Cooperative dynamics of proteins unraveled by network models

Eran Eyal,[1,2] Anindita Dutta[1] and Ivet Bahar[1]*

Recent years have seen a significant increase in the number of computational studies that adopted network models for investigating biomolecular systems dynamics and interactions. In particular, elastic network models have proven useful in elucidating the dynamics and allosteric signaling mechanisms of proteins and their complexes. Here we present an overview of two most widely used elastic network models, the Gaussian Network Model (GNM) and Anisotropic Network Model (ANM). We illustrate their use in (i) explaining the anisotropic response of proteins observed in external pulling experiments, (ii) identifying residues that possess high allosteric potentials, and demonstrating in this context the propensity of catalytic sites and metal-binding sites for enabling efficient signal transduction, and (iii) assisting in structure refinement, molecular replacement and comparative modeling of ligand-bound forms via efficient sampling of energetically favored conformers. *WIREs Comput Mol Sci* 2011 1 426–439
DOI: 10.1002/wcms.44

INTRODUCTION

Most proteins are molecular machines. They achieve their function because they possess the ability to undergo the structural changes required for their function. These structural changes may vary over a broad range of scales, from atomic fluctuations and side chain rotations to collective domain or subunit movements. Yet, even the large-scale movements do not, in principle, alter the native 'fold'; the packing of secondary structural elements and/or the distribution of tertiary and quaternary contacts between residues remain unchanged for the most part. The collective motions allow for cooperative rearrangements of domains with respect to each other, or loop motions, which maintain the overall architecture/fold, or allow the protein to restore its original conformation, similar to an elastic material. Here, cooperative motions refer to conformational changes that collectively engage large portions of the structure and usually lie at the lowest frequency end of the spectrum of modes accessible to the protein under physiological conditions. Not surprisingly, elastic network models (ENMs) have been broadly exploited in unraveling protein dynamics, based on the premise that proteins have 'intrinsic' abilities, uniquely encoded by their three-dimensional (3D) fold, to sample functional fluctuations near their native state.[1,2]

From a statistical thermodynamics perspective, the accessible conformations result from the equilibrium fluctuations near the global free energy minimum. Some directions of fluctuations are more probable than others. These are the 'soft modes' of motions; they evolve along a direction, on the multidimensional energy landscape, which by definition involves a relatively small energy ascent for a given deformation. These same modes are also expected to be most frequently recruited to achieve function, following the hypothesis that structures may have evolved to access most easily their functional movements. This hypothesis motivated a large body of studies aiming at characterizing the soft modes and exploring their relevance to function. A popular approach has been to extract the soft modes from the normal mode analysis (NMA) of ENMs. In parallel, simplified models such as the Gaussian network model (GNM),[3,4] rooted in the statistical thermodynamics of polymer networks,[5] have been broadly used in the last decade. *Theory and Methods* section provides an overview of the theory and assumptions underlying these approaches.

ENM–NMAs and GNM analyses are now providing us with a better understanding of the cooperative machinery and design principles of biomolecular systems, as will be illustrated for a few cases in the section *Learning from Network Models: Cooperative Responses and Communication*. First, the soft modes

*Correspondence to: bahar@pitt.edu
[1]Department of Computational & Systems Biology, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA
[2]Cancer Research Institute, Sheba Medical Center, Ramat Gan, Israel

DOI: 10.1002/wcms.44

are highly cooperative as may be quantified by their degree of collectivity,[6] i.e., they cooperatively entail a large part of the molecule, if not the entire molecule. They are also very robust, as indicated by the following two features: (i) detailed atomic coordinates or energetics have negligible effect on these modes, as originally demonstrated by Tirion,[7] and confirmed in many studies since then. This is presumably due to the fact that they are collectively defined by all the $3N$-6 internal degrees of freedom for a structure of N interaction sites (atoms or pseudoatoms); and (ii) the intrinsic, structure-encoded fluctuations in conformations undergone by enzymes (in the absence of ligand binding) are in close accord with the changes observed in ligand-bound forms, suggesting that enzymes are predisposed to make the conformational rearrangements required for ligand binding.[8,9] A similar behavior is exhibited by antibodies when binding their antigens[10,11] or by allosteric proteins in general.[1] It should be noted that predisposition of the enzyme means the accessibility of the structural transition path, or a given mode of motion, rather than the coexistence of the alternative state. As such, predisposition does not exclude the adaptation of a given protein upon binding its ligand, but implies that the directions of structural change are selected from among those intrinsically accessible prior to deformation.

Second, the accessible modes provide a complete orthonormal basis set for describing the equilibrium motions in the $3N$-6 dimensional space of internal coordinates. Each mode may be viewed as a mechanism of relaxation, or dissipation of energy, if the structure is subjected to a perturbation from the environment. Low-frequency (or soft) modes evidently constitute the most favorable paths of energy dissipation, as they incur the lowest ascent on the energy surface for a given deformation. Or, the accommodation of a given external stress by a soft mode would be expected to entail a lower internal resistance; therefore, soft modes are naturally activated in response to perturbations. The effective forces measured for exerting uniaxial tensions along different directions can indeed be explained by the selective recruitment of the modes that comply with the extension.[12,13] This concept will be elaborated in the section *Anisotropic Response to Mechanical Stress: Selection of Suitable Modes*.

Third, biomolecular structures need to be optimally 'wired' to efficiently transmit signals, to couple or exchange their chemical and mechanical energies, or induce allosteric responses. ENM representation of proteins permits us to utilize graph theory, along with information theoretic approaches and spectral clustering techniques to identify the sites that play a critical role in the flow of information across the 'network' of amino acids in the folded state. The collective fluctuations predicted by the GNM also define signal propagation mechanisms, as derived using a Markovian model in a recent study.[14] Certain 'nodes' are distinguished by their effective signal transmission, or allosteric communication, properties. Examination of the identity of such key sites reveals a striking preponderance of active sites (e.g., catalytic residues and metal-binding sites), suggesting that efficient transmission of signals is an inherent, structure-encoded property of active sites. Conceivably, active sites of proteins evolved to be positioned at sites that lend themselves to efficient communication. Examples will be presented in the section *Reconciling Equilibrium Dynamics and Signal Propagation Stochastics*.

