PROTEINS: Structure, Function, and Bioinformatics 63:1038–1051 (2006)

Multiscale Modeling of Macromolecular Conformational Changes Combining Concepts From Rigidity and Elastic Network Theory

Aqeel Ahmed and Holger Gohlke*
Department of Biology and Computer Science, J. W. Goethe-University, Frankfurt, Germany

ABSTRACT **The development of a two-step approach for multiscale modeling of macromolecular conformational changes is based on recent developments in rigidity and elastic network theory. In the first step, static properties of the macromolecule are determined by decomposing the molecule into rigid clusters by using the graph-theoretical approach FIRST and an all-atom representation of the protein. In this way, rigid clusters are not limited to consist of residues adjacent in sequence or secondary structure elements as in previous studies. Furthermore, flexible links between rigid clusters are identified and can be modeled as such subsequently. In the second step, dynamical properties of the molecule are revealed by the rotations-translations of blocks approach (RTB) using an elastic network model representation of the coarse-grained protein. In this step, only rigid body motions are allowed for rigid clusters, whereas links between them are treated as fully flexible. The approach was tested on a data set of 10 proteins that showed conformational changes on ligand binding. For efficiency, coarse-graining the protein results in a remarkable reduction of memory requirements and computational times by factors of 9 and 27 on average and up to 25 and 125, respectively. For accuracy, directions and magnitudes of motions predicted by our approach agree well with experimentally determined ones, despite embracing in extreme cases >50% of the protein into one rigid cluster. In fact, the results of our method are in general comparable with when no or a uniform coarse-graining is applied; and the results are superior if the movement is dominated by loop or fragment motions. This finding indicates that explicitly distinguishing between flexible and rigid regions is advantageous when using a simplified protein representation in the second step. Finally, motions of atoms in rigid clusters are also well predicted by our approach, which points to the need to consider mobile protein regions in addition to flexible ones when modeling correlated motions. Proteins 2006;63:1038–1051.** © 2006 Wiley-Liss, Inc.

Key words: normal mode analysis; rigidity/flexibility; collective motions; vibrations; proteins; induced fit

INTRODUCTION

Specific functions of biological systems often require conformational transitions of macromolecules. Such changes range from movements of single side-chains and loop rearrangements to large-scale domain motions, as observed in the ribosome,[1] F_0F_1-ATPase,[2] chaperonins,[3] or viruses.[4] In binding events involving macromolecules, molecular motions provide the origin of the plasticity of the binding partners, enabling them to conformationally adapt to each other.[5,6] Thus, being able to describe and predict conformational changes of biological macromolecules is not only important for understanding their impact on biological function but will also have implications for the modeling of (macro)molecular complex formation.[7]

Computational approaches based on atomic models of the biological systems have been well proven for understanding and modeling conformational changes. Although applicable to small-scale transitions, current molecular dynamics simulations are too expensive for studying large-scale molecular motions because of the limited sampling of conformational space during typical simulation times of only a few tens of nanoseconds. Normal mode analysis (NMA)[8] provides an interesting alternative in this case.[9] Here, instead of numerically solving Newton's equations of motion, an analytical solution yields collective variables (normal modes) that describe the dynamics of the system. Despite the harmonic approximation inherent to the method that neglects transitions from one local minimum to another, this approach has been successfully applied to study protein motions.[10,11] In particular, recent studies have shown that biologically relevant motions can be reliably described by considering only a small subset of low-frequency normal modes (in many cases, even a single mode is sufficient).[12]

Initially, in "standard NMA," all-atom models of the macromolecules together with conventional force fields were applied, requiring energy minimization to reach a stationary point on the potential energy surface prior to the NMA step. Furthermore, NMA involves the numeric diagonalization of a $3N$-dimensional matrix (where $3N$ is

*Correspondence to: Holger Gohlke; Marie-Curie-Str. 9, 60439 Frankfurt, Germany. E-mail: gohlke@bioinformatik.uni-frankfurt.de

Received 12 May 2005; Revised 24 October 2005; Accepted 24 October 2005

Published online 21 February 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20907

the number of degrees of freedom of the system), leading to considerable requirements in memory and computational time even in the case of proteins with a few thousand atoms. These limitations have recently been overcome by modeling proteins based on reduced representations[13]; that is, by considering C_α atoms (from now on, referred to as "ANM analysis")[14] or one mass-point per residue[15] only, or even more coarse-grained models[16] such as protein shapes filled with uniformly spaced lattice points[17] or other representations of shape and mass distribution.[18,19] This results in a considerable decrease of the dimensionality of the matrix. Instead of atomic force fields, simplified potentials in terms of Hookean springs are then used, which connect the above "particles" and result in an elastic network.[20] Because, by definition, all springs are in a relaxed state in this network, no energy minimization is required. Convincingly, good agreement between amplitudes and directions of the motions predicted by these simplified models and those based on all-atom representations or experimentally observed ones has been found.[15,20] This finding suggests that it is sufficient to capture the shape and mass distribution of a protein for a reliable description of its large-scale displacements. In that respect, discrete representations of the particle mass distribution based on low-resolution data of macromolecules (e.g., electron densities) have been applied to describe protein motions.[18,19]

An alternative way of coarse-graining is provided by considering proteins to be constructed by (quasi-)rigid bodies that are connected by flexible parts. The rotations-translations of blocks approach (RTB) proposed by Sanejouand et al.[21,22] is a method along these lines, for which a more efficient implementation has been reported recently.[23] Here, the $3N$-dimensional matrix of the entire system is projected into a $6n$-dimensional subspace spanned by the translational and rotational basis vectors of n blocks of amino acids, resulting in a reduction of the dimensionality of the matrix to be diagonalized and, hence, of the computational expense.[22–25] The approach is based on the hypothesis that low-frequency normal modes of proteins can be described as pure rigid body motions of the blocks. A comparison of RTB results with those obtained by standard NMA strongly supports this hypothesis[22,23,26], indicating that RTB is a promising approach for normal mode analysis of large systems.

So far, blocks were constructed by including up to six residues consecutive in sequence into one block,[22,23] by dissecting a protein into clusters of uniform size[24] or by considering whole protein subunits of a virus capsid as rigid.[26] These routes are somewhat counterintuitive for the model that forms the basis of the RTB method, because rigid parts of the protein are not distinguished from flexible regions. An ad hoc approach to overcome this limitation is to place each secondary structure element in its own block, with each residue lying outside the secondary structures being placed in one block.[22] In this case, however, a significant drop in the agreement between experimentally determined and computed directions of conformational change was found compared with, for example, using a one-residue-per-block partitioning. This finding suggests that considering all secondary structure elements as rigid blocks per se may not be appropriate. Moreover, using blocks of the same length as found when partitioning the protein according to the secondary structure elements, but randomly distributed along the protein chain, resulted in the same agreement.[22] This finding again questions the validity of partitioning the protein solely based on secondary structure information. In particular, the influence of nonlocal (tertiary) interactions on determining the flexibility/rigidity of a protein region is completely neglected.

It is of interest that flexibility concepts well grounded in mathematics,[27] engineering,[28] and solid-state physics[29] allow us to accurately and efficiently locate rigid and flexible regions within a macromolecule from a single, static structure. Based on the Molecular Framework Conjecture[30] (which extends Laman's theorem[31] to the subset of all 3D networks with molecule-like properties), the stability of a network of joints (e.g., atoms) connected by struts (e.g., bonds) is related to the average number of struts at the joints (e.g., the mean coordination, or number of bonds, for atoms in the network). When modeling covalent bonds and strong hydrogen-bonds and hydrophobic interactions appropriately as distance constraints between atoms, proteins can be considered as molecular frameworks.[30,32,33] A fast combinational algorithm, the pebble game, has been developed for counting bond-rotational degrees of freedom in 2D and 3D bond networks,[29,34,35] which can be related to regions of flexibility and rigidity. Applied to very large bond networks in amorphous materials, the algorithm on average scales linearly with network size.[34] The FIRST (Floppy Inclusion and Rigid Substructure Topology) software[36] is an implementation of the pebble game along with code that deduces and represents the protein covalent and noncovalent network. FIRST analyses have been used to accurately identify rigid regions as well as collectively and independently moving regions in a series of proteins,[32,36] relate the unfolding of a protein to its loss in structural stability,[33] identify protein folding cores and pathways,[37] provide the starting point for simulating the motions of flexible protein regions,[38–40] and determine the change in protein flexibility on complex formation.[41]

In the current work, we present a new approach that combines concepts from rigidity theory and elastic network theory described above to predict conformational changes from coarse-grained protein models. First, the protein is decomposed into rigid clusters by using FIRST analysis. In this way, the definition of blocks in an ad hoc manner is circumvented. Subsequently, an RTB analysis is carried out for the elastic network model of the protein, in which each rigid cluster is modeled as a block. Although the first step only provides information about which regions of the protein are flexible or rigid ("statics"), information about amplitudes and directions of motions ("dynamics") is obtained from the second step. These results as well as those from ANM analysis and standard NMA using a conventional force-field representation of the

protein are compared with experimentally observed conformational changes for a set of 10 proteins. It is encouraging that, in general, comparable results are found by our approach than if no or a uniform coarse-graining is applied, whereas our method becomes superior if the movement is dominated by loop or fragment motions.

