

9.3 *In Silico* Coarse-Grained Approaches to Structural Dynamics and Function of Proteins and their Assemblies

P Doruker, Bogazici University and Feza Gursey Institute, Istanbul, Turkey

Y Liu, University of Pittsburgh, Pittsburgh, PA, USA

Z Yang, University of California at San Francisco, San Francisco, CA, USA

I Bahar, University of Pittsburgh, Pittsburgh, PA, USA

© 2012 Elsevier B.V. All rights reserved.

9.3.1	Introduction: Protein Dynamics, Allostery and Function	27
9.3.1.1	Structure-Encoded Dynamics of Proteins: A Major Determinant of the Conformational Mechanisms for Biological Function	27
9.3.1.2	Soft Modes Define Accessible Pathways of Reconfiguration Near the Global Energy Minimum	29
9.3.1.3	Coarse-Graining: An Essential Tool for Exploring Collective Dynamics	29
9.3.2	Elastic Network Models: Theory and Methods	31
9.3.2.1	Normal Mode Analysis: Basic Assumptions and Methods	31
9.3.2.2	Gaussian Network Model	31
9.3.2.3	Anisotropic Network Model and Its Hierarchical Coarse-Graining	32
9.3.2.4	Mixed Resolution Models	34
9.3.2.5	Adaptive ANM Methodology	35
9.3.3	Applications to Allosteric Systems and Supramolecular Machines	36
9.3.3.1	Correlation between Experimentally Observed Structural Changes and ANM Modes	36
9.3.3.1.1	Principal component analysis of structural ensembles	36
9.3.3.1.2	Interaction of Hsp70 ATPase domain with nucleotide exchange factors	36
9.3.3.1.3	Triosephosphate isomerase	38
9.3.3.2	Toward Understanding the Molecular Basis and Mechanism of Supramolecular Machinery	39
9.3.3.2.1	Allosteric cycle of the bacterial chaperonin GroEL-GroES	39
9.3.3.2.2	Ribosomal machinery	40
9.3.4	Beyond Structural Dynamics: Extensions and Future Directions	41
9.3.4.1	Bridging between Sequence Correlations and Structural Dynamics	41
9.3.4.1.1	Mutual information theory	41
9.3.4.1.2	HIV-1 Protease sequence and structure correlations	42
9.3.4.1.3	Hsp70 ATPase domain	44
9.3.4.2	Exploring Transitions between Functional States	44
9.3.4.2.1	Using adaptive ANM	44
9.3.4.2.2	Using Monte Carlo simulations integrated with collective modes	48
9.3.4.3	Generation of Conformers for Ensemble Docking	48
	Acknowledgment	49
	References	49

Abbreviations

aANM	adaptive ANM	MC	Monte Carlo
AK	adenylate kinase	MD	molecular dynamics
ANM	anisotropic network model	MDR	multidrug resistance
CG	coarse-grained	MWC	Monod-Wyman-Changeux
DHAP	dihydroxyacetone phosphate	NBD	nucleotide binding domain
ENMs	elastic network models	NEF	nucleotide exchange factor
ET	evolutionary trace	NMA	normal mode analysis
GNM	Gaussian network model	NMR	nuclear magnetic resonance
Hb	hemoglobin	RMSD	root-mean-square deviation
		SBD	substrate binding domain

9.3.1 Introduction: Protein Dynamics, Allostery and Function

9.3.1.1 Structure-Encoded Dynamics of Proteins: A Major Determinant of the Conformational Mechanisms for Biological Function

In recent years, both experimental and computational studies have invited attention to the relevance of the equilibrium dynamics of proteins to the conformational changes that the proteins undergo to achieve their catalytic function or allosteric signaling.^{1–8}

The equilibrium dynamics refer to the collective motions accessible near a global energy minimum. These are computationally assessed by the normal mode analysis (NMA) of equilibrium structures.⁹ Biomolecular systems exhibit a broad range of motions, from high-frequency fluctuations near densely packed centers of energy localization to large-scale domain or subunit movements, also called global motions, which are several orders of magnitude slower. Importantly, the latter most cooperative modes, usually at the low-frequency end of the mode spectrum, are usually insensitive to structural and energetic details but determined by the overall contact topology, or fold, under native state conditions.^{10–12} Models exclusively based on inter-residue contact topology, such as elastic network models (ENMs),^{13–20} have thus found widespread applications in molecular structural biology in approximately the past decade due to their ability to accurately predict these cooperative modes and thereby provide insights into potentially functional changes in conformation while being remarkably efficient from computational time and memory standpoints. The theoretical foundations and extensions of methods based on ENMs are presented in Section 9.3.2.

Functional changes in structure vary over a broad range of time- and length-scales, and they encompass a diversity of events, from physical changes in structure required to recognize and/or bind a substrate protein, ligand, or a DNA/RNA segment to the collective machinery of large complexes and assemblies (e.g., folding of an encapsulated polypeptide by a chaperonin, processing of bound polynucleotides by DNA polymerase or the ribosome, and maturation of viral capsids), or to the chemical changes required for catalytic activity, post-translational modifications, and degradation reactions. Several examples of motions are presented in Section 9.3.3. Chemical reactions usually require precise positioning of the reacting groups and cofactors, if any, and NMAs with ENMs have shown that amino acids near the catalytic sites usually incur minimal change in their positions during the soft modes of motions undergone by the enzyme.²¹ The same regions, while being fixed in space, usually coincide with, or closely neighbor, the global hinge sites that mediate the concerted movements of the surrounding domains while being rigidly affixed in space. Typical examples are cleft regions between two domains, which usually contain the active site in many enzymes and serve as global hinge centers. These regions are therefore ‘translationally rigid’ but ‘rotationally flexible’ and allow for efficient conversion of chemical energy into mechanical energy. They are referred to as ‘key mechanical sites’ due to their critical role in mediating concerted motions. Substrate recognition and binding sites, on the other hand,

usually have a dual character – comprising a highly mobile segment (e.g., a recognition loop or subdomain) and relatively constrained residues coupled to mechanically key sites – so as to enable the long-term effects, also called allosteric effects, induced by substrate binding. Depending on the type of protein, substrate binding may be colocalized with the catalytic site and/or occur at a remote region, called a hot spot. In either case, it usually triggers a change in the structure and/or dynamics of the protein so as to facilitate the biochemical (e.g., catalytic) or biological (e.g., signaling) function of the protein.

The key point is that the previously described functional changes in the structure and dynamics of a given protein are predominantly defined by the protein or by its three-dimensional fold or contact topology.^{4,22,23} These so-called intrinsic motions are indeed sampled by the protein even in the absence of substrate/ligand binding. The protein, or biomolecular structure (composed of multiple proteins), is ‘pre-disposed’ to undergo those changes that appear to be induced upon ligand binding. This predisposition in favor of functional changes in structure, consistent with the Monod-Wyman-Changeux (MWC) description of allostery,²⁴ is proposed to be evolutionarily conferred; that is, structures may have evolved to favor movements (soft modes) that are actually functional so as to efficiently achieve their function. The applications in Section 9.3.3 illustrate this bridge between functional movements naturally observed (e.g., those derived from multiple X-ray structures for the same protein under different functional states) and the soft modes of motion predicted by ENM NMA to be intrinsically accessible to the protein, being uniquely encoded in the native fold/topology.^{22,25,26}

Perhaps these concepts are best illustrated by a protein or even a functional domain that is found to be highly ubiquitous – the ATPase domain of the Hsp70 family of molecular chaperones. The ‘chemical reaction’ site, which in this case is the ATP hydrolysis site, is located at a central cleft region between four subdomains. It also serves as a global hinge site. Nucleotide binding residues at this site are severely constrained in the global modes. The binding of its co-chaperonin, the nucleotide exchange factor (NEF), involves the immobilization of the (originally) most mobile region, subdomain IIB. The structural change stabilized upon NEF binding is that along a soft mode accessible to the ATPase domain in the absence of NEF binding.²⁷ Supporting the evolutionary significance of these structural and dynamic features, ATP binding residues are highly, if not fully, conserved. They cannot sustain any variability in sequence, which would imply variations in structure, at a region where there is a need for a precise proximity and orientation of reactants to ensure chemical reactivity. NEF binding sites, on the other hand, are sequentially variable. Their variation, however, is not random but, rather, occurs in a highly correlated way so as to ensure specificity.²⁸ Such coupled sequence and structural changes are discussed further in Section 9.3.4, which also gives examples of extensions of ENM-based methods to explore conformational changes between sub-states using different variations and combinations of coarse-grained (CG) models/methods and full atomic simulations. This type of hybrid approach holds promise for

simultaneously exploring the cooperative global dynamics of large biomolecular systems while retaining atomic information at functional sites.

9.3.1.2 Soft Modes Define Accessible Pathways of Reconfiguration Near the Global Energy Minimum

The landscape view of the free energy surface near a global minimum is now widely accepted as a conceptually helpful description, consistent with the multiplicity of folding pathways and accessibility of an ensemble of conformers in a multidimensional space of conformations.^{29,30}

In principle, for a protein composed of N interaction sites in a three-dimensional space, the conformational space has dimensions of $3N$, which can be conveniently expressed in terms of $3N-6$ internal and 6 external degrees of freedom. Each degree of freedom is associated with an 'axis' in this space, and the instantaneous state of the protein is described by a $3N$ -dimensional vector, the elements of which represent each point along these axes. The degrees of freedom could be the Cartesian (x, y, z) coordinates of the individual sites or a set of generalized coordinates such as the $N-1$ bond lengths, $N-2$ bond angles, and $N-3$ dihedral angles, plus six variables to fix the absolute position and orientation in space, for a chain molecule of N sites.

Suppose we are interested in examining the time evolution of the conformational state of the protein. To visualize the displacement of the protein in this $3N$ -dimensional space is not practical. Instead, some representative variables, such as the radius of gyration and the number of native contacts, are usually selected to depict the time evolution of conformation in a reduced space of two or three variables. Do these variables provide a good description of the conformational state? What would be the most representative variables if we are interested in examining the equilibrium motions of proteins? The answer is 'the principal axes' of motions, which are readily determined by the NMA of the equilibrium structure. These principal axes define the collective modes of motion.³¹ Each mode thus collectively involves simultaneous changes in the coordinates of multiple sites, hence its representation as a $3N$ -dimensional (eigen)vector. Also, the complete ensemble of modes ($3N-6$ of them in the absence of rigid body motions) form an orthonormal basis set that spans the complete space of conformational changes.

The mode space provides a complete description of collective motions in the neighborhood of a global energy minimum – here, the native state of the protein. In this 'dynamic' description of the energy landscape near the native state, each mode can be viewed as a direction/path of energy ascent away from the original energy minimum. Thus, the protein enjoys $3N-6$ different mechanisms of motions near the native state, each in the form of a harmonic oscillator, with some involving steep ascents in free energy and others involving softer ascents. Steep ascent directions correspond to high-frequency modes because the curvature of the energy landscape along this mode directly scales with the squared frequency of the mode. In contrast, soft ascent directions refer to low-frequency modes. For a given energy to be dissipated, or in the presence of a given perturbation, the molecule will

naturally tend to undergo the largest size displacement along the first mode, succeeded by the second mode, and so on. These global modes indeed explain the structural differences experimentally observed between the alternative structures of a given protein.²²

molecular dynamics This chapter presents CG models and methods that can be advantageously used to determine these soft modes, or the collective changes in conformation most readily accessible to biomolecular systems (see Section 9.3.2), and it demonstrates the relevance of these soft modes to the changes in conformations experimentally observed to be exploited when the protein or biomolecular system performs its function in the cellular environment (see Section 9.3.3). The chapter then demonstrates that these movements are closely related to the evolutionary selections of certain amino acids as highly conserved residues or correlated mutations sites (see Section 9.3.4).

9.3.1.3 Coarse-Graining: An Essential Tool for Exploring Collective Dynamics

The contour plot in **Figure 1** illustrates the projection of the energy landscape near a native structure onto a two-dimensional subspace spanned by two principal axes. The native state is a macrostate composed of three substates – in this case, S1–S3. Suppose S1 is the most stable (lowest energy) microstate. The NMA of this structure using a full atomic model with a detailed force field would yield the two principal axes, p_1 and p_2 . On the other hand, a CG description of the native state energetics, which takes into consideration the outer contours, would identify the principal axes P1 and P2. The ribbon diagrams in **Figure 1** illustrate the conformational motions along these two principal axes for a subunit of the bacterial chaperonin GroEL.

Two points are worth noting here. First, the directions of global modes (or the eigenvectors representative of the softest modes) obtained by either the detailed or the CG model show minimal, if any, difference. This feature, originally noted by Tirion,¹⁹ and confirmed since then in many studies,^{11,12,32,33} is referred to as the robustness of global modes. In principle, the two sets of principal axes (or unit directional vectors) need not be identical. However, numerous applications have demonstrated that they are practically indistinguishable. This is due to the fact that the global modes are cooperatively determined by the entire contact topology or native fold, and the level of detail (structure or energetics) included in the model does not practically alter the shape of the modes provided that the topology of inter-residue contacts is rigorously accounted for. Second, upon deforming the original state along the principal modes of motions, one may encounter, or approach, substates (e.g., S2 and S3) that are separated by relatively low energy barriers. Such transitions require the passage over an energy barrier, when examined at atomic resolution. However, coarse-graining of the structure smoothens out the energy profile, as illustrated in **Figure 2**, allowing for potentially sampling or approaching alternative substates, despite the use of a linear theory. **Figure 2** displays two substates (T and R') assumed by the bacterial chaperonin GroEL subunits. As illustrated later, calculations based on ENMs show that the transition T→R is

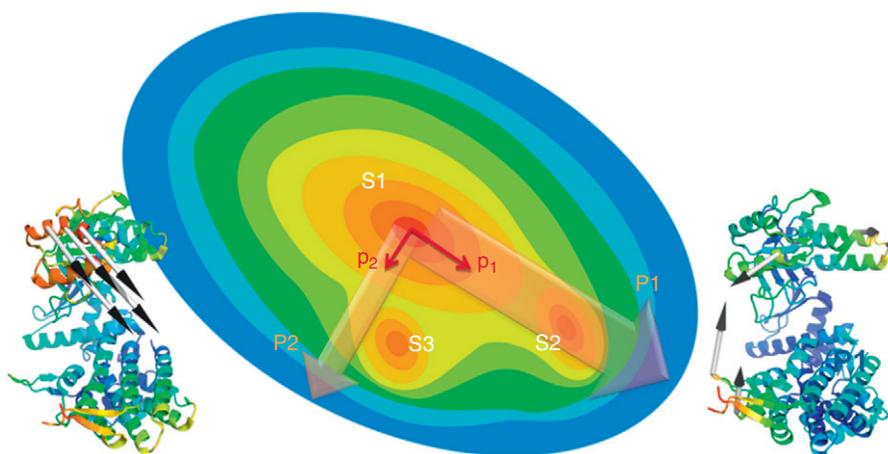


Figure 1 Schematic view of the energy landscape near the native state projected onto the two-dimensional space of principal coordinates. The native state is composed of three substates, S1–S3. The directions of the two softest (global) modes are depicted, based on the atomic NMA of the lowest energy substate S1 (p_1 and p_2 ; dark red arrows) and based on the coarse-grained (CG) NMA of the native (macro)state represented by the broader elliptical region (P1 and P2; transparent red arrows). The two sets of principal modes usually exhibit similar patterns. The ribbon diagrams on the left and right illustrate the movements in the two modes for a subunit of the bacterial chaperonin GroEL. The diagrams are color-coded from red (most mobile) to blue (most constrained, almost rigid). Image generated with the ANM web server¹⁴¹ and PyMOL. (The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC).

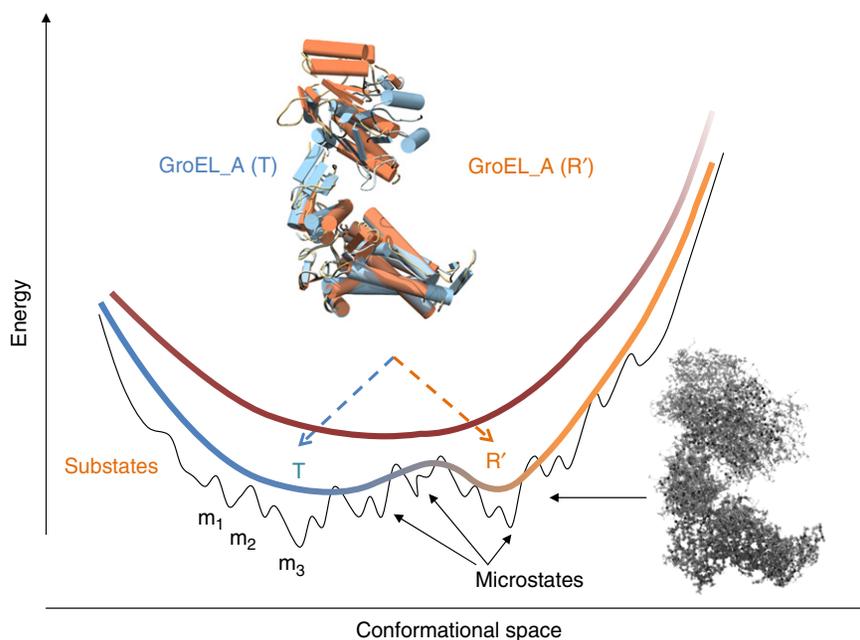


Figure 2 Potential energy at various levels of resolution. Three different energy profiles are shown, corresponding to three hierarchical levels of resolution. The lowermost curve refers to the most detailed (highest resolution) description of the structure, where several microstates (local minima) m_1 , m_2 , m_3 , etc. are distinguished. The diagram at the lower right illustrates such an ensemble of microstates, in which all atoms are explicitly displayed. At an intermediate level of resolution (middle curve), the ruggedness of the energy profile is considerably smoothed out to observe two substates, illustrated here by the T (tense) and R' (relaxed) forms of a given subunit (e.g., subunit A) in the chaperonin GroEL. The fold and secondary structure are practically maintained between the two forms, whereas the relative rearrangements of the domains exhibit differences. At an even lower resolution, the differences between the two substates may become indistinguishable, both being part of a global energy minimum characteristic of the fold shared by the two substates. Image generated with UCSF Chimera.¹⁴²

essentially enabled by one global mode, intrinsically accessible to the T state, modeled as a single energy minimum approximated by a harmonic potential at a CG scale.³⁴

Another important advantage of coarse-graining is the ability to examine long-time and large-scale biomolecular

processes, albeit at low resolution, that are beyond the scope of full atomic (e.g., molecular dynamics (MD)) simulations. Despite significant advances in computational technology, it is still a challenge to explore biologically relevant timescales and cooperative events by conventional simulations, hence the

need for developing CG models and methods. In principle, the level of resolution depends on the scale (length and time) of the process being investigated. The use of low-resolution models is justified when examining slow events during which fast motions are averaged out. However, in many systems, local events and global changes may be coupled, hence the need for considering multiple levels of resolution in integrated approaches, as elaborated later.

