

Dynamic and Structural Analysis of Isotropically Distributed Molecular Ensembles

Jeanine J. Prompers and Rafael Brüschweiler*

Carlson School of Chemistry and Biochemistry, Clark University, Worcester, Massachusetts

ABSTRACT An efficient new method is presented for the characterization of motional correlations derived from a set of protein structures without requiring the separation of overall and internal motion. In this method, termed isotropically distributed ensemble (IDE) analysis, each structure is represented by an ensemble of isotropically distributed replicas corresponding to the situation found in an isotropic protein solution. This leads to a covariance matrix of the cartesian atomic positions with elements proportional to the ensemble average of scalar products of the position vectors with respect to the center of mass. Diagonalization of the covariance matrix yields eigenmodes and amplitudes that describe concerted motions of atoms, including overall rotational and intramolecular dynamics. It is demonstrated that this covariance matrix naturally distinguishes between “rigid” and “mobile” parts without necessitating a priori selection of a reference structure and an atom set for the orientational alignment process. The method was applied to the analysis of a 5-ns molecular dynamics trajectory of native ubiquitin and a 40-ns trajectory of a partially folded state of ubiquitin. The results were compared with essential dynamics analysis. By taking advantage of the spherical symmetry of the IDE covariance matrix, more than a 10-fold speed up is achieved for the computation of eigenmodes and mode amplitudes. IDE analysis is particularly suitable for studying the correlated dynamics of flexible and large molecules. *Proteins* 2002;46:177–189.

© 2001 Wiley-Liss, Inc.

Key words: protein dynamics; principal component covariance analysis; quasi-harmonic analysis; essential dynamics; isotropically distributed ensemble

INTRODUCTION

Detailed characterization of biomolecular motions and of their correlations represents an important step toward the understanding of biomolecular function. To analyze concerted molecular motions that underlie a set of conformations, originating, for example, from a molecular dynamics (MD) simulation, Monte Carlo (MC) simulation, nuclear magnetic resonance (NMR) structure determination, or a cluster of crystal structures, the covariance matrix of the cartesian atomic coordinates is often determined. The matrix elements provide information directly about corre-

lated fluctuations of pairs of atoms.^{1–3} Furthermore, the eigenvectors and eigenvalues of the covariance matrix yield collective motional modes and mode amplitudes, respectively. For n atoms, the dimension of the covariance matrix is either $3n$ or n , depending on whether the magnitudes alone, or the directions of the atomic motions as well, are of interest. Several closely related procedures are known by the terms quasiharmonic analysis (QHA),⁴ principal component analysis (PCA),⁵ and essential dynamics (ED).⁶ On the basis of these methods, important insights were gained into the motional characteristics of proteins and their biological function.⁷

A prerequisite of the QHA, PCA, and ED methods is the separation of overall and internal motion. While translational motions can be easily removed by centering each structure about its center of mass, the elimination of overall rotational motion is not rigorously possible, since biomolecules are not rigid bodies. In practice, overall rotational motion is reduced by superimposing each structure on a reference structure such that the root-mean-square deviation (RMSD) of atomic positions is minimized or that the tensors of inertia are aligned. For the alignment procedure, sometimes only a subset of reference atoms is used that is considered suitable² or that is found to be most rigid by an iterative procedure.³ The atom set selected for alignment significantly affects the outcome of the analysis, in particular for systems that exhibit a high degree of flexibility. Thus, these protocols are suitable primarily for the motional characterization of globular proteins that exhibit restricted intramolecular mobility.

Many proteins are intrinsically unstructured under physiological conditions. It was estimated that at least 30% of eukaryotic proteins contain significant amounts of disordered or partially disordered regions.^{8,9} Alternative computational methods for analysis are required to gain greater insight into the motional and structural ensemble properties of these systems. Because distances do not depend on the alignment, it has been proposed that interatomic distances and their covariances are analyzed

The Supplementary Material referred to in this article can be found at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/46_2/v46_2.html.

Grant sponsor: National Science Foundation; Grant number: MCB-9904875.

*Correspondence to: Rafael Brüschweiler, Carlson School of Chemistry and Biochemistry, Clark University, 950 Main Street, Worcester, MA 01610-1477. E-mail: bruschweiler@nmr.clarku.edu

Received 3 August 2001; Accepted 21 August 2001

instead of cartesian coordinates of atomic positions.³ As for n atoms, the number of interatomic distances grows with $n^2/2$, the computation of a full set of eigenmodes and mode amplitudes is impractical for larger systems.³

A new method that does not involve an alignment procedure is presented for characterization of correlated motions based on a covariance matrix calculated from atomic positions. The idea behind the approach is to represent each member of a set of structures as an isotropic orientational ensemble as if it were part of an isotropic solution. Thus, each structure is represented by an infinitely large number of replicas that probe an isotropic orientational distribution, which leads to a covariance matrix \mathbf{P} with elements

$$P_{ij} = \frac{1}{3} \langle \mathbf{r}_i \cdot \mathbf{r}_j \rangle \quad (1)$$

where \mathbf{r}_i is the position of atom i with respect to the center of mass, and the angular brackets indicate an average over the structures. This approach is called isotropically distributed ensemble (IDE) analysis.

The symmetry induced by the isotropic distribution considerably lowers the computational costs of the IDE analysis. It allows the reconstruction of the $3n$ -dimensional motional eigenmodes from eigenvectors of the $n \times n$ covariance matrix \mathbf{P} , while other approaches (QHA, PCA, ED) require diagonalization of a $3n \times 3n$ matrix.

In the Methods section, eq. 1 is derived, and some useful analytical properties are discussed. The approach is illustrated first for a model system consisting of a rotating methyl group attached to a rigid backbone. The method is then applied to a 5-ns MD trajectory of native ubiquitin and to a 40-ns MD trajectory of a partially folded state of ubiquitin.

METHODS

Covariance Matrix Analysis

Before the IDE approach is explained, we present a brief review of standard covariance matrix analysis of a set of N structures (snapshots) of a molecule in cartesian coordinate space generated, for example, by an MD simulation.^{4–6} To remove overall translation, each structure k defined by the positions of its n atoms $\mathbf{r}_i^{(k)}$ has its coordinate origin at the center of mass, such that

$$\sum_{i=1}^n m_i \mathbf{r}_i^{(k)} = 0$$

In addition, each structure is aligned with respect to its tensor of inertia or with respect to a reference structure for approximate removal of overall rotation. The 3D atomic position vectors $\mathbf{r}_i^{(k)}$ can be combined into a $3n$ -dimensional column vector $\mathbf{R}^{(k)}$. The $3n \times 3n$ cartesian covariance matrix \mathbf{C} is then given by

$$\mathbf{C} = \langle (\mathbf{R} - \langle \mathbf{R} \rangle)(\mathbf{R}^T - \langle \mathbf{R}^T \rangle) \rangle = \langle \mathbf{R}\mathbf{R}^T \rangle - \langle \mathbf{R} \rangle \langle \mathbf{R}^T \rangle \quad (2)$$

where the angular brackets $\langle \dots \rangle$ indicate averaging over the N snapshots $k = 1, \dots, N$. Matrix \mathbf{C} is symmetric and semipositive definite. Determination of eigenvectors and

eigenvalues of \mathbf{C} is the essence of QHA, PCA, and ED. Implementation of these methods in cartesian coordinates in the following discussion is collectively termed ED analysis.

To facilitate comparison of the ED results with IDE results (see Results), a modified ED analysis can be carried out for a reduced $n \times n$ covariance matrix \mathbf{C}_{red} , which is obtained from the $3n \times 3n$ covariance matrix \mathbf{C} of eq. 2 by partial trace formation of 3×3 sub-blocks. The corresponding n -dimensional eigenmodes \bar{c}_i , $\mathbf{C}_{\text{red}}\bar{c}_i = \lambda_i\bar{c}_i$, reflect correlated modulations of groups of atoms, but they do not yield the directions of the motions in 3D space.

IDE Analysis

An MD trajectory of a protein in explicit solvent typically reaches the nanosecond or tens-of-nanosecond time-scale, which precludes sampling of all conceivable conformations. In addition, a molecule undergoes overall reorientational tumbling motion with a correlation time within the nanosecond range, which is usually insufficiently sampled during the simulation. By contrast, many biomolecules in aqueous solution form isotropic solutions with each internal conformation adopting a uniform overall orientational probability distribution. Although it is computationally expensive to extend the length of a MD trajectory to sample a larger portion of conformational space, it is straightforward to generate for each snapshot a large number of replicas that are oriented in a way that approximates an isotropic distribution. Thus, each snapshot $\mathbf{R}^{(k)}$ is replaced by a set of conformations $\{\mathbf{R}_1^{(k)}, \mathbf{R}_2^{(k)}, \dots, \mathbf{R}_M^{(k)}\}_{\text{iso}}$ with the same internal conformation as $\mathbf{R}^{(k)}$ but different overall orientations. For a suitably chosen and sufficiently large number M of orientations, the set of conformations will approximate an isotropic distribution. Application of this procedure to each MD snapshot leads to a new ensemble consisting of $N \cdot M$ structures representing an isotropic solution; i.e., every internal conformation sampled during the MD simulation has a uniform overall orientational probability distribution. Thus, incomplete overall reorientational motion during the MD simulation is overcome in a rigorous way.

