

## 20.2. Molecular-dynamics simulations of biological macromolecules

BY C. B. POST AND V. M. DADARLAT

### 20.2.1. Introduction

Molecular dynamics (MD) is the simulation of motion for a system of particles. Advances in the theory of atomic interactions and the increasing availability of high-power computers have led to rapid development of this field and greater understanding of macromolecular motions. In the earliest molecular-dynamics simulations of protein molecules (McCammon *et al.*, 1977; McCammon & Harvey, 1987), the systems were greatly simplified in order to fit within the computing capabilities of that time. Simplifications included the exclusion of water molecules and even of explicit hydrogen atoms; the effect of hydrogen atoms was built into the heavy-atom properties using so-called extended-atom parameters. Simulation time periods were limited to tens of picoseconds for systems of less than  $10^3$  atoms. Modern simulations, by contrast, are based on improved force fields (MacKerell *et al.*, 1998) and benefit from considerable development in algorithms. In addition, the possible size and time period of simulations have increased by orders of magnitude; large systems of the order of  $10^4$  atoms (including explicit solvent molecules) and nanosecond time periods are accessible. With dedicated computer time, the microsecond regime is possible (Duan & Kollman, 1998). Interestingly, the first 100 ps simulation of an enzyme complex was of hen egg-white lysozyme (Post *et al.*, 1986), the first enzyme whose structure was solved by X-ray crystallography. Then the simulation required several months of dedicated time on a Cray supercomputer, but now it can be accomplished in less than a week on a common workstation.

A consequence of this enormous growth in computing power has been the particularly successful application of molecular dynamics of biological molecules to three-dimensional structure determination and refinement. It is now practical to use molecular dynamics, in combination with crystallographic and NMR data, to search the large conformational space of proteins and nucleic acids to find structures consistent with the data and to improve the agreement with the data. The advantages of molecular dynamics over manual rebuilding and least-squares refinement are the abilities to overcome the local minimum problem in an automated fashion and to search the complex conformational space of a macromolecule more extensively (Brünger *et al.*, 1987).

### 20.2.2. The simulation method

Molecular mechanics, whereby the energy of the system is expressed in classical terms as a function of atomic coordinates, is well established as a useful approach for describing atomic interactions (Brooks *et al.*, 1988; Goodfellow & Levy, 1998). Owing to the size of proteins and nucleic acids, the potential-energy function for large biomolecules is empirically based rather than derived from quantum-mechanical calculations. The total force on each atom,  $\mathbf{F}_i$ , is calculated from the gradient, or the first derivative of this potential energy, with respect to the atomic coordinates. The motion of the atom resulting from the net force is described by Newton's equation of motion,

$$\mathbf{F}_i = m_i \mathbf{a}_i, \quad (20.2.2.1)$$

where  $m_i$  is the mass of atom  $i$  and  $\mathbf{a}_i$  is the acceleration. Integration of equation (20.2.2.1) gives

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t)\Delta t + \mathbf{F}_i(t)(\Delta t)^2/(2m_i), \quad (20.2.2.2)$$

where  $\Delta t$  is the time step in the integration,  $\mathbf{r}_i(t + \Delta t)$  is the atomic position at time  $(t + \Delta t)$ , given the position  $\mathbf{r}_i(t)$  at time  $t$ , and  $\mathbf{v}_i(t)$

is the velocity. The forces on the particles change continuously so that a numerical solution of the equation is required. The Verlet algorithm (Verlet, 1967) or a variation, Leapfrog, is commonly used.

### 20.2.3. Potential-energy function

#### 20.2.3.1. Empirical energy

The central element of simulations is the interaction potential between atoms as a function of atomic position,  $\mathbf{r}$ . The success of simulations in describing the average structure of proteins and other biological features suggests that such relatively simple potential functions adequately represent proteins and nucleic acids. The empirically based components of the energy function,  $E_{\text{empir}}$ , include geometric terms for bond lengths, bond angles and torsion angles, and non-bonding terms for steric van der Waals interactions and electrostatic interactions. A commonly used energy function is

$$E_{\text{empir}} = E_{\text{geom}} + E_{\text{nonb}}, \quad (20.2.3.1)$$

$$E_{\text{geom}} = \sum_{\text{bonds}} (1/2)k_b(b - b_{\text{eq}})^2 + \sum_{\text{angles}} (1/2)k_\theta(\theta - \theta_{\text{eq}})^2 + \sum_{\text{torsions}} (1/2)k_\varphi[1 + \cos(n\varphi - \delta)], \quad (20.2.3.2)$$

$$E_{\text{nonb}} = \sum_{\text{nbpairs}} (q_i q_j / D r_{ij}) + 4\varepsilon_{ij} \left[ (\sigma_{ij}/r_{ij})^{12} - (\sigma_{ij}/r_{ij})^6 \right], \quad (20.2.3.3)$$

where  $k_b$ ,  $k_\theta$  and  $k_\varphi$  are force constants,  $b_{\text{eq}}$  and  $\theta_{\text{eq}}$  are equilibrium values for bond lengths,  $b$ , and angles,  $\theta$ , respectively, and  $\varphi$  is the torsion angle of periodicity  $n$  and phase  $\delta$ . The non-bonded terms depend on the interatomic distance  $r_{ij}$ , the dielectric constant  $D$ , the partial atomic charge  $q_i$ , and the van der Waals parameters  $\varepsilon_{ij}$  and  $\sigma_{ij}$ . The bond-stretching and angle-bending contributions are represented by harmonic potentials, while the energy associated with rotation about a bond, the torsional potential, is modelled by a cosine function [equation (20.2.3.2)]. The electrostatic component of the non-bonded interactions [first term of equation (20.2.3.3)] follows Coulomb's Law, and a Lennard–Jones 6–12 potential function [second term of equation (20.2.3.3)] is used to model steric repulsion and attractive dispersion interactions.  $E_{\text{nonb}}$ , as a sum over pairs of atoms not involved in either a bond or bond angle, requires the use of a pairwise list between atoms. The small contribution from pairs separated by a large distance allows the use of cutoff limits for this list, but at some cost in accuracy.

