

The Consistency of Large Concerted Motions in Proteins in Molecular Dynamics Simulations

B. L. de Groot,* D. M. F. van Aalten,# A. Amadei,* and H. J. C. Berendsen*

*Groningen Biomolecular Sciences and Biotechnology Institute, Department of Biophysical Chemistry, University of Groningen, 9747 AG Groningen, The Netherlands, and #Department of Biochemistry and Molecular Biology, University of Leeds, Leeds LS2 9JT UK

ABSTRACT A detailed investigation is presented into the effect of limited sampling time and small changes in the force field on molecular dynamics simulations of a protein. Thirteen independent simulations of the B1 IgG-binding domain of streptococcal protein G were performed, with small changes in the simulation parameters in each simulation. Parameters studied included temperature, bond constraints, cut-off radius for electrostatic interactions, and initial placement of hydrogen atoms. The essential dynamics technique was used to reveal dynamic differences between the simulations. Similar essential dynamics properties were found for all simulations, indicating that the large concerted motions found in the simulations are not particularly sensitive to small changes in the force field. A thorough investigation into the stability of the essential dynamics properties as derived from a molecular dynamics simulation of a few hundred picoseconds is provided. Although the definition of the essential modes of motion has not fully converged in these short simulations, the subspace in which these modes are confined is found to be reproducible.

INTRODUCTION

Recent studies have provided methods for revealing large concerted motions in proteins from molecular dynamics (MD) computer simulations (Garcia, 1992; Amadei et al., 1993; Hayward et al., 1993; van Aalten et al., 1995b, 1996). These methods divide the configurational subspace of proteins in a high dimensional subspace in which merely constraint-like motions of high frequency occur (which will be referred to from now on as the near-constraints subspace), and a low dimensional subspace in which all biologically relevant motions occur (the essential subspace). In this paper, we investigate how reproducible these two distinct spaces are in multiple simulations of one protein. The essential dynamics (ED) method, introduced by Amadei et al. (1993), is used to extract the definition of both subspaces from MD simulations. The sensitivity of the definition of these spaces toward different force-field parameters used in MD simulations, as well as the speed of convergence of the description of these subspaces, is examined. With this aim, four simulations of a test protein, the B1 IgG-binding domain of streptococcal protein G, were set up, each with one parameter different from seven reference simulations. Apart from these simulations, which were performed using explicit solvent, two simulations were run in vacuo. This protein was chosen because it is a small and fairly globular protein, containing both α -helix and β -strand secondary structure elements. Both x-ray (Gallagher et al., 1994) and NMR (Gronenborn et al., 1991) structures are available, as well as NMR relaxation data (Barchi et al., 1994).

Previous work (Amadei et al., 1993; van Aalten et al., 1995a, 1996) has suggested that a few hundred picoseconds is usually enough to obtain a rough approximation of the essential subspace of a small protein, although there is still an appreciable amount of noise present in the description of both subspaces after such limited sampling time. Here we use a set of 300-ps simulations as well as a 1-ns simulation to investigate the accuracy of the definition of both subspaces. The influence of a number of simulation parameters is also investigated. All simulations are compared to a set of six solvent simulations of 300 ps. These reference simulations were also compared to each other and to a 1-ns simulation of the same protein to gain insight into the convergence of the ED properties in such short simulations. Apart from comparison of properties derived from ED, conventional structural and geometrical properties were evaluated from the trajectories, to examine the stability and overall structural and dynamic behavior of the simulations.

METHODS AND THEORY

Simulation parameters

Simulations were performed with the GROMOS (van Gunsteren and Berendsen, 1987) simulation package. The simulations were started from the crystal structure (Protein Data Bank entry 1PGB; Gallagher et al., 1994). The protein was placed in a truncated octahedral box filled with simple point charge water (Berendsen et al., 1981), except for two simulations that were run in vacuo. The protein consists of 535 (united) atoms. Together with four sodium ions, which were used to compensate for the net charge of -4 (the ions were placed in the box by replacing water molecules at the lowest electrostatic potential) and approximately 1900 water molecules (the number of water molecules varied from simulation to simulation), the total number of atoms approximated

Received for publication 26 April 1996 and in final form 28 June 1996.

Address reprint requests to Dr. H. J. C. Berendsen, Department of Chemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands. Tel.: 31-50-3634323; Fax: 31-50-3634800; E-mail: berendsen@chem.rug.nl.

