

Fluctuation and Cross-correlation Analysis of Protein Motions Observed in Nanosecond Molecular Dynamics Simulations

P. H. Hünenberger, A. E. Mark and W. F. van Gunsteren*

Laboratorium für
Physikalische Chemie
ETH-Zentrum, CH-8092
Zürich, Switzerland

Nanosecond molecular dynamics simulations of bovine pancreatic trypsin inhibitor and lysozyme in water are analyzed in terms of backbone atomic positional fluctuations and dynamical cross-correlations. It is found that although the molecular systems are stable, *B*-factors calculated over a time period as long as 500 ps are not representative for the motions within the proteins. This is especially true for the most mobile residues. On a nanosecond time-scale, the *B*-factors calculated from the simulations of the proteins in solution are considerably larger than those obtained by structure refinement of the proteins in crystals, based on X-ray data. The time evolution of the atomic fluctuations shows that for large portions of the proteins under study, atomic positional fluctuations are not yet converged after a nanosecond. Cross-correlations do not converge faster than the fluctuations themselves. Most display very erratic behavior if the sampling covers less than about 200 ps. It is also shown that inclusion of mobile atoms into the procedure used to remove rigid-body motion from the simulation can lead to spurious correlations between the motions of the atoms at the surface of the protein.

© 1995 Academic Press Limited

Keywords: computer simulation; molecular dynamics; protein dynamics; atomic fluctuations; *B*-factors

*Corresponding author

Introduction

Knowledge of the atomic motions and their collective (or correlated) character in proteins is central to the understanding of their biological function (Huber and Bennett, 1983). Although difficult to obtain experimentally, this information may easily be extracted from molecular dynamics (MD) simulations. During the past ten years, atomic positional fluctuations and the correlations of the motions between atom pairs as observed in MD simulations of proteins have been analyzed to further the understanding of a variety of dynamical properties of proteins (McCammon & Harvey, 1987; Brooks *et al.*, 1988; Ichiye & Karplus, 1991). These include identification of groups of atoms moving in concert, influence of the solvent on the dynamics, distinction between lock-and-key or induced-fit

models of binding, differences in the dynamics between free and bound protein, side-chain motions in enzymatic reactions, allosteric enzyme regulation, interrelationship between different crystallographic conformers, and conformational changes occurring upon thermal unfolding of a protein. A non-exhaustive list of such studies can be found in Table 1.

Atomic positional fluctuation and cross-correlation analysis techniques seem appealing for two reasons. First, atomic positional fluctuations can formally be related to crystallographic temperature factors or *B*-factors, and thus compared with available experimental data. Second, cross-correlation analysis can be used to broaden the scope or time-scale of a MD simulation by extrapolating the motions along directions in which major changes should occur, for example on a longer time-scale or upon substrate binding, without actually simulating these motions. The following questions have, however, to our knowledge never been systematically investigated:

(1) Is it realistic to compare *B*-factors resulting from structure refinement based on crystallographic X-ray data with *B*-factors obtained from simulations

Abbreviations used: MD, molecular dynamics; HEWL, hen egg white lysozyme; BPTI, bovine pancreatic trypsin inhibitor; ms, mean-square; r.m.s., root-mean-square; SPC, simple point charge; SPC/E, extended simple point charge; DCCM, dynamical cross-correlation map.

of solvated proteins, considering the very different environments, time-scales, system sizes and determination techniques?

(2) Is agreement between the X-ray derived and simulated *B*-factors a good indicator for the quality of a simulation of a protein in solution?

(3) What is the minimum averaging time required to obtain a representative sample of the extent of the motions within the protein, so that an extrapolation to longer times becomes meaningful?

(4) What is the influence of the fitting procedure that is generally used to remove rigid-body motions from the trajectory on the results of the analysis?

To address these questions, atomic positional fluctuation and fluctuation correlation analyses have been performed on two MD trajectories, a 1.4 ns simulation of bovine pancreatic trypsin inhibitor (BPTI; Figure 1a) and a 1.1 ns simulation of hen egg white lysozyme (HEWL; Figure 1b), using different averaging time windows, and the results have been compared to experimental data. Both simulations were carried out using explicit water molecules in order to avoid the distortive effects induced by vacuum boundary conditions (Levitt & Sharon, 1988; Komeiji *et al.*, 1991; Harte *et al.*, 1992; Elofsson & Nilsson, 1993; Hayward *et al.*, 1993; Ptaszek *et al.*, 1994). The simulation periods of over a nanosecond are close to the current computational limits of sampling in protein systems. The present work focuses on the description of backbone motions; similar principles, however, naturally apply to side-chain motions.

Results and Discussion

To describe correctly an equilibrium property of a system based on a MD simulation, two conditions must be fulfilled (van Gunsteren *et al.*, 1995): (1) the studied trajectory must be an equilibrium trajectory, and (2) the simulation time should be considerably longer than the relaxation (or build-up) time of the property of interest. The first condition clearly does not imply the second. Whether the former condition is satisfied in our two simulations will be discussed below. The following four subsections are dedicated to the discussion of the second condition, which is the main subject of the present work. The last section deals with the influence of the choice of a structure fitting procedure.

Stability of the simulations

To assess the stability of the two simulations, a collection of properties was monitored as a function of time. The means, standard deviations, maximal and minimal values, drifts (from a linear regression, given per nanosecond) as well as the values calculated from the crystal structure are reported in Table 2. Both simulations are essentially stable.

For BPTI (Figure 1a) the drift (per nanosecond) for most properties is of the same order or smaller than the standard deviation. The radius of gyration

increases by about 0.02 nm when the protein is immersed into water. The initial root-mean-square (r.m.s.) positional deviations measured for all atoms and for C α atoms only, excluding equilibration, are 0.24 and 0.14 nm, respectively. These three parameters do not display significant drifts or variances after the 50 ps equilibration time. A slight decrease (250 kJ mol $^{-1}$ ns $^{-1}$, 0.2%) is observed in the total potential energy, mostly due to a decrease in the protein-protein non-bonded energy. The average apolar exposed surface area is larger (by 4.0 nm 2) than in the X-ray structure, whereas the polar exposed surface area is smaller (by 3.5 nm 2). These changes occur during the 50 ps equilibration period and no further significant changes are observed during the analysis period. The secondary structure is very stable throughout the simulation, as was checked using the DSSP program (Kabsch & Sander, 1983) for configurations taken every 10 ps. Residues 48 to 51 always form an α -helix, which extends up to residue 54 in more than 90% of the configurations and up to residue 55 in 13% of the configurations. Residues 19 to 23 and 30 to 34 always adopt an antiparallel β -sheet structure, to which the pairs 24–29 and 18–35 are associated in 94% of the configurations. A hydrogen-bonded bridge is observed in all configurations between Tyr21 and Phe45. The N-terminal 3 10 helix (residues 3 to 6) makes an exception to this overall high stability. In about 70% of the cases, DSSP describes it as a 3 10 helix from residues 3 to 5 (extended to Leu6 in only 43% of the cases), in 13% of the cases as an α -helix, and the rest of the time as a hydrogen-bonded turn. Interconversion of $i \rightarrow i + 3$ and $i \rightarrow i + 4$ hydrogen bonds occurs in this helix on a picosecond time-scale.

