PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Large-scale comparison of protein essential dynamics from molecular dynamics simulations and coarse-grained normal mode analyses

Aqeel Ahmed,[1] Saskia Villinger,[1] and Holger Gohlke[1,2]*

[1] Department of Biological Sciences, Molecular Bioinformatics Group, Goethe-University, Frankfurt, Germany
[2] Department of Mathematics and Natural Sciences, Heinrich-Heine-University, Düsseldorf, Germany

ABSTRACT

A large-scale comparison of essential dynamics (ED) modes from molecular dynamic simulations and normal modes from coarse-grained normal mode methods (CGNM) was performed on a dataset of 335 proteins. As CGNM methods, the elastic network model (ENM) and the rigid cluster normal mode analysis (RCNMA) were used. Low-frequency normal modes from ENM correlate very well with ED modes in terms of directions of motions and relative amplitudes of motions. Notably, a similar performance was found if normal modes from RCNMA were used, despite a higher level of coarse graining. On average, the space spanned by the first quarter of ENM modes describes 84% of the space spanned by the five ED modes. Furthermore, no prominent differences for ED and CGNM modes among different protein structure classes (CATH classification) were found. This demonstrates the general potential of CGNM approaches for describing intrinsic motions of proteins with little computational cost. For selected cases, CGNM modes were found to be more robust among proteins that have the same topology or are of the same homologous superfamily than ED modes. In view of recent evidence regarding evolutionary conservation of vibrational dynamics, this suggests that ED modes, in some cases, might not be representative of the underlying dynamics that are characteristic of a whole family, probably due to insufficient sampling of some of the family members by MD.

Proteins 2010; 78:3341–3352.

Key words: RCNMA; ENM; intrinsic motion; conformational change; evolution.

INTRODUCTION

Macromolecules are dynamic, and their motions are critical for their functions.[1] For example, conformational changes are required for the functioning of transport and motor proteins,[2–5] catalytic processes of enzymes,[6,7] and molecular mechanisms of protein regulation.[8–10] Conformational changes of proteins have also been observed upon ligand binding, as seen in the cases of HIV-1 protease,[11] aldose reductase,[12] adenylate kinase,[13–15] tyrosine phosphatase,[16,17] and calmodulin.[18,19] These conformational changes range from side chain fluctuations to reorientations of domains and partial unfolding and refolding.[20,21]

Different computational approaches targeting the modeling of protein dynamics are available. Molecular dynamics (MD)[22–24] simulation is the most widely applied and accurate computational technique currently used for this. However, despite an immense increase in computational power and significant methodological improvements, MD simulations are still computationally expensive for exploring conformational space due to slow barrier crossing on rugged energy landscapes of macromolecules.[25,26]

As a result, there have been efforts to develop alternative approaches that are computationally more efficient for describing conformational changes of a protein structure. Among those, coarse-grained normal mode (CGNM) approaches, such as the elastic network model (ENM)[27–30] and the rigid cluster normal mode analysis (RCNMA),[31] are promising in that they provide the directions of intrinsic mobility of biomolecules in terms of harmonic modes.[29,31] The modes can be viewed as possible deformations of proteins and can be sorted by their energetic cost of deformation. More importantly, in agreement with the conformational selection model,[32–35] the conformational changes on ligand binding of many proteins have been found to occur usually along a few low-energy modes of unbound proteins calculated with CGNM approaches.[29–31,36,37]

Additional Supporting Information may be found in the online version of this article.

Aqeel Ahmed's current address is Department of Biochemistry & Molecular Biophysics, The University of Arizona, Tucson, Arizona
Saskia Villinger's current address is Department of NMR-based Structural Biology, Max-Planck-Institute for Biophysical Chemistry, Göttingen, Germany
*Correspondence to: Holger Gohlke, Universitätsstr. 1, 40225 Düsseldorf, Germany.
E-mail: gohlke@uni-duesseldorf.de.
Received 26 May 2010; Revised 13 July 2010; Accepted 25 July 2010
Published online 2 August 2010 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.22841

Furthermore, the calculation of modes only takes seconds for normal-sized proteins and, therefore, can be applied also to large macromolecules. Realizing the potential of these CGNM approaches, different approaches have utilized the directional information provided by them for steering MD simulations,[38–40] incorporating receptor flexibility in docking approaches,[41–43] flexible fitting of molecular structures,[44–47] and efficient generation of pathways of conformational changes.[48–50]

Several applications have demonstrated the successful use of CGNM approaches in describing protein conformational changes.[31,51,52] Still, a recent study[37] has shown that the success of ENM in describing experimental conformational changes from unbound to bound protein conformations strongly depends on the collectivity of the conformational changes. Hence, a large-scale comparison between the essential dynamics (ED) of proteins derived from MD simulations[53,54] and the dynamics computed by CGNM approaches appears in order to investigate the validity and applicability of CGNM approaches.

Different studies have previously[53,55–58] shown striking similarities between normal modes derived from all-atom force-field potentials[59,60] and ED modes from MD simulations. These early studies have used one or a few proteins with limited simulation lengths and have focused on the comparisons of frequency spectra.[56,59,61] For example, Hayward *et al.*[56] have used a 200 ps MD trajectory of bovine pancreatic trypsin inhibitor for such a study. For ENM, however, it has been argued[52] that information on the directionality of biologically relevant conformational changes provided by the eigenvectors has wider applications than information on the magnitude of motions provided by the eigenvalues. This is also reflected by the applications and developments of ENM-based approaches in recent years.[45,46,48,62,63]

Along these lines, this study aims to compare ED modes of proteins observed in MD simulations with normal modes obtained from CGNM methods (ENM and RCNMA) for a large dataset of 335 diverse proteins. To the best of our knowledge, a similar study[64] has been reported recently; however, it was limited to a dataset of only 30 proteins. As for MD simulations, the first five ED modes for each protein were obtained from the Molecular Dynamics Extended Library database (MoDEL).[64,65] There, the modes were extracted from MD trajectories of 10 ns in length. Normal modes were calculated using in-house implementations of the ENM and RCNMA models[31] as detailed below. The three sets of modes were compared in terms of overlap of directions, correlation of relative magnitudes of motions, and spanning coefficients. The CATH[66] classification of protein structures was used to investigate the influence of protein structure similarity/dissimilarity on mode similarity/dissimilarity. Finally, for a smaller protein subset, ED and CGNM modes were also compared against experimentally observed conformational changes.

MATERIALS AND METHODS

Essential dynamics modes and protein data set

ED modes were obtained from the MoDEL database (http://mmb.pcb.ub.es/MODEL, version as of May 2006).[64,65] This version of the MoDEL database stores information derived from MD simulations for more than 400 proteins. The MD simulations were performed with the Amber8 suite of programs at 300 K in the NPT ensemble, and the parm99 force field was used together with TIP3P as a water model. The length of each MD trajectory was 10 ns.

The first five available ED modes of 418 proteins were downloaded from the MoDEL database. ED modes were calculated using all atoms of snapshots extracted every picosecond from trajectories of 10 ns lengths. For reasons of comparison with CGNM results, only C_α atom directions were used in this study. Corresponding experimental structures were obtained from the RCSB Protein Data Bank.[67] For the sake of compatibility, heavy atoms in the ED modes files were compared with heavy atoms in the PDB files using the PDBParser module of Biopython.[68] Where possible, inconsistencies between the two sets were corrected manually. However, 83 of 418 cases were removed from the dataset due to deviating numbers of atoms/residues, empty or corrupt ED modes files, C_α-only structures, bad structural quality or inconsistency with the standard amino acid library, or problems in processing by FIRST.[69] This resulted in a final dataset of 335 protein structures. The PDB structures were then protonated using "tleap" from the Amber suite. Disulfide-bridges involving cysteine residues and protonation states of histidines were adopted from the ED mode files. All structures were then aligned with their respective MD average (reference) structure using C_α atoms. Before this, the average MD structures were minimized in the gas phase to circumvent stereochemical inaccuracies obtained by the averaging process. For the minimization, the conjugate-gradient method with a distance-dependent dielectric of 4r was used until the root-mean square of the elements of the gradient vector is $<10^{-4}$ kcal/mol/Å.

