A Comparison of Techniques for Calculating Protein Essential Dynamics

D. M. F. VAN AALTEN,[1,*] B. L. DE GROOT,[2] J. B. C. FINDLAY,[1] H. J. C. BERENDSEN,[2] and A. AMADEI[2]

[1] *Department of Biochemistry, University of Leeds, Leeds LS2 9JT, United Kingdom*
[2] *Groningen Biomolecular Sciences and Biotechnology Institut (GBB), Department of Biophysical Chemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands*

Received 20 June 1995; accepted 3 May 1996

ABSTRACT

Recently the basic theory of essential dynamics, a method for extracting large concerted motions from protein molecular dynamics trajectories, was described. Here, we introduce and test new aspects. A method for diagonalizing large covariance matrices is presented. We show that it is possible to perform essential dynamics using different subsets of atoms and compare these to the basic C-α analysis. Essential dynamics analyses are also compared to the normal modes method. The stability of the essential space during a simulation is investigated by comparing the two halves of a trajectory. Apart from the analyses in Cartesian space, the essential dynamics in ϕ/ψ torsion angle space is discussed.

Introduction

The recently introduced essential dynamics (ED) method[1] can be used to separate large concerted structural rearrangements from irrelevant fluctuations in a multiparticle system. The method consists of diagonalizing the covariance matrix of atomic fluctuations, and it has been shown that this analysis identifies constraints or approximate constraints (near constraints) in the configurational space.[1] As a first example, this method was applied to a protein, lysozyme.[1] The ED method was able to separate the few large, mainly anharmonic motions (the essential subspace) in a lysozyme molecular dynamics (MD) trajectory from the small, Gaussian fluctuations (the near-constraints subspace). The ED procedure is equivalent to a multidimensional linear least squares fit of the trajectory, where the first eigenvector corresponds to the direction that fits best to the ensemble of configurations, the second to the second best, etc.[2] The principle of such a multidimensional fitting was applied to protein dynamics for the first time by Garcia.[3]

* Author to whom correspondence should be addressed. E-mail: bmb5dva@biovax.leeds.ac.uk

CCC 0192-8651 / 97 / 020169-13

The motions in the essential space are linked to the biological function of the protein. For thermolysin, a large hinge-bending motion was observed that opened and closed the active site.[4] This motion was supported by crystallographic data. Similar results were obtained for the histidine-containing protein (HPr), where the motions were centered around the active site histidine.[5] The SH3 binding domain showed large correlated fluctuations of two loops known to be involved in peptide binding.[6] The ED method was also used to reveal the retinol entry/exit site in the cellular retinol binding protein, where retinol showed strong correlations to three surface loops, which moved apart to create a hole for the retinol.[7] So far, the basic application of ED, involving all atoms or C-a atoms only, has proved useful in investigating protein dynamics with the aim of revealing motions with functional implications. Here, various aspects of the ED method are tested. We prove that even for short simulations, the ED properties derived are stable. Furthermore, we compare ED results to those from the normal modes (NM) analysis. We show the application of ED in dihedral space instead of Cartesian space, propose possible methods to take side chains into account, and present a fast way for diagonalizing large covariance matrices.

Methods

COMPARISON OF ESSENTIAL SPACES

The most straightforward way of comparing essential spaces is to consider inner products of the corresponding eigenvectors. When these projections are close to 1.0 for all the eigenvectors of the essential space, the essential spaces spanned by the two sets of eigenvectors are the same. However, if one wants to compare two sets of eigenvectors, coming from, for instance, two different simulations of the same system,[4] it may happen that two or more eigenvectors in one set are interchanged with respect to the other set. Slight differences in fluctuation in the system may result in a different ordering of the eigenvectors because the eigenvectors are ordered by the size of the eigenvalue.

Here we compare essential subspaces by calculating the average square projection of all eigenvectors of one set onto eigenvectors of another set. To compare eigenvectors calculated from, for instance, a covariance matrix of C-α atoms only to those from a covariance matrix of a different set of atoms, the components not belonging to C-α atoms are deleted from the latter set. The resulting eigenvectors are renormalized and then used for the comparison method described above. Because all the non-C-α components were deleted, the resulting eigenvectors are not exactly orthogonal anymore. The resulting error, however, is usually small, as will be shown below.

SPLITTING COVARIANCE MATRIX

If more than about 1500 particles are used in the ED analysis, it becomes computationally impossible to diagonalize the corresponding covariance matrix (on the average workstation with about 64 MB of memory). Both CPU time and memory needed for diagonalization increase rapidly with the size of the system (approximately proportional to N^2). It is possible, however, to overcome this problem by dividing the overall covariance matrix in submatrices. The protein is divided in a number of groups containing atoms equally distributed over the protein. If the protein is to be divided in K groups, the first atom is included in the first group the Kth atom in the Kth group, and atom $K + 1$ again in the first group. For each of the K groups a covariance matrix is built up and diagonalized. It is not enough to simply combine the eigenvectors obtained in this way because no correlations between the K groups are taken into account. To include this information, projections were made of the subtrajectories (containing the same atoms as the groups) onto the K eigenvector sets. Projections onto eigenvectors are overall coordinates that give information on how the system moves in the directions described by the eigenvectors. As before, the eigenvectors are ordered in such a way that the first vector is the direction in which the group of atoms has the largest overall mean square positional fluctuation. Because approximately the first 10% of the eigenvectors of each subgroup usually span a space in which more than 90% of the motion is concentrated and because all the near-constraints eigenvectors of each subgroup tend to be statistically independent from any large slow motion, we can use only the first Y eigenvectors as an approximation. The number Y is limited by the computational resources; taking a higher Y results in higher accuracy. If these projections are used to build up a new covariance matrix, the original $K*A*3$-dimensional matrix, where A is the number of atoms per subgroup, can be approximated by a $K * Y$ dimensional one. The covariance matrix built up in this way is

diagonalized, now including correlations between the first Y eigenvector coordinates of the K groups. However, the eigenvectors obtained are expressed in overall coordinates instead of the starting Cartesian coordinates. They have the first Y eigenvectors of the K subgroups as a basis set instead of the starting Cartesian basis set. To express the new eigenvectors in the Cartesian basis set, they must be described as linear combinations of the eigenvectors of the K subgroups. This is possible because eigenvectors within one subgroup are always orthogonal (symmetrical covariance matrix), and eigenvectors belonging to different subgroups are orthogonal because these subgroups define orthogonal subspaces (eigenvectors of one group have zero components where eigenvectors from another group have nonzero components). Therefore the K sets of eigenvectors form an orthogonal basis set. The new eigenvectors can be defined as

$$\boldsymbol{\eta}_k = \sum_{j=1}^{N} \sum_{i=1}^{Y} \boldsymbol{\alpha}_{ij}^{k} \boldsymbol{\nu}_{ij}.$$

