International Tables for Crystallography (2006). Vol. F, Section 20.2.8, pp. 492–494.

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 α-lactalbumin (1hfz, 2.3 Å resolution) and trypsin (2ptn, 1.55 Å resolution). [The coordinates for bovine α-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 *CHARMM*22 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 4.0×10^{-6}. The non-bonded pair lists were updated every 10 steps. Structures for analysis were saved every 0.1 ps.

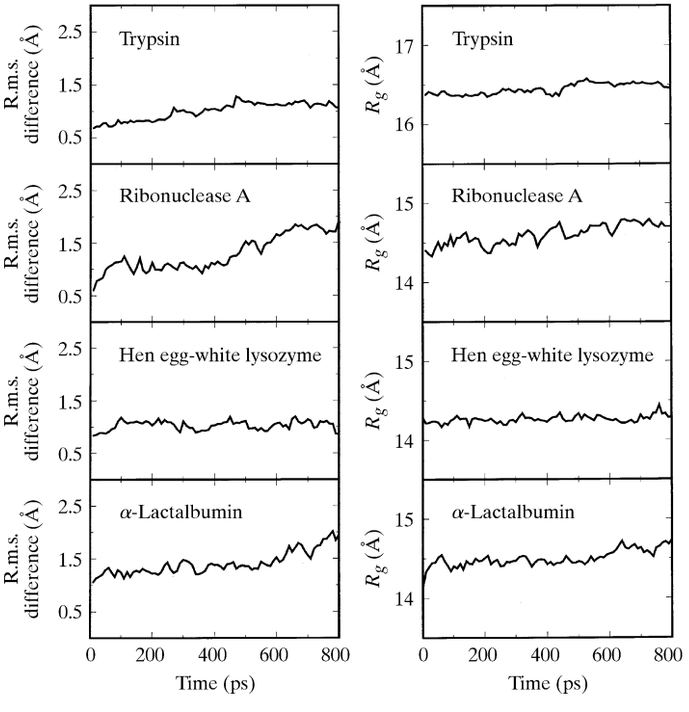

Fig. 20.2.7.1. Structural comparison and radii of gyration of various proteins as a function of time in the molecular-dynamics simulation. Left: r.m.s. coordinate differences averaged over main-chain atoms (N, Cα, C) between the energy-minimized crystallographic structure and the simulation snapshot.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

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 velocites 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

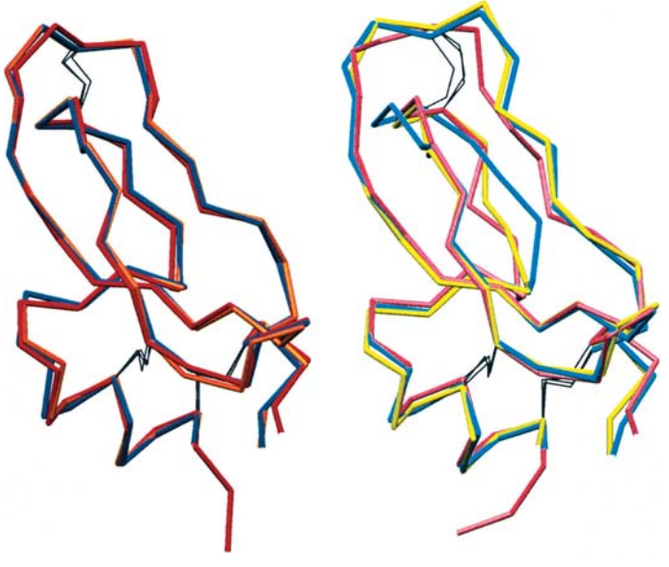

Fig. 20.2.8.1. Cα tracings of BPTI. Left: crystallographic structures determined from data at three different resolutions: 1.1 (red), 1.7 (blue) and 2.7 Å (orange). Right: 10 ps average MD structures from simulations initiated with the energy-minimized crystallographic structure determined at 1.1 (pink),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

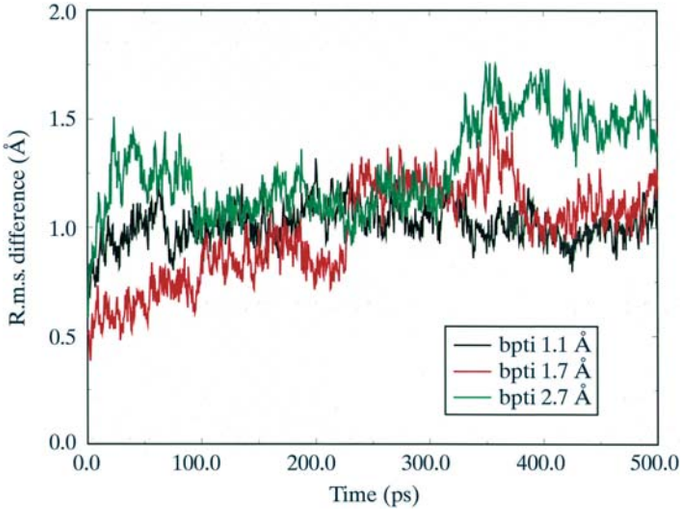

Fig. 20.2.8.2. BPTI r.m.s. coordinate differences between the energy-minimized crystallographic structure and MD snapshots from three simulations. A simulation was initiated from the energy-minimized crystallographic structure determined at 1.1 (black), 1.7 (red) or 2.7 Å (green) resolution. The r.m.s.d.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

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 α-cyclodextrin as a test system. Acta Cryst.* D**51**, 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.