Initial values for the atomic coordinates and velocities are required to begin the molecular-dynamics simulation. While initial coordinates are obtained from the model built into the electron-density map, it is necessary to generate initial velocities computationally. The most common approach is to assign random values for each atom,  $i$ , consistent with the temperature chosen for the system:  $3kT/2 = (1/2) \sum m_i v_i^2$ .

Integration of the equation of motion [equation (20.2.2.1)] also requires specification of the time step  $\Delta t$ . In the case of structure determination, this choice is limited only by the numerical stability of the calculation. Too large a value for  $\Delta t$  results in errors in the integration, manifested by a rapid and unacceptable increase in energy. Whereas a value for  $\Delta t$  of 1 to 2 fs is required for accurate trajectories and strict conservation of the energy, structure-determination protocols can employ larger values and are limited only by the need for numerical stability.

Both the temperature and  $\Delta t$  influence the sampling rate of conformational space. Enhanced sampling increases the rate of

## 20. ENERGY CALCULATIONS AND MOLECULAR DYNAMICS

convergence to the structure solution. Issues related to sampling are detailed in Chapter 18.2.

### 20.2.3.2. Particle mesh Ewald

In an MD simulation, the accurate and rapid calculation of long-range electrostatic interactions is a central issue for the correct physical representation of the system. The Coulombic potential [first term of equation (20.2.3.3)] has been used in most cases, but the Coulombic interactions must be limited for practical reasons to a subset of pair interactions so that long-distance interactions are truncated. More recently, the Ewald method has been implemented to avoid the need for truncating the non-bonded pair list while maintaining computational efficiency. The electrostatic energy is calculated using periodic boundary conditions\* and requires a double summation over all atoms in the central unit cell, as well as their infinite number of periodic images.

The total electrostatic energy of a periodic system with a neutral unit cell containing  $N$  point charges  $q_1, q_2, \dots, q_n$  located at positions  $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N$  is

$$E_{\text{elec}} = \frac{1}{2} \sum_n \sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{n}|}$$

The sum over  $|\mathbf{n}|$  takes into account all the periodic unit cells, and  $\mathbf{n}$  reflects the shape of the unit cell. For a simple periodic cubic lattice,  $\mathbf{n} = (n_x L, n_y L, n_z L)$ , where  $L$  is the length of the unit cell, and  $n_x, n_y$  and  $n_z$  are integers. The prime indicates that in the central unit cell the terms  $i = j$  are not included. The sum is convergent for very large  $|\mathbf{n}|$ .

An efficient way of evaluating the triple sum above is represented by the Ewald method. In this method, each point charge is surrounded by a Gaussian charge distribution of opposite sign and same integrated magnitude as the original charge:

$$\rho_i(\mathbf{r}) = -q_i \beta^3 \exp(-\beta^2 \mathbf{r}^2) / \pi^{3/2}.$$

$\beta$  is inversely proportional to the width of the charge distribution and  $\mathbf{r}$  is the position relative to the centre of the distribution. The result of adding this charge distribution is a screening of the electrostatic interactions, so that the interaction between neighbouring charges is effectively short-ranged. A total screened potential is calculated by summation in the real-space lattice over the initial point-charge distribution together with the screening-charge distribution. To end up with the original point-charge distribution, a cancelling distribution having the opposite sign and same functional form as the Gaussian screening distribution must be added. Together, these Gaussian distributions form a smooth varying function of  $\mathbf{r}$ , which can be approximated by a superposition of continuous functions. The contribution of the cancelling distribution is calculated by adding the Fourier transforms of the distributions in reciprocal space at each point charge and transforming the total back into real space.

The final Ewald expression for the total electrostatic energy contains a real-space sum plus a reciprocal-space sum, minus a self-term introduced by the interaction of the cancelling distribution with itself:

$$E_{\text{elec}} = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \left[ \sum_n q_i q_j \frac{\text{erfc}(\beta |\mathbf{r}_{ij} + \mathbf{n}|)}{|\mathbf{r}_{ij} + \mathbf{n}|} + q_i q_j \Phi_{\text{rec}}(\beta \mathbf{r}_{ij}) \right] - \frac{\beta}{\pi^{1/2}} \sum_{i=1}^N q_i^2 + J(\mathbf{D}, \mathbf{P}, e').$$

$J(\mathbf{D}, \mathbf{P}, e')$  (de Leeuw *et al.*, 1980) depends on the total dipole moment of the unit cell, the shape of the macroscopic crystal lattice and the dielectric constant of the surrounding medium. The term  $\text{erfc}(\beta |\mathbf{r}_{ij} + \mathbf{n}|)$  is the complementary error function which falls to zero as  $(\beta |\mathbf{r}_{ij} + \mathbf{n}|)$  increases. The term  $\Phi_{\text{rec}}(\beta \mathbf{r}_{ij})$  is the reciprocal Ewald sum:

$$\Phi_{\text{rec}}(\beta \mathbf{r}_{ij}) = \frac{1}{\pi V} \sum_{\mathbf{m} \neq 0} \frac{\exp(-\pi^2 \mathbf{m}^2 / \beta^2)}{\mathbf{m}^2} \exp(2\pi i \mathbf{m} \mathbf{r}).$$

$V$  is the volume and  $\mathbf{m}$  is a vector in the reciprocal space. The Ewald sum in reciprocal space is the solution to Poisson's equation, with Gaussian charge densities as sources and with periodic boundary conditions. For large  $\beta$ , the only contribution to the sum in the real space comes from the minimum image terms. When  $\beta$  is large, the charge distributions are sharp and the summation in reciprocal space must be done over a large number of reciprocal-space vectors to achieve convergence.

Calculation of the Ewald sum in the original form is slow for large macromolecular solution simulations. The usual implementation of the Ewald summation for calculating electrostatic interaction is an order  $N^2$  algorithm, or at best,  $N^{3/2}$  (Christiansen *et al.*, 1993) by adjusting  $\beta$  to optimize computational effort. The particle mesh Ewald (PME) method (Essmann *et al.*, 1995; Darden *et al.*, 1993) is an approximation of the Ewald sum that allows an order  $N \log N$  algorithm for the calculation of the total electrostatic energy. This reduction in the number of steps is accomplished by choosing  $\beta$  large enough that atom pairs separated by distances larger than a specified cutoff (9–12 Å) make a negligible contribution to the direct-space sum, which thus becomes an order  $N$  computation. Moreover, the reciprocal-space sum is expressed in terms of the electrostatic structure factors. Each atomic charge distribution is approximated by a gridded charge distribution. The resulting approximate structure factors are calculated by a three-dimensional fast Fourier transform applied to the grid. Using an optimized  $\beta$  and grid density for each simulated system, the PME method can compute the electrostatic energy for a periodic boundary system at the same level of accuracy as the classical Ewald summation in a much shorter time (Darden *et al.*, 1993).

