PROTEINS: Structure, Function, and Genetics 31:116–127 (1998)

Domain Motions in Bacteriophage T4 Lysozyme: A Comparison Between Molecular Dynamics and Crystallographic Data

B.L. de Groot,[1] S. Hayward,[1] D.M.F van Aalten,[2] A. Amadei,[1] and H.J.C. Berendsen[1]*
[1]*Department of Biophysical Chemistry, Groningen Biomolecular Sciences and Biotechnology Institute, the University of Groningen, Groningen, The Netherlands*
[2]*Keck Structural Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York*

ABSTRACT **A comparison of a series of extended molecular dynamics (MD) simulations of bacteriophage T4 lysozyme in solvent with X-ray data is presented. Essential dynamics analyses were used to derive collective fluctuations from both the simulated trajectories and a distribution of crystallographic conformations. In both cases the main collective fluctuations describe domain motions. The protein consists of an N- and C-terminal domain connected by a long helix. The analysis of the distribution of crystallographic conformations reveals that the N-terminal helix rotates together with either of these two domains. The main domain fluctuation describes a closure mode of the two domains in which the N-terminal helix rotates concertedly with the C-terminal domain, while the domain fluctuation with second largest amplitude corresponds to a twisting mode of the two domains, with the N-terminal helix rotating concertedly with the N-terminal domain. For the closure mode, the difference in hinge-bending angle between the most open and most closed X-ray structure along this mode is 49 degrees. In the MD simulation that shows the largest fluctuation along this mode, a rotation of 45 degrees was observed. Although the twisting mode has much less freedom than the closure mode in the distribution of crystallographic conformations, experimental results suggest that it might be functionally important. Interestingly, the twisting mode is sampled more extensively in all MD simulations than it is in the distribution of X-ray conformations. Proteins 31:116–127, 1998.** © 1998 Wiley-Liss, Inc.

Key words: molecular dynamics; X-ray crystallography; essential dynamics; lysozyme; hinge bending

INTRODUCTION

The notion of domain motions in hen lysozyme, inferred from its X-ray structure,[1,2] is more than twenty years old.[3] Although bacteriophage T4 lysozyme (T4L) has a very different structure, the domain character of the protein is even more pronounced.[4] From the differences between crystallographic structures of various mutants of T4L it has been suggested that a hinge-bending mode of T4L is an intrinsic property of the molecule.[5–7] This hypothesis was recently qualitatively supported by studies of T4L in solution.[8] Also from computer simulations domain motions of the wild-type protein have been observed.[9,10] The domain fluctuations are predicted to be essential for the function of the enzyme, allowing the substrate to enter and the products to leave the active site. Crystallographic studies of a mutant T4L,[11] in which a substrate is covalently bound to the enzyme, suggest that the substrate-bound enzyme is locked in a state in which the two domains have closed around the substrate with respect to the unbound state. The unbound enzyme is expected to display a larger hinge-bending angle on average.

More than 200 T4L structures crystallized in more than 25 different crystal forms are present in the Protein Data Bank.[7] Assuming that each crystal structure represents a possible conformation in solution, this provides a unique experimental view on the conformational flexibility of the protein at atomic resolution. Information on conformational freedom of proteins is usually obtained from only a few experimental structures,[12–14] but dynamics of proteins is so complex that these few structures give only an extremely limited view of the dynamics involved. For T4L, the comparatively large number of different experimental conformations should provide us with a more detailed picture of its dynamical

Contract grant sponsor: Supercomputing Resource for Molecular Biology at the European Molecular Biology Laboratory funded by a European Union Human Capital and Mobility Access to Large-Scale Facilities grant; Contract grant number: ERBCHGECT940062.

A. Amadei's present address is Department of Chemistry, University 'La Sapienza,' Rome, Italy.

*Correspondence to: H.J.C. Berendsen, Department of Biophysical Chemistry, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), the University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands. E-mail: berends@chem.rug.nl

Received 15 July 1997; Accepted 18 November 1997

behavior which can then be sensibly compared to an MD simulation.[15] This provides the opportunity to assess the reliability of MD simulations.

T4L is a good system to study, not only for its large number of X-ray conformers but also because it is a rather small domain protein suitable for MD simulation. As domain proteins are usually relatively large, only few MD studies have been published in which domain motions were extensively studied.[9,16–21]

In this study, a detailed comparison is made between the collective (domain) fluctuations in T4L derived from the distribution of X-ray structures and from extensive MD simulations in solvent. Three simulations were conducted, each of one nanosecond, starting from different experimental structures. The essential dynamics (ED) analysis[22] was applied both to the distribution of X-ray and MD structures to separate small-amplitude fluctuations from large-amplitude global fluctuations. The largest-amplitude collective fluctuations from the X-ray distribution and from MD were subjected to domain and hinge-bending analyses[21,23] to monitor domain fluctuations. Collective fluctuations derived from MD can be expected to be affected by limited sampling[24–26] or imperfections in the interatomic interactions or force field. On the other hand, the crystallographic structures may not be representative of solvent-accessible conformations for the wild-type as they may be affected by the different mutations or by crystallization conditions and/or crystal contacts.[7] Despite these reservations a good correspondence between the MD results and X-ray analysis is obtained. Additionally, the detailed analyses of the domain fluctuations in T4L reveal interesting dynamical aspects that may be important for the function of the protein.

METHODS

MD Simulations

Three simulations were performed, each of one nanosecond. The first simulation, of the wild-type protein, started from a high-resolution X-ray structure[27] (PDB entry 2LZM). This simulation will from now on be referenced to as WT. The second simulation (M6I) was of the mutant M6I (methionine 6 replaced by isoleucine) and started from the X-ray structure with largest hinge-bending angle of this mutant[5] (PDB entry 150L, fourth conformer (MGI "D") hinge-bending angle 31 degrees more open than the WT X-ray structure). The coordinates of the three C-terminal residues not present in this crystal structure were taken from the most closed conformation from the same PDB entry. The third simulation started from the same structure, now mutated back to the wild type (WT*). All simulations were performed in a periodic box filled with SPC[28] water molecules (crystallographic water molecules were also included). Polar and aromatic hydrogens were added to the protein. In each of the simulated systems, 8 Cl^- ions were added to compensate the net positive charge on the protein. These ions were introduced by replacing water molecules with the highest electrostatic potential. This added up to a total of 19,195 atoms for the WT simulation and 17,101 for the M6I and the WT* simulation. Prior to the simulations, the structures were energy-minimized for 100 steps using a steepest-descents algorithm. Subsequently the structures were simulated for 10 ps with a harmonic positional restraint on all protein atoms (force constant of 1,000 kJ mol^{-1} nm^{-2}) for an initial equilibration of the water molecules. Production runs of 1 ns started from the resulting structures. All simulations were run at constant volume. The temperature was kept constant at 300 K by weak coupling to a temperature bath[29] (τ = 0.1 ps). A modification[30] of the GROMOS87[31] force field was used with additional terms for aromatic hydrogens[32] and improved carbon–oxygen interaction parameters.[30] SHAKE[33] was used to constrain bond lengths, allowing a time step of 2 fs. A twin-range cutoff method was used for nonbonded interactions. Lennard-Jones and Coulomb interactions within 1.0 nm were calculated every step, whereas Coulomb interactions between 1.0 and 1.8 nm were calculated every ten steps. All simulations were performed with the GROMACS simulation package.[34]

