. These domains have also been studied using high-temperature MD simulations. A recent overview report of the unfolding of the Ig and fibronectin domains has been published (Isralewitz *et al.* 2001).

The unfolding process of CI2 has been characterized using implicit solvation-targeted MD simulations and compared to unrestrained MD simulations at high temperatures (Ferrara *et al.* 2000). The different approaches gave qualitatively similar results, in that the disruption of the secondary and tertiary structures took place in the same way, although the pathways were diverse. These MD simulations showed that the sequence of events occurred as had previously been seen in MD simulations using different force fields and solvation models. Therefore the study concluded that the targeted MD simulations gave realistic pathways.

One MD simulation study has so far recognized unfolding of a protein driven by pressure (Floriano *et al.* 1998). The transition to the molten globule state of metmyoglobin was found to start at 600 MPa and unfolding occurred at approximately 900 MPa. Approximately 30–40% of secondary-structure content was obtained in the molten globule. The study further suggested that the helix formation process followed the $(i, i+2)\rightarrow(i, i+4)$ hydrogen-bond pathway, and the relative stability among the metmyoglobin helices was predicted.

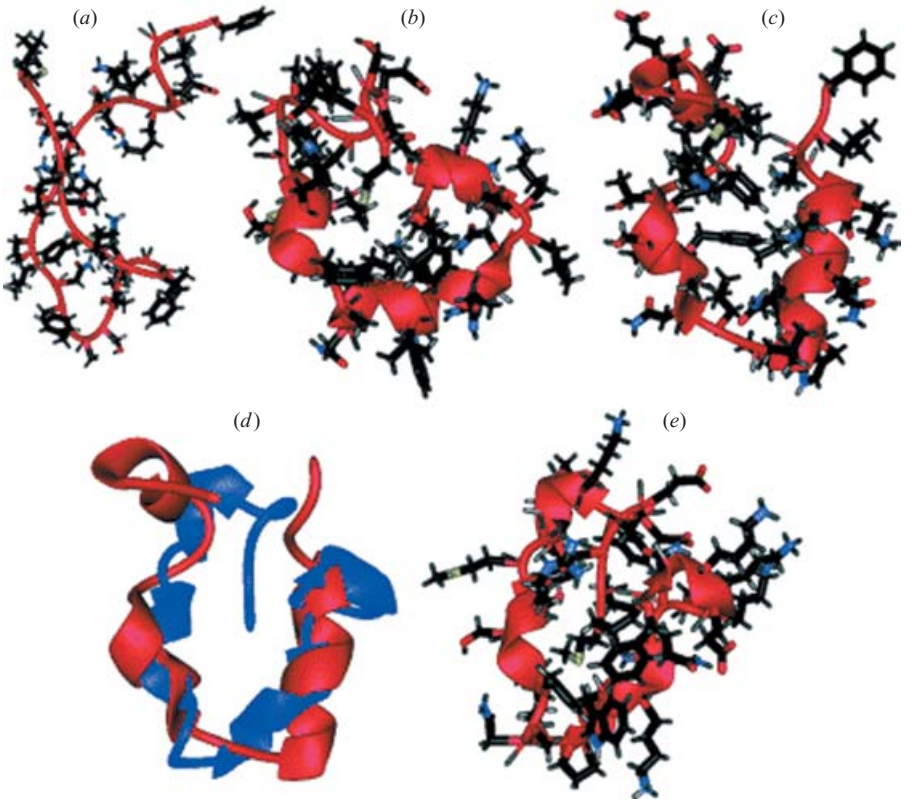


**Fig. 7.** Currently the longest protein simulation (1 μs) of the villin headpiece subdomain. The simulation was started from: (*a*) an unfolded extended conformation; (*b*) a partially folded state at 980 ns; (*c*) native structures; (*e*) a representative structure of the most stable cluster; (*d*) the overlap of the native structure with the most stable cluster structures. Color code [except (*d*)]: red, main chain atoms and oxygen; black, non-main chain carbon; blue, non-main chain nitrogen; gray, hydrogen; yellow, sulfur (Duan & Kollman, 1998). [Reprinted with permission from *Science* **282** (1998), 740–744.] (Copyright 1998 American Association for the Advancement of Science.)

## 11. On the horizon

Presently very impressive computational resources are affordable for single departments, or research groups. Most of the commonly used MD programs have been parallelized to take advantage of inexpensive commodity personal computers. New programs are also being designed for massively parallel MD (Straatsma *et al.* 2000). Global optimization of energy functions for proteins have been created using parallel algorithms (Lee *et al.* 2000a). The study further discusses the efficiency of parallel computation. Multiple-level parallel schemes and new methodology such as multiple time-step integration and non-canonical variable transformation to excite long lengthscale motion have been reviewed (Tuckerman *et al.* 2000). In another investigation the $P^3M$ method was favored for parallel implementations over the FM or Ewald methods (Pollock & Glosli, 1996). As an example of the computational efficiency in parallel MD, timing data for a benchmark simulation of myoglobin in water (Fig. 6) show that standard components can be effectively used to run approximately 8–16 processors in parallel.

Currently the longest MD simulation of a protein carried out is 1 μs (Duan & Kollman, 1998). The study focused on the villin headpiece subdomain (Fig. 7). A massively parallel computer with

256 CPU processors was utilized in this MD simulation. For the computationally demanding long-range non-bonded protein–solvent and solvent–solvent interactions a spherical truncation method with an 8 Å residue-based cut-off was applied, whereas the intramolecular non-bonded interactions of the protein were not truncated. This can be seen as the first step for MD simulations towards reaching lengthy timescales (Daggett, 2000) and in the near future longer simulations are expected to appear, both in terms of the length of a single trajectory and in terms of aggregate time for several simulations analyzed together. Recently some very large systems like GroEl (Ma *et al.* 2000) virus capsids (Phelps *et al.* 2000; Speelman *et al.* 2001) and membrane proteins (de Groot & Grubmüller, 2001) have also been the subject of MD simulations.

These computational advances have also made the MD simulation method available for a wider audience of researchers, and make it possible to tackle large biochemical systems like viral capsids, membrane proteins and protein–DNA complexes. At the same time the timescale within reach for MD simulations has been shifted upwards, allowing improved sampling. This, together with the continuously improving force fields, allows us to simulate proteins, nucleic acids, membranes and complexes of various kinds, with trajectories that usually are within a r.m.s. deviation for the backbone of 2 Å from the X-ray or NMR structure.

The speed of computers will no doubt keep increasing, and this additional power will be used to study the same kinds of system as have been studied in the first 25 years of biomolecular MD, but with better sampling (longer or more simulations), and using better force fields that include polarization effects. There are also very exciting prospects for the study of intriguing biological problems, some of which are already appearing, concerning the regulation of protein activity (Young *et al.* 2001), allostery, water (de Groot & Grubmüller, 2001; Tajkorshid *et al.* 2002) and ion channels (Åqvist & Luzhkov, 2000; Bernèche & Roux, 2001), protein–protein and protein–DNA/RNA interactions in many different cellular contexts.

## 12. Acknowledgements

This work was supported by the Magnus Bergvall foundation and the Swedish Research Council.

## 13. References

Adams, D. J. (1979). Computer simulation of ionic systems: the distorting effects of the boundary conditions. *Chemical Physics Letters* **62**, 329–332.

Alder, B. J. & Wainwright, T. E. (1957). Phase transition for a hard sphere system. *Journal of Chemical Physics* **27**, 1208–1209.

Allen, M. P. & Tildesley, D. J. (1987). *Computer Simulation of Liquids*. New York: Oxford University Press.

Alonso, D. O. V., Alm, E. & Daggett, V. (2000). Characterization of the unfolding pathway of the cell-cycle protein p13suc1 by molecular dynamics simulations: implications for domain swapping. *Structure, Folding & Design* **8**, 101–110.

Amara, P., Field, M. J., Alhambra, C. & Gao, J. (2000). The generalized hybrid orbital method for combined quantum mechanical/molecular mechanical calculations: formulation and tests of the analytical derivatives. *Theoretical Chemical Accounts* **104**, 336–343.

Andersen, H. C. (1980). Molecular dynamics simulations at constant pressure and/or temperature. *Journal of Chemical Physics* **72**, 8205–8212.

Andersen, H. C. (1983). RATTLE: a 'velocity' version of the SHAKE algorithm for molecular dynamics calculations. *Journal of Computational Physics* **52**, 24–34.

Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science* **181**, 223–230.

Antes, I. & Thiel, W. (1999). Adjusted connection atoms for combined quantum mechanical and molecular mechanical methods. *Journal of Physical Chemistry* (*A*) **103**, 9290–9295.

Åqvist, J. & Luzhkov, V. (2000). Ion permeation mechanism of the potassium channel. *Nature* **404**, 881–884.

Åqvist, J., Luzhkov, V. B. & Brandsdal, B. O. (2002). Ligand binding affinities from MD simulations. *Accounts of Chemical Research* **35**, 358–365.

Auffinger, P. & Beveridge, D. L. (1995). A simple test for evaluating the truncation effects in simulations of systems involving charged groups. *Chemical Physics Letters* **234**, 413–415.

Auffinger, P. & Westhof, E. (1998). Simulations of the molecular dynamics of nucleic acids. *Current Opinion in Structural Biology* **8**, 227–236.

Balsera, M., Stepaniants, S., Izrailev, S., Oono, Y. & Schulten, K. (1997). Reconstructing potential energy functions from simulated force-induced unbinding processes. *Biophysical Journal* **73**, 1281–1287.

Banavali, N. K., Im, W. & Roux, B. (2002). Electrostatic free energy calculations using the generalized solvent boundary potential method. *Journal of Chemical Physics* **117**, 7381–7388.

Banba, S. & Brooks III, C. L. (2000). Free energy screening of small ligands binding to an artificial protein cavity. *Journal of Chemical Physics* **113**, 3423–3433.

Banba, S., Guo, Z. Y. & Brooks III, C. L. (2000). Efficient sampling of ligand orientations and conformations in free energy calculations using the $\lambda$-dynamics method. *Journal of Physical Chemistry* (*B*) **104**, 6903–6910.

Banks, J. L., Kaminski, G. A., Zhou, R., Mainz, D. T., Berne, B. J. & Friesner, R. A. (1999). Parametrizing a polarizable force field from *ab initio* data. I. The fluctuating point charge model. *Journal of Chemical Physics* **110**, 741–754.

Bartels, C. (2000). Analyzing biased Monte Carlo and molecular dynamics simulations. *Chemical Physics Letters* **331**, 446–454.