9.3.2 Elastic Network Models: Theory and Methods

9.3.2.1 Normal Mode Analysis: Basic Assumptions and Methods

The internal energy $V(\mathbf{q})$ of a molecular system of $3N$ degrees of freedom (e.g., the x , y , and z coordinates of N interaction sites/atoms) may be written as a series expansion around the equilibrium state $\mathbf{q}^0 = [q_1^0 \ q_2^0 \ q_3^0 \ \dots \ q_{3N}^0]^T$ as

$$V(\mathbf{q}) = V(\mathbf{q}^0) + \sum_i \left[\frac{\partial V}{\partial q_i} \right]^0 (q_i - q_i^0) + \frac{1}{2} \sum_{ij} \left[\frac{\partial^2 V}{\partial q_i \partial q_j} \right]^0 (q_i - q_i^0)(q_j - q_j^0) + \dots \quad [1]$$

where the summations are performed over all coordinates, and the superscript 0 refers to the equilibrium state \mathbf{q}^0 . We assume that the first term represents the zero level, and we note that the first derivative of the potential is by definition zero at \mathbf{q}^0 . $V(\mathbf{q})$ approximated to second order thus becomes

$$V(\mathbf{q}) = \frac{1}{2} \sum_{ij} (q_i - q_i^0) \left[\frac{\partial^2 V}{\partial q_i \partial q_j} \right]^0 (q_j - q_j^0) = \frac{1}{2} \Delta \mathbf{q}^T \mathbf{K} \Delta \mathbf{q} \quad [2]$$

where $\Delta \mathbf{q}$ is the $3N$ -dimensional vector of the instantaneous fluctuations, superscript T designates its transpose, and \mathbf{K} is the positive semidefinite matrix the ij th element of which is $K_{ij} = [\partial^2 V / \partial q_i \partial q_j]^0$. For a solid-like system, assuming that each atom behaves as a harmonic oscillator, the collective dynamics of the structure is governed by $3N$ equations of motion, written in compact form as

$$\mathbf{M}(d^2 \Delta \mathbf{q} / dt^2) + \mathbf{K} \Delta \mathbf{q} = 0 \quad [3]$$

\mathbf{M} is a diagonal matrix composed of N super-elements (3×3 matrices) along the diagonal, with each super-element being equal to $m_i \mathbf{I}_3$, where m_i is the mass of the i th atom and \mathbf{I}_3 is the identity matrix of order 3. The solution to eqn [3] has the form $\Delta \mathbf{q}(t) = \mathbf{a} e^{i\omega t}$, which upon substitution into eqn [3] leads to

$$(-\omega^2 \mathbf{M} + \mathbf{K}) \mathbf{a} = 0 \quad [4]$$

Premultiplication of eqn [4] by $\mathbf{M}^{-1/2}$ and substitution of the variables $\mathbf{u} = \mathbf{M}^{1/2} \mathbf{a}$, $\lambda = \omega^2$, and $\mathbf{H} = \mathbf{M}^{-1/2} \mathbf{K} \mathbf{M}^{-1/2}$ leads to the eigenvalue equation

$$\lambda \mathbf{u} = \mathbf{H} \mathbf{u} \quad [5]$$

NMA is the solution of this equation to obtain the $3N - 6$ nonzero eigenvectors $\mathbf{u}^{(k)}$ of the Hessian matrix \mathbf{H} , along with the corresponding eigenvalues, λ_k . $\mathbf{u}^{(k)}$ is a $3N$ -dimensional vector, the elements of which are organized in three-dimensional vectors $\mathbf{u}_1^{(k)}$, $\mathbf{u}_2^{(k)}$, \dots , $\mathbf{u}_n^{(k)}$, which represent the displacements of the individual atoms away from their equilibrium positions as the structure moves along mode k ; and the eigenvalue λ_k is the corresponding squared frequency. λ_k scales with the curvature of the energy landscape along mode k . Thus, the lower frequency modes (or global modes) have a smaller curvature/stiffness, and they undergo larger excursions from the energy minimum for a given energy increase, hence their 'soft modes' attribute.

9.3.2.2 Gaussian Network Model

The Gaussian network model (GNM) is the simplest ENM,^{14,16} inspired by the work of Tirion.¹⁹ In her seminal study, Tirion showed that the global modes evaluated using a detailed force field and those based on a purely harmonic potential with uniform force constants between all atom pairs (within an interaction range) are practically indistinguishable, thus inviting attention to the insensitivity of these modes to the details of interactions. Inspired by this observation, we introduced the GNM, in which the structure is represented by a set of nodes, representative of individual residues. Nodes are connected by springs of uniform force constant, γ , provided that they are located within a cutoff distance of R_c . **Figure 3**

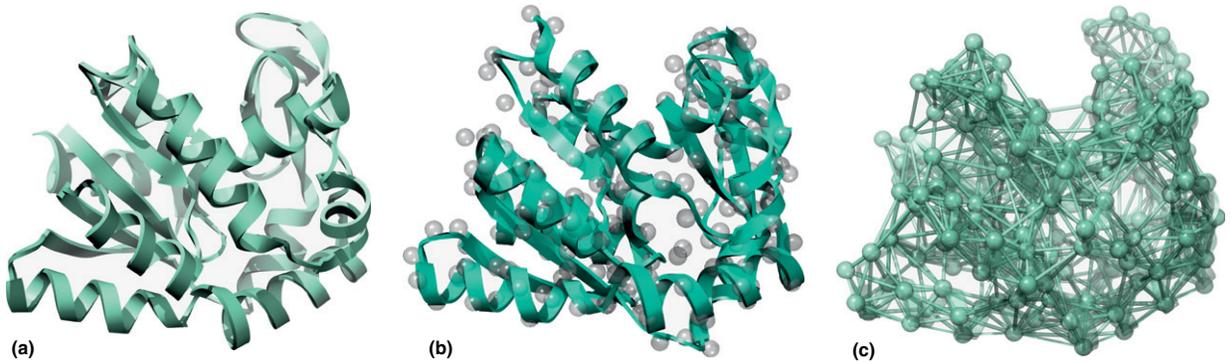


Figure 3 Three alternative representations of a given structure. Adenylate kinase (AK) (PDB code: 1AKE) structure is shown as (a) a ribbon diagram, (b) a ribbon diagram + spheres at the α -carbon positions, and (c) an elastic network model in which the nodes are identified with the α carbons and all pairs of nodes within a cutoff distance are connected (by harmonic springs). Image generated with UCSF Chimera.¹⁴²

displays the GNM representation of adenylate kinase, along with its ribbon diagram. The network nodes are located at the positions of the α carbons. The nodes evidently undergo Gaussian fluctuations $\Delta\mathbf{R}_i$ ($1 \leq i \leq N$) about their mean positions, hence the name GNM. The potential of the molecular system is given in terms of the fluctuations $\Delta\mathbf{R}_i = \mathbf{R}_i - \mathbf{R}_i^0$ in the position vectors of the nodes away from their equilibrium (native state) positions \mathbf{R}_i^0 as

$$V_{\text{GNM}} = \frac{Y}{2} \left[\sum_i \sum_{j>i} (-\Gamma_{ij}) [\Delta\mathbf{R}_j - \Delta\mathbf{R}_i]^2 \right] \quad [6]$$

where $\Delta\mathbf{R}_i - \Delta\mathbf{R}_j = \Delta\mathbf{R}_{ij}$ is the change in the distance vector between residues i and j , and Γ is the Kirchhoff (or connectivity) matrix, the off-diagonal elements of which are defined as $\Gamma_{ij} = -1$ if the distance $|\mathbf{R}_{ij}| \leq R_c$ and zero otherwise, and the diagonal elements are evaluated from the summation $\Gamma_{ii} = -\sum_j \Gamma_{ij}$ over all off-diagonal elements in the i th row (or column). Following the original statistical thermodynamics theory of random polymer networks,³⁵ the cross-correlations between the fluctuations of residues i and j are found from the statistical mechanical average

$$\langle \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j \rangle = \int (\Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j) \exp\left(-\frac{V_{\text{GNM}}}{k_B T}\right) d\{\Delta\mathbf{R}\} / \int \exp\left(-\frac{V_{\text{GNM}}}{k_B T}\right) d\{\Delta\mathbf{R}\} = \frac{3k_B T}{Y} [\Gamma^{-1}]_{ij} \quad [7]$$

where $[\Gamma^{-1}]_{ij}$ is the ij th element of Γ^{-1} , k_B is the Boltzmann constant, and T is the absolute temperature. Note that the determinant of Γ is 0; that is, Γ cannot be inverted. Instead, its pseudo-inverse is evaluated using the $N-1$ nonzero eigenvalues σ_k and eigenvectors $\mathbf{v}^{(k)}$ of Γ . In compact notation, the covariance between residue fluctuations is thus given by a weighted sum of the $N-1$ nonzero modes as

$$\mathbf{C}^{(N)} = \frac{3k_B T}{Y} \Gamma^{-1} = \frac{3k_B T}{Y} \sum_{k=1}^{N-1} [\sigma_k^{-1} \mathbf{v}^{(k)} \mathbf{v}^{(k)T}] \quad [8]$$

where $\mathbf{C}^{(N)}$ is a symmetric $N \times N$ matrix, the ij th element of which is $C_{ij}^{(N)} = \langle \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j \rangle$, and the i th diagonal element $C_{ii}^{(N)}$ is simply the mean-square (MS) fluctuation $\langle (\Delta\mathbf{R}_i)^2 \rangle$ of residue i . The contribution of any subset of modes to cross-correlations or MS fluctuations may be evaluated by performing the summation in eqn [8] over this particular subset. Evidently, the soft modes (smallest eigenvalues) make the largest contribution to the covariance. The eigenvectors are normalized such that the plot of $[\mathbf{v}^{(k)} \mathbf{v}^{(k)T}]_{ii}$ as a function of residue i represents the probability distribution of residue fluctuations in mode k , also called the k th mode profile.

9.3.2.3 Anisotropic Network Model and Its Hierarchical Coarse-Graining

The anisotropic network model (ANM)^{13,15} is an ENM that provides information on the directionality of residue fluctuations. ANM analysis is essentially a CG (residue-level ENM-based) NMA, also proposed independently by Tama and Sanejouand.¹⁸ The structure is represented as a network in which the distances between residues (as opposed to the

distance vectors in the GNM; see eqn [6]) are constrained by harmonic springs of uniform force constant. The N interaction sites are located at α carbon atoms. We note a first, residue-level ENM NMA has been proposed by Hinsen et al.^{17,36} A distance-dependent function has been adopted for the force constant in Hinsen's model, consistent with the weakening of atomic attractions with distance. Cutoff distances of $R_c \geq 13 \text{ \AA}$ have been adopted in previous applications of the ANM. The ANM potential energy is a summation over all harmonic interactions in the structure

$$V_{\text{ANM}} = \frac{Y}{2} \sum_i \sum_j h(R_c - R_{ij}) (R_{ij} - R_{ij}^0)^2 = \frac{1}{2} \Delta\mathbf{R}^T \mathbf{H} \Delta\mathbf{R} \quad [9]$$

where R_{ij} and R_{ij}^0 are the instantaneous and equilibrium distances between nodes i and j , respectively. $h(R_c - R_{ij})$ is the Heaviside step function that is equal to 1 if $R_c \geq R_{ij}$. $\Delta\mathbf{R}$ is a $3N$ -dimensional vector of the positional fluctuations of the N residues, $\Delta\mathbf{R}^T$ is its transpose, and \mathbf{H} is the ($3N \times 3N$) Hessian or force constant matrix (see Section 9.3.2.1). Note that in the ANM, $\mathbf{H} = \mathbf{K}$, and \mathbf{M} is taken as identity matrix of order N . By performing an orthogonal transformation of the ($3N \times 3N$) Hessian matrix (see eqn [5]), the overall motion can be expressed as a sum over the ($3N-6$) normal modes, with the six zero eigenvalues corresponding to the translation and rotation of the whole molecule. The global deformation driven by a specific mode k can be determined from the corresponding $3N$ -dimensional eigenvector $\mathbf{u}^{(k)}$. The diagonalization of \mathbf{H} becomes computationally exhaustive for a supramolecular system such as the ribosome. For this purpose, computationally efficient software package BLZPCK³⁷ with block Lanczos algorithm³⁸ is used to solve the eigenvalue problem for a number of collective modes.

Using the notation introduced in Section 9.3.2.1 and identical unit mass for each residue ($m_i = 1$ for all i) in residue-based coarse-graining, the MS fluctuation of the i th residue is given by

$$\langle (\Delta\mathbf{R}_i)^2 \rangle = (k_B T / Y) \sum_k \frac{|\mathbf{u}_i^{(k)}|^2}{\lambda_k} \quad [10]$$

The force constant γ is the only adjustable parameter in both the GNM and the ANM, whose value is usually determined based on the experimental B-factors. The value of γ does not affect the mode profile (i.e., the eigenvectors) but uniformly rescales the frequencies. Therefore, the relative sizes of the MS fluctuations of residues in the different modes are not dependent on the choice of γ . On the contrary, they are uniquely defined by a given contact topology or by the ENM.

The vast majority of ENM studies adopt this one-node-per-residue description – that is, a residue-based coarse-graining of the structure. However, there are interesting applications of lower resolution models – that is using renormalization techniques – which will be discussed next.

Soon after the introduction of ANM, a hierarchical coarse-graining methodology was proposed that is described here via its application to hemagglutinin A (HA), which is a homotrimeric integral membrane protein of the influenza virus.^{12,39} Each monomer comprises two different chains linked by disulfide bridges (Figure 4(a)). The lower part of the

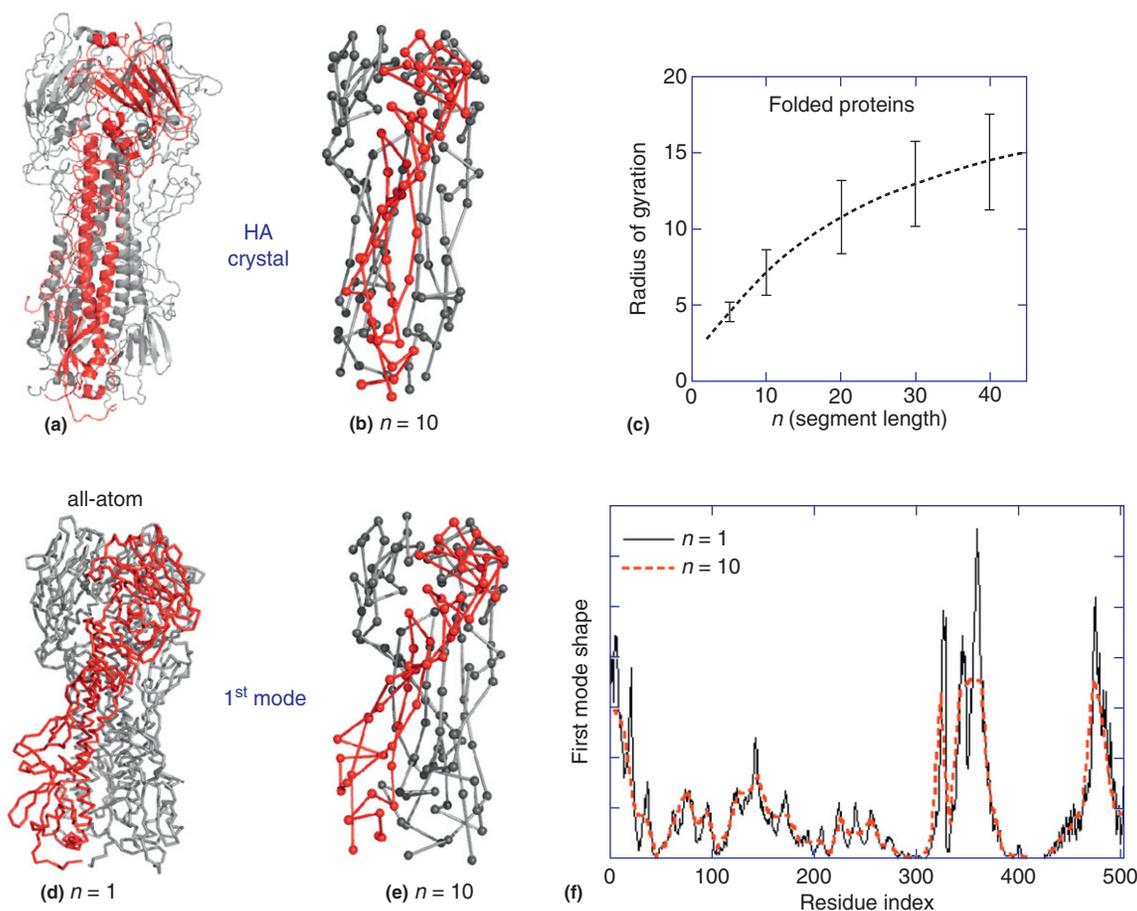


Figure 4 Hierarchical coarse-graining of hemagglutinin A. (a) The X-ray structure of influenza virus hemagglutinin A (HA; PDB code: 2HMG) determined at 3-Å resolution^{143,144} indicates that the ectodomain of the trimeric protein folds into a coiled-coil stem region supporting three receptor binding globular domains (monomer shown in red). (b) Hierarchical coarse-graining of HA structure, where each node/segment contains $n=10$ residues along the chain sequence. (c) The mean radius of gyration of segments in folded proteins as a function of segment size n , used for calibrating the cutoff distance of interaction in CG ANMs. (d) and (e) Twisting motion along the cylindrical axis of HA in the first mode for the original ($n=1$) and a low-resolution ($n=10$) ANM. Harmonic deformations from the native structure are exaggerated for clarity. (f) Similar first mode shapes at two levels of resolution – that is, normalized distribution of the squared displacements of residues in the softest mode, obtained for $n=1$ and $n=10$, in support of the applicability of the coarse-graining $n=10$. See Doruker et al.^{12,39} for more details. Image generated with PyMOL.

molecule is the membrane proximal region, and the top part contains the receptor binding sites.

In general, a protein composed of N residues can be further coarse-grained to s segments/nodes, each representing n residues. In a simple scheme (Figure 4(b)), HA composed of $N=1509$ residues has been coarse-grained at different levels ($n=2, 10, 20, 40$) by retaining every $(nk+1)$ th residue along the linear sequence of the backbone ($k=0, 1, 2, \dots$). The residue-based R_c (~ 13 Å) needs to be scaled for $n>1$ to generate an intact network with proper connectivity. For this purpose, the radius of gyration (R_g) calculated for segments of various lengths in folded proteins (Figure 4(c)) has been utilized to determine the cutoff radius³⁹

$$R_c = 2R_g + 13 \text{ \AA} \quad [11]$$

Once the cutoffs for different levels of resolution are fixed, the corresponding force constants can be determined so as to (uniformly) scale the computed B-factors to match the average value observed in experiments. Notably, the MS fluctuations of

residues display similar trends as the original ($n=1$) model, even for very low resolutions ($n \geq 10$). More important, the first (second) mode shapes indicate excellent fits to the original model with linear correlation coefficients above 0.9 (0.8), up to $n=40$. The first mode shapes of HA are compared for $n=1$ and $n=10$ (Figure 4(f)). The features of the vibrational frequency distribution are very similar at the lower end of the spectrum. Hierarchical coarse-graining has also proven useful for exploring the collective dynamics of larger proteins with different topologies, namely tetrameric β -galactosidase ($N=4092$) and dimeric xanthine dehydrogenase ($N=2664$).

HA serves two essential functions: the binding of influenza virus to target cells and the fusion of viral and endosomal membranes. Visual inspection of the slowest mode deformations indicates consistent behavior in terms of the collective motions of HA for different n values. The first mode (Figures 4(d) and (e)) is a twisting motion along the cylindrical axis of HA, which has been related to HA's function during membrane fusion – that is, fusion pore opening by the cooperative action of

clustered HA molecules.⁴⁰ The second and third modes represent a hinge-bending motion about the same site halfway along the cylindrical axis, where the deformation directions of the two modes are complementary in terms of the cylindrical symmetry of HA. Such a bending motion has been proposed to serve for the close association of viral and target membranes prior to fusion.^{41,42}

Another simple route to mimic the intact protein structure is to build an on-lattice model with a uniform density distribution. When the simple cubic lattice sites that fall within 4 Å of HA's α carbon atoms are retained for further ANM calculations with cutoff adjustments,⁴³ the collective modes for HA present a consistent picture with the original, off-lattice model. In summary, the primary determinant of global dynamics is clearly the overall shape because it is the only distinguishing factor among on-lattice representations of different folded structures besides possible cavities within.