Analysis Techniques

Apart from conventional structural and geometrical analyses to assess the stability of the structures during the simulation, ED[22] analyses were utilized to study large concerted motions. The method yields the directions in configurational space that best describe concerted atomic fluctuations and is related to principal component analysis and quasiharmonic analyses.[35–39] It consists of diagonalization of the covariance matrix of atomic fluctuations, after removal of overall translation and rotation. Resulting eigenvectors are directions in configurational space that represent collective motions. Corresponding eigenvalues define the mean square fluctuation of the motion along these vectors. The method can be applied to any (sub)set of atoms using any set of structures.[22]

An ED analysis was performed on a cluster of X-ray crystallographic structures. Only structures from different crystal forms were included in the analysis. Zhang et. al.[7] described 25 different crystal forms. From their list, a set of 21 PDB entries was constructed, including 38 structures. These entries include 149L,[40] 152L,[40–42] 169L,[7] 172L,[41] 176L,[7] 179L,[41] 2LZM,[27] 137L,[43,44] 150L,[5] 167L,[7,41] 170L,[7] 173L,[7] 177L,[7] 1L97,[6] 151L,[7,40] 168L,[7] 171L,[7] 174L,[7] 178L,[7,41] 216L,[43] and 148L.[11] ED analyses were performed on the cartesian coordinates of the main chain N, C-α and C coordinates. Residues 163 and 164 were excluded from the analysis because their coordinates were absent in many of the PDB entries.

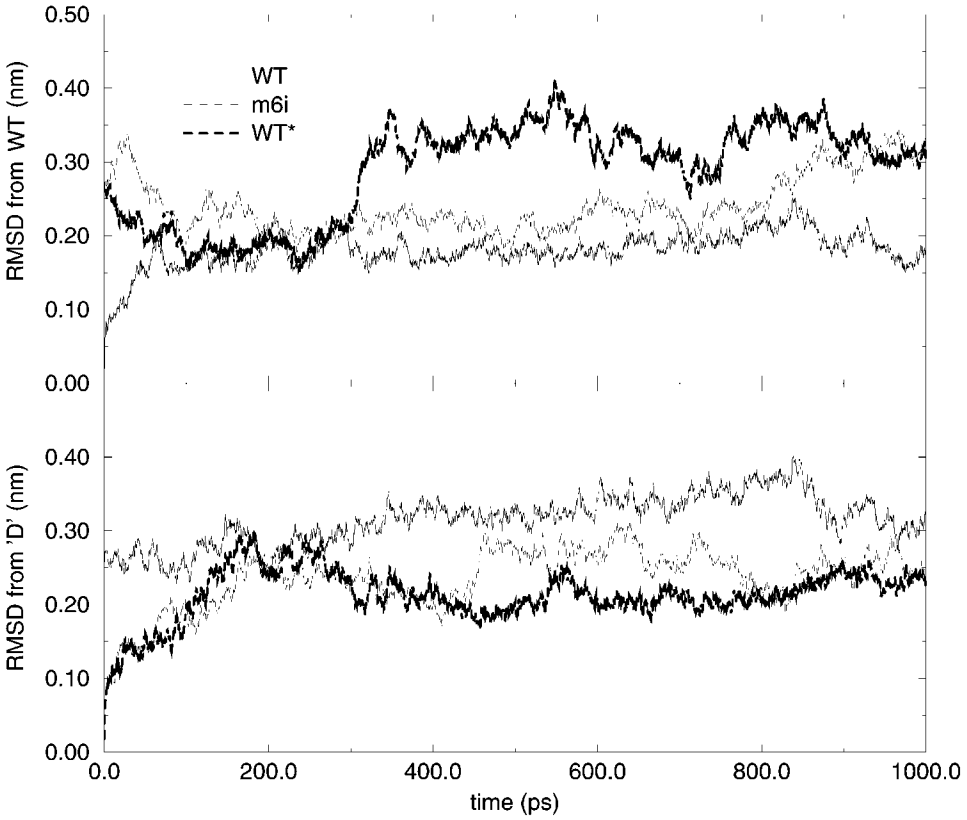

Fig. 1. Root mean square deviatation of C-? atoms from the WT X-ray structure (**upper panel**) and from the most open M6I ('D') X-ray structure (**lower panel**).

The same atoms were used in the ED analyses of the MD simulations. Analyses were performed on each individual MD trajectory (as the potential energies appeared to stabilize in less than 100 ps, the first 100 ps of each trajectory were disregarded) and on a combination of the three simulations. In this combination, the three simulations were not simply concatenated, because the eigenvectors would then be influenced by the differences between the average (starting) structures of each simulation. To remove the bias caused by these static differences, only the fluctuations from the average structure in each simulation were taken into account. This analysis implies the approximation that there are no systematic differences between the individual simulations. This combined analysis will be referenced to as MD_ALL.

ED analyses were carried out using the WHAT IF program.[45] Domains and hinge axes were identified and characterized using the DYNDOM program.[21,23] The method analyzes conformational changes in terms of rotational properties. Dynamic domains are identified by clustering each residue's rotation vector in a particular collective mode of motion.

RESULTS

Figure 1 shows the root mean square deviation (RMSD) during the three free simulations with respect to the WT X-ray structure and to the most open M6I X-ray structure. Deviations from the respective starting structures are relatively large, suggesting large structural fluctuations. The difference between the two starting structures (0.26 nm) is approximately as large as the drifts from the starting structures in each simulation.

Atomic fluctuations in the set of X-ray structures were compared to the crystallographic B-factors averaged over the 38 experimental structures and to the atomic fluctuations calculated from the MD simulations (Fig. 2). There is poor correspondence between the average B-factors and the atomic fluctuations in the distribution of X-ray structures (correlation coefficient of 0.55), but there is good correspondence between the atomic fluctuations in the X-ray and MD distribution (correlation coefficient of 0.85).

Figure 3 shows the eigenvalues of the ED analyses of the set of X-ray structures and of the combination of the three MD simulations (MD_ALL). The eigenvalue curve is very steep in the X-ray analysis, with the first eigenvector contributing 86% to the total

