PROTEINS: Structure, Function, and Genetics 17:412–425 (1993)

Essential Dynamics of Proteins

Andrea Amadei, Antonius B.M. Linssen, and Herman J.C. Berendsen
Department of Biophysical Chemistry and BIOSON Research Institute, the University of Groningen, 9747 AG Groningen, The Netherlands

ABSTRACT **Analysis of extended molecular dynamics (MD) simulations of lysozyme in vacuo and in aqueous solution reveals that it is possible to separate the configurational space into two subspaces: (1) an "essential" subspace containing only a few degrees of freedom in which anharmonic motion occurs that comprises most of the positional fluctuations; and (2) the remaining space in which the motion has a narrow Gaussian distribution and which can be considered as "physically constrained." If overall translation and rotation are eliminated, the two spaces can be constructed by a simple linear transformation in Cartesian coordinate space, which remains valid over several hundred picoseconds. The transformation follows from the covariance matrix of the positional deviations. The essential degrees of freedom seem to describe motions which are relevant for the function of the protein, while the physically constrained subspace merely describes irrelevant local fluctuations. The near-constraint behavior of the latter subspace allows the separation of equations of motion and promises the possibility of investigating independently the essential space and performing dynamic simulations only in this reduced space.**

Key words: normal modes, constraint dynamics, molecular dynamics, lysozyme

INTRODUCTION

Functional proteins are generally stable mechanical constructs that allow certain types of internal motion to enable their biological function. The internal motions may allow the binding of a substrate or coenzyme, the adaptation to a different environment as in specific aggregation, or the transmission of a conformational adjustment to affect the binding or reactivity at a remote site, as in allosteric effects. Such functional internal motions may be subtle and involve complex correlations between atomic motions, but their nature is inherent in the structure and interactions within the molecule. It is a challenge to derive such motions from the molecular structure and interactions, to identify their functional role, and to reduce the complex protein dynamics to its essential degrees of freedom.

In this paper we investigate the correlations between atomic positional fluctuations in a protein, as derived from (nanosecond) molecular dynamics (MD) simulations, both in vacuum and in aqueous environment. By diagonalizing the covariance matrix of the atomic displacements, we find that most of the positional fluctuations are concentrated in correlated motions in a subspace of only a few (not more than 1%) degrees of freedom, while all other degrees of freedom represent much less important, basically independent, Gaussian fluctuations orthogonal to the "essential" subspace. The motion outside the essential subspace can be considered as essentially constrained. This offers the possibility of representing protein dynamics in the essential subspace only.

Our treatment differs from a harmonic or quasiharmonic normal mode analysis[1–4,14] in two ways. First, we do not analyze the motion but rather the positional fluctuations, without involving the atomic masses in the analysis. Our purpose is to identify an "irrelevant" subspace which may be considered essentially constrained. Second, we do not attempt to describe the motion in the "essential" subspace as harmonic, or even as mutually uncoupled, because it is neither harmonic nor uncoupled and such a treatment would restrict the mechanics of a protein to the level of uninteresting vibrations. The projection of a MD trajectory onto normal mode axes as carried out by Horiuchi and Gō[2] bears a resemblance to our analysis of displacements in the essential subspace (be it that the spaces onto which the motion is projected are not the same: they consider a dihedral angle subspace defined by normal modes of low frequency; we retain Cartesian coordinates and define the subspace from the covariance matrix). They find that the motions in the lower modes are restricted in narrower ranges than those derived by the harmonic approximation; we find a similar restriction due to nonlinear behavior and the presence of nonlinear constraints within the essential subspace.

The covariance matrix of atomic displacements has been used previously for quasiharmonic analy-

Received May 25, 1993; revision accepted August 13, 1993.
Address reprint requests to Dr. Herman J.C. Berendsen, Department of Biophysical Chemistry and BIOSON Research Institute, The University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands.

sis,[3,4] for entropy determination,[5–7] and for an analysis of collective motions.[1] We note that Ichiye and Karplus[1] construct the $N \times N$ covariance matrix, where N is the number of atoms in the system, of the vectorial inner products, while we construct the $3N \times 3N$ matrix of Cartesian displacements. The former indicates whether two particles move in the same or opposite directions, while the latter includes more complex correlations such as twist and mutually perpendicular displacements.

This paper introduces the theory, analyzes the dynamics for one biomolecular example, and suggests how dynamic simulation in the essential subspace can be performed in general for proteins. In subsequent papers we shall investigate the essential subspace, the relation between essential dynamics and biological function in more detail for specific proteins, and perform dynamics in the essential space only.

THEORY

We consider the dynamics of a protein in equilibrium in a given environment at a temperature T. Assume that a trajectory in phase space is available from a reliable MD simulation. We first eliminate the overall translational and rotational motion because these are irrelevant for the internal motion we wish to analyze. The precise method of eliminating the overall motion is not important: either the linear and angular moments are removed every step in the simulation, or the molecular axes are constructed each step by a least-squares translational and rotational fit. The result in any case is a Cartesian molecular coordinate system in which the atomic motions can be expressed. The internal motion is now described by a trajectory $\mathbf{x}(t)$, where $\mathbf{x}$ is a $3N$-dimensional vector of all atomic coordinates, represented by a column vector. If desired, $\mathbf{x}$ can represent a subset of atoms.

The correlation between atomic motions can be expressed in the covariance matrix C of the positional deviations:

$$C = \mathrm{cov}(\mathbf{x}) = \langle (\mathbf{x} - \langle \mathbf{x} \rangle)(\mathbf{x} - \langle \mathbf{x} \rangle)^{\mathrm{T}} \rangle \qquad (1)$$

where $\langle\ \rangle$ denote an average over time. The symmetric matrix C can always be diagonalized by an orthogonal coordinate transformation T:

$$\mathbf{x} - \langle \mathbf{x} \rangle = T\mathbf{q} \text{ or } \mathbf{q} = T^{\top}(\mathbf{x} - \langle \mathbf{x} \rangle) \qquad (2)$$

which transforms C into a diagonal matrix $\Lambda = \langle \mathbf{q}\mathbf{q}^{\top} \rangle$ of eigenvalues λ_i:

$$C = T\Lambda T^{\top} \text{ or } \Lambda = T^{\top}CT. \qquad (3)$$

The ith column of T is the eigenvector belonging to λ_i. When a sufficient number of independent configurations (at least $3N+1$) are available to evaluate C, there will be $3N$ eigenvalues, of which at least 6 representing overall translation and rotation are nearly zero. When a number of configurations, S, less than $3N+1$, is analyzed the total number of nonzero eigenvalues is at most $S-1$ since the covariance matrix will not have full rank.

The matrix C has the property of being always connected to the system constraints. In Appendix A we show that a subspace which is forbidden (or almost forbidden) for the motion is always fully defined by a subset of eigenvectors of the matrix C with zero or nearly zero eigenvalues. It is also important to note that the probability distribution of the displacements along the eigenvectors, although linearly uncorrelated, is not necessarily statistically independent. On the other hand, if a linear orthogonal transformation defines a subset of statistically independent generalized coordinates, then the unit vectors corresponding to this subset will always be eigenvectors of the covariance matrix C. The total positional fluctuation $\Sigma_i \langle (x_i - \langle x_i \rangle)^2 \rangle$ can be thought to be built up from the contributions of the eigenvectors:

$$\Sigma_i \langle (\mathbf{x}_i - \langle \mathbf{x}_i \rangle)^2 \rangle = \langle (\mathbf{x} - \langle \mathbf{x} \rangle)^{\top}(\mathbf{x} - \langle \mathbf{x} \rangle) \rangle = \langle \mathbf{q}^{\top} T^{\top} T \mathbf{q} \rangle = \langle \mathbf{q}^{\top} \mathbf{q} \rangle = \Sigma_i \langle q_i^2 \rangle = \Sigma_i \lambda_i. \qquad (4)$$

We choose to sort λ_i in order of decreasing value. Thus the first eigenvectors represent the largest positional deviation, and most of the positional fluctuations reside in a limited subset of the first n eigenvalues, where n is small compared to a total of $3N$.

We now divide the total $\mathbf{q}$-space in an *essential subspace* $q_1, \ldots q_n$ and the remaining space $q_{n+1}, \ldots, q_{3N}$. We denote coordinates in the essential subspace by ξ, and coordinates in the remaining subspace by s. As we shall show, the s-coordinates behave effectively as constraints: they have narrow Gaussian distributions with zero mean and do not contribute significantly to the positional fluctuations. Thus they behave as harmonic oscillators with a large force constant. As we shall show in Appendix B, such degrees of freedom may be treated as full constraints. This means that the mechanics in the essential subspace can be approximated by setting all $s = 0$, the approximation becoming exact if the force constants of the s-coordinates tend to infinity.