METHODS

General Strategy

The rigid cluster NMA (RCNMA) presented in this study consists of two steps. Using an all-atom representation, a rigid cluster decomposition of the protein is obtained by FIRST analysis in the first step. In the second step, an RTB analysis is performed on the basis of the coarse-grained representation of the protein by FIRST consisting of rigid clusters connected by flexible links. This step yields information about amplitudes and directions of motions. Moreover, a reduced C_α atom representation of the protein in an elastic network is used in this step. In total, macromolecular conformational changes are predicted by a multiscale approach, the single steps of which will be detailed in the following sections.

Rigid Cluster Decomposition

Flexible and rigid regions of the proteins are identified by FIRST. Because the algorithm and underlying mathematical rigidity theory have been detailed elsewhere,[30,32,34,36] the approach is only briefly reviewed here. Applying the pebble game algorithm,[34] FIRST identifies and counts the bond-rotational degrees of freedom in a molecular framework, whose vertices represent protein atoms and whose edges represent covalent and noncovalent (hydrogen-bond and hydrophobic) constraints within the protein.[32,33,36] Flexibility in this network results from dihedral rotations of bonds that are not locked in by other bonds. Each bond is assigned by FIRST to be part of either a rigid cluster or a flexible region. A rigid cluster forms a collection of interlocked bonds in which no relative motion can be achieved without a cost in energy. Phrased differently, only rigid body motions (translation and rotation) are allowed for a rigid cluster. Hence, each rigid cluster will form a "block" in the subsequent RTB (see below). Note that in this way, blocks are not limited to comprise one or a few residues or single secondary structure elements but may extend over considerable parts of the macromolecular structure. Underconstrained regions in the network are typically flexible links between rigid clusters. These underconstrained regions are modeled on one-atom-per-block basis in the RTB (in which case only translational motion of the block is considered).

The molecular framework that represents the protein is completely defined by bond constraints between atoms and next-nearest neighbor constraints that define coordination angles between bonded atoms. Biologically important motions are in many cases characterized by low-frequency, large-amplitude structural fluctuations. By including constraints into the network that represent strong forces, high-frequency motions can be effectively quenched, thereby reducing the complexity of the energy landscape. Here, covalent and hydrogen bonds, salt bridges, and hydrophobic interactions are considered to be strong forces. Bond lengths (represented as distance constraints between bonded atoms) and bond coordination angles (represented as distance constraints between next-nearest neighbors of a central atom) are set to their values observed in the input structure. The configuration of double and partial double bonds (peptide bonds) are restricted by additional constraints.[33] The noncovalent interactions are modeled as described in previous FIRST studies.[33,37,41] For further details, see Gohlke et al.[41]

Rigid Cluster Normal Mode Analysis

Standard NMA requires the diagonalization of a $3N$-dimensional Hessian matrix H to obtain the normal modes of the system, where N is either the number of atoms if an atomic force-field-based representation of the molecule is used or the number of all "particles" (e.g., C_α atoms) if the molecule is represented as an elastic network (see below).[14,42] Using a coarse-grained representation of the molecule in terms of n rigid clusters (blocks) as obtained by FIRST analysis, the dimensionality of H can be reduced to $6n$ by following the RTB approach of Sanejouand and coworkers.[21,22] This leads to a reduction of memory requirements by a factor proportional to $(N/n)^2$ and a decrease in the number of operations required for the diagonalization by a factor proportional to $(N/n)^3$. In this approach, only translational and rotational degrees of freedom of the blocks are considered (for single atoms, only translations are considered), whereas no relative motions of elements within a block are allowed. Therefore, H is projected into the subspace spanned by translation/rotation basis vectors of the blocks. Diagonalizing the resulting matrix H_{sub} yields approximate low-frequency modes U_{sub} and eigenvalues Λ. Finally, atomic displacements can be obtained by expanding back the eigenvectors U_{sub} to the Cartesian space. Further details of the calculations are given in the Appendix.

Elastic Network Model

Based on a simplified representation of the potential energy,[14,15,20] the proteins are described as 3D elastic networks for the rigid cluster normal mode analysis. In this study, each amino acid is reduced to a single "particle" (the C_α atom), which acts as a junction in the network. Although an all-atom representation was required to appropriately model the network of constraints used as input for the FIRST analysis, reduced representations have been successfully applied in normal mode calculations to study macromolecular dynamics and conformational changes.[18,19,42] Interactions between these particles are modeled by Hookean springs based on a harmonic pairwise potential,[20] resulting in a total potential energy of the system given by

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

where r_c = 10Å is the cutoff up to which interactions between the C_α atoms are taken into account. Varying the

TABLE I. Protein Data Set Used in This Study

Protein	Open structure[a]	Closed structure[a]	Type of motion[b]
Adenylate kinase	4ake	1ake	H
Alcohol dehydrogenase	8adh	6adh	S
Aspartate aminotransferase	9aat	1ama	S
Calmodulin	1cfd	1cfc	H
Citrate synthase	5csc	6csc	S
CD kinase	1hck	1hlr	H[c]
HIV-1 protease	1hhp	1ajx	H[c]
LAO binding protein	2lao	1lst	H
Thymidylate synthase	3tms	2tsc	S
Tyrosine phosphatase	1ypt	1yts	H[c]

[a]PDB code of crystallographically determined structure.
[b]H = hinge bending; S = shear motion, as classified in the Molecular Movements Database.[43]
[c]Loop or fragment motions dominate the overall movements.

cutoff value in the range of 9–14 Å overall yielded qualitatively similar results in our case. r_{ij} and r_{ij}^{o} are the instantaneous and equilibrium distances between atoms i and j, respectively. $\theta(x)$ is the Heaviside step function that accounts for the cutoff effect of the interaction; 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^{-1} Å^{-2}.

According to the elastic network model,[42] the elements of matrix H are then obtained from the second derivatives of V with respect to the Cartesian coordinates of atoms i and j. In all cases, the H matrix is constructed by using C_α atom coordinates derived from the crystallographically determined protein structures.

Protein Data Set

The multiscale modeling approach presented here was tested on a set of 10 proteins (Table I), for which conformational changes between at least two unbound and bound structures have been experimentally observed. The unbound protein structure will be referred to as "open" in the following and the ligand-bound protein structure as "closed." The conformational changes represent predominantly hinge or predominantly shear motions, as classified in the Database of Macromolecular Movements.[43] Part of this data set was also used in the study of Tama and Sanejouand.[44] In all cases, all non-protein atoms were stripped from the Protein Data Base (PDB[45]) files. Because an all-atom representation was required for FIRST analysis, hydrogens were then added by using the protonate program of Amber.[46] In histidine side-chains, the molecular environment was visually inspected to decide which ring nitrogens to protonate. Using a default assignment of the hydrogen to the N_ε of histidines instead[47] did not yield qualitatively different results (data not shown). All other ionizable side-chains were modeled according to their standard protonation states at neutral pH. No further optimization of the hydrogen bond network was performed, owing to recent findings that the outcome of FIRST is robust even if the number of hydrogen bonds considered in the molecular framework varies by 5–10%.[41]

Comparison With Experiment

Results of the normal mode analyses are compared with the experimentally observed conformational changes in the direction of the motions and their magnitude. First, the overlap I_i between normal mode $\bar{u}_i$ and the conformational change $\Delta\bar{r} = \bar{r}_o - \bar{r}_c$ is calculated according to[48]

$$I_i = \frac{|\bar{u}_i \cdot \Delta\bar{r}|}{(\bar{u}_i \cdot \bar{u}_i)^{1/2}(\Delta\bar{r} \cdot \Delta\bar{r})^{1/2}}, \quad (2)$$

where $\bar{r}_o$ and $\bar{r}_c$ are vectors of the atomic coordinates of the open and closed structures of the protein, respectively. For this, the closed structure was superimposed onto the open structure on the basis of C_α atom positions. An overlap of one indicates that the directions of both kinds of collective C_α displacements are identical.

To measure the similarity in the relative magnitude of the atomic displacements determined experimentally or by normal mode analysis, the correlation coefficient[44] is calculated.

$$C_i = \frac{\bar{A}_i \cdot \Delta\bar{R}}{(\bar{A}_i \cdot \bar{A}_i)^{1/2}(\Delta\bar{R} \cdot \Delta\bar{R})^{1/2}} \quad (3)$$

$\bar{A}_i$ and $\Delta\bar{R}$ are the vectors of mean centered amplitudes of atomic displacements as determined by mode i or experiment, respectively. A correlation coefficient of one indicates that the relative magnitudes of atomic displacements are identical in both cases.

Finally, a collectivity index as proposed by Bruschweiler[49] is calculated according to

$$\kappa = \frac{1}{N}\exp\left(-\sum_{i=1}^{N}\Delta\bar{r}_i^2 \log\Delta\bar{r}_i^2\right), \quad (4)$$

where $\Delta\bar{r}_i$ is the difference in Cartesian coordinates of atom i due to the experimentally determined conformational change between open and closed forms of the proteins. All $\Delta\bar{r}_i$ have been scaled consistently so that $\sum_{i=1}^{N}\Delta\bar{r}_i^2 = 1$. The index describes the degree of collectivity of a conformational change in that it reflects the number of atoms that are affected during the conformational transition. $\kappa = 1$ indicates a conformational change of maximal collectivity (i.e., all $\Delta\bar{r}_i$ are identical). Conversely, if only one atom is affected by the conformational change, κ reaches the minimal value of $1/N$.