### 20.2.3.3. Experimental restraints in the energy function

For the purpose of structure determination, the potential-energy function used for molecular-dynamics calculation incorporates the information from experimental data in the form of non-physical restraint terms. These restraint terms, introduced to bias the conformational sampling toward structures consistent with the experiment, are used in addition to the total potential function and, in some sense, can fulfil the requirements of a physical term in equation (20.2.2.2) (see below). The experimentally based restraint terms are added to the potential-energy function to give a total effective potential,  $E_{\text{tot}} = E_{\text{empir}} + E_{\text{rest}}$ . Whereas structure-determination protocols based on NMR data employ a number of types of restraint terms, data from X-ray crystallography provide a single restraint term,  $E_{\text{rest}} = w E_{\text{Xray}}$ , the residual between the observed and

\* The periodic boundary conditions method allows for the simulation of the atoms in a central unit cell in the field of image unit cells generated by symmetry operations.

calculated structure-factor amplitudes; where

$$wE_{\text{Xray}} = w \sum_{hkl} (|F_o| - k|F_c|)^2,$$

where  $w$  and  $k$  are scale factors.

#### 20.2.4. Empirical parameterization of the force field

Considerable effort has gone into the development of a number of force fields for use in molecular-dynamics simulations of biomolecules (Jorgensen & Tirado-Rives, 1988; MacKerell *et al.*, 1995, 1998; van Gunsteren *et al.*, 1996). The parameters described here are those of the CHARMM22 force field, the force field used in the X-PLOR program. Estimation of the force constants, equilibrium values and non-bonding parameters in equations (20.2.3.2) and (20.2.3.3) involves a self-consistent approach that balances the bonding and non-bonding interaction terms among the macromolecule and solvent molecules (MacKerell *et al.*, 1998). A wide range of data are taken into account during an interactive process of optimization in order to adequately account for the extensive and correlated nature of the parameters in a consistent fashion. Small-molecule model compounds representative of proteins or nucleic acids are considered in detail, and a hierarchical approach is applied to extend the parameters to larger molecules with minimal adjustment at the points of connection.

The empirical basis of the parameters is broad. Gas-phase geometries and crystal structures are used to determine equilibrium bond lengths, bond angles, and dihedral phase and periodicity. Vibrational spectra, primarily from gas-phase infrared and Raman spectroscopy, are used to fit values for the force constants. Torsion-angle terms are estimated from relative energies of different conformers of model compounds, such as 4-ethylimidazole and ethylbenzene, based on gas-phase data. In cases where no satisfactory experimental data are available, *ab initio* calculations are used to obtain the required energy surfaces. Adjustments are made to describe the energy barriers and positions of saddle points, as well as the minimum-energy structures.

Optimization of the non-bonded parameters includes fitting the van der Waals and electrostatic terms of equation (20.2.3.3), while maintaining a balance among the protein-protein, water-water and protein-water interactions. The parameterization of the CHARMM22 force field is based on the water model and water-water interactions of the TIP3P model (Jorgensen *et al.*, 1983). As such, use of this parameter set with another water model will lead to inconsistencies in the balance of intermolecular interactions. Data from dipole moments, heats and free energies of vaporization, solvation and sublimation, and molecular volumes, as well as *ab initio* calculations of interaction energies and geometries are used to optimize intermolecular interactions. Partial charges of atoms are determined by fitting *ab initio* interaction energies and geometries of small-molecule compounds that model the peptide backbone and amino-acid side chains. Magnitudes and directions of dipole-moment values are also used to optimize partial charges. Experimental gas-phase dipole-moment values are used when available, while *ab initio* calculated values are adopted otherwise. The van der Waals parameters are then refined by comparing results of condensed-phase simulations on pure solvents with heats of vaporization and molecular volumes.

The crystallographic restraint term in the potential-energy function,  $E_{\text{rest}}$ , must also be parameterized to optimize the agreement with the experimental structure-factor amplitudes while simultaneously retaining good geometry and non-bonding interactions. Optimization of  $E_{\text{rest}} = wE_{\text{Xray}}$  involves only the estimation of  $w$ . Unlike the parameters in  $E_{\text{empir}}$ ,  $w$  has no physical

basis and is usually chosen to make the force due to  $E_{\text{rest}}$  balance the total force contributed by all terms in  $E_{\text{empir}}$ . As refinement of the structure progresses, these forces, and hence  $w$ , necessarily change since the quality, in terms of geometry and non-bonding interactions, of the structure improves and the crystallographic residual is reduced.

#### 20.2.5. Modifications in the force field for structure determination

Simulated-annealing protocols require modification of the parameters to maintain the correct geometry and local structural integrity of the molecule in order to allow heating to very high, non-physical temperatures for several thousand integration steps. Such modifications are acceptable in the case of structure determination since the primary goal is to define the optimum equilibrium structure in best agreement with the crystallographic or NMR data. Simulations intended to reproduce the fluctuations or dynamic properties of the system must employ carefully defined parameters without such modifications. These modifications include substantial increases in the force constants for bond lengths and angles, *e.g.*, factors of two to ten are used in the parameters specified in the X-PLOR file parallhdg.pro. A number of improper torsional terms are added to maintain proper chirality.

The specific terms in  $E_{\text{nonb}}$  are also modified for the purpose of structure determination. In this methodology, the goal is to converge efficiently to a model that satisfies the experimental data, rather than to obtain an accurate description of the conformational surface, such as estimating fluctuations and equilibrium distributions. Alterations in  $E_{\text{nonb}}$  include the replacement of the computationally expensive  $E_{\text{vdw}}$  by a quartic or harmonic repulsive term, which prevents steric conflict among atoms, but ignores dispersive attraction. The electrostatic term,  $E_{\text{elec}}$ , is frequently excluded altogether, since the  $1/r$  dependence of the Coulombic potential allows charge interactions to dominate the interatomic forces far from the global minimum in a fashion that hinders movement toward the global minimum. Exclusion of this important physical property of biological systems is possible, because the crystallographic structure factors contain sufficient information to reflect adequately the imprint of electrostatics on the average structure.