Basma, M., Sundara, S., Çalgan, D., Vernali, T. & Woods, R. J. (2001). Solvated ensemble averaging in the calculation of partial atomic charges. *Journal of Computational Chemistry* **22**, 1125–1137.

Batcho, P. F., Case, D. A. & Schlick, T. (2001). Optimized particle-mesh Ewald/multiple-time step integration for molecular dynamics simulations. *Journal of Chemical Physics* **115**, 4003–4018.

Bayly, C. I., Cieplak, P., Cornell, W. D. & Kollman, P. A. (1993). A well-behaved electrostatic potential based method using charge restraints for deriving charges – the RESP model. *Journal of Physical Chemistry* **97**, 10269–10280.

Beachy, M., Chasman, D., Murphy, R., Halgren, T. & Friesner, R. (1997). Accurate *ab initio* quantum chemical determination of the relative energetics of peptide conformations and assessment of empirical force fields. *Journal of the American Chemical Society* **119**, 5908–5920.

Berendsen, H. J. C. (1993). Electrostatic interactions. In *Computer Simulations of Biomolecular Systems: Theoretical and Experimental Applications*, pp. 161–181. Leiden: ESCOM.

Berendsen, H. J. C. & Hayward, S. (2000). Collective protein dynamics in relation to function. *Current Opinion in Structural Biology* **10**, 165–169.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics* **81**, 3684–3690.

Berendsen, H. J. C. & Tieleman, D. P. (1998). Molecular dynamics: studies of lipid bilayers. In *Encyclopedia of computational chemistry*, John Wiley and Sons, Chichester, 1639–1650.

Berkowitz, M. & McCammon, J. A. (1982). Molecular dynamics with stochastic boundary conditions. *Chemical Physics Letters* **90**, 215–217.

Berne, B. J. & Straub, J. E. (1997). Novel methods of sampling phase space in the simulation of biological systems. *Current Opinion in Structural Biology* **7**, 181–189.

Bernèche, S. & Roux, B. (2001). Energetics of ion conduction through the $K^+$ channel. *Nature* **414**, 73–77.

Bertrand, H. O., Pullman, A., Zakrzewska, K., Hartmann, B. & Fermandjian, S. (1999). Determination of a set of parameters for the molecular modelling of phosphorothioate DNA. *Theoretical Chemical Accounts* **101**, 269–273.

Beveridge, D. L. & DiCapua, F. M. (1989). Free energy via molecular simulation – applications to chemical and biomolecular systems. *Annual Review of Biophysics and Biophysical Chemistry* **18**, 431–492.

Beveridge, D. L. & McConnell, K. J. (2000). Nucleic acids: theory and computer simulation, Y2K. *Current Opinion in Structural Biology* **10**, 182–196.

Boczko, E. M. & Brooks III, C. L. (1993). Constant-temperature free energy surfaces for physical and chemical processes. *Journal of Physical Chemistry* **97**, 4509–4513.

Boczko, E. M. & Brooks III, C. L. (1995). First principles calculation of the folding free energy of a three-helix bundle protein. *Science* **269**, 393–396.

Bond, C. J., Wong, K.-B., Clarke, J., Fersht, A. R. & Daggett, V. (1997). Characterization of residual structure in the thermally denatured state of barnase by simulation and experiment: description of the folding pathway. *Proceedings of the National Academy of Sciences USA* **94**, 13409–13413.

Bonvin, A. M. J. J., Mark, A. E. & van Gunsteren, W. F. (2000). The GROMOS96 benchmarks for molecular simulations. *Computer Physics Communications* **128**, 550–557.

Boresch, S. & Steinhauser, O. (1999). Rationalizing the effects of modified electrostatic interactions in computer simulations: the dielectric self-consistent field method. *Journal of Chemical Physics* **111**, 8271–8274.

Born, M. & von Karmen, T. (1912). Über Schwingungen in Raumgittern. *Physikalische Zeitschrift* **13**, 297–309.

Bosch, D., Foloppe, N., Pastor, N., Pardo, L. & Campillo, M. (2001). Calibrating nucleic acids torsional energetics in force-field: insight from model compounds. *Journal of Molecular Structure: Theochem* **537**, 283–305.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM: a program for macromolecular energy,

minimization, and dynamics calculations. *Journal of Computational Chemistry* **93**, 2487–2502.

Brooks III, C. L. (1998). Simulations of protein folding and unfolding. *Current Opinion in Structural Biology* **8**, 222–226.

Brooks III, C. L., Brünger, A. T. & Karplus, M. (1985a). Active site dynamics in protein molecules: a stochastic boundary molecular-dynamics approach. *Biopolymers* **24**, 843–865.

Brooks III, C. L. & Case, D. A. (1993). Simulations of peptide conformational dynamics and thermodynamics. *Chemical Reviews* **93**, 2487–2502.

Brooks III, C. L. & Karplus, M. (1983). Deformable stochastic boundaries in molecular dynamics. *Journal of Chemical Physics* **79**, 6312–6325.

Brooks III, C. L., Pettitt, B. M. & Karplus, M. (1985b). Structural and energetic effects of truncating long range interactions in ionic and polar fluids. *Journal of Chemical Physics* **83**, 5897–5908.

Brooks III, C. L., Pettitt, B. M. & Karplus, M. (1988). *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics*, vol. 71. *Advances in Chemical Physics*. New York: John Wiley & Sons.

Brünger, A. T., Brooks, C. L. & Karplus, M. (1985). Active site dynamics of ribonuclease. *Proceedings of the National Academy of Sciences USA* **82**, 8458–8462.

Bursulaya, B. D. & Brooks III, C. L. (1999). Folding free energy surface of a three-stranded $\beta$-sheet protein. *Journal of the American Chemical Society* **121**, 9947–9951.

Bursulaya, B. D. & Brooks III, C. L. (2000). Comparative study of the folding free energy landscape of a three-stranded $\beta$-sheet protein with Explicit and Implicit solvent models. *Journal of Physical Chemistry* (*B*) **104**, 12378–12383.

Caflisch, A. & Karplus, M. (1994). Molecular dynamics simulation of protein denaturation: solvation of the hydrophobic cores and secondary structure of barnase. *Proceedings of the National Academy of Sciences USA* **91**, 1746–1750.

Caflisch, A. & Karplus, M. (1995). Acid and thermal denaturation of barnase investigated by molecular dynamics simulations. *Journal of Molecular Biology* **252**, 672–708.

Caflisch, A. & Karplus, M. (1999). Structural details of urea binding to barnase: a molecular dynamics analysis. *Strucure: Folding and Design* **7**, 477–488.

Çağin, T. & Pettitt, B. M. (1991a). Grand molecular dynamics: a method for open systems. *Molecular Simulations* **6**, 5–26.

Çağin, T. & Pettitt, B. M. (1991b). Molecular dynamics with a variable number of molecules. *Molecular Physics* **72**, 169–175.

Çağin, T. & Pettitt, B. M. (1991c). A method for modeling icosahedral virions – rotational symmetry boundary-conditions. *Journal of Computational Chemistry* **12**, 627–634.

Carlson, H. A., Masukawa, K. M. & McCammon, J. A. (1999). Method for including the dynamic fluctuations of a protein in computer-aided drug design. *Journal of Physical Chemistry* (*A*) **103**, 10213–10219.

Cheatham III, T. E. & Brooks, B. R. (1998). Recent advances in molecular dynamics simulation towards the realistic representation of biomolecules in solution. *Theoretical Chemical Accounts* **99**, 279–288.

Cheatham III, T. E. & Kollman, P. A. (2000). Molecular dynamics simulation of nucleic acids. *Annual Review of Physical Chemistry* **51**, 435–471.

Chen, X. & Tropsha, A. (1999). Generalized linear response method: application to hydration free energy calculations. *Journal of Computational Chemistry* **20**, 749–759.

Cheng, A., Best, S. A., Merz Jr., K. M. & Reynolds, C. H. (2000). GB/SA water model for the Merck molecular force field (MMFF). *Journal of Molecular Graphics and Modelling* **18**, 273–282.

Cheng, A. & Merz Jr., K. M. (1999). Application of a multiple time step algorithm to biomolecular systems. *Journal of Physical Chemistry* (*B*) **103**, 5396–5405.

Cheng, H., Greengard, L. & Rokhlin, V. (1999). A fast adaptive multipole algorithm in three dimensions. *Journal of Computational Physics* **155**, 468–498.

Christianson, L. A., Lucero, M. J., Appella, D. H., Klein, D. A. & Gellman, S. H. (2000). Improved treatment of cyclic $\beta$-amino acids and successful prediction of $\beta$-peptide secondary structure using a modified force field: AMBER*C. *Journal of Computational Chemistry* **21**, 763–773.

Chun, H. M., Padilla, C. E., Chin, D. N., Watanabe, M., Karlov, V. I., Alper, H. E., Soosaar, K., Blair, K. B., Becker, O. M., Caves, L. S. D., Nagle, R., Haney, D. N. & Farmer, B. L. (2000). MBO(N)D: a multibody method for long-time molecular dynamics simulations. *Journal of Computational Chemistry* **21**, 159–184.

Ciccotti, G., Martyna, G. J., Melchionna, S. & Tuckerman, M. E. (2001). Constrained isothermal-isobaric molecular dynamics with full atomic virial. *Journal of Physical Chemistry* (*B*) **105**, 6710–6715.

Cieplak, P., Caldwell, J. & Kollman, P. (2001). Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: aqueous solution free energies of methanol and N-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases. *Journal of Computational Chemistry* **22**, 1048–1057.

Clarke, J., Hounslow, A. M., Bond, C. J., Fersht, A. R. & Daggett, V. (2000). The effects of disulfide bonds on the denatured state of barnase. *Protein Science* **9**, 2394–2404.

Corcho, F. J., Filizola, M. & Pérez, J. J. (2000). Evaluation of the iterative simulated annealing technique in conformational search of peptides. *Chemical Physics Letters* **319**, 65–70.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz Jr., K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. & Kollman, P. A. (1995). A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *Journal of the American Chemical Society* **117**, 5179–5197.

Cramer, C. J. & Truhlar, D. G. (1999). Implicit solvation models: equilibria, structure, spectra, and dynamics. *Chemistry Reviews* **99**, 2161–2200.

Crouzy, S., Baudry, J., Smith, J. C. & Roux, B. (1999). Efficient calculation of two-dimensional adiabatic and free energy maps: application to the isomerization of the C13═C14 and C15═C16 bonds in the retinal of bacteriorhodopsin. *Journal of Computational Chemistry* **20**, 1644–1658.