The lysozyme (Figure 1b) simulation shows larger fluctuations than the BPTI one. Although the average radius of gyration is equal to the one calculated for the crystal structure, it drifts from 1.44 nm at the end of the 50 ps equilibration time to 1.41 nm at the end of the 1.1 ns simulation, with a clear minimum at 400 ps, where it reaches a value of 1.38 nm. The r.m.s. positional deviations, calculated for all atoms and C α atoms only, increase from 0.15 and 0.09 nm, respectively, at 50 ps, to 0.31 and 0.22 nm by the end of the simulation. Corresponding changes are not perceptible in the energies. Whereas the polar surface area is almost equal to the one calculated from the crystal structure and seems relatively stable during the simulation, there is a change in the apolar exposed surface area. The value is about 5 nm 2 above the X-ray value at 50 ps and slowly decreases to finally reach a value of about 2 nm 2 above it. A plot describing the evolution of the secondary structure during this simulation has been presented elsewhere (van Gunsteren *et al.*, 1995; Figure 6)

Simulated *B*-factors calculated using averaging windows of different lengths

In conventional crystallographic refinement of a protein structure, a set of isotropic *B*-factors is

Table 1. Selected atomic fluctuation and cross-correlation studies of proteins

Theme ^a	System ^b	Time ^c (ps)	Reference
Fast local versus slow concerted motions	Cytochrome c Vacuum	<32>	McCammon (1984)
Validity of analytical correlation maps from normal mode analysis of a single structure	BPTI, crambin, RNase A, HEWL NMA	—	McCammon & Harvey (1987) Levitt et al. (1985)
Active-site dynamics	RNase A (active site)	<27>	Brünger et al. (1985)
Side-chain positioning around the substrate	Free and substrate bound forms Solvent	eq. 12	
Dynamics of different conformers	Pig insulin (mono- and dimer)	<80>	Caves et al. (1990)
Secondary structure dynamics	From three crystal structures Vacuum	eq. 20	
Comparison with thermal diffuse scattering results	HIV-1 protease dimer Vacuum and solvent	2 × <40> eq. 20	Harte et al. (1990) Swaminathan et al. (1991)
Domain-domain communication	BPTI	<25>	Harte et al. (1992) Ichiye & Karplus (1991)
Molecular cantilever mechanism	Vacuum, solvent, crystal and NMA Tlp apo- and holo-repressor	<100>	Komeiji et al. (1991)
Relation of these motions to enzyme function	Vacuum and solvent	eq. 50	Komeiji et al. (1994)
Description of the cross-correlation analysis technique	BPTI, native 298 K and reduced 498 K	<50>	Daggett & Levitt (1993)
Collective motions and influence of the environment	Solvent	eq. 100	
Three domains articulation format	Antibody 8F5 (binding domain)	<75>	de la Cruz et al. (1994)
Monomer ability to bind itself, co-repressor and DNA	Free and antigen bound forms Solvent	eq. 50	
Native versus "molten globule" dynamics	Human CD-4 protein (N-terminal part) Vacuum and solvent	<134> eq. 31	Praszek et al. (1994)
Binding to antigen of induced-fit or lock-and-key type			
Binding to gp120 of induced-fit or lock-and-key type			
Binding regulation by allosteric effect			

All studies listed use cross-correlation analysis of atomic motions.

^a Theme, describes the purpose of the study, i.e. not necessarily the conclusion reached.

^b Characteristics of the studies are indicated under System, in particular MD simulation in solvent, vacuum, crystal, or normal-mode analysis (NMA).

^c Time, <...> indicates the simulation time (excluding equilibration) and eq. the equilibration time.

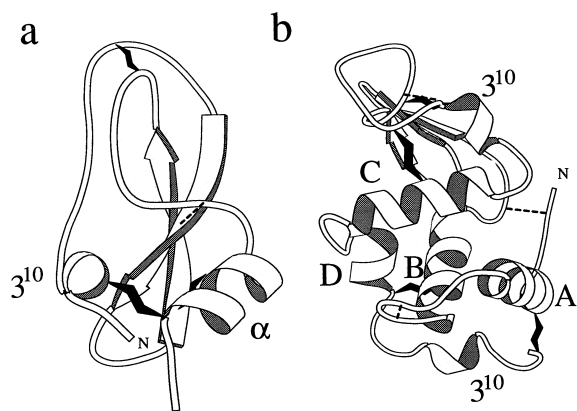


Figure 1. Ribbon models corresponding to the crystal structures of a, BPTI; and b, HEWL. N indicates the N terminus. Secondary structure assignment to residues was made using the program DSSP (Kabsch & Sander, 1983), i.e. a, 3^{10} , 3 to 6; β , 18 to 24 and 29 to 35; α , 48 to 55; and b, A, 4 to 15; B, 24 to 36; β (rear), 42 to 46 and 50 to 53 and 58 to 59; 3^{10} (top), 79 to 84; C, 88 to 99; D, 108 to 115; 3^{10} (bottom), 120 to 125. Black hooks represent disulfide bridges, i.e. a, [5–55], [14–38] and [30–51] and; b, [6–127], [30–115], [64–80] and [76–94]. Broken lines represent β -bridges, i.e. a, 21 to 45; and b, 2 to 39, 20 to 23 and 65 to 79.

obtained by minimizing the R -factor as a function of atomic coordinates and B -factors. The actual procedure utilizes fast-Fourier transforms and performs the minimization in terms of structure factor amplitudes. In other words, atomic coordinates and B -factors are fitted to the electron density map. This means that, in principle, a B -factor represents the spread of density around a location or peak in this map, irrespective of the particular atom that is occupying that location at a certain point in time. In refinement practice, however, B -factors are assigned to particular atoms. This is a correct interpretation as long as different atoms do not occupy the same location at different times. The sizes of the B -factors are representative of the amount of disorder or motion present in the crystal on the time-scale of the diffraction experiment, which ranges from seconds to days. Simulated B -factors can be calculated from a MD trajectory using equation (2) and compared to crystallographic B -factors, assuming equivalence between the simulated mean square positional fluctuation or B -factor of a given atom and a location in the crystal. In Figure 2, the simulated B -factors for BPTI are reported. Figure 2a shows their values using an averaging over nine windows of 50 ps length, each shifted by 150 ps, Figure 2b shows the corresponding values for seven consecutive windows of 200 ps length and Figure 2c for five consecutive windows of 500 ps length (from an extended trajectory) together with the crystallographic B -factors (broken line).

On a 50 ps time-scale, few motions are observed in a consistent manner between all the time windows. For most residues, a sizeable scatter is observed in the values of the simulated B -factors. Only the β -sheet residues consistently display a low mobility. The use

of a longer averaging time (200 ps, 500 ps) increases the overall amount of motion observed for most residues, but generally does not result in a reduction of the scatter between different windows. The most striking examples are the 3^{10} helix, the loops 12 to 18 and 36 to 42 and the three C-terminal residues. Exceptions to this behavior are, however, observed. The atomic positional fluctuations for the β -sheet residues and for residues 43 to 45 remain consistent among the different time-windows at any studied time-scale. The β -factors of the C^α atoms corresponding to the small loop connecting the first two strands of the β -sheet (residues 24 to 29) even seem to converge when increasing the time-scale. In spite of these exceptions, in large portions of the protein the simulated B -factors are not yet converged on a 500 ps time-scale and thus cannot be expected to be representative of the dynamics of the system at a macroscopic time-scale.

Although the mobile regions observed in the simulation on the 500 ps time-scale (mainly the N and C termini, the 3^{10} helix and the three loops connecting secondary structure elements, i.e. residues 12 to 18, 24 to 29 and 36 to 42) roughly correspond to regions with high crystallographic B -factors, agreement residue by residue is not found. The fluctuations observed in certain regions, especially in the C terminus and the loop connecting the 3^{10} helix and the first strand of the β -sheet, are larger than those determined by structure refinement based on X-ray data. A similar effect is observed in the 1.05 ns lysozyme simulation (Figure 3). For the least mobile residues, mainly helices A, B and C, and the three strands of the β -sheet, the simulated values are 1.5 to 2 times larger than the X-ray refinement values. The other regions of the protein, mainly the hairpin turn connecting helices A and B, the small loop between the two first strands of the β -sheet, the main loop region (residues 61 to 78), the β -domain 3^{10} helix and the 34 C-terminal residues have B -factors three to eight times larger than expected based on the crystallographic values.

Similar discrepancies between experimental and simulated B -factors have been reported for other proteins (Caves *et al.*, 1990; Elofsson & Nilsson, 1993). The following explanations can be proposed.