A practical utility of ENM–NMA of proteins is the predictions of collective changes of structures that are expected to be most readily accessible or to generate alternative conformations that satisfy certain requirements, e.g., optimal binding of a substrate.[8,10,15] Conformational changes along the soft modes have indeed been the focus of essential dynamics analysis of molecular dynamics (MD) simulations trajectories[16] and principal component analysis of ensembles of structures.[8] The soft modes are advantageously exploited for deforming high-resolution structures to fit cryo-electron microscopy (EM) images,[17–19] for structural refinement, and for generating alternative conformations that would suitably account for structural changes stabilized in various complexes/assemblies (e.g., upon binding substrate proteins or ligands). *NMA in Structural Refinement* section will summarize recent developments in that direction and illustrate the utility of ENM–NMA in refining homology models.

THEORY AND METHODS

Normal Mode Analysis

Classical NMA originated from spectroscopic analyses of molecules, in which the absorption in the vibrational spectra is given by normal modes. Its first application to proteins dates back to early 1980s.[20–22] Renewed interest in the last decade stems from the fact that the low-frequency modes can be robustly and efficiently derived using simplified models such as ENMs, and despite such simplifications, the global modes accessible to biomolecular systems are robustly determined and they usually relate to functional motions.[23,24]

The potential energy $V(\boldsymbol{q})$ of a molecular system of n degrees of freedom (e.g., atomic coordinates) may be approximated to the second order of the series expansion around the equilibrium state $\boldsymbol{q}^0 = [q_1{}^0\ q_2{}^0\ q_3{}^0 \ldots q_n{}^0]^T$ as:

$$V(\boldsymbol{q}) = \frac{1}{2}\sum_{i,j}\left(q_i - q_i^0\right)\left[\frac{\partial^2 V}{\partial q_i q_j}\right]^0\left(q_j - q_j^0\right)$$

$$= \frac{1}{2}\Delta\boldsymbol{q}^{\mathrm{T}}\,\boldsymbol{K}\,\Delta\boldsymbol{q}, \qquad (1)$$

where $\Delta\boldsymbol{q}$ is the $3n$-dimensional vector of the instantaneous fluctuations, the superscript T designates its transpose, and $\boldsymbol{K}$ is the positive semidefinite matrix of second derivatives of the potential with respect to atomic coordinates. The ijth element of $\boldsymbol{K}$ is $K_{ij} = [\partial^2 V/\partial q_i\,\partial q_j]^0$. If we treat structural points as harmonic oscillators in the absence of other effects, the protein obeys the equation of motion:

$$\boldsymbol{M}\frac{d^2\Delta\boldsymbol{q}}{dt^2} + \boldsymbol{K}\,\Delta\boldsymbol{q} = 0, \qquad (2)$$

where $\boldsymbol{M}$ is a diagonal matrix composed of n superelements $m_i\,\mathbf{I}_3$ along the diagonal, where m_i is the mass of the ith atom and $\mathbf{I}_3$ is the identity matrix of order 3. The general solution to Eq. (2) is a $3n$-dimensional vector of the form $\Delta\boldsymbol{q}(t) = \boldsymbol{a}e^{i\omega t}$, which, upon substitution into Eq. (2), leads to

$$(-\omega^2\boldsymbol{M} + \boldsymbol{K})\,\boldsymbol{a} = 0. \qquad (3)$$

Premultiplication of Eq. (3) by $\boldsymbol{M}^{-\frac{1}{2}}$ yields $\omega^2\boldsymbol{M}^{\frac{1}{2}}\boldsymbol{a} = \boldsymbol{M}^{-\frac{1}{2}}\boldsymbol{K}\boldsymbol{a} = \boldsymbol{M}^{-\frac{1}{2}}\boldsymbol{K}[\boldsymbol{M}^{-\frac{1}{2}}\boldsymbol{M}^{\frac{1}{2}}]\boldsymbol{a}$, which, upon the change of variables, $\boldsymbol{u} = \boldsymbol{M}^{1/2}\boldsymbol{a}$, $\lambda = \omega^2$, and $\boldsymbol{H} = \boldsymbol{M}^{-\frac{1}{2}}\boldsymbol{K}\boldsymbol{M}^{-\frac{1}{2}}$, leads to the eigenvalue equation

$$\lambda\,\boldsymbol{u} = \boldsymbol{H}\boldsymbol{u}. \qquad (4)$$

NMA is the solution of this equation to obtain the $3n$-6 nonzero eigenvectors $\boldsymbol{u}^{(k)}$ of the Hessian $\boldsymbol{H}$, along with the corresponding eigenvalues, λ_k. The eigenvectors are $3n$-dimensional vectors, the elements of which are organized in 3D vectors that represent the displacements $\boldsymbol{u}_1{}^{(k)}$, $\boldsymbol{u}_2{}^{(k)}$, $\ldots$ $\boldsymbol{u}_n{}^{(k)}$ of the n atoms away from their equilibrium positions as the structure reconfigures along a given mode (e.g., 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 have a smaller curvature/stiffness and they undergo larger excursions from the energy minimum for a given energy increase, hence their 'soft modes' attribute.

One clear limitation of this method is its applicability to near-native conditions only, because it implicitly assumes linear dynamics, all interactions being approximated by harmonic potentials. This assumption is strictly valid in the immediate neighborhood of the global energy minimum. Proteins are subject, however, to anharmonic potentials and nonlinear dynamics, which are manifested by the ruggedness of the energy landscape and multiple minima/barriers. Even side-chain bond rotations exhibit two or three isomeric states often approximated by cosine functions. Furthermore, solvent damping effect is not accounted for by conventional NMA. The results from NMA should therefore be interpreted in the context of these approximations. As discussed below, coarse graining presents, in a sense, the advantage of eliminating certain degrees of freedom and resulting in a 'smoother' energy landscape where local minima are overlooked (or overcome) during collective motions, providing access to substates separated by low-energy barriers, but this comes at the cost of loosing atomic level accuracy.

GNM and Anisotropic Network Model

The use of ENMs to represent native proteins has the following two computational advantages: (i) there is no need for energy minimization and the Protein Data Bank (PDB)[25] structure is assumed to represent a global energy minimum, and (ii) a coarse-grained model is adopted for structure and energetics; typically, each node represents a residue and the pairs of nodes within a certain cutoff distance (R_c) are connected by springs of uniform force constant γ, which significantly reduce the size and complexity of $\boldsymbol{H}$. Two ENMs, the GNM[3,4] and the anisotropic network model (ANM),[26–28] have found wide use in the last decade.