The equations of motion in the essential subspace can be obtained by applying the transformation T to the equations of motion of $\mathbf{x}$:

$$M'\ddot{\mathbf{q}} = T^{\top} M\ddot{\mathbf{x}} = -T^{\top} \nabla_x V(\mathbf{x}) = -\nabla_q V(\mathbf{q}). \qquad (5)$$

Here M is the diagonal matrix of particle masses m_i and M' is a transformed mass tensor

$$M' = T^{\top} MT \text{ or } M'_{ij} = \Sigma_k T_{ki} T_{kj} m_k. \qquad (6)$$

In (5) we have used the potential gradient transformation

$$\frac{\partial V(\mathbf{x})}{\partial x_i} = \sum_j \frac{\partial V(\mathbf{q})}{\partial q_j} \frac{\partial q_j}{\partial x_i} = \sum_j T_{ij} \frac{\partial V(\mathbf{q})}{\partial q_j}. \qquad (7)$$

The potential V can be expressed in $\mathbf{q}$, rather than in $\mathbf{x}$, by applying the coordinate transformation. If

the dynamics is approximated by applying full constraints to all s-coordinates (setting all s identically to zero), then V can be expressed in ξ only. The forces in s-space then vanish since the basically independent Gaussian distributions found for the s-coordinates imply that V can be approximated as

$$V(\xi,s) = V(\xi,s = 0) + \tfrac{1}{2}\Sigma_i k_i s_i^2, \quad i = n + 1, 3N - 6 \quad (8)$$

where k_i are the force constants. This means that the equations of motion of the constrained system are entirely restricted to the essential subspace, in spite of the apparent mixing between spaces caused by the mass tensor [Eq. (5)]. For the dynamic equations in the essential subspace only the upper left $n \times n$ block of M' is needed. Thus protein dynamics may be reduced to the essential subspace of very limited dimensionality. This has extremely important consequences for the simplification of protein dynamics, the consequences of which we shall pursue in a separate paper.

METHODS

Analysis was performed on the trajectories of two distinct simulations of hen egg white lysozyme.

A simulation in vacuum was performed by the authors, using the GROMOS simulation package and the GROMOS force field.[10] A starting structure was taken from the Brookhaven Protein Data Bank, entry 3LYZ. Including polar hydrogens, the system contained 1258 atoms. Nonpolar hydrogens were incorporated implicitly by the use of united atoms. In total a simulation of 1 nsec was performed, with a step size of 2 fsec. The temperature was kept at 298 K by coupling to an external temperature bath,[9] with a coupling constant $\tau = 0.01$ psec. Bond lengths were constrained using the procedure SHAKE.[8] Rotational motion around, and translational motion of the center of mass was removed every 0.5 psec to prevent conversion of thermal motions into overall rotational and translational ones. Nonbonded interactions were evaluated using a short cutoff range of 0.8 nm, within which interactions were calculated every time step. Interactions in the range of 0.8–1.2 nm were updated every 20 fsec. During the simulation, configurations were saved every 0.5 psec.

A. Mark kindly offered a 900 psec (100–1000 psec) trajectory of a simulation of lysozyme. This simulation, which included 5,345 water molecules, was performed at 300 K, also using the GROMOS package and the corresponding force field.[10] Here, configurations were saved every 0.05 psec. For further details concerning this calculation we refer to Smith et al.[13]

Before the covariance matrix was built, all configurations were fitted to the first configuration by first fitting the center of mass and next perform a least square fit procedure[11] on the C_α coordinates. Covariance matrices C were constructed from the position coordinates of the atoms (all atoms or C_α atoms only) according to

$$C_{ij} = \frac{1}{s}\sum_t \{x_i(t) - \langle x_i \rangle\}\{x_j(t) - \langle x_j \rangle\}. \quad (9)$$

Where S is the total number of configurations, $t = 1,2, \ldots S$, $x_i(t)$ are the position coordinates with $i = 1,2, \ldots 3N$, and N is the number of atoms from which C is constructed and $\langle x_i \rangle$ is the average of coordinate i over all configurations. Eigenvalues and their corresponding eigenvectors were calculated using the **QL** algorithm.[12] Diagonalization of the C_α matrices, of size 387 by 387, required 24 sec of CPU time on a single processor of a CONVEX 240 while diagonalizing the all-atom covariance matrix of the solvent simulation of size 3792 by 3792 required 20.4 hr on the same machine.

RESULTS AND DISCUSSION

Three different covariance matrices were diagonalized. The corresponding eigenvalues are shown in Figure 1, plotted in descending order against the corresponding eigenvector indices. Figure 1a shows the eigenvalues from the matrix that was constructed from (387) C_α coordinates in the vacuum simulation. In Figure 1b we show the eigenvalues as obtained from the (387) C_α coordinates from the solvent simulation. Finally, Figure 1c shows the eigenvalues obtained by analyzing the covariance matrix constructed from all-atom coordinates (3792) of the protein from the solvent simulation. In this case the first few eigenvalues (average square displacements) are one order of magnitude larger than in the previous panels, because the number of atoms involved in these displacements is approximately 10 times larger. Since the eigenvalues are average square displacements, it is clear from Figure 1 that the configurational space of the protein is not a homogeneous space, in terms of the motion along the eigenvector directions. As can be seen from Figure 1a and b, the eigenvalues from the solvent simulation show a steeper decrease than those from the vacuum simulation. One reason for this may be the fact that the force fields used are not equivalent. In the vacuum force field, full charges have been replaced by dipoles. This produces a weakening of the electrostatic interactions that, as we found, mainly affect the near constraints. As far as the methodology presented here is concerned there are no basic differences between vacuum and solvent simulation; so in the subsequent text we will show only the results obtained from the solvent simulation. At the end of this section the vacuum and solvent results will be compared.

The amount of motion associated to a subspace spanned by the first n eigenvectors can be defined as the corresponding subspace positional fluctuation [Eq. (4), for the summation over n] where the eigen-

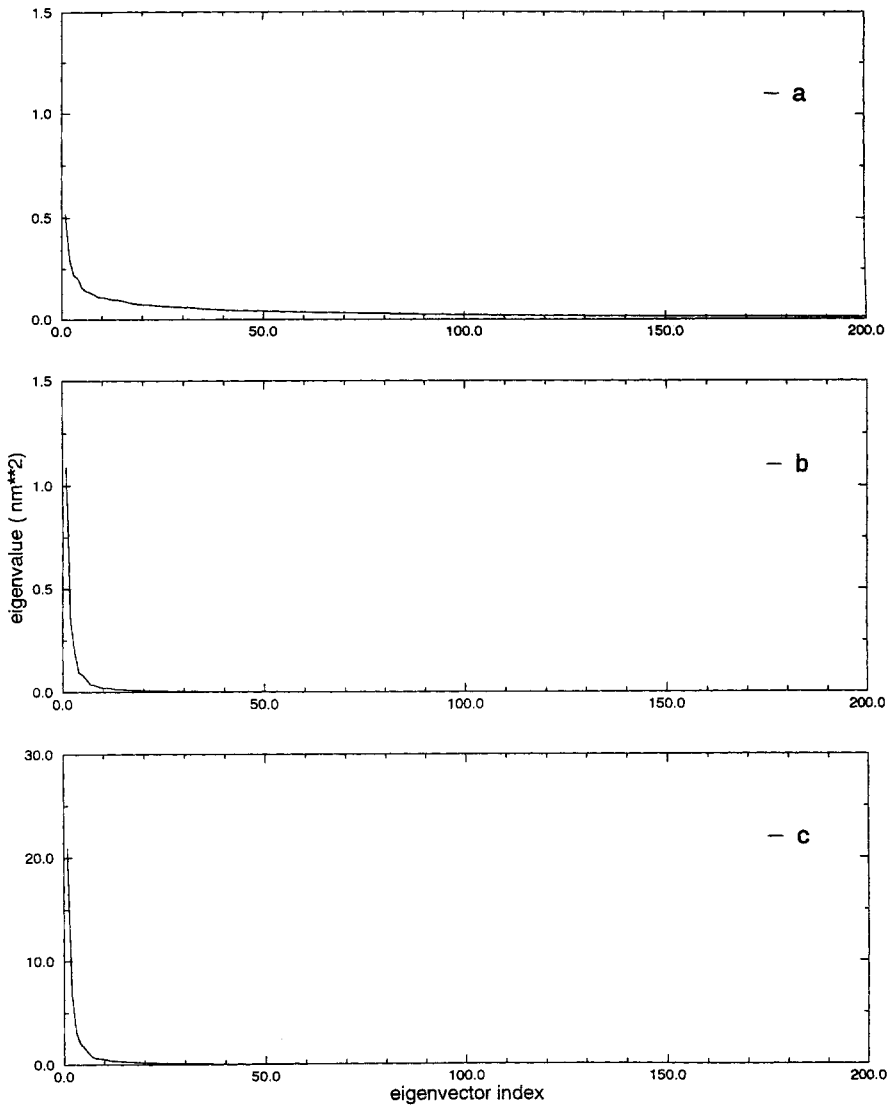

Fig. 1. Eigenvalues, in decreasing order of magnitude, obtained from **(a)** C_α coordinates covariance matrix from the vacuum simulation; **(b)** C_α coordinates covariance matrix from the solvent simulation; **(c)** all atoms coordinates covariance matrix from the solvent simulation.