9.3.2.4 Mixed Resolution Models

Computational efficiency gained by hierarchical coarse-graining is crucial especially for its application to supramolecular assemblies such as the ribosome. However, details at the atomistic level, or at least at the residue level, of the mechanisms would still be desirable for the region of interest, such as those at the active site of an enzyme (e.g., the decoding center and peptidyl transferase centers for the ribosome). Thus, the motivation behind the mixed resolution models is to retain high-level detail for the 'interesting' regions and at the same time to increase the computational efficiency by describing the rest of the structure at lower resolution. Therefore, the system is composed of a mixture of high-resolution (HR; atomistic or residue-based with $n = 1$) and low-resolution (LR; residue-based or lower resolution with $n > 1$) nodes. Here, the critical

step is the determination of cutoff radii and force constants for the interactions between the different-sized node pairs – that is, HR-HR, LR-LR, and LR-HR pairs.

If the HR regions are modeled at the residue level and the rest at lower resolution,⁴⁴ the cutoffs for LR-LR interactions ($n > 1$) can be based on eqn [11]. The force constants and cutoffs for the interface between the HR and LR regions need to be determined by empirical rules. On the other hand, to model the HR region atomistically with the remaining regions at residue level, the atomic cutoff could be adjusted so as to match the lower end of the frequency distribution from atomistic and residue-based ANMs. In this approach, the atomic cutoffs have been found to be system-size dependent, varying from 6 to 9 Å for relatively small proteins.⁴⁵ Although this approach satisfactorily described the collective dynamics for HA and triosephosphate isomerase (TIM; discussed later),^{44,45} the system-size dependence of the parameters and the empirical rules employed are major drawbacks, especially for supramolecules such as the ribosome.

This has led to a more universal approach that is applicable to all system sizes and composition, including protein, DNA, RNA, and possibly other components.⁴⁶ Here, a single atomic cutoff is chosen as the basis for all pairwise interactions independent of node size. Heavy atom pairs (HR-HR interactions) are connected with a uniform spring constant, taken as unity. In the LR region, one node may represent one or multiple consecutive amino acids/nucleotides, necessitating the use of larger force constants to link the heavier nodes. The force constant between two nodes (LR-LR or LR-HR) is therefore taken equal to the total number of interacting atom pairs among them (Figure 5).

In all mixed resolution methods, the Hessian matrix H is mass weighted, equating the total mass of each node to m_i (in eqn [3]). In the universal approach, an atomic cutoff distance

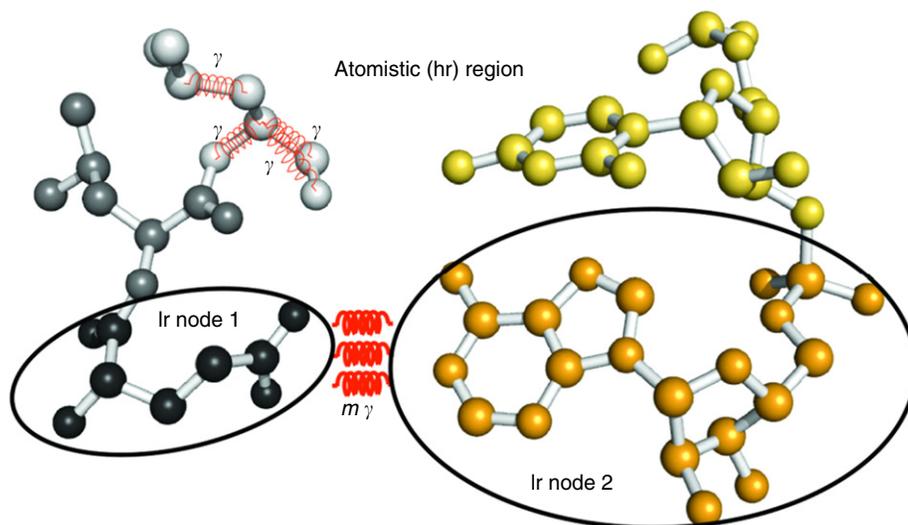


Figure 5 Mixed resolution model. The force constants between any pair of nodes are determined based on atomic cutoff radius. All neighboring pairs of heavy atoms in the high-resolution (hr) region (either amino acids or nucleotides) are linked by a uniform spring (with force constant γ). In the low-resolution (lr) region, nodes may represent one or a group of residues, and any CG node pair (here, lr node 1 and lr node 2) is linked via multiple springs ($m\gamma$), where m is the total number of atomic pair interactions among the nodes that fall within the atomic cutoff (atom pairs residing in the same node are not included in the count). Image generated with PyMOL.

of 10 Å is preferred to avoid extreme mobility of solvent-exposed tip residues with few neighbors (low connectivity). The similarity between two normalized eigenvector sets $\mathbf{v}^{(i)}$ and $\mathbf{w}^{(j)}$ from independent ANM calculations can be determined using overlap definition⁴⁷

$$\text{Overlap} = \left(\frac{1}{k} \sum_{i=1}^k \sum_{j=1}^k (\mathbf{v}^{(i)} \cdot \mathbf{w}^{(j)})^2 \right)^{1/2} \quad [12]$$

A subset of low-frequency modes (e.g., $k=10$) are conveniently used for a comparison of the collective dynamics. The single atomic cutoff of 10 Å has been valid using uniformly coarse-grained TIM models, which exhibit an overlap of 0.94 between the original, residue-based ($R_c = 13$ Å) and the atom-based cutoffs.⁴⁵ Mixed resolution models ($n=1$ or $n=5$ for LR region) for TIM have also produced quite satisfactory overlap values (greater than ~ 0.9) with the residue-based models. Additional validation of mixed resolution models has been performed by performing classical NMA and MD simulations on TIM.^{45,48}

Mixed resolution methodology has been particularly useful in exploring the functional motions of the ribosome. In Wang et al.'s original ANM analysis of the ribosome,⁴⁹ nodes of the network were located at the C $^\alpha$ and P atoms of the amino acid (aa) and nucleotide (nu) residues, respectively. The interaction distance between C $^\alpha$ -C $^\alpha$ pairs was taken as 15 Å and for C $^\alpha$ -P and P-P pairs as 24 Å in the uniform ANM because more distant sites necessitate the use of larger cutoff distances.^{12,39} Ribosome collective dynamics using the single atomistic cutoff of 10 Å for all interactions produced a quite satisfactory overlap of 0.72 with the original residue-based model described previously. Adjustments in the atomic cutoff parameters for the ribosome model (respective values of 10, 13, and 15 Å for the aa-aa, aa-nu, and nu-nu pairs) increased the overlap to 0.87. In the mixed resolution modeling of the ribosome structure, the HR region was selected as the codon and anticodon areas of mRNA and tRNAs at the A and P sites and the decoding center A1492, A1493 residing on 16S rRNA of the small subunit, which resulted in almost perfect overlap (0.99) with the residue-

based model.⁴⁶ Thus, atomic-level details have been obtained for this specific region of interest and for another antibiotic binding site.⁴⁶ More details on the residue-based ribosome models are presented later.

9.3.2.5 Adaptive ANM Methodology

The characterization of the transition mechanism between conformations is difficult, both experimentally and computationally, due to the transient nature of the intermediate, high-energy conformers crossed as the molecule undergoes the transition. In many cases, only the two ending structures are known from experiments.⁵⁰ Furthermore, the passage between the two endpoints does not necessarily involve a single pathway but, rather, multiple pathways. Obtaining a molecular understanding of the most probable transition pathways between the two end structures is a challenging task in the case of large systems in particular.^{51,52} On the other hand, studies indicate that the global transitions proceed, or at least start, via the collective global mode directions that are predicted by NMA. Several ANM-based studies indeed suggest that conformational changes may be described to a good approximation in terms of a few slow modes.^{4,18,20,36,49,53-57} However, given that NMA is valid only in the local region surrounding a potential energy minimum, its application to nonequilibrium events is possible only upon suitable coarse-graining of the energy landscape so as to eliminate/overcome low energy barriers that separate the substates (see Section 9.3.1.3).

Considering these caveats, the authors introduced a methodology called adaptive ANM (aANM), which utilizes the ANM modes to guide the motion of the biomolecular system along the directions intrinsically favored by its instantaneous inter-residue contact topology.³⁴ The essence of the method is to reevaluate the ANM modes at each step, as the structure reconfigures along the soft modes, hence the name adaptive. Details of the methodology may be found in the original study.³⁴ In brief, the aANM method consists of the following steps (Figure 6): Two sets of intermediates are generated, starting from both endpoints, A and B. The recurrence equation for evaluating the k th intermediate

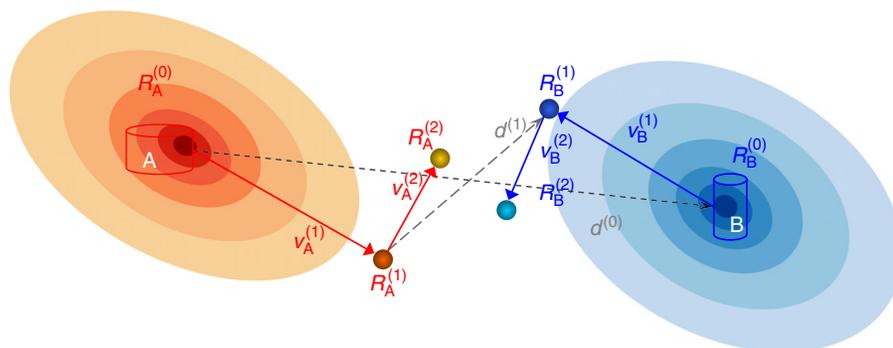


Figure 6 Schematic description of the sampling of the transition paths between two substates A and B using the adaptive ANM methodology. Both substates are modeled as ENMs. An iterative scheme is adopted, in which the two end structures ($R_A^{(0)}$ and $R_B^{(0)}$) undergo stepwise reconfigurations ($\mathbf{v}_A^{(1)}, \mathbf{v}_A^{(2)}, \dots$ and $\mathbf{v}_B^{(1)}, \mathbf{v}_B^{(2)}, \dots$) to successively reduce the original difference vector $\mathbf{d}^{(0)}$. The steps are undertaken along one or more soft modes. Intermediate pairs of conformations generated along the pathway are designated as $(R_A^{(1)}, R_B^{(1)}), (R_A^{(2)}, R_B^{(2)}), \dots$. See the text and Yang, Z.; Majek, P.; Bahar, I. Allosteric transitions of supramolecular systems explored by network models: Application to chaperonin GroEL. *PLoS Comput. Biol.* **2009**, *5*, e1000360, for more details. Copyright by PLoS ONE.

starting from state A is

$$\begin{aligned} \mathbf{R}_A^{(k)} &= \mathbf{R}_A^{(k-1)} + \mathbf{v}_A^{(k)} \\ &= \mathbf{R}_A^{(k-1)} + s_A^{(k)} \sum_{i=1}^{m_A^{(k)}} (\mathbf{d}^{(k-1)} \cdot [\mathbf{u}_A^{(i)}]^{(k)}) [\mathbf{u}_A^{(i)}]^{(k)} \end{aligned} \quad [13]$$

For simplicity, the variables for state A are defined here, and similar expressions hold for state B: $[\mathbf{u}_A^{(i)}]^{(k)}$ designates the normalized displacements (eigenvector) corresponding to the i th mode evaluated at the k th step for the conformation $\mathbf{R}_A^{(k-1)}$, $m_A^{(k)}$ is the number of (low-frequency) eigenvectors ($1 \leq i \leq m_A^{(k)}$) that contribute to the displacement $\mathbf{v}_A^{(k)}$ at step k , and $\mathbf{d}^{(k-1)}$ is the instantaneous (at step $k-1$) difference vector between the two endpoints (starting from the original distance vector of $\mathbf{d}^{(0)} = \mathbf{R}_B^{(0)} - \mathbf{R}_A^{(0)}$). The contribution of a given mode i to the displacement $\mathbf{v}_A^{(k)}$ is proportional to the projection $(\mathbf{d}^{(k-1)} \cdot [\mathbf{u}_A^{(i)}]^{(k)})$ of the instantaneous distance vector onto the eigenvector $[\mathbf{u}_A^{(i)}]^{(k)}$, and $s_A^{(k)}$ is a parameter that scales the step size. $s_A^{(k)}$ and $s_B^{(k)}$ are simultaneously selected at each iteration k as a fraction f of those, $s_{A,m}^{(k)}$ and $s_{B,m}^{(k)}$, that minimize $\mathbf{d}^{(k)}$. The limit $f \rightarrow 0$ refers to infinitesimally small displacements that are strictly accurate but prohibitively expensive (computationally), whereas the other extreme case $f \rightarrow 1$ is the most efficient move but may give rise to unphysical deformations in structural coordinates. Selected $f = 0.2$ in our work as a scaling factor that optimally balances between efficiency and accuracy.³⁴ The number $m_A^{(k)}$ is based on a threshold squared cosine, F_{\min} , that defines the maximal angular departure between the instantaneous displacement direction and that targeted. To this end, we evaluate the cumulative squared cosine at each step k ,

$$[C(m_A^{(k)})]^2 = \sum_{i=1}^{m_A^{(k)}} \cos^2(\mathbf{d}^{(k-1)}, [\mathbf{u}_A^{(i)}]^{(k)}) \quad [14]$$

and we select the minimal number of modes, starting from the low-frequency end of the spectrum, that satisfy the inequality $[C(m_A^{(k)})]^2 \geq F_{\min}$. Note that $C(m_A^{(k)})$ is identical to the correlation cosine between the instantaneous deformation and distance vectors³⁴ – that is,

$$C(m_A^{(k)}) = \cos(\mathbf{d}^{(k-1)}, \mathbf{v}_A^{(k)}) \quad [15]$$

Therefore, the threshold F_{\min} ensures the selection of the smallest subset of modes to drive the deformation $\mathbf{v}_A^{(k)}$ of the molecule toward a direction that does not deviate by more than a specified correlation cosine ($F_{\min}^{1/2}$) from the target direction $\mathbf{d}^{(k-1)}$. The use of the complete set of modes leads, by definition, to $C(3N-6) = 1$. By selecting a subset, we let the molecule undergo a structural change that is not necessarily toward the endpoint but, rather, along the soft coordinates energetically favored by its fold. Overall, the aANM thus involves two parameters, F_{\min} and f . The former controls the direction of motion, and the latter controls its size. Smaller F_{\min} values permit us to proceed via lower energy ascent directions, at the cost of longer excursions; smaller f implies more conservative displacements at each step. This scheme is repeated to generate a series of intermediate conformations until the root-mean-square deviation (RMSD) between the intermediates

becomes sufficiently small (comparable to the resolution of the structures). The total number of iterations, k_{tot} is thus defined by this targeted RMSD. The application of aANM to the allosteric transitions of GroEL is discussed later.

9.3.3 Applications to Allosteric Systems and Supramolecular Machines

9.3.3.1 Correlation between Experimentally Observed Structural Changes and ANM Modes

9.3.3.1.1 Principal component analysis of structural ensembles

Bakan and Bahar performed a comparative study of the global modes of motions indicated by experiments and those predicted by the ANM to make an assessment of the level of correlation between computational predictions and the changes in structures observed experimentally.²² The analysis was performed for a few proteins (HIV-1 reverse transcriptase (RT), p38 MAP kinase, and cyclin-dependent kinase 2) that have been crystallographically resolved in multiple functional forms. For example, there are more than 100 structures for RT, determined in the presence of various inhibitors, in the unbound form, or in the DNA-bound form. The authors also considered a few proteins (e.g., ubiquitin and calmodulin) for which multiple models have been determined by nuclear magnetic resonance (NMR) spectroscopy. The ensemble of structures experimentally (X-ray or NMR) determined for a given system was subjected to a principal component analysis to determine the corresponding dominant directions of structural changes P1, P2, and P3 in the 3N-dimensional space. In parallel, one representative structure from each ensemble was selected to perform an ANM analysis and determine the global modes/eigenvectors, $\mathbf{u}^{(1)}$, $\mathbf{u}^{(2)}$, ..., $\mathbf{u}^{(k)}$. Comparative analysis with dominant modes of motions from experiments and computations revealed a striking correlation between the two sets. A similar analysis performed by Jernigan and collaborators for HIV-1 protease also indicating the same behavior: The different forms observed in experiments for a given protein are essentially conformational variations along the softest modes of motions that are readily available to the native fold, and these modes can be accurately predicted by CG NMA.²⁵ The following sections provide more details for four systems (domains, enzymes, and molecular machines) the allosteric dynamics of which is critical to cellular functions: the ATPase domain in the Hsp70 family of molecular chaperones, the chaperonin GroEL, the enzyme TIM, and the ribosome as the major factory of protein synthesis.

9.3.3.1.2 Interaction of Hsp70 ATPase domain with nucleotide exchange factors

The Hsp70 family of 70-kDa heat shock proteins is a ubiquitous group of molecular chaperones that plays a crucial role in regulating the correct folding and intracellular trafficking of proteins: It binds nascent polypeptide chains and prevents their unwanted aggregation, especially under exposure to heat, stress, or toxic agents, or directs them to cellular biodegradation pathways.^{58–60} Two widely studied members of this family are the *Escherichia coli* DnaK and the human Hsp70.

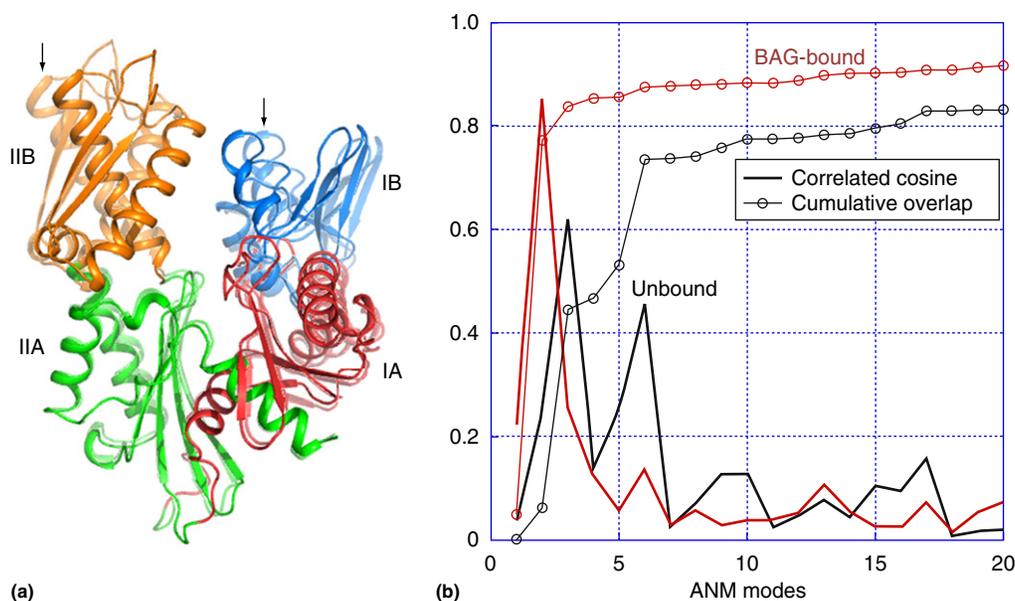


Figure 7 Comparison of experimentally observed and computationally predicted structural changes for the Hsp70 ATPase domain. The experimentally observed changes are illustrated for the NEF-bound and free forms of the Hsp70 ATPase domain (respective PDB files 1HX1⁶² and 1HPM¹⁴⁵ corresponding to the BAG-bound and free forms of eukaryotic chaperone). Computational results are obtained by applying the ANM to the two structures. (a) Structural alignment of NEF-bound and -unbound states of the ATPase domain. The ATPase domain structure is colored by subdomains: IA (red; residues 1–39 and 116–188), IB (blue; residues 40–115), IIA (green; residues 189–228 and 307–385), and IIB (orange; residues 229–306). The unbound ATPase fragment is shown in the foreground, whereas the NEF-bound form (1HX1) is shown in the background. The regions showing the largest deformation are marked by arrows. (b) Results for the unbound (black curves) and BAG-bound (red curves) ATPase domain. The solid thick curves represent the correlation cosine between the experimentally observed structural deformation (between the NEF-bound and -free forms) and the softest 20 ANM modes accessible to the NBD. The thin curves with the circles describe the cumulative overlap, summed over subsets of modes. The results show that a subset of six slow modes accessible to the unbound form ensures the passage to the NEF-bound conformer with an overlap of 0.85. The NEF-bound form exhibits an even stronger potential to be reconfigured back to its closed form, consistent with the preferred conformation of the NBD in the absence of NEF: The two top ranking modes yield a cumulative overlap of 0.87 with the experimental deformation vector.⁶⁸ The ribbon diagram was generated with PyMOL.