Cui, Q., Elstner, M., Kaxiras, E., Frauenheim, T. & Karplus, M. (2001). A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method. *Journal of Physical Chemistry* (*B*) **105**, 569–585.

Czerminski, R. & Elber, R. (1991). Computational studies of ligand diffusion in globins: I. Leghemoglobin. *Proteins: Structure, Function and Genetics* **10**, 70–80.

Daggett, V. (2000). Long timescale simulations. *Current Opinion in Structural Biology* **10**, 160–164.

Daggett, V. (2002). Molecular dynamics simulations of the protein unfolding/folding reaction. *Accounts of Chemical Research* **35**, 422–429.

Daggett, V. & Levitt, M. (1992). A model of the molten globule state from molecular dynamics simulations. *Proceedings of the National Academy of Sciences USA* **89**, 5142–5146.

Daggett, V. & Levitt, M. (1993). Protein unfolding pathways explored through molecular dynamics simulations. *Journal of Molecular Biology* **232**, 600–619.

Daggett, V., Li, A. & Fersht, A. R. (1998). Combined molecular dynamics and $\phi$-value analysis of structure-reactivity relationships in the transition state and unfolding pathway of barnase: structural basis of Hammond and anti-Hammond effects. *Journal of the American Chemical Society* **120**, 12740–12754.

Damm, W. & van Gunsteren, W. F. (2000). Reversible peptide folding: dependence on molecular force field used. *Journal of Computational Chemistry* **21**, 774–787.

Darden, T., Perera, L., Li, L. & Pedersen, L. (1999). New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. *Structure, Folding & Design* **7**, R55–R60.

Darden, T. A., Toukmaji, A. & Pedersen, L. G. (1997). Long-range electrostatic effects in biomolecular simulations. *Journal de Chimie Physique et de Physico-Chimie Biologique* **94**, 1346–1364.

Darden, T., York, D. & Pedersen, L. (1993). Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. *Journal of Chemical Physics* **98**, 10089–10092.

Daura, X., Jaun, B., Seebach, D., van Gunsteren, W. F. & Mark, A. E. (1998a). Reversible peptide folding in solution by molecular dynamics simulation. *Journal of Molecular Biology* **280**, 925–932.

Daura, X., Mark, A. E. & van Gunsteren, W. F. (1998b). Parametrization of aliphatic CHn united atoms of GROMOS96 force field. *Journal of Computational Chemistry* **19**, 535–547.

Daura, X., Mark, A. E. & van Gunsteren, W. F. (1999a). Peptide folding simulations: no solvent required? *Computer Physics Communications* **123**, 97–102.

Daura, X., van Gunsteren, W. F. & Mark, A. E. (1999b). Folding-unfolding thermodynamics of a $\beta$-heptapeptide from equilibrium simulations. *Proteins: Structure, Function and Genetics* **34**, 269–280.

Daura, X., van Gunsteren, W. F., Rigo, D., Jaun, B. & Seebach, D. (1997). Studying the stability of a helical $\beta$-heptapeptide by molecular dynamics simulations. *Chemistry: A European Journal* **3**, 1410–1417.

de Groot, B. L., van Aalten, D. M. F., Scheek, R. M., Amadei, A., Vriend, G. & Berendsen, H. J. C. (1997). Prediction of protein conformational freedom from distance constraints. *Proteins: Structure, Function and Genetics* **29**, 240–251.

de Groot, B. L. & Grubmüller, H. (2001). Water permeation across biological membranes: mechanism and dynamics of aquaporin-1 and GlpF. *Science* **294**, 2353–2357.

Deserno, M. & Holm, C. (1998a). How to mesh up Ewald sums. I. A theoretical and numerical comparison of various particle mesh routines. *Journal of Chemical Physics* **109**, 7678–7693.

Deserno, M. & Holm, C. (1998b). How to mesh up Ewald sums. II. An accurate error estimate for the particle-particle-particle-mesh algorithm. *Journal of Chemical Physics* **109**, 7694–7701.

Duan, Y. & Kollman, P. A. (1998). Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* **282**, 740–744.

Duan, Y., Wang, L. & Kollman, P. A. (1998). The early stage of folding of villin headpiece subdomain observed in a 200-nanosecond fully solvated molecular dynamics simulation. *Proceedings of the National Academy of Sciences USA* **95**, 9897–9902.

Duan, Z.-H. & Krasny, R. (2000). An Ewald summation based multipole method. *Journal of Chemical Physics* **113**, 3492–3495.

Egberts, E., Marrink, S. J. & Berendsen, H. J. C. (1994). Molecular dynamics simulation of a phospholipid membrane. *European Biophysics Journal* **22**, 423–436.

Egorov, S. A., Rabani, E. & Berne, B. J. (1999). On the adequacy of mixed quantum-classical dynamics in condensed phase systems. *Journal of Physical Chemistry B* **103**, 10978–10991.

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. & Pedersen, L. G. (1995). A smooth particle

mesh Ewald method. *Journal of Chemical Physics* **103**, 8577–8593.

Evans, D. J. & Morris, G. P. (1983). Isothermal isobaric molecular dynamics. *Chemical Physics* **77**, 63–66.

Evans, D. J. & Morris, G. P. (1984). Non-Newtonian molecular dynamics. *Computer Physics Reports* **1**, 297–344.

Evans, E. & Ritchie, K. (1997). Dynamic strength of molecular adhesion bonds. *Biophysical Journal* **72**, 1541–1555.

Ewald, P. P. (1921). Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik* **64**, 253–287.

Feenstra, K. A., Hess, B. & Berendsen, H. J. C. (1999). Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *Journal of Computational Chemistry* **20**, 786–798.

Feig, M., MacKerell Jr., A. D. & Brooks III, C. L. (2003). Force field influence on the observation of $\pi$-helical protein structures in molecular dynamics simulations. *Journal of Physical Chemistry* (*B*) **107**, 2831–2836.

Feller, S. E. & MacKerell Jr., A. D. (2000). An improved empirical potential energy function for molecular simulations of phospholipids. *Journal of Physical Chemistry* (*B*) **104**, 7510–7515.

Feller, S. E., Yin, D., Pastor, R. W. & MacKerell Jr., A. D. (1997). Molecular dynamics simulation of unsaturated lipid bilayers at low hydration: parameterization and comparison with diffraction studies. *Biophysical Journal* **73**, 2269–2279.

Feller, S. E., Zhang, Y., Pastor, R. W. & Brooks, B. R. (1995). Constant pressure molecular dynamics simulation: the Langevin piston method. *Journal of Chemical Physics* **103**, 4613–4621.

Ferrara, P., Apostolakis, J. & Caflisch, A. (2000). Targeted molecular dynamics simulations of protein unfolding. *Journal of Physical Chemistry* (*B*) **104**, 4511–4518.

Ferrara, P. & Caflisch, A. (2000). Folding simulations of a three-stranded antiparallel $\beta$-sheet peptide. *Proceedings of the National Academy of Sciences USA* **97**, 10780–10785.

Ferrenberg, A. M. & Swendsen, R. H. (1989). Optimized Monte Carlo data analysis. *Physical Review Letters* **63**, 1195–1198.

Fersht, A. R. (1995). Characterizing transition states in protein folding: an essential step in the puzzle. *Current Opinion in Structural Biology* **5**, 79–84.

Field, M. J., Albe, M., Bret, C., Proust-De Martin, F. & Thomas, A. (2000). The Dynamo library for molecular simulations using hybrid quantum mechanical and molecular mechanical potentials. *Journal of Computational Chemistry* **21**, 1088–1100.

Field, M. J., Bash, P. A. & Karplus, M. (1990). A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *Journal of Computational Chemistry* **11**, 700–733.

Figueirido, F., Levy, R. M., Zhou, R. H. & Berne, B. J. (1997). Large scale simulation of macromolecules in solution: combining the periodic fast multipole method with multiple time step integrators. *Journal of Chemical Physics* **106**, 9835–9849. [Correction (1997), **107**, 7002.]

Floriano, W. B., Nascimento, M. A. C., Domont, G. B. & Goddard III, W. A. (1998). Effects of pressure on the structure of metmyoglobin: molecular dynamics predictions for pressure unfolding through a molten globule intermediate. *Protein Science* **7**, 2301–2313.

Foloppe, N. & MacKerell Jr., A. D. (2000). All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *Journal of Computational Chemistry* **21**, 86–104.

Forester, T. R. & Smith, W. (1998). SHAKE, rattle, and roll: efficient constraint algorithms for linked rigid bodies. *Journal of Computational Chemistry* **19**, 102–111. [Erratum (2000), **21**, 157.]

Forrest, D. K. & Sansom, M. S. P. (2000). Membrane simulations: bigger and better? *Current Opinion in Structural Biology* **10**, 174–181.

Fulton, K. F., Main, E. R. G., Daggett, V. & Jackson, S. E. (1999). Mapping the interactions present in the transition state for unfolding/folding of FKBP12. *Journal of Molecular Biology* **291**, 445–461.

Gao, J. L., Amara, P., Alhambra, C. & Field, M. J. (1998). A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations. *Journal of Physical Chemistry* (*A*) **102**, 4714–4721.

García-Archilla, B., Sanz-Serna, J. M. & Skeel, R. D. (1998). In *Numerical Analysis 1997* (eds. D. F. Griffiths & G. A. Watson), pp. 318–331. London: Pitman.

Gilquin, B., Guilbert, C. & Perahia, D. (2000). Unfolding of hen egg lysozyme by molecular dynamics simulations at 300 K: insight into the role of the interdomain interface. *Proteins: Structure, Function and Genetics* **41**, 58–74.

Gilson, M. K., Given, J. A., Bush, B. L. & McCammon, J. A. (1997). The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophysical Journal* **72**, 1047–1069.

Greengard, L. & Rokhlin, V. (1987). A fast algorithm for particle simulations. *Journal of Computational Physics* **73**, 325–348.

Gresh, N., Guo, H., Salahub, D. R., Roques, B. P. & Kafafi, S. A. (1999). Critical role of anisotropy for the dimerization energies of two protein-protein recognition motifs: *cis-N*-methylacetamide versus a beta-sheet conformer of alanine dipeptide. A joint *ab initio*, density functional theory, and molecular mechanics investigation. *Journal of the American Chemical Society* **121**, 7885–7894.

Grootenhuis, P. D. J. & Haasnoot, C. A. G. (1993). A CHARMM based force-field for carbohydrates using

the CHEAT approach – carbohydrate hydroxyl-groups represented by extended atoms. *Molecular Simulation* **10**, 75–95.