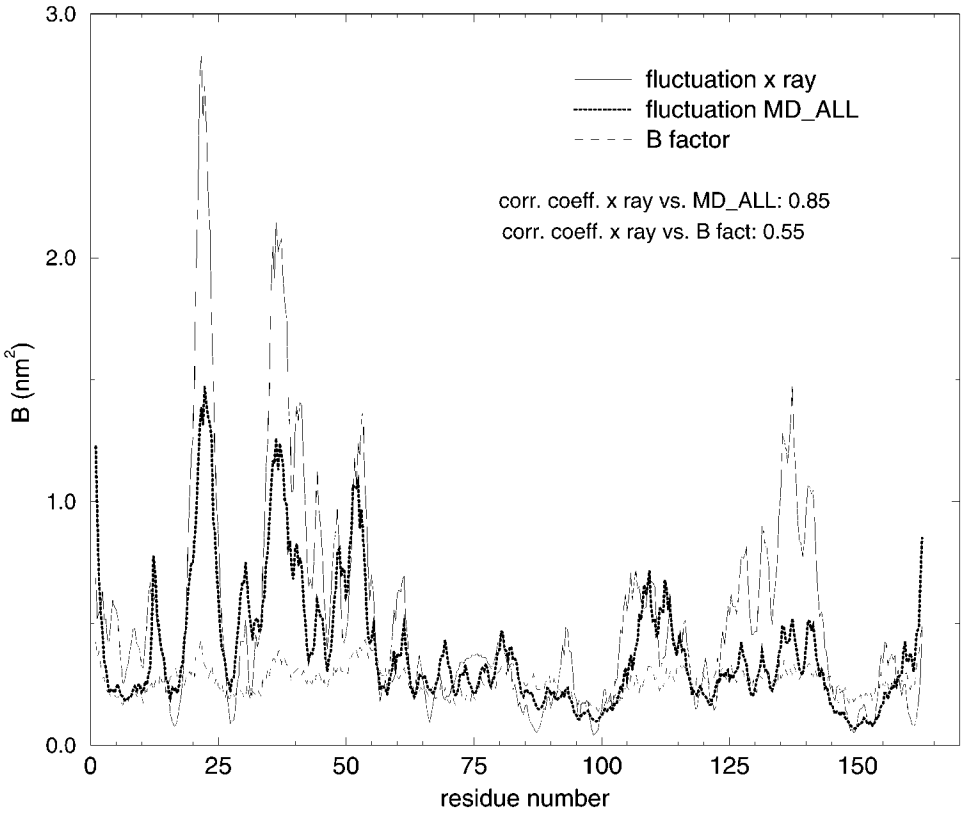

Fig. 2. Atomic fluctuations (expressed in isotropic B-factors: $B = (\Delta r)^2 \cdot 8\pi^2/3$, with $(\Delta r)^2$ being the calculated atomic mean square fluctuation) of main chain atoms in the X-ray cluster compared to the B-factors averaged over the 38 crystal structures and to the atomic fluctuations averaged over the three MD simulations.

mean square fluctuation. For MD_ALL, the eigenvalue curve is less steep and therefore more eigenvectors are required to achieve the same level of approximation of the total mean square fluctuation.

The domain analysis[21,23] was performed on the motions along single eigenvectors to ascertain whether these main modes of correlated fluctuation correspond to domain motions. Table I and Figure 4 show that the two most dominant of these modes extracted from the distribution of X-ray structures clearly correspond to the motion of two quasi-rigid bodies with respect to each other.[23] For both modes there are two distinguishable domains. The C-terminal domain is largest and ranges from approximately residue 75 to the C-terminus. The smaller N-terminal domain ranges from approximately residue 13 to 65. The first ten N-terminal residues are not statically part of the N- or C-terminal domain, but fluctuate correlated with either of the two domains: with the C-terminal domain in the first eigenvector and with the N-terminal domain in the second. The transition between the N- and C-terminal domains is located between residues 65 and 75, in the middle of the interdomain helix. The flexible link between the first ten residues and the N-terminal domain consists of residues 11 and 12.

The assignment of residues to the domains given above was used to extract the axes around which the domains rotate with respect to each other. The calculated interdomain screw-axes are shown as arrows in Figure 4 for the first and second eigenvectors from the ED analysis of the X-ray cluster. Both axes are 'effective hinge axes'[23] as they pass near the residues shown to be involved in the interdomain motion (see Table I). The first eigenvector corresponds (mainly) to a closure motion[21] (defined by an effective hinge axis perpendicular to the line connecting the centers of mass of the two domains) (Table I, Fig. 4a). The angular difference between the most open (PDB entry 178L[41]) and the most closed configuration (PDB entry 152L) is as much as 47 degrees (Table I). From the clustering of the endpoints of the rotation vectors in Figure 4a it is visible that the ten N-terminal residues rotate together with the C-terminal domain. The second eigenvector consists (mainly) of a twisting of the two domains, with the effective hinge axis being more parallel to the line connecting the two centers of mass. (Table I, Fig. 4b). From the clustering of the atoms in Figure 4B it can be seen that the first ten residues now rotate more concertedly with the N-terminal domain.

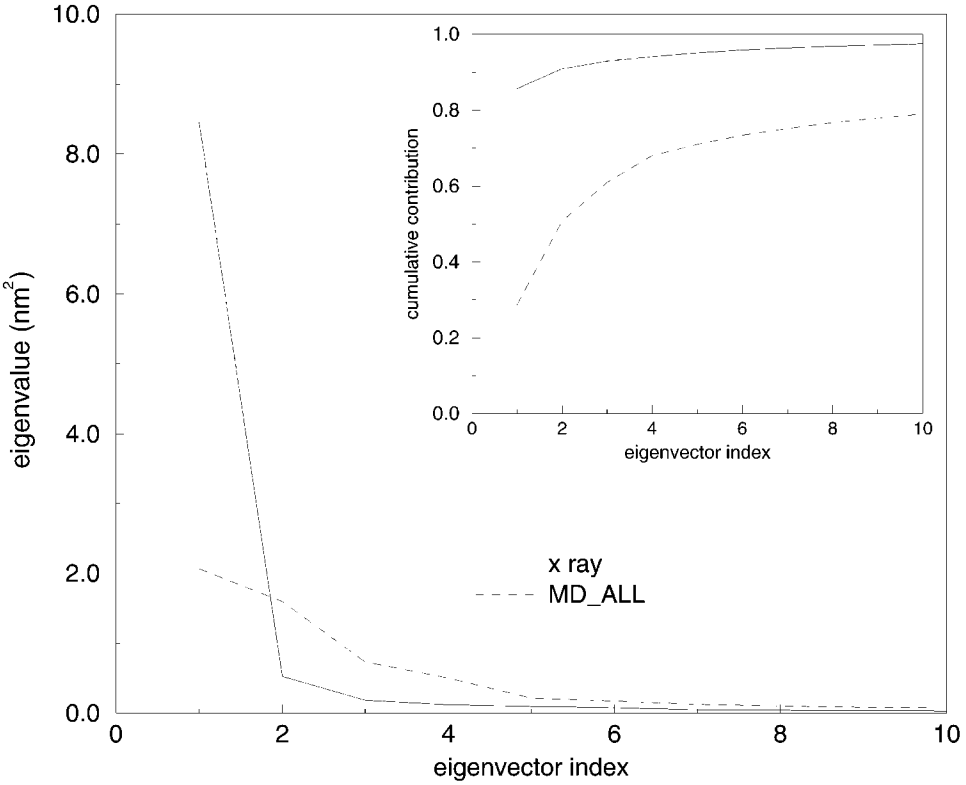

Fig. 3. Eigenvalues obtained from the essential dynamics analyses of the cluster of X-ray structures and of the combination of MD simulations. The **inset** shows the cumulative contribution of the eigenvectors to the total mean square fluctuation.