values are ordered in descending order. In Figure 2 we show this relative subspace positional fluctuation (with respect to the total positional fluctuation) versus the increasing number of eigenvectors that span the subspace. In Figure 2a we show the results as obtained from the C_α matrix eigenvectors. Figure 2b shows the results from the all-atom analysis. From Figure 2a it can be seen that 90% of the total motion is described by the first 20 eigenvectors out of 387. If we analyze the motion due to all atoms (Fig. 2b) we see that the first 35 eigenvectors out of 3,792 contribute to 90% of the overall motion. This shows that most of the internal motion of the protein is confined within a subspace of very small dimension.

To have a closer look at the motion along the eigenvector directions one can project the trajectory onto these individual eigenvectors. In Figure 3 some projections of the C_α trajectory on the eigenvectors obtained from the C_α covariance matrix are plotted against time. It is clear from this figure that all motions that have not yet reached their equilibrium fluctuation belong to the first 10 eigenvectors. Figure 4 shows the sampling distribution functions for the displacements along the same eigenvectors, as well as the corresponding Gaussian functions with the same variance and average value. Obviously the only non-Gaussian distributions are again found within the first 10 eigenvectors. As is shown in Appendix B, Gaussian distribution functions are ex-

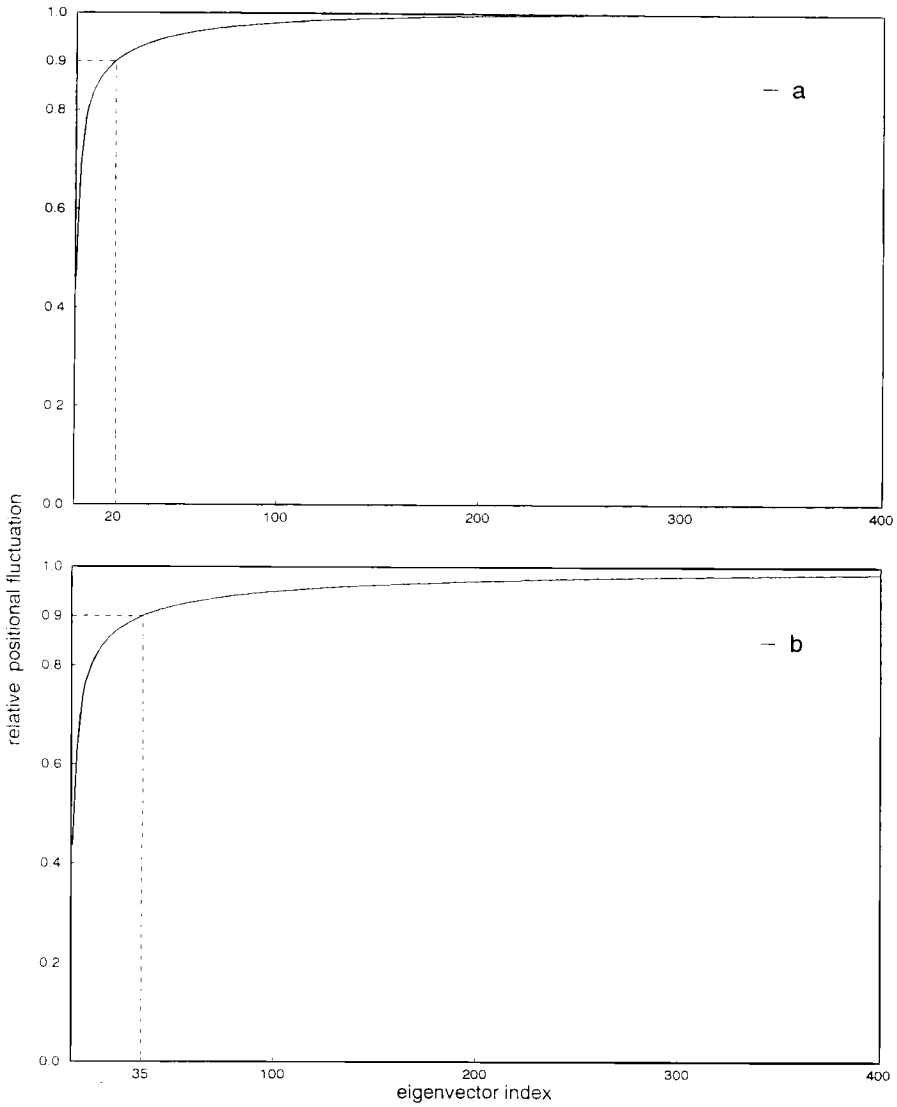

Fig. 2. (**a**) Relative positional fluctuation (see text) of the motions along the eigenvectors obtained from the C_α coordinates covariance matrix (solvent simulation). (**b**) Relative positional fluctuation of the motions along the eigenvectors obtained from the all atoms coordinates covariance matrix (solvent simulation).

pected for independent and harmonic (near constraint) motions that are already equilibrated. Figures 5 and 6 show the same, but now projections and distributions have been evaluated using the eigenvectors that were obtained from the all atoms covariance matrix. Just as in the case for the C_α analysis we find that all the motions that have not yet reached equilibrium fluctuation are confined within the first 10 eigenvectors. Also the only non-Gaussian distributions appear within the same eigenvectors.

We also noticed a great similarity between the motions along the first few eigenvectors of the C_α matrix and those along the first few eigenvectors derived from the all atoms matrix. To investigate this similarity further, we extracted the components from the all-atom eigenvectors that corresponded to the C_α coordinates and normalized the vectors that we obtained in this way. In Figure 7 the projections of these vectors on the eigenvectors of the C_α matrix are plotted. It is clear that the first 8 extracted vectors correspond to the first 8 C_α matrix eigenvectors. It should be mentioned that the length of these extracted vectors is approximately 20% of the whole length of the corresponding all-atom eigenvector, whereas the number of C_α atoms is about 10% of the total number of atoms in the protein. This indicates that the essential internal motion of the protein mainly involves the backbone atoms. We also noted that the displacements along the first 5 eigenvectors