(1) The configuration space sampled in the experiment is incompletely sampled in the simulation since: (i) crystallographic B -factors derived from X-ray data include static disorder due to averaging over a collection of molecules (Frauenfelder *et al.*, 1979); (ii) in the crystal, restricted oscillation of the whole molecule around its lattice position occurs, which is not accounted for in the MD results due to the fitting procedure (Edwards *et al.*, 1990); (iii) the simulation time (nanosecond) is much shorter than the data acquisition time (>seconds).

(2) Crystallographic B -factors form an incomplete measure of the motion since: (i) they are a measure of the spread of the electron density as a function of position in the crystal, irrespective of which particular atom is contributing to the electron

Table 2. Data concerning the time evolution of some protein properties during the HEWL and BPTI simulations

	Mean	s.d.	Min.	Max.	Drift	X-ray
A. BPTI <50 to 1400 ps>						
R_{gyr} (nm)	1.123	0.0084	1.094	1.153	+0.0062	1.105
All r.m.s. (nm)	0.240	0.0188	0.199	0.301	-0.0164	(0)
C^{α} r.m.s. (nm)	0.141	0.0161	0.101	0.181	+0.0123	(0)
E_{pot} tot (MJ/mol)	-118.3	0.35	-119.5	-116.8	-0.25	—
E_{NB} P-P (MJ/mol)	-4.7	0.27	-5.6	-3.7	-0.28	—
E_{NB} P-S (MJ/mol)	-9.1	0.34	-10.2	-7.8	+0.06	—
E_{NB} S-S (MJ/mol)	-106.2	0.32	-107.4	-105.0	-0.04	—
SAS apol (nm ²)	22.4	0.44	20.9	23.3	+0.04	18.4
SAS pol (nm ²)	10.2	0.23	9.6	10.8	+0.03	13.7
SAS tot (nm ²)	32.6	0.52	31.2	33.8	+0.07	32.1
C_p (cal/(K mol))	742.0	19.97	672.5	779.6	+1.00	473.6
B. HEWL <50 to 1100 ps>						
R_{gyr} (nm)	1.419	0.0124	1.381	1.456	-0.0118	1.419
All r.m.s. (nm)	0.266	0.0339	0.153	0.320	+0.0716	(0)
C^{α} r.m.s. (nm)	0.182	0.0282	0.087	0.244	+0.0937	(0)
E_{pot} tot (MJ/mol)	-200.4	0.44	-202.1	-198.6	+0.071	—
E_{NB} P-P (MJ/mol)	-11.9	0.31	-12.9	-10.7	+0.035	—
E_{NB} P-S (MJ/mol)	-13.8	0.42	-15.5	-12.4	-0.044	—
E_{NB} S-S (MJ/mol)	-178.5	0.41	-180.0	-176.7	+0.043	—
SAS apol (nm ²)	37.4	0.95	35.6	41.1	-1.59	35.2
SAS pol (nm ²)	20.9	0.37	20.0	21.9	+0.33	21.3
SAS tot (nm ²)	58.3	1.04	56.4	62.0	-1.26	56.4
C_p (cal/(K mol))	1137.8	43.50	1061.8	1308.3	-79.9	1029.5

Values for BPTI were monitored from 50 to 1400 ps and for HEWL from 50 to 1100 ps. For these intervals, the following quantities are given: Mean, mean value; s.d., standard deviation; Min, minimal value; Max, maximal value; Drift, drift from a linear regression over the trajectory, indicated per nanosecond; X-ray, value calculated using the X-ray structures (BPTI: Wlodawer *et al.*, 1987; HEWL: Ramanadham *et al.*, 1987). The properties are: R_{gyr} , radius of gyration; All r.m.s./ C^{α} r.m.s., all atom/ C^{α} atom root-mean-square atomic positional deviation from the X-ray structure; E_{pot} tot, total (all force field terms) potential energy; E_{NB} P-P/ E_{NB} P-S/ E_{NB} S-S, non-bonded (Coulomb plus van der Waals) energy between protein and protein/protein and solvent/solvent and solvent atoms; SAS apol/SAS pol/SAS tot, apolar/polar/total solvent accessible surface area calculated using the algorithm of Lee & Richards (1971) with a probe radius of 1.4 Å and a density of about 5 dots/Å²; C_p , heat-capacity value calculated from the solvent accessible surface areas using an increment of 0.45 cal K⁻¹ mol⁻¹ Å⁻² for apolar and -0.26 cal K⁻¹ mol⁻¹ Å⁻² for polar surface area (Murphy *et al.*, 1992).

density (i.e. they result from fitting on the electron density map); (ii) they are, in general, restricted to a given maximum value during structure refinement, whereas the simulated B -factors are true atomic positional fluctuations, which may in principle grow infinitely with atomic mobility; (iii) systematic errors in the data, e.g. due to absorption, extinction and thermal diffuse scattering, may not have been corrected.

(3) Crystal packing forces restrict the motions in the protein, especially for surface regions (loops, N and C termini). The MD simulations concern proteins in solution.

(4) The force field used in the simulations can be biased in favor of more or less mobility than is present in reality.

Reason (1) tends toward simulated B -factors that are lower than experimental X-ray B -factors. Reasons (1) (i) and (1) (ii) are often invoked (Levitt & Sharon, 1988; Caves *et al.*, 1990; Komeiji *et al.*, 1994) to explain this observation in short simulations. In reality, the short averaging time (reason (1) (iii)) is probably the dominant factor. Reasons (2) and (3) point in the opposite direction and may explain our finding

that mobility is larger in the simulation even on the short nanosecond timescale. We calculated the correlation coefficients between three different sets of crystallographic B -factors for the protein BPTI (Brookhaven Protein Data Bank entries 4PTI, 5PTI and 6PTI). Using linear regression for residues 5 to 55, we found the following correlation coefficient values: 0.35 (4PTI versus 5PTI); 0.22 (4PTI versus 6PTI); and 0.64 (5PTI versus 6PTI). Recently, it has been shown by comparison of electron-density maps obtained from anisotropic B -factor refinement and from time-averaging MD refinement using exact data that the former technique yields a worse representation of the exact data than the latter (Schiffer *et al.*, 1995). The influence of the force field (reason (4)) is difficult to assess but should not be underestimated. This is especially true when performing nanosecond simulations with a force field tested on much shorter time-scales. If, however, the substantial difference between a crystal and an aqueous environment (reason (3)) is the real explanation for the larger mobility observed in the simulations, simulated B -factors from a solvent simulation and B -factors derived from crystallographic X-ray data should be compared very cautiously.

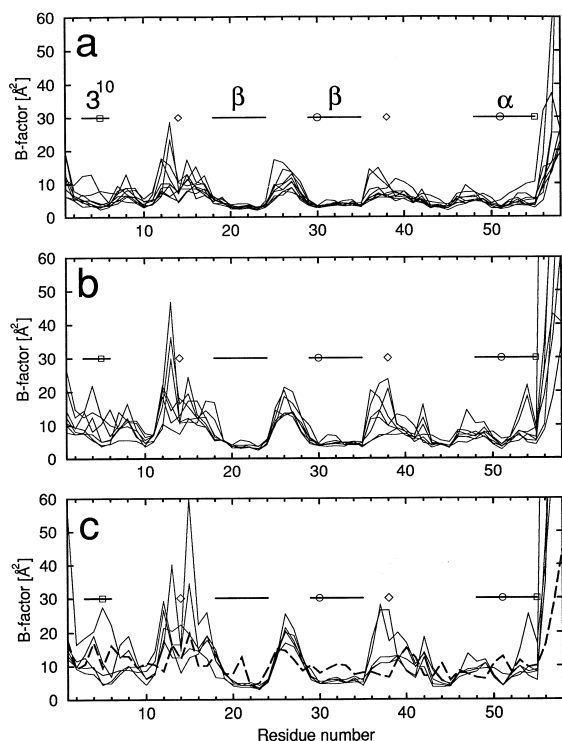


Figure 2. Simulated B -factors for the C^α atoms of BPTI calculated according to equation (2): a, for nine time windows of 50 ps length starting after 50 ps (equilibration period), each shifted by 150 ps forward in time; b, for seven consecutive windows of 200 ps length starting after 50 ps; c, for five consecutive windows of 500 ps length starting after 50 ps (taken from a trajectory extended to 2550 ps). The broken line in c represents B -factors obtained from crystallographic structure refinement based on X-ray data (Wlodawer *et al.*, 1987). Secondary structure elements and disulfide bonds are indicated at the top (see also Figure 1a), where \square , [5–55]; \diamond , [14–38] and \circ , [30–51].