TABLE I. Domain Analyses of the Two Modes With Largest Amplitudes From X-Ray and MD_ALL[a]

	X-ray, e.v. 1	X-ray, e.v. 2	MD_ALL, e.v. 1	MD_ALL, e.v. 2
Domain A	14–66	1–65	15–63	12–66
Domain B	1–10, 81–162	74–162	1–12, 75–162	70–162
Connecting regions	11–13, 67–80	66–73	13–14, 64–74	1–11, 67–69
Angle of rotation	47.1	16.0	39.4	34.5
Residues near axis	12, 13, 29, 71, 74–76	6, 7, 49, 50, 66, 67, 160	13, 29, 59, 102, 119	12, 67, 69, 70
Angle ‖	66.5	37.1	29.2	75.3
% closure motion	84.1	36.5	23.9	93.5

[a]Residues were marked near to the effective hinge axis if their Cα atoms were found within 3 Å of the axis. Angle ‖ denotes the angle between the effective hinge axis and the line connecting the two centers of mass of the two domains.

Domains were identified also from the first two modes of the MD_ALL analysis (Table I, Fig. 5) and there is good correspondence with the domain demarcation obtained from the X-ray analysis. Again, residues 11–14 and 65–80 form the dynamical links between the two domains. The dynamic behavior of the N-terminal helix is less pronounced than in the analysis of X-ray structures. It is assigned to the C-terminal domain along the first mode (twist) and is identified as intermediate region along the second mode, which describes a closure mode.

All X-ray and MD structures were projected onto the plane spanned by the two eigenvectors with largest eigenvalue from the distribution of X-ray conformations to compare the kind and extent of fluctuation in the X-ray structures and MD (Fig. 6). All MD simulations fluctuate significantly in this plane, indicating that the main modes of collective fluctuation in the X-ray cluster are accessible during MD. This is in agreement with previous findings.[15] There are differences between the regions sampled in this plane by X-ray and MD, however. The WT simulation shows a 25-degree opening of the structure along the first X-ray eigenvector with respect to its starting configuration, but does not reach any of the most open configurations observed in the X-ray cluster. The M6I simulation starts from a more open configuration and closes 29 degrees, reaching a

hinge-bending angle almost equal to that of the WT X-ray structure. Both the M6I simulation and the WT simulation spend most of the time at a hinge-bending angle between 7 and 19 degrees more open than the WT X-ray structure. The WT* simulation initially closes and also reaches a conformation similar to that of the WT X-ray structure. After that it opens up again and reaches a conformation with a hinge-bending angle 45 degrees more open than that of the WT X-ray structure, slightly more open than the X-ray structure with largest hinge-bending angle. Along this first eigenvector there seem to be two distinct clouds in the cluster of X-ray structures with only two configurations in between. This is consistent with a two-state mechanism postulated on the basis of these structures.[7] The simulations do not support this hypothesis, however, and indicate that intermediate structures are equally accessible.

The position along the second X-ray eigenvector, which mainly describes a twisting mode, fluctuates uncoupled from the position along the first eigenvector, both in the X-ray cluster and in the three MD simulations. The amplitude of the fluctuation in this direction is larger in each of the three simulations than in the cluster of crystal structures.

Tables II and III list inner products between eigenvectors obtained from the cluster of X-ray structures and those obtained from MD. This provides a quantitative measure of the overlap in modes of motion derived from the two techniques. Table II shows that the two eigenvectors with largest eigenvalue from the X-ray analysis are to a large extent present in the space spanned by the first five eigenvectors obtained from each simulation. This means that the modes of domain motion extracted from the differences between the X-ray conformations are also among the most dominant ones in the simulations. It is interesting to note that the overlap between eigenvectors extracted from the combination of the three MD simulations (MD_ALL) and the X-ray eigenvectors is larger than the average of the overlaps between the X-ray eigenvectors and those extracted from each of three simulations individually. When the MD simulations are compared to each other in the same fashion (Table III), the overlap is on average lower than with the X-ray structures (Table II). Therefore, the main modes of motion derived from each of the MD simulations are more similar to the main collective fluctuations derived from the X-ray cluster than to those from the other MD simulations.

In order to compare with the qualitative results of Mchaourab et. al.,[8] fluctuations of the distances between selected pairs of α-carbon atoms were monitored along the two most prominent modes of collective fluctuation derived from the cluster of X-ray structures (Table IV). The pairs were selected to study the difference in conformation between the protein free in solution and covalently bound to a substrate. The fluctuations of the distances between pairs 35–137, 22–137, 4–71, and 4–60 are mainly ruled by the fluctuation along the eigenvector with largest eigenvalue, describing a closure motion. The observed spin–spin interactions[8] are consistent with a shift along the closure mode (toward closing) upon substrate binding. The distance between residues 35 and 109, however, hardly changes upon 'substrate release,' although a fluctuation along the closure mode significantly influences the distance between this pair. The distance between residues 22 and 109 does change upon 'substrate release,' but the fluctuation of the distance is much more connected with the twisting mode than with the closure mode, suggesting that substrate binding may also affect the twisting mode.

A Web page has been dedicated to the visualization of the dynamical information presented here (http://rugmd0.chem.rug.nl/~degroot/t4l.html).

DISCUSSION AND CONCLUSIONS

The collective fluctuations in T4L comprise, for the largest part, domain motions. The most dominant modes of fluctuation in the X-ray analysis as well as in all MD analyses correspond to external motions of the domains with respect to each other. Moreover, the main modes of fluctuation obtained from the cluster of X-ray structures are very similar to those obtained from simulation. The amount of overlap between X-ray and MD modes is larger than between modes of two similar MD runs. This is remarkable because it has been observed previously[24–26,46] that the definition of single eigenvectors in an ED analysis has not converged in simulations over time periods in the order of nanoseconds. A possible explanation for this phenomenon lies in the domain character of the protein, which causes two modes of domain motion to dominate over all other fluctuations. The domain fluctuations observed in the X-ray cluster are among the most extensively sampled directions in all MD simulations. The incomplete mutual overlap between MD modes is mainly due to insufficient sampling statistics, suggesting that longer MD simulations will show an even larger amount of overlap with the cluster of X-ray structures. The most important conclusion from the comparison of structural variability in X-ray and MD-generated structures is that MD indeed samples the important, physically relevant space, thus validating the MD method for application to protein dynamics.

The domain fluctuations in the MD simulations indicate that both the wild-type protein and the M6I mutant fluctuate significantly along the domain modes derived from the X-ray cluster. This is consistent with the hypothesis by Zhang et. al.[7] that domain motions are an intrinsic property of the T4L molecule. The results by Mchaourab et. al.[8] further support this finding. From the simulated data there