The dataset of 335 protein structures is diverse with respect to protein size, function, origin, subcellular localization, and structure determination method. The proteins contain, on average, 121 residues, with a minimum of 20 and a maximum of 349 residues. The size distribution of the dataset is shown in Figure 1. The distribution is positively skewed with a peak in the range of 60 to 80 residues.

Elastic network model

The ENM model has been successfully applied in CGNM calculations.[51,52,70] In the ENM model, the

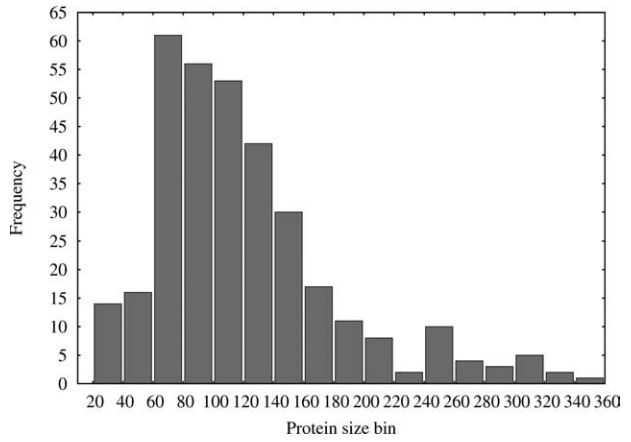

Figure 1

Frequency distribution of the protein size, in terms of the number of residues, for the dataset of 335 proteins.

proteins are described as 3D elastic networks based on a simplified representation of the potential energy.[71–73] Usually, each C_α atom of an amino acid acts as a junction in the network. Interactions between these particles are modeled by Hookean springs based on a harmonic pairwise potential,[71] resulting in a total potential energy of the system given by

$$V = \frac{\gamma}{2}\sum_i \sum_j \theta(r_c - r_{ij}^0)(r_{ij} - r_{ij}^0)^2 \qquad (1)$$

r_{ij} and r_{ij}^0 are the instantaneous and equilibrium distances between atoms i and j. $\theta(x)$ is the Heaviside step function; it is 1 if $x > 0$ and 0 otherwise; γ is a phenomenological force constant assumed to be the same for all pairwise interactions; it is set to 1 kcal/mol/Å^2. $r_c = 10$ Å is the cutoff up to which interactions between the C_α atoms are taken into account. These values of γ and r_c are identical to the ones used in a previous study by us.[31]

According to the ENM,[29] the elements of a $3N \times 3N$ Hessian matrix H (where N is the number of C_α atoms) are obtained from the second derivatives of V with respect to the Cartesian coordinates of atoms i and j. H is diagonalized to get the normal modes.

Rigid cluster normal mode analysis

The RCNMA approach[31] adds another level of coarse-graining to ENM by first identifying rigid clusters and flexible regions within protein structures. In a subsequent step, dynamical properties of the molecule are revealed by the rotations-translations of blocks approach[74] using an ENM representation of the coarse-grained protein. In this step, only rigid body motions are allowed for rigid clusters while links between them are treated as fully flexible. This considerably reduces the dimensionality of H. In this study, RCNMA is performed as previously described.[31]

Flexible and rigid regions of proteins are identified by FIRST,[69] which identifies and counts the bond-rotational degrees of freedom in a molecular framework of atoms connected by covalent and noncovalent constraints (hydrogen bonds, salt bridges, and hydrophobic interactions) based on rigidity theory.[69,75,76] Parameters used for FIRST analysis, that is, a hydrogen bond energy cutoff $E_{cut} = -1.0$ kcal/mol and a distance cutoff for hydrophobic interaction $d_{HC} = 0.25$ Å, were consistent with our previous study.[31] No profound change in the results was observed by changing these parameters (data not shown).

The all-atom representation of proteins needed for the FIRST analysis is reduced to a C_α-only representation in the next step of RCNMA. Each rigid cluster forms a block in the subsequent rotations and translations of block[74,77] approach, and flexible regions are modeled on a one-residue-per-block basis (in which case only translational motion of the "block" is considered). Interactions between these particles are modeled as in ENM [Eq. (1)], and the same parameters are used, that is, $r_c = 10$ Å and $\gamma = 1$ kcal/mol/Å^2. The $3N \times 3N$ matrix H is, therefore, reduced to a $6n \times 6n$ dimensional matrix H_{sub} by projecting H into the subspace spanned by translation/rotation basis vectors of n blocks according to:

$$H_{sub} = P^t H P \qquad (2)$$

with P being an orthogonal $3N \times 6n$ projection matrix of the infinitesimal translation/rotation eigenvectors of each block. This leads to a reduction of the memory requirement proportional to $(N/n)^2$ and the computational time proportional to $(N/n)^3$. Diagonalization of the resulting matrix H_{sub} yields the normal modes U_{sub} and eigenvalues Λ:

$$H_{sub} U_{sub} = U_{sub} \Lambda \qquad (3)$$

Atomic displacements can be obtained by expanding back the normal modes U_{sub} from the subspace spanned by the translation/rotation eigenvectors of the blocks to the Cartesian space ($U = PU_{\mathrm{sub}}$).

ED and CGNM comparison

The directions and relative magnitudes of motions described by the first five ED modes were compared with results from CGNM analyses. As done previously,[30,31] the overlap of mode directions and the correlation of magnitudes of motions [see (*i*) and (*ii*)] between two sets of modes were calculated for each structure in the protein dataset. Distributions of maximal overlap, maximal correlation, and the mode number involved in maximal overlap between the two sets of modes were analyzed for the dataset. It was further analyzed how well the subspace spanned by each ED mode is described by the 10% and 25% lowest frequency CGNM modes by calculating the "spanning coefficient" [see (*iii*)]. To analyze the level

of coarse-graining for RCNMA based on the rigid cluster decomposition from FIRST, the dimensionality reduction of H [see (*iv*)] was calculated.

(i) The overlap I_{in}[78] of the ith CGNM mode $\vec{u}_i$ with the nth ED mode $\vec{v}_n$ ($n = 1, 2, \ldots 5$) was calculated according to:

$$I_{in} = \frac{|\vec{u}_i \cdot \vec{v}_n|}{(\vec{u}_i \cdot \vec{u}_i)^{1/2} \cdot (\vec{v}_n \cdot \vec{v}_n)^{1/2}} \quad (4)$$

An overlap of 1 indicates that the directions of the collective atom displacements along the ED mode and the CGNM mode are identical. From a comparison of all pairs of modes ($\vec{u}_i$, $\vec{v}_n$), the CGNM mode with maximal overlap was selected for further analysis.