Here $\boldsymbol{\eta}_k$ is the kth new (Cartesian) eigenvector, $\boldsymbol{\nu}_{ij}$ is the ith eigenvector from the jth subgroup, and α_{ij}^{k} are the corresponding coefficients (components of the eigenvectors of the reduced analysis). Because the trace of a matrix is unaffected by a transformation like diagonalization, the sum of the traces of the K diagonalized submatrices is the same as the trace of the (diagonalized) full matrix. This means that the sum of the eigenvalues from the K sets is equal to the sum of the eigenvalues from the full matrix. From the definition of the elements of the covariance matrix it follows that the diagonal elements of the covariance matrix constructed from the projections are just the eigenvalues from the K subgroups.

It is possible to estimate the error produced by the approximation presented here. The correlations between the projections not taken into account can be expressed in terms of the (known) diagonal elements using the Cauchy–Schwarz inequality,

$$|\lambda_{ij}| \leq \sqrt{\lambda_{ii}\lambda_{jj}}\,.$$

As stated above, the diagonal elements C_{ii} and C_{jj}, are just the eigenvalues of the K subsets. The maximum possible value of C_{ij} that is neglected is then equal to the square root of the product of the largest eigenvalue of the K sets (the first eigenvalue from one of the K sets) and the largest eigenvalue that is neglected (the $Y + 1$ eigenvalue from one of the K sets); hence, if Y is large (and the $Y + 1$ eigenvalue close to zero) the error will be close to zero. In general, this is an upper limit significantly larger than the real largest neglected C_{ij} element. This is caused by the tendency of subgroup eigenvectors with very small corresponding eigenvalues to be statistically independent from any large slow motion. In fact every eigenvector with a very small eigenvalue corresponds to a high frequency Gaussian fluctuation stable (equilibrated) in a short time and hence independent from any large slow structural change. This implies that any C_{ij} element in the matrix involving a near-constraint eigenvector of one of the subgroups and an essential eigenvector of another subgroup should be almost zero.

STABILITY OF THE ED ANALYSIS

Because the ED analysis is mostly used on short (< 1.0 ns) trajectories, it is useful to have an analysis of the statistical relevance and stability of the motions observed. The stability of the ED properties is investigated in two ways: 1) by studying the differences in fluctuation along the essential eigenvectors for the two halves of the trajectory; 2) by calculating two sets of eigenvectors from the two halves of the trajectory and studying the inner products between these sets. This gives a cross-projection matrix that, if the sets are similar, should only have high numbers near the diagonal.

NM ANALYSIS

Like ED, NM analysis[8] reveals large conformational changes in proteins. The methods, however, differ on a few essential points. First, NM is based on the shape of the potential energy function in a (local) minimum. This minimum is assumed to be a harmonic well, restricting the method to prediction of only harmonic vibrations. Eigenvectors are obtained by diagonalization of the Hessian matrix, which contains derivatives of the forces with respect to every coordinate as elements. The ED method extracts large concerted motions from an MD trajectory, in which many of such minima are sampled, and therefore puts no restrictions to the shape of the potential energy function. As will be shown below, ED can be used on any set of atoms selected from an MD simulation, thus implicitly including solvent effects; NM is restricted to an ''all atom'' analysis, where usually no solvent ef-

fects are taken into account. Although ED is a more general analysis than NM, the latter is a useful method for conformational analysis because it is much less time consuming. The basic NM algorithm[8] was implemented in the GROMOS suite of programs.[9] The crystal structures of SH3[10] and HPr[11] were energy minimized by an initial cycle of steepest descent, followed by multiple cycles of conjugate gradients until no energy difference could be detected. The eigenvectors obtained by NM analysis were converted to eigenvectors only containing C-α coordinates, by deleting all non-C-α components and renormalizing. These eigenvectors were then projected on those from the C-α ED analysis by the projection method described above.

MD SIMULATIONS

We chose two protein MD trajectories to test the theories described above. The first is a 300-ps (3000 frames) equilibrium trajectory of HPr (85 residues, 785 protein atoms) described before.[11] The second is a 200-ps (5000 frames) equilibrium trajectory of SH3 (57 residues, 583 protein atoms.[6] Both MD simulations were performed with GROMOS[9] using periodic boundaries with a truncated octahedron box filled with simple point charge water molecules.[12]

Results

STABILITY OF ED PROPERTIES

If one is interested in obtaining the ED properties of a protein, a sufficiently long trajectory should be sampled of the system in equilibrium. Because it might be argued that the initial equilibration of a system by MD is a nonphysical process, one needs to be sure that this process is excluded from the trajectory to be analyzed by ED. Here we show, using two methods for comparing the two halves of both the SH3 and HPr trajectories, that the essential subspace obtained from one half of the trajectories is similar to that calculated from the other half. Figure 1 shows the fluctuations in the two halves of these trajectories projected onto the eigenvectors calculated over the whole trajectories. Although there appear to be some differences in dynamic behavior along these eigenvectors, all the eigenvectors that show large fluctuation in one-half of the trajectory also show significant fluctuation in the other half. To further investigate the differences between the two halves of the trajectories, eigenvectors (C-α atoms only) were calculated from the two halves independently. The two resulting sets were then cross-projected (Fig. 2A, B). The plots show that all essential eigenvectors of one set have high projections on essential eigenvectors of the other set. Also the near-constraints eigenvectors give high projections close to the diagonal, indicating the sets are similar. No high projections are observed far from the diagonal.

To estimate the amount of noise in this cross-projection method, two further such projections were calculated. First, eigenvectors sets (C-α atoms only) were calculated from the first 57 residues of HPr and all residues (also 57) of SH3 and subsequently cross-projected (Fig. 2C). Because two differently folded polypeptides are now compared, the eigenvectors are expected to be different as indeed reflected by Figure 2C. The cross-projection of a randomly generated set of eigenvectors onto another shows, however, an even larger degree of spread of the projections (Fig. 2D). So, even though two significantly different polypeptides are compared Figure 2C, some degree of overlap is observed. This is probably due to the fact that although the folds are different, a lot of the constraints are essentially the same.