Grubmüller, H. (1995). Predicting slow structural transitions in macromolecular systems – conformational flooding. *Physical Review* (*E*) **52**, 2893–2906.

Grubmüller, H. & Tavan, P. (1998). Multiple time step algorithms for molecular dynamics simulations of proteins: how good are they? *Journal of Computational Chemistry* **19**, 1534–1552.

Guenot, J. & Kollman, P. A. (1992). Molecular dynamics studies of a DNA-binding protein: 2. An evaluation of implicit and explicit solvent models for the molecular dynamics simulation of the *Escherichia coli trp* repressor. *Protein Science* **1**, 1185–1205.

Gullingsrud, J., Braun, R. & Schulten, K. (1999). Reconstructing potentials of mean force through time series analysis of steered molecular dynamics simulations. *Journal of Computational Physics* **151**, 190–212.

Guo, Z. Y., Brooks III, C. L. & Boczko, E. M. (1997). Exploring the folding free energy surface of a three-helix bundle protein. *Proceedings of the National Academy of Sciences USA* **94**, 10161–10166.

Halgren, T. A. (1999). MMFF VII. Characterization of MMFF94, MMFF94s and other widely available force fields for conformational energies and for intermolecular-interaction energies and geometries. *Journal of Computational Chemistry* **20**, 730–748.

Halgren, T. A. & Damm, W. (2001). Polarizable force fields. *Current Opinion in Structural Biology* **11**, 236–242.

Hart, R. K., Pappu, R. V. & Ponder, J. W. (2000). Exploring the similarities between potential smoothing and simulated annealing. *Journal of Computational Chemistry* **21**, 531–552.

Harvey, S. C. (1989). Treatment of electrostatic effects in macromolecular modeling. *Proteins: Structure, Function and Genetics* **5**, 78–92.

Henchman, R. H. & Essex, J. W. (1999a). Free energy of hydration using restrained electrostatic potential derived charges via free energy perturbations and linear response. *Journal of Computational Chemistry* **20**, 499–510.

Henchman, R. H. & Essex, J. W. (1999b). Generation of OPLS-like charges from molecular electrostatic potential using restraints. *Journal of Computational Chemistry* **20**, 483–498.

Hendrix, D. A. & Jarzynski, C. (2001). A 'fast-growth' method of computing free energy differences. *Journal of Chemical Physics* **114**, 5974–5981.

Hernández, B., Luque, F. J. & Orozco, M. (2000). Mixed QM/MM molecular electrostatic potential. *Journal of Computer-Aided Molecular Design* **14**, 329–339.

Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. (1997). LINCS: a linear constraint solver for molecular interactions. *Journal of Computational Chemistry* **18**, 1463–1472.

Hockney, R. W. & Eastwood, J. W. (1981). *Computer Simulation Using Particles*. New York: McGraw-Hill.

Hoover, W. G. (1985). Canonical dynamics: equilibrium phase-space distributions. *Physical Reviews* (*A*) **31**, 1695–1697.

Hummer, G. (2001). Fast-growth thermodynamic integration: error and efficiency analysis. *Journal of Chemical Physics* **114**, 7330–7337.

Hünenberger, P. H. (2000). Optimal charge-shaping functions for the particle-particle-paticle-mesh ($P^3M$) method for computing electrostatic interactions in molecular simulations. *Journal of Chemical Physics* **113**, 10464–10476.

Hünenberger, P. H. & McCammon, J. A. (1999). Ewald artifacts in computer simulations of ionic solvation and ion-ion interaction: a continuum electrostatics study. *Journal of Chemical Physics* **110**, 1856–1872.

Hünenberger, P. H., Mark, A. E. & van Gunsteren, W. F. (1995). Computational approaches to study protein unfolding: hen egg white lysozyme as a case study. *Proteins: Structure, Function and Genetics* **21**, 196–213.

Im, W., Bernèche, S. & Roux, B. (2001). Generalized solvent boundary potential for computer simulations. *Journal of Chemical Physics* **114**, 2924–2937.

Isralewitz, B., Gao, M. & Schulten, K. (2001). Steered molecular dynamics and mechanical functions of proteins. *Current Opinion in Structural Biology* **11**, 224–230.

Izaguirre, J. A., Reich, S. & Skeel, R. D. (1999). Longer time steps for molecular dynamics. *Journal of Chemical Physics* **110**, 9853–9864.

Jarzynski, C. (1997). Nonequilibrium equality for free energy differences. *Physical Reviews Letters* **78**, 2690–2693.

Ji, J., Çağin, T. & Pettitt, B. M. (1992). Dynamic simulations of water at constant chemical-potential. *Journal of Chemical Physics* **96**, 1333–1342.

Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society* **118**, 11225–11236.

Juffer, A. H. & Berendsen, H. J. C. (1993). Dynamic surface boundary conditions. A simple boundary model for molecular dynamics simulations. *Molecular Physics* **79**, 623–644.

Kaminski, G. A., Friesner, R. A., Tirado-Rives, J. & Jorgensen, W. L. (2001). Evaluation and reparametrization of the OPLS-AA force field for proteins via coparison with accurte quantum chemical calculations on peptides. *Journal of Physical Chemistry* (*B*) **105**, 6474–6487.

Kamiya, N. & Higo, J. (2001). Repeated-annealing sampling combined with multicanonical algorithm for conformational sampling of bio-molecules. *Journal of Computational Chemistry* **22**, 1098–1106.

KARPLUS, M. (2000). Aspects of protein reaction dynamics: deviations from simple behavior. *Journal of Physical Chemistry* (*B*) **104**, 11–27.

KARPLUS, M. (2003). Molecular dynamics of biological macromolecules: a brief history and perspective. *Biopolymers* **68**, 350–358.

KARPLUS, M. & MCCAMMON, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nature Structural Biology* **9**, 646–652.

KARPLUS, M. & PETSKO, G. A. (1990). Molecular-dynamics simulations in biology. *Nature* **347**, 631–639.

KAWATA, M. & MIKAMI, M. (1999). Acceleration of the canonical molecular dynamics by the particle mesh Ewald method combined with the multiple time-step integrator algorithm. *Chemical Physics Letters* **313**, 261–266. [Corrigendum (2000), **317**, 515.]

KAWATA, M. & MIKAMI, M. (2000). Computationally efficient canonical molecular dynamics simulations by using a multiple time-step integrator algorithm combined with the particle mesh Ewald method and with the fast multipole method. *Journal of Computational Chemistry* **21**, 201–217.

KAWATA, M. & NAGASHIMA, U. (2001). Particle mesh Ewald method for three-dimensional systems with two-dimensional periodicity. *Chemical Physics Letters* **340**, 165–172.

KAZMIRSKI, S. & DAGGETT, V. (1998a). The structural and dynamic properties of unfolded proteins: simulations of the 'molten coil' state of BPTI. *Journal of Molecular Biology* **277**, 487–506.

KAZMIRSKI, S. & DAGGETT, V. (1998b). Non-native interactions in protein folding intermediates: molecular dynamics simulation of hen lysozyme. *Journal of Molecular Biology* **284**, 793–806.

KAZMIRSKI, S. L., LI, A. & DAGGETT, V. (1999). Analysis methods for comparison of nuclear dynamics trajectories: applications to protein unfolding pathways and denatured ensembles. *Journal of Molecular Biology* **290**, 283–304.

KIMURA, S. R., BROWER, R. C., ZHANG, C. & SUGIMORI, M. (2000). Surface of active polarons: a semiexplicit solvation method for biomolecular dynamics. *Journal of Chemical Physics* **112**, 7723–7734.

KING, G. & BARFORD, R. A. (1993). Calculation of electrostatic free-energy differences with a time-saving approximate method. *Journal of Physical Chemistry* **97**, 8798–8802.

KLAMT, A. & SCHÜÜRMANN, G. (1993). Cosmo – a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *Journal of the Chemical Society – Perkin Transactions* (*2*) **5**, 799–805.

KLEINJUNG, J., BAYLEY, P. & FRATERNALI, F. (2000). Leap-dynamics: efficient sampling of conformational space of proteins and peptides in solution. *FEBS Letters* **470**, 257–262.

KNELLER, G. R. & MÜLDERS, T. (1996). Nosé–Andersen dynamics of partially rigid molecules: coupling all degrees of freedom to heat and pressure baths. *Physical Reviews* (*E*) **54**, 6825–6837.

KOLB, A. & DÜNWEG, B. (1999). Optimized constant pressure stochastic dynamics. *Journal of Chemical Physics* **111**, 4453–4459.

KOLLMAN, P. A. (1993). Free energy calculations: applications to chemical and biochemical phenomena. *Chemical Reviews* **93**, 2395–2417.

KOLLMAN, P. A. (1996). Advances and continuing challenges in achieving realistic and predictive simulations of the properties of organic and biological molecules. *Accounts of Chemical Research* **29**, 461–469.

KOLLMAN, P. A., KUHN, B., DONINI, O., PERÄKYLÄ, M., STANTON, R. & BAKOWIES, D. (2001). Elucidating the nature of enzyme catalysis utilizing a new twist on an old methodology: quantum mechanical-free energy calculations on chemical reactions in enzymes and aqueous solution. *Accounts of Chemical Research* **34**, 72–79.

KOLLMAN, P. A., MASSOVA, I., REYES, C., KUHN, B., HUO, S., CHONG, L., LEE, M., LEE, T., DUAN, Y., WANG, W., DONINI, O., CIEPLAK, P., SRINIVASAN, J., CASE, D. A. & CHEATHAM III, T. E. (2000). Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Accounts of Chemical Research* **33**, 889–897.

KONG, X. J. & BROOKS III, C. L. (1996). $\lambda$-dynamics: a new approach to free energy calculations. *Journal of Chemical Physics* **105**, 2414–2423.

KOUWIJZER, M. L. C. E. & GROOTENHUIS, P. D. J. (1995). Parametrization and application of CHEAT95, an extended atom force-field for hydrated (oligo)saccharides. *Journal of Physical Chemistry* **99**, 13426–13436.

KRATKY, K. W. (1980). New boundary condition for computer experiments of thermodynamic systems. *Journal of Computational Physics* **37**, 205–217.

KRÄUTLER, V., VAN GUNSTEREN, W. F. & HÜNENBERGER, P. H. (2001). A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *Journal of Computational Chemistry* **22**, 501–508.