Standard Normal Mode Analysis and Spanning Coefficient

For comparison with the RCNMA results, standard NMA is performed on a subset of six proteins (Table I), by using the Amber 8 suite of programs[46] and the Cornell et al. force field[50] for the atomic description of the proteins. Before NMA, protein structures are minimized in the gas phase by using the conjugate-gradient method with a distance-dependent dielectric of $4r$ (to approximately account for solvation effects, with r being the distance between two atoms) until the root-mean square of the elements of the gradient vector is $<10^{-4}$ kcal mol^{-1} Å^{-1}. The root-mean-square deviation (RMSD) of the backbone

atom positions between the starting and minimized structures amounts to 1.8 Å, averaged over all six cases. Frequencies and normal modes are then computed for these minimized structures.

Standard normal modes are used to evaluate the quality of eigenvectors obtained by our approach. For this, a "spanning coefficient"[23] is computed as the sum of the square of expansion coefficients $c_{ij} = \bar{u}_i \cdot \bar{v}_j$

$$P_j = \sum c_{ij}^2, \quad (5)$$

where the summation runs over all RCNMA eigenvectors $\bar{u}_i$, and $\bar{v}_j$ is the j-th standard normal mode. A spanning coefficient of 1 indicates that $\bar{v}_j$ can be perfectly described in the subspace considered in the RCNMA approach.

RESULTS AND DISCUSSION

Rigid Cluster Decomposition by FIRST Analysis

In the first step of our approach, the protein is decomposed into rigid clusters, which provides the input for the subsequent RCNMA. In previous studies, decomposition schemes were used, which resulted in blocks consisting of residues consecutive in sequence,[22,23] secondary structure elements,[22] clusters of uniform size,[24] or whole protein subunits.[26] These schemes are counterintuitive with respect to the model that forms the basis of the RTB method.[21] Instead, we sought for a decomposition in which structurally rigid protein regions are distinguished from flexible ones between them and which takes into account that nonlocal interactions have an important influence on protein rigidity/flexibility. The molecular framework approach FIRST provides this information about local flexibility characteristics from a single 3D structure in about 1 s of computational time.

The conformational flexibility predicted by FIRST will depend on whether a ligand-free (open) or ligand-bound (closed) form will be analyzed because of the additional constraints provided by the ligand and/or conformational changes of the protein. Although cases of increased flexibility[51–58] or flexibility transferred to other parts of the protein[59–64] have been described, in most of the instances known to date ligand binding reduces the flexibility of the macromolecule.[36] Thus, as we anticipate that underpredicting rigidity (i.e., underestimating the size of rigid clusters) is less detrimental to the subsequent RCNMA than the opposite, we only analyze the open forms of the proteins by FIRST. Along these lines, the hydrogen-bond energy cutoff was set to $E_{cut} = -1.0$ kcal mol^{-1}, which results in on average 5% less constraints in the molecular framework and, hence, a more floppy protein than if E_{cut} is set to -0.6 kcal mol^{-1} as in previous studies.[36,41]

Results of the rigid cluster decomposition by FIRST for the 10 proteins are given in Table II. The number of rigid clusters obtained is related to the number of residues of the protein. Hence, the values indicate the fraction of "particles" that need to be considered in the RCNMA with respect to the number of "particles" if no coarse-graining would have been performed. The decrease is by a factor of 2.3 in CDK and thymidylate synthase and 4.8 in tyrosine phosphatase, with values around 3 for most of the other cases.

TABLE II. Results of the Rigid Cluster Decomposition

Protein	No. of residues	No. of blocks[a]	Size of largest block[b]
Adenylate kinase	214	0.27	32 (0.14)
Alcohol dehydrogenase	374	0.35	134 (0.36)
Aspartate aminotransferase	401	0.37	128 (0.32)
Calmodulin	148	0.35	41 (0.27)
Citrate synthase	860	0.41	198 (0.23)
CD kinase	296	0.44	53 (0.17)
HIV-1 protease	198	0.41	116 (0.58)
LAO binding protein	238	0.37	106 (0.44)
Thymidylate synthase	264	0.44	118 (0.44)
Tyrosine phosphatase	278	0.21	179 (0.64)

[a]Number of rigid clusters with respect to the number of residues of the protein as obtained by FIRST analysis using a hydrogen bond energy cutoff $E_{cut} = -1.0$ kcal mol^{-1}.
[b]Size of the largest rigid cluster as obtained by FIRST analysis using $E_{cut} = -1.0$ kcal mol^{-1}. In parentheses, the size of the largest block with respect to the number of residues of the protein is given.

Thus, performing a rigid cluster decomposition by FIRST overall results in a coarse-graining of the proteins (and, hence, a considerable gain in efficiency regarding the subsequent RCNMA) that is comparable with the one used in RTB analyses, where up to six residues were included into one block.[22] However, in the latter case, all blocks had the same size and were constructed of residues consecutive in sequence, irrespective of whether the parts of the protein are rigid or flexible. In contrast, the FIRST analysis provides a more realistic mixed coarse-graining such that rigid protein regions will be modeled at low resolution in the subsequent RCNMA, whereas flexible regions will be modeled in atomic (i.e., high-resolution) detail. Along these lines, Figure 1 displays the rigid cluster distribution for the 10 proteins in terms of the cumulative summation of relative rigid cluster sizes as a function of the number of blocks per number of residues of a protein. The dotted diagonal line would be obtained with a one-residue-per-block decomposition. In contrast, the curves of the 10 proteins are characterized by a steep increase at the beginning (indicating a decomposition into a few large rigid clusters) but then run parallel to the diagonal (indicating that the remaining blocks are of size one). It is of interest that in tyrosine phosphatase or adenylate kinase, about 80% of all residues are composed of rigid clusters of size >1. Yet, both proteins exhibit rather different characteristics of the decomposition (Fig. 2). In tyrosine phosphatase [Fig. 2(a)], 64% of all residues are part of the largest rigid cluster (Table II), which spreads throughout the protein and comprises all secondary structure elements except helices α3 and α4. Thus, the picture of a dominating rigid core flanked by flexible loops is provided. In contrast, in adenylate kinase [Fig. 2(b)], the largest rigid cluster only contains 14% of all residues (Table II) and comprises the lid domain. All helical regions of the protein form additional isolated rigid clusters as does the central β-sheet. Although similar results can be expected from a decomposition based only on secondary

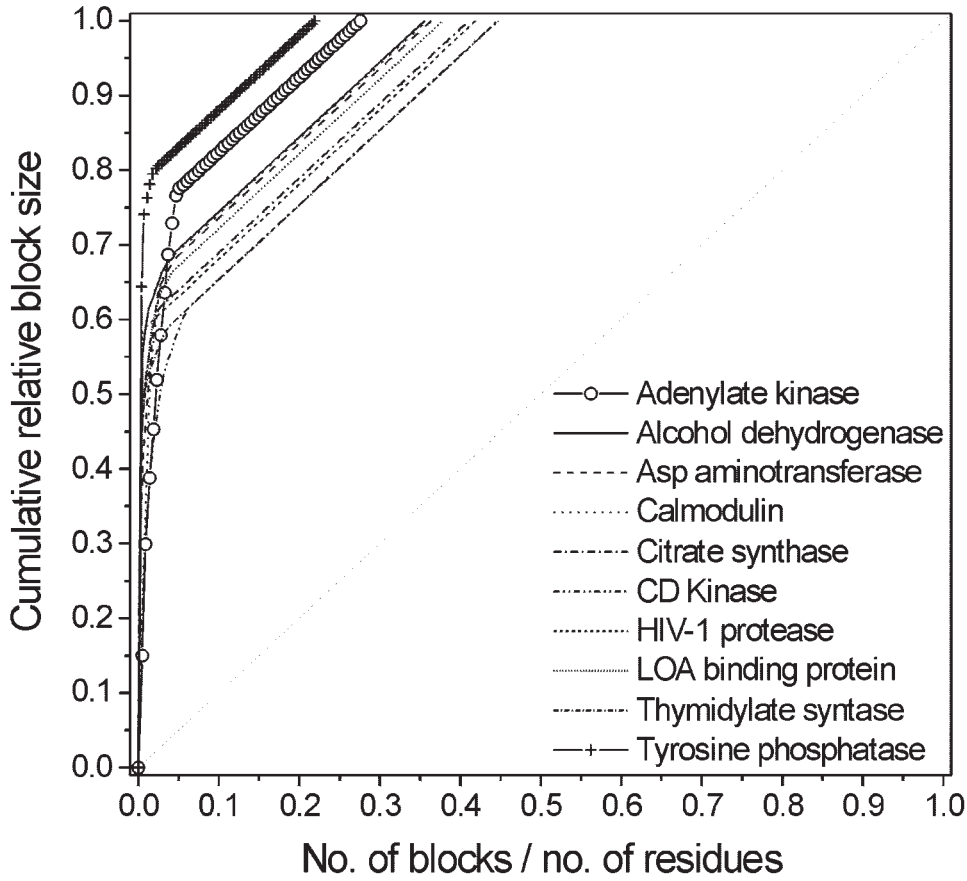

Fig. 1. The cumulative summation of the relative rigid cluster size (i.e., the rigid cluster size for the number of all residues) as a function of the number of rigid clusters related to the number of all residues for the proteins, determined by FIRST analysis using $E_{cut} = -1.0$ kcal mol^{-1}. The dotted line indicates the curve that would result from a decomposition on a one-residue-per-block basis.

structure information, we note that in this case, neither the whole lid domain nor the whole β-sheet would have been considered as a rigid cluster. This finding indicates the importance of taking into account the influence of nonlocal interactions on protein rigidity. Finally, in all other cases, between 60 and 70% of the protein residues are part of rigid clusters of size >1, with relative sizes of the largest blocks in between those found for tyrosine phosphatase and adenylate kinase.