USING DIFFERENT SETS OF ATOMS FOR ED

In most previous ED analyses[1,4,5] the covariance matrix was constructed from C-α atomic displacements only. We investigate how other atoms from the backbone or side chains can be included to gain information on the dynamics of these groups, without losing accuracy in the C-α–C-α correlations and without significantly increasing the computational costs as is the case with the all atom analysis.[1] Two approximations of describing side chains are used: inclusion of the terminal atom of all side chains in the construction of the covariance matrix and inclusion of the coordinates of the geometrical center of the side chain. Both these options lead to a doubling of the size of the covariance matrix as compared to the C-α covariance matrix, because there are now two particles per residue. Figure 3 shows the comparison of the C-α components of the eigenvectors derived from these covariance matrices with the eigenvectors of the pure C-α analysis. Inclusion of the side chain

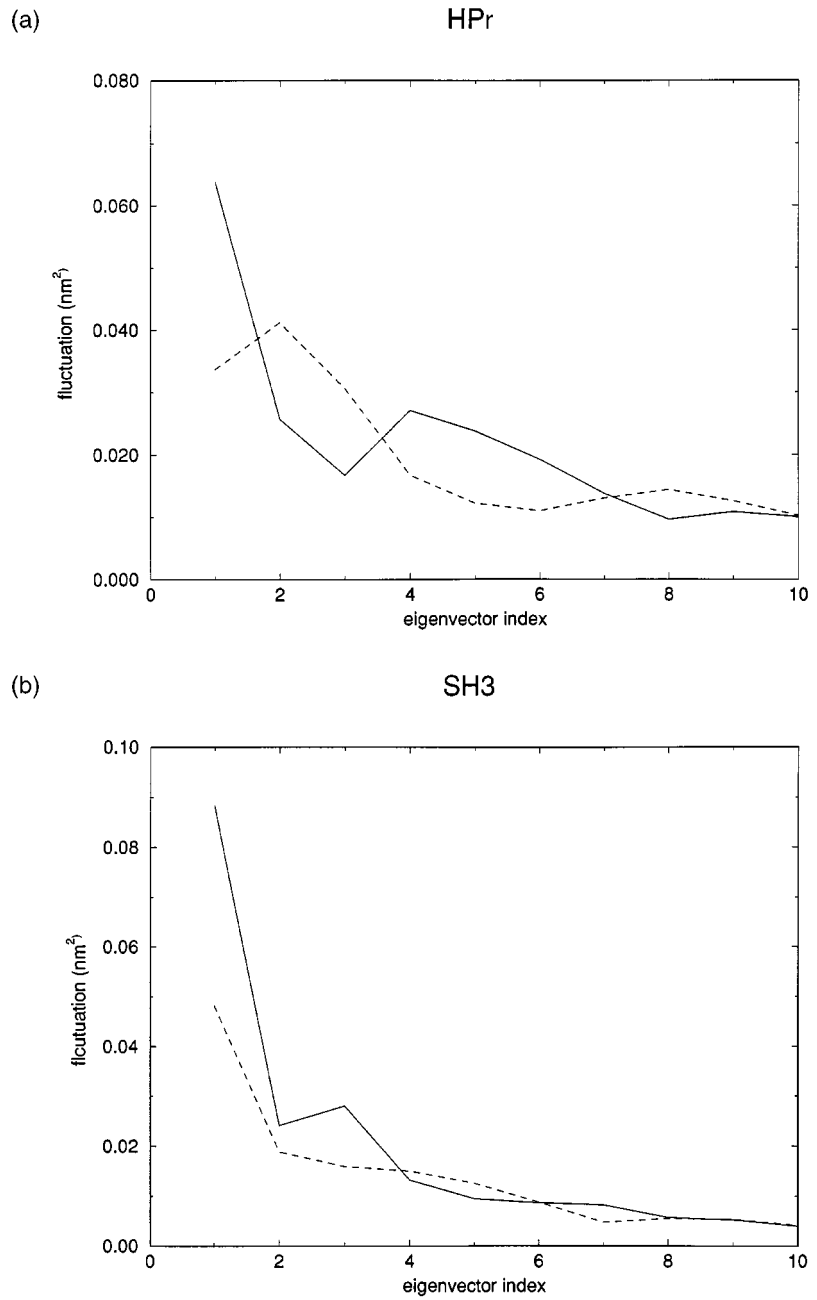

FIGURE 1. Comparison of the fluctuations along the first 10 essential eigenvectors in the two halves of the (A) HPr and (B) SH3.

information does not lead to significant deviation in the C-α components. A slightly closer agreement is obtained when the geometrical center of the side chain is used. This is probably due to the fact that the geometrical center of the side chain is defined by more than one atom, which leads to the reduction of the noise in the atomic correlations. Moreover, the geometrical center is closer to the backbone, leading to stronger correlation with the C-α atoms. The inclusion of this side chain information only leads to a small increase in the computational cost of diagonalizing the covariance matrix and introduces less noise in the C-α components compared to the application of ED using all atoms of the protein (Fig. 3).

To allow more detailed geometrical evaluations of structures in the essential space, such as secondary structure, hydrogen bonds, and dihedral