Gaussian Network Model

Following the original statistical thermodynamics theory of random polymer networks,[5] the network nodes are assumed to undergo Gaussian fluctuations $\Delta\boldsymbol{R}_i$ $(1 \le i \le N)$ about their mean positions. Likewise, the distance vectors between nodes undergo Gaussianly distributed fluctuations $\Delta\boldsymbol{R}_{ij} = \boldsymbol{R}_{ij} - \boldsymbol{R}_{ij}{}^0 = \Delta\boldsymbol{R}_j - \Delta\boldsymbol{R}_i$ (Figure 1). The coordinates of α-carbons are used to define the spatial position of the nodes. The GNM potential is given in terms of the fluctuations $\Delta\boldsymbol{R}_i = \boldsymbol{R}_i - \boldsymbol{R}_i{}^0$ in the position vectors of the nodes as[3]

$$V_{\mathrm{GNM}} = \frac{\gamma}{2}\left[\sum_i\sum_{j,j>i}(-\Gamma_{ij})[\Delta\boldsymbol{R}_j - \Delta\boldsymbol{R}_i]^2\right]$$

$$= \frac{\gamma}{2}[\Delta\boldsymbol{X}^T\boldsymbol{\Gamma}\Delta\boldsymbol{X} + \Delta\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{\Gamma}\Delta\boldsymbol{Y} + \Delta\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\Gamma}\Delta\boldsymbol{Z}], \qquad (5)$$

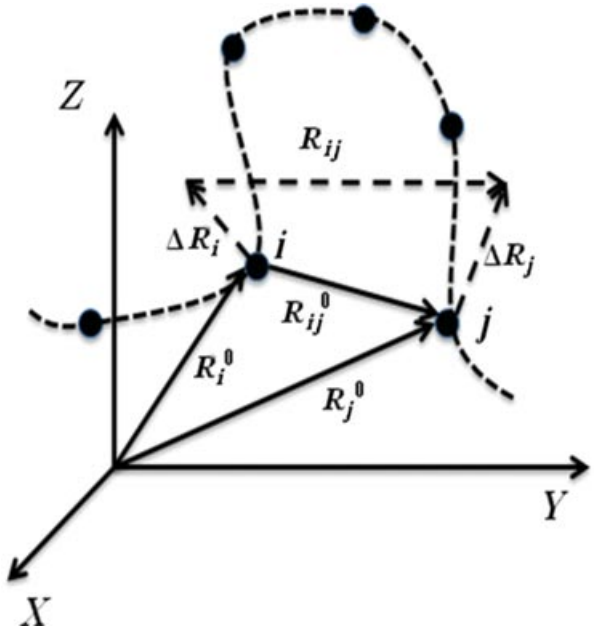

FIGURE 1 | Schematic representation of equilibrium fluctuations. A portion of the protein backbone is displayed by the dotted curve. Filled dots refer to interaction sites (e.g., C^α-atoms) that are adopted as the network nodes. $\boldsymbol{R}_i^0$ and $\boldsymbol{R}_j^0$ designate the equilibrium positions of residues *i* and *j*; $\boldsymbol{R}_i$ and $\boldsymbol{R}_j$ are their instantaneous position vectors. The original and instantaneous separations are indicated by the solid line and dashed line, respectively. The fluctuations in the position vectors are given by $\Delta\boldsymbol{R}_j$ and $\Delta\boldsymbol{R}_i$. $\boldsymbol{R}_{ij}^0$ and $\boldsymbol{R}_{ij}$ designate the equilibrium and instantaneous distance vectors between residues *i* and *j*. The change in the interresidue distance vector is related to the fluctuations in residue positions as $\Delta\boldsymbol{R}_{ij} = \boldsymbol{R}_{ij} - \boldsymbol{R}_{ij}^0 = \Delta\boldsymbol{R}_j - \Delta\boldsymbol{R}_i$, and may be expressed as a weighted sum over the contributions $\Delta\boldsymbol{R}_{ij}^{(k)}$ of individual modes.

where $\boldsymbol{\Gamma}$ is the Kirchhoff (or connectivity) matrix, the off-diagonal elements of which are defined as $\Gamma_{ij} = -1$ if $|\boldsymbol{R}_{ij}| \le R_c$ and zero otherwise, and the diagonal elements are evaluated from the summation $\Gamma_{ii} = -\boldsymbol{\Sigma}_j \Gamma_{ij}$ over all off-diagonal elements in the *i*th row (or column); $\Delta\boldsymbol{X}$, $\Delta\boldsymbol{Y}$, and $\Delta\boldsymbol{Z}$ are the *N*-dimensional vectors of the respective *X*-, *Y*-, and *Z*-components of the fluctuation vectors $\Delta\boldsymbol{R}_1$, $\Delta\boldsymbol{R}_2$, ..., and $\Delta\boldsymbol{R}_N$ corresponding to the *N* residues.

The cross-correlations between the fluctuations of residues *i* and *j* are found from the statistical mechanical average[3]

$$\langle \Delta\boldsymbol{R}_i \cdot \Delta\boldsymbol{R}_j \rangle = \int (\Delta\boldsymbol{R}_i \cdot \Delta\boldsymbol{R}_j) \exp(-V/k_BT) d\{\Delta\boldsymbol{R}\} / \times \int \exp(-V/k_BT) d\{\Delta\boldsymbol{R}\} = \frac{3k_BT}{\gamma}[\boldsymbol{\Gamma}^{-1}]_{ij}, \quad (6)$$

where $[\Gamma^{-1}]_{ij}$ is the *ij*th element of $\boldsymbol{\Gamma}^{-1}$, k_B is the Boltzmann constant, and T is the absolute temperature. Note the determinant of $\boldsymbol{\Gamma}$ is 0, i.e., $\boldsymbol{\Gamma}$ cannot be inverted. Instead, its pseudoinverse is calculated using the $N-1$ nonzero eigenvalues σ_k and eigenvectors $\boldsymbol{v}^{(k)}$ of $\boldsymbol{\Gamma}$. In compact notation,

$$\boldsymbol{C}^{(N)} = \frac{3k_BT}{\gamma}\boldsymbol{\Gamma}^{-1} = \frac{3k_BT}{\gamma}\sum_{k=1}^{N-1}\left[\sigma_k^{-1}\boldsymbol{v}^{(k)}\boldsymbol{v}^{(k)\mathrm{T}}\right], \quad (7)$$