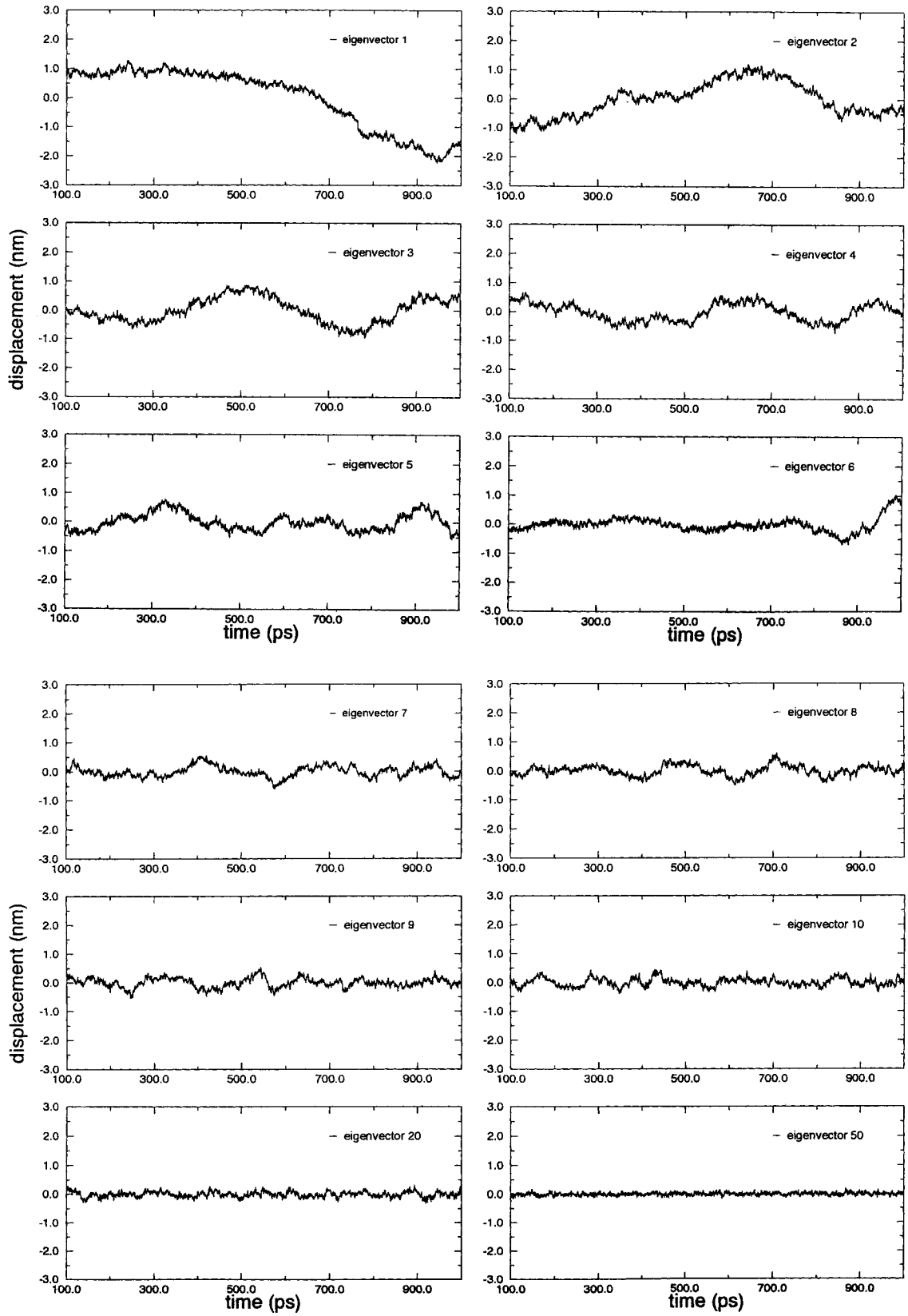

Fig. 3. Motions along several eigenvectors obtained from the C_α coordinates covariance matrix (solvent simulation).

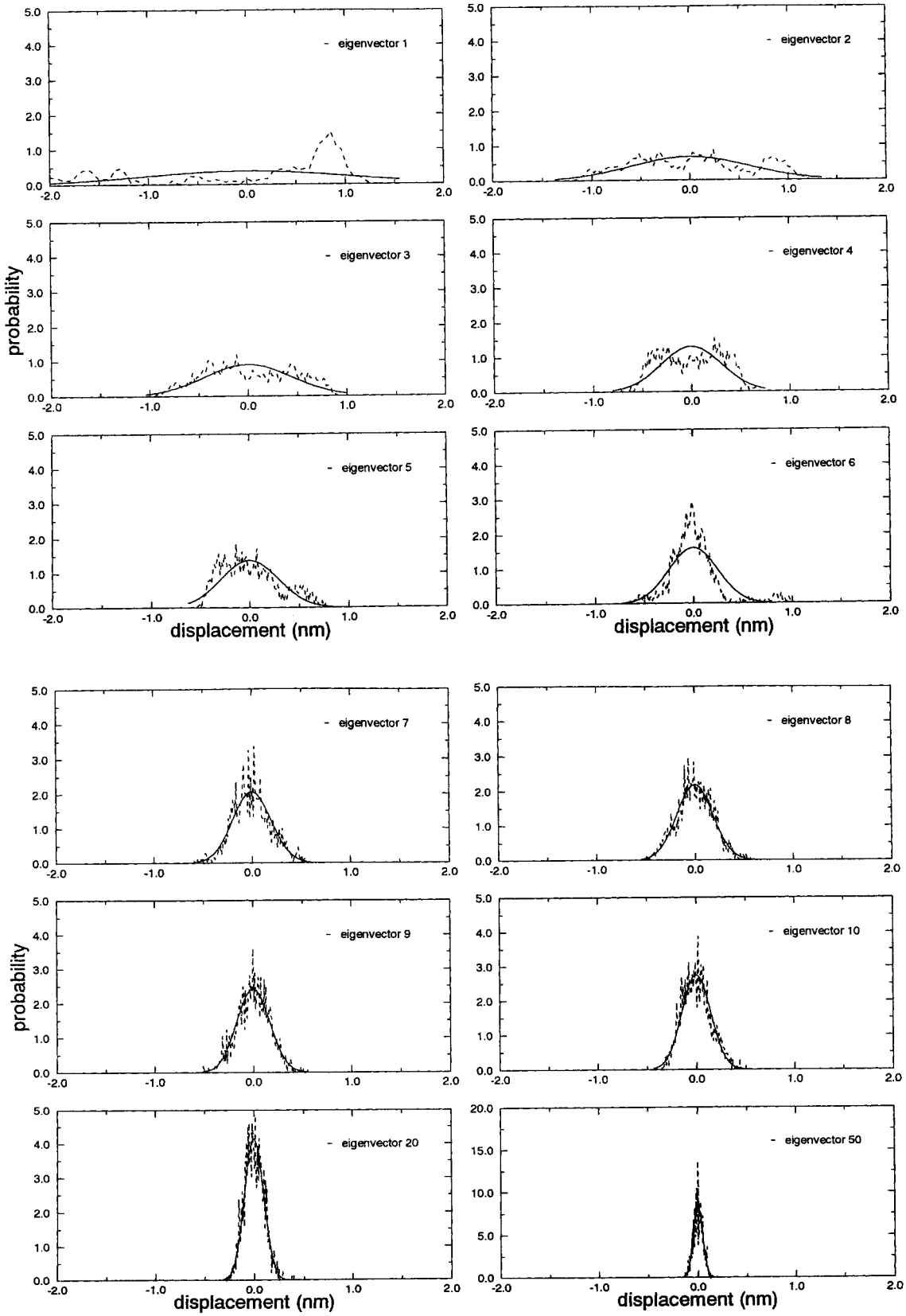

Fig. 4. Probability distributions for the displacements along several eigenvectors obtained from the C_α coordinates covariance matrix (solvent simulation). Solid line: Gaussian distributions derived from the eigenvalues of the corresponding eigenvectors. Dashed line: Sampling distributions.