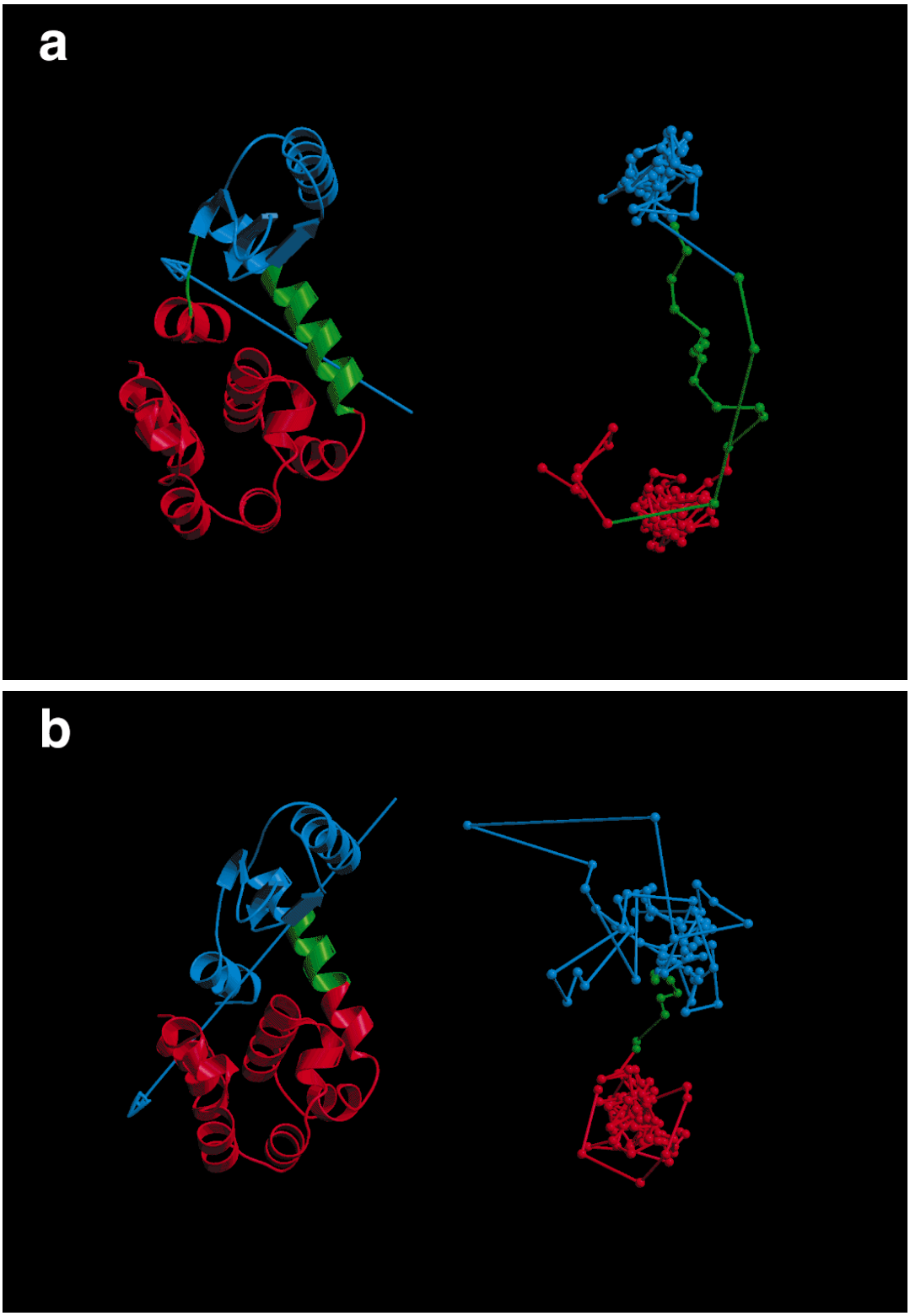

Fig. 4. Backbone structure (**left**) and rotation vectors (**right**) of T4L. End points, depicted as beads, of rotation vectors per residue (the beads are connected as in the amino acid sequence) were used to identify the domains. The colors indicate the different clusters (domains) that were assigned (red, blue) and the interdomain regions (green) based on the rotation vectors. The arrow indicates the direction of rotation of the red domain relative to the blue domain by the thumb rule of the right hand. **a:** Eigenvector 1 from the X-ray cluster. Displayed is the most open conformation with the arrow indicating the closure motion. **b:** Eigenvector 2 from the distribution of X-ray conformations. Domain analyses were performed by DYNDOM.[23] Plots were made with MOLSCRIPT[47] and Raster3D.[48,49]

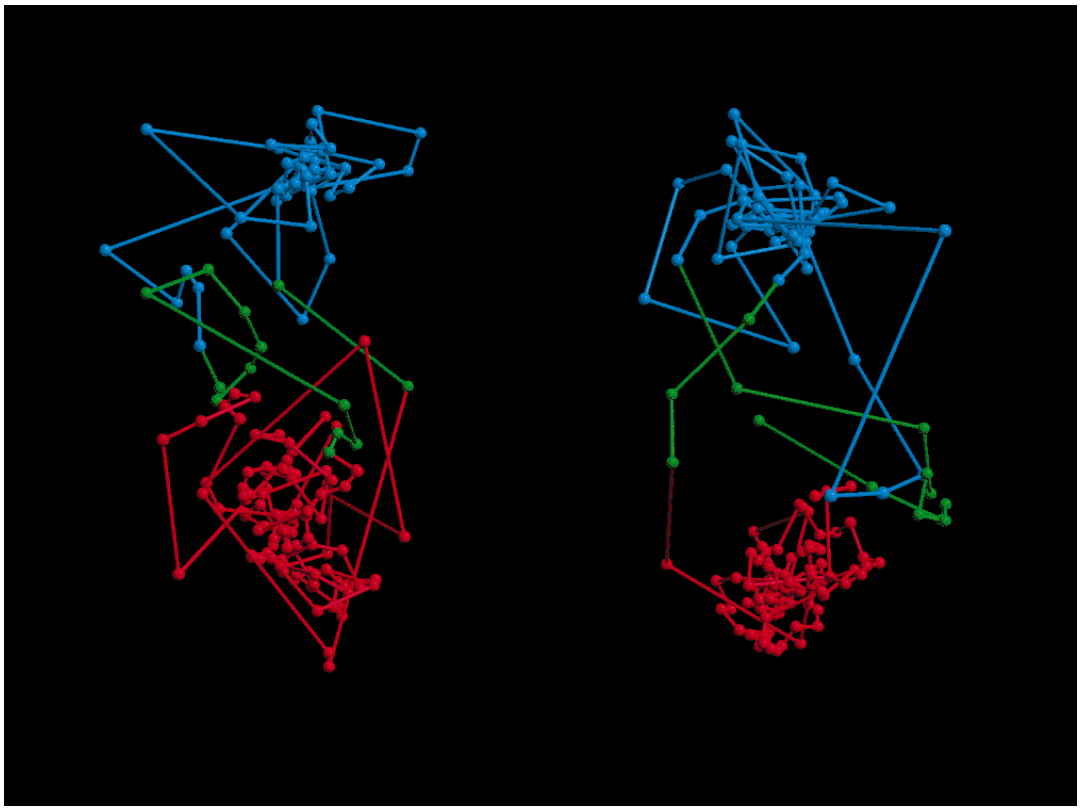

Fig. 5. Domain identification from the first (**left**) and second (**right**) mode from MD_ALL. As in Figure 4, colors indicate the different clusters (domains) that were assigned (red, blue) and the interdomain regions (green) from the rotation vectors. Domain analyses were performed by DYNDOM.[23] This plot was made with MOLSCRIPT[47] and Raster3D.[48,49]

is no evidence for the proposed two-state mechanism[7] for the main hinge-bending mode. The WT and M6I simulations do show a preference for intermediate hinge-bending angles for this mode (angles between 7 and 19 degrees more open than the WT X-ray structure), but the WT* simulation also indicates that more open configurations are easily accessible. Since there is no topological difference between the WT and the WT* simulations, a lack of sufficient sampling seems the most probable cause for the apparent difference between these simulations. Since also the differences between the M6I simulation and the WT and WT* simulations are not larger than the difference between the WT and WT* simulations, the conclusion that the hinge bending properties of the M6I mutant are close to those of the WT protein seems justified. This supports our assumption that the combination of the three MD trajectories for ED analysis (MD_ALL) is valid.