Correlation between simulated and experimental B -factors over different averaging windows

The lysozyme trajectory has been analyzed in terms of simulated atomic B -factors calculated from equation (2) for the C^α atoms, using windows of 50, 100, 200, 500 and 1000 ps, time-shifted by 50 ps. For 50, 200 and 1000 ps, the correlation coefficient, assuming a linear relationship between the simulated B -factors and the experimental crystallographic B -factors, is reported in Figure 4b. It is clear that using a sampling window of 50 ps the correlation coefficients are unreliable, ranging from 0.33 to 0.80. Only when the sampling is extended to 500 ps (not shown) do the results become relatively stable, with correlation coefficients ranging from 0.65 to 0.73. Increasing the sampling time not only removes the strong variation of the correlation coefficients, but also increases the correlation on average in this case. The average correlation coefficients using observation windows of increasing length are: 0.59 for

50 ps; 0.65 for 100 ps; and 0.69 for 200, 500 and 1000 ps.

Such correlation coefficients are sometimes presented in the literature as evidence of the quality of a trajectory. Correlation of the simulated B -factors to the X-ray B -factors for the 1.35 ns BPTI simulation results in a correlation coefficient of 0.85. However, the linear regression parameters are mostly determined by the large simulated B -factors ($\geq 94 \text{ \AA}^2$) of the three C-terminal residues. If these residues are removed from the comparison, the correlation coefficient drops to 0.60. Figure 4a is the analog of Figure 4b for the BPTI trajectory, with residues 56 to 58 removed from the fit. Although BPTI displays the same global behavior as lysozyme, the correlation tends to decrease during the course of the simulation. The motions initially sampled are more characteristic of the crystal structure, reflecting the starting configuration, whereas motions characteristic of the solvent environment require a longer equilibration and sampling time to be observed. When the correlation of the simulated B -factors with other (5PTI and 6PTI) sets of crystallographic B -factors is calculated, the same pattern as that shown in Figure 4a for 4PTI is obtained. The correlation coefficients for the 1.35 ns window are 0.60 (4PTI), 0.50 (5PTI) and 0.45 (6PTI).

Thus, good agreement (i.e. a high correlation coefficient) between simulated and crystallographically refined B -factors does not necessarily indicate a high quality simulation. Linear regression parameters can be dominated by highly mobile residues in loops and chain termini, short sampling times (<200 ps) lead to extreme scatter, and initial sampling of still highly “crystal-like” motions can strongly bias the comparison.

Build-up time of fluctuations and cross-correlations

Motions in proteins occur on a very broad spectrum of time-scales, ranging from femtoseconds to seconds (McCammon & Harvey, 1987). Only the fastest motions (up to a nanosecond) are accessible by current MD simulations. Some insight into the time-scales of subnanosecond backbone motions can be gained by examining the time development of simulated B -factors. Typical examples for BPTI are given in Figure 5a for residues 50 to 55 (C-terminal part of the α -helix) and in Figure 5b for residues 3 to 7 (3^{10} helix). In agreement with previous work (Swaminathan *et al.*, 1982), it is found that for all residues, the local backbone librations are sampled on a subpicosecond time-scale, giving rise to B -factors which are very uniform as a function of the residue sequence number (between 0.73 and 1.69 \AA^2 for all C^α atoms at 1 ps but already spread between 1.27 and 9.47 \AA^2 at 5 ps). These motions in the picosecond range form the main contribution to the B -factors when calculated from a relatively short simulation (Ichiye & Karplus, 1991). This is no longer

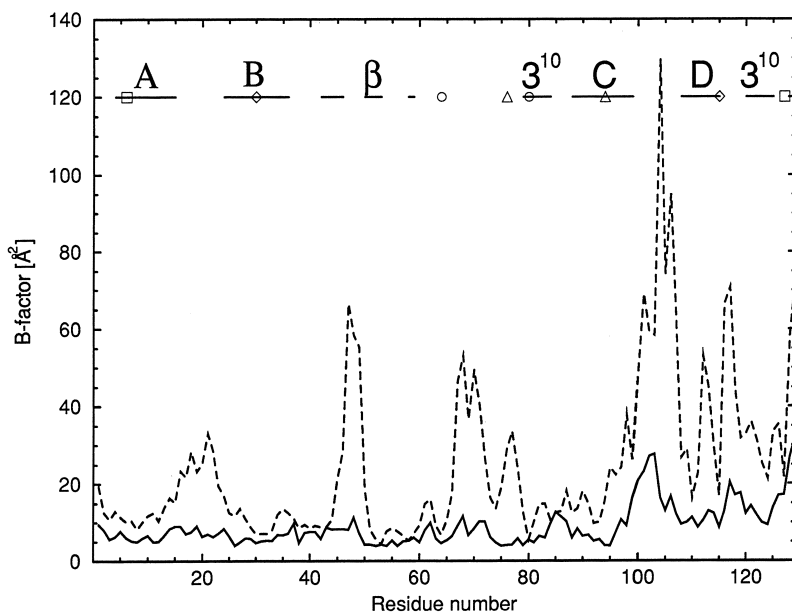


Figure 3. Simulated B -factors for the C^α atoms of HEWL, calculated according to equation (2) over the 1050 ps period from 50 to 1100 ps (---), together with B -factors obtained from crystallographic structure refinement based on X-ray data (Ramanadham *et al.*, 1987) (—). Secondary structure elements and disulfide bonds are indicated at the top (see also Figure 1b): \square , [6–127]; \diamond , [30–115]; \circ , [64–80] and \triangle , [76–94].

the case in nanosecond simulations, where slower, more collective motions clearly constitute the dominant contribution to the simulated B -factors. C^α fluctuations over 5 ps contribute on average only 13% of the ones calculated over 1.35 ns.

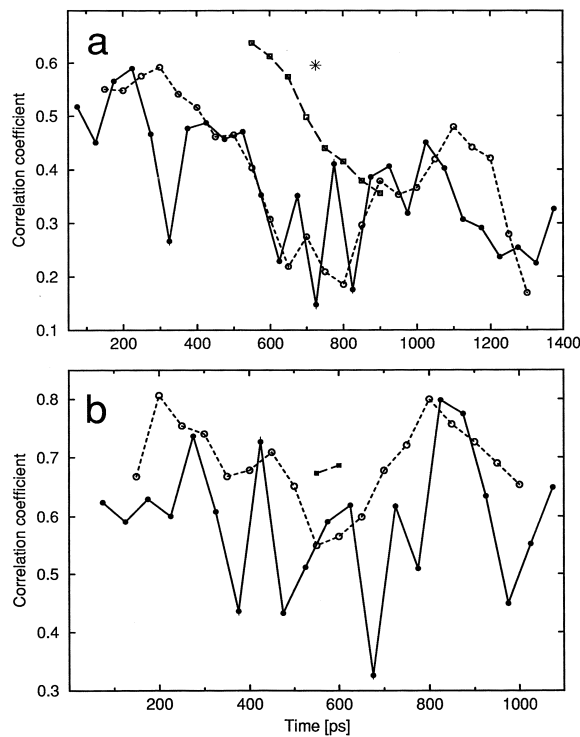


Figure 4. Correlation coefficient assuming a linear relationship between the C^α simulated B -factors and those obtained from crystallographic structure refinement based on X-ray data: a, for the BPTI simulation; and b, for the HEWL simulation, averaged over time windows of length 50 ps (\bullet), 200 ps (\circ), 1000 ps (\blacksquare) and 1350 ps (*). Successive time windows are shifted by 50 ps.