Hsp70 proteins are composed of a nucleotide binding domain (NBD), also called ATPase domain, and a substrate binding domain (SBD). The NBD is composed of four subdomains (**Figure 7(a)**): subdomains IA, IIA, IB, and IIB. The substrate binding affinity of the SBD is regulated by the ATPase domain via the allosteric ATPase cycle.⁵⁸ In the absence of substrate binding, the NBD is usually in the ATP-bound state. Substrate binding stimulates the hydrolysis of ATP (which is enhanced by the assistance of co-chaperones, known as J proteins). ATP hydrolysis provides the energy needed for subsequent conformational changes that assist in processing the bound, partially folded polypeptide. This process is terminated upon binding of a NEF, another co-chaperone, to the ATPase domain. NEFs stabilize an open conformer of the ATPase domain (**Figure 7(a)**),^{61,62} thus facilitating the release of the product (ADP) from ATP hydrolysis, as well as the insertion of a new ATP molecule.⁶³

Although the precise geometry of complex formation with the NBD varies among different types of NEFs – namely, GrpE,⁶⁴ BAG,⁶⁵ HspBP1,⁶⁶ and Sse1⁶⁷ – they all share some common features, including the recognition of subdomain IIB and the stabilization of an open conformer to facilitate ADP/ATP exchange.⁶⁸ The structural change ‘apparently induced’ upon NEF binding has been shown by both NMR experiments²⁷ and molecular computations⁶⁸ to be an intrinsic feature of the ATPase domain structure. NEF binding essentially

exploits the predisposition of the ATPase domain to undergo a conformational transition to the open form. An unambiguous verification of this property is provided by comparing (1) the experimentally observed structural change $d^{(0)}$ between the NEF-free and NEF-bound forms of the NBD and (2) ANM-predicted global modes of deformation, intrinsically accessible to the NBD in the absence of NEF binding.⁶⁸ The results are presented in **Figure 7(b)**. The top-ranking 20 ANM modes at the low-frequency end of the mode spectrum calculated for unbound NBD are compared with the $3N$ -dimensional deformation vector $d^{(0)}$ found from optimal structural alignment⁶⁹ of the open (NEF-bound state) and closed (unbound) conformations. The ordinate displays the correlation cosine, $u^{(k)} \cdot d^{(0)} / |u^{(k)}| |d^{(0)}|$, between each mode $u^{(k)}$ ($k = 1, 20$) and $d^{(0)}$ and also the cumulative overlap ($\sum_k \cos^2(d^{(0)}, u^{(k)})$). Strikingly, the third ANM mode exhibits a correlation cosine of 0.62, alone, with the deformation vector. This number is remarkably high, given that a randomly selected $3N$ -dimensional vector (as a control) would, by definition, yield a correlation cosine of $(1/3N)^{1/2} = 0.029$ in the current case of $N = 380$ residues in NBD. Thus, mode 3, which is intrinsically accessible to the NEF-free NBD, exhibits a 200 times stronger correlation with experimental deformation compared to a random $3N$ -dimensional directional vector.

ANM calculations were repeated with the open conformations as well to examine whether the open structure of

the ATPase domain, observed in the presence of NEF, has a tendency to restore its conformation back to the closed form in the absence of NEF. Results displayed in **Figure 7(b)** show that the ANM slow modes sampled in the open form have even higher correlations with $d^{(0)}$; mode 2 exhibits a correlation cosine of 0.85, and the first two modes contribute approximately 77% toward the overall cumulative overlap. These results provide evidence for the strong intrinsic tendency of the open form to change back to the closed form in the absence of NEF.

Overall, the previous analysis shows that a small subset of slow modes is exploited by the NBD to optimize its interaction with NEF. The NBD appears to have structurally evolved to favor the collective conformational changes that are required for its biological activity.

9.3.3.1.3 Triosephosphate isomerase

TIM is a crucial enzyme in the glycolytic pathway catalyzing the interconversion between dihydroxyacetone phosphate

(DHAP) and D-glyceraldehyde 3-phosphate. TIM is fully active as a homodimer (~ 500 residues), even though no cooperativity or allostery has been reported between the two catalytic sites located in the middle of each TIM barrel (**Figure 8(a)**).⁷⁰ Loop 6 (residues 166–176) plays a key role during catalysis by closing over the ligand by a 7-Å displacement of the loop's tip between the open and closed states in the crystal structures.

ANM has revealed, for the first time, large-scale domain motions of the enzyme that are coupled with loop dynamics.⁴⁵ A mixed resolution model has been adopted in these calculations, in which the HR region comprises the active site residues and the functionally important loop 6 on one monomer with the remaining residues described at low resolution to maintain the overall shape of the enzyme. In both uniform and mixed resolution models, a correlation was observed between the global motions (first and fourth modes) of the overall enzyme and the local loop 6 opening/closure dynamics.⁴⁵ The alternative conformations that occur during the fluctuations along the first mode (**Figures 8(b)** and **(d)**)

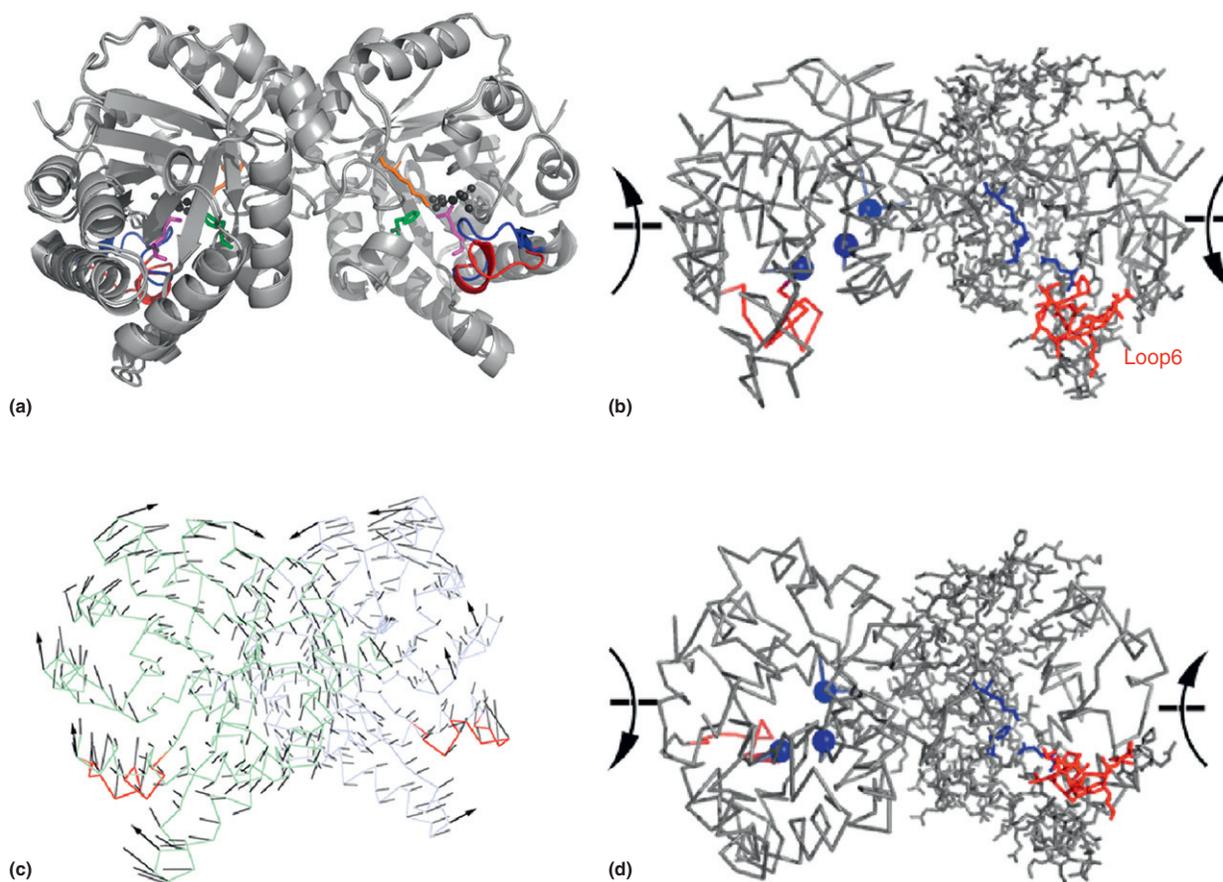


Figure 8 Triosephosphate isomerase dynamics. (a) Aligned X-ray structures of apo (PDB code: 8TIM) and ligand-bound (1TPH) TIM show the loop 6 in open (red) and closed (blue) conformations, respectively. The ligand (black spheres) is phosphoglycolohydroxamate, an analog of a reaction intermediate.¹⁴⁶ Active site residues are Lys13 (orange), His95 (green), and Glu165 (magenta). (b) and (d) Alternative conformations in the first mode of the mixed resolution model emphasize that the loop opening/closure is an intrinsic feature of the apo enzyme coupled to collective dynamics (a counter-rotation of the subunits indicated by arrows). The ratio of the high-resolution (atomistic) to low-resolution ($n=1$) nodes is 1290/323. Loop 6 is red and active site residues are blue in both conformations. (c) The first essential mode from a 60-ns MD simulation of the dimer is a counter-rotation of the subunits. Deformation vectors (loop 6 shown in red) indicate that the loop closure is coupled to the global motion, in conformity with the results from ANM. Image generated with PyMOL.

indicate proper loop closure (in accord with known ligand-bound structure) for both atomistic and residue-based regions of the mixed resolution model.⁴⁶ In contrast, ANM calculations performed for the high-resolution region have only resulted in a low overlap value (0.25 using $k=4$ in eqn [12]) with an all-atom ANM of TIM based on the first four modes, which emphasizes the necessity of intact three-dimensional structure for correct collective dynamics.⁴⁵ Thus, loop opening/closure emerges as an intrinsic dynamics feature of the dimeric enzyme, which is consistent with the experimental finding that loop closure is not ligand gated in the apo state.⁷¹

When the available crystal structures for TIM are aligned, different conformations of loop 6 are predominantly observed (as in **Figure 8(a)**) without a clear indication of the collective motions described by ANM. As a result, MD simulations^{72–74} have mainly concentrated on the loop region without taking the flexibility of the whole protein into account. However, MD simulations (60 ns) on fully flexible, dimeric TIM⁴⁸ have presented a complementary and consistent view to the ENM. The essential dynamics of MD trajectories was extracted by principal component analysis of the covariance matrix of internal displacements.⁷⁵ The first mode (representing 34% of the overall motion) was found to be a counter-rotation of the two subunits, coupled to the closing/opening motion of loop 6 (**Figure 8(c)**). In contrast, MD simulations performed on a single monomer of TIM have not shown significant collective motions and proper loop closure. This may partially explain why the catalytic activity is highly suppressed in a monomeric TIM designed with almost the same TIM-barrel fold.⁷⁶

Elastic network modeling has provided other insights in the context of the relationship between intrinsic conformational flexibility and catalytic activity.⁷ With respect to the first step of the catalytic reaction, significant changes in the interatomic distances are observed in the slowest ANM mode that could promote proton transfer from DHAP to Glu165.⁷⁷ In addition to the catalytic reaction, deamidation of Asn71 and Asn15 takes place at the dimer interface of TIM, which leads to subunit dissociation, unfolding, and subsequent degradation of the enzyme in mammals. A GNM study identified a network of coupled motions between the deamidation and catalytic sites of TIM, which may affect the deamidation barrier height.⁷⁸

9.3.3.2 Toward Understanding the Molecular Basis and Mechanism of Supramolecular Machinery

9.3.3.2.1 Allosteric cycle of the bacterial chaperonin GroEL-GroES

Many allosteric systems are multimeric and packed in spatially symmetric ways. A classic example is hemoglobin (Hb), a tetramer usually referred to as a dimer of a dimer, given the structural and dynamic correspondence between the pairs of α subunits α_1 and α_2 and β subunits β_1 and β_2 . The transition of Hb between its relaxed (R) and tense (T) forms occurs via the coupled movements of the two dimers, which give rise to a conformational switch at the $\alpha_1\beta_2$ and $\alpha_2\beta_1$ interface. ENM-NMA calculations demonstrated that this conformational change is achieved by moving along the second softest mode.⁵⁴

Concerted motion of multiple subunits is a feature observed in many allosteric systems. In many cases, the structures are multimeric, composed of identical subunits arranged in a symmetrical geometry. The soft modes that simultaneously induce the same global change in all subunits maintain the symmetry of the structure. These modes are nondegenerate (i.e., no other symmetrically related mode complements them) and instrumental in achieving the transition between two symmetrically arranged end structures. A typical example is the bacterial chaperonin GroEL (**Figure 9**). GroEL is a large (on the order of megadaltons) molecular machine that assists the folding of a number of *E. coli* proteins via an ATP-regulated allosteric mechanism. It consists of two rings, *cis* and *trans*, each comprising seven subunits with a sevenfold symmetry around the cylindrical cavity that encapsulates the substrate (misfolded or partially folded protein or peptide). Structures along the allosteric cycle of GroEL highlight the cooperative changes accompanying its transitions between unliganded (T form of all subunits in both rings or T/T form of the GroEL), ATP-bound (to all seven subunits on one of the rings, called the *cis* ring; R/T), co-chaperonin (GroES)- and ATP-bound (R'/T), ADP-bound (to all *cis* ring subunits; R''/T), and both ATP- and ADP-bound (to the respective *trans* and *cis* rings; R''/R) states of GroEL. **Figure 9** illustrates the conformations R''/R and T/R viewed from the side (**Figure 9(a)**) and from the top (**Figure 9(b)**). **Figures 9(c)** and **(d)** show the coupled conformational changes of pairs of subunits between the R'' and T states. The subunits in **Figure 9(c)** are colored by their extent of mobility (as in **Figures 9(a)** and **(b)**, with blue being the most rigid and red the most mobile regions); those in **Figure 9(d)** display the three domains (apical, yellow; intermediate, red; and equatorial, blue) on each subunit. A pair of subunits is shown in each case to illustrate how the adjacent subunits in a given ring undergo concerted changes in structure – that is, the change in one subunit could not be undertaken if the adjacent subunits were not undergoing the same type of conformational change.

The experimentally observed change between the R'' and T forms of the entire ring conforms closely to the global (softest) mode predicted by ANM for a single subunit. The bars in **Figure 9(e)** display the correlation cosine between the ANM modes $u^{(k)}$ (for $k < 25$) accessible to the R'' state and the experimentally observed difference $d^{(0)}$ between the R'' and T forms of GroEL. Strikingly, the first lowest frequency mode accessible to the single subunit yields a correlation cosine of approximately 0.8 with the change undergone in the GroEL complex during its allosteric cycle. The cumulative correlation squared, $\sum_k \cos^2(d^{(0)}, u^{(k)})$, over all $3N - 6$ modes is by definition 1. The observed correlation of 0.8 is therefore higher by a factor of approximately 10^5 compared to that expected from a random vector in $3N$ -dimensional space. **Figure 9(f)** illustrates how the RMSD between the two endpoints of the transition R'' \rightarrow T decreases from its original value of approximately 12.4 Å to approximately 7 Å by moving along only mode 1. The other two curves display the further decrease achieved by additional movements along two more modes. The right plot in **Figure 9(f)** refers to the reverse transition, T \rightarrow R''. Mode 1 has practically no effect in this case, whereas modes 3 and 4 are functional. The passage from R'' to T is a

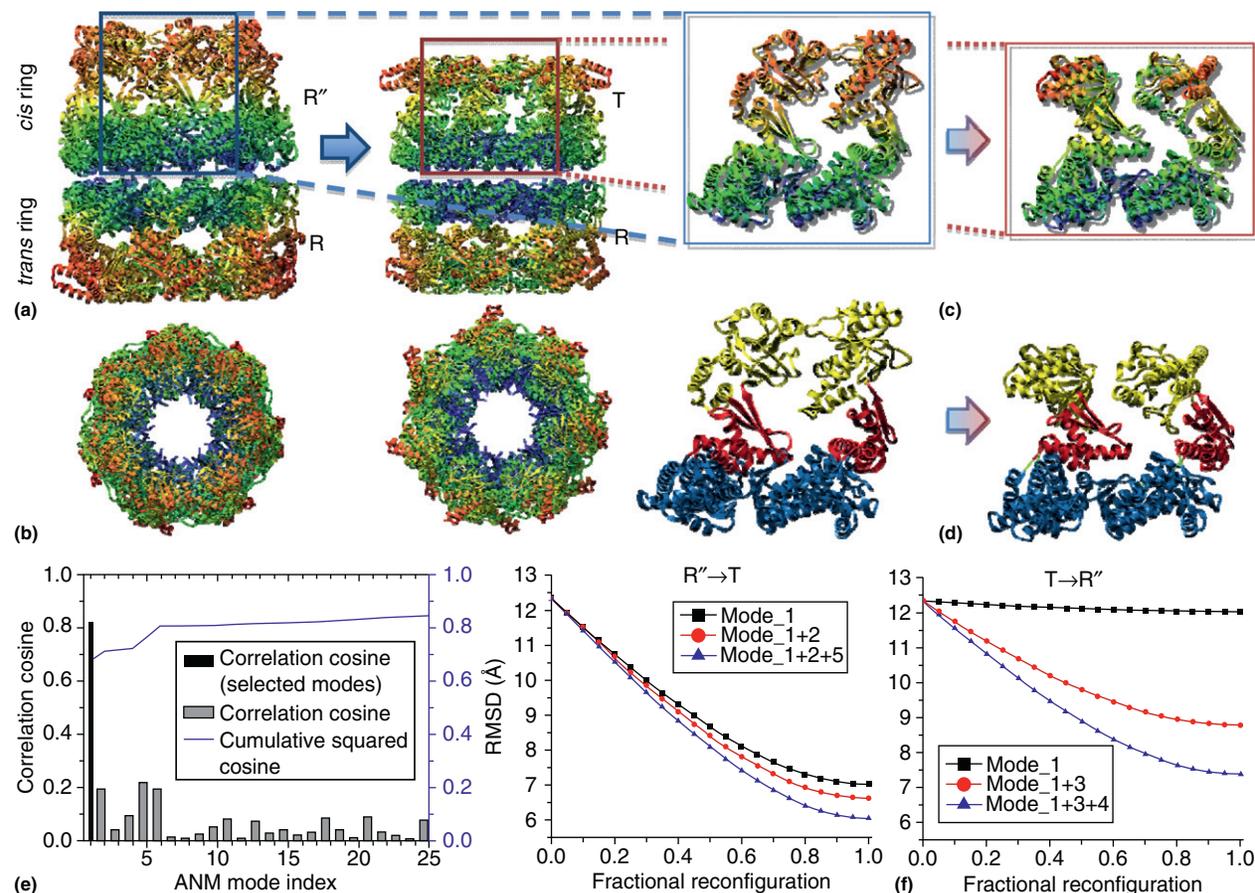


Figure 9 Functional changes in GroEL structure and relation to intrinsic dynamics of individual subunits. Structural transition of GroEL between the R''/R to the T/R, viewed from the side (a) and from the top (b), color-coded according to the mobilities in the softest (global mode) (blue, most rigid; red, most mobile). Two subunits are enlarged in panel c, and they are colored by their domains in panel d. (e) Correlation cosine between the experimentally observed deformation of a subunit during the transition R'' \rightarrow T and the ANM modes accessible to the state R''. (f) Decrease in the RMSD between the two endpoints as the structure reconfigures along the indicated one, two, or three soft modes in the forward (left) and reverse (right) directions. More details may be found in Yang et al.³⁴ Image generated with UCSF Chimera.¹⁴²

'natural' change that occurs spontaneously upon releasing all seven ADPs (of the R'' form). The high correlation between the ANM-predicted softest mode and the structural change naturally observed is noteworthy. The opposite change, on the other hand, is driven by ATP binding, consistent with the computed lower intrinsic propensity of the subunit to undergo this change spontaneously.