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

Anisotropic Network Model

The ANM potential, V_{ANM}, is similar in form to V_{GNM}, except for the replacement of the scalar product $[\Delta\boldsymbol{R}_j - \Delta\boldsymbol{R}_i]^2 \equiv [(\boldsymbol{R}_{ij} - \boldsymbol{R}_{ij}^0) \bullet (\boldsymbol{R}_{ij} - \boldsymbol{R}_{ij}^0)]$ in Eq. (6) by $[|R_{ij}| - |R_{ij}^0|]^2$ where $|R_{ij}|$ designates the magnitude of R_{ij}. Thus, V_{GNM} penalizes the change in the orientation of $\Delta\boldsymbol{R}_{ij}$ even if the magnitude of the distance vectors is maintained, whereas V_{ANM} is exclusively based on distance changes. The second derivative of V_{ANM} leads to expressions of the form[26]

$$\begin{aligned}\partial^2 V_{\mathrm{ANM}}/\partial X_i\, \partial Y_j|_{\mathrm{R0}} &= -\gamma\, \Gamma_{ij}\left(X_j^0 - X_i^0\right)\left(Y_j^0 - Y_i^0\right)/\left(R_{ij}^c\right)^2 \\ &= -\gamma\, \Gamma_{ij}\left(X_{ij}^0\, Y_{ij}^0\right)/\left(R_{ij}^0\right)^2 \end{aligned} \quad (8)$$

in terms of the components X_i^0, Y_i^0, and Z_i^0 of $\boldsymbol{R}_i^0$. This permits us to express the elements of ANM Hessian $\boldsymbol{H}$ in closed form, and easily evaluate the ANM eigenvector $\boldsymbol{u}^{(k)}$ and eigenvalue λ_k, $(1 \le k \le 3N\text{-}6)$, which provide information on the shape and frequencies of normal modes subject to V_{ANM}. $\boldsymbol{u}^{(k)}$ is a $3N$-dimensional vector, composed of 3D vectors $\boldsymbol{u}_1^{(k)}$, $\boldsymbol{u}_2^{(k)}$, ... $\boldsymbol{u}_N^{(k)}$, that describe the displacements of all *N* residues in mode *k*. The change in interresidue distance induced by mode *k* is given in terms of the elements of $\boldsymbol{u}^{(k)}$ as

$$\Delta\boldsymbol{R}_{ij}^{(k)} = R_{ij}^{(k)} - R_{ij}^0 = (k_BT/\lambda_k)^{1/2}\left[\boldsymbol{u}_i^{(k)} - \left[\boldsymbol{u}_i^{(k)}\right]. \quad (9)$$

Likewise, a $3N \times 3N$ covariance matrix, $\boldsymbol{C}^{(3N)}$, composed of blocks of size 3×3 associated with the three components of each fluctuation vector $\Delta\boldsymbol{R}_j$ replaces $\boldsymbol{C}^{(N)}$. $\boldsymbol{C}^{(3N)}$ scales with the inverse of the

Hessian as

$$\boldsymbol{C}^{(3N)} = k_B T \boldsymbol{H}^{-1} = k_B T \sum_{k=1}^{3N-6} \left[\lambda_k^{-1} \boldsymbol{u}^{(k)} \boldsymbol{u}^{(k)\mathrm{T}}\right]. \quad (10)$$

The GNM/ANM results depend on the overall 'fold' rather than specific interactions or detailed atomic coordinates. As such, they provide information on the global dynamics favored by the particular fold/architecture, rather than local changes in structure and interactions. The main utility of GNM/ANM is indeed the ability to efficiently explore global dynamics (collective motions in the low-frequency regime), which have been demonstrated in multiple studies to be insensitive to structural details or detailed force fields. Results on the high-frequency modes have limited utility: they permit us to identify potentially conserved sites and folding nuclei, but provide no information on the mechanisms of local motions.[29–32] It should also be noted that protein conformers used as input in, or generated as outputs from, ENM–NMA are not energy minimized with a molecular mechanics force field and hence may not be directly utilized in docking studies that require atomic level precision. Efforts are underway for designing hybrid methodologies that combine the information on global dynamics from ENMs and local motions from MD simulations[33] to efficiently explore multiscale processes. Also, sequence information is not included in ENMs and hence single mutations that result in loss of function are not captured by ENM analysis of the mutant, unless there are structural data for the mutant, which differ from that of the wild-type protein. Clearly, NMAs with ENMs also suffer from the same limitations (e.g., inadequate description of nonlinear effects, solvent effect, etc.) as atomic NMA. Yet, the coarse-graining of the structure and interactions allows for efficient sampling of domain rearrangements and cooperative movements near native state conditions. These motions could be hardly accessible (due to local energy barriers) should a detailed, full atomic description of the structure and dynamics be implemented.

Markovian Propagation of Allosteric Signals

Network models provide a useful tool for investigating the communication and signaling properties of complex systems. A widely used approach is model signal transduction as a Markovian stochastic process. We recently explored the possibility of probing the stochastics of signal propagation in proteins using a discrete time, discrete state Markov model.[14] The strength of interactions (or weight of edges) in this network model (or the nonzero elements of $\boldsymbol{\Gamma}$) could be assumed to be uniform (as in the GNM) as a first approximation. A slightly refined model, however, is to take account of residue specificity at a coarse-grained scale. To this aim, we define affinities $a_{ij} = N_{ij}/\sqrt{(N_i N_j)}$ as the weights of the edges, where N_{ij} is the total number of heavy atom contacts between residues i and j with the denominator correcting for bias due to size (Figure 2). The degree d_j of each node is given by $\Sigma_i\, a_{ij}$. The Kirchhoff matrix corresponding to this model is then $\boldsymbol{\Gamma} = \mathbf{D} - \mathbf{A}$, where $\mathbf{D}$ is the diagonal matrix of the degrees and $\mathbf{A}$ is the matrix of affinities. The conditional probability of transmitting a signal from residue j to residue i in one time step is simply $m_{ij} = d_j^{-1} a_{ij}$ (where m_{ij} is the ijth element of the Markov transition matrix $\mathbf{M}$).[14]

A metric of efficiency of signal transmission in network models is the hit time, which is the average number of steps it takes to transmit information from broadcaster i to receiver j. This is described by enumerating all possible ways to get from residue i to residue j, weighted by the transition/conditional probability of signal flow. The recursive equation to evaluate the average hit time between nodes i and j is given by:

$$\begin{aligned} H(j,i) &= \sum_{k=1}^{n} [1 + H(j,k)] m_{ki} \\ &= \sum_{k=1}^{n} m_{ki} + \sum_{k=1,\, k\neq j}^{n} H(j,k) m_{ki} \\ &= 1 + \sum_{k=1,\, k\neq j}^{n} H(j,k) m_{ki}. \end{aligned} \quad (11)$$