#### 20.2.6. Internal dynamics and average structures

It is most often the goal of the structural biologist to define a single average structure of a macromolecule. The well recognized internal motions arising from thermal fluctuations of a macromolecule may be necessary for function, but, nonetheless, the methods of structure determination generally aim to model a single average structure. Internal motions range from the high frequency, small amplitude motions (*i.e.* those modelled by crystallographic  $B$  values) to low frequency, larger amplitude motions of loops and whole domains. Some studies (Kuriyan *et al.*, 1986; Post, 1992) have examined the validity of the assumptions about fast timescale motions made by the methods of structure determination. It is reasonable that some of the differences between the structure solutions of a protein obtained by NMR spectroscopy and X-ray crystallography are due to differences in the effects of internal motions. The application of molecular-dynamics algorithms for structure determination has allowed the use of protocols that account for effects of internal motions by employing time-averaged restraints (Schiffer *et al.*, 1995).

### 20.2.7. Assessment of the simulation procedure

Equation (20.2.3.1) is a reasonable representation of the energy function of proteins. This point is illustrated here with results from 0.8 ns molecular-dynamics simulations of hen egg-white lysozyme (PDB entry 1lzt, 1.97 Å resolution), bovine pancreas ribonuclease A (5rsa, 2.0 Å resolution), bovine  $\alpha$ -lactalbumin (1hfs, 2.3 Å resolution) and trypsin (2ptn, 1.55 Å resolution). [The coordinates for bovine  $\alpha$ -lactalbumin were kindly provided by K. R. Acharya prior to their publication (Pike *et al.*, 1996).]

The proteins were fully hydrated, and the simulations were calculated with the *CHARMM22* force field, using truncated octahedron periodic boundary conditions. The four proteins were overlaid with bulk water from an equilibrated simulation, and a 10 ps trajectory was calculated for the rearrangement of the water molecules around the protein with fixed protein atoms. The number of water molecules (4000 to 6000) required to hydrate the protein varied with the protein size (123–223 residues) and shape, with at least four layers of water molecules between the peripheral protein atoms and the walls of the boxes. The simulations were performed at constant pressure and temperature ( $T = 300$  K and  $p = 1$  atm) using the extended-system algorithms (Hoover, 1985; Nose, 1984) implemented in *CHARMM*. The 300 K constant temperature was maintained by coupling to an external bath, with a coupling constant of 25 ps. The *SHAKE* algorithm (Ryckaert *et al.*, 1977) was used to constrain bond lengths between hydrogen atoms and heavy atoms, allowing for a time step of 2 fs in the integration of the equations of motion. A non-bonded cutoff of 12 Å was used for the Lennard-Jones potential calculation. The electrostatic forces and energies were computed using the PME method (Darden *et al.*, 1993; Essmann *et al.*, 1995). The PME charge grid spacing was 0.7 Å, and the charge grid was interpolated with the direct sum tolerance set to  $4.0 \times 10^{-6}$ . The non-bonded pair lists were updated every 10 steps. Structures for analysis were saved every 0.1 ps.

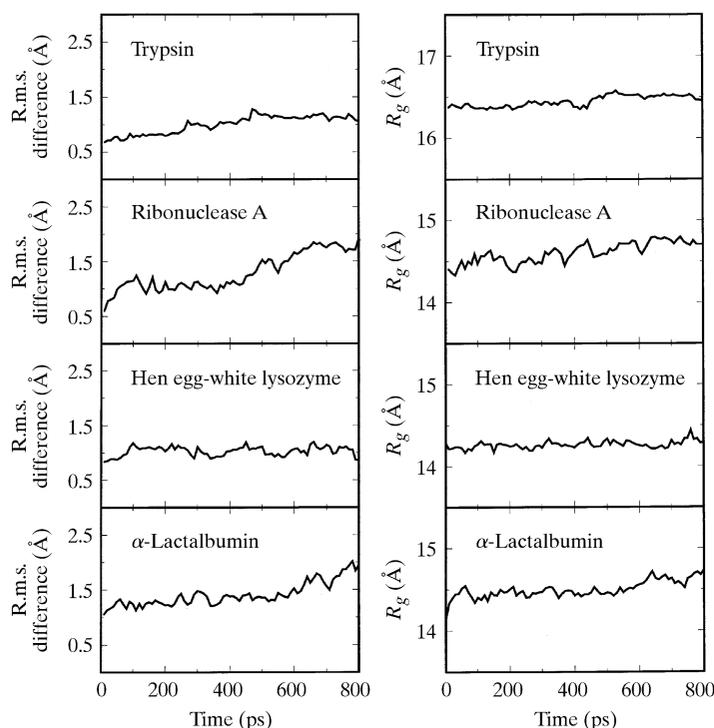


Fig. 20.2.7.1. Structural comparison and radii of gyration of various proteins as a function of time in the molecular-dynamics simulation. Left: r.m.s. coordinate differences averaged over main-chain atoms (N, C $\alpha$ , C) between the energy-minimized crystallographic structure and the simulation snapshot. Right: radii of gyration ( $R_g$ ).

Fig. 20.2.7.1 shows the deviation during the simulation period of the main-chain coordinates between the simulation structures and the crystallographic starting structure. The r.m.s. coordinate deviations in the case of lysozyme are particularly stable and small: approximately 1.0 Å over the course of the trajectory. Other r.m.s. values are more typical and range from 1.0 to 2.0 Å. The time dependence of other properties, such as the radius of gyration, can also be used to follow the stability and behaviour of a trajectory. These time series, also shown in Fig. 20.2.7.1, are constant. Jumps in such a time series can be used to detect conformational transitions (Post *et al.*, 1989).