The time-domain between a few picoseconds and about 500 ps is characterized for most residues by a stepwise and asynchronous increase in the

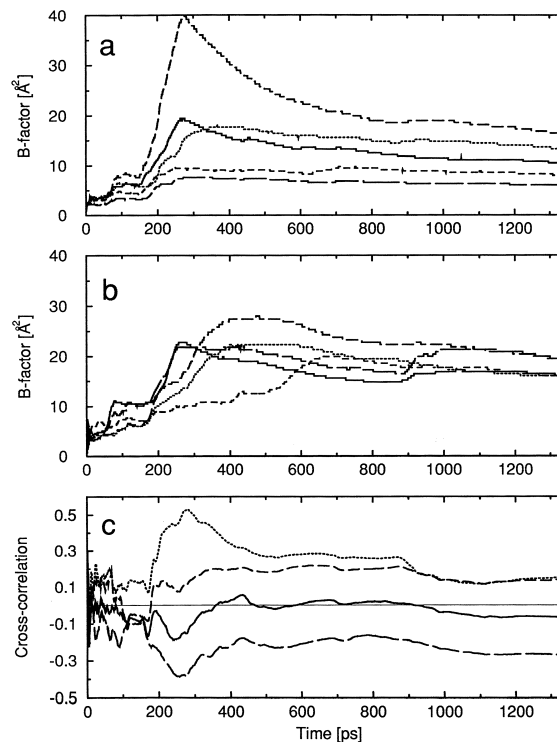


Figure 5. Time development of C^α simulated B -factors for the BPTI simulation, calculated starting after 50 ps (equilibration): a, for α -helix residues 50 (---), 51 (---), 53 (—), 54 (---) and 55 ($\cdot\cdot\cdot$); and b, for 3^{10} helix residues 3 (—), 4 (---), 5 (---), 6 ($\cdot\cdot\cdot$) and 7 (---). In c, the time development of cross-correlation coefficients of equation (3) between C^α atom pairs surrounding disulfide bridge [5–55] i.e. atom pairs 5–52 (---), 5–53 (—), 5–54 (---) and 5–55 ($\cdot\cdot\cdot$) is shown.

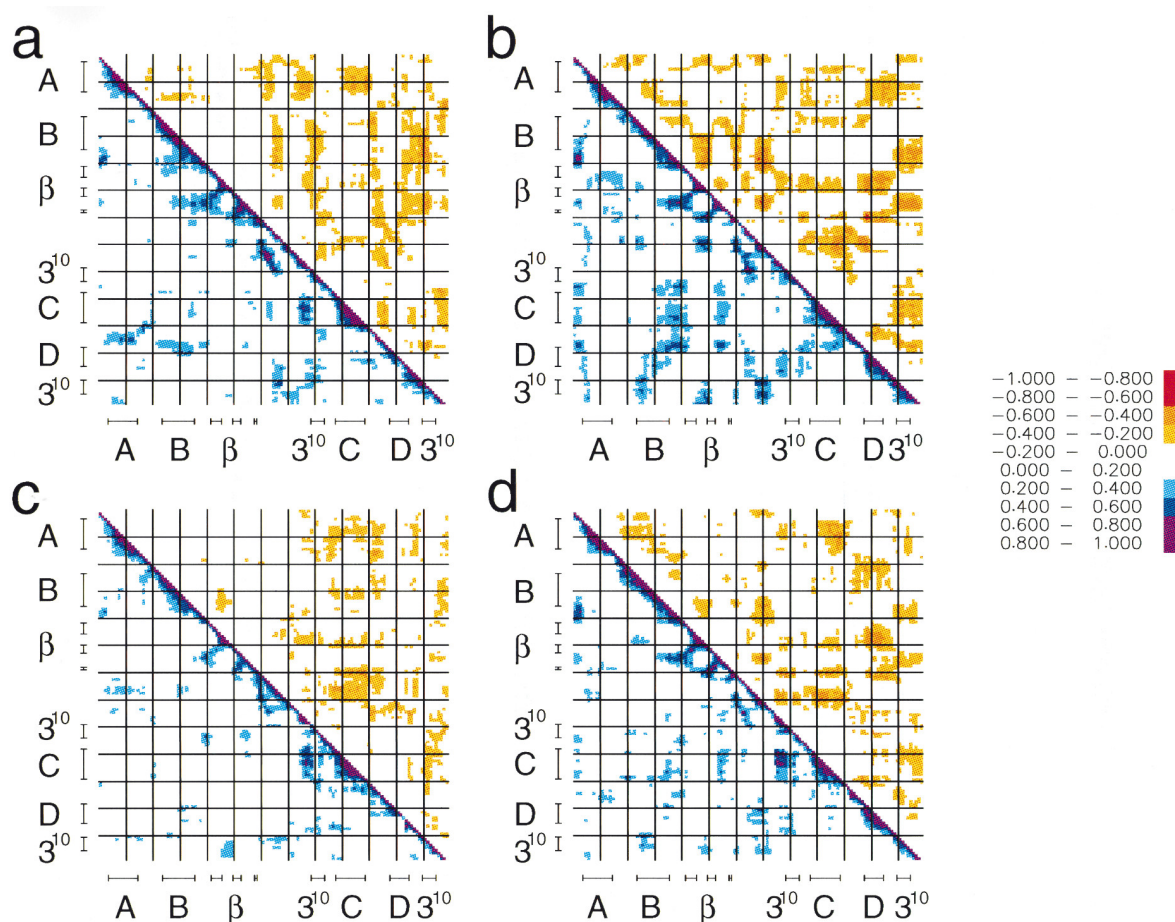


Figure 6. Dynamical cross-correlation maps for the C^z atom pairs of HEWL according to equation (3), using time averaging periods of: a, $\langle 100$ to 300 ps); b, $\langle 400$ to 600 ps); c, $\langle 600$ to 800 ps); and d, $\langle 100$ to 1100 ps). Positive correlations are in the lower left triangle, negative correlations in the upper right triangle. Only correlation coefficients with an absolute value greater than 0.2 are displayed. For the interpretation of the major features, see Harte *et al.* (1992). Secondary structure elements are indicated (see also Figure 1b).

fluctuations, possibly (e.g. Figure 5a, residue 54) with overshooting of the proper mean values. The two parts of BPTI presented in Figure 5a and b differ, however, with respect to the long-time behavior. Whereas the fluctuations in the α -helix (Figure 5a) seem to be largely converged after about 800 ps, the fluctuations in the 3^{10} helix (Figure 5b) still show a jump after 800 ps.

In Figure 5c, the time development of the cross-correlations between the positional fluctuations of selected atom pairs belonging to the residues surrounding disulfide bridge (5–55) linking the two previously described regions, are displayed. The behavior shown is typical for most of the correlations between pairs of atoms within the protein. Erratic behavior is observed between 0 and about 200 ps, with 37% of the 1711 C^z – C^z correlations changing sign at least once between 50 and 200 ps. Absolute values of the cross-correlation coefficients remain relatively low. From this time-point on, cross-correlations begin to diverge and after about 500 ps are converging towards their nanosecond values.