9.3.3.2.2 Ribosomal machinery

The ribosome is a molecular machine that synthesizes proteins based on the genetic information encoded by mRNA. The bacterial 70S ribosomal complex is a supramolecular assembly of the small (30S) subunit composed of 16S rRNA and 22 proteins and the large (50S) subunit with 23S and 5S rRNAs and 34 proteins (Figure 10(a)). During the process of translation – composed of initiation, elongation, termination, and recycling steps – the ribosome interacts critically with other key factors, such as mRNA, tRNAs, initiation factors, and elongation factors (EF-G and EF-Tu). Cryo-electron microscopy and X-ray crystallography have provided valuable information on ribosome conformational dynamics and its interactions with ligands, recently at high resolution.⁷⁹ Single-

molecule fluorescence resonance energy transfer studies have played a key role in the proposal of a mechanism for tRNA translocation through the ribosome in the elongation step. This multistep process of translocation, which presents the features of a complex energy landscape, is intricately coupled to the ratcheting and unratcheting motions of the ribosome.⁸⁰

The ratchet-like rotation of the 30S subunit relative to 50S, first observed by cryo-electron microscopy,⁸¹ has emerged as an intrinsic motion of the 70S complex in all CG ENM studies performed to date^{46,49,82–84} (Figures 10(b) and (c)) and also in a CG MD simulation.⁸⁴ The deformations of tRNAs appear to be coupled to the ratchet-like rotation (Figure 10(d)). Moreover, another ANM mode⁸⁵ reveals a 30S head rotation around the neck, consistent with experimental results.⁷⁹ Both of these collective modes may be involved at certain stages of the translocation mechanism of tRNAs residing in the A and P sites. High anticorrelated fluctuations of the L1 and L7/12 stalks in slow modes bear functional significance. Specifically, the closure/opening of the intersubunit interface due to L1 stalk mobility may be linked to the exit of the E-tRNA, and the mobility of the L7/L12 stalk may be required for the

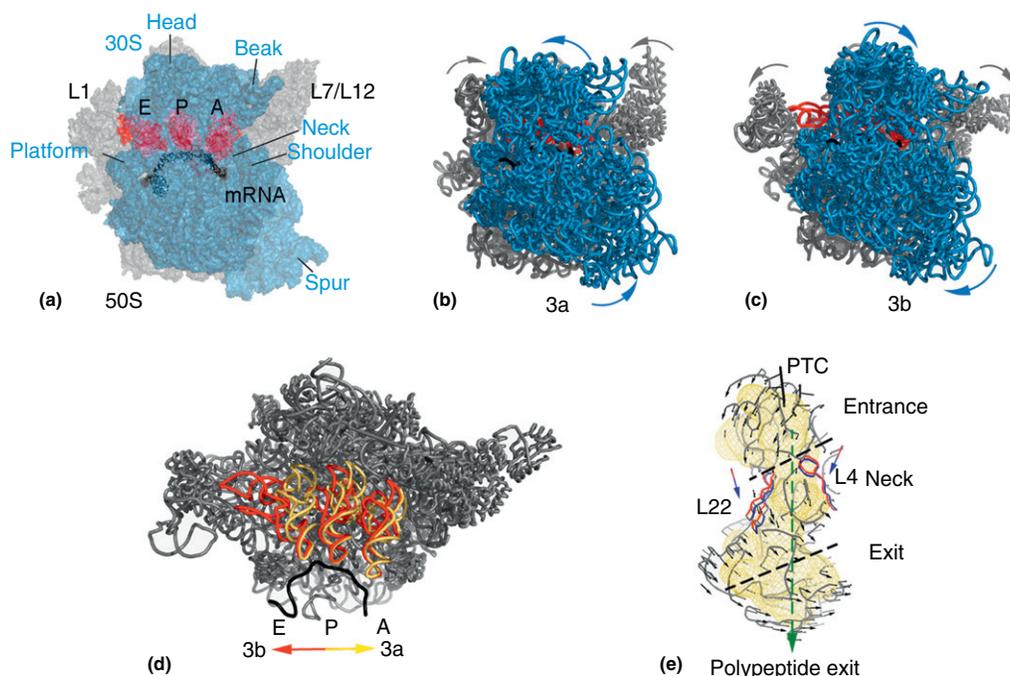


Figure 10 Ratchet-like rotation of the ribosome, an intrinsic dynamics feature predicted by the ANM. (a) Functionally important regions of the bacterial ribosome are shown on the crystal structure of the bacterial complex 70S. The small subunit 30S (in transparent blue) and the large subunit 50S (gray) form the 70S ribosomal complex with the mobile stalks (L1 and L7/L12) on the 50S. At the interface between the two subunits, the mRNA (black) bound to the neck region of 30S and the tRNAs bound to the A, P, and E sites are shown. The 70S complex, which is formed by using the PDB structures 1JGO and 1GIY,^{147,148} has been used in ANM calculations presented here.^{49,85} (b and c) Ratchet-like rotation of the 30S subunit with respect to the 50S is shown by generating two alternative conformations (3a and 3b) about the native state using the third mode from ANM (deformations exaggerated). The anticorrelated motion of the stalks (L1 and L7/L12) and the high deformation of the E-tRNA toward the exit direction are coupled with ratchet-like rotation. (d) Alternative conformations of tRNAs (yellow and red) in the third mode are shown after alignment of the mRNAs in both conformations. Small subunit 30S is not shown for clarity. Deformations of tRNA in the A and P sites seem relevant for the translocation mechanism. (e) Ribosomal tunnel elements amount to 379 CG residues that fall within 10 Å of the tunnel wall (yellow mesh extracted by the program ^{3V}¹⁴⁹) in the 50S crystal structure. Deformation vectors of the tunnel elements are shown by the action of the third ANM mode from the intact 70S structure.⁸⁸ Alternative conformations of ribosomal proteins L4 and L22 (red and blue) display a swinging motion perpendicular to the polypeptide exit route, at the narrowest region of the tunnel. Image generated with PyMOL.

recruitment of elongation factors (Figures 10(b)–(d)), in line with experimental findings.^{86,87} Intrinsic flexibility of the neck region in 30S seems to be a requirement for the binding of mRNA and tRNAs.⁴⁹ Common features, mainly ratchet-like rotation and stalk motions, survive in a hypothetical model excluding ribosomal proteins, which implies that ribosome’s built-in catalytic activity (carried out by the rRNA) is evolutionarily conserved.⁸⁵

Structural studies have shown that proteins are synthesized at the peptidyl transferase center and emerge through a tunnel (with ~15 Å diameter and ~100 Å length), both located in the large subunit 50S. The collective deformations of the polypeptide tunnel have been extracted based on the ANM calculations for the intact 70S ribosome.⁸⁸ Orientational cross-correlations of residues clearly divide the tunnel into an entrance region and the rest comprising the neck and the exit regions, in line with the experimentally determined polypeptide folding regions.⁸⁹ The motions of ribosomal proteins L4 and L22 located at the narrowest part of the tunnel may be related to the polypeptide gating mechanism.⁹⁰ The third mode in Figure 10(e) (i.e., ratchet-like rotation) represents an overall downward movement that may direct a nascent polypeptide toward the exit. Again, the intrinsic dynamics of the ribosome (several of the first five

modes and their linear combinations) bear functional implications in terms of guiding the nascent polypeptide through the tunnel and the gating mechanism.

9.3.4 Beyond Structural Dynamics: Extensions and Future Directions

9.3.4.1 Bridging between Sequence Correlations and Structural Dynamics

9.3.4.1.1 Mutual information theory

In recent years, there have been many advancements in the development and implementation of high-throughput sequencing techniques. Sequence analyses usually aim at identifying evolutionarily conserved residues, based on the hypothesis that functional sites ought to be conserved. However, a second group of residues, not conserved but evolving in a correlated way, has provided almost equally valuable information on functional mechanisms.

The coevolution of residue pairs, also referred to as correlated mutations,⁹¹ is now being broadly explored to obtain information on a range of issues, including the possible prediction of native contacts in proteins^{92–94} and the identification of probable docking sites at protein-protein

interfaces.^{95,96} Concepts from statistics and information theory are utilized, such as Pearson correlation statistics⁹⁷ and mutual information (MI) theory.⁹⁸ Other approaches have been borrowed from established statistical thermodynamics concepts, such as statistical coupling analysis introduced by Lockless and Ranganathan.⁹⁹ A comparison of various methods can be found in the literature.^{100,101}

MI is one of the earliest techniques employed for detecting coevolving residues.^{102,103} It quantifies the degree to which two random variables co-vary. The MI conveyed by two discrete random variables X and Y is defined as

$$I(X, Y) = \sum_{\text{all } x} \sum_{\text{all } y} P(X=x, Y=y) \log \frac{P(X=x, Y=y)}{P(X=x)P(Y=y)} \quad [16]$$

where $P(X=x, Y=y)$ is the joint probability of occurrence of the values/properties x and y for the respective variables X and Y , and $P(X=x)$ and $P(Y=y)$ are the corresponding singlet probability distributions. In the application of the MI theory to correlated mutations analysis, the properties x and y are the types of amino acids, and the random variables X and Y designate particular positions/sites along the sequence. If X and Y are independent (i.e., the types of amino acids observed at sequence positions X and Y are not correlated), then the joint probability $P(X=x, Y=y)$ reduces to the product $[P(X=x)P(Y=y)]$, which defines the lower bound, 0, for $I(X, Y)$. In the other extreme case of a random variable X completely determined by Y (or vice versa), the joint probability reduces to the singlet probability, and eqn [16] simply becomes the entropy of X (or Y), which is the largest MI value obtainable for the examined system. The latter case corresponds to a perfect coevolution of the two residues.

9.3.4.1.2 HIV-1 Protease sequence and structure correlations

HIV-1 protease has been under extensive study for decades as an important target for AIDS therapy.¹⁰⁴ This enzyme is a homodimer of 99 residues per monomer, with the active site located at the core of the dimerization interface. Numerous HIV-1 protease inhibitors have been developed and used in U.S. Food and Drug Administration-approved drugs, although these efforts have been impeded by the rapid mutations of the enzyme that exhibit multidrug resistance (MDR).^{105,106} Compared to mutations arising from phylogenetic variations, drug-induced mutations usually occur at much higher rates and over a broader range of residues, depending on the selective pressure imposed by drug treatment.

Our group conducted a two-prong study.¹⁰³ First, we retrieved the covariance patterns of amino acids in different sequences (HIV-1 protease mutants) from the Stanford HIV Drug Resistance Database.¹⁰⁷ Second, we performed a GNM analysis of the global dynamics of the protease to determine the potentially functional mechanisms of collective motions. This study demonstrated (1) the possibility of distinguishing between the correlated substitutions resulting from neutral mutations and those induced by MDR upon appropriate clustering analysis of sequence covariance data and (2) a connection between global dynamics and functional substitution of amino acids.

The Stanford HIV Drug Resistance Database provides a wealth of information on HIV-1 reverse transcriptase and HIV-1 protease mutants, annotated with the treatment history of patients from whom the virus samples have been drawn. We retrieved the sequence data on HIV-1 mutants, which have been organized into two groups, namely those from patients treated by different regimens (by at least one drug) and those from untreated patients. Sequences in each group were optimally aligned, containing approximately 8000 sequences with 99 columns (residues), and then subjected to MI analysis. A 99×99 symmetric matrix, referred to as the MI map, was obtained, which in turn was analyzed using a spectral graph partitioning method (minimization of normalized cut).¹⁰⁸ The method permitted us to identify two clusters of residue pairs with maximal correlations within each cluster and minimal correlations across the clusters (**Figure 11(a)**). Closer examination of the residues in each cluster revealed that one of the clusters was essentially composed of residues involved in MDR acquisition (MDR cluster), as evidenced by comparison with experimentally known sites;¹⁰³ the other included those residues differentiated in the various HIV-1 protease subtypes,¹⁰⁹ and it was called the phylogenetic variation (PGV) cluster. The residues identified by this type of correlated mutation analysis to belong to these two clusters are shown in blue and orange, respectively, along the sequence alignment of representative members from subgroups in **Figure 11(c)** and explicitly shown by the same colors in the ribbon diagram displayed in **Figure 11(d)**. The high correlation between the MDR residues is consistent with their cooperative conferral of multiple drug resistance.¹¹⁰ Clustering on the MI map generated for the untreated data, on the other hand, yielded one single cluster – the PGV (**Figure 11(b)**).

Figure 11(c) displays the global mode profile obtained by the GNM for HIV-1 protease. Calculations were performed for the dimer, but results are shown for only a monomer because the two monomers exhibit the same profile. The question we raised, then, was whether the residues belonging to the two clusters exhibited any distinctive features with regard to the structure and dynamics of the enzyme. As shown in **Figure 11(c)**, the MDR cluster residues essentially occupy sites close to the central symmetry axis of the enzyme. Importantly, their positions overlap with the global hinge region identified by the GNM analysis of the enzyme (minima in **Figure 11(c)**), and they may thus be involved in the coordination of the global motions. The PGV residues, on the other hand, occupy solvent-exposed and highly mobile regions, which may presumably sustain substitutions without significant consequences on the stability of the structure or the mechanism of global motions.

This study thus revealed the close link between global dynamics and sequence correlation patterns. In summary, drug-resistance mutations occurred at regions close to the global hinge, which neighbors the catalytic center and cooperatively affects the global dynamics of the protease. The correlated mutations corresponding to phylogenetic variations, on the other hand, have preferentially occurred on relatively more mobile (or less constrained) regions of the structure such that the key catalytic and mechanical sites have not been affected, consistent with basic evolutionary selection principles.

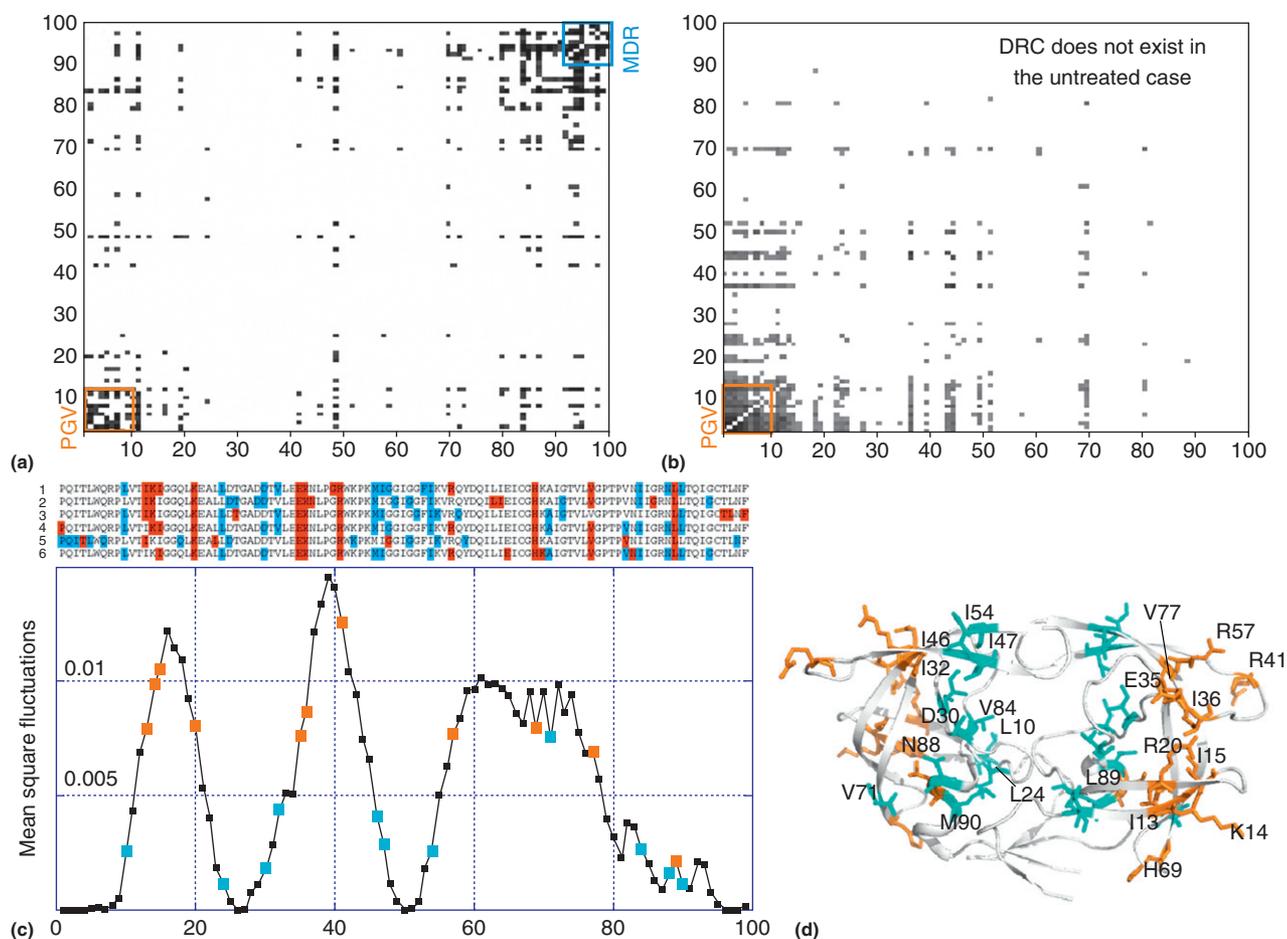


Figure 11 Results from mutual information (MI) analysis and GNM calculations for HIV-1 protease. (a and b) MI maps for two data sets of structures: sequences (mutants) from patients who have been subjected to a given anti-HIV-1 drug treatment regimen (a) and those from untreated patients (b). Residue numbers along the two axes have been reordered (columns/rows shuffled) in both cases so as to achieve optimal clustering of subsets of correlated residues. For visual clarity, the top ranking (highest MI) pairs of amino acids (500 out of a total of 99×99 pairs) are displayed. Two distinctive types of correlated mutations are seen in panel (a), clustered at the bottom left and top right portions of the map. The second (top right) is nonexistent in panel (b). The latter is found, upon comparison with experimental data, to predominantly include the pairs of correlated substitutions that have been induced in response to drug treatment (multidrug resistance (MDR) mutations). The second cluster (bottom left), on the other hand, refers to the evolutionary changes observed between HIV-1 protease subtypes (phylogenetic variations (PGV)). (c) GNM slowest mode profile as a function of residue index. The sequence alignment above the curve refers to six data sets corresponding to different combinations of drug treatment (1, treated; 2, untreated; 3, treated by IDV only; 4, treated by IDV and other drugs; 5, treated by NFV only; and 6, treated by NFV and other drugs). Two extreme subsets of 12 residues rank-ordered according to the spectral bipartitioning of the MI matrix are shown in blue (DRC) and orange (PGV). (d) The locations of the two clusters on the three-dimensional structure of HIV-1 protease. The MDR sites are shown in cyan, and the PGV sites are shown in orange. See Liu et al.¹⁰³ for more details. Ribbon diagram generated with PyMOL.