While the covariance matrix belonging to this new ensemble can, in principle, be calculated according to eq. 2, in practice the computational effort becomes quickly large. For example, for a molecule with $n = 1,000$ atoms with an original set of $N = 1,000$ structures each of them represented by $M = 1,000$ replica structures that approximate an isotropic distribution, computation of \mathbf{C}_{iso} requires about $9n^2NM/2 > 4 \cdot 10^{12}$ multiplications.

Fortunately, averaging of covariance matrix \mathbf{C} over an isotropic distribution of structures can be carried out analytically. Consider an individual structure $\mathbf{R}^{(k)}$, which is rotated about the origin $\mathbf{R}'^{(k)} = \mathbf{U}(\alpha, \beta, \gamma)\mathbf{R}^{(k)}$, where $\mathbf{U}(\alpha, \beta, \gamma)$ is an orthogonal rotation matrix and α, β, γ are the three Euler angles. Each set of structures $\mathbf{R}'^{(k)}$ contributes an isotropically averaged matrix $\mathbf{C}^{(k)}$ to the IDE covariance matrix \mathbf{K}

$$\mathbf{K} = \mathbf{C}_{\text{iso}} = \frac{1}{N} \sum_{k=1}^N \mathbf{C}^{(k)} \quad (3)$$

where

$$\mathbf{C}^{(k)} = \langle \mathbf{R}^{(k)} \mathbf{R}^{(k)T} \rangle_{\text{iso}} = \langle \mathbf{U}(\alpha, \beta, \gamma) \mathbf{R}^{(k)} \mathbf{R}^{(k)T} \mathbf{U}(\alpha, \beta, \gamma)^T \rangle_{\alpha\beta\gamma} \quad (4)$$

The angular brackets indicate an isotropic average over the Euler angles α, β, γ and the relationship $\langle \mathbf{R}^{(k)} \rangle_{\text{iso}} = \langle \mathbf{U}(\alpha, \beta, \gamma) \mathbf{R}^{(k)} \rangle_{\alpha\beta\gamma} = 0$ was used.

Matrix $\mathbf{R}^{(k)} \mathbf{R}^{(k)T}$ consists of $n^2 3 \times 3$ sub-blocks $\mathbf{B}_{ij}^{(k)}$ with elements $B_{ij,\mu\nu}^{(k)} = r_{i\mu}^{(k)} r_{j\nu}^{(k)}$ ($\mu, \nu = x, y, z$). All off-diagonal elements of $B_{ij,\mu\nu}^{(k)}$ (i.e., $\mu \neq \nu$) average to zero for an isotropic distribution. For example, if for a given orientation $r_{ix}^{(k)} r_{jy}^{(k)} = a$, an equally probable orientation is obtained by a 180° rotation about the x -axis, for which $r_{ix}^{(k)} r_{jy}^{(k)} = -a$, canceling the former contribution. By contrast, because the trace remains invariant under rotation the three diagonal elements of $\mathbf{B}_{ij}^{(k)}$ become identical and are equal to one-third times the matrix trace of $\mathbf{B}_{ij}^{(k)}$, i.e., $B_{ij,\mu\mu}^{(k)} = \text{Tr}(\mathbf{B}_{ij}^{(k)})/3$. Consequently, the isotropic average $\langle \mathbf{B}_{ij}^{(k)} \rangle_{\text{iso}}$ can be expressed as

$$\langle \mathbf{B}_{ij}^{(k)} \rangle_{\text{iso}} = \langle \mathbf{U}(\alpha, \beta, \gamma) \mathbf{B}_{ij}^{(k)} \mathbf{B}_{ij}^{(k)T} \mathbf{U}(\alpha, \beta, \gamma)^T \rangle_{\alpha\beta\gamma} = \frac{1}{3} \text{Tr}\{\mathbf{B}_{ij}^{(k)}\} \mathbf{1}_3 \quad (5)$$

where $\mathbf{1}_3$ is the 3×3 unity matrix. Using the relationship $\text{Tr}\{\mathbf{B}_{ij}^{(k)}\} = r_{ix}^{(k)} r_{jx}^{(k)} + r_{iy}^{(k)} r_{jy}^{(k)} + r_{iz}^{(k)} r_{jz}^{(k)} = \mathbf{r}_i^{(k)} \cdot \mathbf{r}_j^{(k)}$, matrix $\mathbf{C}^{(k)}$ can be written as the direct matrix product $\mathbf{C}^{(k)} = \frac{1}{3} \mathbf{P}^{(k)} \otimes \mathbf{1}_3$ where $\mathbf{P}^{(k)}$ is an $n \times n$ matrix with elements $P_{ij}^{(k)} = \mathbf{r}_i^{(k)} \cdot \mathbf{r}_j^{(k)}$. According to eq. 3, averaging of $\mathbf{C}^{(k)}$ over all N snapshots yields the isotropically averaged $3n \times 3n$ covariance matrix

$$\mathbf{K} = \frac{1}{3} \langle \mathbf{P}^{(k)} \rangle_k \otimes \mathbf{1}_3 \quad \text{with elements} \quad K_{i\mu, j\nu} = \frac{1}{3} \langle \mathbf{r}_i^{(k)} \cdot \mathbf{r}_j^{(k)} \rangle_k \delta_{\mu\nu} \quad (6)$$

where the angular brackets indicate averaging over the N snapshots. The particular form of eq. 6 implies that each eigenvalue is threefold degenerate, a direct consequence of the isotropic symmetry of \mathbf{K} . To obtain the eigenvectors and eigenvalues of \mathbf{K} , it is sufficient to diagonalize the $n \times n$ matrix

$$\mathbf{P} = \frac{1}{3} \langle \mathbf{P}^{(k)} \rangle_k \quad \text{with elements} \quad P_{ij} = \frac{1}{3} \langle \mathbf{r}_i \cdot \mathbf{r}_j \rangle \quad (7)$$

which yields

$$\mathbf{P} \bar{p}_i = \lambda_i \bar{p}_i \quad (8)$$

In the following analysis, we assume that the eigenvalues are sorted such that λ_1 is the smallest and λ_n is the largest eigenvalue. From the n -dimensional eigenvectors \bar{p}_i the $3n$ -dimensional orthogonal eigenvectors $\bar{P}_{i,1}, \bar{P}_{i,2}, \bar{P}_{i,3}$, that are threefold degenerate with eigenvalue λ_i , can be calculated using

$$\bar{P}_{i,1} = \bar{p}_i \otimes \bar{e}_1, \quad \bar{P}_{i,2} = \bar{p}_i \otimes \bar{e}_2, \quad \bar{P}_{i,3} = \bar{p}_i \otimes \bar{e}_3 \quad (9)$$

where the vectors $\bar{e}_1, \bar{e}_2, \bar{e}_3$ form an (arbitrary) orthonormal basis in 3D space.

Computation of matrix \mathbf{P} (eq. 7) for $N = 1,000$ structures with $n = 1,000$ atoms requires about $3n^2N/2 = 1.5 \cdot 10^9$ multiplications, which is much less than the $4 \cdot 10^{12}$ multiplications used when the isotropic distribution is constructed explicitly. Since the off-diagonal elements of \mathbf{P} correspond to motional covariances and the diagonal elements to motional variances, the cross-correlation coefficient r_{ij} of the motions of atom i and j is determined by

$$r_{ij} = \frac{P_{ij}}{(\bar{P}_{ii} \bar{P}_{jj})^{1/2}} \quad (10)$$

Certain modes, such as overall rotational modes, can be selectively eliminated from eq. 10 by using the spectral representation for the modified matrix \mathbf{P}' with elements

$$P'_{uv} = \sum_i \lambda_i \bar{p}_{i,u} \bar{p}_{i,v}$$

where the sum ranges only over modes that are to be included.

Covariance matrices \mathbf{K} and \mathbf{P} do not depend on how well the original MD trajectory samples overall rotation, as each snapshot has been replaced by an isotropic orientational distribution. \mathbf{K} and \mathbf{P} also do not contain explicit information about timescales. For example, no information can be gained about anisotropic overall tumbling behavior from matrices \mathbf{K} and \mathbf{P} alone. Such information can be retrieved by projecting MD snapshot k taken at time t , $\mathbf{R}^{(k)} = \mathbf{R}(t)$, on the eigenmodes

$$a_{i,j}(t) = \bar{P}_{i,j} \cdot \mathbf{R}(t) \quad i = 1, \dots, n; \quad j = 1, 2, 3 \quad (11)$$

where coefficients $a_{i,j}(t)$ correspond to the excursion (amplitude) of structure $\mathbf{R}(t)$ along mode $\bar{P}_{i,j}$. Conversely, $\mathbf{R}(t)$ can be represented as a linear combination of the $\bar{P}_{i,j}$:

$$\mathbf{R}(t) = \sum_{ij} a_{i,j}(t) \bar{P}_{i,j} \quad (12)$$

The sum of the squares of the three projection coefficients

$$a_i^2(t) = a_{i,1}^2(t) + a_{i,2}^2(t) + a_{i,3}^2(t) \quad (13)$$

reflects the total excursion squared of snapshot $\mathbf{R}(t)$ along mode \bar{p}_i .