Substituting $\mathbf{M} = \mathbf{D}\mathbf{A}^{-1}$ and $\Gamma = \mathbf{D} - \mathbf{A}$, $H(j,i)$ can be expressed in terms of the elements of $\boldsymbol{\Gamma}^{-1}$ as[14]

$$\begin{aligned} H(j,i) = \sum_{k=1}^{N} \{ &[\boldsymbol{\Gamma}^{-1}]_{ki} - [\boldsymbol{\Gamma}^{-1}]_{ji} - [\boldsymbol{\Gamma}^{-1}]_{kj} \\ &+ [\boldsymbol{\Gamma}^{-1}]_{jj} \} d_k, \end{aligned} \quad (12)$$

or, using Eq. (6),

$$\begin{aligned} H(j,i) = (3k_B T/\gamma)^{-1} \sum_{k=1}^{N} [&\langle \Delta R_k . \Delta R_i \rangle - \langle \Delta R_j . \Delta R_i \rangle \\ &- \langle \Delta R_k . \Delta R_j \rangle + \langle \Delta R_j . \Delta R_j \rangle]\, \mathrm{d}_k. \end{aligned} \quad (13)$$

$H(j,i)$ depends on the direction of signal transmission, i.e., $H(j,i) \neq H(i,j)$.

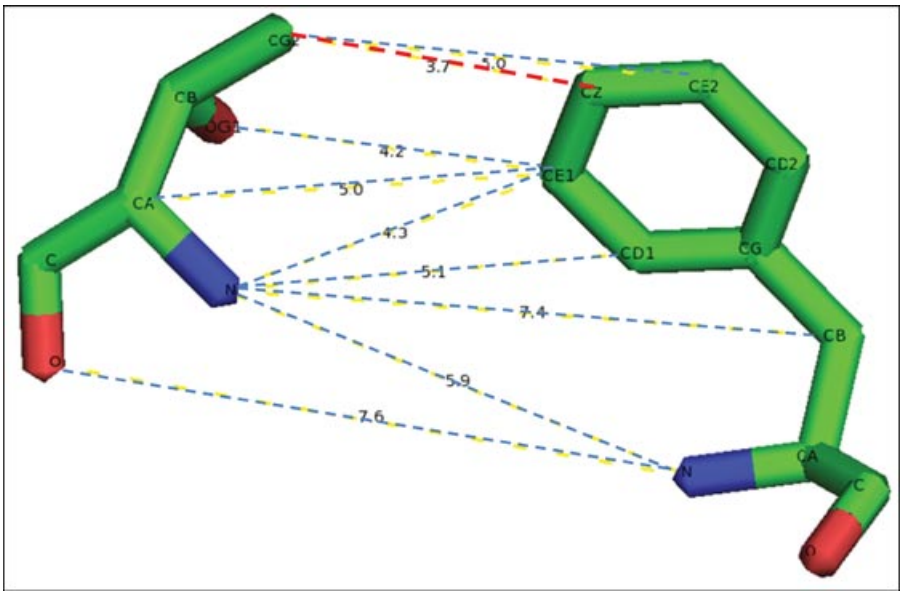

FIGURE 2 | Schematic description of the evaluation of affinity matrix elements. Two residues within interaction range are displayed, composed of $N_i = 7$ and $N_j = 11$ atoms. The affinity is evaluated based on atom–atom contacts that are closer than a cutoff separation, e.g., 4.0 Å. In the present case, there is only one pair of atom within this interaction rage, such that $N_{ij} = 1$, and the affinity between this pair becomes $a_{ij} = N_{ij}/\sqrt{(N_i N_j)} = 1/(77)^{\frac{1}{2}}$ (see *Markovian Propagation of Allosteric Signals* section).

Commute time, $\tau\ (i,\ j) \equiv H(j,i) + H(i,\ j)$, is another metric, which has no directionality.[14] $\tau\ (i,\ j)$ scales with the fluctuations in the distance vector $\Delta \boldsymbol{R}_{ij}$ as

$$\tau(i,\ j) \propto \{[\boldsymbol{\Gamma}^{-1}]_{ii} + [\boldsymbol{\Gamma}^{-1}]_{jj} - 2[\boldsymbol{\Gamma}^{-1}]_{ij}\} = \langle \Delta R_{ij}.\Delta R_{ij}\rangle, \quad (14)$$

where the proportionality constant is $(3k_B T/\gamma)^{-1}\ \boldsymbol{\Sigma}_k d_k$. Eq. (14) is readily obtained using Eq. (13) twice, for $H(j,i)$ and $H(i,\ j)$. Eqs (13) and (14) permit us to bridge between statistical physical quantities such as mean-square fluctuations or cross-correlations with graph-theoretic concepts such as hitting or commute times. We will illustrate the utility of these concepts for analyzing signal transduction behavior of proteins by way of application to two cases in the section *Reconciling Equilibrium Dynamics and Signal Propagation Stochastics*.

LEARNING FROM NETWORK MODELS: COOPERATIVE RESPONSES AND COMMUNICATION

Anisotropic Response to Mechanical Stress: Selection of Suitable Modes

Biomolecules are often subjected to mechanical pressures—e.g., within muscle fibers and microtubules. This, together with recent advances in single-molecule manipulation techniques, led to increased interest in the mechanical response of proteins to external pulling forces applied at specific positions.[34] A database of results from such experiments has been recently constructed.[34,35]

Theoretical studies for elucidating the relationship between mechanical stability and biological function have been traditionally performed using steered MD.[36–40] More recently, network models and analytical treatments have been utilized. One example is the ANM analysis of the anisotropic response to deformations in different pulling directions,[12] presented below in some detail. Although this approach assumes that the unfolding path correlates with the 'weak/soft directions' near the native state, other studies went on to longer and more detailed simulations of the process and demonstrated explicitly the relation between normal modes and the unfolding process by interactively considering the specific interactions being broken. An example is the work of Dietz and Reif[41] wherein the fractional force that each 'bond' experiences so as to withstand a uniaxial tension was evaluated, which also yielded good agreement with experimental data on green fluorescent protein (GFP). In another study, snapshots from stochastic simulations of stretching have been analyzed.[42] ENMs also assist in identifying folding nuclei[32,43] and exploring the order of events in thermal unfolding. A recent application of GNM