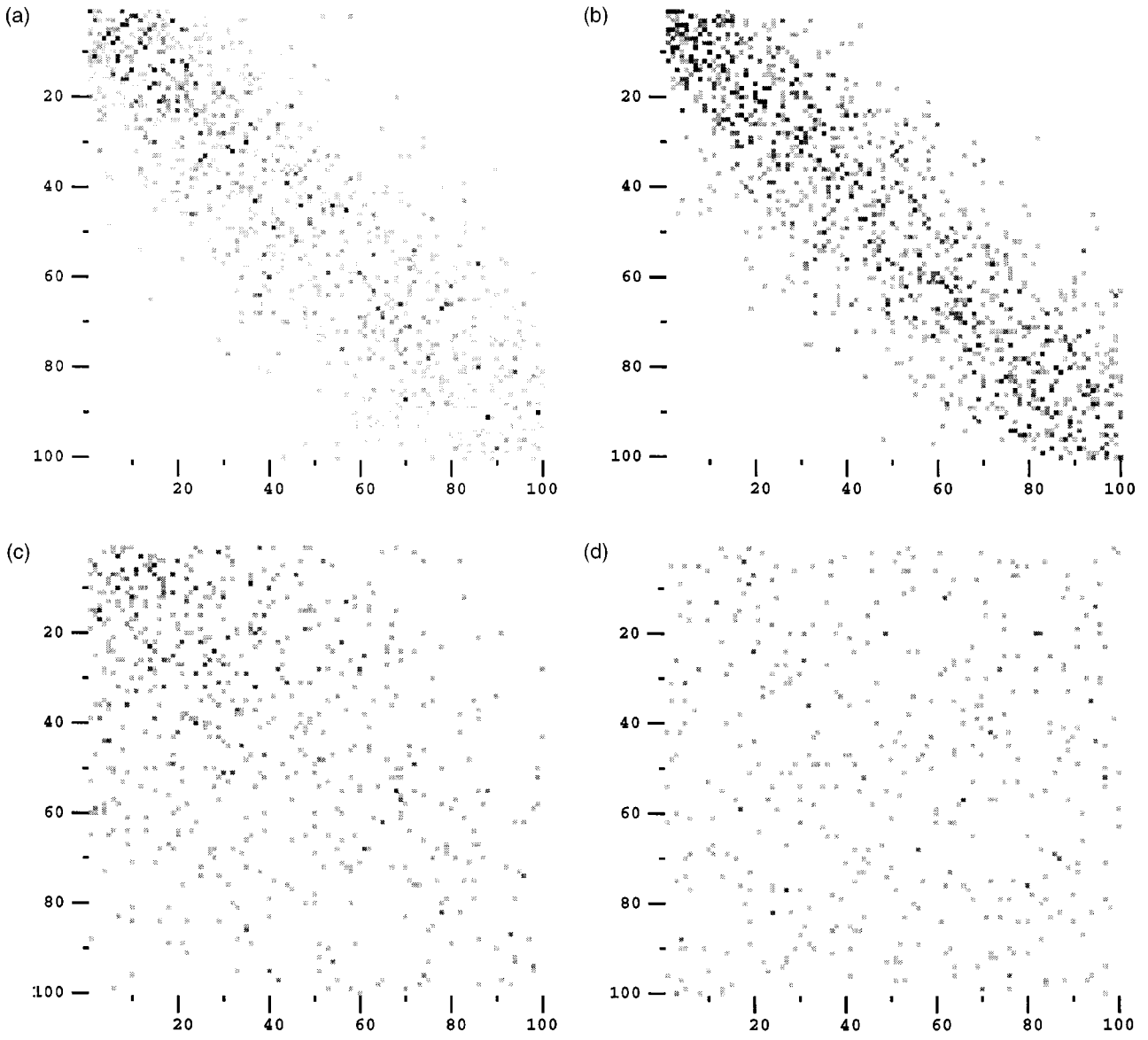

FIGURE 2. Projection of one set of eigenvectors onto another. (A) Eigenvectors calculated from the first half of the HPr trajectory onto those calculated from the second half. (B) Eigenvectors calculated from the first half of the SH3 trajectory onto those calculated from the second half. (C) Eigenvectors calculated from the first 57 residues of the HPr trajectory onto those calculated from the SH3 trajectory. (D) One set of random eigenvectors onto another set of random eigenvectors.

angle analyses, it is necessary to include all backbone atoms in the construction of the covariance matrix. The backbone carbonyl carbon and oxygen, nitrogen, hydrogen, and C-α carbon atoms were used to build the covariance matrix. Although this leads to a fivefold increase of the dimension of the matrix, it leads to no significant deviation in the C-α components (Fig. 3) whereas a wealth of additional information is obtained.

ANALYSIS OF THE COVARIANCE MATRIX SPLITTING METHOD

The HPr and SH3 trajectories were split into four and three groups of atoms, respectively, and covariance matrices were constructed as described in the Methods section. The all atom eigenvectors resulting from combining the data of the submatrices were compared to the all atom eigenvectors

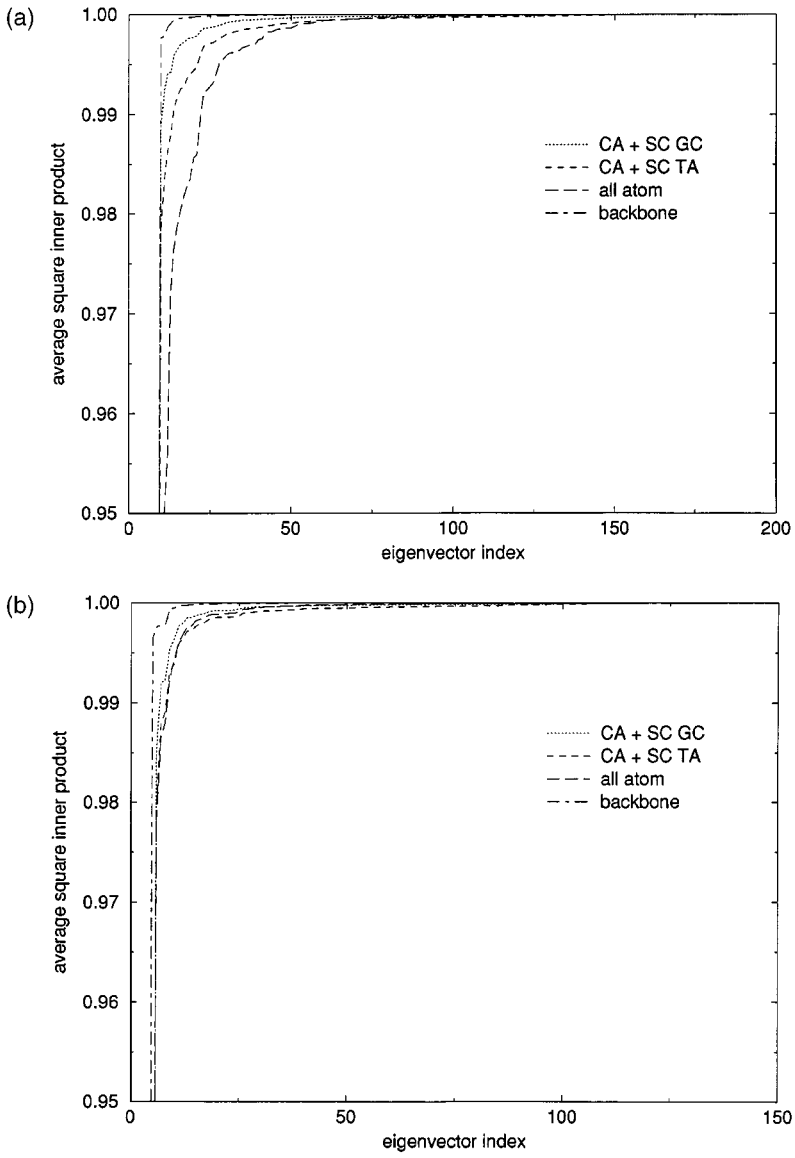

FIGURE 3. Comparison of various (A) HPr and (B) SH3 ED analyses to the C-α ED analysis, using the projection method described in the Methods section. CA, C-α; SC, side chain; GC, geometrical center; TA, terminal atom.