In a recent study, Arnold and Ornstein also presented results from MD simulations on native T4L and the M6I mutant.[10] They found that in all their simulations the protein went to a more compact conformation and concluded that a conformation more closed than the WT crystal structure would be the most stable configuration in solution. These findings are not supported by our results. We observe that in all simulations, the large majority of sampled conformations displays a more open conformation than the WT X-ray structure. A possible explanation of this apparent discrepancy is the difference between simulation protocols used. We have used a periodic box filled with a large number of solvent molecules (approximately 5,000), whereas Arnold and Ornstein used a shell of solvent containing approximately 2,200 water molecules. Protein dynamics in simulations using a shell of solvent molecules might be affected by surface tension effects in such small droplets, resulting in unrealistically compact structures. Since in both cases three simulations have been performed, with consistent results, limited statistics can probably be ruled out as a possible explanation for this observation. Interestingly, Arnold and Ornstein reported that the conformational change toward more compact structures did reveal the domain character of the protein, suggesting once again that domain motions are among the most prominent collective fluctuations of T4L.

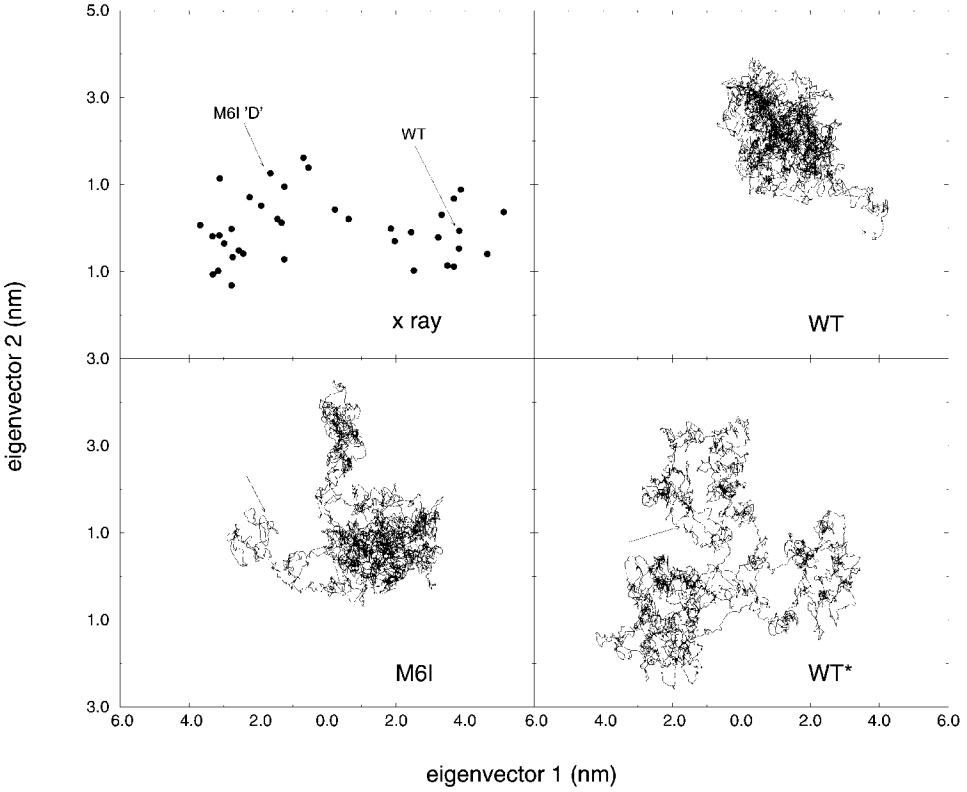

Fig. 6. Projections (in nm) onto the plane spanned by the two eigenvectors with largest eigenvalues extracted from the cluster of X-ray structures. **Upper left panel:** X-ray structures. **Upper right panel:** Structures from the WT simulation. **Lower left panel:** Structures from the M6I simulation. **Lower right panel:** Structures from the WT* simulation. The arrows indicate the starting structures of each simulation. In the horizontal direction, structures differ from each other along the closure mode (structures to the left are more open than those on the right); the vertical direction depicts the twisting mode.

TABLE II. Summed Squared Inner Products Between One Eigenvector of One Set and the First Five of Another: MD Eigenvectors Compared to X-Ray Eigenvectors

	X-ray eigenvector	
e.v. 1–5	1	2
WT	0.72	0.69
M6I	0.80	0.81
WT*	0.91	0.76
MD_ALL	0.92	0.77

TABLE III. Summed Squared Inner Products Between One Eigenvector of One Set and the First Five of Another: MD Eigenvectors From Different Simulations Compared to Each Other

	WT eigenvector			WT* eigenvector			M6I eigenvector	
e.v. 1–5	1	2	e.v. 1–5	1	2	e.v. 1–5	1	2
M6I	0.34	0.40	WT	0.64	0.54	WT	0.65	0.60
WT*	0.61	0.45	M6I	0.85	0.76	WT*	0.58	0.66

The domain modes obtained from MD and the cluster of X-ray structures are essentially similar (Figs. 4 and 5., Tables I and II). The protein consists of two domains: an N-terminal domain comprising residues 15 to 65 and a C-terminal domain that ranges from residue 80 to the C-terminus. Residues 70–75, residing in the C-terminal half of the interdomain helix, form the dynamical bridge between the two domains. The behavior of the ten N-terminal residues is complex. In the main domain fluctuation derived from the X-ray cluster, mainly a hinge-bending mode describing a closure motion between the two domains, this N-terminal helix rotates con-

TABLE IV. Fluctuation of Distances Between Pairs of α-Carbon Atoms Along the First (Closure) and Second (Twist) Collective Mode of Fluctuation Derived From the Cluster of X-Ray Structures[a]

Pair	Spin labeling, upon substrate binding distance	X-ray, fluctuation affected by
35–137	Decreases	Closure
22–109	Increases	Twist (closure)
22–137	Decreases	Closure
4–71	Decreases	Closure
4–60	Increases	Closure
35–109	—	Closure (twist)

[a]Selection of the pairs after Mchaourab et al.[8]

certedly with the C-terminal domain. Along the collective fluctuation with second largest amplitude however, which mainly consists of a twisting of the two domains, this helix appears to be part of the N-terminal domain. The two main modes of collective fluctuation obtained from MD basically form a linear combination of the first two modes from the X-ray cluster. Therefore, the dynamical behavior of the N-terminal helix is influenced by both the N- and C-terminal domains in these modes and the assignment to either domain is less evident (Fig. 4, Table I). Thus, the N-terminal helix is not a static part of either of the two domains but rather adapts its dynamical behavior to the kind of domain motion. Upon opening, contacts with residues 93–97 and the C-terminal residues push the N-terminal helix away from its original position. The flexible loop connecting it to the N-terminal domain (the rotation is concentrated around GLU11 and GLY12) allows it to move concertedly with the C-terminal domain along the closure mode. The absence of such a steric effect in the twisting mode causes the helix to move concertedly with the N-terminal domain in this mode.