For an internally static molecule, matrix \mathbf{P} of eq. 7 has elements $\mathbf{r}_i \cdot \mathbf{r}_j/3$, which, besides the factor one-third, is identical to the metric matrix used in distance geometry.^{10,11} For this special case, matrix \mathbf{P} has three positive eigenvalues and $n - 3$ zero eigenvalues. According to distance geometry theory, the normalized eigenvectors $\bar{P}_{n-2}, \bar{P}_{n-1}, \bar{P}_n$ belonging to the three positive eigenvalues contain the x, y, z -coordinates of the individual atoms if they are scaled by $(3\lambda_{n-2})^{1/2}$, $(3\lambda_{n-1})^{1/2}$, and $(3\lambda_n)^{1/2}$, respectively. According to eq. 8, the three positive eigenvalues determine the radius of gyration, R_g ,

$$R_g^2 = \frac{1}{n} \sum_{i=1}^n \mathbf{r}_i \cdot \mathbf{r}_i = \frac{3}{n} (\lambda_{n-2} + \lambda_{n-1} + \lambda_n) \quad (14)$$

provided that the center of mass coincides with the geometric center.

In the presence of internal motions, matrix \mathbf{P} of eq. 7 generally has $n - 1$ positive eigenvalues and one zero eigenvalue. The zero eigenvalue reflects the absence of overall translational motion. R_g^2 averaged over the whole ensemble is

$$\langle R_g^2 \rangle = \frac{3}{n} \sum_{i=1}^n \lambda_i = \frac{3}{n} \text{Tr}\{\mathbf{P}\} \quad (15)$$

The separability of internal and overall motion is reflected in the separability index:

$$g = \sum_{i=1}^n \lambda_i / \sum_{i=1}^{n-3} \lambda_i \quad (16)$$

The larger g , the better overall and internal motions fulfill separability. In the limiting case of a structure undergoing infinitesimally small amplitude motion, separability is perfectly fulfilled and $g \rightarrow \infty$, while for macromolecules with a finite amount of motion g is finite.

To each eigenmode \bar{q}_i ($= \bar{p}_i$ or \bar{c}_i), the collectivity κ_i can be assigned that describes the effective number of atoms that is significantly affected by this mode^{12,13}

$$\kappa_i = \frac{1}{n} \exp \left\{ - \sum_{l=1}^n |\bar{q}_{i,l}|^2 \log |\bar{q}_{i,l}|^2 \right\} \quad (17)$$

where $|\bar{q}_{i,l}|^2$ is the l th element squared of eigenmode \bar{q}_i .

The overlap of the subspaces spanned by the eigenmodes \bar{p}_i and \bar{c}_i of the IDE covariance matrix \mathbf{P} and the ED covariance matrix \mathbf{C} can be defined as

$$O_{pc}^{(r)} = \frac{1}{r} \text{Tr} \left\{ \left(\sum_{j=0}^{r-1} \bar{p}_{n-3-j} \cdot \bar{p}_{n-3-j}^T \right) \cdot \left(\sum_{j=0}^{r-1} \bar{c}_{n-j} \cdot \bar{c}_{n-j}^T \right) \right\} \quad \text{with } r < n - 3 \quad (18)$$

where the three largest IDE eigenvectors (associated with overall rotational motion) were excluded (see Results). The overlap measure $O_{pc}^{(r)}$ is based on the well-known trace metric that provides a convenient similarity measure of matrices of arbitrary dimension. The overlap of eigenvectors of two ED covariance matrices \mathbf{C} and \mathbf{C}' is defined analogously (see eq. 7 of ref. 13).

MD Simulations of Ubiquitin

Two MD simulations were carried out using the program CHARMM 24^{14,15} under periodic boundary conditions. The first trajectory corresponds to native ubiquitin and the second to a partially folded state of ubiquitin. For the

native trajectory, an all-atom representation of the x-ray structure of ubiquitin¹⁶ was embedded in a cubic box including 2,909 explicit water molecules. The simulation was performed at a temperature of 300 K for 6 ns. More details on the first 1.5 ns of the trajectory and the full 6-ns trajectory have been described elsewhere.^{17,18} The first 1 ns of the MD simulation was used for equilibration, and a total of 1,000 snapshots with a time increment of 5 ps were analyzed from the final 5 ns of the trajectory (segment 1–6 ns).

For the trajectory of the partially folded state, ubiquitin was simulated under the conditions experimentally known to induce the A-state, i.e., a 60%/40% (v/v) methanol–water mixture at low pH.^{19–21} An all-atom representation of the protein was protonated and immersed in a cubic box containing a total of 1,727 water molecules and 1,102 methanol molecules. During the first part of the 70-ns trajectory, the temperature was varied to guide and speed up the conformational transformation of the native state into a state resembling the A-state, while the final 42.8 ns (segment 27.2–70.0 ns) of the trajectory were performed at 300 K. More details on the first 33 ns of this simulation have been described elsewhere.¹⁸ In the present study, 800 snapshots of the final 40 ns of the trajectory (segment 30–70 ns), sampled at a time increment of 50 ps, were analyzed. The MD trajectories were processed using MMTK,²² IDE and ED analyses were implemented in Python.

RESULTS

Rotation of Methyl Group Attached to a Rigid Backbone

The IDE method is illustrated first for a model system consisting of a rigid backbone structure with a rotating methyl group attached. The backbone is represented by 10 consecutive C α atoms of residues 24–33 of the α -helix of the MD snapshot at 3.5 ns of native ubiquitin, while the methyl group is represented by the side-chain methyl hydrogens of Ala28. For the first calculation, the methyl hydrogens were considered fixed; i.e., no rotation about the threefold symmetry axis was permitted. Diagonalization of matrix \mathbf{P} defined in eq. 7 constructed from this single structure yields three eigenmodes with nonzero eigenvalues, presented in Table I. The sum of these eigenvalues determines the radius of gyration, $R_g = 4.8$ Å, according to eq. 14. These three modes are completely determined by the 3D structure, and vice versa. They reflect rigid-body overall rotational motion and are termed structural modes.

Next, an ensemble of three structures was generated from the structure described above by cyclic permutations of the three methyl hydrogen positions corresponding to methyl jump motion between three rotameric states. Matrix \mathbf{P} yields five eigenmodes with nonzero eigenvalues, presented in Table I. The three largest modes correspond to the overall rotational (structural) modes with eigenvalues slightly smaller than for the static case. The two smallest amplitude modes are degenerate; they describe the internal rotation of the methyl hydrogen atoms.

TABLE I. Rotation of a Methyl Group Attached to a Rigid Backbone

IDE static ^a	IDE ^b	ED C ^α ^c	ED all ^d
222.818103	221.738092	—	—
54.921450	54.208955	—	—
23.777203	22.423066	—	—
0	1.573322	1.573322	1.573435
0	1.573322	1.573322	1.573349
0	0	0	0.000123
0	0	0	0.000065

IDE, isotropically distributed ensemble; ED, essential dynamics.

^a3 λ values (\AA^2) of the IDE analysis of the static structure of a helical backbone with a fixed methyl group.

^b3 λ values (\AA^2) of the IDE analysis of a rigid backbone with a rotating methyl group.

^c3 λ values (\AA^2) of the ED analysis of a rigid backbone and a rotating methyl group with alignment using all C^α atoms.

^d3 λ values (\AA^2) of the ED analysis of a rigid backbone and a rotating methyl group with alignment using all atoms.

For comparison, the same conformational ensemble was also analyzed by the ED method, using two different superposition (alignment) schemes. In the first case, the three structures were aligned on the static C^α atoms; i.e., no additional rotation was required. Matrix **C** of eq. 2 yields 2 eigenmodes with nonzero eigenvalues describing the rotation of the methyl hydrogen atoms (Table I). The eigenvalues of these modes are identical to the eigenvalues for the rotation of the methyl group of the IDE analysis.

In the second case, the three structures were aligned using a mass-weighted superposition of all 13 atoms (10 C^α and 3 H atoms). The resulting RMSD between the structures is 0.28 \AA . Matrix **C** now yields four eigenmodes with nonzero eigenvalues (Table I). The eigenvalues of the two largest modes are very similar to the eigenvalues of the internal modes of the previous analyses, but they are slightly larger and no longer degenerate. In addition, two much smaller amplitude modes emerge. These modes do not reflect actual internal motion, but rather spurious overall motion of the helix caused by the alignment process itself.

This study was extended to two independently rotating methyl groups attached to the same rigid backbone used above. The IDE analysis then yields two twofold degenerate modes for the methyl rotations and the ED analysis with the alignment on the C^α atoms gives identical results. For the ED analysis with the alignment on all atoms the four largest modes are similar to these modes, but three additional smaller modes appear representing spurious overall rotation of the helix caused by the alignment process. These examples illustrate how the IDE analysis naturally separates between global and internal motions without requiring any a priori knowledge about static and mobile parts. The IDE results are equivalent to the ED analysis, provided that the latter uses the static C^α atoms for alignment.

Backbone Dynamics of Native Ubiquitin

NMR relaxation-related aspects of the 6-ns MD trajectory of native ubiquitin were recently discussed.¹⁸ The

simulation is stable and provides useful insights into the structure and dynamics of the native state of ubiquitin. The radius of gyration remains rather constant throughout the 6-ns trajectory: $R_g = 11.67 \pm 0.08 \text{ \AA}$. The trajectory forms a good example of a globular protein state for which overall and internal motions are expected to be separable in good approximation.