(ii) Similarly, a correlation coefficient C_{in}[30] of the ith CGNM mode $\vec{u}_i$ with the nth ED mode $\vec{v}_n$ ($n = 1, 2, \ldots 5$) was calculated according to:

$$C_{in} = \frac{\vec{A}_i \cdot \vec{B}_n}{(\vec{A}_i \cdot \vec{A}_i)^{1/2} (\vec{B}_n \cdot \vec{B}_n)^{1/2}} \quad (5)$$

where $\vec{A}_i$ and $\vec{B}_i$ are the vectors of mean centered amplitudes of atomic displacements as determined from vectors $\vec{u}_i$ and $\vec{v}_n$. A correlation coefficient of 1 indicates that the relative magnitudes of atomic displacements along the ED mode and the CGNM mode are identical. From a comparison of all pairs of modes ($\vec{u}_i$,$\vec{v}_n$), the CGNM mode with maximal correlation was selected for further analysis.

(iii) The "spanning coefficient" S_n^k[79] was computed as the sum of the square of the expansion coefficients:

$$S_n^k = \sum_i^k (\vec{u}_i \cdot \vec{v}_n)^2 \quad (6)$$

Here, the sum over the first k nonzero CGNM modes was computed to determine the percentage of CGNM modes needed for describing each of the first five ED modes. A spanning coefficient of 1 indicates that the subspace spanned by the nth ED mode can be completely described by the subspace considered by the k CGNM modes. The "mean spanning coefficient" is the average of spanning coefficients of the first five ED modes of all proteins in the dataset.

(iv) The dimensionality reduction D was calculated based on the reduction of the H matrix dimension due to considering rigid blocks in RCNMA:

$$D = 1 - \left(\frac{6n + 3m - 6}{3N - 6}\right) \quad (7)$$

where n is the number of blocks of size >2 and m is the number of blocks of size 1 (note that for simplicity blocks of size of two are not considered *per se* in the H_{sub} matrix and are decomposed instead into two blocks, each of size one). A dimensionality reduction of 1 indicates that all C_α atoms are in one rigid block, whereas 0 indicates that every block is of size 1. In that case, RCNMA becomes equal to ENM.

Similarity of motions in protein classes and folds: ED versus CGNM modes

To analyze the similarity of the dynamics within different protein classes or folds, the dataset of proteins was classified according to CATH. Of 335 proteins in our dataset, 320 proteins were found in the CATH database.[66] Overlap and correlation results were sorted for these proteins according to different protein classes or folds, and mean values and standard deviations were calculated accordingly.

In addition, to analyze the locality or collectivity of motion within different classes, the collectivity index [Eq. (8)] was used, which describes the number of atoms that are affected by a mode or conformational change. The collectivity index proposed by Bruschweiler[80] is calculated according to:

$$\kappa = \frac{1}{N} \exp\left(-\sum_{i=1}^{N} \Delta\vec{r}_i^2 \log \Delta\vec{r}_i^2\right) \quad (8)$$

where N is the number of atoms, and $\Delta\vec{r}_i$ is the relative displacement of the mode or the difference in Cartesian coordinates of atom i if an experimentally determined conformational change of the protein is considered. All values of $\Delta\vec{r}_i$ have been scaled consistently such that $\sum_{i=1}^{N} \Delta\vec{r}_i^2 = 1$. A value of $\kappa = 1$ indicates a mode or conformational change of maximal collectivity, that is, all $\Delta\vec{r}_i$ are identical. Conversely, if only one atom is affected by the mode or conformational change, κ reaches the minimal value of $1/N$.

Comparison of ED and CGNM modes with experimentally observed conformational changes

To compare ED and CGNM modes with experimentally observed conformational changes, MolmovDB[81] was searched for alternate conformations to the structures in the dataset. Only those pairs of alternate conformations that have at least 99% sequence similarity and have an RMSD > 1 Å were selected. Pairs of alternate conformations were visually inspected and rejected if the conformational change only occurred in the *N*- or *C*-terminal region. This leads to a subset of 13 pairs of proteins, which are listed in Table IV.

The maximal overlap [Eq. (4)] and maximal correlation [Eq. (5)] between the experimentally observed conformational change, $\Delta\vec{r} = \vec{r}_0 - \vec{r}_c$, and ED and CGNM modes, respectively, were computed for the dataset. The $\vec{r}_o$ and $\vec{r}_c$ are vectors of C_α atom coordinates of a structure in the dataset and its alternate conformation, respectively.

Table I
Comparison of ED Modes With ENM and RCNMA Results

	Maximal overlap[a]			Maximal correlation[a]			Mean spanning coefficient[b]	
Methods	Mean[c]	<0.4[d] (%)	>0.5[d] (%)	Mean[c]	<0.4[d] (%)	>0.5[d] (%)	10%	25%
ENM	0.65 (0.31–0.93)	3	83	0.73 (0.22–0.98)	3	94	0.68 (30)	0.84 (85)
RCNMA	0.64 (0.34–0.95)	3	80	0.74 (0.26–0.98)	2	94	0.59 (20)	0.78 (58)

[a]Maximal overlap [Eq. (4)] or maximal correlation [Eq. (5)] between ENM or RCNMA modes and ED modes.
[b]Average of spanning coefficients [Eq. (6)] of the first five ED modes of all proteins in the dataset using 10% and 25% of all available ENM modes. The average number of modes used in each case is given in parentheses.
[c]Average over all 335 proteins in the dataset with lowest and highest values given in parentheses.
[d]Percentage of maximal overlap or maximal correlation values <0.4 and >0.5.

RESULTS AND DISCUSSION

Influence of the reference structure: average versus open

ED modes[53] are based on a reference structure, that is, the structure obtained by averaging all conformations along the trajectory, which can be conformationally different from the experimental starting structure (termed "open" structure here) and can have stereochemical inaccuracies. In contrast, CGNM analyses are usually based on experimental structures.[71] To analyze the influence of using different reference structures, average or open, in CGNM analysis, ED modes were compared with CGNM modes computed from either structure. Expectedly, ED modes correlate better with CGNM modes in both, directions and amplitudes of motions, if the average structure is used for CGNM. For example, for ENM the mean maximal overlap [Eq. (4)] and mean maximal correlation [Eq. (5)] values are 0.65 and 0.73, respectively, using the average structure. These values decrease by 0.10 and 0.08 if the open structure is used instead. The lower values are mainly due to those proteins that show large conformational differences between the open and the average structures. For example, the 49 of 335 protein structures for which the maximal overlap decreases by at least 0.2 have a mean RMSD between open and average structure of 3.11 Å. For comparison, the mean RMSD over all proteins is 2.06 Å.

As a further test, average structures were minimized (see Materials and Methods) to remove stereochemical inaccuracies obtained by the averaging process. The mean maximal overlap and correlation values between ED and CGNM modes were found to be almost unaffected compared to the use of nonminimized average structures, which can be explained[51,82] by the coarse-grained nature of CGNM. Given that in general very similar results are obtained for CGNM from both the open and average structures, and for the sake of a fair comparison with ED modes, the average structure will be used henceforth as a reference in this study. The average structure has also been used previously in ENM for the sake of comparison.[64]

As for the decomposition of the structure into rigid clusters and flexible regions by FIRST, however, the MD average structures generally result in more flexible decompositions than the open ones. This can be explained by the fact that FIRST requires input at an atomic level, which makes FIRST more sensitive to the accuracy of the input structure.[69] It will be discussed in more detail below to what extent RCNMA results are influenced by this.