The large amount of overlap between the domain fluctuations in the cluster of X-ray structures and the MD simulations is the main reason for the close agreement of the atomic fluctuations in both clusters (Fig. 2). The much smaller correlation between the fluctuations in the cluster of X-ray structures and the averaged B-factors, together with the significantly lower average level of the B-factors, suggests that the main domain motions are significantly suppressed in most of the crystal environments included in this analysis. Although the pattern of thermal factors in some cases (especially those in 176L_A, 176L_B, and to a lesser extent also 2LZM (WT)) does suggest some degree of domain fluctuation,[27] we can conclude that, at least for flexible proteins, B-factors may be a less reliable indication of motional freedom in solution than fluctuations derived from MD.

Apart from the similarities between the fluctuations in MD and the X-ray cluster, there are also a few discrepancies. One of the most striking differences is in the shapes of the eigenvalue curves (Fig. 3). For the X-ray cluster there is one dominating collective fluctuation (the closure mode), which accounts for 86% of the total fluctuation, and the first ten eigenvectors together represent 98% of the fluctuation. In MD, the first mode only contributes 29% to the total fluctuation and the first ten together represent 79%. This is not the result of the fact that there are far fewer structures present in the X-ray cluster (38) than in the MD cluster (27,000) (when a subset of 38 structures, equally spaced in time, is taken from the MD_ALL cluster, the first eigenvector contributes 32% and the first ten eigenvectors 85% to the total fluctuation). This indicates that in the MD, a larger number of collective fluctuations than in X-ray make a significant contribution to the total fluctuation. The difference in sampled regions in the two main directions from the X-ray cluster is illustrated by Figure 6. Both the WT and M6I simulations do not sample the complete range of hinge-bending angles along the main closure mode derived from the X-ray cluster. The WT* simulation, however, indicates that this is the result of limited sampling, since in this simulation almost the complete range that is present in the X-ray cluster is sampled in one nanosecond. For the eigenvector with the second largest eigenvalue derived from the cluster of X-ray structures, the twisting mode, the fluctuation in all three simulations is larger than in the X-ray cluster (Fig. 6). Limited sampling in MD cannot be the explanation for this observation since this direction is oversampled with respect to the X-ray cluster. Also, the effect of mutations in the cluster of X-ray structures is not likely to be the reason for this discrepancy since one could expect the mutations to result in a larger fluctuation, rather than smaller, with respect to the WT protein. If one assumes that in 25 different crystal forms all conformational freedom has been sampled, then only the effect of crystallization conditions and/or crystal contacts or the used force field in MD remain as possible explanations for this difference. Further studies (e.g., NMR) will be necessary to distinguish which is the main effect.

The investigation of the fluctuation of distances between selected pairs of α-carbon atoms (Table IV) shows that for four out of the six investigated pairs, the experimentally observed changes in distances in solution are in accordance with an opening along the closure mode upon transition from the substrate-bound state to the substrate-free state. The fluctuation of the distance between residues 22 and 109, which is found to change upon 'substrate release,' is more connected with the twisting mode than with the closure mode, however. This suggests that the twisting mode is affected by the presence of the substrate. Another distance, between residues 35 and 109, does not seem to change much upon 'substrate release,' but is affected substantially by the closure mode. A possible explanation for this observation is a partial compensation by a change along the twisting mode, which also makes a significant contribution to the fluctuation of this distance. This is a further indication that not only the closure motion but also the twisting mode is relevant for the function of this protein and that the two modes are concertedly involved in the dynamics of substrate binding. Interestingly, all MD simulations display a larger extent of fluctuation along the twisting mode than is observed in the cluster of X-ray structures (Fig. 6).

In summary, we conclude that T4 lysozyme exhibits a mixture of two hinge-bending modes (a closure and a twist), which are both involved in the dynamic

response to substrate binding. Furthermore, we have shown that MD simulations of this protein provide reliable predictions of its functional dynamics.

ACKNOWLEDGMENTS

S.H. is a recipient of a BIOTECH fellowship from the European Union.