IDE analysis was performed on the 76 backbone C^α atoms using 1,000 snapshots from the final 5 ns of the MD simulation with a time increment of 5 ps. For comparison, reduced ED analyses were performed on the same atoms as described in the Methods section, using three different alignment schemes, with the snapshot at 3.5 ns used as reference (half-time snapshot of the 1–6-ns segment). In the first scheme, the superposition is done using all atoms as (including hydrogens) of both the backbone and the side-chains. In the second scheme, the superposition is done using all C^α atoms. In the third scheme, superposition is done using the C^α atoms belonging to regular secondary structural elements (C^α atoms of residues 1–7, 10–17, 40–45, 48–50, and 64–72 of the β -sheet and residues 23–34 of the α -helix). In the following discussion, mode amplitudes, mode collectivities, mode overlaps, cross-correlation coefficients, and mode projection coefficients are evaluated for the IDE analysis, as well as for the reduced ED analyses and compared with each other.

Figure 1 shows that the mode collectivities κ versus the eigenvalues λ for the four analyses (κ values were calculated using eq. 17). The IDE results differ from the reduced ED results mostly in terms of the largest amplitude modes. IDE exhibits three very large amplitude modes with collectivities around 0.6 indicating that about 60% of all C^α atoms are effectively involved in each of these modes. These modes are separated from all other modes by a mode amplitude gap of two orders of magnitude. These three modes are the structural modes predominantly reflecting rigid-body overall rotational motion already encountered in the previous section. The separability index defined in eq. 16 applied to the IDE eigenvalues is $g = 169.4$. The structural modes are not dominantly manifested in the ED analyses, since overall rotational motion is removed to a significant extent (but not completely) by the alignment process. The eigenvalues and collectivities for the remaining “internal” modes exhibit notably similar characteristics for all four analyses. As noted previously, the collectivities κ in a globular protein are lowest for the smallest and the largest amplitude modes.^{18,23}

The differences in κ and λ values between the IDE and the three ED analyses are comparable in magnitude to the differences between the three ED analyses using different alignment schemes. Some of the largest amplitude modes, $\bar{q}_i = \bar{p}_i$ and \bar{c}_i , are displayed in Figure A1 (see Supplementary Material), by plotting the squared amplitudes $|\bar{q}_{i,k}|^2$ as a function of the residue number of the C^α atoms. By skipping the three structural modes of the IDE analysis (modes 74–76), which collectively affect large stretches of the backbone, the next four largest modes (modes 70–73) were found to exhibit behavior similar to that of the four largest modes (modes 73–76) of the different ED analyses.

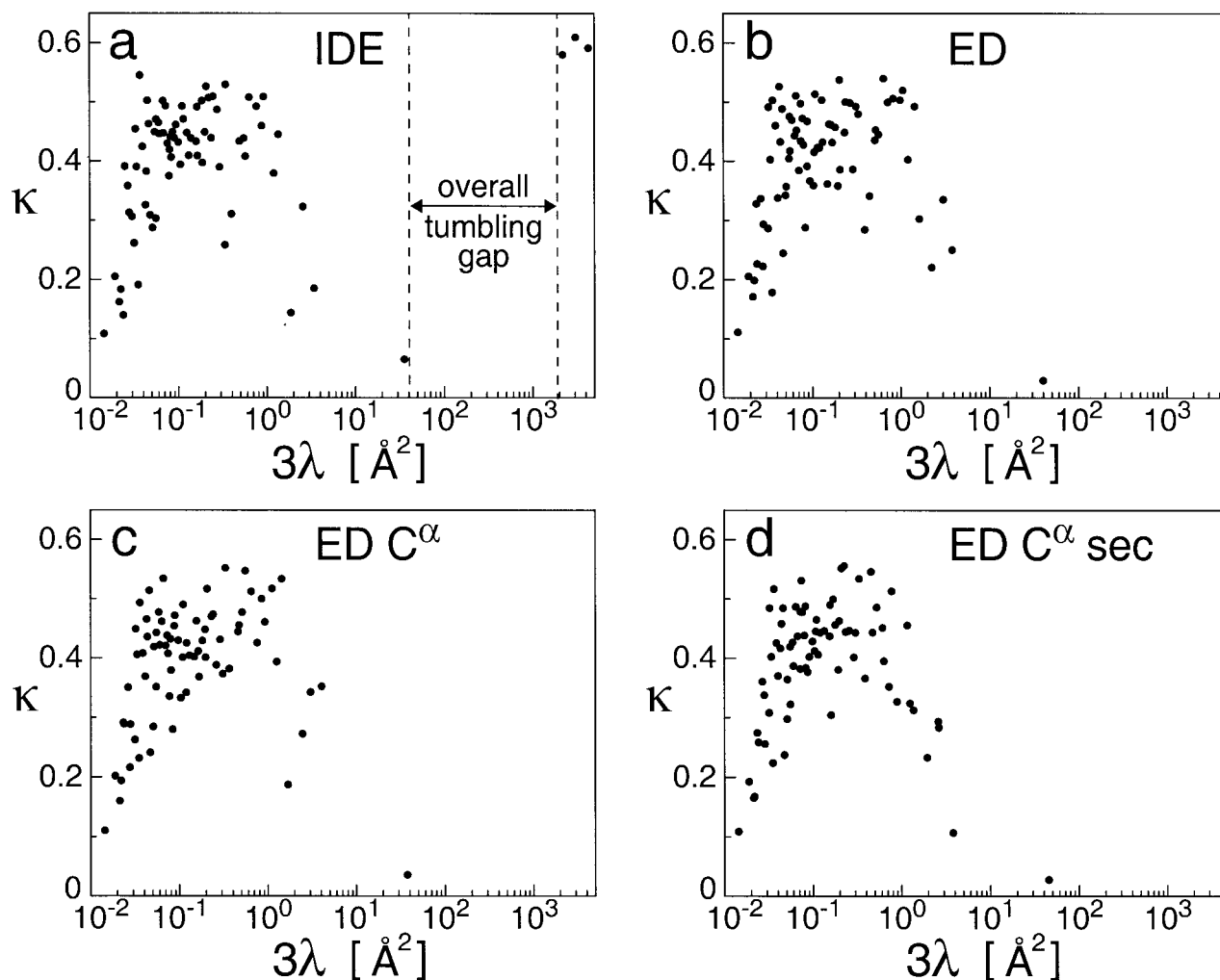


Fig. 1. Mode collectivities κ versus $3 \cdot \lambda$ eigenvalues for the four different analyses of the C^α atoms of the native state. (a) Isotropically distributed ensemble (IDE) analysis, (b) essential dynamics (ED), with alignment on all atoms, (c) ED with alignment on all C^α atoms, and (d) ED with alignment on the secondary structure C^α atoms. κ , which varies between 0 and 1, is a measure for the effective number of C^α atoms that are significantly affected by a certain mode (eq. 17).

The modes of the ED analysis with the alignment based on all atoms come closest to the internal modes of the IDE analysis.

The subspaces spanned by subsets of the eigenvectors were compared with each other by applying eq. 18 to the eigenmodes of the four analyses with the results shown in Figure 2. In case of the IDE analysis the three overall rotational modes were not included (see eq. 18). Figure 2a shows the overlaps between the IDE modes and the modes of the three different ED analyses. The subspaces spanned by six or more of the largest amplitude modes exhibit a very similar overlap pattern, while the subspaces spanned by the first five modes show notable differences. The ED analysis that uses all atoms of ubiquitin for alignment is on average again closest to the IDE analysis (largest overlaps).

For comparison, the mode overlaps between different ED analyses are shown in Figure 2b. The ED analyses that use all atoms and all C^α atoms for alignment, respectively, show a uniformly high overlap. In contrast, the overlap behavior between these two analyses and the ED analysis

with the alignment based on the C^α atoms belonging to regular secondary structural elements resembles for the large amplitude modes the overlap behavior between the IDE and ED analyses of Figure 2a.

It is possible to perform the ED analysis without orientationally aligning each snapshot (but with removal of center of mass motion). In this case, the overall tumbling motion of ubiquitin during the MD simulation will contribute to the covariance matrix. In Figure A2 (see Supplementary Material), the collectivities and mode amplitudes are plotted for parts of the trajectory of different lengths. The gap between overall rotational modes and internal modes is clearly reduced as compared to the IDE analysis. The overall rotational modes, as well as the internal modes, show a stronger dependence on the trajectory length than for the (standard) ED and IDE analyses. For a given set of snapshots, the amplitudes of internal IDE modes are more similar to the internal modes of the ED analysis than to the ones of the ED analysis without orientational alignment.