© 1996 by the Biophysical Society

0006-3495/96/10/1707/07 \$2.00

6500. After energy minimization, a HEATUP procedure (van Aalten et al., 1996) of 25 ps was performed to equilibrate the structure. In short, this involves a slow increase in the temperature, cut-off radius, and time step, combined with positional restraining. The simulations were then continued for 275 ps, of which the last 250 ps was used for ED analyses (all other analyses were performed on the full 275-ps trajectories, to include differences in the equilibration period). One simulation was extended to 1 ns. In total, 13 simulations have been performed and are identified below.

1. 275°K: This simulation was performed at a constant temperature of 275°K instead of 300°K. All simulations were kept at a constant temperature by coupling to an external temperature bath (Berendsen et al., 1984), using a coupling constant of 0.1 ps.

2. NO_SHAKE: This simulation was performed without SHAKE (Ryckaert et al., 1977), and covalent bond interactions were described by harmonic potentials. In this simulation, a time step of 1 fs was applied. In other simulations SHAKE was used to constrain bond lengths, allowing a time step of 2 fs.

3. CUT_OFF: This simulation was performed with a twin range cut-off method with radii of 10 and 14 Å instead of 8 and 10 Å for the other simulations. For the short range, the pair list was updated every time step; for the long range, this list was updated every 10 steps.

4. HPLACE: This simulation was started from a structure in which the positions of the hydrogens were generated using an algorithm that optimizes hydrogen bond networks throughout the structure (Hooft et al., 1996). Other simulations were started with standard GROMOS hydrogen placement, which uses standard hydrogen positioning.

5. REF_1 till REF_6: Six reference simulations were performed. These simulations differed in the initial velocities used.

6. REF_7: Identical to the other reference simulations, but this simulation was extended to 1 ns.

7. VAC_1 and VAC_2: Additionally, two simulations in vacuo were performed for comparison. VAC_1 was carried out with reduced charges to mimic the screening effect of the solvent; VAC_2 was performed with full charges.

Comparison techniques

Two types of techniques were used to identify differences between the simulations. First, a number of standard structural analyses were performed to check overall stability. Subsequently, ED analysis was used to compare the dynamic behavior of the protein in the different simulations. ED analyses were performed on each trajectory and compared to the reference simulations. Programs used were those available in the molecular modeling program WHAT IF (Vriend, 1990). Accessible surface calculations and secondary structure evaluations were performed by dictionary

of secondary structures in proteins (DSSP) (Kabsch and Sander, 1983).

Overlap between eigenvector sets

Overlap between multiple sets of eigenvectors is calculated by two methods. First, the overlap between two essential subspaces is calculated as the sum of all of the squared inner products between all pairs of eigenvectors from both essential subspaces, divided by the dimension of that space (see also the next subsection). This definition of the overlap has the disadvantage that it concentrates on similarities between two compared sets, and not on differences. Therefore, we have defined another measure of the overlap between two sets of eigenvectors. It is defined as the product of the square inner product and the difference in eigenvector index (i.e., a difference in relative contribution to the overall fluctuation; eigenvalues and corresponding eigenvectors are sorted according to decreasing value), averaged over all pairs of eigenvectors from both sets. This will result in a positive number, which will be close to zero if the sets are similar. This quantity can be regarded as a penalty function: a high inner product between an eigenvector with a high eigenvector index from one set and an eigenvector with a low eigenvector index from the other set gives a high contribution to this penalty. Significant differences in the dynamic behavior of two simulations (a motion that is accessible in one simulation but not in another) are therefore immediately evident.

Convergence of trajectories

ED analyses can be used to gain insight into the convergence of MD trajectories, because only a few coordinates are usually required to describe the relevant dynamics of a protein in MD simulations. Here, the overlap between essential subspaces (as defined by the 10 eigenvectors with the largest eigenvalues) obtained from pairs of simulations is taken as a measure of the similarity of two trajectories in terms of collective motions. This overlap is defined as the cumulative mean square inner product between 10 eigenvectors obtained from one set of eigenvectors and 10 from another. This results in a number between zero, when there is no overlap, and one, when the two sets are identical. In practice, the lower limit for this value is not zero, even if the actual overlap between the two essential subspaces is negligible, because there will always be some projection of the eigenvectors of one set into the essential subspace of the other. The overlap of real interest is not the overlap between two sets of eigenvectors obtained from MD, each containing noise, but the overlap of one of such sets with the fully converged set of eigenvectors, as would be obtained after infinite simulation time. This overlap is underestimated by the method described above, because in the comparison of two MD eigenvector sets, both sets contain an appreciable amount of noise, making the overlap smaller than in the case where only one of the sets contains this noise.