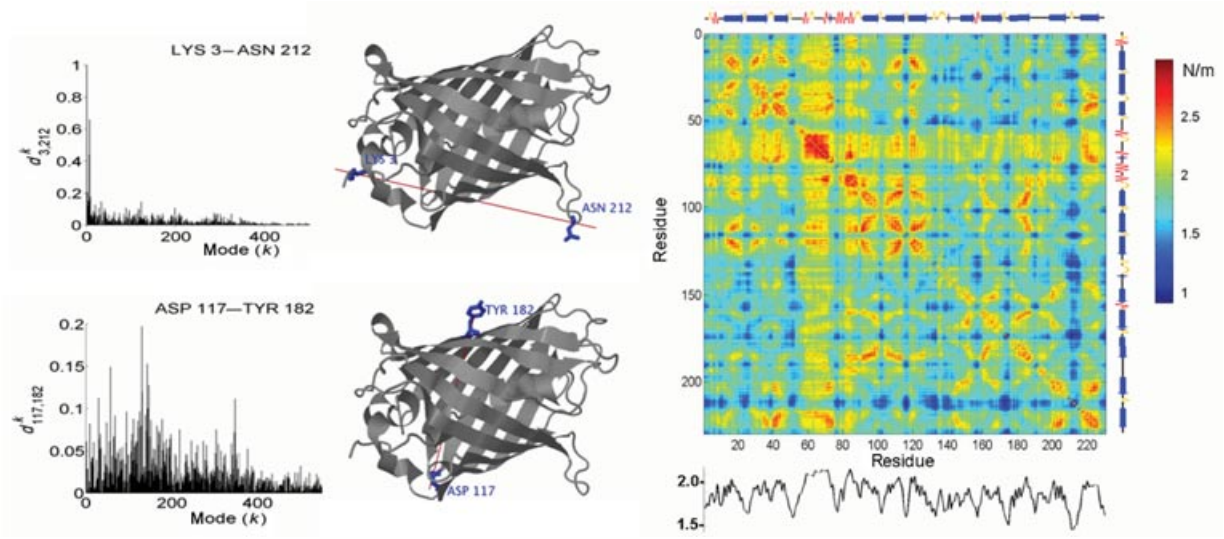

FIGURE 3 | (a) Comparison of mechanical stress (uniaxial tension) experiments with anisotropic network model (ANM) predictions for green fluorescent protein (GFP). The plots on the left display the normalized contributions $d_{ij}^{(k)}$ of ANM modes $k = 1$, $3N$-6 (abscissa) to the deformation of the molecule in response to uniaxial tension exerted at residues $(i, j) =$ (K3, N212) (top) and (D117,Y182) (bottom). When pulling apart the residues Lys3–Asn212 (upper diagram), the deformation is largely accommodated one soft mode (note the peak at mode 1). In the case of Asp117–Tyr182 (lower diagram), the response is distributed over a broader range of modes, including high-frequency modes that entail high energy cost. The force required to unfold the molecule by stretching in the Lys3–Asn112 direction was reported[46] to be five times weaker than that needed in the Asp117–Tyr182 direction, consistent with the mode distributions. (b) Complete mechanical resistance map generated for GFP by ANM. The entries in the map represent the effective force constants $<\kappa_{ij}>$ (Eq. 15) in response to uniaxial tensions exerted at each of pair of residues. The secondary structure of the protein is shown along the upper abscissa and on the right (blue arrows, β-strand; zigzag line, α-helix). The profile at the lower part of the map displays the mean resistance per residue, averaged over all values in a given column. See our study[12] for more details.

wherein native contacts are iteratively broken to simulate unfolding yielded results in accord with those observed in MD and Monte Carlo simulations.[43]

The ANM permits us to calculate an effective force constant $<\kappa_{ij}>$ for the stiffness/resistance of the molecule against deformation under uniaxial tension applied to residues i and j.[12] The idea is simple: if these residues already undergo long distance fluctuations by virtue of intrinsically accessible slow modes, the molecule is more 'yielding'. $<\kappa_{ij}>$ is the 'average force constant' evaluated using[12]

$$< \kappa_{ij} >= \sum_k d_{ij}^{(\kappa)} \lambda_k \Big/ \sum_k d_{ij}^{(\kappa)}, \qquad (15)$$

where $d_{ij}^{(\kappa)}$ represents the contribution of mode k to the extension of the interresidue distance R_{ij}, given by the projection of $\Delta \boldsymbol{R}_{ij}^{(k)}$ onto $\boldsymbol{R}_{ij}^{0}$, i.e., $d_{ij}^{(k)} = \Delta \boldsymbol{R}_{ij}^{(k)} \cos(\boldsymbol{R}_{ij}^{0}, \Delta \boldsymbol{R}_{ij}^{(k)})$.

We recently analyzed[12] the mechanical behavior observed by single-molecule atomic force microscopy for a series of proteins.[44–46] Figure 3 illustrates the results obtained for GFP. The bars represent the computed contributions $d_{ij}^{(k)}$ of different modes to the extensions along the pairs Lys3–Asn212 (top) and Asp117–Tyr182 (bottom). In the former case, the low-frequency modes make a larger contribution (e.g., $\boldsymbol{d}_{3-212}{}^{(k)} > 0.6$ for $k = 1$), consistent with the relatively low (~115 pN) pulling force observed in experiments.[46] In the latter case, on the contrary, $\boldsymbol{d}_{117-182}{}^{(k)}$ values are more uniformly distributed over a broader range of modes, including higher frequency modes, which are energetically expensive. Consistent with these predictions, the GFP is destabilized at 548 pN upon pulling these two residues.[46] Similar calculations performed for all residue pairs led to the resistance map ($<\kappa_{ij}>$ values) displayed in Figure 3. The lower profile is the average over all entries in each column, thus providing a measure of mechanical resistance for each individual amino acid. Similar calculations performed for various pulling directions and various proteins yielded results in good agreement with experiments despite the simplicity of the theory and the differences in various experimental conditions.[12]

In principle, the forces predicted by the ANM refer to behavior in the immediate neighborhood of the global energy minimum or to the resistance to initiating a motion in a particular direction. Experiments, on the contrary, probe the unfolding behavior. The

consistency between theory and experiments points to a correlation between the original stiffness/resistance to exiting from the folded state and the effective cost of unfolding. The approach of Dietz and Reif,[41] on the contrary, assumes that a single bond breakage is equivalent to complete unfolding. Both theoretical approaches are based on native contact topology and cannot be used for unfolding processes wherein the transition state is far from the native state. They yield relative forces for different residue pairs, rather than absolute values; and they lack sequence information. Thus due to calibrations that depend on the protein and on experimental conditions alike, these models cannot compare values obtained for different proteins or mutants. Despite such limitations, coarse-grained (CG) models[42,47] and in particular ENM-based approaches[12,13,41,42] appear as useful tools for estimating the anisotropic response of proteins to external stress, and may assist in designing or interpreting single-molecule experiments at a very low computing time and memory cost.