Influence of the level of coarse-graining: ENM versus RCNMA

FIRST[69] decomposes a protein structure into rigid clusters and flexible regions based on rigidity theory.[69,75,76,83] RCNMA utilizes this information and considers each rigid cluster as a single node with six degrees of freedom in an elastic network representation of the protein. This not only reduces the dimensionality of the problem, and hence, the memory requirements and computational times, but also simplifies and emphasizes important movements of mobile regions.[31] When applying RCNMA, caution is required as an overly rigid representation of a protein might lead to an underestimation of motion. Average structures obtained from MD trajectories, which are used here as reference structures, generally result in more flexible decompositions than the respective open structures: on average, the largest rigid cluster comprises 16% of the residues of the average structure, whereas it comprises 25% of the residues of the open structure. As a more general measure for the level of coarse-graining, the dimensionality reduction [Eq. (7)] has been introduced. Here, a dimensionality reduction of, on average, 0.26 for the average structures is found, whereas it is 0.32 for the open structures, which is in agreement with our previous results.[31] For larger proteins, the dimensionality reduction is even more pronounced. For example, for proteins with >200 residues in the dataset, this value amounts to 0.45.

Compared with ED modes, both ENM and RCNMA perform, on average, similar in terms of the maximal overlap of mode directions and correlation of amplitudes of motions (Table I). However, there are some differences on the level of individual proteins. Figure S1 in Supporting Information shows differences between the maximal overlap values for modes either obtained from ENM or

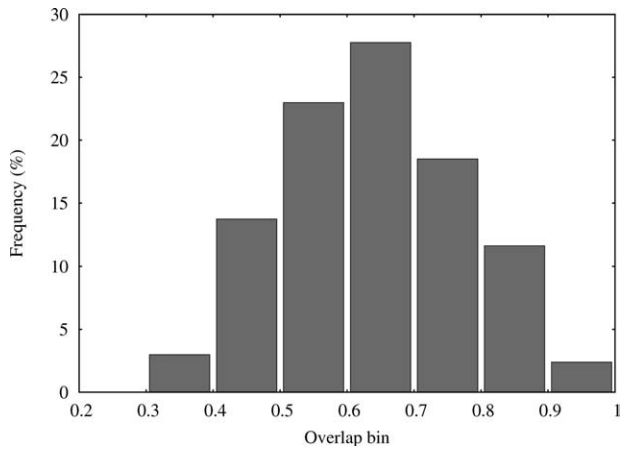

Figure 2

Frequency distribution of maximal overlap values of ENM modes with ED modes.

RCNMA with ED modes as a function of the dimensionality reduction. Differences in overlap values occur in both negative and positive directions and are mainly in the range between 0.05 and 0.2, indicating that using a coarse-grained protein representation does not deteriorate the agreement largely. This is also corroborated by the fact that there is no correlation between dimensionality reduction and overlap difference values and that both positive and negative overlap differences are observed even for the highest levels of coarse graining. Finally, no difference between ENM and RCNMA results were found if the average structures were used. For simplicity, we, thus, present ENM results from here onward unless stated otherwise. Details on RCNMA results are provided in the Supporting Information (Figs. S2–S5 in Supporting Information).

COMPARISON BETWEEN ED AND ENM MODES

The first five ED modes of each protein of the dataset were compared with ENM modes in terms of overlap, correlation, and spanning coefficient between the two sets of modes. Despite underlying differences between ED and normal mode methods, high maximal overlap and maximal correlation values between the two sets of modes were observed. Table I shows maximal overlap and maximal correlation values averaged over 335 proteins of ED modes with ENM modes. Only 3% of the proteins have overlap values <0.4, indicating an unsatisfactory agreement of mode directions, whereas 83% of the proteins have maximal overlap values >0.5, and more than 30% of the proteins have maximal overlap values >0.7 (Fig. 2). These high overlap values demonstrate that the essential motions extracted from MD trajectories can likewise be obtained from a CGNM method, albeit at a much lower computational expense. Furthermore, good overlap values on such a large and diverse dataset support the argument that the CGNM approaches are useful in describing motions of proteins with different and complex architectures, as long as they describe collective motions.[51,52,84] These collective modes, derived from either ED or ENM, have been shown previously to relate to biologically important conformational changes.[30,51,57,85–87]

In addition, the frequency distribution of ENM modes involved in the maximal overlap (Fig. 3) shows that these modes are among the lowest-frequency ones. Around 94% of the overlapping modes are among the first five nonzero modes of ENM. Interestingly, the probability of maximal mode involvement with ED decreases exponentially among the first five nonzero ENM modes, that is, the first (fifth) nonzero lowest-frequency modes are considered in 45% (3%) of all cases. This result can be helpful in designing approaches that make use of the directional information provided by normal modes. Interestingly, a similar trend is reported for experimental conformational changes on a large dataset of ~4,000 proteins.[88] Contrary to ENM, the frequency distribution of the first five ED modes involved in the maximal overlap does not show any single mode dominance, that is, the first (fifth) nonzero lowest-frequency modes are considered in 21% (18%) of all cases. This is probably due to the presence of anharmonic modes in ED, which are associated with crossing energy barriers during MD simulation and reside among the first few modes.[55,56]

Correlations of the amplitudes of motions described by ED and ENM modes are even higher than overlap values (Table I), with a mean value around 0.73, more than 94% of the cases with a correlation value >0.5, and still

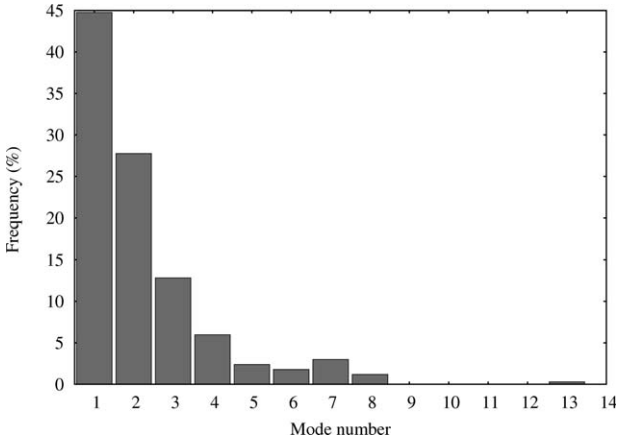

Figure 3

Frequency distribution of ENM mode numbers involved in the maximal overlap with ED modes. Mode 1 is the first nonzero frequency mode.

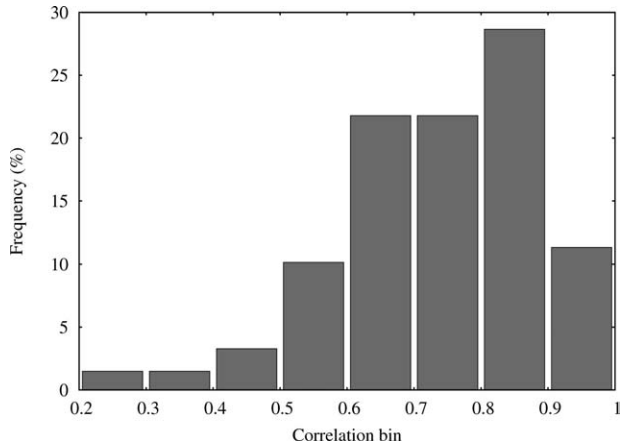

Figure 4

Frequency distribution of maximal correlation values of ENM modes with ED modes.

more than 40% of the cases with a correlation value >0.80 (Fig. 4). The finding that the correlations of the amplitudes of motions are higher than the overlap of the directions of collective atom displacements is rooted in the fact that an individual mode is only a single piece of the overall motion of a molecule. As such, it may be deceiving when it comes to characterizing motional correlations of structurally mobile regions: positive correlation ascribed to one pair of residues from one low-frequency mode can be negated or even reversed by a group of modes that display a negative correlation.[89] Thus, positive or negative motional correlations observed in a single ENM mode need not necessarily coincide perfectly with those observed in a single ED mode. Still, in comparison with ED modes, the above results emphasize that low-frequency modes of ENM do describe well not only magnitudes of motions but also directions of motions.