KUMAR, S., BOUZIDA, D., SWENDSEN, R. H., KOLLMAN, P. A. & ROSENBERG, J. M. (1992). The weighted histogram analysis method for free-energy calculations on biomolecules. 1. The method. *Journal of Computational Chemistry* **13**, 1011–1021.

KUTTEH, R. (1999). New methods for incorporating nonholonomic constraints into molecular dynamics simulations. *Journal of Chemical Physics* **111**, 1394–1409.

LABERGE, L. J. & TULLY, J. C. (2000). A rigorous procedure for combining molecular dynamics and Monte Carlo simulation algorithms. *Chemical Physics* **260**, 183–191.

LADURNER, A. G., ITZHAKI, L. S., DAGGETT, V. & FERSHT, A. R. (1998). Synergy between simulation and experiment in describing the energy landscape of protein

folding. *Proceedings of the National Academy of Sciences USA* **95**, 8473–8478.

Lahiri, A., Nilsson, L. & Laaksonen, A. (2001). Exploring the Idea of self-guided dynamics. *Journal of Chemical Physics* **114**, 5993–5999.

Lamb, M. L. & Jorgensen, W. L. (1997). Computational approaches to molecular recognition. *Current Opinion in Chemical Biology* **1**, 449–457.

Langley, D. R. (1998). Molecular dynamics simulations of environment and sequence dependent DNA conformations: the development of the BMS nucleic acid force field and comparison with experimental results. *Journal of Biomolecular and Structural Dynamics* **16**, 487–509.

Lavery, R., Zakrzewska, K. & Sklenar, H. (1995). JUMNA (JUnction Minimization of Nucleic-Acids). *Computer Physics Communications* **91**, 135–158.

Lazaridis, T. & Karplus, M. (1997). New view of protein folding reconciled with the old through multiple unfolding simulations. *Science* **278**, 1928–1931.

Lazaridis, T. & Karplus, M. (1999). Effective energy function for proteins in solution. *Proteins: Structure, Function and Genetics* **35**, 133–152.

Lazaridis, T. & Karplus, M. (2000). Effective energy functions for protein structure prediction. *Current Opinion in Structural Biology* **10**, 139–145.

Lazaridis, T. & Karplus, M. (2003). Thermodynamics of protein folding: a microscopic view. *Biophysical Chemistry* **100**, 367–395.

Ledauphin, V. & Vergoten, G. (2000). New nonbonded interactions calculation strategy for rectangular systems. I. Preliminary molecular dynamics study of solvated Na ion. *Biopolymers* **57**, 373–382.

Lee, J., Pillardy, J., Czaplewski, C., Arnautova, Y., Ripoll, D. R., Liwo, A., Gibson, K. D., Wawak, R. J. & Scheraga, H. A. (2000a). Efficient parallel algorithms in global optimization of potential energy functions for peptides, proteins, and crystals. *Computer Physics Communications* **128**, 399–411.

Lee, M. R., Duan, Y. & Kollman, P. A. (2000b). Use of MM/PB-SA in estimating the free energies of proteins: application to native, intermediates, and unfolded villin headpiece. *Proteins: Structure, Function and Genetics* **39**, 309–316.

Lee, S.-H., Palmo, K. & Krimm, S. (2001). A new formalism for molecular dynamics in internal coordinates. *Chemical Physics* **265**, 63–85.

Leimkuhler, B. & Reich, S. (2001). A reversible averaging integrator for multiple time-scale dynamics. *Journal of Computational Physics* **171**, 95–114.

Levitt, M. (1983). Computer simulation of DNA double-helix dynamics. *Cold Spring Harbor Symposium on Quantative Biology* **47**, 251–262.

Levitt, M. (2001). The birth of computational structural biology. *Nature Structural Biology* **8**, 392–393.

Levitt, M., Hirshberg, M., Sharon, R. & Daggett, V. (1995). Potential energy function and parameters for simulations of the molecular dynamics of proteins and nucleic acids in solution. *Computer Physics Communications* **91**, 215–231.

Lévy, B. & Enescu, M. (1998). Theoretical study of methylene blue: a new method to determine partial atomic charges; investigation of the interaction with guanine. *Journal of Molecular Structure: Theochem* **432**, 235–245.

Levy, R. M. & Gallicchio, E. (1998). Computer simulations with explicit solvent: recent progress in the thermodynamic decomposition of free energies and in modeling electrostatic effects. *Annual Review of Physical Chemistry* **49**, 531–567.

Li, A. & Daggett, V. (1994). Charaterization of the transition state of protein unfolding: chymotrypsin inhibitor 2. *Proceedings of the National Academy of Sciences USA* **91**, 10430–10434.

Li, A. & Daggett, V. (1996). Identification and characterization of the unfolding transition state of chymotrypsin inhibitor 2 by molecular dynamics simulations. *Journal of Molecular Biology* **257**, 412–429.

Li, A. & Daggett, V. (1998). The unfolding of barnase: charaterization of the major intermediate. *Journal of Molecular Biology* **275**, 677–694.

Liu, H., Mark, A. E. & van Gunsteren, W. F. (1996). Estimating the relative free energy of different states with respect to a single reference state. *Journal of Physical Chemistry* **100**, 9485–9494.

Lu, H., Isralewitz, B., Krammer, A., Vogel, V. & Schulten, K. (1998). Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophysical Journal* **75**, 662–671.

Lu, H., Krammer, A., Isralewitz, B., Vogel, V. & Schulten, K. (2000). Computer modeling of force-induced titin domain unfolding. *Advances in Experimental and Medical Biology* **481**, 143–160.

Lu, H. & Schulten, K. (1999a). Steered molecular dynamics simulations of conformational changes of immunoglobulin domain I27 interpret atomic force microscopy observations. *Chemical Physics* **247**, 141–153.

Lu, H. & Schulten, K. (1999b). Steered molecular dynamics simulations of force-induced protein domain unfolding. *Proteins: Structure, Function and Genetics* **35**, 453–463.

Lu, H. & Schulten, K. (2000). The key event in force-induced unfolding of titin's immunoglobulin domains. *Biophysical Journal* **79**, 51–65.

Lu, N. & Kofke, D. A. (2001). Accuracy of free-energy perturbation calculations in molecular simulation. I. Modeling. *Journal of Chemical Physics* **114**, 7303–7311.

Lüdemann, S. K., Lounnas, V. & Wade, R. C. (2000). How do substrates enter and products exit the buried active site of cytochrome P450cam? 1. Random expulsion molecular dynamics investigation of ligand access

channels and mechanisms. *Journal of Molecular Biology* **303**, 797–811.

MA, J. & KARPLUS, M. (1997). Molecular switch in signal transduction: reaction paths of the conformational changes in ras p21. *Proceedings of the National Academy of Sciences USA* **94**, 11905–11910.

MA, J., SIGLER, P. B., XU, Z. & KARPLUS, M. (2000). A dynamic model for the allosteric mechanism of GroEL. *Journal of Molecular Biology* **302**, 303–313.

MACKERELL JR., A. D. & BANAVALI, N. K. (2000). All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *Journal of Computational Chemistry* **21**, 105–120.

MACKERELL JR., A. D., BASHFORD, D., BELLOTT, M., DUNBRACK JR., R. L., EVANSECK, J. D., FIELD, M. J., FISCHER, S., GAO, J., GUO, H., HA, S., JOSEPH-MCCARTHY, D., KUCHNIR, L., KUCZERA, K., LAU, F. T. K., MATTOS, C., MICHNICK, S., NGO, T., NGUYEN, D. T., PRODHOM, B., REIHER III, W. E., ROUX, B., SCHLENKRICH, M., SMITH, J. C., STOTE, R., STRAUB, J., WATANABE, M., WIÓRKIEWICZ-KUCZERA, J., YIN, D. & KARPLUS, M. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry* (*B*) **102**, 3586–3616.

MACKERELL JR., A. D., WIÓRKIEWICZ-KUCZERA, J. & KARPLUS, M. (1995). An all-atom empirical energy function for the simulation of nucleic acids. *Journal of the American Chemical Society* **117**, 11946–11975.

MALEVANETS, A. & KAPRAL, R. (2000). Solute molecular dynamics in a mesoscale solvent. *Journal of Chemical Physics* **112**, 7260–7269.

MANNFORS, B., PALMO, K. & KRIMM, S. (2000). A new electrostatic model for molecular mechanics force fields. *Journal of Molecular Structure* **556**, 1–21.

MARCHI, M., BORGIS, D. & LEVY, N. (2001). A dielectric continuum molecular dynamics method. *Journal of Chemical Physics* **114**, 4377–4385.

MARCHI, M. & PROCACCI, P. (1998). Coordinates scaling and multiple time step algorithms for simulation of solvated proteins in the NPT ensemble. *Journal of Chemical Physics* **109**, 5194–5202.

MARK, A. E. & VAN GUNSTEREN, W. F. (1992). Simulation of the thermal denaturation of hen egg white lysozyme: trapping the molten globule state. *Biochemistry* **31**, 7745–7748.

MARSZALEK, P. E., LU, H., LI, H., CARRION-VAZQUEZ, M., OBERHAUSER, A. F., SCHULTEN, K. & FERNANDEZ, J. M. (1999). Mechanical unfolding intermediates in titin modules. *Nature* **402**, 100–103.

MARTÍ-RENOM, M. A., STOTE, R. H., QUEROL, E., AVILÉS, F. X. & KARPLUS, M. (1998). Refolding of potato carboxypeptidase inhibitor by molecular dynamics simulations with disulfide bond constraints. *Journal of Molecular Biology* **284**, 145–172.

MARTYNA, G. J., KLEIN, M. L. & TUCKERMAN, M. (1992). Nosé–Hoover chains: the canonical ensemble via continuous dynamics. *Journal of Chemical Physics* **97**, 2635–2643.

MARTYNA, G. J., TOBIAS, D. J. & KLEIN, M. L. (1994). Constant pressure molecular dynamics algorithms. *Journal of Chemical Physics* **101**, 4177–4189.

MAYOR, U., JOHNSON, C. M., DAGGETT, V. & FERSHT, A. R. (2000). Protein folding and unfolding in microseconds to nanoseconds by experiments and simulations. *Proceedings of the National Academy of Sciences USA* **97**, 13518–13522.

MCCAMMON, J. A. (1977). Molecular dynamics study of the bovine pancreatic trypsin inhibitor. In *Models for Protein Dynamics 1976* (ed. H. J. C. Berendsen), pp. 137–152. Orsay, France: CECAM, Université de Paris IX.

MCCAMMON, J. A., GELIN, B. R. & KARPLUS, M. (1977). Dynamics of folded proteins. *Nature* **267**, 585–590.