Comparison of RCNMA With Experimentally Determined Conformational Changes

For the 10 proteins, RCNMA is performed by using the rigid cluster decomposition as obtained by FIRST. Although FIRST analysis only provides information about which protein regions are flexible or rigid, information about amplitudes and directions of motions are obtained in this step. Only the ligand-free (open) forms of the proteins are analyzed here, because it was found previously[44] that almost always a better description of the conformational change is given by the mode most involved in this change if this mode is obtained from the open form instead of the closed one. This finding has been explained in that (simplified potential-based) NMA captures properties of the proteins that for the most part are properties of the mass distribution. The open protein forms usually show better separated domains and, hence, more distinguishable congregations of masses. In addition, ANM analysis (to which RCNMA becomes identical if only blocks consisting of one residue are used), RTB analysis with three sequentially consecutive residues forming one block (denoted RTB-3 from now on), and standard NMA are performed.

The computed results are compared with experimentally determined conformational changes of the proteins in terms of the directions and magnitudes of the motions (Fig. 3). Previous studies strongly suggest that NMA based on elastic network model representations[15,42,44] and/or coarse-grained models[16,22,23] of proteins can reliably describe biologically relevant motions by considering only a small subset of low-frequency normal modes. It is not clear, however, whether this still holds if up to 64% of the protein is modeled as one rigid body (Table 2).

Direction of motions

In Table III, the mode most involved in the conformational change as judged by the overlap value (Eq. 2)[48] is given. Overall, for 7 of 10 cases, overlap values ≥ 0.6 are found by using RCNMA. Comparing these results with

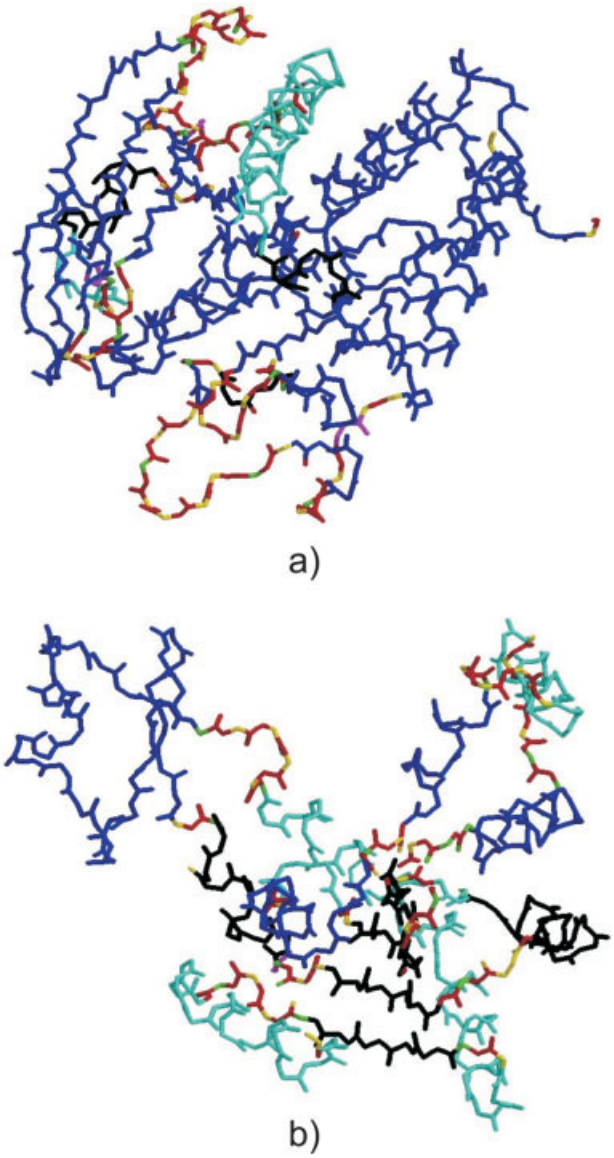

Fig. 2. Rigid cluster decomposition of tyrosine phosphatase (**a:** PDB code: 1ypt) and adenylate kinase (**b:** PDB code: 4ake). Rigid clusters are colored in blue, cyan, black, yellow, red, and green. In the adenylate kinase case, multiple medium-size rigid clusters are found. In contrast, in the tyrosine phosphatase case, one rigid cluster that comprises 64% of all protein atoms dominates the decomposition.

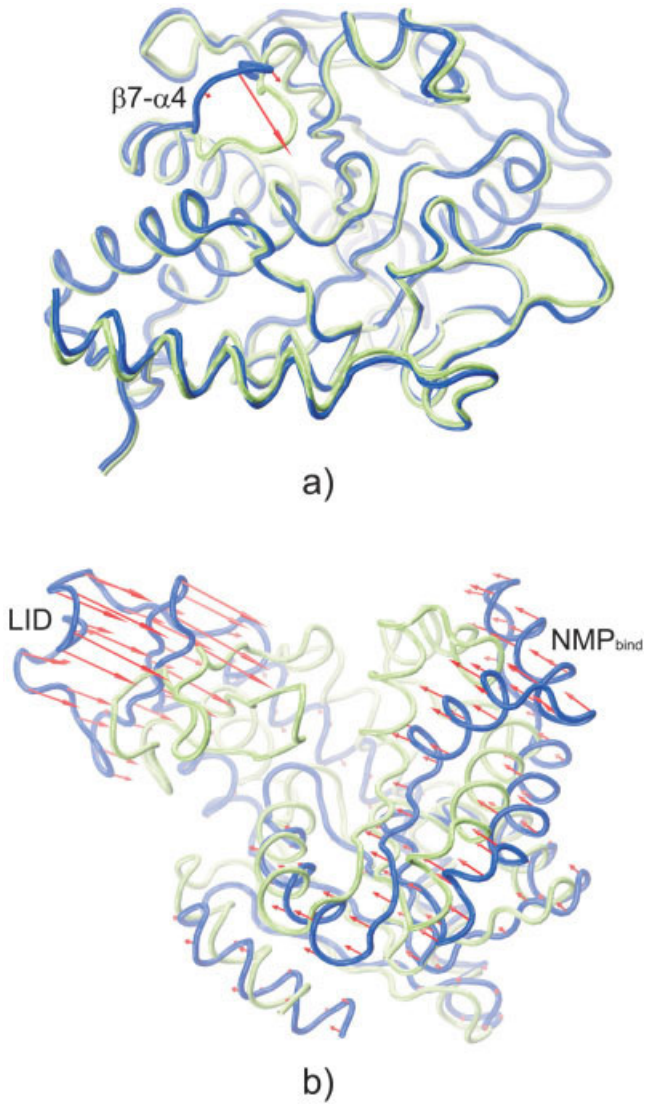

Fig. 3. Superimposition of open (blue) and closed (green) conformations of tyrosine phosphatase (**a**) and adenylate kinase (**b**). In addition, the amplitudes and directions of motions as predicted by the modes most involved in the conformational changes, respectively, are depicted as red arrows. In both cases, the amplitudes of the motions were scaled for best graphical representation.

those obtained by ANM analysis, only in the citrate synthase case the latter performs slightly better. In contrast, RCNMA better describes the direction of the conformational change in CDK, HIV-1 protease, and tyrosine phosphatase and yields comparable results in all other cases. RCNMA results are also compared with RTB-3 analyses. RTB-3 uses the same level of coarse-graining that has been found on average also by a rigid cluster decomposition with FIRST (Table II): however, rigid and flexible parts of the protein are not explicitly distinguished for block formation. Again, better overlap values are found by RTB-3 in the citrate synthase case. However, RCNMA outperforms RTB-3 in alcohol dehydrogenase, CDK, HIV-1 protease, and tyrosine phosphatase.

By the two comparisons, the influence of a mixed coarse-graining (RCNMA) with regard to either no (ANM) or a uniform (RTB-3) coarse-graining is probed. As indicated by the results, distinguishing between flexible and rigid protein regions may in general be advantageous, although we note that this finding is still based on a limited number of test cases. It is further enlightening to analyze the observed differences for the type of motion found on going from the open to the closed conformation. As such, cases that are predominantly characterized by hinge bending or shear motions are predicted equally well by all approaches. Thereby, better agreement between predicted and experimentally determined directions of

TABLE III. Comparison of NMA Results With Experimentally Determined Conformational Changes in Directions of Motions

Protein	RCNMA		ANM[a]	RTB-3[b]	Standard NMA
	Overlap[c]	Flexible/rigid[d]	overlap[c]	overlap[c]	overlap[c]
Adenylate kinase	0.81 (1)	0.65/0.85	0.81 (1)	0.82 (1)	0.86 (1)
Alcohol dehydrogenase	0.73 (3)	0.67/0.80	0.73 (3)	0.61 (2)	-[e]
Aspartate aminotransferase	0.49 (6)	0.32/0.57	0.51 (6)	0.52 (7)	-
Calmodulin	0.84 (2)	0.83/0.84	0.84 (2)	0.85 (2)	0.57 (4)
Citrate synthase	0.80 (3)	0.77/0.82	0.88 (3)	0.86 (3)	-
CD kinase	0.39 (2)	0.46/0.31	0.26 (2)	0.28 (1)	-
HIV-1 protease	0.61 (2)	0.70/0.00	0.53 (2)	0.55 (5)	0.62 (2)
LAO binding protein	0.84 (1)	0.86/0.83	0.85 (1)	0.86 (1)	0.66 (1)
Thymidylate synthase	0.48 (2)	0.52/0.13	0.46 (1)	0.44 (4)	0.25 (25)
Tyrosine phosphatase	0.60 (1)	0.67/0.20	0.38 (1)	0.37 (1)	0.44 (2)

[a]Anisotropic network model analysis[42] (i.e., no block formation was considered).
[b]Rotations-translations of blocks analysis[22] with three sequentially consecutive residues forming one block.
[c]Overlap (Eq. 2) of the mode most involved in the conformational change, considering all C_α atoms of the protein. In parentheses, the number of this mode is given.
[d]Overlap (Eq. 2) of the mode most involved in the conformational change. For the first value, only C_α atoms in flexible protein regions were considered; for the second value, only C_α atoms in rigid clusters of the protein were considered.
[e]Because of the size of the systems, no standard NMA could be performed.

movements are found for hinge motions than for shear motions. This may be related to the issue of mass distribution discussed above, because in the case of shear motions, domains may be less well separated in the open form than in the hinge-bending motions.