Reconciling Equilibrium Dynamics and Signal Propagation Stochastics

Equations (13) and (14) establish the connection between information theoretic concepts (hit and commute times) and physics-based equilibrium dynamics (e.g., mean-square fluctuations in interresidue distances). Previous work demonstrated that hit times are relatively small when targeting catalytic residues, consistent with the efficient communication requirements of active sites.[14] A more recent study also shows that metal-binding residues follow the same trend.[48]

Figure 4 illustrates these concepts for two example cases. Panels (a)–(c) display the results for a xylose isomerase that binds a cobalt ion. Panel (a) shows the hit time map of the metal-binding protein. It can be seen that signal propagation properties essentially depend on the target site (abscissa). For a clearer understanding of the signaling properties of individual residues, we define two single-body properties: mean hitting times and mean commute times, $<H_r(j)> = \mathbf{\Sigma}_i\, H(j,i)/N$ and $<\tau(j)> = \mathbf{\Sigma}_i\, \tau(j,i)/N$ for each residue. These properties provide a measure of the allosteric potential of residues. Panel (b) displays the mean hitting times for xylose isomerase. Metal-binding residues are E267, E231, D338, and D295 in this case. The hitting time profile shows that all these residues (indicated by red dots) are located at minima, i.e., they communicate efficiently with all other residues or they are very sensitive to signals conveyed by other residues. Panel (c) shows that in addition to having small $<H_r(j)>$ values (abscissa), their variances in $H(j,i)$ (for $1 \leq i \leq N$) are small, suggesting that these residues are not only fast but also precise in their communication with other residues.

Panels (a′)–(c′), on the contrary, present the results for an enzyme, human carbonic anhydrase II, with respect to its catalytic residues H64, Q92, H94, H96, and H119. Again, the active residues appear to be highly predisposed to efficient communication with the rest of the protein, in addition to be precisely tuned (i.e., have minimal variance in hitting times).

These results suggest that the positions of the active residues in the 3D structure may have been evolutionarily optimized to ensure efficient communication. In order to examine to which extent the sites with high allosteric potential (or enhanced communication properties) are evolutionarily conserved, we compared in Figure 5 the color-coded diagrams based on communication properties (panels a and a′) and conservation properties (panels b and b′). Conservation scores were obtained using the ConSurf server.[49] The comparisons of panels (a) and (b), and panels (a′) and (b′) clearly demonstrates the correlation between conservation patterns and efficient communication propensities, i.e., those residues distinguished by their efficient signaling properties are also evolutionarily conserved presumably due to their functional role.

There are currently several studies that unravel the structural origins of molecular signaling processes using NMA. A recent study by Beratan and coworkers[50] demonstrated, e.g., that local structural changes triggered by ligand binding can translate into signaling responses at remote allosteric sites and different motions mediate different signaling pathways. The connection between signaling pathways and collective motions predicted by ENM–NMA has been also demonstrated in an extensive study of GroEL–GroES machinery.[51] Such analyses could be particularly useful in structure-based design of inhibitors/regulators of allosteric proteins.

NMA in Structural Refinement

Soon after the pioneering articles in the 1980s[20–22] that demonstrated the utility of NMA for investigating the equilibrium dynamics of proteins, it was suggested that the new methodology could assist in refining X-ray structures.[52–55] The refinement of B-factors, previously based on a single fitted parameter per atom, was replaced by a summation of the anisotropic fluctuations over a selected set of low-frequency normal modes. The main conceptual advantage of a NMA-based refinement over traditional