derived directly from an all atom covariance matrix. Figure 4 shows the inner products resulting from the projection of eigenvectors of one set onto corresponding eigenvectors of the other set. For the essential eigenvectors, inner products close to 1.0 were obtained. For both HPr and SH3 only the first 100 eigenvectors of the submatrices were used for the projections and subsequent construction of a new covariance matrix. The square root of the product of the largest eigenvalue taken into account and the largest neglected eigenvalue (corresponding to the upper limit of the neglected matrix element) was comparable to the 31st eigenvalue for HPr and the 24th eigenvalue for SH3, suggesting that higher index eigenvectors might be inaccurate. In practice we find that even more eigenvectors obtained from the splitting and direct method are similar. Hence, the splitting method can be used to reliably reconstruct the eigenvectors from the essential space. Roughly, one could say that if 10% of the total number of coordinates are used in the approximation, about 10% of the total number of eigenvectors can be approximated reliably.

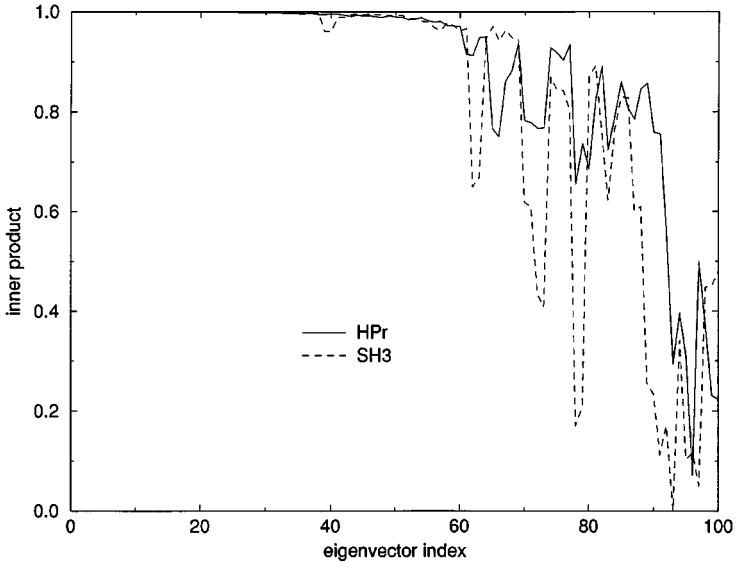

FIGURE 4. Projection of the eigenvectors resulting from an all atom covariance matrix onto corresponding eigenvectors obtained with the covariance matrix splitting method.

ED IN DIHEDRAL SPACE

In all analyses presented above the atomic coordinates in Cartesian space were used to build the covariance matrix. For the atoms of the protein backbone, N, H, C-α, C, and O, this results in 15 degrees of freedom per residue. However, the real degrees of freedom in a protein backbone are rotation around the N—C-α (ϕ) and the C-α—C (ψ) bonds. A 3-D structure of the backbone of a protein can be generated from a set of ϕ and ψ angles, assuming fixed bond lengths, bond angles, and a planar peptide bond. Therefore, there are 2 degrees of freedom per residue in dihedral angle space instead of 15 in Cartesian space. Moreover, if these dihedral angles are the real degrees of freedom in an amino acid, there could be only a few, localized components with a large value in an eigenvector derived from a dihedral angle covariance matrix, indicating hinge regions in the protein. We investigated the essential degrees of freedom in dihedral angle space for HPr and SH3. Instead of using the correlated fluctuations around average Cartesian coordinates for the construction of the covariance matrix, fluctuations in the ϕ and ψ angles in the protein trajectory were used. Diagonalization of this matrix yields a set of eigenvectors and eigenvalues. The eigenvectors describe correlated rotations around certain ϕ and ψ angles in the protein backbone. The eigenvalues indicate the amplitude of these correlated rotations. In Figure 5 the eigenvector components of the first four eigenvectors of the dihedral analysis are compared to the ϕ/ψ fluctuations in the four eigenvectors with the largest eigenvalues of the Cartesian backbone analysis. The eigenvectors of the dihedral angles analysis are much ''cleaner'' than those of the Cartesian analysis; that is, there is less noise and there are only a few large components (angles) showing a large concerted rotation. There is a significant amount of similarity between the dihedral angle eigenvectors and the ϕ/ψ fluctuations in the Cartesian analysis. Angles that fluctuate in the dihedral angle analysis also show a large fluctuation in the Cartesian analysis. This shows that the dihedral angle analysis identifies 3-D motions similar to the Cartesian space analysis. Figure 5 also shows that when there is a dihedral angle showing a large rotation in an eigenvector, there is an adjacent dihedral angle that shows an opposite rotation, partially compensating the first one. These large rotations always occur at residues that also show a large fluctuation in their ϕ/ψ angles in Cartesian space (Fig. 5), but the interesting mechanism of one dihedral angle compensating for the other cannot be observed in the Cartesian space analysis.

There are some drawbacks in the dihedral ED analysis. The cumulative, normalized eigenvalue curves of the HPr and SH3 dihedral space eigenvalues compared to those of the Cartesian back-

bone analysis are shown in Figure 6. For the essential eigenvectors, the dihedral angle space curves are less steep than the eigenvalue curves of the Cartesian space analyses. This indicates that the dimensionality of the essential space is larger than that of the essential Cartesian space. In addition, visualization of structural rearrangements described by the eigenvectors showed unphysical structures. This is probably due to the fact that the bond angles and the peptide bond torsion angles are not completely constrained. An error of a few degrees in one dihedral angle may lead to deviations of several angstroms in Cartesian space.

NM ANALYSIS

Eigenvectors resulting from the NM analysis of SH3 and HPr were compared using the cumulative inner product method described in the Methods section. The cumulative inner products of the first 10 eigenvectors of ED projected onto the first 10 of the NM analysis (the 10 lowest frequency modes) are shown in Table I. The cumulative inner products for the ED eigenvectors from the two halves of the trajectories are similar to the inner products of ED with NM eigenvectors. This suggests there is significant overlap between the essential spaces derived from normal modes and essential dynamics. Longer simulations or simulations started from a structure distinct from the structure used for NM are likely to give smaller a overlap between the eigenvectors obtained from both techniques.