Other experimental properties have been compared in the literature with those calculated from molecular-dynamics trajectories. Of particular interest is comparing time-dependent properties measured by NMR spectroscopy. An approach to calculating NMR relaxation rates was recognized early on when development of both the molecular-dynamics simulations of proteins and a model-independent theory for NMR relaxation was started (Levy, Karplus & McCammon, 1981; Levy, Karplus & Wolynes, 1981; Lipari & Szabo, 1982). Since then, the common practice of isotopic labelling of proteins for NMR structure determination has allowed the measurement of numerous NMR relaxation rates, particularly rates that characterize the motion of backbone atoms. Long simulations have been conducted to compare the calculated and experimental values (Abseher *et al.*, 1995; Chatfield *et al.*, 1998; Smith *et al.*, 1995). In a particularly long simulation, an 11 ns trajectory period was used to estimate relaxation rates associated with the motions of the vectors N—H, C $\alpha$ —H and C—H methyl groups from alanine and leucine (Chatfield *et al.*, 1998). Trends in the general order of mobility of these vectors are reproduced, although a residue-by-residue comparison shows some differences.

### 20.2.8. Effect of crystallographic atomic resolution on structural stability during molecular dynamics

The variation in r.m.s. deviation between the initial crystallographic structure and the simulation coordinates for different protein trajectories (Fig. 20.2.7.1) raises the question of whether the atomic resolution of the starting X-ray structure influences the magnitude of this deviation. In order to investigate this issue, we calculated trajectories for bovine pancreatic trypsin inhibitor (BPTI), starting with crystallographic structures determined from data at three different atomic resolutions: 1bpi at 1.1 Å resolution (Parkin *et al.*, 1999), 6pti at 1.7 Å resolution (Wlodawer *et al.*, 1987) and 1bhc at 2.7 Å resolution (Hamiaux *et al.*, 1999). The errors in the atomic coordinates estimated from the Luzzati plots are 0.06 Å for the 1.1 Å resolution structure and 0.41 Å for the 2.7 Å resolution structure. The protocol described in the previous section was followed for simulations starting with each of the three crystallographic structures over a 500 ps simulation time. The net charge of +6 e on BPTI was neutralized by adding six chloride anions to the solvated protein system, thus accomplishing the ideal conditions for a PME calculation for the electrostatic interaction. The truncated octahedra contain approximately 3700 water molecules, and the total number of atoms in the simulations is over 12 000. The simulations were carried out on an eight-node IBM/SP2 and required 4.5 h of CPU time per 10 ps of dynamics run.

Root-mean-square differences (r.m.s.d.'s) in atomic coordinates were calculated between all pairs of coordinates from the X-ray structures, the energy-minimized X-ray structures and the 10 ps average MD structure obtained near 300 ps of the simulation period. In Table 20.2.8.1, the upper diagonal r.m.s.d. values are the main-chain-atom differences, while the lower diagonal ones are the side-chain-atom differences. The r.m.s.d.'s between the three X-ray structures range from 0.4–0.5 Å for the main-chain atoms and 1.4–

## 20.2. MOLECULAR-DYNAMICS SIMULATIONS OF BIOLOGICAL MACROMOLECULES

 Table 20.2.8.1. *R.m.s. coordinate differences between crystallographic structures and average MD structures*

The upper half of the matrix contains values for main-chain atoms, while the lower half contains values for side-chain heavy atoms.

		X-ray structure			Energy-minimized X-ray structure			Average MD structure		
		1.1 Å	1.7 Å	2.7 Å	1.1 Å	1.7 Å	2.7 Å	1.1 Å	1.7 Å	2.7 Å
X-ray structure	1.1 Å		0.54	0.51	0.4	0.61	0.7	0.97	1.09	1.13
	1.7 Å	1.5		0.41	0.6	0.42	0.57	1.11	1.14	1.11
	2.7 Å	1.62	1.43		0.59	0.57	0.48	1.12	1.12	1.11
Energy-minimized X-ray structure	1.1 Å	0.57	2.81	2.06		0.59	0.63	1.01	1.15	1.13
	1.7 Å	1.5	0.79	1.68	1.47		0.58	1.12	1.13	1.13
	2.7 Å	1.72	1.65	0.99	1.78	1.77		1.28	1.11	1.11
Average MD structure	1.1 Å	1.89	2.18	2.45	1.94	2.87	2.66		1.06	1.08
	1.7 Å	2.53	2.28	2.49	2.52	2.33	2.52	2.22		0.87
	2.7 Å	2.15	2.05	1.9	2.17	2.13	1.85	2.17	2.15	

1.6 Å for the side-chain atoms. The larger r.m.s. values for averages over side-chain atoms imply alternative side-chain orientations. This degree of structural deviation is visualized with an overlay of the three X-ray structures of BPTI shown on the left-hand side of Fig. 20.2.8.1. Energy minimization with respect to the *CHARMM* force field alters the main-chain atoms of the X-ray structures by approximately 0.4 Å and increases the differences between two X-ray structures to 0.6–0.7 Å. The differences in the main-chain atoms between an MD average structure and an energy-minimized X-ray structure are only somewhat larger: 1.0–1.1 Å. Interestingly, these values are comparable to the values obtained when comparing the three MD average structures: 0.9–1.1 Å. The main-chain structural differences among the three 10 ps average MD structures are shown on the right-hand side of Fig. 20.2.8.1. The general trends observed for main-chain atoms are also found for side-chain atoms. Thus, the differences between the X-ray structures increase somewhat as a result of energy minimization, and the differences between MD average structures and X-ray structures (1.9–2.9 Å)

are similar to those between two X-ray structures (1.5–2.8 Å) or two MD average structures (approximately 2.2 Å).

Similar r.m.s. values are found if the starting velocities for a simulation are varied while maintaining the same starting coordinates (Caves *et al.*, 1998); the r.m.s.d.'s obtained from 120 ps MD simulations were 0.7–1.1 Å for the main-chain atoms with respect to the reference X-ray structure and 0.8–1.5 Å between MD individual trajectory averages. The results given in Table 20.2.8.1, together with those of Caves *et al.* (1998), suggest that sampling on the nanosecond timescale largely reflects the conformational variations due to thermal fluctuations that result from a potential-energy surface with multiple minima separated by low barriers (Cooper, 1976). In this context, MD simulations starting with different X-ray structures offer a more extensive sampling of the conformational space of the specific protein than simulations carried out from a single X-ray structure, although this conclusion remains to be demonstrated by a more thorough analysis. Our results do not support the conclusion that overall