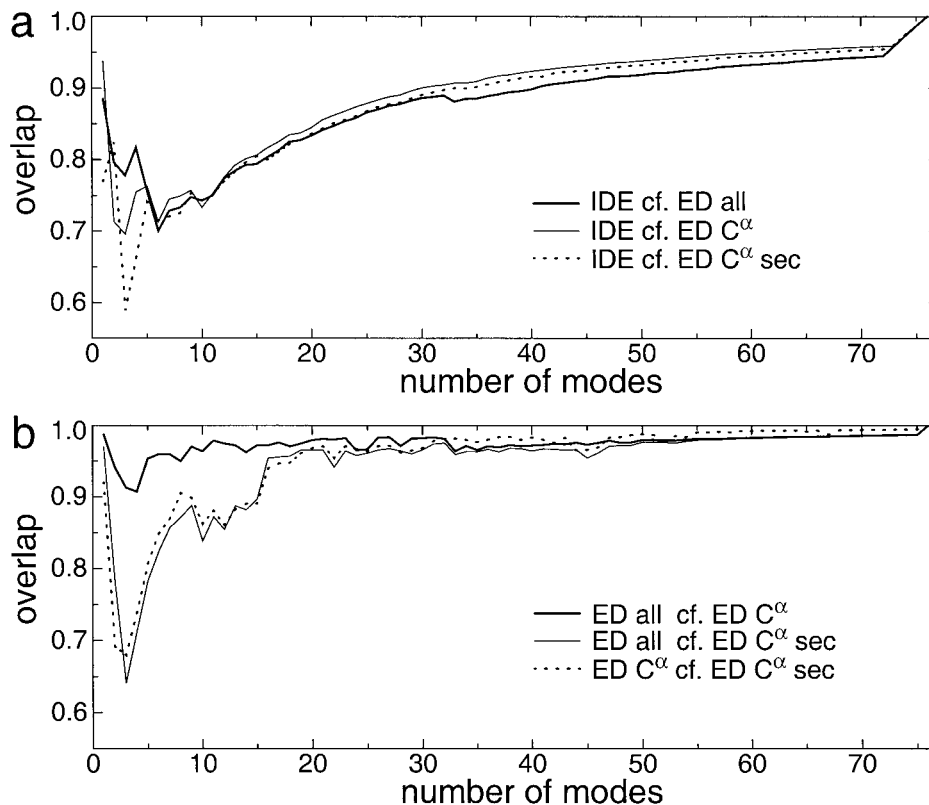


Fig. 2. Mode overlaps as a function of the number of modes using eq. 18. **a:** Overlap between the isotropically distributed ensemble (IDE) modes (neglecting the three overall rotational modes) and the essential dynamics (ED) modes: IDE compared with ED with alignment on all atoms (thick line), IDE compared with ED with alignment on all C $^{\alpha}$ atoms (thin line), and IDE compared with ED with alignment on the secondary structure C $^{\alpha}$ atoms (dotted line). **b:** Overlap between the modes of the ED analyses with different alignment schemes: alignment on all atoms compared with alignment on all C $^{\alpha}$ atoms (thick line), alignment on all atoms compared with alignment on secondary structure C $^{\alpha}$ atoms (thin line), and alignment on all C $^{\alpha}$ atoms compared with alignment on secondary structure C $^{\alpha}$ atoms (dotted line).

To assess the amount and the distance range of correlated motions in proteins, Ichiye and Karplus¹ plotted the cross-correlation coefficient for atom pairs as a function of their average distance. Figure 3a shows the corresponding scatterplots for the IDE analysis. The plots from left to right include all modes, all modes except for the three largest (structural) modes, and all modes except for the five largest modes. For comparison, the corresponding plots are shown in Figure 3b for the different ED analyses. They are similar to those published for other globular proteins^{1–3} and confirm that the magnitudes of cross-correlations at larger distances tend to decrease when the alignment is performed on only the regular secondary structural elements.

The plot for the IDE analysis including all modes (Fig. 3a, left plot) is very distinct from all other plots. At short distances, the cross-correlations are positive, with many close to 1; at long distances, they become negative, with many close to -1 . These features result from the overall rotational modes. They affect adjacent atoms in an in-phase manner, whereas atom pairs separated by longer distances, which tend to lie on opposite sides of the rotation center (center of mass), undergo overall rotation in an anti-phase manner. Since no motion was excluded,

this figure displays a least-biased view of motional correlation effects produced during the simulation of ubiquitin modeled as an isotropic solution.

When the three overall rotational modes are removed from the IDE matrix, the results become similar to those in the ED analyses. However, the distribution of cross-correlations is wider than for the ED analyses. When two additional modes are removed from the IDE matrix (Fig. 3a, far right), the distribution becomes narrower and very similar to the plots in Figure 3b. These two modes mainly affect the C $^{\alpha}$ atoms of the C-terminal residues 74–76, but they also have a small effect on some residues in the core (see Fig. A1, Supplementary Material). For the ED analyses, the largest correlation effects are found when the alignment is done on all C $^{\alpha}$ atoms (Fig. 3b, middle).

The excursion of individual atoms along a given mode during the simulation can be monitored by projecting individual snapshots of the MD simulation on this mode (see eqs. 11 and 12). Figure 4a shows the projection coefficients $a_{i,1}(t)$, $a_{i,2}(t)$, and $a_{i,3}(t)$ of IDE modes $i = 76$, 73, and 38 with collectivities $\kappa_{76} = 0.591$, $\kappa_{73} = 0.065$, and $\kappa_{38} = 0.462$ as a function of simulation time t . For a statistically converged mode, the projection coefficient is expected to fluctuate around zero and to cross the zero line

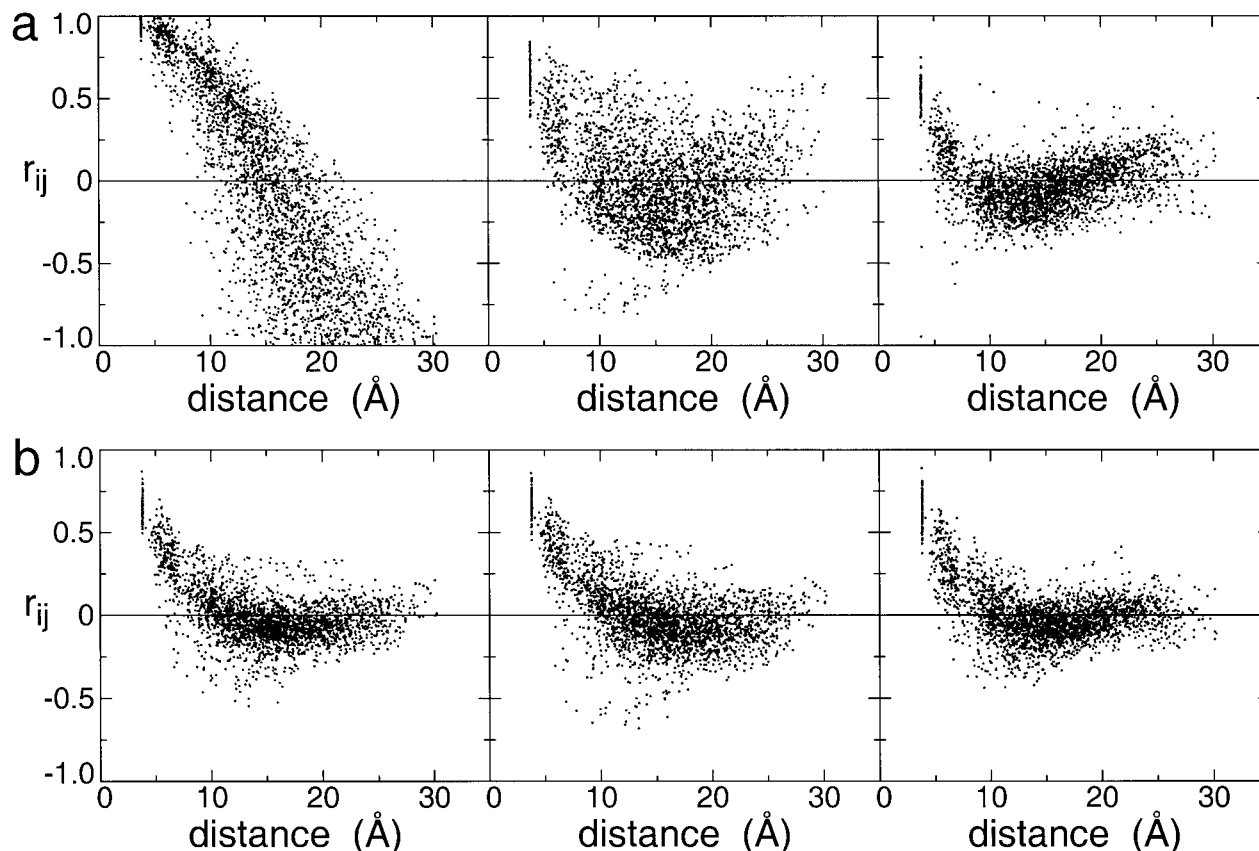


Fig. 3. Cross-correlation coefficient r_{ij} as a function of the average interatomic distance. **a**: From left to right: cross-correlations calculated from the isotropically distributed ensemble (IDE) matrix, including all modes, without the three largest (overall rotational) modes, and without the five largest modes. **b**: From left to right: cross-correlations calculated from the covariance matrices of the essential dynamics (ED) analyses with alignment on all atoms, alignment on all C^α atoms, and alignment on the secondary structure C^α atoms.

multiple times. For mode 76, which is in good approximation an overall rotational mode, the projection coefficients $a_{76,1}$, $a_{76,2}$, $a_{76,3}$ do not cross zero, indicating that this mode is far from converged and therefore the overall rotational diffusion of ubiquitin during the 5-ns simulation is undersampled. For mode 76, the squared amplitude $a_{76}^2(t)$ defined in eq. 13 is almost constant. The same applies for the two other structural modes 74 and 75 (not shown), which reflects the fact that the overall shape of native ubiquitin does only little change over the course of the trajectory. Mode 73, the largest amplitude “internal” mode, is better sampled, but it has still not converged. For this mode, the total excursion squared, $a_{73}^2(t)$, significantly varies in time in a nonoscillatory diffusive way. In contrast, for mode 38 the projection coefficients indicate that the protein fluctuates along this mode on a considerably faster timescale and that this mode is statistically well sampled during the MD simulation.