Discussion and Conclusions

We introduced a number of new techniques, enhancing the possibilities of using ED to analyze correlated motions in proteins. We showed that even relatively short (about 200 ps) trajectories can be used to derive a stable essential space. Recently reports were made about the stability of correlated fluctuations of atoms in MD simulations.[13,14] It was shown that for a short simulation on myoglobin, different covariance matrices were obtained if the two halves of the simulation were considered separately. We show here, however, in two different ways, that a few hundred picoseconds of simulation on two proteins produce an approximately converged definition of the essential subspace in which all relevant fluctuations are confined.

The observation of the significant amount of overlap between eigenvectors derived from an NM analysis and those from ED indicates that the two methods essentially yield the same kind of information. There are, however, a few crucial differences between NM and ED, even though some are not directly apparent from the proteins we chose for our comparison. First, we recently demonstrated that water molecules and in particular crystallographic water molecules are of great importance in accurately determining protein essential dynamics.[4,6] Water molecules are normally not taken into account in NM analysis. However, both proteins used here do not contain internal crystallographic water molecules, which may cause a higher degree of overlap than for proteins with crystallographic water molecules. Second, the simulations used here are relatively short. During longer simulations, multiple minima are sampled by MD, resulting in anharmonic modes that cannot be described by NM, because of the assumption that all modes are harmonic. In our previous ED analyses, we saw that the protein is able to sample regions in the essential subspace distant from the crystal structure. The NM results depend heavily on the local structure, and the relatively short simulations used here only sample the essential space close to the crystal structure; this accounts for the higher overlap between ED and NM eigenvectors than would be found for longer simulations or simulations started from a different structure. Apart from these theoretical differences, there is also the difference in freedom of choosing the set of atoms used in the analysis. NM analyses are restricted to all atoms; as shown above, there is no such restriction for ED.

Usually, a description of the correlations between C-α or backbone atoms is sufficient to get information about the relevant motions of the protein. Restriction of the analysis to C-α or backbone atoms has the advantage that the analysis is less influenced by statistical noise and hence yields a more reliable description of the essential space than an all atom analysis does. However, in certain cases a detailed description of the correlation between all atoms in the protein might be desirable to provide additional information. Up to now, this was limited by the size of the protein because of the computational costs involved in diagonalizing the matrix. The same limit may restrict the analysis of backbone atoms for very large proteins. We

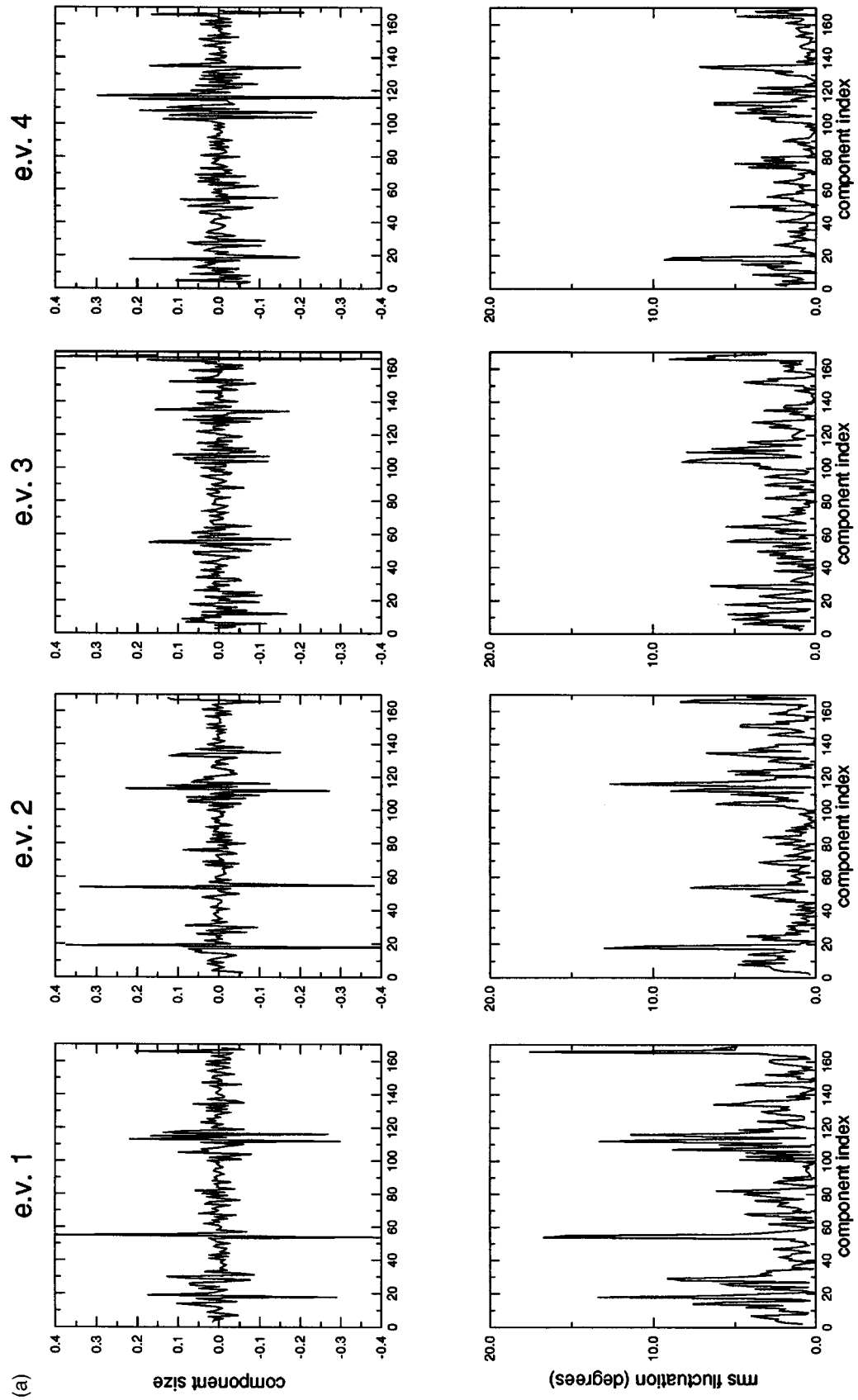

FIGURE 5. Comparison of eigenvector components (upper graphs) of the ED analysis in dihedral angle space to the mean square fluctuation of ϕ / ψ angles in the eigenvector trajectories coming from the Cartesian backbone analysis (lower graphs) for (A) HPr and (B) SH3.