9.3.4.1.3 Hsp70 ATPase domain

The change in the conformation of Hsp70 ATPase domain structure upon NEF binding was discussed previously. A molecular understanding of the mechanism of NEF assisting the nucleotide exchange process requires a more detailed investigation of the two groups of functional residues: those residues at the NEF binding interface and those at the nucleotide binding pocket. We have conducted a combined analysis of the sequential, structural, and dynamical properties of these residues.⁶⁸ **Figure 12(a)** shows the structural locations of NEF binding residues (based on three mammalian NBD-NEF complexes: BAG, HspBP1, and Sse1) and the nucleotide binding residues. The NEF binding residues are mostly exposed and abundant in polar amino acids, whereas the nucleotide binding residues are buried in the core region and contain a large number of glycines (Gly12, Gly201, Gly202, Gly203, Gly230, Gly338, and Gly339). GNM calculations reveal the fundamentally distinct character of the two groups of residues with regard to the dynamic properties of the NBD. **Figure 12(b)** highlights the enhanced mobility of NEF-contacting residues (peaks in the global mode profile). The nucleotide binding residues, on the other hand, usually occupy minima in the mode profile, indicative of their constrained mobility in the soft modes. Subdomain IIB contains a major portion of NEF binding residues, and these residues – and subdomain IIB in general – are the most mobile region of the molecule. Their high mobility allows them to efficiently recognize the co-chaperone NEF; on the other hand, the abundance of glycines at the nucleotide binding site is in accord with the global hinge-bending role of this site, where residues are rotationally flexible but spatially immobile.

Comparison of the intrinsic dynamics (weighted average of the 10 slowest modes; **Figure 12(b)**) between the unbound and bound forms suggests that the mobility of NEF binding residues is suppressed upon NEF binding, although the overall shape of the global mode is closely maintained. The depression in the mobility of NEF binding sites essentially indicates the stabilization of the NEF-bound (open) conformation. In particular, certain residues in subdomain IIB (Arg247, Lys248, Lys250, Asp285, Ser286, Gly290, and Asp292) and in subdomain IA (Asp32 and Gln33) exhibit the largest reduction in mobility. Despite these differences, the similarity of the two curves is also apparent in that the loci of peaks and minima are maintained. Note that both curves are normalized and represent a probability distribution of mobilities in this representative subset of soft modes.

Figure 12(c) shows an interesting relation between the level of conservation of a given residue and its intrinsic dynamics. The level of conservation was quantified using the evolutionary trace (ET) method.¹¹¹ More conserved residues have higher ET rank (i.e., rank 1 is the most conserved). Overall, the results disclose an inverse correlation between the extent of mobility and the level of conservation. Nucleotide binding residues are usually highly conserved, whereas NEF binding residues tend to be more variable (except for a few outliers such as Asp292 and Leu274).

Closer examination of mutational trends at the NEF binding sites, however, reveals that the mutations are not random; instead, they are correlated with each other. The MI map shown in **Figure 12(d)** indicates that NEF-contacting

residues account for a large proportion of top-ranking correlated pairs. The bars below the MI map represent averages over each column, revealing that most correlated residues essentially coincide with NEF binding residues. **Figures 12(e)** and **(f)** provide a closer view of two portions of the MI map, which involve residues 246–305 and 16–75. These two regions contain 90% of all NEF-contacting residues, and most of them are highly correlated with each other. In **Figure 12(e)**, note the remarkably high coevolutionary patterns of the pairs Thr265-Thr273, Arg258-Tyr288, and Thr273-Tyr288. In **Figure 12(f)**, the pairs Glu27-Arg258, Glu27-Tyr288, Gln33-Thr273, and Gln33-Glu283 exhibit remarkably high correlations, despite their long-distance separation on the structure. These results suggest that the correlated mutations at those sites are associated with allosteric effects and may affect binding of NEF residues at distant sites. These residues presumably coevolve to perform specific NEF-dependent recognition and binding activities, which differ between Hsp70 subfamilies. Interestingly, Bukau and co-workers¹¹² found that a signature loop (Ala276-Arg302) in DnaK subdomain IIB along with three interacting residue pairs (Met259-Val59, Glu264-Arg56, and Glu267-Lys55) at the nucleotide binding cleft mark the divergence of Hsp70 subfamilies such that variations of these residues may account for the selectivity of NEF binding (in this case, GrpE and BAG). The phylogeny-based variations of these residues are therefore contributing to the strong coevolution detected within these regions.

9.3.4.2 Exploring Transitions between Functional States

9.3.4.2.1 Using adaptive ANM

To explore the suitability of aANM for predicting the transitions $T \rightarrow R \rightarrow R'' \rightarrow T$ of GroEL rings, we first examined the transition $R'' \rightarrow T$ before proceeding to the intact chaperonin. Results for the transition $R'' \rightarrow T$ are shown in **Figure 13**. **Figure 13(a)** displays the change in energy involved in different F_{\min} . It can be seen that lower F_{\min} values allow for larger excursions away from the targeted direction by recruiting relatively smaller numbers of low-frequency modes (**Figure 13(c)**). They consequently require a larger number of steps to be undertaken to reach the target, whereas the accompanying energy increase is relatively small. Higher F_{\min} values, on the other hand, permit the target to be reached faster but with a higher energy cost. The limit $F_{\min} = 1$ corresponds to pure interpolation by recruiting all modes. The reaction coordinate in **Figure 13(a)** is the projection of the cumulative displacement $\mathbf{v}^{(n)} = \mathbf{R}^{(n)} - \mathbf{R}_A^{(0)}$ on the original distance vector $\mathbf{d}^{(0)}$ – that is, $x(n) = \mathbf{d}^{(0)} \cdot \mathbf{v}^{(n)} / |\mathbf{d}^{(0)}|^2$, with R'' and T representing the respective limits $x(n) = 0$ and 1 . The peak in the energy profile tends to be closer to the T state, especially when lower F_{\min} values (which entail lower energy barriers) are adopted. This may be related to the recruitment of higher modes (steeper ascent along the energy surface) near state T , as opposed to the first mode near R'' .

A series of conformations visited along the transition pathway are displayed in **Figure 13(b)** for $F_{\min} = 0.5$. If the equatorial domain is fixed as a reference, the apical domain closes down and twists about the axial direction, and the intermediate domain moves away. An interesting

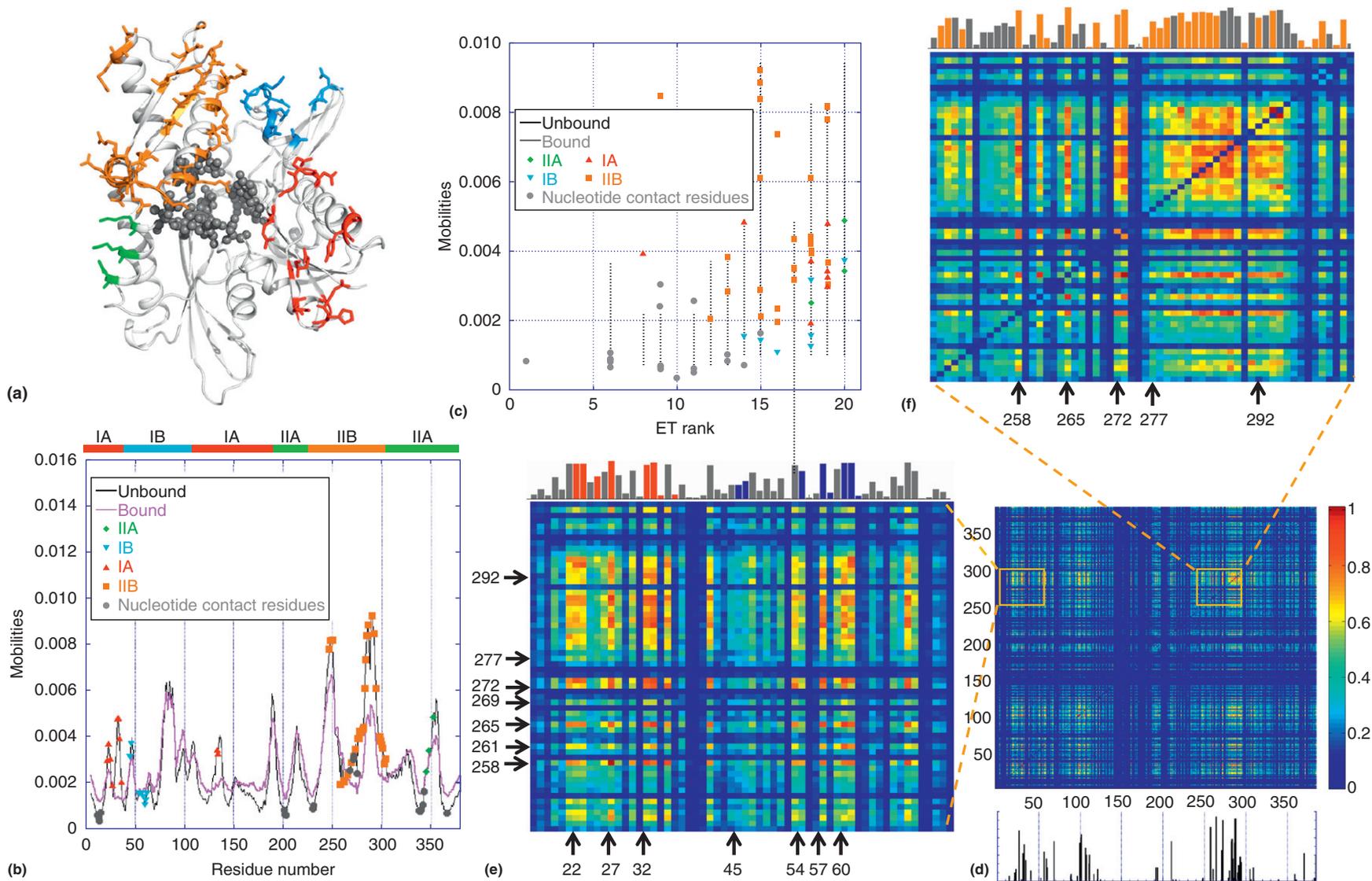


Figure 12 Bridging between sequence correlations and structural dynamics of Hsp70 ATPase domain (NBD), in relation to its NEF binding properties. (a) NEF-contacting residues are shown in stick representation and colored by subdomains (see **Figure 2(a)**). Nucleotide binding residues (G12-Y15, G201-G203, G230, E231, E268, K271, R272, S275, G338-S340, R342, I343, and D366) are shown as gray spheres. (b) Weighted-average mobility profiles based on the top-ranking (softest) 10 GNM modes, calculated for the unbound NBD (black) and for the NEF-bound (purple) structures, averaged over three mammalian complexes (BAG, HspBP1, and Sse1). Labels are colored in accord with the scheme in panel (a). The horizontal bars along the upper abscissa indicate the ranges of the four subdomains IA, IB, IIA, and IIB. Bound and unbound NBD exhibit similar profiles, consistent with the intrinsic dynamics of the NBD irrespective of co-chaperone (NEF) binding. NEF-binding residues tend to occupy peak positions, whereas nucleotide-binding residues are at the minima (global hinge site). (c) Evolutionary trace (ET) rank vs. mobilities (from panel (b)) for all residues in the Hsp70 ATPase domain. (d) MI map for the Hsp70 ATPase domain sequence. The color-coded bar on the right indicates the level of correlation between the evolution of residue pairs. The map displays the contribution of each residue to the most correlated residue pairs (top 5%, 722 pairs) in the MI matrix. NEF binding sites exhibit the highest correlations. (e and f) Close-up views of two regions of MI maps, which contain a large number of NEF-binding residues: residues 246–305 (subdomain IIB) and 16–75 (containing NEF-contacting segments in subdomains IA and IB) (e) and residues 246–305 (f). The bar plot displays the average MI (over all entries in a given column) for residues in these segments; those corresponding to NEF binding residues are colored by their subdomain. For more details, see Liu et al.⁶⁸ Ribbon diagram generated with PyMOL.

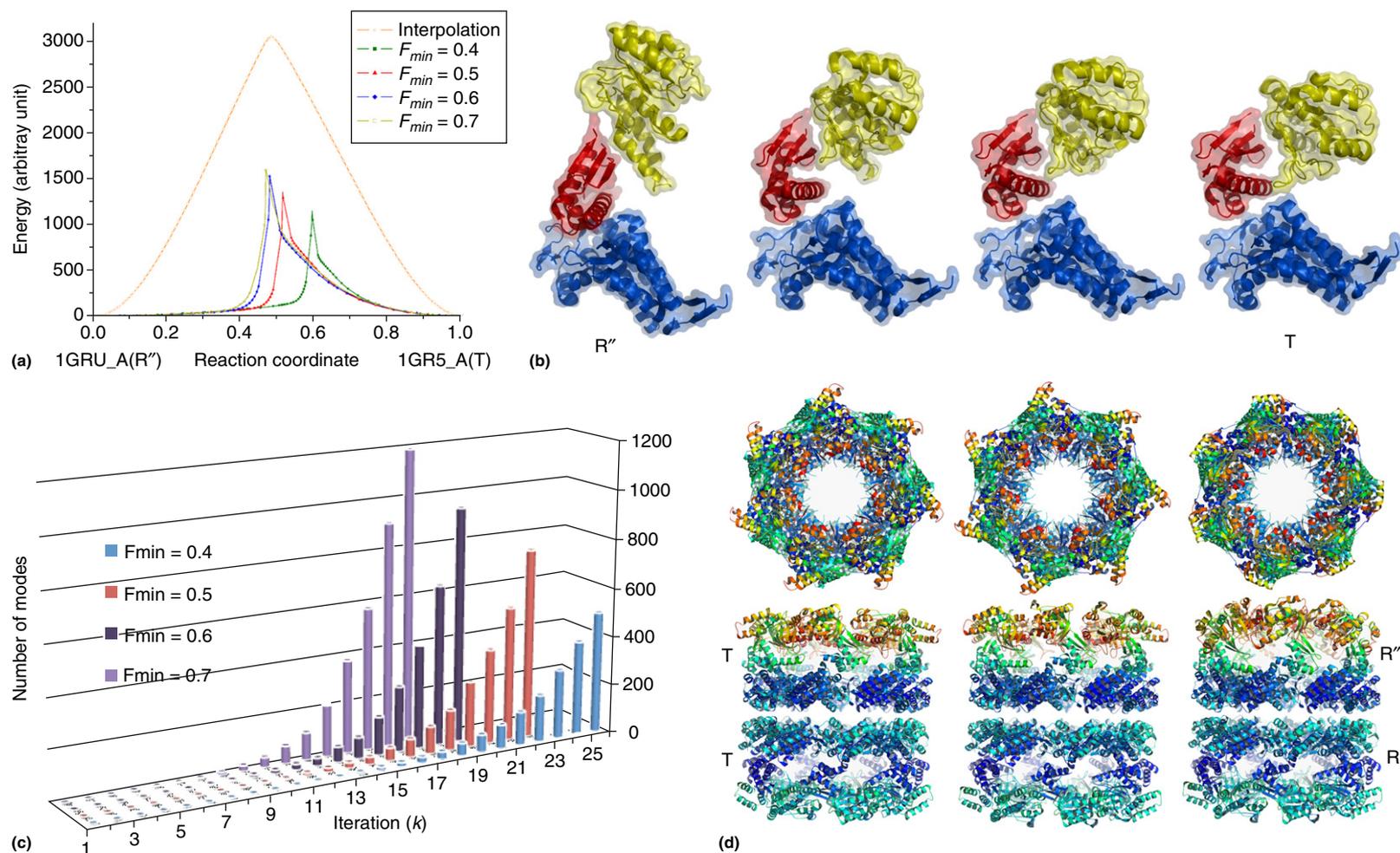


Figure 13 Pathways sampled by aANM. (a) The energy profile for alternative pathways in arbitrary units for a single subunit of GroEL from R'' to T. Note the significantly lower energy barrier compared to the interpolation (orange curve) between the endpoints. The reaction coordinate refers to the normalized projection of the instantaneous displacement on the original distance vector. (b) Series of conformations sampled along the reaction coordinate. The diagrams are colored by domains (equatorial, blue; intermediate, red; apical, yellow). (c) The number of modes recruited in each iteration in various F_{min} from R'' to T. (d) Top and side views of intact GroEL sampled along the transition from T/T to R''/R, colored according to the mobilities of residues (red, most mobile; blue, almost fixed). Note that the equatorial domains of *cis* ring subunits are almost fixed, whereas the largest motions occur at the apical domains of the same subunits. Image generated with UCSF Chimera.¹⁴²

observation is the sequential order of events. First, the equatorial and intermediate domains almost stick to each other and move coherently as a single rigid body, whereas the apical domain undergoes an upward tilting and simultaneous twisting. The movement of the apical domain is completed in the first half of the transition pathway from R' to T. Then, slight rearrangements in the relative positions of the equatorial and intermediate domains occur, which expose the top portion of the apical domain to bind the substrate and the co-chaperonin GroES. This sequence of events is consistent with the two-stage transition explored by targeted MD simulations.¹¹³

The number of modes involved in the transition between substates is an issue that has been raised in a number of studies.^{114,115} Figure 13(c) displays the number of modes recruited in each iteration during the reconfiguration of a given subunit from substate R' to substate T. An important observation is the increased involvement of higher frequency modes as we proceed away from the original state. In other words, the slow modes play an important role during the initial stages of deformation, and they continue to play a role throughout the entire trajectory, although they are gradually complemented by increasingly larger subsets of higher-frequency modes. The dominance of low-frequency modes is consistent with the previously noted driving role of global motions in GroEL allostery.¹¹⁶

The previously performed calculations for a single subunit demonstrated that the transition pathways between the functional forms of the single subunit can be delineated by the aANM algorithm. We then explored the allosteric transitions of the intact chaperonin. Figure 13(d) displays three snapshots along the evolution of the intact structure from T to R viewed from the top and the side. The *trans* ring is generally observed to undergo moderate changes between the T and R states. The *cis* ring, on the other hand, undergoes concerted rotations and extensions at the apical and intermediate domains. Notably, the intermediate domains move toward the cleft between neighboring equatorial domains, whereas the apical domains extend along the vertical (cylindrical axis) direction accompanied by a rotation about the same axis. These motions result in the enlargement of the central cavity along with the exposure of the flexible loop K15-T36 on the apical domains of the *cis* ring to bind the GroES.