Backbone Dynamics of a Partially Folded State of Ubiquitin

A structural analysis of the first part of the trajectory of the partially folded state of ubiquitin was described in previous work.¹⁸ It was concluded that during the final 5 ns of this part of the simulation (segment 28–33 ns) the

N-terminal half of the simulated state is similar to the A-state characterized experimentally by circular dichroism (CD)¹⁹ and NMR,^{20,21} in that it contains the N-terminal antiparallel β -sheet and the central helix. In contrast to the experimental studies, the simulation lacked the helical propensity in the C-terminal half as was also found in earlier MD studies.²⁴ The trajectory used in this analysis was extended beyond the 33 ns by an additional 37 ns, reaching a total of 70 ns. During the final 37 ns of the simulation time, the secondary structural elements in the N-terminal half are mostly stable and remain essentially unchanged compared to the previously analyzed 5 ns segment. In the C-terminal half formation of helical structure is not observed over the whole trajectory.

From the final 40 ns of the MD simulation of the partially folded state of ubiquitin, a series of selected snapshots are shown in Figure 5. These snapshots illustrate the expanded range of conformations adopted by ubiquitin during the simulation, which includes some conformers that are rather compact and some that are extended. The highly dynamic character of the protein under these conditions is also reflected in fluctuations of the radius of gyration R_g (eq. 14) displayed in Figure 6 indicating distinct cycles of expansion and collapse. The average R_g is 19 Å, with a standard deviation of ± 3 Å.

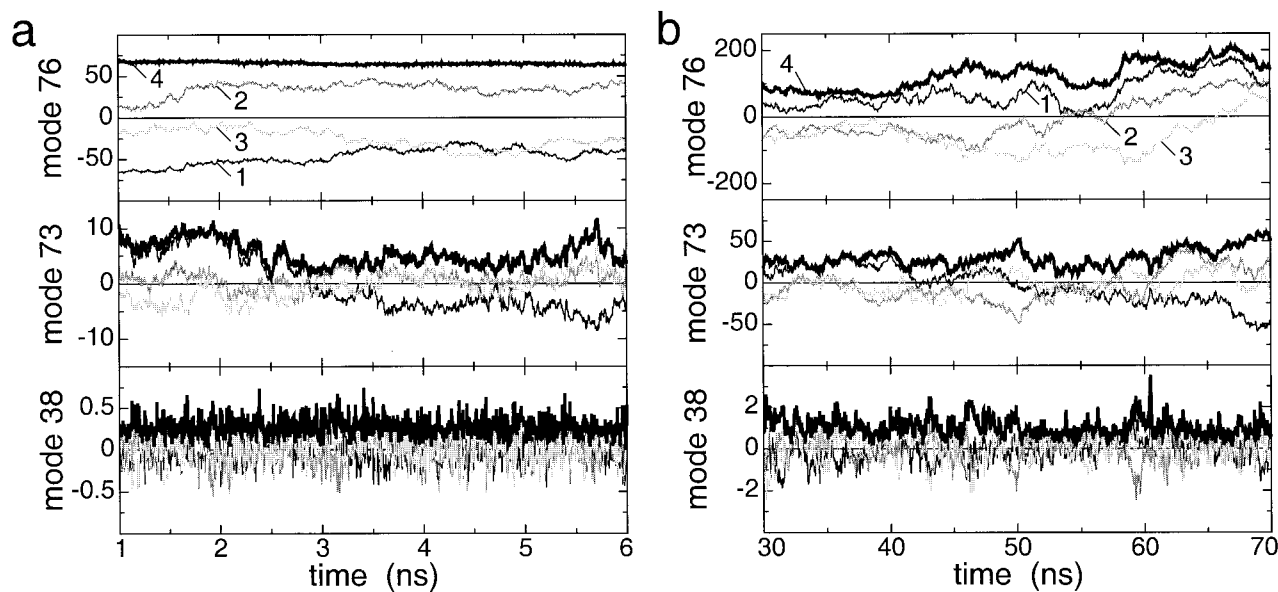


Fig. 4. Projection coefficients $a_{i,1}(t)$ (black), $a_{i,2}(t)$ (gray), and $a_{i,3}(t)$ (light gray) of isotropically distributed ensemble (IDE) modes $i = 76, 73,$ and 38 as a function of the simulation time for (a) the native state and (b) the partially folded state. The thick black line labeled 4 corresponds to $(a_{i,1}(t)^2 + a_{i,2}(t)^2 + a_{i,3}(t)^2)^{1/2}$.

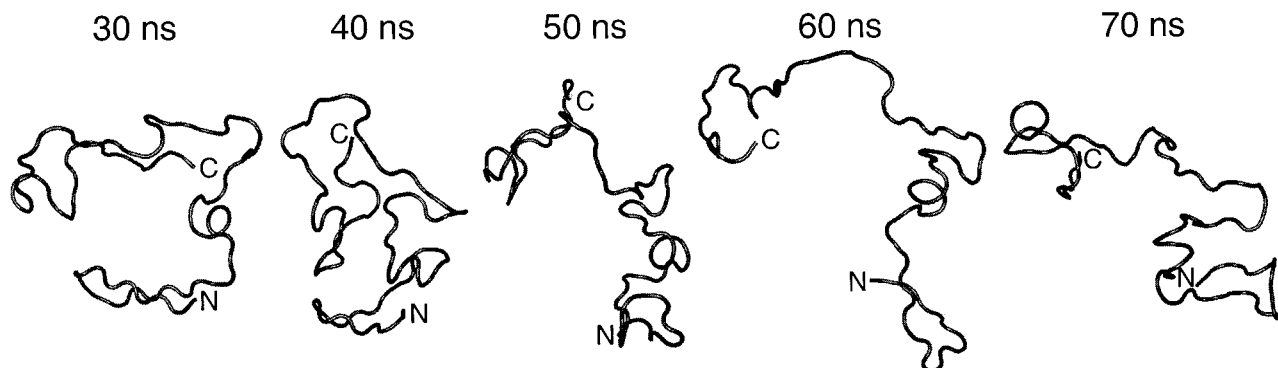


Fig. 5. Backbone conformations of snapshots taken at 30, 40, 50, 60, and 70 ns of the trajectory of the partially folded state of ubiquitin. The structures were superimposed on the C^α atoms, using the structure at 30 ns as a reference structure and are shown in a ribbon representation using the visual molecular dynamics (VMD) software.²⁵

The IDE method was applied to the 76 C^α atoms using 800 snapshots from the final 40 ns of the MD simulation with a time increment of 50 ps. The mode collectivities κ versus the eigenvalues λ are shown in Figure 7a, together with the results for the native state discussed in the previous section. For the partially folded state there is no clear gap between the overall rotational modes and the internal modes, which is in sharp contrast to the native state. The separability calculated according to eq. 16 is $g = 9.5$, which is about a factor 18 smaller than the separability index found in the native state, suggesting that separability of internal and overall motions is not possible in this case. The high-amplitude modes of the partially folded state have a much higher collectivity than do the largest amplitude internal modes of the native state, in agreement with previous observations based on quasi-harmonic analysis of the backbone N and H^N atoms

and reorientational quasi-harmonic analysis of the N— H^N internuclear vectors.¹⁸

For comparison, ED analyses were performed on the C^α atoms of the same snapshots of the partially folded state using three different alignments: superposition using all C^α atoms on the snapshot at 30 ns, 50 ns, and 70 ns, respectively. Resulting mode collectivities κ and eigenvalues λ are shown in Figure 7b, together with the IDE results. The IDE analysis yields two additional modes with larger amplitudes and collectivities (upper right corner in Figure 7b) that are reminiscent of the three structural modes observed for native ubiquitin. However, as shown below, these two modes do not represent pure overall rotational motion. For the small-amplitude modes, the κ versus λ profiles of the IDE analysis and the three ED analyses are very similar; however, for larger amplitude modes ($3\lambda > 10 \text{ \AA}^2$), the patterns begin to differ, particu-

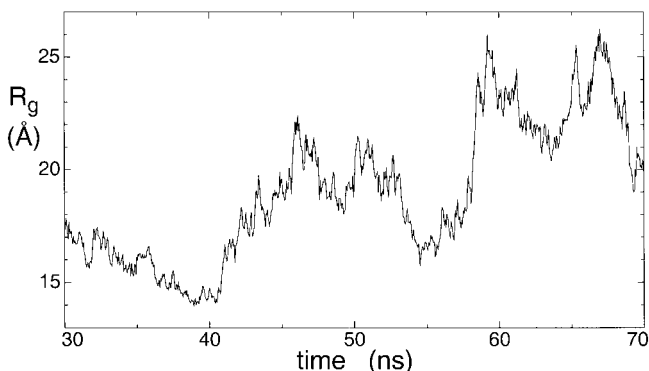


Fig. 6. Radius of gyration of the partially folded state as a function of the simulation time. Data are shown for the last 40 ns of the trajectory (segment 30–70 ns).

larly in the collectivity. The different alignment procedures for the ED analyses lead to differences comparable to the differences between the IDE and ED analyses.