In turn, all three cases where RCNMA performs superior to ANM and RTB-3 (CDK, HIV-1 protease, and tyrosine phosphatase) are predominantly characterized by loop or fragment motions. It may not be surprising that these types of motion are less well predicted if the flexible loops are constructed from rigid blocks as in the RTB-3 case. It is more difficult to conceive, however, why better results are obtained from RCNMA in those cases than if no coarse-graining is applied. One may speculate that using a coarse-grained protein representation in the RTB step leads to a less rugged potential energy surface and, hence, facilitates the modeling of these motions when using a harmonic approximation. In any case, being able to model these types of motions with reasonable computational efficiency provides an interesting way of incorporating protein flexibility into docking algorithms, particularly, because ligand binding sites are often formed by loop regions.

Finally, larger overlap values are found by RCNMA in four out of six cases for which also standard NMA calculations were performed; in the remaining two cases, RCNMA and standard NMA perform comparably.

To test how well the experimentally determined conformational change can be described collectively in the subspace spanned by the set of eigenvectors of the RCNMA, the spanning coefficient (Eq. 5) was determined (Table IV). For 9 of 10 cases, values are close to or larger than 0.8, indicating that the conformational change is well described by the RCNMA eigenvectors.

In Table III, overlap values are also presented, which were calculated separately for rigid protein regions (i.e., rigid clusters with size > 1) and flexible ones. In all but four cases, those obtained for rigid clusters are even larger than the values of flexible parts. This indicates that directions of motions of atoms within rigid clusters are well represented by a single low-frequency mode even if these motions arise from rigid body movements of the whole cluster. The finding can be attributed to the fact that biologically relevant motions are typically delocalized and do not require gross changes in the internal structure of the rigid clusters. In contrast, in HIV-1 protease, thymidylate synthase, and tyrosine phosphatase, the computed and experimentally observed directions of motions of rigid protein regions show little or no overlap (in contrast to flexible regions where a considerable overlap is observed). It should be noted, however, that for these proteins, the experimentally determined movements of rigid parts are only of the same magnitude as the uncertainty in the atomic positions of the experimental structures. Thus, "noisy" atomic displacements result that cannot be represented by a mode describing collective motions.

Collectivity of motion

Except for aspartate aminotransferase, the mode most involved in the conformational change is one of the five with lowest frequencies, as found by RCNMA, ANM, and RTB-3 analyses. This is in line with previous reports[12] that biologically relevant motions can be reliably described by considering one or only a small subset of low-frequency normal modes. These motions are also typically delocalized and involve mainly collective motions of residues throughout the protein [Fig. 3(b)]. In that respect, it is encouraging that even for motions that are rather localized as indicated by values of the collectivity index (Eq. 4; Table IV) below 0.16, overlap values close to or above 0.5 have been found by RCNMA. This suggests that, for example, the direction of the motion of the β7-α4 loop (residues 350–360) in tyrosine phosphatase is adequately described by a single mode [Fig. 3(a)]. It should be noted, however, that this only holds as long as the corresponding motion is linear but not if a nonlinear path of atomic

TABLE IV. Spanning Coefficients, Collectivity Indices, and Correlation Coefficients for RCNMA Results

Protein	Spanning coefficient[a]	Collectivity[b]	Correlation[c]
Adenylate kinase	0.98	0.48	0.84
Alcohol dehydrogenase	0.91	0.48	0.66
Aspartate aminotransferase	0.97	0.43	0.68
Calmodulin	0.99	0.78	0.75
Citrate synthase	0.91	0.14	0.86
CD kinase	0.74	0.31	0.60
HIV-1 protease	0.79	0.74	0.84
LAO binding protein	0.91	0.74	0.65
Thymidylate synthase	0.91	0.16	0.66
Tyrosine phosphatase	0.81	0.12	0.72

[a]Spanning coefficient (Eq. 5) of the conformational change in the subspace spanned by the eigenvectors of the RCNMA.
[b]Collectivity index (Eq. 4) of the mode most involved in the conformational change (see Table III).
[c]Correlation coefficient of the magnitude of the C_α displacements (Eq. 3), using the mode most involved in the conformational change (see Table III).

motions can be expected for displacements, leading to changes in the directions of the motions during the conformational transition. Again, being able to distinguish between rigid protein regions and flexible links between them as given by our approach clearly provides an advantage over previous RTB analyses.[22–24,26] This is particularly true if rather localized motions are to be modeled.

Relative magnitude of motions

Finally, correlation coefficients (Eq. 3) are given in Table IV, which measure the similarity in the magnitude of the atomic displacements determined experimentally or given by the mode most involved in the conformational change. In all cases, values > 0.6 are found, suggesting that some information about the amplitudes of atomic displacements is provided by the mode even in those cases in which the directions of the motions are less well described, as has been stated previously.[44] Along these lines, Figure 4 shows displacements of C_α atoms as a function of the residue number, obtained from experimentally determined conformational changes or by displacing atoms along the direction of the normal mode most involved in this change. Adenylate kinase and tyrosine phosphatase were chosen exemplarily here because they represent, respectively, delocalized and localized conformational changes. Convincingly, in both cases, theoretical and experimental curves are strikingly similar for both residues located in flexible regions and rigid clusters. The latter indicates that magnitudes of atomic motions in rigid protein parts are well described by the rigid body motions of these clusters, as has been found also for the directions of these motions above.

Comparison of RCNMA With Standard NMA

The RCNMA approach introduced here is an approximation to standard NMA calculations, which use an atomic force-field representation of the macromolecule. On this account, RCNMA results will now be compared with those from standard NMA.

Comparing experimentally determined conformational changes and RCNMA results in terms of the directions of observed and calculated movements (see above) already suggests that the atomic displacements involved in the biologically relevant motions, in general, can be well represented in the subspace considered in the RCNMA approach. Similarly, the low-frequency subspace spanned by standard NMA normal vectors is fairly described by eigenvectors obtained from the RCNMA approach, as indicated in Figure 5. Here, spanning coefficients (Eq. 5) calculated for each standard normal mode with a frequency $< 20\ \mathrm{cm}^{-1}$ are depicted exemplarily for the three proteins adenylate kinase, calmodulin, and thymidylate synthase. These proteins represent different degrees of coarse-graining, as indicated by relative numbers of blocks after FIRST analysis of 0.27, 0.35, and 0.44 (Table II), respectively. For most of the vectors, spanning coefficients > 0.6 are found. Only in those cases where the largest rigid cluster contains $> 50\%$ of all protein atoms [HIV-1 protease, tyrosine phosphatase (Table II)], standard NMA vectors can be less well represented in the subspace spanned by RCNMA vectors. This is indicated by spanning coefficients between 0.4 and 0.6 (data not shown). Although similar findings have already been reported with coarse-graining levels of up to three sequentially consecutive residues,[22,23] this result is particularly encouraging in our case. It shows that low-frequency vibrational motions can be fairly described collectively even if a considerably more coarse-grained protein representation based on a rigid cluster decomposition by FIRST is applied.