RESULTS

Results from structural analyses are summarized in Table 1. Apart from the two simulations in vacuo, the observed average properties in the simulations that were performed with different parameters do not differ significantly from those observed in the 300-ps reference simulations. When these properties are plotted as a function of time (data not shown), no significant drift is observed in any of the simulations (apart from those performed in vacuum). Hence, all solvent simulations are stable in terms of these properties in a time window of 300 ps. The simulation performed at lower temperature (275°K) does show a lower total mean square fluctuation than most other simulations, as expected, but one of the reference simulations (REF_1) shows an even lower total fluctuation. This suggests that the spread in the observed fluctuation for the reference simulations of 300 ps covers the difference that might have been caused by the lower temperature.

Compared to other proteins (van Aalten et al., 1995b, 1996; Scheek et al., 1995; Amadei et al., 1993), the total sum of fluctuations and the largest eigenvalues are relatively small, indicating a rather rigid molecule. This is in agreement with recent studies (Kuszewski et al., 1994; Gronenborn et al., 1991), in which this domain was reported as highly stable.

ED analyses were performed on each trajectory. Only α carbon coordinates were used in the covariance analysis. It has been shown that this approach identifies all large-scale concerted motions in proteins (Amadei et al., 1993). It has the advantage over an all-atom analysis (besides saving CPU time in all analyses) that backbone dynamics equilibrates faster than the dynamics in the full coordinate space (apparent correlations between backbone and side-chain motions introduce noise in an ED analysis on all atoms of a simulation of a few hundred picoseconds).

As an illustration of the typical overlap between two eigenvector sets obtained from 300-ps simulations in sol-

vent, an inner products matrix is shown in Fig. 1 A for two eigenvector sets obtained from REF_1 and REF_2. All high inner products are found close to the diagonal, meaning that directions in configurational space have a similar amount of freedom in the two simulations. The same qualitative picture is found for all combinations of the solvent simulations. Fig. 1 A shows that the eigenvectors spanning the essential subspace (e.g., defined as the first 10 eigenvectors) of one set of eigenvectors show the largest inner products with the essential eigenvectors extracted from another simulation, and that the projections outside the essential subspace are mainly concentrated in those eigenvectors