To analyze how well each of the five modes of ED can be described by ENM modes collectively and to explore the minimal set of the most contributing ENM modes in the low-frequency range, the spanning coefficient [Eq. (6)] was calculated with a varying mode number (Fig. 5). The number of ENM modes used in Eq. (6) is given in percent of all available ENM modes for a protein. It was found that only a relatively small number of normal modes are needed to describe the space spanned by low-frequency ED modes: the space spanned by the first 10% and 25% of the ENM modes describes, on average, around 68% and 84% of all five modes of ED, respectively (Table I). If calculated separately, almost no difference in the mean spanning coefficients for each of the five ED modes were found (Table S2 in Supporting Information). This again shows that these ED modes, on average, exhibit comparable behavior when compared with ENM modes. In the case of 10% (on average, 30) of the modes, a rather broad distribution of points shows that not all of the five ED modes are well represented. Still, the results emphasize that the two methods, which differ considerably in their underlying techniques and the level of coarse graining, show not only a high mean maximal overlap of 0.65 but also a good overlap between the two important subspaces defined by the first five ED modes and, on average, 30 (10% of all) or 85 (25% of all) ENM modes. Furthermore, it shows how much dynamic information a single protein structure can provide with almost no computational cost. For normal mode-based approaches, this result is helpful in deciding on the number of modes to be considered to explore the essential conformational space.

Similarity of motion in protein classes and folds: ED versus ENM modes

To analyze the dynamic similarities or dissimilarities within different protein classes and folds based on ED and ENM modes, the proteins in the dataset were classified according to CATH. Maximal overlap and correlation of amplitudes of motions described by ED and ENM were sorted for these proteins according to different classes and folds, and the mean and the standard deviation values were calculated (see Table S1). With respect to the maximal overlap and the correlation of amplitudes between ED and ENM modes, no prominent differences among different classes, that is, "mainly-α," "mainly-β," "α-β," and "low secondary structure content," were found (Table II) when considering standard deviations of 0.1 for all classes. This is in agreement with a recent study on a smaller dataset.[64] This finding shows that, on average, ED and ENM analyses are equally applicable to proteins of different classes.

In addition, the collectivity of the modes involved in maximal overlap between ED and ENM methods were

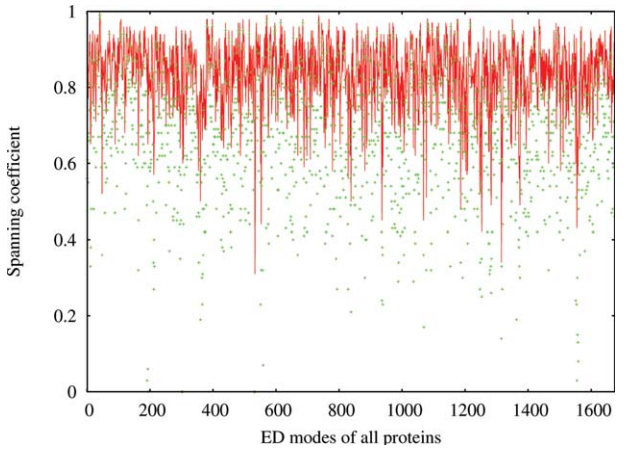

Figure 5

Spanning coefficients [Eq. (6)] of the first five ED modes of all proteins in the dataset using 10 (green points) and 25% (red line) of all available ENM modes. Abscissa values 1 to 5 relate to the first five ED modes of the first protein in the dataset, values 6 to 10 to the second protein, and so forth.

Table II
Mean Results for Different Protein Classes

Class No.[a]	Collectivity index[b] ED	Collectivity index[b] ENM	Maximal overlap[c]	Maximal correlation[c]
1 (90)	0.41 (±0.11)	0.34 (±0.14)	0.64/0.65	0.73/0.72
2 (103)	0.35 (±0.10)	0.32 (±0.16)	0.64/0.65	0.75/0.73
3 (122)	0.35 (±0.10)	0.30 (±0.13)	0.62/0.63	0.74/0.75
4 (5)	0.42 (±0.13)	0.38 (±0.05)	0.65/0.67	0.75/0.76

[a]Protein classes "mainly-α," "mainly-β," 'α-β," and "some secondary structure content" are numbered 1 to 4. The numbers of domains in the respective class are given in parentheses.
[b]Mean collectivity index [Eq. (8)] of modes involved in maximal overlaps between ED and ENM modes with standard deviations in parentheses.
[c]Mean of the maximal overlap [Eq. (4)] or correlation values [Eq. (5)] of RCNMA/ENM with ED modes.

sorted according to CATH classes. Notably, lower collectivity indices of ED modes for mainly-β and α-β classes were found compared to the other two classes (Table II). The higher collectivity indices in the mainly-α class are probably due to a lower packing as compared with mainly-β or α-β classes.[90] The lower packing provides the required space for collective motions of atoms. Similarly, the high collectivity index in the low secondary structure content class could be attributed to the underlying flexibility of the structures due to the lack of secondary structure, such that motions again involve many atoms.

As for the topology level, proteins of the same topology show, in general, a similar maximal overlap [Eq. (4)] between ED and ENM modes with a standard deviation of around 0.1 (see Table S1). However, for some members with the same topology or of the same homologous superfamily, much lower overlap values were observed compared with the remaining topology or family members. To investigate whether this denotes a limitation of ED or ENM, three pairs of proteins, one of each of the three main CATH classes, were selected such that each protein in the pair differs highly from the other in the maximal overlap or correlation values (values are highlighted in bold in Table S1), although both proteins have the same topology and belong to the same homologous superfamily. These selected pairs are listed in Table III. Assuming that proteins with the same topology and residing in the same superfamily have similar dynamics,[91] the modes derived from either ED or ENM of the two selected proteins in each pair were compared in terms of maximal overlap and maximal correlation. Interestingly, in all three cases the maximal overlaps and correlations in amplitudes of motions obtained from ENM were found to be higher than those of ED (Table III). As such, the mean maximal overlap (correlation) values are 0.31 (0.57) and 0.56 (0.84) using ED and ENM, respectively.

It is worth mentioning here that functional modes are usually among the most robust modes,[92] even if the sequence varies.[93] Along these lines, Leo-Macias *et al.*[94] have concluded that, to a significant extent, the structural response of a protein topology to sequence changes takes place by means of collective deformations along combinations of a small number of low-frequency modes. Recently, it has also been argued that dynamics and functional promiscuity are building blocks of protein evolvability.[95] In accordance with this view, the results presented here for three selected proteins show that ENM better describes these robust and evolutionary conserved modes than ED. This very likely is due to underlying MD simulations of only 10 ns length, which are too short to capture these modes due to slow barrier crossing on rugged energy landscapes.[25,26]

Comparison of ED and ENM modes with experimentally observed conformational changes

So far, ED and ENM modes have been compared with respect to each other. From an application point of view, however, it is interesting to investigate which of the methods better describe experimentally observed conformational changes. For this, 13 proteins were selected from the dataset for which alternate conformations were available in the MolmovDB[81] (see Materials and Methods). Maximal overlap values [Eq. (4)] were then calculated between the experimental conformational changes and ED or ENM modes (Fig. 6). In general, smaller maximal overlap values are observed for both, ED and ENM, when compared with the experimental conformational changes, than if ED and ENM are compared with each other. This may be attributed to the nature of the conformational changes observed for the 13 proteins, which are mainly loop movements and helix deformations (Table IV) and, hence, rather localized motions. Still, ENM performs better than ED in describing experimentally observed conformational changes: in 7 of 13 cases ENM outperforms ED, whereas the reverse is true in only three cases (in the remaining three cases ED and ENM only differ negligibly, that is, by less than 0.04 in their maximal overlap values). An identical trend is

Table III
Comparison of ED and CGNM Modes for Selected Protein Pairs

PDB codes[a]	Maximal overlap[b] ED	Maximal overlap[b] ENM	Maximal Correlation[c] ED	Maximal Correlation[c] ENM
1ahq/1cof (43%)	0.28	0.59	0.58	0.92
1ccr/1co6 (50%)	0.26	0.46	0.41	0.72
1idi/1ntn (66%)	0.38	0.64	0.72	0.87

[a]PDB codes of selected protein pairs in which both proteins have the same topology and are of the same homologous superfamily[66] but strongly differ in their maximal overlap values between ED and ENM. In parentheses, the sequence identities of the protein pairs are given.
[b]Highest overlap [Eq. (4)] between two sets of modes of each pair of proteins using either ED or ENM modes.
[c]Highest correlation [Eq. (5)] between two sets of modes of each pair of proteins using either ED or ENM modes.