Notably, the aANM trajectories of the intact GroEL structures retain the sevenfold rotational symmetry. In each ring, the movements of all subunits are identical and coherent. We noted that not all slow modes in the lower end of the spectrum contribute to the transition. Instead, a subset of modes that induce cylindrically symmetric changes effectively contribute. For example, in the first iteration, although 51 modes are involved, the contributions of modes 2–9 are almost zero, and there are large contributions from the nondegenerate

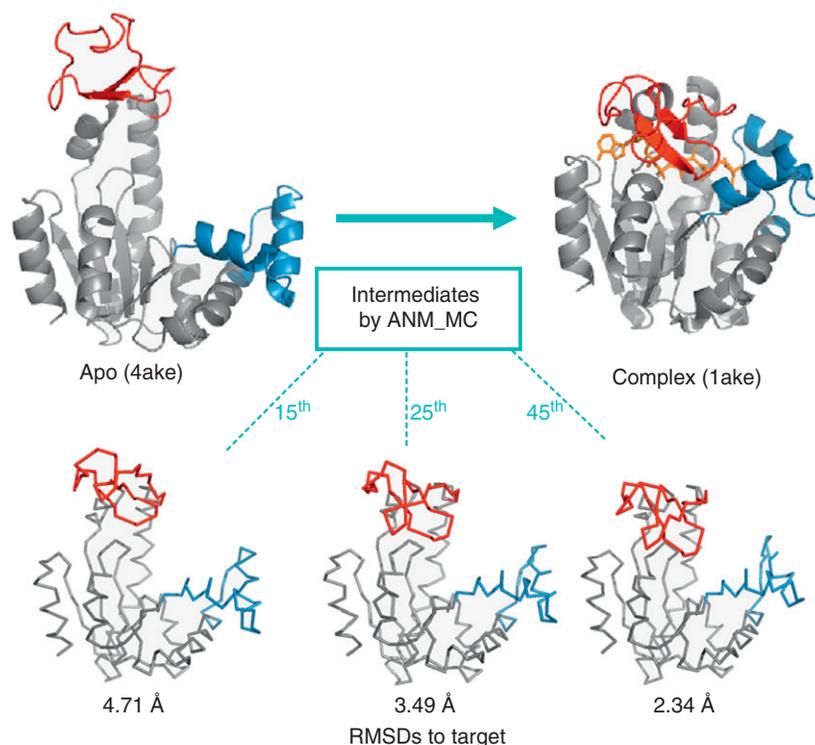


Figure 14 Conformational change between the apo and ligand-bound forms of adenylate kinase observed by ANM_MC simulations. Crystal structures of apo adenylate kinase (PDB code¹⁵⁰) and its complex (1AKE¹⁵¹) with AP₅A (orange) are shown at the top. AK comprises three domains: a core domain (gray), an ATP binding or LID domain (red), and an AMP binding domain (blue). Three intermediates detected³⁴ on the transition pathway of AK (bottom) are from the 15th (with 2.96 Å RMSD to 2AK2 crystal structure¹⁵²), 25th (2.36 Å RMSD to 1DVR A chain¹⁵³), and 45th (2.38 Å RMSD to 1E4Y A chain¹⁵⁴) iterations of ANM_MC simulation. The initial RMSD of 7.13 Å between open and closed conformations decreases to 2.34 Å at the 45th iteration. The closure of the LID precedes that of the AMP binding domain. Image generated with PyMOL.

modes 1, 10, 11, 28, and 51. Our results support the MWC view^{24,117} of preexisting cooperative modes being selectively used for controlling allostery. Likewise, GroEL engineering experiments provide evidence for the concerted nature of the allosteric transition.¹¹⁸

9.3.4.2.2 Using Monte Carlo simulations integrated with collective modes

Inasmuch as the information for the conformational changes that occur upon ligand binding is implicit in the folded structure of the unbound state, ENMs present a computationally efficient means to extract this information.^{18,22,23} However, new hybrid techniques have to be devised in order to extract realistic data along conformational transitions described by the purely harmonic collective modes of ANM.^{119,120} Energy minimization – by molecular mechanics or MD or Monte Carlo (MC) simulations – presents a plausible route.

A hybrid methodology combining MC simulations with the collective modes from ANM has been proposed for analyzing conformational transitions of proteins.¹²¹ The original CG MC technique^{122–124} is based on the virtual bond model,¹²⁵ with every residue being represented by two sites – one on the α carbon and the other at the side chain centroid. Knowledge-based potentials extracted from crystal structures of folded proteins are used for calculating the energy of a conformation including both short- and long-range potentials.^{126,127} The original MC algorithm based on local moves of randomly chosen sites has been modified into an iterative procedure (ANM-MC), in which a perturbation along an ANM harmonic mode is followed by a certain number (100–1000) of MC local moves to relax the deformed structure. In each cycle, ANM modes are updated, and the protein is deformed (preferably with an RMSD of ~ 0.2 Å) along the collective mode presenting the highest correspondence to the open-to-closed transition direction.

Using the ANM-MC algorithm, the initial RMSDs of 7.13 and 3.55 Å between the apo and bound states of adenylate kinase (AK) and Hb have been decreased to 2.3 and 1.9 Å, respectively. For AK, certain intermediates on the transitional pathway (shown in Figure 14) have been identified by comparison with the other available crystal structures of AK, in line with previous studies based on ENMs.¹¹⁹ During the transition, the closure of the ATP binding domain (also called the LID) over the core domain precedes the AMP binding domain closure, as originally noted by Temiz et al.¹²⁸ Hemoglobin presents three major conformational states: a tense, apo state (T), an O₂-bound relaxed state (R), and a CO-bound relaxed state (R2). ANM-MC has indicated that the R state is an intermediate along the transition from T to R2, similar to a previous computational study.¹²⁹ The allosteric T-to-R2 transition has also been achieved down to an RMSD of 2.4 Å by using only the slowest mode from ENM.⁵⁴ For both proteins, the first two modes were effective in driving the collective conformational changes during the first half of the simulations, after which none of the collective modes exhibited high overlaps with the transition direction.

A major achievement in this line of work would be to devise techniques that could predict plausible bound

conformations using only the apo structures. Preliminary results with ANM-MC indicate that successful end-structure predictions could be made by constraining the radius of gyration of some hinge-bending proteins. An ENM-based approach making use of several distance constraints for the end state has proven successful in maintaining the associated transitions for a set of 16 proteins.^{130,131}

9.3.4.3 Generation of Conformers for Ensemble Docking

Inasmuch as protein conformational dynamics emerges as a key player in ligand binding, the inclusion of receptor flexibility has been emphasized in structure-based drug design for obtaining higher success rates compared to those of studies with rigid target structure. Several alternative methodologies exist for incorporating protein flexibility (side chain and/or main chain) in docking but at the expense of increased CPU requirements.^{132,133} The ensemble docking algorithm presents a computationally feasible solution to this problem, in which a set of conformers from experiments (X-ray and NMR) or computations (MD and MC simulations, normal modes) describe the intrinsic flexibility of the receptor for docking.¹³⁴

Because MD simulations become computationally expensive for the conformational sampling of large systems, ENMs may provide a possible alternative. Realistic conformers for

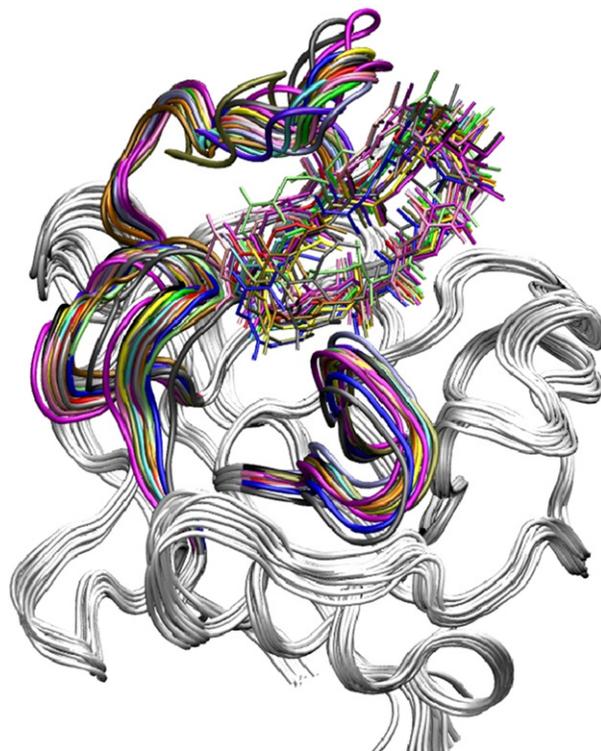


Figure 15 Cyclophilin A conformers generated by atomistic ANM and subsequent energy minimization. Models are shown together with their docked ligand cyclosporine A. The starting protein conformation is extracted from the crystal structure (1BCK) for the CypA-cyclosporine A complex.¹⁵⁵ AutoDock v.4¹⁵⁶ software is used for docking (rigid protein, flexible ligand with 17 rotatable bonds) with the Lamarckian genetic algorithm. Image generated with VMD.¹⁵⁷

cyclophilin A (CyPA, 165 residues) have been generated by deforming the native structure along low-frequency ANM modes and performing subsequent energy minimization using implicit salvation.¹³⁵ In a previous study on CyPA, a graph-theoretical approach was used to generate conformers by considering the flexibility of each loop region independently, in contrast to ANM modes that describe collective deformations for the global structure including the loops.¹³⁶ The conformers based on either atomistic or residue-based (uniform coarse-graining) ANM present reasonable energies (comparable or lower than that of the starting X-ray structure) and internal geometries (i.e., torsional angle preferences and atom-atom contacts). All CyPA conformers have been subject to an ensemble docking experiment to the cyclic ligand cyclosporine A, as a result of which some conformers presented stronger binding modes, again compared to the crystal structure of the complex (Figure 15). For the specific case of CyPA, conformers generated from atomic and CG ANM modes are equally successful in producing docked poses of the ligand. The conformers exhibit moderate conformational changes with maximum RMSDs of approximately 2 Å for the loops surrounding the binding pocket. Proteins presenting larger conformational changes (overall or loops) may be dealt with by performing ANM and energy minimizations in cycles. Conformations along TIM's loop closure have been mimicked by performing this cyclic procedure along first mode.⁴⁶

Cavasotto and co-workers¹³⁷ used the ENM for ligand docking to cAMP-dependent protein kinase, in which relatively high-frequency modes (mode numbers >140) have been considered for loop rearrangements at binding sites. Target flexibility introduced along the second normal mode for metalloproteinases has been effective for determining the inhibitor binding modes.¹³⁸ ENMs are also being implemented in protein-protein docking algorithms for describing global conformational changes of partners.^{139,140} Thus, an increasing number of studies support the utility of using these simplified models for efficiently sampling the conformational space of the target protein and generating most readily accessible conformations that would facilitate the search for the optimal binding pose of the lead compounds.

Acknowledgment

Ivet Bahar is supported by NIH grant 5R01LM007994-07.

References

- [1] Changeux, J. P.; Edelstein, S. J. Allosteric mechanisms of signal transduction. *Science* **2005**, *308*, 1424–1428.
- [2] Houseweart, M. K.; Cleveland, D. W. Intermediate filaments and their associated proteins: Multiple dynamic personalities. *Curr. Opin. Cell Biol.* **1998**, *10*, 93–101.
- [3] Bahar, I.; Lezon, T. R.; Yang, L. W.; Eyal, E. Global dynamics of proteins: Bridging between structure and function. *Annu. Rev. Biophys. Biomol. Struct.* **2010**, *39*, 23–42.
- [4] Bahar, I.; Chennubhotla, C.; Tobi, D. Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. *Curr. Opin. Struct. Biol.* **2007**, *17*, 633–640.

- [5] Bahar, I.; Lezon, T. R.; Bakan, A.; Shrivastava, I. H. Normal mode analysis of biomolecular structures: Functional mechanisms of membrane proteins. *Chem. Rev.* **2010**, *110*, 1463–1497.
- [6] Tokuriki, N.; Tawfik, D. S. Protein dynamism and evolvability. *Science* **2009**, *324*, 203–207.
- [7] Hammes-Schiffer, S.; Benkovic, S. J. Relating protein motion to catalysis. *Annu. Rev. Biochem.* **2006**, *75*, 519–541.
- [8] Smock, R. G.; Gierasch, L. M. Sending signals dynamically. *Science* **2009**, *324*, 198–203.
- [9] Cui, Q.; Bahar, I. *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems*; CRC Press: Boca Raton, FL, 2006.
- [10] Nicolay, S.; Sanejouand, Y. H. Functional modes of proteins are among the most robust. *Phys. Rev. Lett.* **2006**, *96*, 078104.
- [11] Tama, F.; Brooks, C. L. Symmetry, form, and shape: Guiding principles for robustness in macromolecular machines. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 115–133.
- [12] Doruker, P.; Jernigan, R. L.; Bahar, I. Dynamics of large proteins through hierarchical levels of coarse-grained structures. *J. Comput. Chem.* **2002**, *23*, 119–127.
- [13] Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* **2001**, *80*, 505–515.
- [14] Bahar, I.; Atilgan, A. R.; Erman, B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold. Des.* **1997**, *2*, 173–181.
- [15] Doruker, P.; Atilgan, A. R.; Bahar, I. Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: Application to alpha-amylase inhibitor. *Proteins* **2000**, *40*, 512–524.
- [16] Haliloglu, T.; Bahar, I.; Erman, B. Gaussian dynamics of folded proteins. *Phys. Rev. Lett.* **1997**, *79*, 3090–3093.
- [17] Hinsen, K. Analysis of domain motions by approximate normal mode calculations. *Proteins* **1998**, *33*, 417–429.
- [18] Tama, F.; Sanejouand, Y. H. Conformational change of proteins arising from normal mode calculations. *Protein Eng.* **2001**, *14*, 1–6.
- [19] Tirion, M. M. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* **1996**, *77*, 1905–1908.
- [20] Delarue, M.; Sanejouand, Y. H. Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: The elastic network model. *J. Mol. Biol.* **2002**, *320*, 1011–1024.
- [21] Yang, L. W.; Bahar, I. Coupling between catalytic site and collective dynamics: A requirement for mechanochemical activity of enzymes. *Structure* **2005**, *13*, 893–904.
- [22] Bakan, A.; Bahar, I. The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 14349–14354.
- [23] Tobi, D.; Bahar, I. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 18908–18913.
- [24] Monod, J.; Wyman, J.; Changeux, J. P. On the nature of allosteric transitions: A plausible model. *J. Mol. Biol.* **1965**, *12*, 88–118.
- [25] Yang, L.; Song, G.; Carriquiry, A.; Jernigan, R. L. Close correspondence between the motions from principal component analysis of multiple HIV-1 protease structures and elastic network modes. *Structure* **2008**, *16*, 321–330.
- [26] Yang, L. W.; Eyal, E.; Bahar, I.; Kitao, A. Principal component analysis of native ensembles of biomolecular structures (PCA_NEST): Insights into functional dynamics. *Bioinformatics* **2009**, *25*, 606–614.
- [27] Bhattacharya, A.; Kurochkin, A. V.; Yip, G. N.; Zhang, Y.; Bertelsen, E. B.; Zwietering, E. R. Allostery in Hsp70 chaperones is transduced by subdomain rotations. *J. Mol. Biol.* **2009**, *388*, 475–490.
- [28] Liu, Y.; Bahar, I. Toward understanding allosteric signaling mechanisms in the ATPase domain of molecular chaperones. *Pac. Symp. Biocomput.* **2010**, 269–280.
- [29] Ansari, A.; Berendzen, J.; Bowne, S. F.; Frauenfelder, H.; Iben, I. E.; Sauke, T. B.; Shyamsunder, E.; Young, R. D. Protein states and proteinquakes. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 5000–5004.
- [30] Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. The energy landscapes and motions of proteins. *Science* **1991**, *254*, 1598–1603.
- [31] Kitao, A.; Go, N. Investigating protein dynamics in collective coordinate space. *Curr. Opin. Struct. Biol.* **1999**, *9*, 164–169.
- [32] Bahar, I.; Rader, A. J. Coarse-grained normal mode analysis in structural biology. *Curr. Opin. Struct. Biol.* **2005**, *15*, 586–592.

- [33] Chennubhotla, C.; Rader, A. J.; Yang, L. W.; Bahar, I. Elastic network models for understanding biomolecular machinery: From enzymes to supramolecular assemblies. *Phys. Biol.* **2005**, *2*, S173–S180.
- [34] Yang, Z.; Majek, P.; Bahar, I. Allosteric transitions of supramolecular systems explored by network models: Application to chaperonin GroEL. *PLoS Comput. Biol.* **2009**, *5*, e1000360.
- [35] Flory, P. J. Statistical thermodynamics of random networks. *Proc. R. Soc. London* **1976**, *351*, 351–380.
- [36] Hinsen, K.; Thomas, A.; Field, M. J. Analysis of domain motions in large proteins. *Proteins* **1999**, *34*, 369–382.
- [37] Marques, O.; Sanejouand, Y. H. Hinge-bending motion in citrate synthase arising from normal mode calculations. *Proteins* **1995**, *23*, 557–560.
- [38] Grimes, R.; Lewis, J.; Simon, H. A shifted block Lanczos algorithm for solving sparse symmetric generalized eigenproblems. *SIAM J. Matrix Anal. Appl.* **1994**, *15*, 228–272.
- [39] Doruker, P.; Jernigan, R.; Navizet, I.; Hernandez, R. Important fluctuation dynamics of large protein structures are preserved upon coarse-grained renormalization. *Int. J. Quantum Chem.* **2002**, *90*, 822–837.
- [40] Isin, B.; Doruker, P.; Bahar, I. Functional motions of influenza virus hemagglutinin: A structure-based analytical approach. *Biophys. J.* **2002**, *82*, 569–581.
- [41] Tatulian, S. A.; Hinterdorfer, P.; Baber, G.; Tamm, L. K. Influenza hemagglutinin assumes a tilted conformation during membrane fusion as determined by attenuated total reflection FTIR spectroscopy. *EMBO J.* **1995**, *14*, 5514–5523.
- [42] White, J. M.; Hoffman, L. R.; Arevalo, J. H.; Wilson, I. A. Attachment and Entry of Influenza Virus into Host Cells: Pivotal Roles of Hemagglutinin. In *Structural Biology of Viruses*; Chiu, W.; Burnett, R. M.; Garcea, R. L., Eds.; Oxford University Press: New York, 1997; pp 80–104.
- [43] Doruker, P.; Jernigan, R. L. Functional motions can be extracted from on-lattice construction of protein structures. *Proteins* **2003**, *53*, 174–181.
- [44] Kurkuoglu, O.; Jernigan, R. L.; Doruker, P. Mixed levels of coarse-graining of large proteins using elastic network model succeeds in extracting the slowest motions. *Polymer* **2004**, *45*, 649–657.
- [45] Kurkuoglu, O.; Jernigan, R. L.; Doruker, P. Loop motions of triosephosphate isomerase observed with elastic networks. *Biochemistry* **2006**, *45*, 1173–1182.
- [46] Kurkuoglu, O.; Turgut, O. T.; Cansu, S.; Jernigan, R. L.; Doruker, P. Focused functional dynamics of supramolecules by use of a mixed-resolution elastic network model. *Biophys. J.* **2009**, *97*, 1178–1187.
- [47] Amadei, A.; Ceruso, M. A.; Di, N. A. On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations. *Proteins* **1999**, *36*, 419–424.
- [48] Cansu, S.; Doruker, P. Dimerization affects collective dynamics of triosephosphate isomerase. *Biochemistry* **2008**, *47*, 1358–1368.
- [49] Wang, Y.; Rader, A. J.; Bahar, I.; Jernigan, R. L. Global ribosome motions revealed with elastic network model. *J. Struct. Biol.* **2004**, *147*, 302–314.
- [50] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- [51] Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291–318.
- [52] van der Vaart, A.; Karplus, M. Simulation of conformational transitions by the restricted perturbation-targeted molecular dynamics method. *J. Chem. Phys.* **2005**, *122*, 114903.
- [53] Ma, J. Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure* **2005**, *13*, 373–380.
- [54] Xu, C.; Tobi, D.; Bahar, I. Allosteric changes in protein structure computed by a simple mechanical model: Hemoglobin T ↔ R2 transition. *J. Mol. Biol.* **2003**, *333*, 153–168.
- [55] Zheng, W.; Doniach, S. A comparative study of motor-protein motions by using a simple elastic-network model. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 13253–13258.
- [56] Zheng, W.; Brooks, B. R. Probing the local dynamics of nucleotide-binding pocket coupled to the global dynamics: Myosin versus kinesin. *Biophys. J.* **2005**, *89*, 167–178.
- [57] Zheng, W.; Brooks, B. R.; Thirumalai, D. Allosteric transitions in the chaperonin GroEL are captured by a dominant normal mode that is most robust to sequence variations. *Biophys. J.* **2007**, *93*, 2289–2299.
- [58] Bukau, B.; Horwich, A. L. The Hsp70 and Hsp60 chaperone machines. *Cell* **1998**, *92*, 351–366.
- [59] Young, J. C.; Agashe, V. R.; Siegers, K.; Hartl, F. U. Pathways of chaperone-mediated protein folding in the cytosol. *Nature Rev. Mol. Cell Biol.* **2004**, *5*, 781–791.
- [60] Hartl, F. U.; Hayer-Hartl, M. Converging concepts of protein folding *in vitro* and *in vivo*. *Nature Struct. Mol. Biol.* **2009**, *16*, 574–581.
- [61] Polier, S.; Dragovic, Z.; Hartl, F. U.; Bracher, A. Structural basis for the cooperation of Hsp70 and Hsp110 chaperones in protein folding. *Cell* **2008**, *133*, 1068–1079.
- [62] Sondermann, H.; Scheutler, C.; Schneider, C.; Hohfeld, J.; Hartl, F. U.; Moarefi, I. Structure of a Bag/Hsc70 complex: Convergent functional evolution of Hsp70 nucleotide exchange factors. *Science* **2001**, *291*, 1553–1557.
- [63] Kabani, M. Structural and functional diversity among eukaryotic Hsp70 nucleotide exchange factors. *Protein Pept. Lett.* **2009**, *16*, 623–660.
- [64] Harrison, C. J.; Hayer-Hartl, M.; Di, L. M.; Hartl, F.; Kuriyan, J. Crystal structure of the nucleotide exchange factor GrpE bound to the ATPase domain of the molecular chaperone DnaK. *Science* **1997**, *276*, 431–435.
- [65] Alberti, S.; Esser, C.; Hohfeld, J. BAG-1: a nucleotide exchange factor of Hsc70 with multiple cellular functions. *Cell Stress Chaperones* **2003**, *8*, 225–231.
- [66] McLellan, C. A.; Raynes, D. A.; Guerriero, V. HspBP1, an Hsp70 cochaperone, has two structural domains and is capable of altering the conformation of the Hsp70 ATPase domain. *J. Biol. Chem.* **2003**, *278*, 19017–19022.
- [67] Andreasson, C.; Fiaux, J.; Rampelt, H.; Mayer, M. P.; Bukau, B. Hsp110 is a nucleotide-activated exchange factor for Hsp70. *J. Biol. Chem.* **2008**, *283*, 8877–8884.
- [68] Liu, Y.; Gierasch, L. M.; Bahar, I. Role of Hsp70 ATPase domain intrinsic dynamics and sequence evolution in enabling its functional interactions with NEFs. *PLoS Comput. Biol.* **2010**, *6*, e1000931.
- [69] Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. A* **1976**, *32*, 922–923.
- [70] Schnackez, K. D.; Gracy, R. W. Probing the catalytic sites of triosephosphate isomerase by ³¹P-NMR with reversibly and irreversibly binding substrate analogues. *Eur. J. Biochem.* **1991**, *199*, 231–238.
- [71] Williams, J. C.; McDermott, A. E. Dynamics of the flexible loop of triosephosphate isomerase: The loop motion is not ligand gated. *Biochemistry* **1995**, *34*, 8309–8319.
- [72] Brown, F. K.; Kollman, P. A. Molecular dynamics simulations of 'loop closing' in the enzyme triose phosphate isomerase. *J. Mol. Biol.* **1987**, *198*, 533–546.
- [73] Joseph, D.; Petsko, G. A.; Karplus, M. Anatomy of a conformational change: Hinged 'lid' motion of the triosephosphate isomerase loop. *Science* **1990**, *249*, 1425–1428.
- [74] Massi, F.; Wang, C.; Palmer III, A. G. Solution NMR and computer simulation studies of active site loop motion in triosephosphate isomerase. *Biochemistry* **2006**, *45*, 10787–10794.
- [75] Amadei, A.; Linssen, A. B.; Berendsen, H. J. Essential dynamics of proteins. *Proteins* **1993**, *17*, 412–425.
- [76] Thanki, N.; Zeelen, J. P.; Mathieu, M.; Jaenicke, R.; Abagyan, R. A.; Wierenga, R. K.; Schliebs, W. Protein engineering with monomeric triosephosphate isomerase (monoTIM): The modelling and structure verification of a seven-residue loop. *Protein Eng.* **1997**, *10*, 159–167.
- [77] Jernigan, R. L.; Yang, L.; Song, G.; Kurkuoglu, O.; Doruker, P. Elastic Network Models of Coarse-Grained Proteins Are Effective for Studying the Structural Control Exerted over Their Dynamics. In *Coarse-Graining of Condensed Phase and Biomolecular Systems*; Voth, G. A., Ed.; CRC Press: Boca Raton, FL, 2009; pp 237–254.
- [78] Konuklar, F. A.; Aviyente, V.; Haliloglu, T. Coupling of structural fluctuations to deamidation reaction in triosephosphate isomerase by Gaussian network model. *Proteins* **2006**, *62*, 715–727.
- [79] Berk, V.; Cate, J. H. Insights into protein biosynthesis from structures of bacterial ribosomes. *Curr. Opin. Struct. Biol.* **2007**, *17*, 302–309.
- [80] Munro, J. B.; Sanbonmatsu, K. Y.; Spahn, C. M.; Blanchard, S. C. Navigating the ribosome's metastable energy landscape. *Trends Biochem. Sci.* **2009**, *34*, 390–400.
- [81] Frank, J.; Agrawal, R. K. A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature* **2000**, *406*, 318–322.