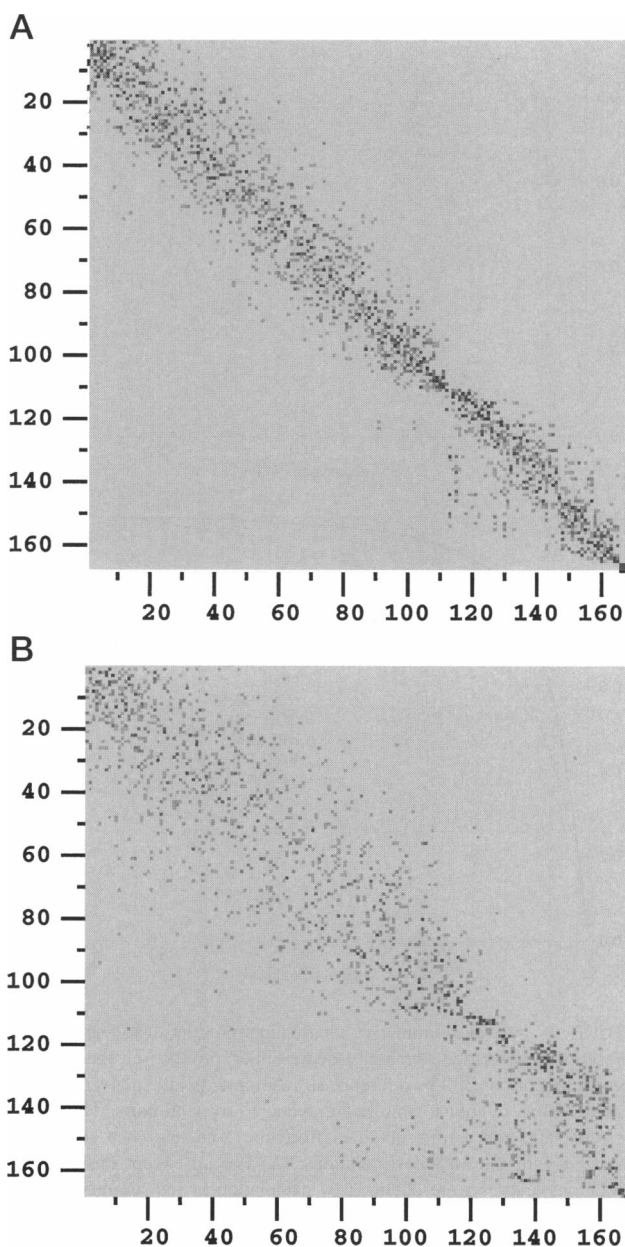


FIGURE 1 Squared inner products matrices. (A) Inner products between eigenvectors extracted from REF_1 (y axis) and REF_2 (x axis). (B) Inner products between REF_1 (y axis) and VAC_2 (x axis).

TABLE 1 Structural properties

Type	σ^2	RMSD	NRC	HBO	ACC	GYR
275K	0.244	1.19	8.04	44.6	3773	1.018
NO_SHAKE	0.337	1.02	8.95	42.9	3879	1.025
CUT_OFF	0.406	1.15	9.63	44.9	3856	1.015
HPLACE	0.366	1.23	10.61	43.3	3884	1.025
REF_1	0.242	0.93	10.80	41.1	3761	1.016
REF_2	0.351	1.12	8.42	45.8	3840	1.016
REF_3	0.366	1.27	8.40	45.6	3883	1.021
REF_4	0.394	1.11	9.10	43.9	3777	1.018
REF_5	0.365	1.28	8.29	46.0	3849	1.023
REF_6	0.331	1.24	10.70	42.4	3904	1.025
REF_7	0.532	1.45	10.03	44.0	3850	1.024
VAC_1	0.399	1.79	10.47	50.5	3607	0.997
VAC_2	1.261	3.83	18.22	35.3	3600	1.031

σ^2 , total mean square fluctuations (nm^2); RMSD, root mean square deviation from crystal structure (\AA); NRC, number of residues adopting random coil conformation; HBO, number of main chain hydrogen bonds; ACC, total solvent-accessible surface (\AA^2); GYR, radius of gyration (nm). NRC, HBO, and ACC were calculated with DSSP (Kabsch and Sander, 1983).

that still have a significantly high eigenvalue. The simulations in vacuo show also high inner products further from the diagonal (Fig. 1 *B*), indicating that the simulations in vacuo are more different from the solvent simulations than the solvent simulations are from each other.

For all reference simulations, the noise as discussed in the Methods and Theory section, which causes overlap between eigenvectors from the essential subspace obtained from one simulation and near-constraint eigenvectors from another set, is not homogeneously spread over all near-constraint eigenvectors (Fig. 2 *A*). Instead, it is concentrated in the