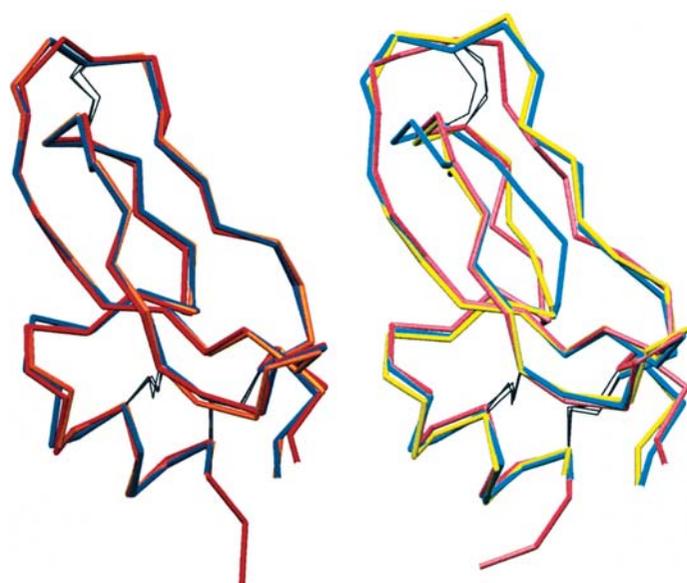


Fig. 20.2.8.1.  $C\alpha$  tracings of BPTI. Left: crystallographic structures determined from data at three different resolutions: 1.1 (red), 1.7 (blue) and 2.7 Å (orange). Right: 10 ps average MD structures from simulations initiated with the energy-minimized crystallographic structure determined at 1.1 (pink), 1.7 (cyan) or 2.7 Å (yellow) resolution. The 10 ps average is over coordinates from 290 to 300 ps.

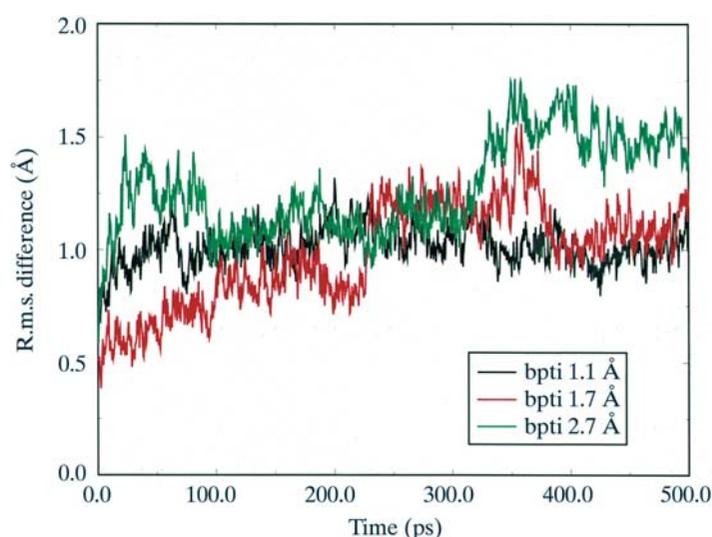


Fig. 20.2.8.2. BPTI r.m.s. coordinate differences between the energy-minimized crystallographic structure and MD snapshots from three simulations. A simulation was initiated from the energy-minimized crystallographic structure determined at 1.1 (black), 1.7 (red) or 2.7 Å (green) resolution. The r.m.s.d. is averaged over the main-chain atoms N,  $C\alpha$  and C.

## 20. ENERGY CALCULATIONS AND MOLECULAR DYNAMICS

r.m.s.d.'s between MD average structures and the starting X-ray structures correlate with atomic resolution.

The r.m.s.d.'s between main-chain atoms in the starting X-ray structures and simulation snapshots as a function of time are presented in Fig. 20.2.8.2. The 1.1 Å resolution structure has the most stable trajectory during the 500 ps trajectory, with an average r.m.s. value of 1.01 (9) Å. The 1.7 Å resolution structure has an r.m.s. value of 0.98 (22) Å. In this simulation, the r.m.s.d.'s fluctuate more widely from the average value, with small differences in the first 200 ps, larger ones between 200 and 400 ps, and again smaller ones in the last 100 ps. For the 2.7 Å resolution structure, the average over the simulation is 1.28 (21) Å. From the results presented here, it is concluded that the higher-resolution structures are more stable during MD simulations and have a shorter equilibration period (50 ps for 1.1 Å resolution and

over 300 ps for 2.7 Å resolution). This conclusion is consistent with larger errors in the atomic coordinates of X-ray structures determined from lower-resolution data.

### Acknowledgements

This work was supported by grants to CBP from the NIH (R01-GM39478, AI39639). CBP was supported by a Research Career Development Award from the NIH (K04-GM00661) and VMD is a DOE/SLOAN Postdoctoral Fellow in computational biology. The computing facilities shared by the Structural Biology group were supported by grants from the Lucille P. Markey Foundation and the Purdue University Academic Reinvestment Program.