MCCAMMON, J. A. & HARVEY, S. C. (1987). *Dynamics of Proteins and Nucleic Acids*. Cambridge: Cambridge University Press.

MELCHIONNA, S., CICCOTTI, G. & HOLIAN, B. L. (1993). Hoover NPT dynamics for systems varying in shape and size. *Molecular Physics* **78**, 533–544.

MERTZ, J. E., TOBIAS, D. J., BROOKS, C. L. & SINGH, U. C. (1991). Vector and parallel algorithms for the molecular-dynamics simulation of macromolecules on shared memory computers. *Journal of Computational Chemistry* **12**, 1270–1277.

MEZEI, M. (2000). Comment on 'Molecular dynamics simulations in the grand canonical ensemble: formulation of a bias potential for umbrella sampling'. *Journal of Chemical Physics* **112**, 1059–1060.

MIRANKER, A. & KARPLUS, M. (1991). Functionality maps of binding sites: a multiple copy simultaneous search method. *Proteins: Structure, Function and Genetics* **11**, 29–34.

MIYAMOTO, S. & KOLLMAN, P. A. (1992). SETTLE – an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry* **13**, 952–962.

MÖLLHOFF, M. & STERNBERG, U. (2001). Molecular mechanics with fluctuating atomic charges – a new force field with a semi-empirical charge calculation. *Journal of Molecular Modelling* **7**, 90–102.

MONARD, G. & MERZ JR., K. M. (1999). Combined quantum mechanical/molecular mechanical methodologies applied to biomolecular systems. *Accounts of Chemical Research* **32**, 904–911.

MORDASINI, T. Z. & MCCAMMON, J. A. (2000). Calculations of relative hydration free energies: a comparative study using thermodynamic integration and an extrapolation method based on a single reference state. *Journal of Physical Chemistry* (*B*) **104**, 360–367.

Morishita, T. (2000). Fluctuation formulas in molecular-dynamics simulations with the weak coupling heat bath. *Journal of Chemical Physics* **113**, 2976–2982.

Nelson, M., Humphrey, A., Gursoy, A., Dalke, A., Kale, R. D., Skeel, R. D. & Schulten, K. (1996). NAMD – a parallel, object-oriented molecular dynamics program. *International Journal of Supercomputing Applications and High Performance Computing* **10**, 251–268.

Némethy, G., Gibson, K. D., Palmer, K. A., Yoon, C. N., Paterlini, G., Zagari, A., Rumsey, S. & Sheraga, H. A. (1992). Energy parameters in polypeptides. 10. Improved geometrical parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides. *Journal of Physical Chemistry* **96**, 6472–6484.

Norberg, J. & Nilsson, L. (1995a). A conformational free energy landscape of ApApA from molecular dynamics simulations. *Journal of Physical Chemistry* **100**, 2550–2554.

Norberg, J. & Nilsson, L. (1995b). NMR relaxation times, dynamics, and hydration of a nucleic acid fragment from molecular dynamics simulations. *Journal of Physical Chemistry* **99**, 14876–14884.

Norberg, J. & Nilsson, L. (1996). Glass transition in DNA from molecular dynamics simulations. *Proceedings of the National Academy of Sciences USA* **93**, 10173–10176.

Norberg, J. & Nilsson, L. (2000). On the truncation of long-range electrostatic interactions in DNA. *Biophysical Journal* **79**, 1537–1553.

Norberg, J. & Nilsson, L. (2002). Molecular dynamics applied to nucleic acids. *Accounts of Chemical Research* **35**, 465–472.

Nosé, S. & Klein, M. L. (1983). Constant pressure molecular dynamics for molecular systems. *Molecular Physics* **50**, 1055–1076.

Okuyama-Yoshida, N., Nagaoka, M. & Yamabe, T. (1998). Transition-state optimization on free energy surface: toward solution chemical reaction ergodography. *International Journal of Quantum Chemistry* **70**, 95–103.

Okuyama-Yoshida, N., Kataoka, K., Nagaoka, M. & Yamabe, T. (2000). Structure optimization via free energy gradient method: application to glycine zwitterion in aqueous solution. *Journal of Chemical Physics* **113**, 3519–3524.

Oliva, B., Daura, X., Querol, E. & Avilés, F. X. (2000). A generalized Langevin dynamics approach to model solvent dynamics effects on proteins via a solvent-accessible surface. The carboxypeptidase A inhibitor protein as a model. *Theoretical Chemical Accounts* **105**, 101–109.

Ono, S., Nakajima, N., Higo, J. & Nakamura, H. (1999). The multicanonical weighted histogram analysis method for the free-energy landscape along structural transition paths. *Chemical Physics Letters* **12**, 247–254.

Ono, S., Nakajima, N., Higo, J. & Nakamura, H. (2000). Peptide free-energy profile is strongly dependent on the force field: comparison of C96 and AMBER95. *Journal of Computational Chemistry* **21**, 748–762.

Onuchic, J. N., Luthey-Schulten, Z. & Wolynes, P. G. (1997). Theory of protein folding: the energy landscape perspective. *Annuual Review of Physical Chemistry* **48**, 545–600.

Orozco, M. & Luque, F. J. (2000). Theoretical methods for the description of the solvent effect in biomolecular systems. *Chemical Reviews* **100**, 4187–4225. [Corrections (2001). **101**, 203.]

Ota, N. & Brünger, A. T. (1997). Overcoming barriers in macromolecular simulations: non-Boltzmann thermodynamic integration. *Theoretical Chemical Accounts* **98**, 171–181.

Ota, N., Stroupe, C., Ferreira-da-Silva, J. M. S., Shah, S. A., Mares-Guia, M. & Brünger, A. T. (1999). Non-Boltzmann thermodynamic integration (NBTI) for macromolecular systems: relative free energy of binding of trypsin to benzamidine and benzylamine. *Proteins: Structure, Function and Genetics* **37**, 641–653.

Ott, K.-H. & Meyer, B. (1996). Parametrization of GROMOS force field for oligosaccharides and assessment of efficiency of molecular dynamics simulations. *Journal of Computational Chemistry* **17**, 1068–1084.

Paci, E. & Karplus, M. (1999). Forced unfolding of Fibronectin type 3 modules: an analysis by biased molecular dynamics simulations. *Journal of Molecular Biology* **288**, 441–459.

Paci, E. & Karplus, M. (2000). Unfolding proteins by external forces and temperature: the importance of topology and energetics. *Proceedings of the National Academy of Sciences USA* **97**, 6521–6526.

Pappu, R. V., Hart, R. K. & Ponder, J. W. (1998). Analysis and application of potential energy smoothing and search methods for global optimization. *Journal of Physical Chemistry* (*B*) **102**, 9725–9742.

Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham III, T. E., DeBolt, S., Ferguson, D., Seibel, G. & Kollman, P. (1995). AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications* **91**, 1–41.

Pérez, S., Imberty, A., Engelsen, S. B., Gruza, J., Mazeau, K., Jimenez-Barbero, J., Poveda, A., Espinosa, J.-F., van Eyck, B. P., Johnson, G., French, A. D., Kouwijzer, M. L. C. E., Grootenuis, P. D. J., Bernardi, A., Raimondi, L., Senderowitz, H., Durier, V., Vergoten, G. & Rasmussen, K. (1998). A comparison and chemometric analysis of several molecular mechanics force fields and parameter sets applied to carbohydrates. *Carbohydrate Research* **314**, 141–155.

Pettitt, B. M., Makarov, V. A. & Andrews, B. K. (1998). Protein hydration density: theory, simulations and crystallography. *Current Opinion in Structural Biology* **8**, 218–221.

Phelps, D. K., Speelman, B. & Post, C. B. (2000). Theoretical studies of viral capsid proteins. *Current Opinion in Structural Biology* **10**, 170–173.

Phillips, S. C., Essex, J. W. & Edge, C. M. (2000). Digitally filtered molecular dynamics: the frequency specific control of molecular dynamics simulations. *Journal of Chemical Physics* **112**, 2586–2597.

Pomès, R., Eisenmesser, E., Post, C. B. & Roux, B. (1999). Calculating excess chemical potentials using dynamic simulations in the fourth dimension. *Journal of Chemical Physics* **111**, 3387–3395.

Pollock, E. L. & Glosli, J. (1996). Comments on $P^3M$, FMM, and the Ewald method for large periodic Coulombic systems. *Computer Physics Communications* **95**, 93–110.

Procacci, P. & Berne, B. J. (1994). Multiple time scale methods for constant-pressure molecular dynamics simulations of molecular systems. *Molecular Physics* **83**, 255–272.

Radmer, R. J. & Kollman, P. A. (1997). Free energy calculation methods: a theoretical and empirical comparison of numerical errors and a new method for qualitative estimates of free energy changes. *Journal of Computational Chemistry* **18**, 902–919.

Rahman, A. (1964). Correlations in the motion of atoms in liquid argon. *Physical Reviews* **136A**, 405–411.

Rahman, A. & Stillinger, F. H. (1971). Molecular dynamics study of liquid water. *Journal of Chemical Physics* **55**, 3336–3359.

Reuter, N., Dejaegere, A., Maigret, B. & Karplus, M. (2000). Frontier bonds in QM/MM methods: a comparison of different approaches. *Journal of Physical Chemistry* (*A*) **104**, 1720–1735.

Rick, S. W., Stuart, S. J. & Berne, B. J. (1994). Dynamical fluctuating charge force-fields – application to liquid water. *Journal of Chemical Physics* **101**, 6141–6156.

Ridard, J. & Lévy, B. (1999). Effective atomic charges in alanine dipeptide. *Journal of Computational Chemistry* **20**, 473–482.

Rizzo, R. C. & Jorgensen, W. L. (1999). OPLS all-atom model for amines: resolution of the amine hydration problem. *Journal of the American Chemical Society* **121**, 4827–4836.

Roux, B. & Simonson, T. (1999). Implicit solvent models. *Biophysical Chemistry* **78**, 1–20.

Rozanska, X. & Chipot, C. (2000). Modeling ion–ion interaction in proteins: a molecular dynamics free energy calculation of the guanidinium-acetate association. *Journal of Chemical Physics* **112**, 9691–9694.

Ryckaert, J. P., Arialdi, G. & Melchionna, S. (2001). Molecular dynamics of polymers with explicit but frozen hydrogens. *Molecular Physics* **99**, 155–165.

Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. (1977). Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **23**, 327–341.

Sagui, C. & Darden, T. A. (1999). Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Annual Review of Biophysical and Biomolecular Structures* **28**, 155–179.