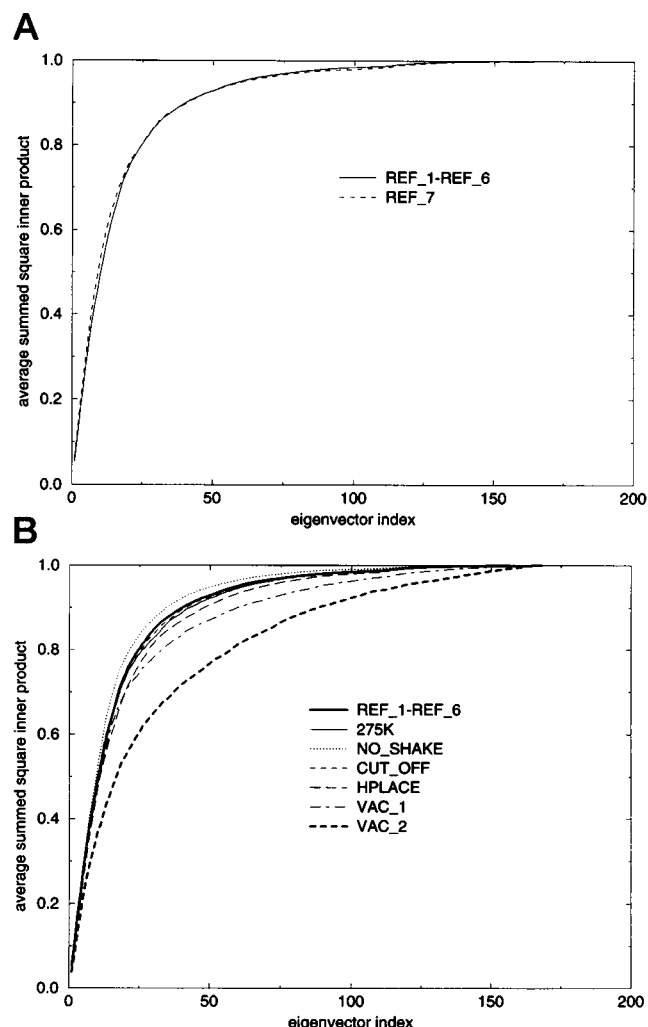


FIGURE 2 Average cumulative square inner products. (*A*) The solid line represents the average summed square inner product of the first ten eigenvectors of one 300-ps reference simulation with all eigenvectors from another 300-ps reference simulation, averaged over all pairs. The dashed line represents the results obtained from the 1-ns reference simulation, compared with the 300-ps simulations. (*B*) For all except the reference simulations, curves were obtained by calculation of the summed squared inner product between all eigenvectors from a single simulation with the first 10 eigenvectors of each reference simulation, divided by 10, averaged over all reference simulations. For the reference simulations, the average cumulative squared inner products of the first 10 eigenvectors of each set with all eigenvectors from all other reference sets were calculated. The average over all pairs is plotted.

near-constraints that still have an appreciable eigenvalue (eigenvectors 11–50), leaving a negligible overlap with all other eigenvectors. This indicates that the definitions of the essential subspaces from all reference simulations are similar. The overlap of all 300-ps reference simulations (REF_1 through REF_6) with the reference simulation of 1 ns (REF_7) is not significantly higher than the overlap between the 300-ps simulations mutually, although a more converged (i.e., containing less statistical noise) description of the essential subspace was to be expected from this longer simulation. This indicates that the convergence of the definition of the essential subspace is initially fast and does not increase significantly in the time window from 300 ps to 1 ns. This validates the use of relatively short simulations to gain insight into the dynamic properties of such systems.

Fig. 2 *B* shows the average cumulative square inner products of all eigenvectors of all sets with the first 10 eigenvectors from the reference simulations. All curves add up to one, because all eigenvectors of one set are always able to rebuild a (subset of) eigenvector(s) of another set (both sets of vectors span the same space). The curves are steep, indicating a high degree of overlap of the first 10 eigenvectors of each of the reference simulations in the essential subspaces obtained from the other simulations. The simulations that were run with parameters different from the reference simulations all show an equal amount of overlap with the reference simulations, as do the reference simulations among each other. For the simulations that were run in vacuo, the measured overlap is significantly smaller, especially VAC_2. The summed average square inner products (here taken as the overlap of the 10 eigenvectors, with highest eigenvalues from one set compared to the reference sets) and values for the penalty function as described in the Methods and Theory section are summarized in Table 2. All values are averages over the comparison with the reference simulations. For the reference simulations, values were obtained by comparison of one reference simulation compared to all others, and subsequent averaging over all. The values in the second column of this table are

TABLE 2 ED properties

Type	MSI		PENAL	
	Mean	σ	Mean	σ
275K	0.483	0.040	1.172	0.087
NO_SHAKE	0.524	0.037	1.051	0.084
CUT_OFF	0.498	0.017	1.158	0.070
HPLACE	0.472	0.024	1.203	0.103
REFS_1_6	0.494	0.040	1.155	0.087
REF_7	0.530	0.065	1.169	0.132
VAC_1	0.478	0.033	1.517	0.068
VAC_2	0.372	0.038	1.967	0.039

MSI, Average and root mean square fluctuation of summed square inner product between the first 10 C- α eigenvectors from ED analyses of each trajectory with each of the first 10 eigenvectors from all reference simulations. PENAL, Average and root mean square fluctuation of penalty function (see Methods and Theory section) between eigenvectors from all simulations and reference simulations.

obtained by summation of the square inner products between the first 10 eigenvectors of one set with all of the first 10 eigenvectors of another set, and subsequent division by 10. A value of 1 is obtained when two sets are identical. It should be noted that the first 10 eigenvectors span only $10/(56 \cdot 3) = 5.95\%$ of the total space. The fact that the essential subspaces from the different simulations overlap for approximately 50% means that similarities between the essential subspaces of the individual trajectories are significant. The amounts of overlap of the essential subspaces of any combination of (except for the second simulation in vacuo) simulations are similar. The penalty function (Table 2), which is more sensitive to differences between eigenvector sets than the measure of overlap in terms of a cumulative inner product (see Methods and Theory section), also gives similar values for all solvent simulations. The two simulations in vacuo, however, give significantly higher values for this penalty function, demonstrating dynamic differences caused by the presence of solvent. Based on these data, there are no detectable systematic differences between the various methods of simulation, apart from the second simulation in vacuo.