REFERENCES

1. Blake, C.C.F., Koenig, D.F., Mair, G.A., North, A.C.T., Phillips, D.C., Sarma, V.R. Structure of hen egg-white lysozyme, a three dimensional fourier synthesis at 2 Å resolution. Nature 206:757–761, 1965.
2. Diamond, R. Real-space refinement of the structure of hen egg-white lysozyme. J. Mol. Biol. 82:371–391, 1974.
3. McCammon, J.A., Gelin, B., Karplus, M., Wolynes, P.G. The hinge-bending mode in lysozyme. Nature 262:325–326, 1976.
4. Matthews, B.W., Remington, S.J. The three-dimensional structure of the lysozyme from bacteriophage T4. Proc. Nat. Acad. Sci. U.S.A. 71:4178–4182, 1974.
5. Faber, H.R., Matthews, B.W. A mutant T4 lysozyme displays five different crystal conformations. Nature 348:263–266, 1990.
6. Dixon, M.M., Nicholson, H., Shewchuk, L., Baase, W.A., Matthews, B.W. Structure of a hinge-bending bacteriophage T4 lysozyme mutant, Ile3→ Pro. J. Mol. Biol. 227:917–933, 1992.
7. Zhang, X.-J., Wozniak, J.A., Matthews, B.W. Protein flexibility and adaptability seen in 25 crystal forms of T4 lysozyme. J. Mol. Biol 250:527–552, 1995.
8. Mchaourab, H.S., Oh, K.J., Fang, C.J., Hubbell, W.L. Conformation of T4 lysozyme in solution: Hinge-bending motion and the substrate-induced conformational transition studied by site-directed spin labeling. Biochemistry 36:307–316, 1997.
9. Arnold, G.E., Manchester, J.I., Townsend, B.D., Ornstein, R.L. Investigation of domain motions in bacteriophage T4 lysozyme. J. Biomol. Struct. Dyn. 12:457–474, 1994.
10. Arnold, G.E., Ornstein, R.L. Protein hinge bending as seen in molecular dynamics simulations of native and M6I mutant T4 lysozymes. Biopolymers 41:533–544, 1997.
11. Kuroki, R., Weaver, L.H., Matthews, B.W. A covalent enzyme-substrate intermediate with saccharide distortion in a mutant T4 lysozyme. Science 262:2030–2033, 1993.
12. Berry, M.B., Meador, B., Bilderback, T., Liang, P., Glaser, M., Phillips, Jr., G.N. The closed conformation of a highly flexible protein: The structure of *E. coli* adenylate kinase with bound AMP and AMPPNP. Proteins 19:183–198, 1994.
13. Waksman, G., Krishna, T.S.R., Williams, Jr., C.H., Kuriyan, J. Crystal structure of *eschericia coli* thioredoxin reductase refined at 2 Å resolution: Implications for a large conformational change during catalysis. J. Mol. Biol. 236: 800–816, 1994.
14. Sim, K.K., Song, H.K., Shin, D.H., Hwang, K.Y., Suh, S.W. The crystal structure of a triacylglycerol lipase from *pseudomonas cepacia* reveals a highly open conformation in the absence of a bound inhibitor. Structure 5:173–186, 1997.
15. Van Aalten, D.M.F., Conn, D.A., De Groot, B.L., Findlay, J. B.C., Berendsen, H.J.C., Amadei, A. Protein dynamics derived from clusters of crystal structures. Biophys. J. 73:2891–2896, 1997.
16. Verma, C.S., Caves, L.S.D., Hubbard, R.E., Roberts, G.C.K. Domain motions in dihydrofolate reductase: A molecular dynamics study. J. Mol. Biol. 266:776–796, 1997.
17. Kasper, P., Sterk, M., Christen, P., Gehring, H. Molecular dynamics simulation of domain movements in aspartate aminotransferase. Eur. J. Biochem. 240:751–755, 1996.
18. Komeiji, Y., Uebayasi, M., Yamato, I. Molecular dynamics simulations of trp apo- and holorepressors: Domain structure and ligand–protein interaction. Proteins 20:248–258, 1994.
19. Van der Spoel, D., De Groot, B.L., Hayward, S., Berendsen, H.J.C., Vogel, H.J. Bending of the calmodulin central helix: A theoretical study. Prot. Sci. 5:2044–2053, 1996.
20. Van Aalten, D.M.F., Amadei, A., Linssen, A.B.M., Eğsink, V.G.K., Vriend, G., Berendsen, H.J.C. The essential dynamics of thermolysin: Confirmation of hinge-bending motion and comparison of simulations in vacuum and water. Proteins 22:45–54, 1995.
21. Hayward, S., Kitao, A., Berendsen, H.J.C. Model free methods of analyzing domain motions in proteins from simulation: A comparison of normal mode analysis and molecular dynamics simulation of lysozyme. Proteins 27: 425–437, 1997.
22. Amadei, A., Linssen, A.B.M., Berendsen, H.J.C. Essential dynamics of proteins. Proteins 17:412–425, 1993.
23. Hayward, S., Berendsen, H.J.C. Systematic analysis of domain motions in proteins from conformational change: New results on citrate synthase and T4 lysozyme. Proteins 30:144–154, 1998.
24. Clarage, J.B., Romo, T., Andrews, B.K., Pettitt, B.M., Phillips, Jr., G.N. A sampling problem in molecular dynamics simulations of macromolecules. Proc. Nat. Acad. Sci. U.S.A. 92:3288–3292, 1995.
25. Balsera, M.A., Wriggers, W., Oono, Y., Schulten, K. Principal component analysis and long-time protein dynamics. J. Phys. Chem. 100:2567–2572, 1996.
26. De Groot, B.L., Van Aalten, D.M.F., Amadei, A., Berendsen, H.J.C. The consistency of large concerted motions in proteins in molecular dynamics simulations. Biophys. J. 71: 1554–1566, 1996.
27. Weaver, L.H., Matthews, B.W. Structure of bacteriophage T4 lysozyme refined at 1.7Å resolution. J. Mol. Biol. 193:189–199, 1987.
28. Berendsen, H.J.C., Postma, J.P.M., Van Gunsteren, W.F., Hermans, J. Interaction models for water in relation to protein hydration. In: 'Intermolecular Forces.' Pullman, B. (ed.). Dordrecht: D. Reidel Publishing Company, 1981:331–342.
29. Berendsen, H.J.C., Postma, J.P.M., DiNola, A., Haak, J.R. Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81:3684–3690, 1984.
30. Van Buuren, A.R., Marrink, S.J., Berendsen, H.J.C. A molecular dynamics study of the decane/water interface. J. Phys. Chem. 97:9206–9212, 1993.
31. Van Gunsteren, W.F., Berendsen, H.J.C. 'Gromos manual. BIOMOS, Biomolecular Software,' Groningen: Laboratory of Physical Chemistry, University of Groningen 1987.
32. Van Gunsteren, W., Billeter, S., Eising, A., et al. 'Biomolecular simulation: The GROMOS96 manual and user guide.' Groningen: Biomos b.v., 1996.
33. Ryckaert, J.P., Ciccotti, G., Berendsen, H.J.C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. J. Comp. Phys. 23:327–341, 1977.
34. Van der Spoel, D., Berendsen, H.J.C., Van Buuren, A.R., et al. 'Gromacs User Manual.' Groningen, 1995. Internet: http://rugmd0.chem.rug.nl/~gmx
35. Levy, R., Srinivasan, A., Olson, W., MsCammon, J. Quasi-harmonic method for studying very low frequency modes in proteins. Biopolymers 23:1099–1112, 1984.
36. Garcia, A.E. Large-amplitude nonlinear motions in proteins. Phys. Rev. Lett. 68:2696–2699, 1992.
37. Kitao, A., Hirata, F., Go, N. The effects of solvent on the conformation and the collective motions of protein: Normal mode analysis and molecular dynamics simulations of mellitin in water and in vacuum. Chem. Phys. 158:447–472, 1991.
38. Hayward, S., Kitao, A., Hirata, F., Go, N. Effect of solvent on collective motions in globular proteins. J. Mol. Biol. 234:1207–1217, 1993.
39. Hayward, S., Go, N. Collective variable description of native protein dynamics. Annu. Rev. Phys. Chem. 46:223–250, 1995.
40. Zhang, X.J., Matthews, B.W. Conservation of solvent-binding sites in 10 crystal forms of T4 lysozyme. Prot. Sci. 3:1031–1039, 1994.

41. Matsumura, M., Signor, G., Matthews, B.W. Substantial increase of protein stability by multiple disulphide bonds. Nature 342:291–293, 1989.
42. Jacobson, R.H., Matsumura, M., Faber, H.R., Matthews, B.W. Structure of a stabilizing disulfide bridge mutant that closes the active-site cleft of T4 lysozyme. Prot. Sci. 1:46–57, 1992.
43. Blaber, M., Zhang, X.J., Matthews, B.W. Structural basis of amino acid α-helix propensity. Science 260:1637–1640, 1993.
44. Blaber, M., Zhang, X.J., Lindstrom, J.D., Pepiot, S.D., Baase, W.A., Matthews, B.W. Determination of α-helix propensity within the context of a folded protein: Sites 44 and 131 on bacteriophage T4 lysozyme. J. Mol. Biol. 235:600–624, 1994.
45. Vriend, G. WHAT IF: a molecular modeling and drug design program. J. Mol. Graph. 8:52–56, 1990.
46. De Groot, B.L., Amadei, A., Scheek, R.M., Van Nuland, N.A.J., Berendsen, H.J.C. An extended sampling of the configurational space of hpr from *E. coli*. Proteins 26:314–322, 1996.
47. Kraulis, P.J. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. J. Appl. Cryst. 24:946–950, 1991.
48. Bacon, D.J., Anderson, W.F. A fast algorithm for rendering space-filling molecule pictures. J. Mol. Graph. 6:219–220, 1988.
49. Merritt, E.A., Murphy, M.E.P. Raster3D version 2.0: A program for photorealistic molecular graphics. Act. Cryst. D. 50:869–873, 1994.