## References

### 20.1

- Berendsen, H. J. C., van Gunsteren, W. F., Zwinderman, H. R. J. & Geurtsen, R. (1986). *Simulations of proteins in water*. *Ann. N. Y. Acad. Sci.* **482**, 269–285.
- 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*. *J. Chem. Phys.* **81**, 3684–3690.
- Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F. & Hermans, J. (1981). *Interaction models for water in relation to protein hydration*. In *Intermolecular forces*, edited by B. Pullman, pp. 331–342. Dordrecht: Reidel.
- Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *The Protein Data Bank: a computer-based archival file for macromolecular structures*. *J. Mol. Biol.* **112**, 535–542.
- Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Crystallographic R-factor refinement by molecular dynamics*. *Science*, **235**, 458–460.
- Fennen, J., Torda, A. E. & van Gunsteren, W. F. (1995). *Structure refinement with molecular dynamics and a Boltzmann-weighted ensemble*. *J. Biomol. NMR*, **6**, 163–170.
- Fujinaga, M., Gros, P. & van Gunsteren, W. F. (1989). *Testing the method of crystallographic refinement using molecular dynamics*. *J. Appl. Cryst.* **22**, 1–8.
- Gros, P. & van Gunsteren, W. F. (1993). *Crystallographic refinement and structure-factor time-averaging by molecular dynamics in the absence of a physical force field*. *Mol. Simul.* **10**, 377–395.
- Gros, P., van Gunsteren, W. F. & Hol, W. G. J. (1990). *Inclusion of thermal motion in crystallographic structures by restrained molecular dynamics*. *Science*, **249**, 1149–1152.
- Gunsteren, W. F. van & Berendsen, H. J. C. (1990). *Computer simulations of molecular dynamics: methodology, applications and perspectives in chemistry*. *Angew. Chem. Int. Ed. Engl.* **29**, 992–1023.
- Gunsteren, W. F. van, Berendsen, H. J. C., Hermans, J., Hol, W. G. J. & Postma, J. P. M. (1983). *Computer simulation of the dynamics of hydrated protein crystals and its comparison with X-ray data*. *Proc. Natl Acad. Sci. USA*, **80**, 4315–4319.
- Gunsteren, W. F. van, Billeter, S. R., Eising, A. A., Hünenberger, P. H., Krüger, P., Mark, A. E., Scott, W. R. P. & Tironi, I. G. (1996). *Biomolecular simulation: the GROMOS96 manual and user guide*. Vdf Hochschulverlag, Zürich, Switzerland.
- Gunsteren, W. F. van, Bonvin, A. M. J. J., Daura, X. & Smith, L. J. (1997). *Aspects of modelling biomolecular structures on the basis of spectroscopic or diffraction data*. In *Modern techniques in protein NMR*, edited by Krishna & Berliner. Plenum.
- Gunsteren, W. F. van, Brunne, R. M., Gros, P., van Schaik, R. C., Schiffer, C. A. & Torda, A. E. (1994). *Accounting for molecular mobility in structure determination based on nuclear magnetic resonance spectroscopic and X-ray diffraction data*. *Methods Enzymol.* **239**, 619–654.
- Gunsteren, W. F. van, Kaptein, R. & Zuiderweg, E. R. P. (1984). *Use of molecular dynamics computer simulations when determining protein structure by 2D-NMR*. In *Proceedings of the NATO/CECAM workshop on nucleic acid conformation and dynamics*, edited by W. K. Olson, pp. 79–97. France: CECAM.
- Gunsteren, W. F. van & Karplus, M. (1981). *Effect of constraints, solvent and crystal environment on protein dynamics*. *Nature (London)*, **293**, 677–678.
- Gunsteren, W. F. van & Karplus, M. (1982). *Protein dynamics in solution and in crystalline environment: a molecular dynamics study*. *Biochemistry*, **21**, 2259–2274.
- Harvey, T. S. & van Gunsteren, W. F. (1993). *The application of chemical shift calculation to protein structure determination by NMR*. In *Techniques in protein chemistry IV*, pp. 615–622. New York: Academic Press.
- Heiner, A. P., Berendsen, H. J. C. & van Gunsteren, W. F. (1992). *MD simulation of subtilisin BPN' in a crystal environment*. *Proteins*, **14**, 451–464.
- Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R. & van Gunsteren, W. F. (1985). *A protein structure from nuclear magnetic resonance data, lac repressor headpiece*. *J. Mol. Biol.* **182**, 179–182.
- 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*. *Comput. Phys. Commun.* **91**, 215–231.
- 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*. *J. Comput. Phys.* **23**, 327–341.
- Scheek, R. M., Torda, A. E., Kemmink, J. & van Gunsteren, W. F. (1991). *Structure determination by NMR: the modelling of NMR parameters as ensemble averages*. In *Computational aspects of the study of biological macromolecules by nuclear magnetic resonance spectroscopy*, edited by J. C. Hoch, F. M. Poulsen & C. Redfield, NATO ASI Series, Vol. A225, pp. 209–217. New York: Plenum Press.
- Schiffer, C. A., Gros, P. & van Gunsteren, W. F. (1995). *Time-averaging crystallographic refinement: possibilities and limitations using  $\alpha$ -cyclodextrin as a test system*. *Acta Cryst.* **D51**, 85–92.
- Schiffer, C. A., Huber, R., Wüthrich, K. & van Gunsteren, W. F. (1994). *Simultaneous refinement of the structure of BPTI against NMR data measured in solution and X-ray diffraction data measured in single crystals*. *J. Mol. Biol.* **241**, 588–599.
- Shi, Y.-Y., Yun, R.-H. & van Gunsteren, W. F. (1988). *Molecular dynamics simulation of despentapeptide insulin in a crystalline environment*. *J. Mol. Biol.* **200**, 571–577.

## REFERENCES

## 20.1 (cont.)

- Smith, P. E. & van Gunsteren, W. F. (1995). *Reaction field effects on the simulated properties of liquid water*. *Mol. Simul.* **15**, 233–245.
- Torda, A. E., Brunne, R. M., Huber, T., Kessler, H. & van Gunsteren, W. F. (1993). *Structure refinement using time-averaged J-coupling constant restraints*. *J. Biomol. NMR*, **3**, 55–66.
- Torda, A. E., Scheek, R. M. & van Gunsteren, W. F. (1990). *Time-averaged nuclear Overhauser effect distance restraints applied to tendamistat*. *J. Mol. Biol.* **214**, 223–235.
- Vijay-Kumar, S., Bugg, C. E. & Cook, W. J. (1987). *Structure of ubiquitin refined at 1.8 Å resolution*. *J. Mol. Biol.* **194**, 531–544.