Fig. 3 shows the kinds of motion that correspond to the most prominent eigenvectors extracted from the reference simulations. For each of the first six eigenvectors, the motion is concentrated in a few specific places that move concertedly.

DISCUSSION AND CONCLUSIONS

The results presented in this paper show, considering both overall structural and dynamic properties, that all solvent simulations that were studied behave essentially similarly. Only the simulations that were performed in vacuo showed

significantly different behavior from the reference simulations, although even there, the overlap of the essential subspace is still substantial. This is in agreement with previous findings (van Aalten et al., 1995a, 1996). Of course, all analyses were concentrated on one protein; other proteins may behave differently.

Of the overall quantities, the largest differences were found in the total mean square fluctuation and the root mean square deviation (RMSD) from the crystal structure (Table 1). As has been noted before (Storch and Daggett), the RMSD from a single structure (e.g., the crystal structure) is not necessarily a useful quantity for judging the stability of a simulation when large structural rearrangements are occurring, provided that these rearrangements are reversible. Moreover, backbone RMSD values of the structures in the NMR cluster with respect to the crystal structure are in the same range as the observed values for MD.

From the 300-ps solvent simulations, 275°K and CUT_OFF deviate most from all other simulations. The lower total mean square fluctuation of all protein atoms in the simulation at lower temperatures compared to most other simulations can be partially explained by the fact that less thermal motion is present in this simulation. Fast thermal fluctuations can be expected to have a connection to slower, larger fluctuations, which might be reflected in the fact that in the ED analysis, the mean square fluctuations along all essential eigenvectors for this simulation are among the lowest. The CUT_OFF simulation shows the highest overall fluctuation (Table 1). No obvious explanation can be provided for this observation. Because REF_1 is also quite different with respect to the other reference simulations (it exhibits the lowest total mean square fluctuation of all simulations; Table 1), these differences are believed to be based on statistical rather than systematic reasons. In a recent paper (Janezic and Brooks, 1995), where the effects of different protein models on normal modes results were studied, the only significant sensitivity was reported for the description of electrostatic interactions.

For all geometrical properties, the differences between different reference simulations are as high as the differences between the reference simulations and the solvent simulations that were run with adapted parameters. On the basis of these data, therefore, the simulations with different parameters do not differ in a significant way from the reference simulations.

The differences in dynamic behavior of the 10 solvent simulations as revealed by ED also do not appear to be significant. Because there is no systematic connection between the observed dynamic differences and the different simulation parameters (the reference simulations differ as much from each other as from the other solvent simulations), we get the impression that limited sampling time is the most important reason for the presence of these differences. This leads us to the conclusion that the dynamic properties of the simulated protein are not detectably sensitive toward small differences in the force field or in the choice of starting structure in the time span considered here.

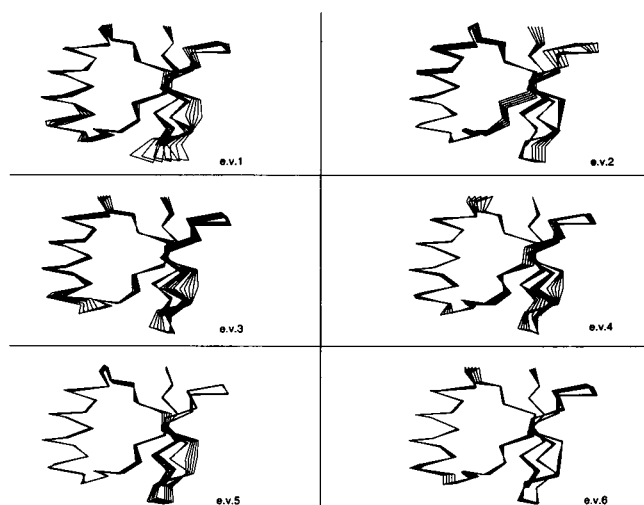


FIGURE 3 Snapshots of single-eigenvector motions. Structures corresponding to the minimum and maximum sampled positions along the first six eigenvectors of an ED analysis of solvent trajectories combined are shown, together with three intermediate positions, equally spaced between the minimum and maximum.