Modes 72–76 behave markedly differently in the κ versus λ relationships for the four analyses. The same applies for the squared amplitudes. In Figure 8, the squared amplitudes $|\bar{q}_{i,k}|^2$ of eigenmodes $i = 76, 75, 74, 73, 72, 66,$ and 61 are plotted as a function of the residue number for the four analyses of the partially folded state. The largest mode of the IDE analysis, mode 76, has a distinctive diffusive behavior with a low amplitude in the middle of the sequence and increasing amplitudes toward the termini (Fig. 8a). This mode has no equivalent in the ED analyses. Modes 61 and 66 are two of the largest modes, which have similar κ and λ values in all four analyses; for these modes, the squared amplitudes are also very similar. Each of the IDE and ED modes plotted in Figure 8 affects a large number of residues (large collectivities). This behavior is very distinct from the native state where the largest internal modes become remarkably localized (see also Supplementary Material).^{18,23}

The projection coefficients $a_{i,1}(t)$, $a_{i,2}(t)$, and $a_{i,3}(t)$ of IDE modes $i = 76, 73,$ and 38 of the partially folded state as a function of the simulation time are shown in Figure 4b. Comparison with the projection coefficients for the native state (Fig. 4a) suggests that the largest mode (mode 76) is almost purely global for the native state, whereas it is significantly mixed for the partially folded state. The sum of the squared projection coefficients of mode 76 for the partially folded state is not constant, which reflects the fact that the shape of the protein undergoes major changes during the simulation. It means that overall and internal motions are not separable for the partially folded state. Thus, analysis protocols that rely on structural alignment are not generally applicable to such mobile systems.

DISCUSSION

Characterization of correlated motions in molecular ensembles with increased flexibility using cartesian covariance matrices is impaired by the nonseparability of internal and overall rotational motion. Hence, the search for the ideal superposition scheme is elusive. For systems

exhibiting sufficiently small fluctuations, such a separation is approximately fulfilled.^{3,26}

As a remedy, covariance analysis can be performed in internal coordinates that do not depend on the overall orientation, such as dihedral angles⁴ and distances.³ Such approaches have other drawbacks, however: local dihedral angle fluctuations do not directly reflect larger scale changes of the shape of the molecule, while eigenvectors of distance covariance matrices of macromolecules are computationally very expensive and their interpretation in cartesian space is less straightforward.

The IDE method introduced in this study uses the scalar products of atomic position vectors with respect to the center of mass. Since the scalar product of two vectors depends only on their lengths and relative orientation, this method bypasses the problem of finding an optimal superposition. It was demonstrated that the scalar products entering the IDE matrix emerge from the cartesian covariance matrix of atomic positions in a structural ensemble exhibiting an isotropic distribution of orientations. Thus, the IDE approach implies an isotropic orientational distribution of each structure as if it were part of an isotropic solution undergoing anisotropic or isotropic overall rotational tumbling. Hence, IDE exhibits a maximal rotation property, contrary to the zero rotation property of normal mode analysis.²⁷

The spherical symmetry created by the isotropic distribution induces a three-fold degeneracy of each eigenmode, which allows for a significant computational speedup. Instead of diagonalization of a $3n \times 3n$ covariance matrix used in ED procedures, IDE requires only diagonalization of a $n \times n$ matrix. The $3n$ -dimensional eigenvectors of the full IDE matrix can be easily reconstructed by direct products with a 3D basis set. For ED, the diagonalization cost scales with $(3n)^3 = 27n^3$, whereas for IDE the diagonalization cost scales only with n^3 . Thus, the isotropic redistribution of snapshots saves a factor of 27 in computational time or, if the computational time is kept constant, the IDE approach allows analysis of a macromolecule with three times more atoms, as compared with ED methods.

An attractive property of the IDE analysis is the fact that it naturally distinguishes between rigid and mobile parts. This was illustrated here using a rotating methyl group attached to a rigid helical protein backbone. The three largest amplitude modes (structural modes) reflect the average structure, while the two lower amplitude modes selectively involve the methyl group, without affecting the backbone atoms. Exactly the same internal modes are found in the ED approach when the structures are superimposed on the C^α atoms of the rigid backbone only. The ED approach requires a priori knowledge or intuition about the rigid and the mobile portions of the molecule, while the IDE approach provides such insight as output.

For the native simulation of ubiquitin, the IDE analysis shows a characteristic gap between the three largest eigenvalues and the rest. This gap is reflected in the large separability index $g = 169.4$, indicating that internal

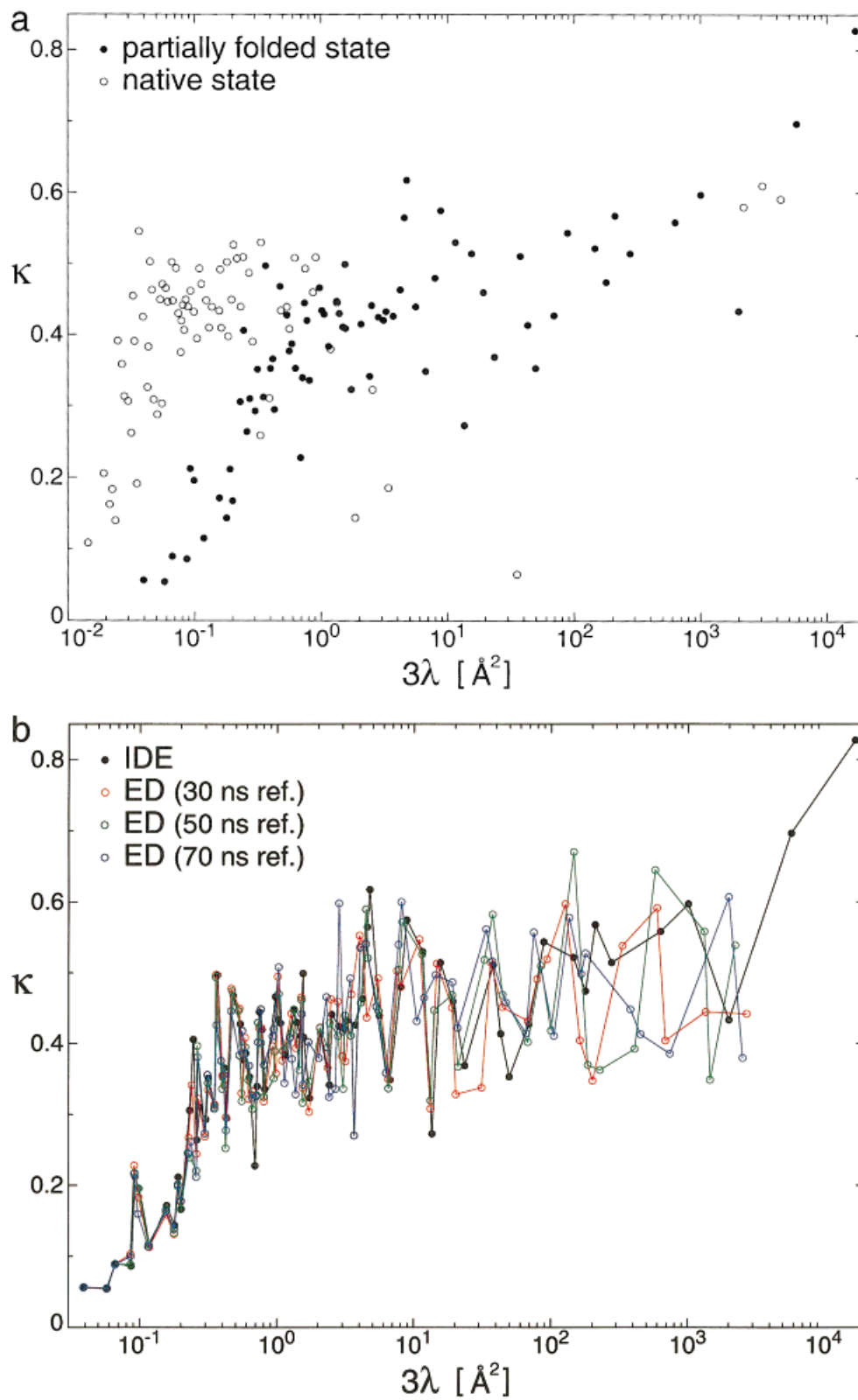


Fig. 7. Mode collectivities κ versus $3 \cdot \lambda$ eigenvalues for (a) isotropically distributed ensemble (IDE) analyses of the partially folded state (filled circles) and the native state (open circles) and (b) the IDE analysis of the partially folded state (black) and the ED analyses of the partially folded state with the 30-ns structure (red), the 50-ns structure (green), and the 70-ns structure (blue) as reference structure for the alignment.