Sakane, S., Yezdimer, E. M., Liu, W., Barriocanal, J. A., Doren, D. J. & Wood, R. H. (2000). Exploring the *ab initio*/classical free energy perturbation method: the hydration free energy of water. *Journal of Chemical Physics* **113**, 2583–2593.

Sankararamakrishnan, R., Konvicka, K., Mehler, E. L. & Weinstein, H. (2000). Solvation in simulated annealing and high-temperature molecular dynamics of proteins: a restrained water droplet model. *International Journal of Quantum Chemistry* **77**, 174–186.

Schäfer, H., van Gunsteren, W. F. & Mark, A. E. (1999). Estimating relative free energies from a single ensemble: hydration free energies. *Journal of Computational Chemistry* **20**, 1604–1617.

Schlenkrich, M., Brinkmann, J., MacKerell Jr., A. D. & Karplus, M. (1996). An empirical potential energy function for phospholipids: criteria for parameter optimization and applications. In *Biological Membranes: A Molecular Perspective from Computation and Experiment* (eds. K. M. Merz & B. Roux), pp. 31–81. Boston: Birkhäuser.

Schlick, T. (1999). Computational molecular biophysics today: a confluence of methodological advances and complex biomolecular applications. *Journal of Computational Physics* **151**, 1–8.

Schlick, T., Barth, E. & Mandziuk, M. (1997). Biomolecular dynamics at long timesteps: bridging the time scale gap between simulation and experimentation. *Annual Review of Biophysical and Biomolecular Structures* **26**, 181–222.

Schlick, T., Skeel, R. D., Brunger, A. T., Kalé, L. V., Board Jr., J. A., Hermans, J. & Schulten, K. (1999). Algorithmic challenges in computational molecular biophysics. *Journal of Computational Physics* **151**, 9–48.

Schlitter, J., Engels, M., Kruger, P., Jacoby, E. & Wollmer, A. (1993). Targeted molecular dynamics simulation of conformational change – application to the T→R transition in insulin. *Molecular Simulation* **10**, 291–308.

Schuler, L. D., Daura, X. & van Gunsteren, W. F. (2001). An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *Journal of Computational Chemistry* **22**, 1205–1218.

Schymkowitz, J. W. H., Rousseau, F., Irvine, L. R. & Itzhaki, L. S. (2000). The folding pathway of the cell-cycle regulatory protein p13suc1: clues for the

mechanism of domain swapping. *Structure, Folding & Design* **8**, 89–100.

Scott, W. R. P., Hünenberg, P. H., Tironi, I. G., Mark, A. E., Billeter, S. R., Fennen, J., Torda, A. E., Huber, T., Krüger, P. & van Gunsteren, W. F. (1999). The GROMOS biomolecular simulation program package. *Journal of Physical Chemistry* (*A*) **103**, 3596–3607.

Sen, S. & Nilsson, L. (1999a). Some practical aspects of free energy calculations from molecular dynamics simulations. *Journal of Computational Chemistry* **20**, 877–885.

Sen, S. & Nilsson, L. (1999b). Structure, interaction, dynamics and solvent effects on the DNA-EcoRI complex in aqueous solution from molecular dynamics simulation. *Biophysical Journal* **77**, 1782–800.

Sessions, R. B., Dauber-Osguthorpe, P. & Osguthorpe, D. J. (1989). Filtering molecular dynamics trajectories to reveal low-frequency collective motions: phospholipase A2. *Journal of Molecular Biology* **210**, 617–633.

Sham, Y. Y., Chu, Z. T., Tao, H. & Warshel, A. (2000). Examining methods for calculations of binding free energies: LRA, LIE, PDLD-LRA, and PDLD/S-LRA calculations of ligands binding to an HIV protease. *Proteins: Structure, Function and Genetics* **39**, 393–407.

Shea, J.-E. & Brooks III, C. L. (2001). From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. *Annual Review of Physical Chemistry* **52**, 499–535.

Sheinerman, F. B. & Brooks III, C. L. (1998a). Calculations on folding of segment B1 of streptococcal protein G. *Journal of Molecular Biology* **278**, 439–456.

Sheinerman, F. B. & Brooks III, C. L. (1998b). Molecular picture of folding of a small $\alpha/\beta$ protein. *Proceedings of the National Academy of Sciences USA* **95**, 1562–1567.

Shimada, J., Kaneko, H., Takada, T., Kitamura, S. & Kajiwara, K. (2000). Conformation of amylose in aqueous solution: small-angle x-ray scattering measurements and simulations. *Journal of Physical Chemistry* (*B*) **104**, 2136–2147.

Shinoda, W. & Mikami, M. (2001). Self-guided molecular dynamics in the isothermal-isobaric ensemble. *Chemical Physics Letters* **335**, 265–272.

Shobana, S., Roux, B. & Andersen, O. S. (2000). Free energy simulations: thermodynamics reversibility and variability. *Journal of Physical Chemistry* (*B*) **104**, 5179–5190.

Shroll, R. M. & Smith, D. E. (1999). Molecular dynamics simulations in the grand canonical ensemble: formation of a bias potential for umbrella samplng. *Journal of Chemical Physics* **110**, 8295–8302.

Simmerling, C., Miller, J. L. & Kollman, P. A. (1998). Combined locally enhanced sampling and particle mesh Ewald as a strategy to locate the experimental structure of a nonhelical nucleic acid. *Journal of the American Chemical Society* **120**, 7149–7155.

Simonson, T. (2001). Macromolecular electrostatics: continuum models and their growing pains. *Current Opinion in Structural Biology* **11**, 243–252.

Singh, U. C. & Kollman, P. A. (1986). 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. *Journal of Computational Chemistry* **7**, 718–730.

Smith, P. E., Blatt, H. D. & Pettitt, B. M. (1997). On the presence of rotational Ewald artifacts in the equilibrium and dynamical properties of a zwitterionic tetrapeptide in aqueous solution. *Journal of Physical Chemistry* (*B*) **101**, 3886–3890.

Smith, P. E. & Pettitt, B. M. (1995). Efficient Ewald electrostatic calculations for large systems. *Computer Physics Communications* **91**, 339–344.

Smith, P. E. & Pettitt, B. M. (1996). Ewald artifacts in liquid state molecular dynamics simulations. *Journal of Chemical Physics* **105**, 4289–4293.

Smith, P. E. & van Gunsteren, W. F. (1994). Predictions of free-energy differences from a single simulation of the initial-state. *Journal of Chemical Physics* **100**, 577–585.

Smondyrev, A. M. & Berkowitz, M. L. (1999). United atom force field for phospholipid membranes: constant pressure molecular dynamics simulations of dipalmitoylphophatidicholine/water system. *Journal of Computational Chemistry* **20**, 531–545.

Sosa, C. P., Hewitt, T., Lee, M. R. & Case, D. A. (2001). Vectorization of the generalized Born model for molecular dynamics on shared-memory computers. *Journal of Molecular Structure: Theochem* **549**, 193–201.

Souaille, M. & Roux, B. (2001). Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Computer Physics Communications* **135**, 40–57.

Speelman, B., Brooks, B. R. & Post, C. B. (2001). Molecular dynamics simulations of human rhinovirus and an antiviral compound. *Biophysical Journal* **80**, 121–129.

Spieser, S. A. H., van Kuik, J. A., Kroon-Batenburg, L. M. J. & Kroon, J. (1999). Improved carbohydrate force field for GROMOS: ring and hydroxymethyl group conformational and exo-anomeric effect. *Carbohydrate Research* **322**, 264–273.

Steinbach, P. J. & Brooks, B. R. (1993). Protein hydration elucidated by molecular dynamics simulation. *Proceedings of the National Academy of Sciences USA* **90**, 9135–9139.

Steinbach, P. J. & Brooks, B. R. (1994). New spherical-cutoff methods for long-range forces in macromolecular simulation. *Journal of Computational Chemistry* **15**, 667–683.

Stern, H. A., Kaminski, G. A., Banks, J. L., Zhou, R., Berne, B. J. & Friesner, R. A. (1999). Fluctuating charge, polarizable, dipole, and combined models: parameterization from *ab initio* quantum chemistry. *Journal of Physical Chemistry* (*B*) **103**, 4730–4737.

Stocker, U. & van Gunsteren, W. F. (2000). Molecular dynamics simulation of hen egg white lysozyme: a test of the GROMS96 force field against nuclear magnetic resonance data. *Proteins: Structure, Function and Genetics* **40**, 145–153.

Stouch, T. R., Ward, K. B., Altieri, A. & Hagler, A. T. (1991). Simulations of lipid crystals. Characterization of potential energy functions and parameters for lecithin molecules. *Journal of Computational Chemistry* **12**, 1033–1046.

Straatsma, T. P. & McCammon, J. A. (1992). Computational alchemy. *Annual Review of Physical Chemistry* **43**, 407–435.

Straatsma, T. P., Philippopoulos, M. & McCammon, J. A. (2000). NWChem: exploiting parallelism in molecular simulations. *Computer Physics Communications* **128**, 377–385.

Stultz, C. M. & Karplus, M. (1999). MCSS functionality maps for a flexible protein. *Proteins: Structure, Function and Genetics* **37**, 512–529.

Sugita, Y. & Okamoto, Y. (1999). Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters* **314**, 141–151.

Tajkorshid, E., Nollert, P., Jensen, M. O., Mierke, L. J. W., O'Connell, J., Stroud, R. M. & Schulten, K. (2002). Control of the selectivity of the aquaporin water channel family by global orientational tuning. *Science* **296**, 525–530.

Tieleman, D. P. & Berendsen, H. J. C. (1996). Molecular dynamics simulations of a fully hydrated dipalmitoylphosphatidylcholine bilayer with different macroscopic boundary conditions and parameters. *Journal of Chemical Physics* **105**, 4871–4880.

Tirado-Rives, J. & Jorgensen, W. L. (1991). Molecular-dynamics simulations of the unfolding of an alpha-helical analog of ribonuclease A S-peptide in water. *Biochemistry* **30**, 3864–3871.

Tirado-Rives, J. & Jorgensen, W. L. (1993). Molecular dynamics simulations of the unfolding of apomyoglobin in water. *Biochemistry* **32**, 4175–4184.

Tirado-Rives, J., Orozco, M. & Jorgensen, W. L. (1997). Molecular dynamics simulations of the unfolding of barnase in water and 8 M aqueous urea. *Biochemistry* **36**, 7313–7329.