Cross-correlations

If the atomic fluctuations of the system are not sampled in a reproducible way the cross-correlations between the fluctuations of two atoms clearly will also not be reproducible. A dynamical cross-correlation matrix thus depends on the time-scale over which data on the correlations were collected. It has been reported that maps calculated using times of 10 ps or less display few off-diagonal correlations (Harte *et al.*, 1992; Komeiji *et al.*, 1994). But are the correlations observed on longer time-scales reliable? Three typical examples of cross-correlation matrices calculated for the lysozyme trajectory using a sampling time of 200 ps are presented in Figure 6a to c. For comparison, the map calculated over 1 ns is presented in Figure 6d. Positive correlations are shown in the lower left triangle, negative in the upper right triangle. In all maps, helices give rise to positive correlation triangles along the diagonal. Three plumes of positive correlation emerging from the diagonal around 42 to 46, 50 to 53 and 58 to 59 are characteristic for the triple-stranded β -sheet. A

positive correlation zone surrounds disulfide bond regions 64–80 and 76–94, although the former is almost absent in Figure 6c and the latter is surrounded by a negative correlation zone in Figure 6b. The two other disulfide bonds do not appear clearly in any of the maps. The hydrogen-bonded bridge linking Val2 and Asn39 throughout the simulation also gives rise to a positively correlated zone (only weak in Figure 6c). None of these observations is surprising. Correlation due to covalent bonds or hydrogen bonds is a well-known effect and does not provide much useful information about the dynamics of the system. Very similar information is available from a simple distance map. Unfortunately, it is difficult to draw any other conclusion consistent with all the maps presented in Figure 6. Correlation patterns can be very pronounced, as in Figure 6b, or relatively weakly differentiated, as in Figure 6c (although the simulated *B*-factors corresponding to both maps are very similar). Examples where a correlation is positive in one map and negative in another are not rare. Helix A is positively correlated with the loop between helices C and D in Figure 6a, but negatively in Figure 6c. The same is observed for loop 61 to 78 versus helices C, D and the C-terminal 3^{10} helix. Helix B is positively correlated with helix D and the C-terminal 3^{10} helix in Figure 6b, but negatively in Figure 6a. The β -domain 3^{10} helix is positively correlated with helix D in Figure 6b, but negatively in Figure 6c. The results presented here question the validity of comparing dynamical cross-correlation matrices calculated from distinct simulations, such as simulations of a free protein against that to which a ligand is bound, or a native protein against a mutant.

Influence of the fitting procedure

Fitting the structure of a reference conformation in order to remove rigid-body motion is a necessary initial procedure when performing fluctuation analysis. Translational and rotational diffusion of the protein would otherwise result in spurious covariances between atom pairs and, over long simulation times, the (interesting) internal fluctuations would be biased or disappear in the noise. There is unfortunately no unique procedure for performing a “physically meaningful” fit. It may seem better to perform the fitting with respect to backbone atoms rather than side-chain atoms, because the former are less subject to random fluctuations, but such choices are subjective.

In a previous work involving BPTI (Ichiye & Karplus, 1991) it was reported that cross-correlations first decrease with increasing interatomic distances to become negative, then increase again for distances longer than 1.5 nm, reaching values up to 0.5 at the maximum backbone distance of 3.0 nm. The authors considered these correlations as being of most interest. The effect is clearly evident in Figure 7a, which displays cross-correlations as a function of interatomic distance when the fit is performed on all

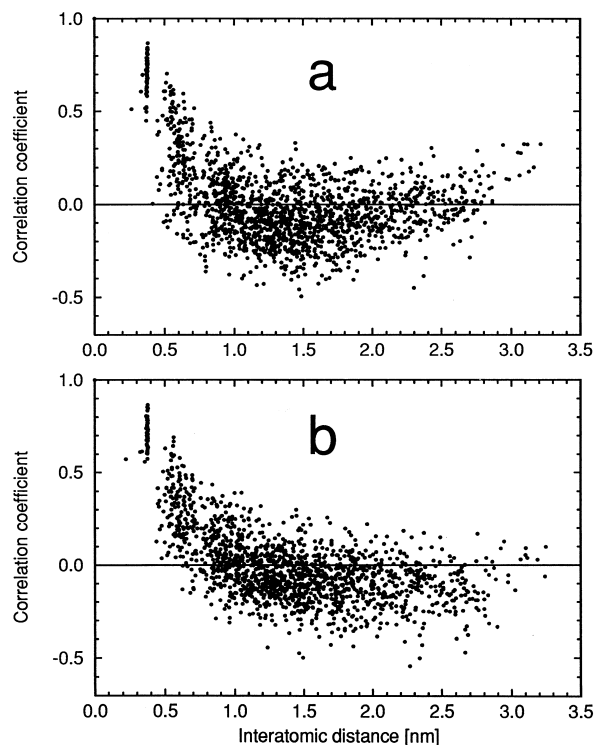


Figure 7. Cross-correlation coefficients for the C_{α} atoms of BPTI calculated from the MD simulation over the period from 50 to 1400 ps as a function of the average interatomic distance: a, with least-squares fitting of consecutive structures using all C_{α} atoms; and b, with least-squares fitting of consecutive structures using 22 low mobility (secondary structure) C_{α} atoms (residues 19 to 24, 29 to 35, 44 to 52).

backbone C^{α} atoms for the whole (1.35 ns) current BPTI trajectory. The coefficients start from unity at a distance of zero, stay large (about 0.5 to 0.9) for C^{α} atoms separated by a single peptide group along the chain and then decay rapidly towards negative values, becoming positive again for interatomic distances beyond 2.5 nm. The effect is, however, no longer evident in Figure 7b, which displays the analogous distribution with the fit performed on a subset of 22 C^{α} atoms with small fluctuations during the simulation. The initial decay is slower (less pairs with negative correlations between 0.5 and 1.0 nm) and the trend towards positive correlations for large interatomic distances has disappeared. These long-range correlations and those observed by Ichiye & Karplus (1991), who report fitting on backbone atoms, are thus likely to be artifactual. Clearly, if a loop moves clockwise at the surface of the protein, any fitting procedure involving all backbone atoms will induce a (smaller) counter-clockwise rotation of the rest of the protein. This will in turn generate covariances between the loop and the rest of the protein proportional to the distance of the atoms under consideration from the center of mass and the negative cosine of the angle they describe with the latter. If this value is large in comparison to the covariance induced by internal motions, it will result in a net change in the cross-correlation coefficient,

positive for surface atoms lying at opposite sides of the molecule, negative for atoms lying at the same side. Calculation of simulated B -factors is also affected by the fitting procedure. The restricted fitting procedure leads to more resolved peaks, whereas the fitting on all atoms enables low mobility residues to pick up some of the motion of more mobile residues.

Conclusion

In the present work, two nanosecond MD trajectories of BPTI and lysozyme in explicit solvent are analyzed in terms of backbone fluctuations and cross-correlations of the latter between atom pairs. It is found that although the simulations are stable, simulated B -factors determined even on a time-scale of 500 ps give an unreliable image of the atomic motions. This is especially true for the most mobile residues. Correlation with experimental B -factors increases with increasing averaging time. However, on a nanosecond time-scale, the B -factors calculated from a solvent simulation are considerably larger than those derived from X-ray crystallographic electron density maps. Although excessive mobility due to the use of a force field developed and tested using shorter time-scales cannot be ruled out, the results suggest that, due to the very different environments, time-scales and determination techniques involved, a meaningful comparison between the B -factors calculated from a solvent simulation and experimental crystallographic B -factors may not be possible. Moreover, the fact that fluctuations monitored early in the simulation are apparently more representative of the initial X-ray structure than of the solvent environment make such a comparison an unreliable criterion for assessing the quality or the convergence of a simulation of a protein in solution.

From the time-evolution of the simulated B -factors, it is evident that atomic positional fluctuations in large sections of the proteins studied do not converge within a nanosecond. If the positional fluctuations of individual atoms have not converged, clearly, the cross-correlations between the fluctuations of atom pairs will not be reliable. In fact, such correlations display very erratic behavior when the sampling is less than about 200 ps. Of course fluctuations and cross-correlations can and do converge more rapidly within a local frame of reference such as elements of secondary structure. The slow convergence properties of atomic fluctuations and cross-correlations are likely to propagate into the convergence properties of any derived quantity. Thus, a critical examination of the dependence of a property of interest on the averaging time should be an integral part of analyses such as quasiharmonic or essential mode determination and entropy calculation from the covariance matrices.