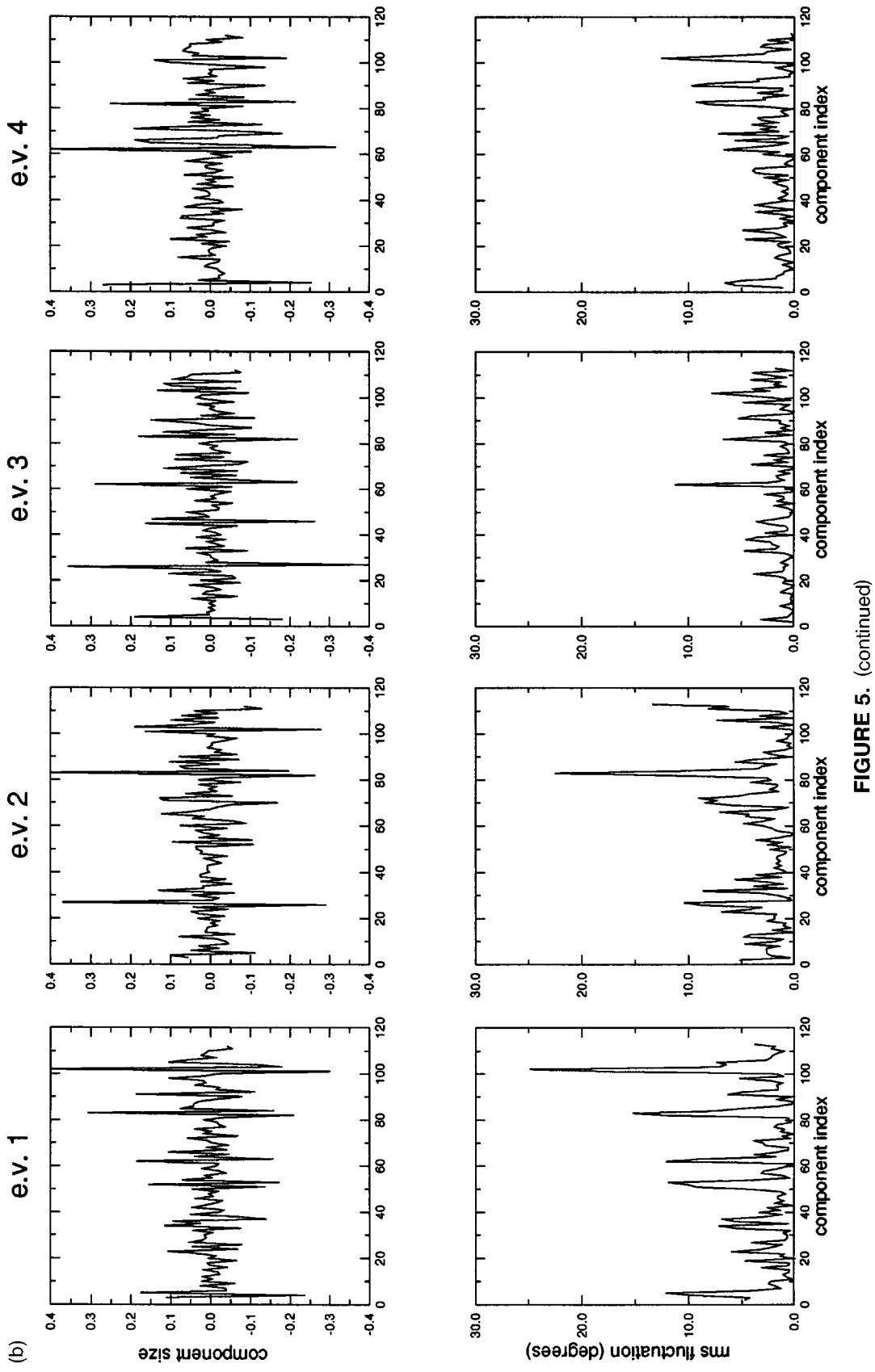

FIGURE 5. (continued)

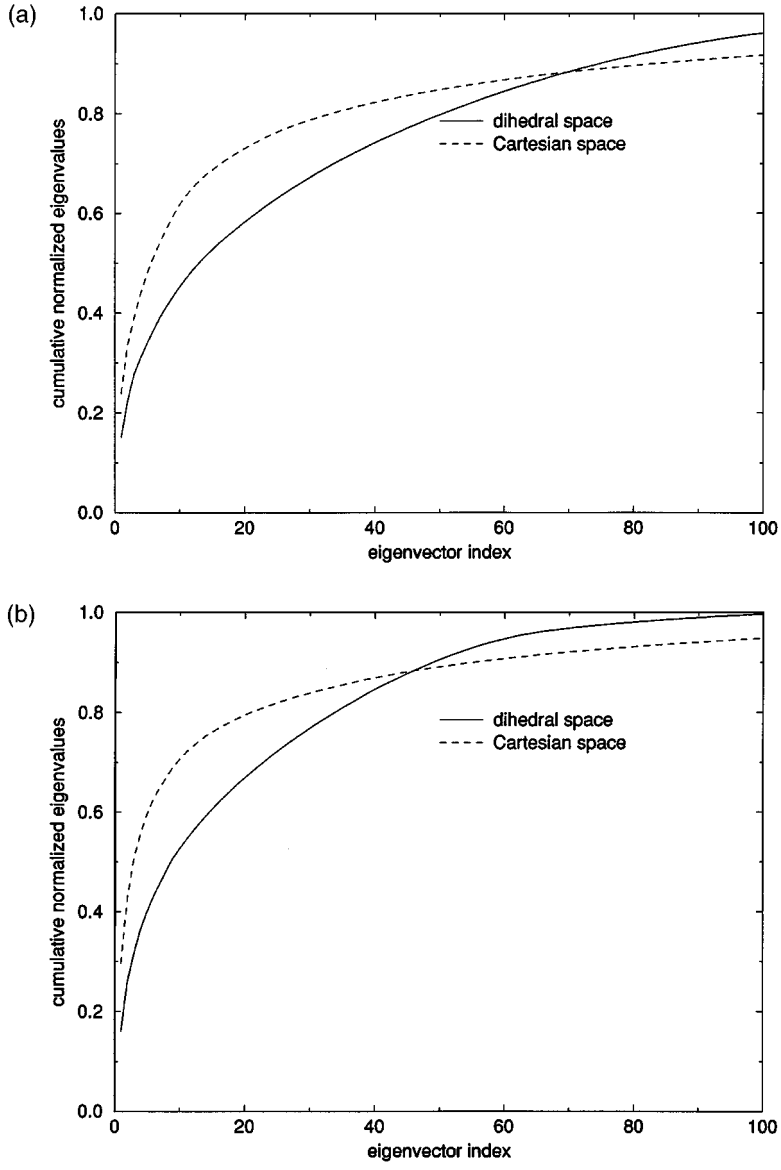

FIGURE 6. Comparison of the cumulative normalized eigenvalue curves of the Cartesian backbone ED analysis and the dihedral angle space ED analysis for (A) HPr and (B) SH3.