Figure 6 depicts approximate frequencies calculated by RCNMA as a function of frequencies obtained by standard NMA. In all cases, RCNMA frequencies are larger than their corresponding frequencies from standard NMA. This indicates that modeling the proteins as a collection of rigid bodies connected by flexible links (and using an elastic network representation with a phenomenological force constant $\gamma = 1\ \mathrm{kcal\ mol}^{-1}\ \mathrm{Å}^{-2}$ (Eq. 1) in that case) results in steeper wells of the potential energy surface and, hence,

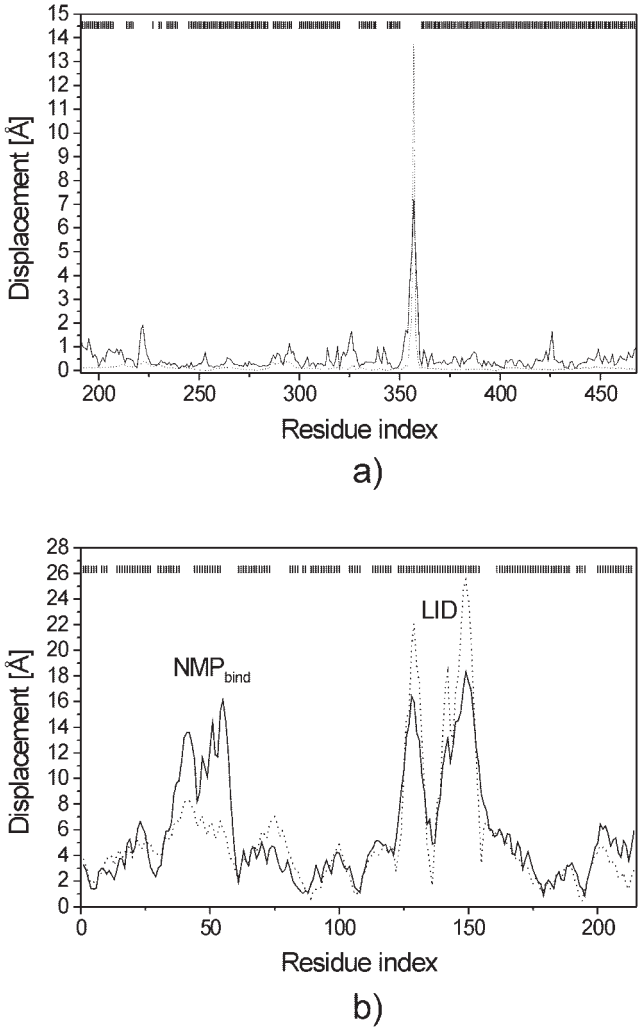

Fig. 4. Displacements of C_α atoms as a function of the residue number for tyrosine phosphatase (**a**) and adenylate kinase (**b**). The displacements were obtained either from experimentally determined conformational changes between the open and closed forms of the proteins (straight line) or by displacing atoms along the direction of the normal mode most involved in this change (dotted line). The theoretical curve was scaled with respect to the experimental one so that the area under the square of the curves is identical.[14] Residues located in rigid clusters, as determined by FIRST analysis, are marked by vertical dashes at the top of the figures.

stiffer proteins. Similar findings have also been reported by studies of Tama et al.[22] and Li et al.,[23] although it should be noted that in these cases, the protein was modeled by an atomic force-field representation in both the rigid block and standard NMA approach. Thus, coarse-graining alone already led to stiffer protein potential energy surfaces. For modes with frequencies < 20 cm^{-1}, which are most interesting for describing biologically relevant motions, RCNMA frequencies correlate very well linearly with those from standard NMA (as indicated by r^2 values > 0.99 obtained for all six proteins). The scaling factors are in the range of 2.6–3.5 in these cases. Furthermore, they neither seem to depend on the structure or sequence of the proteins, as also found in previous work,[22,23] nor on the size of the largest rigid cluster identified by FIRST analysis (data not shown). However, an inverse correlation ($r^2 = 0.75$) for the relative number of rigid clusters is observed (Fig. 7), indicating that higher scaling factors will result if the protein is coarse-grained into less, but larger, rigid clusters.

It should be noted that these findings provide an attractive route to estimate approximate frequencies of low-frequency normal modes from RCNMA via a scaling factor determined from Figure 7 for a given relative number of rigid clusters of the considered protein. These approximate frequencies may possibly be used to estimate (free) energies of conformational protein changes along normal mode directions, as recommended for flexible docking approaches[65,66] or recently applied in a model to describe conformational changes coupled to ligand binding.[67] However, one should not expect a quantitative agreement. This is particularly true for properties that depend also on nonnegligible contributions from high-frequency modes, such as vibrational entropies.[23]

CONCLUSION

NMA-based approaches have received renewed interest in the biomolecular field over the past years because they allow characterization of structural and dynamical properties and describe and predict even large-scale motions of macromolecules. The trend has been particularly driven by the development of physically motivated coarse-grained models of macromolecules.[13–15,17,20,21,24] These models still capture the essential molecular properties that determine low-frequency collective movements of the systems.

In this spirit, we have introduced a two-step approach for multiscale modeling of macromolecular conformational changes in this study. Our approach is based on recent developments in rigidity theory,[34,36] elastic network theory,[14,20] and work by Sanejouand et al.[21,22] First, static properties of the protein are determined by decomposing the molecule into rigid clusters by using the graph-theoretical approach FIRST. For this step, the molecule is represented on an atomic level. The rigid clusters are not limited to consist of residues adjacent in sequence or secondary structure elements as in previous work.[22,23,26] Instead, they may span considerable parts of the macromolecule, taking into account that nonlocal interactions influence the rigidity of a protein. Compared to decompositions that result in rigid clusters of uniform size,[24] this step also identifies flexible links between rigid clusters. Hence, these links may be modeled at a more fine-grained level subsequently. This is particularly advantageous if the movement is dominated by loop or fragment motions. Second, dynamical properties of the macromolecule are revealed by an RTB analysis, in which only rigid-body motions are allowed for rigid clusters, whereas links between them are treated as fully flexible.

For efficiency of our method, coarse-graining the protein by rigid cluster decomposition leads to a decrease of the dimensionality of the Hessian that needs to be diagonalized in the second step up to a factor of 5. In turn, this results in a remarkable reduction of memory requirements and computational times by factors of 9 and 27 on average

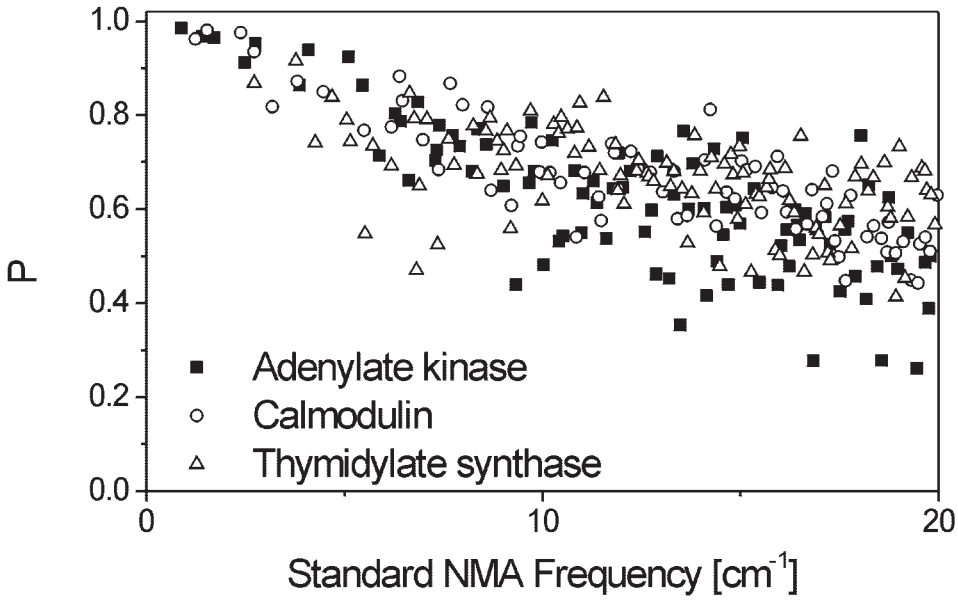

Fig. 5. Spanning coefficients (Eq. 5) as a function of the frequencies of standard NMA vectors for the proteins adenylate kinase (■), calmodulin (○), and thymidylate synthase (△).

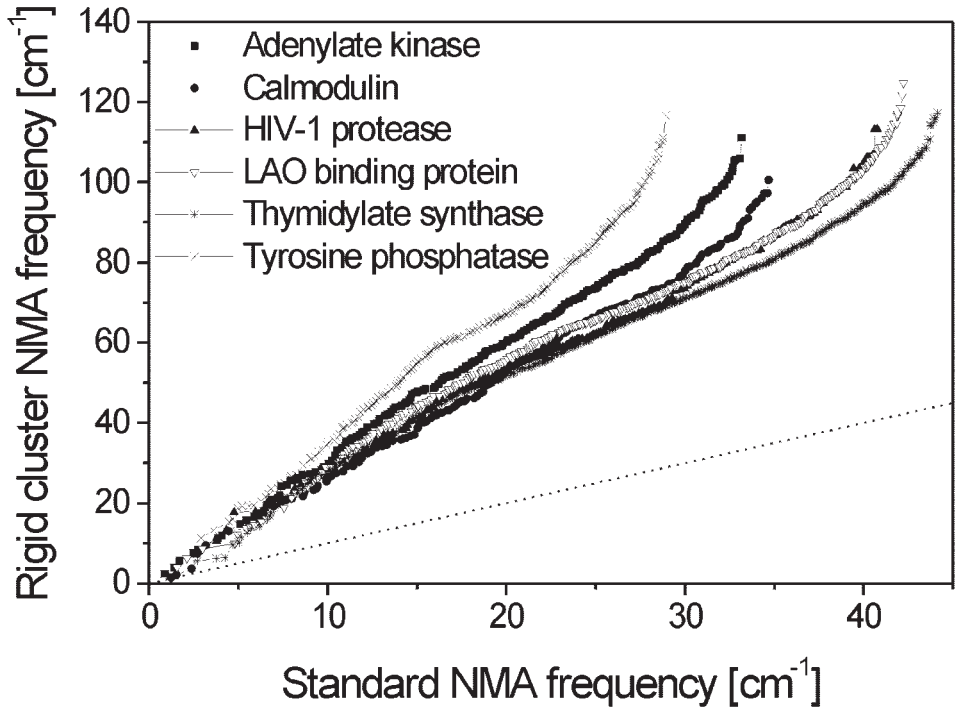

Fig. 6. Frequencies calculated with the RCNMA approach as a function of frequencies calculated with standard NMA for six proteins. The phenomenological force constant γ (Eq. 1) was set to values of 1 kcal mol^{-1} Å^{-2} in all cases. The dashed line indicates an ideal correlation.

and up to 25 and 125, respectively. Similar reductions of the dimensionality have been obtained with the RTB approach considering sequentially consecutive residues as blocks.[22,68] In the current approach, the protein is represented as an elastic network based on C_α atoms in the RTB step. However, an atomic force-field-based representation of the macromolecule can also be included in the RTB step. In this case, the reductions of computational requirements are similar to the ones found in the this study (data not shown).