- [82] Tama, F.; Valle, M.; Frank, J.; Brooks III, C. L. Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 9319–9323.
- [83] Wang, Y.; Jernigan, R. L. Comparison of tRNA motions in the free and ribosomal bound structures. *Biophys. J.* **2005**, *89*, 3399–3409.
- [84] Trylska, J.; Tozzini, V.; McCammon, J. A. Exploring global motions and correlations in the ribosome. *Biophys. J.* **2005**, *89*, 1455–1463.
- [85] Kurkcuoglu, O.; Doruker, P.; Sen, T. Z.; Kloczkowski, A.; Jernigan, R. L. The ribosome structure controls and directs mRNA entry, translocation and exit dynamics. *Phys. Biol.* **2008**, *5*, 046005.
- [86] Valle, M.; Zavialov, A.; Sengupta, J.; Rawat, U.; Ehrenberg, M.; Frank, J. Locking and unlocking of ribosomal motions. *Cell* **2003**, *114*, 123–134.
- [87] Agrawal, R. K.; Heagle, A. B.; Penczek, P.; Grassucci, R. A.; Frank, J. EF-G-dependent GTP hydrolysis induces translocation accompanied by large conformational changes in the 70S ribosome. *Nature Struct. Biol.* **1999**, *6*, 643–647.
- [88] Kurkcuoglu, O.; Kurkcuoglu, Z.; Doruker, P.; Jernigan, R. L. Collective dynamics of the ribosomal tunnel revealed by elastic network modeling. *Proteins* **2009**, *75*, 837–845.
- [89] Lu, J.; Deutsch, C. Folding zones inside the ribosomal exit tunnel. *Nature Struct. Mol. Biol.* **2005**, *12*, 1123–1129.
- [90] Berisio, R.; Schluenzen, F.; Harms, J.; Bashan, A.; Auerbach, T.; Baram, D.; Yonath, A. Structural insight into the role of the ribosomal tunnel in cellular regulation. *Nature Struct. Biol.* **2003**, *10*, 366–370.
- [91] Halperin, I.; Wolfson, H.; Nussinov, R. Correlated mutations: Advances and limitations. A study on fusion proteins and on the Cohesin-Dockerin families. *Proteins* **2006**, *63*, 832–845.
- [92] Eyal, E.; Frenkel-Morgenstern, M.; Sobolev, V.; Pietrovski, S. A pair-to-pair amino acids substitution matrix and its applications for protein structure prediction. *Proteins* **2007**, *67*, 142–153.
- [93] Fariselli, P.; Olmea, O.; Valencia, A.; Casadio, R. Prediction of contact maps with neural networks and correlated mutations. *Protein Eng.* **2001**, *14*, 835–843.
- [94] Eyal, E.; Pietrovski, S.; Bahar, I. Rapid assessment of correlated amino acids from pair-to-pair (P2P) substitution matrices. *Bioinformatics* **2007**, *23*, 1837–1839.
- [95] Pazos, F.; Helmer-Citterich, M.; Ausiello, G.; Valencia, A. Correlated mutations contain information about protein-protein interaction. *J. Mol. Biol.* **1997**, *271*, 511–523.
- [96] Pazos, F.; Valencia, A. *In silico* two-hybrid system for the selection of physically interacting protein pairs. *Proteins* **2002**, *47*, 219–227.
- [97] Gobel, U.; Sander, C.; Schneider, R.; Valencia, A. Correlated mutations and residue contacts in proteins. *Proteins* **1994**, *18*, 309–317.
- [98] Cover, T.; Thomas, J. *Elements of Information Theory*; Wiley-Interscience: New York, 1991.
- [99] Lockless, S. W.; Ranganathan, R. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* **1999**, *286*, 295–299.
- [100] Fodor, A. A.; Aldrich, R. W. Influence of conservation on calculations of amino acid covariance in multiple sequence alignments. *Proteins* **2004**, *56*, 211–221.
- [101] Lee, B. C.; Kim, D. A new method for revealing correlated mutations under the structural and functional constraints in proteins. *Bioinformatics* **2009**, *25*, 2506–2513.
- [102] Clarke, N. D. Covariation of residues in the homeodomain sequence family. *Protein Sci.* **1995**, *4*, 2269–2278.
- [103] Liu, Y.; Eyal, E.; Bahar, I. Analysis of correlated mutations in HIV-1 protease using spectral clustering. *Bioinformatics* **2008**, *24*, 1243–1250.
- [104] Brik, A.; Wong, C. H. HIV-1 protease: Mechanism and drug discovery. *Org. Biomol. Chem.* **2003**, *1*, 5–14.
- [105] Clavel, F.; Hance, A. J. HIV drug resistance. *N. Engl. J. Med.* **2004**, *350*, 1023–1035.
- [106] Velazquez-Campoy, A.; Muzammil, S.; Ohtaka, H.; Schon, A.; Vega, S.; Freire, E. Structural and thermodynamic basis of resistance to HIV-1 protease inhibition: Implications for inhibitor design. *Curr. Drug Targets Infect. Disord.* **2003**, *3*, 311–328.
- [107] Rhee, S. Y.; Gonzales, M. J.; Kantor, R.; Betts, B. J.; Ravela, J.; Shafer, R. W. Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res.* **2003**, *31*, 298–303.
- [108] Shi, J.; Malik, J. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 888–905.
- [109] Gonzales, M. J.; Machekano, R. N.; Shafer, R. W. Human immunodeficiency virus type 1 reverse-transcriptase and protease subtypes: Classification, amino acid mutation patterns, and prevalence in a northern California clinic-based population. *J. Infect. Dis.* **2001**, *184*, 998–1006.
- [110] Ohtaka, H.; Schon, A.; Freire, E. Multidrug resistance to HIV-1 protease inhibition requires cooperative coupling between distal mutations. *Biochemistry* **2003**, *42*, 13659–13666.
- [111] Lichtarge, O.; Sowa, M. E. Evolutionary predictions of binding surfaces and interactions. *Curr. Opin. Struct. Biol.* **2002**, *12*, 21–27.
- [112] Brehmer, D.; Rudiger, S.; Gassler, C. S.; Klostermeier, D.; Packschies, L.; Reinstein, J.; Mayer, M. P.; Bukau, B. Tuning of chaperone activity of Hsp70 proteins by modulation of nucleotide exchange. *Nature Struct. Biol.* **2001**, *8*, 427–432.
- [113] Ma, J.; Sigler, P. B.; Xu, Z.; Karplus, M. A dynamic model for the allosteric mechanism of GroEL. *J. Mol. Biol.* **2000**, *302*, 303–313.
- [114] Petrone, P.; Pande, V. S. Can conformational change be described by only a few normal modes? *Biophys. J.* **2006**, *90*, 1583–1593.
- [115] Yang, L.; Song, G.; Jernigan, R. L. How well can we understand large-scale protein motions using normal modes of elastic network models? *Biophys. J.* **2007**, *93*, 920–929.
- [116] Hyeon, C.; Lorimer, G. H.; Thirumalai, D. Dynamics of allosteric transitions in GroEL. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 18939–18944.
- [117] Horovitz, A.; Yifrach, O. On the relationship between the Hill coefficients for steady-state and transient kinetic data: A criterion for concerted transitions in allosteric proteins. *Bull. Math. Biol.* **2000**, *62*, 241–246.
- [118] Yifrach, O.; Horovitz, A. Mapping the transition state of the allosteric pathway of GroEL by protein engineering. *J. Am. Chem. Soc.* **1998**, *120*, 13262–13263.
- [119] Maragakis, P.; Karplus, M. Large amplitude conformational change in proteins explored with a plastic network model: Adenylate kinase. *J. Mol. Biol.* **2005**, *352*, 807–822.
- [120] Isin, B.; Schulten, K.; Tajkhorshid, E.; Bahar, I. Mechanism of signal propagation upon retinal isomerization: Insights from molecular dynamics simulations of rhodopsin restrained by normal modes. *Biophys. J.* **2008**, *95*, 789–803.
- [121] Kantarci-Carsibasi, N.; Haliloglu, T.; Doruker, P. Conformational transition pathways explored by Monte Carlo simulation integrated with collective modes. *Biophys. J.* **2008**, *95*, 5862–5873.
- [122] Haliloglu, T.; Bahar, I. Coarse-grained simulations of conformational dynamics of proteins: Application to apomyoglobin. *Proteins* **1998**, *31*, 271–281.
- [123] Haliloglu, T. Coarse-grained simulations of the conformational dynamics of proteins. *Comput. Theor. Polymer Sci.* **1999**, *9*, 255–260.
- [124] Bahar, I.; Erman, B.; Haliloglu, T.; Jernigan, R. L. Efficient characterization of collective motions and interresidue correlations in proteins by low-resolution simulations. *Biochemistry* **1997**, *36*, 13512–13523.
- [125] Flory, P. J. *Statistical Mechanics of Chain Molecules*; Wiley-Interscience: New York, 1969.
- [126] Bahar, I.; Jernigan, R. L. Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation. *J. Mol. Biol.* **1997**, *266*, 195–214.
- [127] Bahar, I.; Kaplan, M.; Jernigan, R. L. Short-range conformational energies, secondary structure propensities, and recognition of correct sequence-structure matches. *Proteins* **1997**, *29*, 292–308.
- [128] Temiz, N. A.; Meirovitch, E.; Bahar, I. *Escherichia coli* adenylate kinase dynamics: Comparison of elastic network model modes with mode-coupling (15)N-NMR relaxation data. *Proteins* **2004**, *57*, 468–480.
- [129] Srinivasan, R.; Rose, G. D. The T-to-R transformation in hemoglobin: A reevaluation. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 11113–11117.
- [130] Zheng, W.; Brooks, B. R. Normal-modes-based prediction of protein conformational changes guided by distance constraints. *Biophys. J.* **2005**, *88*, 3109–3117.
- [131] Zheng, W.; Brooks, B. R. Modeling protein conformational changes by iterative fitting of distance constraints using reoriented normal modes. *Biophys. J.* **2006**, *90*, 4327–4336.
- [132] Carlson, H. A. Protein flexibility is an important component of structure-based drug discovery. *Curr. Pharm. Des.* **2002**, *8*, 1571–1578.
- [133] Teague, S. J. Implications of protein flexibility for drug discovery. *Nature Rev. Drug Discov.* **2003**, *2*, 527–541.
- [134] Totrov, M.; Abagyan, R. Flexible ligand docking to multiple receptor conformations: A practical alternative. *Curr. Opin. Struct. Biol.* **2008**, *18*, 178–184.
- [135] Akten, E. D.; Cansu, S.; Doruker, P. A docking study using atomistic conformers generated via elastic network model for cyclosporin A/cyclophilin A complex. *J. Biomol. Struct. Dyn.* **2009**, *27*, 13–26.

- [136] Zavodszky, M. I.; Lei, M.; Thorpe, M. F.; Day, A. R.; Kuhn, L. A. Modeling correlated main-chain motions in proteins for flexible molecular recognition. *Proteins* **2004**, *57*, 243–261.
- [137] Cavasotto, C. N.; Kovacs, J. A.; Abagyan, R. A. Representing receptor flexibility in ligand docking through relevant normal modes. *J. Am. Chem. Soc.* **2005**, *127*, 9632–9640.
- [138] Floquet, N.; Marechal, J. D.; Badet-Denisot, M. A.; Robert, C. H.; Dauchez, M.; Perahia, D. Normal mode analysis as a prerequisite for drug design: Application to matrix metalloproteinases inhibitors. *FEBS Lett.* **2006**, *580*, 5130–5136.
- [139] Schneidman-Duhovny, D.; Nussinov, R.; Wolfson, H. J. Automatic prediction of protein interactions with large scale motion. *Proteins* **2007**, *69*, 764–773.
- [140] May, A.; Zacharias, M. Energy minimization in low-frequency normal modes to efficiently allow for global flexibility during systematic protein-protein docking. *Proteins* **2008**, *70*, 794–809.
- [141] Eyal, E.; Yang, L. W.; Bahar, I. Anisotropic network model: Systematic evaluation and a new web interface. *Bioinformatics* **2006**, *22*, 2619–2627.
- [142] Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D.M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera: A visualization system for exploratory research and analysis. *J. Comp. Chem.* **2004**, *25*, 1605–1612.
- [143] Wilson, I. A.; Skehel, J. J.; Wiley, D. C. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. *Nature* **1981**, *289*, 366–373.
- [144] Weis, W. I.; Brunger, A. T.; Skehel, J. J.; Wiley, D. C. Refinement of the influenza virus hemagglutinin by simulated annealing. *J. Mol. Biol.* **1990**, *212*, 737–761.
- [145] Wilbanks, S. M.; McKay, D. B. How potassium affects the activity of the molecular chaperone Hsc70: II. Potassium binds specifically in the ATPase active site. *J. Biol. Chem.* **1995**, *270*, 2251–2257.
- [146] Zhang, Z.; Sugio, S.; Komives, E. A.; Liu, K. D.; Knowles, J. R.; Petsko, G. A.; Ringe, D. Crystal structure of recombinant chicken triosephosphate isomerase-phosphoglycolohydroxamate complex at 1.8-Å resolution. *Biochemistry* **1994**, *33*, 2830–2837.
- [147] Yusupov, M. M.; Yusupova, G. Z.; Baucom, A.; Lieberman, K.; Earnest, T. N.; Cate, J. H.; Noller, H. F. Crystal structure of the ribosome at 5.5 Å resolution. *Science* **2001**, *292*, 883–896.
- [148] Yusupova, G. Z.; Yusupov, M. M.; Cate, J. H.; Noller, H. F. The path of messenger RNA through the ribosome. *Cell* **2001**, *106*, 233–241.
- [149] Voss, N. R.; Gerstein, M.; Steitz, T. A.; Moore, P. B. The geometry of the ribosomal polypeptide exit tunnel. *J. Mol. Biol.* **2006**, *360*, 893–906.
- [150] Muller, C. W.; Schlauderer, G. J.; Reinstein, J.; Schulz, G. E. Adenylate kinase motions during catalysis: An energetic counterweight balancing substrate binding. *Structure* **1996**, *4*, 147–156.
- [151] Muller, C. W.; Schulz, G. E. Structure of the complex between adenylate kinase from *Escherichia coli* and the inhibitor Ap5A refined at 1.9 Å resolution: A model for a catalytic transition state. *J. Mol. Biol.* **1992**, *224*, 159–177.
- [152] Schlauderer, G. J.; Schulz, G. E. The structure of bovine mitochondrial adenylate kinase: Comparison with isoenzymes in other compartments. *Protein Sci.* **1996**, *5*, 434–441.
- [153] Schlauderer, G. J.; Proba, K.; Schulz, G. E. Structure of a mutant adenylate kinase ligated with an ATP-analogue showing domain closure over ATP. *J. Mol. Biol.* **1996**, *256*, 223–227.
- [154] Muller, C. W.; Schulz, G. E. Crystal structures of two mutants of adenylate kinase from *Escherichia coli* that modify the Gly-loop. *Proteins* **1993**, *15*, 42–49.
- [155] Kallen, J.; Mikol, V.; Taylor, P.; Walkinshaw, M. D. X-ray structures and analysis of 11 cyclosporin derivatives complexed with cyclophilin A. *J. Mol. Biol.* **1998**, *283*, 435–449.
- [156] Morris, G.; Goodsell, D.; Halliday, R.; Huey, R.; Hart, W.; Belew, R.; Olson, A. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639–1662.
- [157] Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.