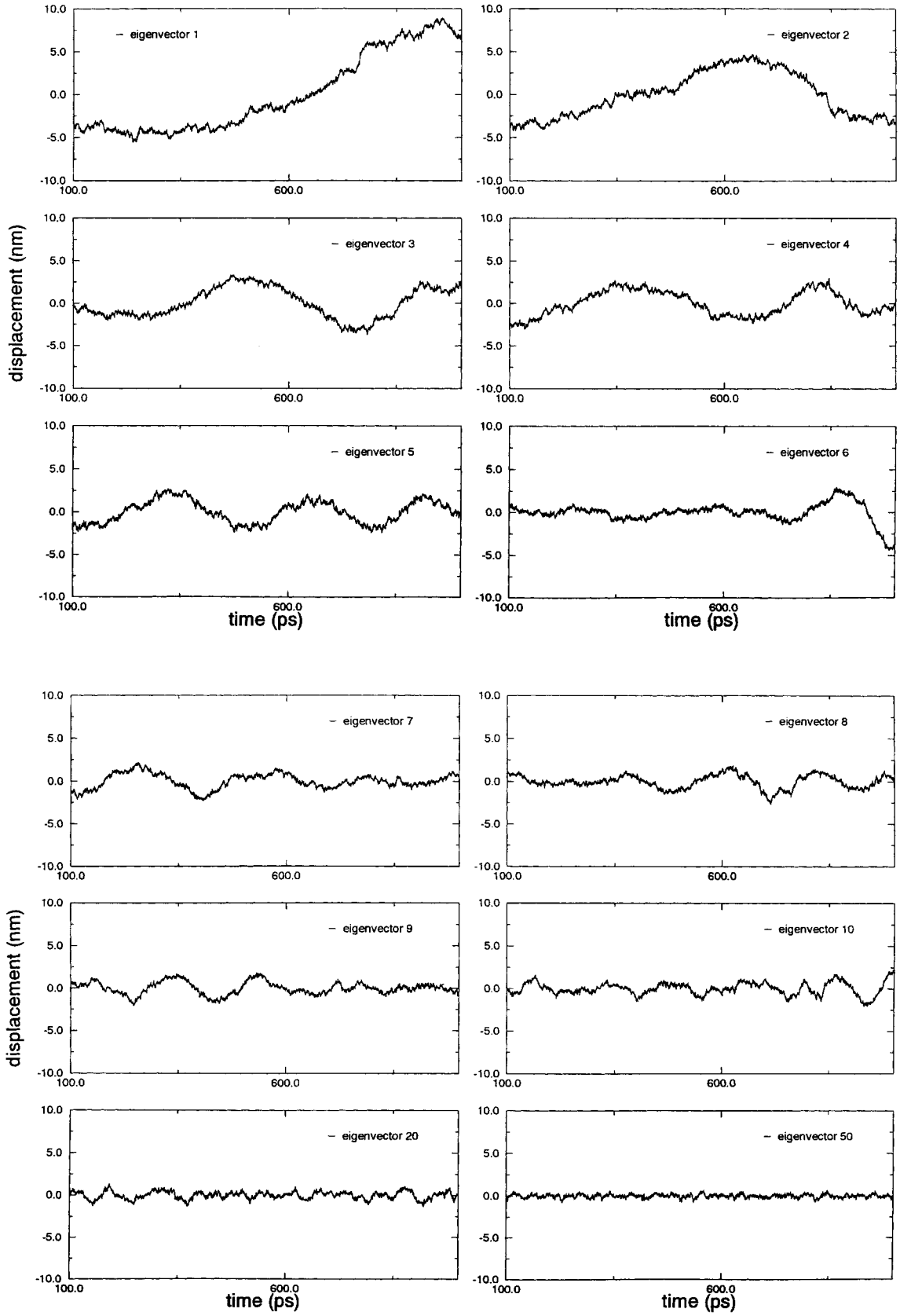

Fig. 5. Motions along several eigenvectors obtained from the all atoms coordinates covariance matrix (solvent simulation).

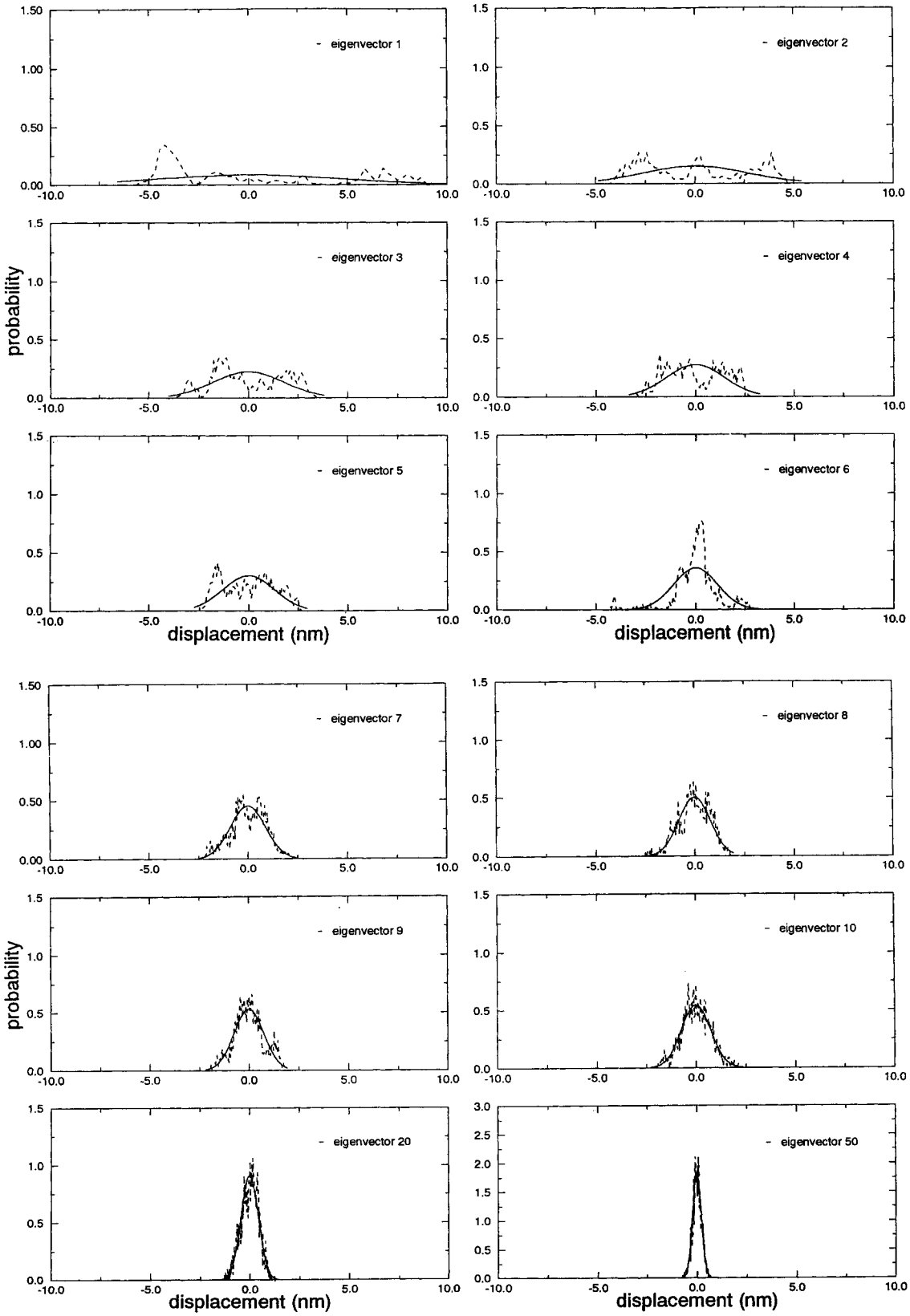

Fig. 6. Probability distributions for the displacements along several eigenvectors obtained form the all atoms coordinates covariance matrix (solvent simulation). Solid line: Gaussian distributions derived from the eigenvalues of the corresponding eigenvectors. Dashed line: Sampling distributions.

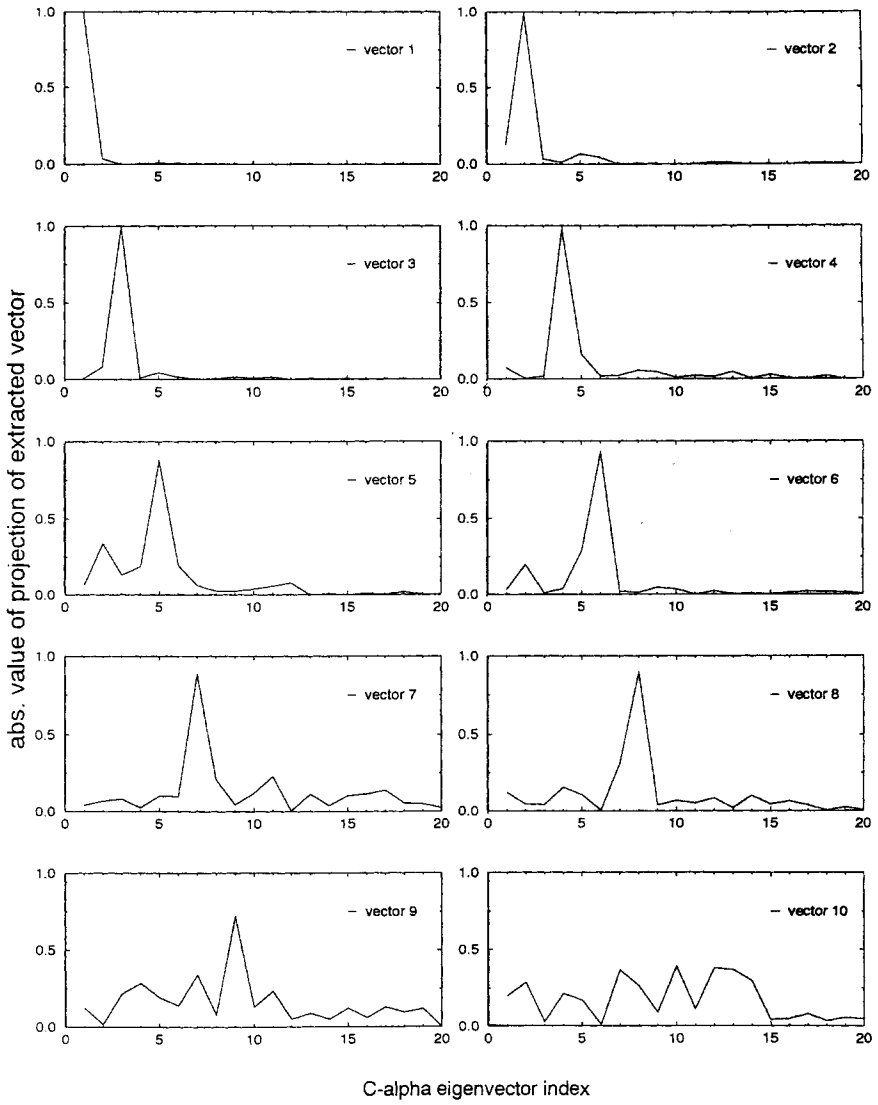

Fig. 7. Absolute values of the projections of the (normalized) extracted vectors coming from the first 10 eigenvectors of the all atoms coordinates covariance matrix (see text) on the eigenvectors obtained from the C_α coordinates covariance matrix (solvent simulation).