For accuracy, our approach was tested on a data set of 10 proteins that show conformational changes on ligand binding. It is encouraging that in directions and magnitudes of the motions, movements predicted by the RCNMA

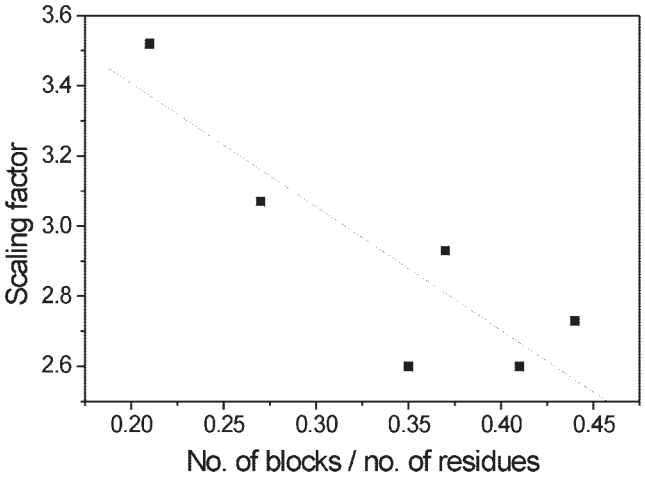

Fig. 7. Scaling factor obtained by relating approximate RCNMA frequencies versus frequencies from standard NMA as a function of the number of rigid clusters with respect to the number of all residues for six proteins. The squared correlation coefficient of the inverse correlation is 0.75.

agree well with experimentally observed ones. The results of our method are in general comparable with when no or a uniform coarse-graining is applied and become superior if the movement is dominated by loop or fragment motions. Furthermore, very good agreement between predicted and observed motions is also found for the rigid parts of the macromolecules, in which atomic movements originate from rigid body motions of the whole cluster. Finally, a linear correlation has been found when comparing frequencies determined by RCNMA with those from standard NMA (up to 20 cm^{-1}), which allows to estimate frequencies of biologically relevant motions in a semiquantitative way using our approach.

These findings lead to the following conclusions. 1) Even if considerable parts of the proteins are considered to be rigid (in the extreme cases, the largest rigid cluster consists of more than 50% of all residues), motional characteristics of the overall molecule are correctly predicted. This indicates the quality of the decomposition of the protein into rigid and flexible regions by FIRST. 2) Obtaining in general comparable and, in the case of loop or fragment motions, even better results with a coarse-grained protein representation compared with no or a uniform coarse-graining provides an indication that explicitly distinguishing between flexible and rigid regions is advantageous when using an elastic network model representation of the protein. Two reasons may be given for this. On the one hand, in the FIRST analysis, an all-atom representation of the protein together with a more elaborate representation of nonbonded interactions is used. Thus, one may anticipate that flexible/rigid regions within the protein are better characterized as when only interactions of equal strength between C_α atoms are considered. On the other hand, using a more coarse-grained protein representation in the RTB step leads to a less rugged potential energy surface. This facilitates the modeling of large-scale motions when using a harmonic approximation. 3) The fact that motions of atoms in rigid clusters are well predicted by our approach confirms that collective macromolecular movements may not require gross changes in the internal structure of these clusters. This points to the need to consider mobile protein parts (which may well move as rigid clusters in total) in addition to flexible ones (where changes in the internal structure of a collection of atoms may occur) when modeling correlated motions. In that respect, RCNMA differs from a recently published approach.[38–40] Here, protein flexibility is also analyzed by FIRST initially. Then, however, only flexible regions are explored by random-walk sampling of rotatable bonds for generating available conformations.

Considering the success of our approach in describing and predicting biologically relevant motions efficiently and accurately, we hope that it will aid in the understanding of the interplay between structure, dynamics, and function of complex molecular systems. Further potential applications of our method include atomic-level molecular dynamics simulations with amplified collective motions[69] and modeling of induced-fit effects in flexible docking approaches.

ACKNOWLEDGMENT

A. Ahmed gratefully acknowledges a fellowship from the "Stiftung zur Förderung der internationalen wissenschaftlichen Beziehungen der J. W. Goethe-Universität", Germany. We are grateful to Qiang Cui (University of Wisconsin) for helpful discussions and to Leslie A. Kuhn (Michigan State University) for granting a FIRST license to us.

APPENDIX
Rotations-Translations of Blocks Approach[21,22]

Given a coarse-grained representation of a protein in terms of n rigid clusters (blocks), the dimensionality of the Hessian matrix H can be reduced from $3N$ (with N being the number of atoms or "particles" in the case of no rigid cluster decomposition) to $6n$ by projecting H into the subspace spanned by translation/rotation basis vectors of the blocks according to

$$H_{sub} = P^t H P \tag{6}$$

Here, P is an orthogonal $3N \times 6n$ projection matrix of the infinitesimal translation/rotation eigenvectors of each block[23,70]

$$P^{\mu}_{J j_\nu} = \sqrt{m_j / M_J}\,\delta_{\mu\nu} \quad \mu = 1,2,3$$

$$P^{\mu}_{J j_\nu} = \sum_{\alpha\beta} (I_J)^{-1/2}_{(\mu-3),\alpha} \sqrt{m_j}(r_j - r^0_J)_\beta \varepsilon_{\alpha\beta\nu} \quad \mu = 4,5,6 \tag{7}$$

J and j label blocks and atoms, respectively, and μ the translation (μ = 1,2,3) and rotation (μ = 4,5,6) of each block. m_j and r_j are the mass and Cartesian coordinate of atom j; M_J, I_J, and $r_J{}^0$ are the total mass, moment of inertia, and center of mass of block J. δ is the Kronecker delta and ε is the permutation symbol. $\alpha, \beta, \nu = \{1, 2, 3\}$.

Diagonalizing H_{sub} yields approximate low-frequency normal modes U_{sub} and eigenvalues Λ.

$$H_{sub}U_{sub} = U_{sub}\Lambda \quad (8)$$

Finally, atomic displacements can be obtained by expanding back the eigenvectors U_{sub} from the subspace spanned by translation/rotation basis vectors of the blocks to the Cartesian space (U).

$$U = PU_{sub} \quad (9)$$

Although the projection matrix P has a nominal dimension of $3N \times 6n$, it is a very sparse matrix, consisting only of $3m_J \times 6$ nonzero elements for block J with m_J atoms. The matrix H is also sparse because interactions between "particles" are only taken into account up to a cutoff distance in the elastic network model representation of the macromolecules applied in this study (see below). Hence, sparse-matrix representations were used in our implementation of the RTB method. We note that a more memory-efficient implementation of the RTB approach has been reported recently,[23] in which only those elements of the matrix H are computed on the fly, which are required for a given block pair. With this, however, a significant speedup can be expected for only those cases in which the matrix H cannot be stored entirely in the computer memory. This does not apply for the systems investigated in this study.