TABLE I.
Average Squared Inner Products for Projection of the First 10 (C-α) Eigenvectors of a Set onto First 10 of Another.

	HPr (A)	HPr (B)	HPr (NM)		SH3 (A)	SH3 (B)	SH3 (NM)
HPr (A)	1.0			SH3 (A)	1.0		
HPr (B)	0.39	1.0		SH3 (B)	0.50	1.0	
HPr (NM)	0.39	0.40	1.0	SH3 (NM)	0.47	0.44	1.0

An overlap of 0.4 –0.5 is considered to be significant because only a small amount of statistical / computational noise considerably decreases the overlap between essentially identical eigenvectors.

described a method for obtaining an approximation of high dimensional eigenvectors by splitting the covariance matrix into submatrices, significantly reducing the amount of memory and CPU time needed. Application to all atom covariance matrices constructed from the SH3 and HPr trajectories showed that the error introduced by this method is acceptable; only nonessential degrees of freedom are affected, which are in most cases irrelevant. Various methods are known for reducing the computational cost in diagonalizing large matrices in the normal modes analysis (overview in ref. 15). The method presented here is unique in that no assumptions are made concerning blocks of atoms in the matrix with correlations relatively independent of the rest of the protein. The eigenvectors of the basis set are built using sets of atoms spread out over the protein, implicitly taking into account correlations between domains in the protein.

We investigated with which set of atoms the most information is obtained at the least computational cost. Although the basic C-α analysis gives a good picture of the large concerted motions in the backbone, more information can be obtained by including all backbone atoms, which facilitates various standard structure analyses to be performed. Approximate side chain–side chain and side chain–backbone correlations can be obtained by including one extra particle per residue, being the geometrical center of the side chain or its terminal atom. Extra information at no significant extra cost is obtained, while there is no distortion of the C-α essential space.

We showed that it is possible to perform ED in ϕ/ψ space, as also recently demonstrated for NM analysis.[16] The ϕ/ψ angles involved in the first few essential eigenvectors are similar to those involved in the Cartesian backbone analysis. The eigenvectors are cleaner; ϕ/ψ fluctuations are more detailed; and the covariance matrix is smaller, requiring less computational resources. However, there are also some problems. The essential space is of larger dimensionality than that of Cartesian space analyses, and the reconstruction of 3-D atom trajectories using the dihedral angle eigenvectors is not straightforward. In addition, the essential eigenvectors in ϕ/ψ space do not necessarily correspond to large fluctuations in 3-D space. Because we want to study correlated displacements of atoms in three dimensions, it is simpler and possibly also better to use atomic fluctuations in Cartesian space to construct the covariance matrix. However, when it is necessary to investigate dihedral angle fluctuations in detail, as in locating a possible hinge point in a protein, it may be useful to perform ED in ϕ/ψ space in addition.

Acknowledgment

We thank Nico van Nuland for providing us with the HPr full solvent simulation.

References

1. A. Amadei, A. B. M. Linssen, and H. J. C. Berendsen, *Proteins: Struct. Funct. Genet.*, **17**, 412 (1993).
2. A. Amadei, A. B. M. Linssen, B. L. de Groot and H. J. C. Berendsen, In *Modelling of Biomolecular Structures and Mechanisms*, A. Pullman et al., Eds., Kluwer Academic Publishers, The Netherlands, 1995, p. 85.
3. A. E. Garcia, *Phys. Rev. Lett.*, **68**, 2696 (1992).
4. D. M. F. van Aalten, A. Amadei, G. Vriend, A. B. M. Linssen, G. Venema, H. J. C. Berendsen, and V. G. H. Eijsink, *Proteins: Struct. Funct. Genet.*, **22**, 45 (1995).
5. R. M. Scheek, N. A. J. van Nuland, B. L. de Groot, A. B. M. Linssen, and A. Amadei, *J. Biomol. NMR*, **6**, 106 (1995).
6. D. M. F van Aalten, A. Amadei, R. Bywater, J. B. C. Findlay, H. J. C. Berendsen, C. Sander, and P. F. Stouten, *Biophys. J.*, **70**, 684 (1996).
7. D. M. F. van Aalten, J. B. C. Findlay, A. Amadei, and H. J. C. Berendsen, *Protein Eng.*, **8**, 1129 (1996).
8. M. Levitt, C. Sander, and P. S. Stern, *J. Mol. Biol.*, **181**, 423 (1985).
9. W. F. Van Gunsterent and H. J. C. Berendsen, BIOMOS, Biomolecular Software, Laboratory of Physical Chemistry, University of Groningen, The Netherlands, 1987.
10. A. Musacchio, M. Noble, R. Pauptit, R. Wierenga, and M. Saraste, *Nature*, **359**, 851 (1992).
11. N. A. J. van Nuland, I. W. Hangyi, R. C. van Schaik, H. J. C. Berendsen, W. F. van Gunsteren, R. M. Scheek, and G. T. Robillard, *J. Mol. Biol.*, **237**, 544 (1994).
12. H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, *J. Phys. Chem.*, **91**, 6269 (1987).
13. J. B. Clarage, T. Romo, B. K. Andrews, B. M. Pettitt, and G. N. Phillips, *PNAS*, **92**, 3288 (1995).
14. D. L. D Caspar, *Structure*, **3**, 327 (1995).
15. S. Hayward and N. Go, *Annu. Rev. Phys. Chem.*, **46**, 223 (1995).
16. A. Kitao, S. Hayward, and N. Go, *Biophys. Chem.*, **52**, 107 (1994).