produced a large motion in the active site of the molecule. Figure 8 shows a superposition of 10 sequential projections of the C_α motion onto the first eigenvector, each separated by 100 psec (compare with Fig. 3). The catalytic site residues Glu-35 and Asp-52 are rigid, but the entrance to the active site cleft, including residues involved in substrate binding (59, 62, 63, 101, 107),[16] shows extensive flexibility. This motion, which also involves other loops in the protein, possibly affects the association and dissociation of substrates and products.

Figure 9 shows the trajectory projected on four planes, each defined by two all atoms matrix eigenvectors. In the planes of Figure 9a and b (respectively, eigenvectors 1 and 2 and eigenvectors 2 and 3) the trajectories are confined within narrower ranges than those expected from independent motions, suggesting the presence of a coupled force field. In Figure 9c and d (respectively, eigenvectors 1 and 50 and eigenvectors 20 and 50) the trajectories fill the expected ranges almost completely. This means that we are dealing with basically independent motions. We analyzed the vacuum simulation and compared the motion in the essential space with that of the solvent simulation. The motion in the vacuum simulation appears to be largely restricted to the carboxy terminal strand; the motion near the active site is no longer present. Therefore pure vacuum simulations appear not to be suitable for the study of biologically relevant motions.

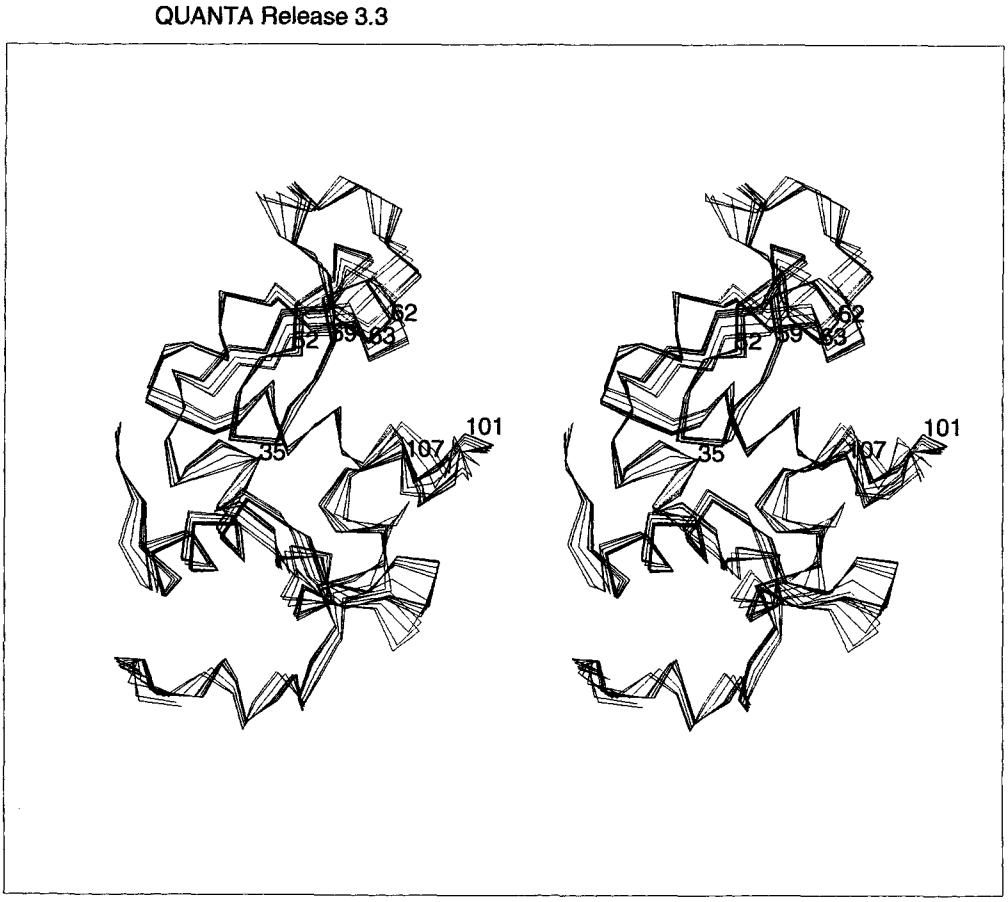

Fig. 8. Superposition of 10 configurations obtained by projecting the C_α motion onto the first eigenvector. Configurations are separated by 100 psec. Residues involved in the catalytic reaction (35 and 52) and in the binding of the substrate (59, 62, 63, 101, and 107) are indicated.

CONCLUSION

The analysis given in this article shows that the essential dynamics of lysozyme, and presumably of other globular proteins, can be described in a subspace of very small dimension (less than 1% of the original Cartesian space) consisting of linear combinations of Cartesian degrees of freedom defined in a molecule-fixed coordinate system. All other degrees of freedom can be considered as corresponding to irrelevant Gaussian fluctuations, behaving like near-constraints. The essential subspace itself is defined by the near-constraints, which are related to the mechanical structure of the molecule in a given conformation. We have strong evidence from inspection of the motion in the essential subspace of a few proteins studied up to now (lysozyme, thermolysin, and a subtilisin analog) that these motions are related to the functional behavior of the proteins such as opening and closing of the active site and hinge-bending motions between two domains enclosing the active site. The analysis of this behavior will be the subject of a subsequent study. A (major) conformation change to a different folded conformation may alter the characteristics of the essential subspace, while unfolding will lead to an increase of its dimensionality. If the unimportant fluctuations are replaced by full constraints, the motion in the essential subspace can be fully described by equations of motion within this subspace, thus reducing the complexity of protein dynamics considerably. The description in terms of the essential subspace also allows the evaluation of thermodynamic properties: (1) the eigenvalues of the near-constraints can be used to evaluate the entropy contribution of the near constraint degrees of freedom. (2) The potential energy within the essential subspace (which can be obtained by sampling) could be used for deriving the essential properties of the protein.

ACKNOWLEDGMENTS

We thank Dr. A.E. Mark from ETH in Zürich for offering us a 900 psec trajectory of lysozyme in solvent. We are also grateful to Daan van Aalten who applied the method presented in this paper to a few