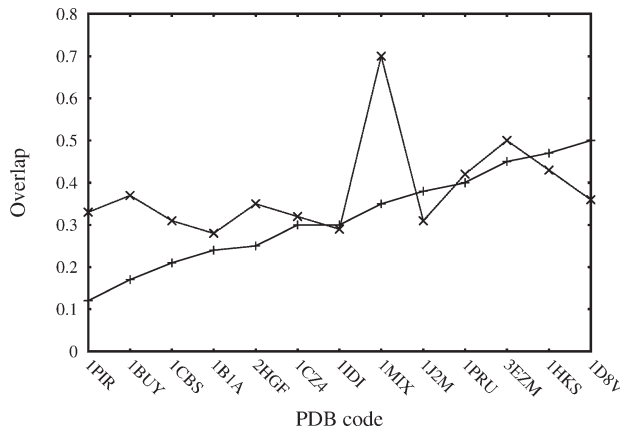

Figure 6

Maximal overlap [Eq. (4)] between ED (+) or ENM (×) modes and the conformational change vectors obtained from two alternate conformations of the selected proteins (Table IV).

observed for RCNMA results (Fig. S5 in Supporting Information), corroborating previous observations by us[31] and others.[96] This shows that experimentally observed conformational changes are better described by low frequency modes of CGNM than by ED modes, which confirms that CGNM modes are in general robust,[82,92] even to high levels of coarse-graining,[97] in describing functional modes.

In general, no correlation between the types of experimentally observed motions (Table IV) and overlap values between ED and ENM can be observed. In the only case that is governed by global collective conformational changes (PDB ID 1mix/1miz), the overlap between ED and ENM is 0.65 and, hence, close to average. However, this is also the case where ED shows a considerably worse overlap with the experimental conformational change than ENM (Fig. 6). Although this single observation must be interpreted with caution, it may point to the fact that global collective conformational changes are better described by ENM than by ED. Finally, the ENM and ED performance in terms of describing experimentally observed conformational changes does not depend on the size of the protein: the correlation coefficient between protein size, measured in terms of the number of residues, and the overlap value is $r = 0.17$ (0.04) for ENM (ED). This is in agreement with a previous study by Yang *et al.*[37]

CONCLUSION

Specific functions of biological systems often require conformational transitions of macromolecules. Of the different computational approaches targeting the modelling of protein motions, MD simulation is one of the most widely applied and accurate computational techniques currently being used. However, MD simulations are computationally expensive and explore limited conformational space due to slow barrier crossing on the rugged energy landscape of macromolecules.[25,26] Hence, there have been efforts to develop alternative approaches that are computationally more efficient. Among these are CGNM approaches, in particular the ENM and the RCNMA, which provide directions of intrinsic motions in terms of harmonic modes.[29,31]

To explore the validity and the applicability of CGNM approaches, a large-scale comparison of ED modes from MD simulations and normal modes from CGNM was performed on a dataset of 335 proteins. Low-frequency normal modes from ENM correlate very well with ED modes in terms of directions of motions (average maximal overlap: 0.65) and relative amplitudes of motions (average maximal overlap: 0.73). Notably, a similar performance has been found if normal modes from RCNMA are used, despite a higher level of coarse graining. On average, the space spanned by the first quarter of ENM modes describes 84% of the space spanned by the five ED modes. Furthermore, no prominent differences for ED and CGNM modes among different protein structure classes (CATH classification) were found. These findings demonstrate the general potential of CGNM approaches for describing intrinsic motions of proteins with little computational cost. For selected cases, CGNM modes are even more robust among proteins that have the same topology or are of the same homologous superfamily than ED modes. In view of recent[94,98,99] evidence regarding evolutionary conservation of vibrational dynamics, this suggests that ED modes, in some cases, might not be representative of the underlying dynamics that is

Table IV

Selected Dataset With Experimentally Observed Conformational Changes

PDB codes[a]	No. residues	RMSD[b]	Collectivity index[c]	Description of motion[d]
1cbs/1blr	137	2.84/3.33	0.37	Helix deforms; loops move
1idi/1idg	74	5.64/6.73	0.42	Change everywhere; uncorrelated
1b1a/1be1	137	3.28/5.09	0.51	Helixes rearrange; loops move
1buy/1eer	156	2.43/2.58	0.30	Helix deforms; loops move
1pir/1fx9	124	2.27/2.54	0.41	Loops move
2hgf/1bht	91	1.38/1.38	0.37	Small loop moves
1mix/1miz	186	1.21/2.73	0.43	Domain moves; correlated
1pru/1prv	56	1.14/5.98	0.52	Change everywhere; uncorrelated
1j2m/1j2n	89	3.65/5.30	0.58	Helix turns; loop moves
3ezm/2ezm	97	16.06/12.47	0.76	Scramble/folding
1hks/1hkt	97	1.28/2.31	0.17	Loop moves
1d8v/1cf5	243	1.48/2.26	0.53	Change everywhere; uncorrelated
1cz4/1cz5	176	1.01/3.24	0.23	Loop moves

[a]Selected PDB codes from the dataset of 335 proteins for which an alternate conformation is available in the MolmovDB.
[b]RMSD of starting experimental/MD average structure with the target experimental structure, in Å.
[c]Collectivity index [Eq. (8)] of the conformational change of both proteins.
[d]Description of the conformational change obtained by visual inspection.

characteristic for a whole family, probably due to insufficient sampling of some of the family members by MD.

Intrinsic motions of a protein relate to its function according to the conformational selection model[32–35] and to allosteric regulation[100,101] and evolvability[94,95] of the structure. Being able to predict intrinsic motions of proteins with little computational cost will be very helpful in the development of computational approaches that allow efficient conformational sampling of biomacromolecules for both understanding biological processes at the molecular level and applications in macromolecular complex prediction and structure-based drug design. CGNM approaches are promising in that sense.

ACKNOWLEDGMENTS

We thank Manuel Rueda (University of Barcelona) for help with the MoDEL database and Daniel Cashman (Heinrich-Heine-University, Düsseldorf) for critically reading the manuscript.