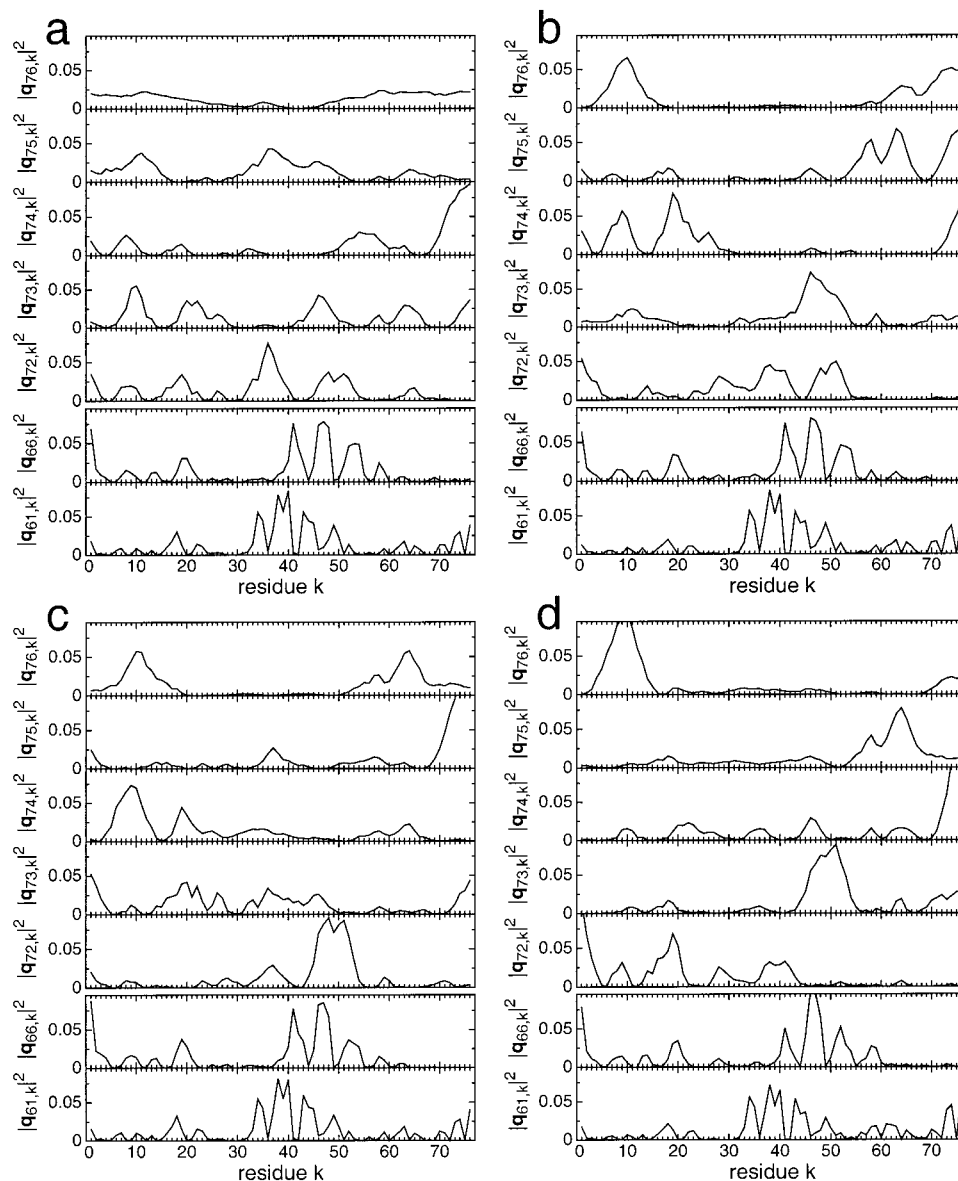


Fig. 8. Squared amplitudes $|\bar{q}_{i,k}|^2$ of eigenmodes $i = 76, 75, 74, 73, 72, 66,$ and 61 as a function of the residue number for (a) isotropically distributed ensemble (IDE) analysis of the partially folded state, and (b–d) essential dynamics (ED) analyses of the partially folded state with (b) the 30-ns structure, (c) the 50-ns structure, and (d) the 70-ns structure as reference structure for the alignment.

motions do not much alter the shape of the protein, or, equivalently, that overall and internal motions are well separable. The first plot in Figure 3a obtained from the IDE analysis applied to the 5-ns MD simulation of native ubiquitin displays the total correlations of motions between C^α atom pairs as a function of their average distances. Obviously, the strong anticorrelations of atoms at larger distances reflects the presence of rotational modes that affect most of the protein. These modes are, in fact, the three highest-amplitude modes exhibiting high collectivities, which can be seen by removal of these modes from the IDE covariance matrix, yielding correlations close to the ED results.

Partially folded and unfolded proteins exhibit characteristic correlated dynamics that are qualitatively different from those observed for globular protein states.¹⁸ Analysis of the eigenmodes of the 40-ns simulation of a partially folded state of ubiquitin demonstrates that the superposition problem becomes severely ill-posed, making methods desirable that do not suffer from this limitation. The IDE method applied to this state yields an eigenmode amplitude distribution with a separability index of 9.5, which is 18 times smaller than the one for the native simulation. The small separability index also reflects that no rigid 3D structure exists that is compatible with covariance matrix \mathbf{P} . For the partially folded state, IDE and ED methods

provide similar results for the small amplitude modes and their collectivities (Fig. 7b). These modes do not mix with modes that significantly alter the shape of the molecule and they are insensitive to the superposition scheme. By contrast, for the largest 20 modes, IDE and ED significantly differ as do the ED analyses using different superposition methods. For the partially folded state, separability of overall rotational motion and internal motion is thus a poor approximation, making covariance analysis based on superpositions of structures inapplicable. Since IDE does not rely on the alignment step, it provides a least biased way of characterizing such larger amplitude motions. The IDE method is applicable to ensembles of other macromolecular systems including RNA and synthetic polymers in solution.

ACKNOWLEDGMENTS

J.J.P. is a recipient of a Human Frontier Science Program postdoctoral fellowship.

SUPPLEMENTARY MATERIAL

Figure A1 shows the squared amplitudes $|\bar{q}_{i,k}|^2$ as a function of the residue number for selected modes of the IDE and ED analyses of the native state of ubiquitin, and Figure A2 shows IDE results and ED results, both with and without orientational alignment of each snapshot of the native state of ubiquitin, using variable trajectory lengths.

REFERENCES

1. Ichiye T, Karplus M. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins* 1991;11:205–217.
2. Hünenberger PH, Mark AE, van Gunsteren WF. Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations. *J Mol Biol* 1995;252:492–503.
3. Abseher R, Nilges M. Are there non-trivial dynamic cross-correlations in proteins? *J Mol Biol* 1998;279:911–920.
4. Karplus M, Kushick JN. Method for estimating the configurational entropy of macromolecules. *Macromolecules* 1981;14:325–332.
5. Hayward S, Kitao A, Hirata F, Gō N. Effect of solvent on collective motions in globular protein. *J Mol Biol* 1993;234:1207–1217.
6. Amadei A, Linssen ABM, Berendsen HJC. Essential dynamics of proteins. *Proteins* 1993;17:412–425.
7. Berendsen HJC, Hayward S. Collective protein dynamics in relation to function. *Curr Opin Struct Biol* 2000;10:165–169.
8. Wright PE, Dyson HJ. Intrinsically unstructured proteins: reassessing the protein structure-function paradigm. *J Mol Biol* 1999;293:321–331.
9. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang C, Kissinger CR, Bailey RW, Griswold MD, Chiu W, Garner EC, Obradovic Z. Intrinsically disordered protein. *J Molec Graphics* 2001;19:26–59.
10. Blumenthal L. Theory and applications of distance geometry. Cambridge: Cambridge University Press; 1953.
11. Crippen GM, Havel TF. Distance geometry and molecular conformation. Taunton: Research Studies Press; 1988.
12. Brüschweiler R. Collective protein dynamics and nuclear spin relaxation. *J Chem Phys* 1995;102:3396–3403.
13. Prompers JJ, Lienin SF, Brüschweiler R. Collective reorientational motion and nuclear spin relaxation in proteins. In: Altman RB, Dunker AK, Hunter L, Lauderdale K, Klein TE, editors. *Biocomputing: proceedings of the 2001 Pacific symposium*. Singapore: World Scientific; 2001. p 79–88.
14. Brooks RB, Bruccoleri RE, Olafson BD, States D, Swaminathan S, Karplus M. CHARMM—a program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 1983;4:187–217.
15. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 1998;102:3586–3616.
16. Vijay-Kumar S, Bugg CE, Cook WJ. Structure of ubiquitin refined at 1.8 Å resolution. *J Mol Biol* 1987;194:531–544.
17. Lienin SF, Bremi T, Brutscher B, Brüschweiler R, Ernst RR. Anisotropic intramolecular backbone dynamics of ubiquitin characterized by NMR relaxation and MD computer simulation. *J Am Chem Soc* 1998;120:9870–9879.
18. Prompers JJ, Scheurer C, Brüschweiler R. Characterization of NMR-relaxation active motions of a partially folded A-state analogue of ubiquitin. *J Mol Biol* 2001;305:1085–1097.
19. Wilkinson KW, Mayer AN. Alcohol-induced conformational changes of ubiquitin. *Arch Biochem Biophys* 1986;250:390–399.
20. Stockman BJ, Euvrard A, Scahill TA. Heteronuclear three-dimensional NMR spectroscopy of a partially denatured protein: the A-state of human ubiquitin. *J Biomol NMR* 1993;3:285–296.
21. Brutscher B, Brüschweiler R, Ernst RR. Backbone dynamics and structural characterization of the partially folded A state of ubiquitin by ^1H , ^{13}C , and ^{15}N nuclear magnetic resonance spectroscopy. *Biochemistry* 1997;36:13043–13053.
22. Hinsen K. The molecular modeling toolkit: a new approach to molecular simulations. *J Comp Chem* 2000;21:79–85.
23. Lienin SF, Brüschweiler R. Characterization of collective and anisotropic reorientational protein dynamics. *Phys Rev Lett* 2000;84:5439–5442.
24. Alonso DOV, Daggett V. Molecular dynamics simulations of protein unfolding and limited refolding: characterization of partially unfolded states of ubiquitin in 60% methanol and in water. *J Mol Biol* 1995;247:501–520.
25. Humphrey W, Dalke A, Schulten K. VMD—visual molecular dynamics. *J Mol Graph* 1996;14:33–38.
26. Karplus M, Ichiye T. Comment on “Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations.” *J Mol Biol* 1996;263:120–122.
27. Wilson E Jr, Decius JC, Cross PC. Molecular vibrations. The theory of infrared and Raman vibrational spectroscopy. New York: Dover; 1955.