Only the simulations that were performed in vacuo are significantly different from the reference simulations in solvent. To support these findings, further simulations of other proteins are necessary.

As already observed recently (Amadei et al., 1993, 1996; van Aalten et al., 1995b), single simulations of a few hundred picoseconds of a protein in water seem to yield an acceptable approximation of the essential subspace, although for a fully converged description of this subspace, longer simulations are required. This is supported by the observation that similarities between ED analyses on individual trajectories that were studied here are relatively high (Fig. 1, Table 2), considering the fact that in the comparison as presented in Table 2, two sets that each contain noise are always compared. As explained in the Methods and Theory section, the overlap (defined as the summed square inner products between vectors from two essential subspaces) between eigenvectors obtained from each of the reference simulations with the fully converged set of eigenvectors can be expected to be larger than the overlap between eigenvectors obtained from two short simulations. The measured overlap of approximately 0.5 between eigenvector sets obtained from multiple simulations (Table 2) therefore means that for each set, the overlap with the fully converged set of eigenvectors is even higher. This means that in a relatively short simulation, a good approximation of the true essential subspace is reached, within the limitations of the force field. The fact that within the essential subspace the individual eigenvectors are not identical in all simulations (although the subspace itself has approximately converged) indicates that in this subspace, the region that has been visited during a single short simulation is only a small fraction of the complete available subspace. This is in agreement with previous findings (Amadei et al., 1996; de Groot et al., 1996a,b).

Further studies have shown that convergence increases only slowly with simulation time (Fig. 2 A, Table 2), making predictions about the minimum time required to obtain a fully converged description of the dynamics impossible.

In a recent study (Clarage et al., 1995), two halves of a short MD simulation of myoglobin were compared. It was concluded that a few hundred picoseconds is not sufficient to obtain equilibrated dynamics. In another study (Balsera et al., 1996), two halves of a simulation of 470 ps of G-actin (375 residues) showed significant differences in terms of principal component analysis, analogous to essential dynamics analysis. We have shown here, and before (Amadei et al., 1993, 1996; van Aalten et al., 1995b), that within a few hundred picoseconds, the definitions of both the essential and the near-constraints subspaces are approximately stable, whereas motions within the essential subspace are still equilibrating. In the present study (as also noted before by us; van Aalten et al., 1995b) and in others (Janezic et al., 1995), the overlap found between the essential subspaces as derived from short simulations is substantial.

The initial description of the essential subspace as derived from a relatively short MD simulation can be used to obtain a more refined definition of this space in an extrapolation method (Amadei et al., 1996). In such a method, an

adapted form of MD is performed, with constraints in the approximated essential subspace. These constraints are chosen such that the system itself determines the regions of the space that it samples. Dynamic coupling between the accessible modes of motion will automatically result in motion in the true essential subspace of that system. Analysis of the cloud of structures thus produced will then yield a more accurate description of that space. The procedure may be repeated until no changes are detectable, to obtain a completely converged definition of the modes spanning the essential subspace. We are currently investigating such methods (de Groot et al., 1996a,b).