It is also shown that special care should be taken with respect to the fitting procedure used to remove rigid-body motion from the trajectory. Fitting using a set of atoms that includes the most mobile ones in

the protein leads to spurious cross-correlations between surface atoms.

The results presented here should not be unexpected. From nuclear magnetic resonance experiments, it is known that significant motions in proteins occur on a time-scale of microseconds or longer. This is not to say that fluctuation and cross-correlation analysis cannot be very useful tools in the elucidation of protein function and dynamics. It should be remembered that the only observables which can validly be compared to experiment are those for which the relaxation or build-up time are within the time-scale reachable by computer simulation, which currently is in the order of nanoseconds. Reliable results will only be obtained if the relevant motions are adequately sampled.

Methods

Molecular model and computational procedure

Both simulations were performed using a modified version of the standard GROMOS87 force field (van Gunsteren & Berendsen, 1987), using an increased value of the C_{12} parameter for the interaction between carbon and water oxygen (Mark *et al.*, 1994). Non-polar hydrogen atoms were treated using a united atom carbon model. Aromatic hydrogen atoms were handled explicitly with C/H atomic partial charges of $-0.14e/+0.14e$ (Smith *et al.*, 1995). The disulfide bonds were not reduced. Atomic charges were chosen appropriate to pH 7.0 (total charge $+6e$ for BPTI and $+8e$ for HEWL).

The crystal structure of BPTI was taken from the entry 4PTI (Wlodawer *et al.*, 1987) of the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977). The protein was placed in a truncated octahedron box corresponding to a cube of edge length 5.40 nm containing 2371 extended simple point charge (SPC/E) water molecules (Berendsen *et al.*, 1987) resulting in a total of 7717 atoms. The system was equilibrated during 10 ps at 277 K using the standard GROMOS87 force field. The aromatic hydrogen atoms were then generated and the system was allowed to relax for a further 40 ps at 300 K with the improved force field parameters.

The crystal structure of triclinic HEWL (Ramanadham *et al.*, 1987), entry 2LZT of the Brookhaven Protein Data Bank, served as a starting structure for the lysozyme simulation. The protein was placed in a truncated octahedron box corresponding to a cube of edge length 6.71 nm containing 4463 simple point charge (SPC) water molecules (Berendsen *et al.*, 1981), resulting in a total of 14,710 atoms. The system was equilibrated during 50 ps with progressive relaxation of positional restraints and regular increase in the temperature from 250 K to 300 K.

For both simulations, bond lengths were constrained by application of the SHAKE procedure (Ryckaert *et al.*, 1977) with a relative tolerance of 10^{-4} . Non-bonded interactions were handled by means of a twin-range method (van Gunsteren & Berendsen, 1990), using short and long range cut-off radii of 0.8 and 1.4 nm, respectively. The non-bonded pair list was updated every 10 fs. A time step of 2 fs was chosen for integrating the equations of motion. The temperature was maintained at 300 K by weakly coupling the protein and the water separately to an external temperature bath (Berendsen *et al.*, 1984) with a coupling constant $\tau_T = 0.1$ ps. The pressure was held by first-order

relaxation (weak coupling) to 1 atm using isotropic coordinate scaling (Berendsen *et al.*, 1984), with a coupling constant $\tau_p = 0.5$ ps. The coordinates were saved for analysis every 0.1 ps for the BPTI simulation (total 1.4 ns) and every 0.5 ps for the HEWL simulation (total 1.1 ns).

Analysis procedure

Analysis of the two trajectories was performed in terms of B -factors, B_i ; covariance matrices, c_{ij} ; and dynamical cross-correlation matrices (DCCM) C_{ij} , over various averaging time windows. These concepts are defined below. Unless specified otherwise, all configurations were translated and rotated by means of a least-square-fitting procedure using all backbone C^α atoms to align on the equilibrated starting configuration.

If the position vectors of two atoms i and j in the (fitted) structure at time t are $\mathbf{r}_i(t)$ and $\mathbf{r}_j(t)$, respectively, the corresponding covariance matrix element, c_{ij} , is calculated as:

$$c_{ij} = \langle (\mathbf{r}_i - \langle \mathbf{r}_i \rangle)(\mathbf{r}_j - \langle \mathbf{r}_j \rangle) \rangle = \langle \mathbf{r}_i \mathbf{r}_j \rangle - \langle \mathbf{r}_i \rangle \langle \mathbf{r}_j \rangle$$

$$= \frac{\Delta t}{t_{\text{aver}}} \left[\sum_{t=0}^{t_{\text{aver}} - \Delta t} \mathbf{r}_i(t) \mathbf{r}_j(t) - \frac{\Delta t}{t_{\text{aver}}} \left(\sum_{t=0}^{t_{\text{aver}} - \Delta t} \mathbf{r}_i(t) \right) \right. \\ \left. \times \left(\sum_{t=0}^{t_{\text{aver}} - \Delta t} \mathbf{r}_j(t) \right) \right] \quad (1)$$

where angle brackets denote time averages; t_{aver} is the averaging time and Δt the time interval between two frames (0.1 ps for BPTI and 0.05 ps for HEWL). The diagonal elements of the covariance matrix ($i=j$) correspond to the mean-square (m.s.) fluctuations of atom i . They can be converted to simulated atomic B -factors, B_i , using the relation (e.g. see Trueblood, 1992; Kidera & Go, 1992; Kidera *et al.*, 1992):

$$B_i = (8\pi^2/3)c_{ii} = (8\pi^2/3)(\langle \mathbf{r}_i^2 \rangle - \langle \mathbf{r}_i \rangle^2) \quad (2)$$

Covariances can be used to estimate the entropy of the system (Karplus & Kushick, 1981; Schlitter, 1993), in the interpretation of diffuse scattering experiments (Caspar *et al.*, 1988; Caspar & Badger, 1991; Clarage *et al.*, 1992), in quasiharmonic mode determination (see Case, 1994) and in "essential" mode determination (Amadei *et al.*, 1993). The cross-correlation (or normalized covariance) matrix elements, C_{ij} , are defined by:

$$C_{ij} = \frac{c_{ij}}{c_{ii}^{1/2} c_{jj}^{1/2}} = \frac{\langle \mathbf{r}_i \mathbf{r}_j \rangle - \langle \mathbf{r}_i \rangle \langle \mathbf{r}_j \rangle}{[(\langle \mathbf{r}_i^2 \rangle - \langle \mathbf{r}_i \rangle^2)(\langle \mathbf{r}_j^2 \rangle - \langle \mathbf{r}_j \rangle^2)]^{1/2}} \quad (3)$$

Cross-correlation coefficients range from a value of -1 (completely anticorrelated motions) to a value of $+1$ (completely correlated motions). Unlike covariance matrix elements, they do not bear any information about the magnitude of the motions, which can be small local oscillations as well as large scale collective motions. Both matrices will reflect correlation of displacements along a straight line. In other words, two atoms moving exactly in phase and with the same period, but along perpendicular lines will have a cross-correlation and covariance of zero.

Acknowledgements

Financial support from Unilever N.V., Vlaardingen, The Netherlands, is gratefully acknowledged, as is Dr Florian

Müller-Plathe for helpful discussions and advice, and a referee for his or her constructive comments.