## 20.2

- Abseher, R., Ludemann, S., Schreiber, H. & Steinhauser, O. (1995). *NMR cross-relaxation investigated by molecular dynamics simulation: a case study of ubiquitin in solution*. *J. Mol. Biol.* **249**, 604–624.
- Brooks, C. L. III, Karplus, M. & Pettitt, B. M. (1988). *Proteins: a theoretical perspective of dynamics, structure and thermodynamics*. *Adv. Chem. Phys.* **71**, 1–259.
- Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Crystallographic R factor refinement by molecular dynamics*. *Science*, **235**, 458–460.
- Caves, L. S. D., Evanseck, J. D. & Karplus, M. (1998). *Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin*. *Protein Sci.* **7**, 649–666.
- Chatfield, D. C., Szabo, A. & Brooks, B. R. (1998). *Molecular dynamics of staphylococcal nuclease: comparison of simulation with 15N and 13C NMR relaxation data*. *J. Am. Chem. Soc.* **120**, 5301–5311.
- Christiansen, D., Peram, J. W. & Petersen, H. G. (1993). *On the fast multipole method for computing the energy of periodic assemblies of charged and dipolar particles*. *J. Comput. Phys.* **107**, 403–405.
- Cooper, A. (1976). *Thermodynamic fluctuations in protein molecules*. *Proc. Natl Acad. Sci. USA*, **92**, 2740–2741.
- Darden, T., York, D. & Pedersen, L. (1993). *Particle mesh Ewald (PME): a N log(N) method for Ewald sums in large systems*. *J. Chem. Phys.* **98**, 10089–10092.
- 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.
- Essmann, U., Perrera, L., Berkovitz, M. L., Darden, T., Lee, H. & Pedersen, L. G. (1995). *A smooth particle mesh Ewald method*. *J. Chem. Phys.* **103**, 8577–8593.
- Goodfellow, J. M. & Levy, R. M. (1998). *Theory and simulation*. *Curr. Opin. Struct. Biol.* **8**, 209–210.
- Gunsteren, W. F. van, Billeter, S. R., Eising, A. A., Hünenberger, P. H., Mark, A. E., Scott, W. R. P. & Tironi, I. G. (1996). *Biomolecular simulation: the GROMOS96 manual and user guide*. Vdf Hochschulverlag, Zurich, Switzerland.
- Hamiaux, C., Prangé, T., Riès-Kautt, M., Ducruix, A., Lafont, S., Astier, J. P. & Veessler, S. (1999). *The dcameric structure of bovine pancreatic trypsin inhibitor (BPTI) crystallized from thiocyanate at 2.7 Å resolution*. *Acta Cryst.* **D55**, 103–113.
- Hoover, W. G. (1985). *Canonical dynamics: equilibrium phase-space distributions*. *Phys. Rev. A*, **31**, 1695–1697.
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. (1983). *Comparison of simple potential functions for simulating liquid water*. *J. Chem. Phys.* **79**, 926–935.
- Jorgensen, W. L. & Tirado-Rives, J. (1988). *The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin*. *J. Am. Chem. Soc.* **110**, 1657–1666.
- Kuriyan, J., Petsko, G. A., Levy, R. M. & Karplus, M. (1986). *Effect of anisotropy and anharmonicity on protein crystallographic refinement: an evaluation by molecular dynamics*. *J. Mol. Biol.* **190**, 455–479.
- Leeuw, S. W. de, Perram, J. W. & Smith, E. R. (1980). *Simulation of electrostatic systems in periodic boundary conditions. I. Lattice sums and dielectric constants*. *Proc. R. Soc. London Ser. A*, **373**, 27–56.
- Levy, R. M., Karplus, M. & McCammon, J. A. (1981). *Increase of <sup>13</sup>C NMR relaxation times in proteins due to picosecond motional averaging*. *J. Am. Chem. Soc.* **103**, 994–996.
- Levy, R. M., Karplus, M. & Wolynes, P. G. (1981). *NMR relaxation parameters in molecules with internal motion: exact Langevin trajectory results compared with simplified relaxation models*. *J. Am. Chem. Soc.* **103**, 5998–6011.
- Lipari, G. & Szabo, A. (1982). *Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. I. Theory and range of validity*. *J. Am. Chem. Soc.* **104**, 4546–4559.
- McCammon, J. A., Gelin, B. R. & Karplus, M. (1977). *Dynamics of folded proteins*. *Nature (London)*, **267**, 585–590.
- McCammon, J. A. & Harvey, S. C. (1987). *Dynamics of proteins and nucleic acids*. Cambridge University Press.
- MacKerell, A., Wiorkiewicz-Kuczera, J. & Karplus, M. (1995). *An all atom empirical energy function for the simulation of nucleic acids*. *J. Am. Chem. Soc.* **117**, 11946–11975.
- MacKerell, A. D. Jr, Bashford, D., Bellott, M., Dunbrack, R. L. Jr, 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, W. E. III, Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J. & Karplus, M. (1998). *All-atom empirical potential for molecular modeling and dynamics studies of proteins*. *J. Phys. Chem. B*, **102**, 3586–3616.
- Nose, S. (1984). *A unified formulation of the constant temperature molecular dynamics methods*. *J. Chem. Phys.* **81**, 511–519.
- Parkin, S., Rupp, B. & Hope, H. (1999). *The structure of bovine pancreatic trypsin inhibitor at 125 K: definition of carboxyl-terminal residues glycine-57 and alanine 58*. In preparation.
- Pike, A. C., Brew, K. & Acharya, K. R. (1996). *Crystal structures of guinea-pig, goat and bovine α-lactalbumin highlight the enhanced conformational flexibility of regions that are significant for its action in lactose synthase*. *Structure*, **4**, 691–703.
- Post, C. B. (1992). *Internal motional averaging and three-dimensional structure determination by NMR*. *J. Mol. Biol.* **224**, 1087–1101.
- Post, C. B., Brooks, B. R., Karplus, M., Dobson, C. M., Artymiuk, P. J., Cheetham, J. C. & Phillips, D. C. (1986). *Molecular dynamics simulations of native and substrate-bound lysozyme*. *J. Mol. Biol.* **190**, 455–479.
- Post, C. B., Dobson, C. M. & Karplus, M. (1989). *A molecular dynamics analysis of protein structural elements*. *Proteins*, **5**, 337–354.
- 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*. *J. Comput. Phys.* **23**, 327–341.
- Schiffer, C. A., Gros, P. & van Gunsteren, W. F. (1995). *Time-averaging crystallographic refinement: possibilities and limitations using α-cyclodextrin as a test system*. *Acta Cryst.* **D51**, 85–92.
- Smith, P. E., van Schaik, R. C., Szyperski, T., Wuthrich, K. & van Gunsteren, W. F. (1995). *Internal mobility of the basic pancreatic trypsin inhibitor in solution: a comparison of NMR spin relaxation measurements and molecular dynamics simulation*. *J. Mol. Biol.* **246**, 356–365.
- Verlet, L. (1967). *Computer experiments on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules*. *Phys. Rev.* **159**, 98–103.
- Wlodawer, A., Nachman, J., Gilliland, G. L., Gallagher, W. & Woodward, C. (1987). *Structure of form III crystals of bovine pancreatic trypsin inhibitor*. *J. Mol. Biol.* **198**, 469–480.