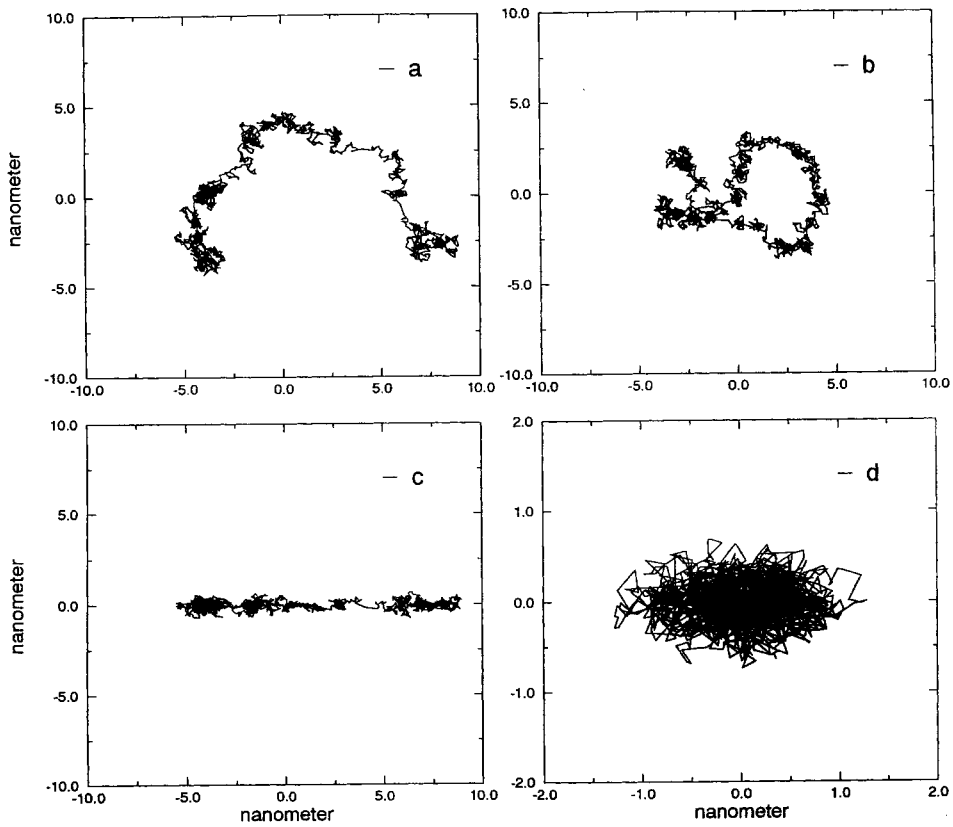

Fig. 9. Projections of the trajectory (solvent simulation) on planes defined by two eigenvectors from the all atoms coordinates covariance matrix. **(a)** Horizontal axis: displacement along first eigenvector. Vertical axis: displacement along second eigenvector. **(b)** Horizontal axis: displacement along second eigenvector. Vertical axis: displacement along third eigenvector. **(c)** Horizontal axis: displacement along first eigenvector. Vertical axis: displacement along 50th eigenvector. **(d)** Horizontal axis: displacement along 20th eigenvector. Vertical axis: displacement along 50th eigenvector.

other proteins. Finally, we gratefully acknowledge the Italian foundation "Cenci Bolognetti," Istituto Pasteur, for financially supporting Dr. A. Amadei in this research project.

REFERENCES

1. Ichiye, T., Karplus, M. Collective motions in proteins; a covariance analysis of atomic fluctuations in molecular dynamics and normal modes simulations. Proteins 11: 205–217, 1991.
2. Horiuchi, T., Gō, N. Projection of Monte Carlo and molecular dynamics trajectories onto the normal mode axes: human lysozyme. Proteins 10:106–116, 1991.
3. Teeter, M.M., Case, A.D. Harmonic and quasiharmonic descriptions of crambin. J. Phys. Chem. 94:8091–8097, 1990.
4. Perahia, D., Levy, R.M., Karplus, M. Motions of an alpha-helical peptide: Comparison of molecular and harmonic dynamics. Biopolymers 29:645–677, 1990.
5. Edholm, O., Berendsen, H.J.C. Entropy estimation from simulations of non-diffusive systems. Mol. Phys. 51:1011–1028, 1984.
6. Di Nola, A., Berendsen, H.J.C., Edholm, O. Free energy determination of polypeptide conformations generated by molecular dynamics. Macromolecules 17:2044–2050, 1984.
7. Karplus, M., Kushick, J.N. Method for estimating the configurational entropy of macromolecules. Macromolecules 14:325–332, 1981.
8. 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. Comput. Phys. 23:327–341, 1977.
9. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., Haak, J.R. Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81:3684–3690, 1984.
10. van Gunsteren, W.F., Berendsen, H.J.C. Groningen Molecular Simulations (GROMOS) Library Manual (Biomos, Groningen, 1987).
11. McLachlan, A.D. Gene duplications in the structural evolution of chymotrypsin. J. Mol. Biol. 128:49–79, 1979.
12. Press, W.H., Flannery, B.P., Teukolsky, S.A., Vettering, W.T. "Numerical Recipes." Cambridge University Press, 1987:357–359.
13. Smith, P.E., Brunne, R.M., Mark, A.E., van Gunsteren, W.F. Dielectric properties of trypsin inhibitor and lysozyme calculated from molecular dynamics simulations. J. Phys. Chem., 97:2009–2014, 1993.
14. Kitao, A., Hirata, F., Gō, N. The effects of solvent on the conformation and collective motions of protein: normal mode analysis and molecular dynamics simulations of melittin in water and in vacuum. Chem. Phys. 158:447–472, 1991.
15. Gallavotti, G. "Meccanica Elementare, 2nd ed. Boringhieri, 1986.
16. Stryer, L. "Biochemistry," 3rd ed. San Francisco: W.H. Freeman, 1988.

APPENDIX A

In this Appendix we show that every exact or approximate holonomic constraint is associated with

an exactly or approximately zero eigenvalue of the covariance matrix.

If we have P holonomic near-constraints in our system we may express them using an implicit form for the relation between Cartesian coordinates:

$$|G_i(\mathbf{x})| \leq \varepsilon \qquad i=1,2,\ldots P \tag{A1}$$

with $\varepsilon \geq 0$. If these constraints can be linearized around the average position of $\mathbf{x}$, we can replace the previous inequality by

$$|G_i(\langle\mathbf{x}\rangle) + \nabla G_i \cdot (\mathbf{x}-\langle\mathbf{x}\rangle)| \leq \varepsilon. \tag{A2}$$

Where ∇G_i is taken at $\langle\mathbf{x}\rangle$. Since

$$|G_i(\langle\mathbf{x}\rangle)| \leq \varepsilon \tag{A3}$$

we must have

$$|\nabla G_i \cdot (\mathbf{x}-\langle\mathbf{x}\rangle)| \leq 2\varepsilon. \tag{A4}$$

If we multiply the covariance matrix at the right side by ∇G_i we obtain

$$\begin{aligned} C\nabla G_i &= \langle(\mathbf{x}-\langle\mathbf{x}\rangle)(\mathbf{x}-\langle\mathbf{x}\rangle)^\top\rangle\nabla G_i \\ &= \langle(\mathbf{x}-\langle\mathbf{x}\rangle)(\mathbf{x}-\langle\mathbf{x}\rangle)^\top\nabla G_i\rangle \\ &= \langle(\mathbf{x}-\langle\mathbf{x}\rangle)[\nabla G_i\cdot(\mathbf{x}-\langle\mathbf{x}\rangle)]\rangle. \end{aligned} \tag{A5}$$

Because we have near constraints ε tends to zero, hence

$$\begin{aligned} \lim_{\varepsilon\to 0} C\nabla G_i &= \lim_{\varepsilon\to 0}\langle(\mathbf{x}-\langle\mathbf{x}\rangle)[\nabla G_i\cdot(\mathbf{x}-\langle\mathbf{x}\rangle)]\rangle \\ &= \langle(\mathbf{x}-\langle\mathbf{x}\rangle)\,\mathbf{0}\rangle \\ &= 0. \end{aligned} \tag{A6}$$

From (A6) it follows that every linear combination of gradients ∇G_i tends to be an eigenvector with an almost zero eigenvalue of C. Thus every time we have in a system P (linear) almost constraints as defined by (A2), then we will always obtain P corresponding eigenvectors with almost zero eigenvalues from C.

APPENDIX B

In this Appendix we show why it is possible to separate the motions in ξ and s subspaces.

The configurational probability density in the Cartesian coordinates of the particles of the system is given by

$$\rho(\mathbf{x}) = \frac{e^{-\beta V(\mathbf{x})}}{\int e^{-\beta V(\mathbf{x})}d\mathbf{x}} \tag{B1}$$

with $\beta = 1/kT$, and $V(\mathbf{x})$ is the potential energy expressed as a function of the particles coordinates. If we express $\mathbf{x}$ as

$$\mathbf{x} = T\mathbf{q} \tag{B2}$$

where T is the orthogonal transformation that diagonalizes the covariance matrix C and

$$\mathbf{q} = (\xi;s); \qquad \xi=\xi_i, \qquad i=1,2,\ldots n \tag{B3}$$

then the configurational probability density in $\mathbf{q}$-space is

$$\rho(\mathbf{q}) = \frac{e^{-\beta V(\mathbf{q})}}{\int e^{-\beta V(\mathbf{q})}d\mathbf{q}}. \tag{B4}$$

This follows from the fact that the Jacobian of an orthogonal transformation equals 1 or -1. Now we expand $V(\mathbf{q}) = V(\xi;s)$ around $s = 0$ to second order:

$$\begin{aligned} V(\xi;s) &= V(\xi;0) + \sum_i \frac{\partial V}{\partial s_i}(\xi;0)s_i \\ &\quad + \frac{1}{2}\sum_{i,j}\frac{\partial^2 V}{\partial s_i \partial s_j}(\xi;0)s_i s_j. \end{aligned} \tag{B5}$$