REFERENCES

1. Karplus M, McCammon JA. Dynamics of proteins: elements and function. Annu Rev Biochem1983; 1983;52:263–300.
2. Quiocho FA, Ledvina PS. Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes. Mol Microbiol 1996;20:17–25.
3. Perutz MF, Wilkinson AJ, Paoli M, Dodson GG. The stereochemical mechanism of the cooperative effects in hemoglobin revisited. Annu Rev Biophys Biomol Struct 1998;27:1–34.
4. Karplus M, Gao YQ. Biomolecular motors: the F1-ATPase paradigm. Curr Opin Struct Biol 2004;14:250–259.
5. Vallee RB, Hook P. Molecular motors: A magnificent machine. Nature 2003;421:701–702.
6. Hammes GG. Multiple conformational changes in enzyme catalysis. Biochemistry (Mosc.) 2002;41:8221–8228.
7. Benkovic SJ, Hammes-Schiffer S. A perspective on enzyme catalysis. Science 2003;301:1196–1202.
8. Huse M, Kuriyan J. The conformational plasticity of protein kinases. Cell 2002;2002:275–282.
9. Doyle DA. Structural changes during ion channel gating. Trends Neurosci 2004;27:298–302.
10. Swartz KJ. Towards a structural view of gating in potassium channels. Nat Rev Neurosci 2004;5:905–916.
11. Wlodawer A, Vondrasek J. Inhibitors of HIV-1 protease: A major success of structure-assisted drug design. Annu Rev Biophys Biomol Struct 1998;27:249–284.
12. Wilson DK, Tarle I, Petrash JM, Quiocho FA. Refined 1.8 A structure of human aldose reductase complexed with the potent inhibitor zopolrestat. Proc Natl Acad Sci USA 1993;90:9847–9851.
13. Schlauderer GJ, Schulz GE. The structure of bovine mitochondrial adenylate kinase: comparison with isoenzymes in other compartments. Protein Sci 1996;5:434–441.
14. Schlauderer GJ, Proba K, Schulz GE. Structure of a mutant adenylate kinase ligated with an ATP-analogue showing domain closure over ATP. J Mol Biol 1996;256:223–227.
15. Muller CW, Schlauderer GJ, Reinstein J, Schulz GE. Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding. Structure 1996;4:147–156.
16. Stuckey J, Schubert H, Fauman E, Zhang Z-Y, Dixon J, Saper M. Crystal structure of Yersinia protein tyrosine phosphatase at 2.5 A and the complex with tungstate. Nature 1994;370:571–575.
17. Schubert HL, Fauman EB, Stuckey JA, Dixon JE, Saper MA. A ligand-induced conformational change in the Yersinia protein tyrosine phosphatase. Protein Sci 1995;4:1904–1913.
18. Vandonselaar M, Hickie RA, Quail JW, Delbaere LTJ. Trifluoperazine-induced conformational change in Ca(2+)-calmodulin. Nat Struct Biol 1994;1:795–801.
19. Kuboniwa H, Tjandra N, Grzesiek S, Ren H, Klee CB, Bax A. Solution structure of calcium-free calmodulin. Nat Struct Biol 1995;2:768–776.
20. Goh CS, Milburn D, Gerstein M. Conformational changes associated with protein-protein interactions. Curr Opin Struct Biol 2004;14:104–109.
21. Chen R, Mintseris J, Janin J, Weng Z. A protein-protein docking benchmark. Proteins 2003;52:88–91.
22. McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. Nature 1977;267:585–590.
23. Hansson T, Oostenbrink C, van Gunsteren W. Molecular dynamics simulations. Curr Opin Struct Biol 2002;12:190–196.
24. Adcock SA, McCammon JA. Molecular dynamics: survey of methods for simulating the activity of proteins. Chem Rev 2006;106: 1589–1615.
25. Cheatham TE, Young MA. Molecular dynamics simulation of nucleic acids: successes, limitations, and promise. Biopolymers 2000;56:232–256.
26. Moraitakis G, Purkiss A, Goodfellow J. Simulated dynamics and biological macromolecules. Rep Prog Phys 2003;66:383–406.
27. Hinsen K. Analysis of domain motions by approximate normal mode calculations. Proteins 1999;33:417–429.
28. Hinsen K, Thomas A, Field MJ. Analysis of domain motions in large proteins. Proteins 1999;34:369–382.
29. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J 2001;80:505–515.
30. Tama F, Sanejouand YH. Conformational change of proteins arising from normal mode calculations. Protein Eng 2001;14:1–6.
31. Ahmed A, Gohlke H. Multiscale modeling of macromolecular conformational changes combining concepts from rigidity and elastic network theory. Proteins 2006;63:1038–1051.
32. Ma B, Kumar S, Tsai CJ, Nussinov R. Folding funnels and binding mechanisms. Protein Eng 1999;12:713–720.
33. Tsai C-J, Kumar S, Ma B, Nussinov R. Folding funnels, binding funnels, and protein function. Prot Sci 1999;8:1181–1190.
34. Tsai C-J, Ma B, Nussinov R. Folding and binding cascades: Shifts in energy landscapes. Proc Natl Acad Sci USA 1999;96:9970–9972.
35. Berger C, Weber-Bornhauser S, Eggenberger J, Hanes J, Pluckthun A, Bosshard HR. Antigen recognition by conformational selection. FEBS Lett 1999;450:149–153.
36. Trakhanov S, Vyas NK, Luecke H, Kristensen DM, Ma J, Quiocho FA. Ligand-free and -bound structures of the binding protein (LivJ) of the Escherichia coli ABC leucine/isoleucine/valine transport system: trajectory and dynamics of the interdomain rotation and ligand specificity. Biochemistry 2005;44:6597–6608.
37. Yang L, Song G, Jernigan RL. How well can we understand large-scale protein motions using normal modes of elastic network models? Biophys J 2007;93:920–929.
38. Zhang Z, Shi Y, Liu H. Molecular dynamics simulations of peptides and proteins with amplified collective motions. Biophys J 2003;84: 3583–3593.
39. Tatsumi R, Fukunishi Y, Nakamura H. A hybrid method of molecular dynamics and harmonic dynamics for docking of flexible ligand to flexible receptor. J Comput Chem 2004;25:1995–2005.
40. He J, Zhang Z, Shi Y, Liu H. Efficiently explore the energy landscape of proteins in molecular dynamics simulations by amplifying collective motions. J Chem Phys 2003;119:4005–4017.
41. Cavasotto CN, Kovacs JA, Abagyan RA. Representing receptor flexibility in ligand docking through relevant normal modes. J Am Chem Soc 2005;127:9632–9640.