REFERENCES

1. Frank J, Agrawal RK. A ratchet-like inter-subunit reorganization of the ribosome during translocation. Nature 2000;406:318–322.
2. Sabbert D, Engelbrecht S, Junge W. Intersubunit rotation in active F-ATPase. Nature 1996;381:623–625.
3. Xu Z, Horwich AL, Sigler PB. The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. Nature 1997;388:741–750.
4. Allison SL, Schalich J, Stiasny K, Mandl CW, Kunz C, Heinz FX. Oligomeric rearrangement of tick-borne encephalitis virus envelope proteins induced by an acidic pH. J Virol 1995;69:695–700.
5. Goh C-S, Milburn D, Gerstein M. Conformational changes associated with protein-protein interactions. Curr Opin Struct Biol 2004;14:104–109.
6. Teague SJ. Implications of protein flexibility for drug discovery. Nat Rev Drug Discov 2003;2:527–541.
7. Carlson HA. Protein flexibility and drug design: How to hit a moving target. Curr Opin Chem Biol 2002;6:447–452.
8. Case DA. Normal mode analysis of protein dynamics. Curr Opin Struct Biol 1994;4:285–290.
9. Ma J. Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. Structure 2005;13: 373–380.
10. Go N, Noguti T, Nishikawa T. Dynamics of a small globular protein in terms of low-frequency vibrational modes. Proc Natl Acad Sci U S A 1983;80:3696–3700.
11. Brooks B, Karplus M. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. Proc Natl Acad Sci U S A 1983;80:6571–6575.
12. Zheng W, Doniach S. A comparative study of motor-protein motions by using a simple elastic-network model. Proc Natl Acad Sci U S A 2003;100:13253–13258.
13. Tozzini V. Coarse-grained models for proteins. Curr Opin Struct Biol 2005;15:144–150.
14. Bahar I, Atilgan AR, Erman B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Fold Design 1997;2:173–181.
15. Hinsen K. Analysis of domain motions by approximate normal mode calculations. Proteins 1998;33:417–429.
16. Doruker P, Jernigan RL, Bahar I. Dynamics of large proteins through hierarchical levels of coarse-grained structures. J Comp Chem 2002;23:119–127.
17. Doruker P, Jernigan RL. Functional motions can be extracted from on-lattice construction of protein structures. Proteins 2003; 53:174–181.
18. Tama F, Wriggers W, Brooks CL III. Exploring global distortions of biological macromolecules and assemblies from low-resolution structural information and elastic network theory. J Mol Biol 2002;321:297–305.
19. Ming D, Kong Y, Lambert MA, Huang Z, Ma J. How to describe protein motion without amino acid sequence and atomic coordinates. Proc Natl Acad Sci U S A 2002;99:8620–8625.
20. Tirion MM. Large amplitude elastic motions in proteins from single-parameter atomic analysis. Phys Rev Lett 1996;77:1905–1908.
21. Durand P, Trinquier G, Sanejouand Y-H. A new approach for determining low-frequency normal modes in macromolecules. Biopolymers 1994;34:759–771.
22. Tama F, Gadea FX, Marques O, Sanejouand Y-H. Building-Block Approach for determining low-frequency normal modes of macromolecules. Proteins 2000;41:1–7.
23. Li G, Cui Q. A coarse-grained normal mode approach for macromolecules: an efficient implementation and application to Ca^{2+} -ATPase. Biophys J 2002;83:2457–2474.
24. Schuyler AD, Chirikjian GS. Normal mode analysis of proteins: a comparison of rigid cluster modes with C(alpha) coarse graining. J Mol Graph Model 2004;22:183–193.
25. Li G, Cui Q. Analysis of functional motions in Brownian molecular machines with an efficient block normal mode approach: myosin-II and Ca2+ -ATPase. Biophys J 2004;86:743–763.
26. Tama F, Brooks CL III. The mechanism and pathway of pH induced swelling in cowpea chlorotic mottle virus. J Mol Biol 2002;318:733–747.
27. Graver J, Servatius B, Servatius H. Combinatorial rigidity (graduate studies in mathematics). Providence, RI: American Mathematical Society; 1993.
28. Maxwell JC. On the calculation of the equilibrium and stiffness of frames. Philos Mag 1864;27:294–299.
29. Thorpe MF, Jacobs DJ, Chubynsky NV, Rader AJ. Generic rigidity of network glasses. In: Thorpe MF, Duxbury PM, editors. Rigidity theory and applications. New York: Kluwer Academic/Plenum Publishers; 1999. p 239–277.
30. Tay T-S, Whiteley W. Recent advances in generic rigidity of structures. Struct Topol 1985;9:31–38.
31. Laman G. On graphs and rigidity of plane skeletal structures. J Eng Math 1970;4:331– 340.
32. Jacobs DJ, Kuhn LA, Thorpe MF. Flexible and rigid regions in proteins. In: Thorpe MF, Duxbury PM, editors. Rigidity theory and applications. New York: Kluwer Academic/Plenum Publishers; 1999. p 357–384.
33. Rader AJ, Hespenheide BM, Kuhn LA, Thorpe MF. Protein unfolding: rigidity lost. Proc Natl Acad Sci USA 2002;99:3540–3545.
34. Jacobs DJ, Thorpe MF. Generic rigidity percolation: the pebble game. Phys Rev Lett 1995;75:4051–4054.
35. Jacobs DJ, Thorpe MF. Generic rigidity percolation in two dimensions. Phys Rev E 1996;53:3683–3693.
36. Jacobs DJ, Rader AJ, Kuhn LA, Thorpe MF. Protein flexibility predictions using graph theory. Proteins 2001;44:150–165.
37. Hespenheide BM, Rader AJ, Thorpe MF, Kuhn LA. Identifying protein folding cores from the evolution of flexible regions during unfolding. J Mol Graph Model 2002;21:195–207.
38. Lei M, Zavodszky MI, Kuhn LA, Thorpe MF. Sampling protein conformations and pathways. J Comput Chem 2004;25:1133–1148.
39. Thorpe MF, Lei M, Rader AJ, Jacobs DJ, Kuhn LA. Protein flexibility and dynamics using constraint theory. J Mol Graph Model 2001;19:60–69.
40. Zavodszky MI, Lei M, Thorpe MF, Day AR, Kuhn LA. Modeling correlated main-chain motions in proteins for flexible molecular recognition. Proteins 2004;57:243–261.
41. Gohlke H, Kuhn LA, Case DA. Change in protein flexibility upon complex formation: analysis of ras-raf using molecular dynamics and a molecular framework approach. Proteins 2004;56:322–337.
42. 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.
43. Echols N, Milburn D, Gerstein M. MolMovDB: analysis and visualization of conformational change and structural flexibility. Nucleic Acids Res 2003;31:478–482.

44. Tama F, Sanejouand YH. Conformational change of proteins arising from normal mode calculations. Protein Eng 2001;14:1–6.
45. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucleic Acids Res 2000;28:235–242.
46. Case DA, Darden TA, Cheatham TE III, Simmerling CL, Wang J, Duke RE, Luo R, Merz KB, Wang B, Pearlman DA, Crowley M, Brozell S, Tsui V, Gohlke H, Mongan J, Hornak V, Cui G, Beroza P, Schafmeister C, Caldwell JW, Ross WS, Kollman PA. Amber 8.0. La Jolla, CA: The Scripps Research Institute; 2004.
47. Kyte J. Structure in protein chemistry. New York: Garland Science; 1995.
48. Marques O, Sanejouand YH. Hinge-bending motion in citrate synthase arising from normal mode calculations. Proteins 1995;23: 557–560.
49. Bruschweiler R. Collective protein dynamics and nuclear-spin relaxation. J Chem Phys 1995;102:3396–3403.
50. Cornell WD, Cieplak CI, Bayly IR, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 1995;117:5179–5197.
51. Yu L, Zhu CX, Tse-Dinh YC, Fesik SW. Backbone dynamics of the C-terminal domain of Escherichia coli topoisomerase I in the absence and presence of single-stranded DNA. Biochemistry 1996;35:9661–9666.
52. Olejniczak ET, Zhou MM, Fesik SW. Changes in the NMR-derived motional parameters of the insulin receptor substrate 1 phosphotyrosine binding domain upon binding to an interleukin 4 receptor phosphopeptide. Biochemistry 1997;36:4118–4124.
53. Zidek L, Novotny MV, Stone MJ. Increased protein backbone conformational entropy upon hydrophobic ligand binding. Nat Struct Biol 1999;6:1118–1121.
54. Loh AP, Pawley N, Nicholson LK, Oswald RE. An increase in side chain entropy facilitates effector binding: NMR characterization of the side chain methyl group dynamics in Cdc42Hs. Biochemistry 2001;40:4590–4600.
55. Arumugam S, Gao G, Patton BL, Semenchenko V, Brew K, Van Doren SR. Increased backbone mobility in β-barrel enhances entropy gain driving binding of N-TIMP-1 to MMP-3. J. Mol. Biol. 2003;327:719–734.
56. Forman-Kay JD. The "dynamics" in the thermodynamics of binding. Nat Struct Biol 1999;6:1086–1087.
57. Kay LE, Muhandiram DR, Wolf G, Shoelson SE, Forman-Kay JD. Correlation between binding and dynamics at SH2 domain interfaces. Nat Struct Biol 1998;5:156.
58. Kay LE, Muhandiram DR, Farrow NA, Aubin Y, Forman-Kay JD. Correlation between dynamics and high affinity binding in an SH2 domain interaction. Biochemistry 1996;35:361–368.
59. Stivers JT, Abeygunawardana C, Mildvan AS, Whitman CP. ^{15}N NMR relaxation studies of free and inhibitor-bound 4-oxalocrotonate tautomerase: backbone dynamics and entropy changes of an enzyme upon inhibitor binding. Biochemistry 1996;35:16036–16047.
60. Hodsdon ME, Cistola DP. Ligand binding alters the backbone mobility of intestinal fatty acid-binding protein as monitored by 15N NMR relaxation and 1H exchange. Biochemistry 1997;36: 2278–2290.
61. Lee AL, Kinnear SA, Wand AJ. Redistribution and loss of side chain entropy upon formation of a calmodulin-peptide complex. Nat Struct Biol 2000;7:72–77.
62. Canino LS, Shen T, McCammon JA. Changes in flexibility upon binding: application of the self-consistent pair contact probability method to protein-protein interactions. J Chem Phys 2002;117: 9927–9933.
63. Tsai C-J, Kumar S, Ma B, Nussinov R. Folding funnels, binding funnels, and protein function. Protein Sci 1999;8:1181–1190.
64. Tsai C-J, Ma B, Nussinov R. Folding and binding cascades: shifts in energy landscapes. Proc Natl Acad Sci USA 1999;96:9970–9972.
65. Wei BQ, Weaver LH, Ferrari AM, Matthews BW, Shoichet BK. Testing a flexible-receptor docking algorithm in a model binding site. J Mol Biol 2004;337:1161–1182.
66. Zacharias M. Rapid protein-ligand docking using soft modes from molecular dynamics simulations to account for protein deformability: binding of FK506 to FKBP. Proteins 2004;54:759–767.
67. 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.
68. Delarue M, Sanejouand YH. Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: the elastic network model. J Mol Biol 2002;320:1011–1024.
69. Zhang Z, Shi Y, Liu H. Molecular dynamics simulations of peptides and proteins with amplified collective motions. Biophys J 2003;84:3583–3593.
70. Wilson EB, Decius JC, Cross PC. Molecular vibrations. New York: Dover Publications; 1980.