We first shall investigate when the experimental distributions are compatible with (B5). If we combine Eqs. (B3),(B4), and (B5) we obtain

$$\rho(\xi;s) = \frac{e^{-\beta V(\xi;0)}e^{-\beta\Delta V(\xi;s)}}{\int e^{-\beta V(\xi;0)}d\xi \int e^{-\beta\Delta V(\xi;s)}ds} \tag{B6}$$

where

$$\Delta V(\xi;s) = \sum_i \frac{\partial V}{\partial s_i}(\xi;0)s_i + \frac{1}{2}\sum_{i,j}\frac{\partial^2 V}{\partial s_i\partial s_j}(\xi;0)s_i s_j. \tag{B7}$$

If we obtain from a simulation for the $3N-6-n$ (there are $3N-6$ degrees of freedom in the molecular axes system) near-constraints

$$\rho(s) \cong \Pi_i\,(1/2\pi\sigma_i^2)^{1/2}\,e^{-s_i^2/2\sigma_i^2} \quad \sigma_i^2 = \lambda_i \tag{B8}$$

then this means that the displacements distribution of each s_i is given by an independent Gaussian. Using (B6) we may express $\rho(s)$ as

$$\rho(s) = \int \rho(\xi;s)d\xi = \frac{e^{-\beta V(\xi;0)}e^{-\beta\Delta V(\xi;s)}d\xi}{\int e^{-\beta V(\xi;0)}d\xi \int e^{-\beta\Delta V(\xi;s)}ds}. \tag{B9}$$

The independent Gaussians in Eq. (B8) are obtained only if

$$e^{-\beta\Delta V(\xi;s)} \cong e^{-\Sigma_i s_i^2/2\sigma_i^2} \tag{B10}$$

hence

$$\frac{\partial V}{\partial s_i}(\xi;0)s_i \cong 0 \quad \text{and} \quad \frac{\partial^2 V}{\partial s_i\partial s_j}(\xi;0) \cong (kT/\sigma_i^2)\delta_{ij}. \tag{B11}$$

This means that the potential energy can be approximated by

$$\begin{aligned} V(\xi;s) &\cong V(\xi;0) + \frac{1}{2}\sum_i (kT/\lambda_i)s_i^2 \\ &\quad i=n+1,\ldots 3N-6. \end{aligned} \tag{B12}$$

It should be noted that Eq. (B12) is not only valid when $(\partial V/\partial s_i)\cong 0$ but it will always be valid when $kT/\lambda_i \to \infty$. In this limit we have $s_i \to 0$ and then

$(\partial V/\partial s_i)s_i \to 0$. Using Eqs. (B6) and (B12) we may express the probability density in ξ,s-space as

$$\rho(\xi;s) \cong \rho(\xi)\,\Pi_i\,(1/2\pi\lambda_i)^{1/2}\,e^{-s_i^2/2\lambda_i} \tag{B13}$$

with

$$\rho(\xi) = \frac{e^{-V(\xi;0)/kT}}{\int e^{-V(\xi;0)/kT}d\xi}. \tag{B14}$$

Equation (B14) shows that the equilibrium density in the essential subspace is approximated by a Boltzmann distribution determined by the potential energy on the ξ-hyperplane where all s-coordinates are constrained to zero.

We shall now show that in the limit where λ_i, with $i=n+1, \ldots 3N-6$, tends to zero, Eq. (B14) also implies that the dynamics in the ξ-hyperplace can be described independently from any other coordinates. We call $\hat{\xi}_i$ with $i=1, \ldots, n$ the unit vectors corresponding to the hyperplane ξ coordinates and $\hat{s}_i$ with $i=1, \ldots,$ (3N–6–n) the unit vectors corresponding to the near-constraints coordinates s. It is always possible to define a new set of unit vectors, $\hat{s}'_j$ with $j=1, \ldots 3N-6-n$, linearly independent from the $\hat{\xi}_i$ unit vectors, such that[15]

$$\hat{\xi}_i^{\top} M \hat{s}_j' = \hat{s}_j'^{\top} M \hat{\xi}_i = 0 \tag{B15}$$

where M is the diagonal mass tensor. Then choosing as a basis set $\{\hat{\xi}_i,\hat{s}'_j\}$ we can express $\mathbf{x}$ and $\mathbf{q}$ in the coordinates corresponding to this basis set and we call this representation $\mathbf{q}'$, hence

$$\mathbf{x}(t) = \Theta_1\mathbf{q}'(t) \tag{B16}$$

$$\mathbf{q}(t) = \Theta_2\mathbf{q}'(t). \tag{B17}$$

Where Θ_1 and Θ_2 are the corresponding transformation matrices which are in general not orthogonal. We denote coordinates q'_i corresponding to $\hat{\xi}$ by ξ'_i and q'_j corresponding to $\hat{s}'$ by s'_j, so that $\mathbf{q}'=(\xi';s')$. Since $\mathbf{x}$ and $\mathbf{q}$ are represented in orthogonal basis sets we must have

$$x_l(t) = \Sigma_i(\hat{\xi}_i \cdot \hat{\mathbf{x}}_l)\xi'_i(t) + \Sigma_j(\hat{s}'_j \cdot \hat{\mathbf{x}}_l)s'_j(t) \tag{B18}$$

$$\xi_l(t) = \Sigma_i\,(\hat{\xi}_i \cdot \hat{\xi}_l)\xi'_i(t) + \Sigma_j(\hat{s}'_j \cdot \hat{\xi}_l)s'_j(t) \tag{B19}$$

$$s_l(t) = \Sigma_i(\hat{\xi}_i \cdot \hat{s}_l)\xi'_i(t) + \Sigma_j(\hat{s}'_j \cdot \hat{s}_l)s'_j(t) \tag{B20}$$

where $i=1, \ldots,n$, and $j=1, \ldots 3N-6-n$, and $\hat{\mathbf{x}}_l$ is the unit vector corresponding to x_l in the physical Cartesian space $\mathbf{x}$. Because by definition $\hat{\xi}_i \cdot \hat{\xi}_l = \delta_{il}$ and $\hat{\xi}_i \cdot \hat{s}_l = 0$, Eqs. (B19) and (B20) reduce to

$$\xi_l(t) = \xi'_l(t) + \Sigma_j(\hat{s}'_j \cdot \hat{\xi}_l)s'_j(t) \tag{B21}$$

$$s_l(t) = \Sigma_j(\hat{s}'_j \cdot \hat{s}_l)s'_j(t). \tag{B22}$$

Combining Eqs. (B12), (B21), and (B22) we obtain

$$V(\xi';s') \cong V(\xi';0) + \frac{1}{2}\sum_{ij}\kappa_{ij}s'_is'_j \tag{B23}$$

where κ_{ij} are force constants. Using Eq. (B16) in the equations of motion expressed in $\mathbf{x}$:

$$\Gamma\ddot{\mathbf{q}}' = -\nabla_{q'}V(\mathbf{q}') \tag{B24}$$

with

$$\Gamma = \Theta_1^{\top}M\Theta_1. \tag{B25}$$

From Eq. (B15) it follows that Γ is a block-diagonal matrix such that

$$\begin{bmatrix}\Gamma^{\xi'} 0\\ 0\,\Gamma^{s'}\end{bmatrix}\begin{bmatrix}\ddot{\xi}'\\ \ddot{s}'\end{bmatrix} = \begin{bmatrix}-\nabla_{\xi'}V(\xi';s')\\ -\nabla_{s'}V(\xi';s')\end{bmatrix}. \tag{B26}$$

Where $\Gamma^{\xi'}$ is an $n \times n$ matrix and $\Gamma^{s'}$ is a $(3N-6-n)\times(3N-6-n)$ matrix with

$$\Gamma^{\xi'}_{ij} = \hat{\xi}_i^{\top}M\hat{\xi}_j \quad i,j = 1, \ldots n \tag{B27a}$$

$$\Gamma^{s'}_{ij} = \hat{s}'^{\top}_i M\hat{s}'_j \quad i,j = 1, \ldots 3N-6 \tag{B27b}$$

From Eq. (B23) we have

$$-\nabla_{\xi'}V(\xi';s') \cong -\nabla_{\xi'}V(\xi';0). \tag{B28}$$

Combining equations (B26) and (B28) we finally obtain

$$\Gamma_{\xi'}\ddot{\xi} \cong -\nabla_{\xi'}V(\xi';0). \tag{B29}$$

Eq. (B29) implies that the motion in the ξ-subspace can be approximately described independently from the s'-coordinates, the approximation becoming exact as each $kT/\lambda_i \to \infty$, with $i=n+1, \ldots 3N-6$.