42. May A, Zacharias M. Protein-protein docking in CAPRI using ATTRACT to account for global and local flexibility. Proteins 2007;69:774–780.
43. May A, Zacharias M. Protein-ligand docking accounting for receptor side chain and global flexibility in normal modes: evaluation on kinase inhibitor cross docking. J Med Chem 2008;51:3499–3506.
44. Delarue M, Dumas P. On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models. Proc Natl Acad Sci USA 2004;101:6957–6962.
45. Hinsen K, Reuter N, Navaza J, Stokes DL, Lacapere JJ. Normal mode-based fitting of atomic structure into electron density maps: application to sarcoplasmic reticulum Ca-ATPase. Biophys J 2005;88:818–827.
46. Tama F, Miyashita O, Brooks CL. Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis. J Mol Biol 2004;337:985–999.
47. Tama F, Miyashita O, Brooks CL. Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM. J Struct Biol 2004;147:315–326.
48. Kim MK, Jernigan RL, Chirikjian GS. Efficient generation of feasible pathways for protein conformational transitions. Biophys J 2002;83:1620–1630.
49. Miyashita O, Onuchic JN, Wolynes PG. Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. Proc Natl Acad Sci USA 2003;100:12570–12575.
50. Miyashita O, Wolynes PG, Onuchic JN. Simple energy landscape model for the kinetics of functional transitions in proteins. J Phys Chem B 2005;109:1959–1969.
51. Bahar I, Rader A. Coarse-grained normal mode analysis in structural biology. Curr Opin Struct Biol 2005;15:586–592.
52. Ma J. Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. Structure 2005;13:373–380.
53. Amadei A, Linssen AB, Berendsen HJ. Essential dynamics of proteins. Proteins 1993;17:412–425.
54. van Aalten DM, Amadei A, Linssen AB, Eijsink VG, Vriend G, Berendsen HJ. The essential dynamics of thermolysin: confirmation of the hinge-bending motion and comparison of simulations in vacuum and water. Proteins 1995;22:45–54.
55. Hayward S, Kitao A, Go N. Harmonic and anharmonic aspects in the dynamics of BPTI: a normal mode analysis and principal component analysis. Protein Sci 1994;3:936–943.
56. Hayward S, Kitao A, Go N. Harmonicity and anharmonicity in protein dynamics: a normal mode analysis and principal component analysis. Proteins 1995;23:177–186.
57. Kitao A, Go N. Investigating protein dynamics in collective coordinate space. Curr Opin Struct Biol 1999;9:164–169.
58. Dauber-Osguthorpe P. Low frequency motion in proteins comparison of normal mode and molecular dynamics of Streptomyces griseus protease A. J Comput Phys 1999;151:169–189.
59. Brooks B, Karplus M. Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme. Proc Natl Acad Sci USA 1985;82:4995–4999.
60. McCammon JA, Gelin BR, Karplus M, Wolynes PG. The hinge-bending mode in lysozyme. Nature 1976;262:325–326.
61. Brooks B, Karplus M. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. Proc Natl Acad Sci USA 1983;80:6571–6575.
62. Leherte L, Vercauteren DP. Collective motions of rigid fragments in protein structures from smoothed electron density distributions. J Comput Chem 2008;29:1472–1489.
63. Song G, Jernigan R. An enhanced elastic network model to represent the motions of domain-swapped proteins. Proteins 2006;63: 197–209.
64. Rueda M, Chacon P, Orozco M. Thorough validation of protein normal mode analysis: a comparative study with essential dynamics. Structure 2007;15:565–575.
65. Rueda M, Ferrer-Costa C, Meyer T, Perez A, Camps J, Hospital A, Gelpi JL, Orozco M. A consensus view of protein dynamics. Proc Natl Acad Sci USA 2007;104:796–801.
66. Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM. CATH—a hierarchic classification of protein domain structures. Structure 1997;5:1093–1108.
67. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The Protein Data Bank—a computer-based archival file for macromolecular structures. J Mol Biol 1977;112:535–542.
68. Hamelryck T, Manderick B. PDB file parser and structure class implemented in Python. Bioinformatics 2003;19:2308–2310.
69. Jacobs D, Rader AJ, Kuhn L, Thorpe MF. Protein flexibility predictions using graph theory. Proteins 2001;44:150–165.
70. Ma J. New advances in normal node analysis of supermolecular complexes and applications to structural refinement. Curr Protein Pept Sci 2004;5:119–123.
71. Tirion M. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. Phys Rev Lett 1996;77:1905.
72. Hinsen K. Analysis of domain motions by approximate normal mode calculations. Proteins 1998;33:417–429.
73. Bahar I, Atilgan AR, Erman B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Folding Design 1997;2:173–181.
74. Durand P, Trinquier G, Sanejouand Y-H. A new approach for determining low-frequency normal modes in macromolecules. Biopolymers 1994;34:759–771.
75. Jacobs DJ, Thorpe MF. Generic rigidity percolation: the pebble game. Phys Rev Lett 1995;75:4051–4054.
76. Jacobs DJ, Kuhn LA, Thorpe MF. Flexible and rigid regions in proteins. In: Thorpe MF, Duxbury PM, editors. Rigidity theory and applications. New York, NY: Kluwer Academic/Plenum Publishers; 1999. pp 357–384.
77. Tama F, Gadea FX, Marques O, Sanejouand YH. Building-block approach for determining low-frequency normal modes of macromolecules. Proteins 2000;41:1–7.
78. Marques O, Sanejouand YH. Hinge-bending motion in citrate synthase arising from normal mode calculations. Proteins 1995;23:557–560.
79. Li G, Cui Q. A coarse-grained normal mode approach for macromolecules: an efficient implementation and application to Ca2+-ATPase. Biophys J 2002;83:2457–2474.
80. Bruschweiler R. Collective protein dynamics and nuclear-spin relaxation. J Chem Phys 1995;102:3396–3403.
81. Echols N, Milburn D, Gerstein M. MolMovDB: analysis and visualization of conformational change and structural flexibility. Nucleic Acids Res 2003;31:478–482.
82. Lu M, Ma J. The role of shape in determining molecular motions. Biophys J 2005;89:2395–2401.
83. Tay TS, Whiteley W. Recent advances in generic rigidity of structures. Struct Topol 1984;9:31–38.
84. Chennubhotla C, Rader AJ, Yang LW, Bahar I. Elastic network models for understanding biomolecular machinery: from enzymes to supramolecular assemblies. Phys Biol 2005;2:S173–S180.
85. Kurkcuoglu O, Robert I, Doruker P. Collective dynamics of large proteins from mixed coarse-grained elastic network model. QSAR Comb Sci 2005;24:443–448.
86. Berendsen H, Hayward S. Collective protein dynamics in relation to function. Curr Opin Struct Biol 2000;10:165–169.
87. Petrone P, Pande V. Can conformational change be described by only a few normal modes? Biophys J 2006;90:1583–1593.
88. Krebs WG, Alexandrov V, Wilson CA, Echols N, Yu H, Gerstein M. Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic. Proteins 2002;48:682–695.
89. Van Wynsberghe A, Cui Q. Interpreting correlated motions using normal mode analysis. Structure 2006;14:1647–1653.

90. Lobanov M, Bogatyreva N, Galzitskaya O. Radius of gyration as an indicator of protein structure compactness. Mol Biol 2008;42:623–628.
91. Keskin O, Jernigan RL, Bahar I. Proteins with similar architecture exhibit similar large-scale dynamic behavior. Biophys J 2000;78: 2093–2106.
92. Nicolay S, Sanejouand YH. Functional modes of proteins are among the most robust. Phys Rev Lett 2006;96:078104.
93. Zheng W, Brooks B, Thirumalai D. Low-frequency normal modes that describe allosteric transitions in biological nanomachines are robust to sequence variations. Proc Natl Acad Sci USA 2006;103: 7664–7669.
94. Leo-Macias A, Lopez-Romero P, Lupyan D, Zerbino D, Ortiz AR. An analysis of core deformations in protein superfamilies. Biophys J 2005;88:1291–1299.
95. Tokuriki N, Tawfik DS. Protein dynamism and evolvability. Science 2009;324:203–207.
96. Liu L, Koharudin LM, Gronenborn AM, Bahar I. A comparative analysis of the equilibrium dynamics of a designed protein inferred from NMR, X-ray, and computations. Proteins 2009; 77:927–939.
97. Doruker P, Jernigan RL, Bahar I. Dynamics of large proteins through hierarchical levels of coarse-grained structures. J Comput Chem 2002;23:119–127.
98. Maguid S, Fernandez-Alberti S, Echave J. Evolutionary conservation of protein vibrational dynamics. Gene 2008;422:7–13.
99. Maguid S, Fernandez-Alberti S, Ferrelli L, Echave J. Exploring the common dynamics of homologous proteins. Application to the globin family. Biophys J 2005;89:3–13.
100. Kern D, Zuiderweg ER. The role of dynamics in allosteric regulation. Curr Opin Struct Biol 2003;13:748–757.
101. Popovych N, Sun S, Ebright RH, Kalodimos CG. Dynamically driven protein allostery. Nat Struct Mol Biol 2006;13:831–838.