Tironi, I. G., Sperb, R., Smith, P. E. & van Gunsteren, W. F. (1995). A generalized reaction field method for molecular dynamics simulations. *Journal of Chemical Physics* **102**, 5451–5459.

Tobias, D. J. (2001). Electrostatic calculations: recent methodological advances and applications to membranes. *Current Opinion in Structural Biology* **11**, 253–261.

Tobias, D. J., Tu, K. C. & Klein, M. L. (1997). Atomic-scale molecular dynamcis of lipid membranes. *Current Opinion in Colloid Interface Science* **2**, 15–26.

Toukmaji, A. Y. & Board, J. A. (1996). Ewald summation techniques in perspective: a survey. *Computer Physics Communications* **95**, 73–92.

Tu, K. C., Tobias, D. J. & Klein, M. L. (1995). Constant-pressure and temperature molecular-dynamics simulations of crystals of the lecithin fragments – glycerylphosphorylcholine and dilauroylglycerol. *Journal of Physical Chemistry* **99**, 10035–10042.

Tuckerman, M. E., Martyna, G. J. & Berne, B. J. (1992). Reversible multiple time scale molecular dynamics. *Journal of Chemical Physics* **97**, 1990–2001.

Tuckerman, M. E. & Martyna, G. J. (2000). Understanding modern molecular dynamics: techniques and applications. *Journal of Physical Chemistry* (*B*) **104**, 159–178. [Additions and corrections (2001), **105**, 7598.]

Tuckerman, M. E., Mundy, C. J. & Martyna, G. J. (1999). On the classical statistical mechanics of non-Hamiltonian systems. *Europhysics Letters* **45**, 149–155.

Tuckerman, M. E., Yarne, D. A., Samuelson, S. O., Hughes, A. L. & Martyna, G. J. (2000). Exploiting multiple levels of parallelism in molecular dynamics based calculations via modern techniques and software paradigms on distributed memory computers. *Computer Physics Communications* **128**, 333–376.

Valleau, J. P. & Cohen, L. K. (1980). Primitive model electrolytes. I. Grand canonical Monte Carlo computations. *Journal of Chemical Physics* **72**, 5935–5941.

van der Vaart, A., Bursulaya, B. D., Brooks III, C. L. & Merz Jr., K. M. (2000). Are many-body effects important in protein folding? *Journal of Physical Chemistry* (*B*) **104**, 9554–9563.

van Gunsteren, W. F. & Berendsen, H. J. C. (1987). GROningen MOlecular Simulation (GROMOS), library manual. Groningen: Biomos BV.

van Gunsteren, W. F. & Berendsen, H. J. C. (1990). Computer simulation of molecular dynamics: methodology, applications, and perspectives in chemistry. *Angewandte Chemie International Edition* **29**, 992–1023.

van Schaik, R. C., Berendsen, H. J. C., Torda, A. E. & van Gunsteren, W. F. (1993). A structure refinement method based on molecular dynamics in four spatial dimensions. *Journal of Molecular Biology* **234**, 751–762.

Villà, J. & Warshel, A. (2001). Energetics and dynamics of enzymatic reactions. *Journal of Physical Chemistry* (*B*) **105**, 7887–7907.

Vorobjev, Y. N. & Hermans, J. (1999). A critical analysis of methods of calculation of a potential in simulated polar liquids: strong arguments in favor of 'molecule-based' summation and of vacuum boundary conditions in Ewald summation. *Journal of Physical Chemistry* (*B*) **103**, 10234–10242.

Wall, I. D., Leach, A. R., Salt, D. W., Ford, M. G. & Essex, J. W. (1999). Binding constants of neuraminidase inhibitors: an investigation of the linear interaction

energy method. *Journal of Medicinal Chemistry* **42**, 5142–5152.

WALSER, R., HÜNENBERGER, P. H. & VAN GUNSTEREN, W. F. (2001). Comparison of different schemes to treat long-range electrostatic interactions in molecular dynamics simulations of a protein crystal. *Proteins: Structure, Function and Genetics* **43**, 509–519.

WANG, J., CIEPLAK, P. & KOLLMAN, P. A. (2000). How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry* **21**, 1049–1074.

WANG, J. & KOLLMAN, P. A. (2001). Automatic parameterization of force field by systematic search and genetic algorithms. *Journal of Computational Chemistry* **22**, 1219–1228.

WANG, S. S. & KRUMHANSL, J. A. (1972). Superposition approximation. II. High density fluid argon. *Journal of Chemical Physics* **56**, 4287–4290.

WANG, W., DONINI, O., REYES, C. M. & KOLLMAN, P. A. (2001). Biomolecular simulations: recent developments in force field, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. *Annual Review of Biophysical and Biomolecular Structures* **30**, 211–243.

WANG, W., WANG, J. & KOLLMAN, P. A. (1999). What determines the van der Waals coefficient $\beta$ in the LIE (Linear Interaction Energy) method to estimate binding free energies using molecular dynamics simulations? *Proteins: Structure, Function and Genetics* **34**, 395–402.

WARSHEL, A. & LEVITT, M. (1976). Theoretical studies of enzymatic reactions: dielectric, electrostatic and steric stabilization of carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology* **103**, 227–249.

WEBER, W., HÜNENBERGER, P. H. & MCCAMMON, J. A. (2000). Molecular dynamics simulations of a polyalanine octapeptide under Ewald boundary conditions: influence of artificial periodicity on peptide conformation. *Journal of Physical Chemistry* (*B*) **104**, 3668–3675.

WEINER, P. K. & KOLLMAN, P. A. (1981). AMBER: assisted model building with energy refinement. A general program for modeling molecules and their interactions. *Journal of Computational Chemistry* **2**, 287.

WEISER, J., SHENKIN, P. S. & STILL, W. C. (1999). Fast, approximate algorithm for detection of solvent-inaccessible atoms. *Journal of Computational Chemistry* **20**, 586–596.

WILLIAMS, D. E. (2001). Improved intermolecular force field for molecules containing H, C, N, and O atoms, with application to nucleoside and peptide crystals. *Journal of Computational Chemistry* **22**, 1154–1166.

WILLIAMS, M. A., THORNTON, J. M. & GOODFELLOW, J. M. (1997). Modelling protein unfolding: hen egg-white lysozyme. *Protein Engineering* **10**, 895–903.

WINN, P. J., FERENCZY, G. G. & REYNOLDS, C. A. (1997). Toward improved force fields. 1. Multipole-derived atomic charges. *Journal of Physical Chemistry* (*A*) **101**, 537–545.

WINN, P. J., FERENCZY, G. G. & REYNOLDS, C. A. (1999). Towards improved force field: III. Polarization through modified atomic charges. *Journal of Computational Chemistry* **20**, 704–712.

WOLF, D., KEBLINSKI, P., PHILLPOT, S. R. & EGGEBRECHT, J. (1999). Exact method for the simulation of Coulombic systems by spherically truncated, pairwise $r^{-1}$ summation. *Journal of Chemical Physics* **110**, 8254–8282.

WONG, K.-B., CLARKE, J., BOND, C. J., NEIRA, J. L., FREUND, S. M. V., FERSHT, A. R. & DAGGETT, V. (2000). Towards a complete description of the structural and dynamic properties of the denatured stat of barnase and the role of residual structure in folding. *Journal of Molecular Biology* **296**, 1257–1282.

WONG, K.-Y. & PETTITT, B. M. (2000). A new boundary condition for computer simulations of interfacial systems. *Chemical Physics Letters* **326**, 193–198.

WOOD, R. H., YEZDIMER, E. M., SAKANE, S., BARRIOCANAL, J. A. & DOREN, D. J. (1999). Free energies of solvation with quantum mechanical interaction energies from classical mechanical simulations. *Journal of Chemical Physics* **110**, 1329–1337.

WOODS, R. J. & CHAPPELLE, R. (2000). Restrained electrostatic potential atomic partial charges for condensed-phase simulations of carbohydrates. *Journal of Molecular Structures: Theochem* **527**, 149–156.

WOODS, R. J., DWEK, R. A., EDGE, C. J. & FRASER-REID, B. (1995). Molecular mechanical and molecular dynamical simulations of glycoproteins and oligosaccharides. 1. GLYCAM-93 parameter development. *Journal of Physical Chemistry* **99**, 3832–3846.

WU, X. W. & WANG, S. M. (1998). Self-guided molecular dynamics simulation for efficient conformational search. *Journal of Physical Chemistry* (*B*) **102**, 7238–7250.

YEH, I.-C. & BERKOWITZ, M. L. (1999). Ewald summation for systems with slab geometry. *Journal of Chemical Physics* **111**, 3155–3162.

YORK, D. M. & KARPLUS, M. (1999). A smooth solvation potential based on the conductor-like screening model. *Journal of Physical Chemistry* (*A*) **103**, 11060–11079.

YOUNG, M. A., GONFLONI, S., SUPERTI-FURGA, G., ROUX, B. & KURIYAN, J. (2001). Dynamic coupling between the SH2 and SH3 domains of c-Src and Hck underlies their inactivation by C-terminal tyrosine phosphorylation. *Cell* **105**, 115–126.

ZHANG, L. Y., GALLICCHIO, E., FRIESNER, R. A. & LEVY, R. M. (2001). Solvent models for protein-ligand binding: comparison of implicit solvent Poisson and surface generalized Born models with explicit solvent simulations. *Journal of Computational Chemistry* **22**, 591–607.

ZHANG, Y., LEE, T.-S. & YANG, W. (1999). A pseudobond approach to combining quantum mechanical and

molecular mechanical methods. *Journal of Chemical Physics* **110**, 46–54.

Zhang, Y., Liu, H. & Yang, W. (2000). Free energy calculation on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined *ab initio* QM/MM potential energy surface. *Journal of Chemical Physics* **112**, 3483–3492.

Zhou, J., Reich, S. & Brooks, B. R. (2000). Elastic molecular dynamics with self-consistent flexible constraints. *Journal of Chemical Physics* **112**, 7919–7929.

Zhou, R., Berne, B. J. & Germain, R. (2001b). The free energy landscape for $\beta$ hairpin folding in explicit water. *Proceedings of the National Academy of Sciences USA* **98**, 14931–14936.

Zhou, R., Harder, E., Xu, H. & Berne, B. J. (2001a). Efficient multiple time step method for use with Ewald and particle mesh Ewald for large biomolecular systems. *Journal of Chemical Physics* **115**, 2348–2358.

Zwanzig, R. W. (1954). High-temperature equation of state by a perturbation method. I. Nonpolar gases. *Journal of Chemical Physics* **22**, 1420–1426.