References

- Amadei, A., Linssen, A. B. M. & Berendsen, H. J. C. (1993). Essential dynamics of proteins. *Proteins: Struct. Funct. Genet.* **17**, 412–425.
- 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* (Pullman, B., ed.), pp. 331–342, Reidel, Dordrecht.
- 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., Grigera, J. R. & Straatsma, T. P. (1987). The missing term in effective pair potentials. *J. Phys. Chem.* **91**, 6269–6271.
- Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., 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.
- Brooks, C. L., III, Karplus, M. & Pettitt, B. M. (1988). Proteins: a theoretical perspective of dynamics, structure, and thermodynamics. *Advan. Chem. Phys.* **71**, 1–259.
- Brünger, A. T., Brooks, C. L., III, & Karplus, M. (1985). Active site dynamics of ribonuclease. *Proc. Natl Acad. Sci. USA*, **82**, 8458–8462.
- Case, D. A. (1994). Normal mode analysis of protein dynamics. *Curr. Opin. Struct. Biol.* **4**, 285–290.
- Caspar, D. L. D. & Badger, J. (1991). Plasticity of crystalline proteins. *Curr. Opin. Struct. Biol.* **1**, 877–882.
- Caspar, D. L. D., Clarage, J., Salunke, D. M. & Clarage, M. (1988). Liquid-like movements in crystalline insulin. *Nature*, **332**, 659–662.
- Caves, L. S. D., Nguyen, D. T. & Hubbard, R. E. (1990). Conformational variability of insulin: a molecular dynamics analysis. In *Molecular Dynamics: An Overview of Applications in Molecular Biology* (Goodfellow, J. M., ed.), pp. 27–68, MacMillan Press, New York.
- Clarage, J. B., Clarage, M. S., Phillips, W. C., Sweet, R. M. & Caspar, D. L. D. (1992). Correlations of atomic movements in lysozyme crystals. *Proteins: Struct. Funct. Genet.* **12**, 145–157.
- Daggett, V. & Levitt, M. (1993). Protein unfolding pathways explored through molecular dynamics simulations. *J. Mol. Biol.* **232**, 600–619.
- de la Cruz, X., Mark, A. E., Tormo, J., Fita, I. & van Gunsteren, W. F. (1994). Investigation of shape variations in the antibody binding site by molecular dynamics computer simulation. *J. Mol. Biol.* **236**, 1186–1195.
- Edwards, C., Palmer, S. B., Emsley, P., Helliwell, J. R., Glover, I. D., Harris, G. W. & Moss, D. S. (1990). Thermal motion in protein crystals estimated using laser-generated ultrasound and Young's modulus measurements. *Acta Crystallogr. sect. A*, **46**, 315–320.
- Elofsson, A. & Nilsson, L. (1993). How consistent are molecular dynamics simulations? Comparing structure and dynamics in reduced and oxidized *Escherichia coli* thioredoxin. *J. Mol. Biol.* **233**, 766–780.
- Frauenfelder, H., Petsko, G. A. & Tsernoglou, D. (1979). Temperature-dependent X-ray diffraction as a probe of protein structural dynamics. *Nature*, **280**, 558–563.
- Harte, W. E., Jr, Swaminathan, S., Mansuri, M. M., Martin, J. C., Rosenberg, I. E. & Beveridge, D. L. (1990).

- Domain communication in the dynamical structure of human immunodeficiency virus 1 protease. *Proc. Natl Acad. Sci. USA*, **87**, 8864–8868.
- Harte, W. E., Jr, Swaminathan, S. & Beveridge, D. L. (1992). Molecular dynamics of HIV-1 protease. *Proteins: Struct. Funct. Genet.* **13**, 175–194.
- Hayward, S., Kitao, A., Hirata, F. & Go, N. (1993). Effect of solvent on collective motions in globular protein. *J. Mol. Biol.* **234**, 1207–1217.
- Huber, R. & Bennett, W. S., Jr (1983). Functional significance of flexibility in proteins. *Biopolymers*, **22**, 261–279.
- Ichiye, T. & Karplus, M. (1991). Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins: Struct. Funct. Genet.* **11**, 205–217.
- Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
- Karplus, M. & Kushick, J. N. (1981). Method for estimating the configurational entropy of macromolecules. *Macromolecules*, **14**, 325–332.
- Kidera, A. & Go, N. (1992). Normal mode refinement: crystallographic refinement of protein dynamic structure. I. Theory and test by simulated diffraction data. *J. Mol. Biol.* **225**, 457–475.
- Kidera, A., Inaka, K., Matsushima, M. & Go, N. (1992). Normal mode refinement: crystallographic refinement of protein dynamic structure. II. Application to human lysozyme. *J. Mol. Biol.* **225**, 477–486.
- Komeiji, Y., Uebayasi, M., Someya, J. & Yamato, I. (1991). Molecular dynamics simulation of *trp*-apopressor in a solvent. *Protein Eng.* **4**, 871–875.
- Komeiji, Y., Uebayasi, M. & Yamato, I. (1994). Molecular dynamics simulations of *trp* apo- and holopressors: domain structure and ligand-protein interaction. *Proteins: Struct. Funct. Genet.* **20**, 248–258.
- Lee, B. & Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379–400.
- Levitt, M. & Sharon, R. (1988). Accurate simulation of protein dynamics in solution. *Proc. Natl Acad. Sci. USA*, **85**, 7557–7561.
- Levitt, M., Sander, C. & Stern, P. S. (1985). Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. *J. Mol. Biol.* **181**, 423–447.
- Mark, A. E., van Helden, S. P., Smith, P. E., Janssen, L. H. M. & van Gunsteren, W. F. (1994). Convergence properties of free energy calculations: α -cyclodextrin complexes as a case study. *J. Am. Chem. Soc.* **116**, 6293–6302.
- McCammon, J. A. (1984). Protein dynamics. *Rep. Prog. Phys.* **47**, 1–46.
- McCammon, J. A. & Harvey, S. C. (1987). *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, Cambridge.
- Murphy, K. P., Bhakuni, V., Dong Xie & Freire, E. (1992). Molecular basis of co-operativity in protein folding. III. *J. Mol. Biol.* **227**, 293–306.
- Ptaszek, L. M., Vijayakumar, S., Ravishanker, G. & Beveridge, D. L. (1994). Molecular dynamics studies of the human CD4 protein. *Biopolymers*, **34**, 1145–1153.
- Ramanadham, M., Sieker, L. C. & Jensen, L. H. (1987). SRSQ refinement of triclinic lysozyme. *Acta Crystallogr. sect. A*, **43**, 13 (Suppl.).
- 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 Crystallogr. sect. D*, **51**, 85–92.
- Schlitter, J. (1993). Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem. Phys. Letters*, **215**, 617–621.
- Smith, L. J., Mark, A. E., Dobson, C. M. & van Gunsteren, W. F. (1995). Comparison of MD simulations and NMR experiments for hen lysozyme: analysis of local fluctuations, cooperative motions and global changes. *Biochemistry*, in the press.
- Swaminathan, S., Ichiye, T., van Gunsteren, W. F. & Karplus, M. (1982). Time dependence of atomic fluctuations in proteins: analysis of local and collective motions in bovine pancreatic trypsin inhibitor. *Biochemistry*, **21**, 5230–5241.
- Swaminathan, S., Harte, W. E., Jr & Beveridge, D. L. (1991). Investigation of domain structure in proteins via molecular dynamics simulation: application to HIV-1 protease dimer. *J. Am. Chem. Soc.* **113**, 2717–2721.
- Trueblood, K. N. (1992). Diffraction studies of molecular motions in crystals. In *Accurate Molecular Structures* (Domenicano, A. & Hargittai, I., eds), pp. 199–219, Oxford University Press, Oxford.
- van Gunsteren, W. F. & Berendsen, H. J. C. (1987). *Groningen Molecular Simulation (GROMOS) Library Manual (Biomos)*, Nijenborgh 4, 9747 AG Groningen, The Netherlands.
- van Gunsteren, W. F. & Berendsen, H. J. C. (1990). Computer simulation of molecular dynamics: methodology, applications and perspectives. *Angew. Chem. Int. Ed. Engl.* **29**, 992–1023.
- van Gunsteren, W. F., Hünenberger, P. H., Mark, A. E., Smith, P. E. & Tironi, I. G. (1995). Computer simulation of protein motion. *Comp. Phys. Commun.* in the press.
- Wlodawer, A., Deisenhofer, J. & Huber, R. (1987). Comparison of two highly refined structures of bovine pancreatic trypsin inhibitor. *J. Mol. Biol.* **193**, 145–156.

Edited by R. Huber

(Received 7 April 1995; accepted 27 June 1995)