REFERENCES

- Amadei, A., A. B. M. Linssen, and H. J. C. Berendsen. 1993. Essential dynamics of proteins. *Proteins Struct. Funct. Genet.* 17:412–425.
- Amadei, A., A. B. M. Linssen, B. L. de Groot, D. M. F. van Aalten, and H. J. C. Berendsen. 1996. An efficient method for sampling the essential subspace of proteins. *J. Struct. Dyn.* 13:615–626.
- Balsera, M. A., W. Wriggers, Y. Oono, and K. Schulten. 1996. Principal component analysis and long time protein dynamics. *J. Phys. Chem.* 100:2567–2572.
- Barchi, J. J., B. Grasberger, A. M. Gronenborn, and G. M. Clore. 1994. Investigation of the backbone dynamics of the IgG-binding domain of the streptococcal protein G by heteronuclear two-dimensional ^1H - ^{15}N nuclear magnetic resonance spectroscopy. *Protein Sci.* 3:15–21.
- Berendsen, H. J. C., J. P. M. Postma, A. DiNola, and J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.
- Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, and J. Hermans. 1981. In *Intermolecular Forces*. B. Pullmann, editor. Reidel, Dordrecht. 331.
- Clarage, J. B., T. Romo, B. K. Andrews, B. M. Pettitt, and G. N. Phillips, Jr. 1995. A sampling problem in molecular dynamics simulations of macromolecules. *Proc. Natl. Acad. Sci. USA.* 92:3288–3292.
- de Groot, B. L., A. Amadei, R. M. Scheek, N. A. J. van Nuland, and H. J. C. Berendsen. 1996a. An extended sampling of the configurational space of HPr from *E. coli*. *Proteins Struct. Funct. Genet.* In press.
- de Groot, B. L., A. Amadei, D. M. F. van Aalten, and H. J. C. Berendsen. 1996b. Toward an exhaustive sampling of the configurational spaces of the two forms of the peptide hormone guanylin. *J. Biomol. Struct. Dyn.* 13:741–751.
- Gallagher, T., P. Alexander, P. Bryan, and G. L. Gilliland. 1994. Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR. *Biochemistry.* 33: 4721–4729.
- Garcia, A. E. 1992. Large-amplitude nonlinear motions in proteins. *Phys. Rev. Lett.* 68:2696–2699.
- Gronenborn, A. M., D. R. Filipula, N. Z. Essig, A. Achari, M. Whitlow, P. T. Wingfield, and G. M. Clore. 1991. A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. *Science.* 253:657–661.
- Hayward, S., A. Kitao, F. Hirata, and N. Gō. 1993. Effect of solvent on collective motions in globular proteins. *J. Mol. Biol.* 234:1207–1217.
- Hooft, R. W. W., G. Vriend, and C. Sander. 1996. *Proteins Struct. Funct. Genet.* In press.
- Janezic, D., and B. R. Brooks. 1995. Harmonic analysis of large systems. II. Comparison of different protein models. *J. Comp. Chem.* 16: 1543–1553.
- Janezic, D., M. Venable, and B. R. Brooks. 1995. Harmonic analysis of large systems. Comparison with molecular dynamics. *J. Comp. Chem.* 16:1554–1566.
- Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* 22:2577–2637.

- Kuszewski, J., G. M. Clore, and A. M. Gronenborn. 1994. Fast folding of a prototypic polypeptide: the immunoglobulin binding domain of streptococcal protein G. *Protein Sci.* 3:1945–1952.
- Ryckaert, J. P., G. Cicciotti, and H. J. C. Berendsen. 1977. Numerical integration of the cartesian equations of motion of a system with constraints; molecular dynamics of *n*-alkanes. *J. Comp. Phys.* 23: 327–341.
- Scheek, R. M., N. A. J. van Nuland, B. L. de Groot, and A. Amadei. 1995. Structure from NMR and molecular dynamics: distance restraining inhibits motion in the essential subspace. *J. Biomol. NMR.* 6:106–111.
- Storch, E. M., and V. Daggett. 1995. Molecular dynamics simulation of cytochrome B5: implications for protein-protein recognition. *Biochemistry.* 34:9682–9693.
- van Aalten, D. M. F., A. Amadei, R. Bywater, J. B. C. Findlay, H. J. C. Berendsen, C. Sander, and P. F. W. Stouten. 1996. A comparison of structural and dynamic properties of different simulation methods applied to SH3. *Biophys. J.* 70:684–692.
- van Aalten, D. M. F., A. Amadei, G. Vriend, A. B. M. Linssen, G. Venema, H. J. C. Berendsen, and V. G. H. Eijssink. 1995a. The essential dynamics of thermolysin—confirmation of hinge-bending motion and comparison of simulations in vacuum and water. *Proteins Struct. Funct. Genet.* 22:45–54.
- van Aalten, D. M. F., J. B. C. Findlay, A. Amadei, and H. J. C. Berendsen. 1995b. Essential dynamics of the cellular retinol binding protein—evidence for ligand induced conformational changes. *Protein Eng.* 8:1129–1136.
- van Gunsteren, W. F., and H. J. C. Berendsen. 1987. Gromos Manual. BIOMOS, Biomolecular Software. Laboratory of Physical Chemistry, University of Groningen, Groningen, The Netherlands.
- Vriend, G. 1990. WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.* 8:52–56.