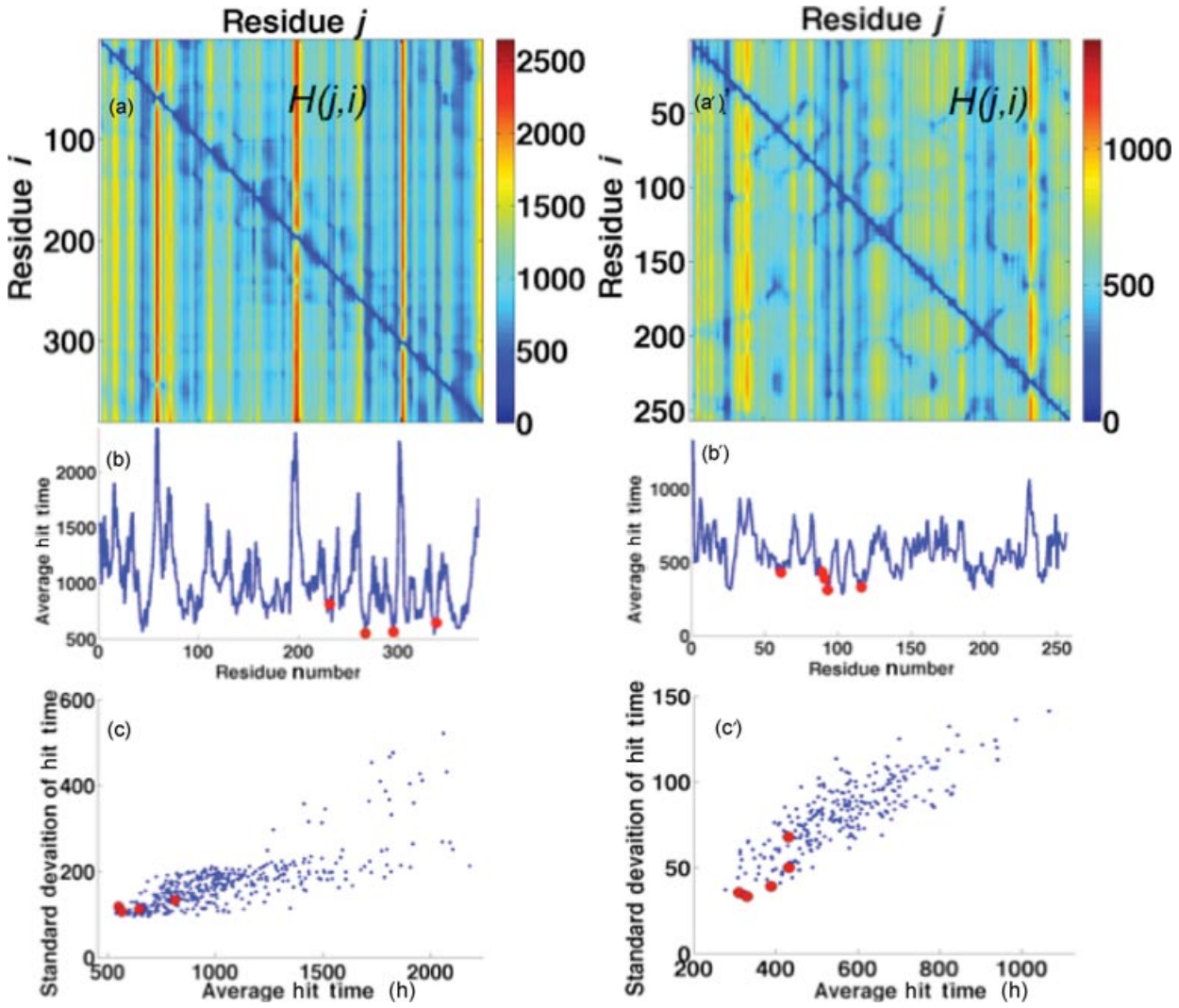

FIGURE 4 | Efficient signal propagation properties of active sites illustrated for a metal-binding protein (a–c) and an enzyme (a′–c′). Hit time maps for cobalt-binding xylose isomerase [Protein Data Bank (PDB) code: 1a0c] and for human carbonic anhydrase II (PDB code: 1a42) are displayed in the respective panels (a) and (a′). Average hit time profiles as a function of residue number are shown in panels (b) and (b′) with metal-binding and catalytic residues labeled by the red dots in the two respective cases. Panels (c) and (c′) display the average hit time versus standard deviation for each residue for the two cases. The residues that participate in metal binding (b) and catalysis (b′) (shown in red markers) exhibit small average hit time and small-to-moderate standard deviation, indicative of their fast and precise communication properties.

refinements is that it takes account of intrinsic motions and correlations. The essence of the refinement procedure is to minimize the difference (R-factor) between observed structure factors and those calculated based on model parameters. In a typical NMA-based refinement, the model is evaluated using the position vectors, $R^0 + \sum_k c_k u_k$, perturbed along normal modes (k), where c_k is the contribution of mode k to be fitted. As low-frequency modes induce the largest displacements, a subset of slow modes is usually sufficient, and thus the number of fitting parameters is usually significantly smaller than the number of atoms. NMA-based fitting with a small number of parameters may outperform standard refinements with many more parameters[52,54,56] while avoiding overfitting. NMA refinement also allows for large movements beyond the capability of other refinement methods[57] and is of particular utility for large structures and complexes that are inherently flexible.[58–60] In an application to the potassium channel KcsA, e.g., Chen et al.[56] were able to resolve residues that were originally missing and obtain anisotropic displacement parameters related to functional motions.

An additional advantage of NMA refinement is that it inherently decomposes the atomic fluctuations into their internal and external components. The

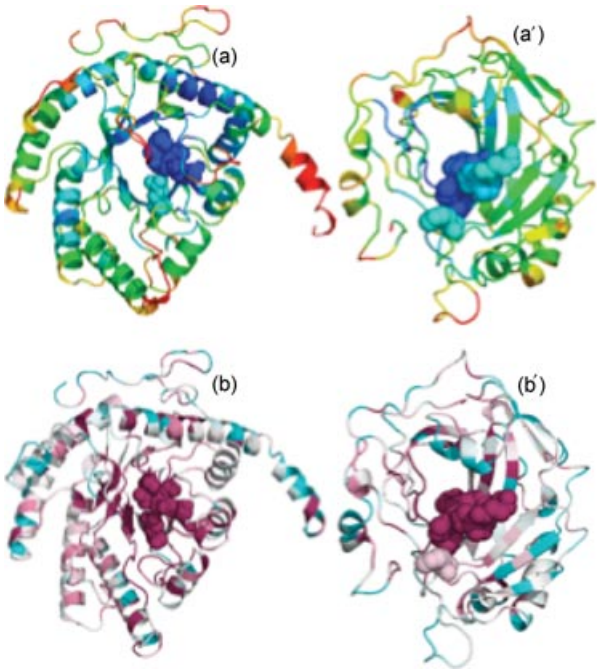

FIGURE 5 | Correlation between signaling efficiency and evolutionary conservation. Panels (a; 1a0c) and (a′; 1a42) show the ribbon diagrams color coded by the distance of residues from the origin in Figure 4 (c) and (c′). The regions colored blue refer to the fastest and most precise communication sites. Panels (b) and (b′) are color coded by conservation score of each residue as obtained from the Consurf server, the most conserved sites being colored purple. The metal-binding and catalytic sites are shown in space-filling representation.

internal component refers to the collective vibrational motions of the molecule (corresponding to the nonzero eigenvalues), which are usually of interest for assessing functional mechanisms. The external component, on the contrary, reflects the static disorder and rigid body motion in the crystal as well as uniform thermal noise and inaccuracies in the experiments. By separating the two components, it has been demonstrated[55,56,58–61] that rigid body motions make a significant contribution to the B-factors observed in X-ray crystallographic structures.

In recent years, it has been shown[62] that NMA can assist in molecular replacement (MR), the most popular and rapid method for solving structures by X-ray crystallography. The method is based on using a structural template. The authors showed for three different examples (metallodextrin-binding protein, glutamine-binding protein, and HIV-1-protease) that perturbing the template structures along one or two low-frequency modes yields new conformers that allow for successful MR, whereas the original structures fail to do so. This suggests that many structures that cannot be resolved currently using MR—despite the existence of apparently suitable templates—will be resolved upon NMA-based perturbations of their templates. This will save effort, time, and resources, setting the stage for application of NMA-based approaches.

With the advances in cryo-EM techniques, there is a great interest in refining the low-resolution (sometimes in the order of 15–25 Å) data. A useful strategy to this aim is to dock known higher resolution structures (or substructures) onto EM density maps, but often the dynamic nature of the molecules is overlooked. There is a need for tools that consider the flexibility of the molecule during refinement, and NMA has been naturally employed for this purpose.[17,19,63,64] The objective function to be minimized is the difference between computed and measured electron densities, using NMA soft modes as search directions. Applications of this methodology confirmed that large conformational changes need to be accounted for in order to have a good fit to EM density maps. An important contribution was the development of Quantized Elastic Deformational Model,[65,66] wherein the protein is treated as an elastic object with mass distribution taken from the density maps. An elegant application to fatty acid synthase[67] demonstrated that structures along electron density maps conform to the first two softest modes. This study highlighted the strength of combining NMA with multiple model refinement for highly flexible proteins, as well as the ability to improve the resolution of the final EM reconstructions.

Chacon et al.[68] showed that alternative conformations for several important well-studied systems (bacterial RNA polymerase, the 30S and 50S subunits of the ribosome, and chaperonin) could be detected within single low-resolution EM structures, consistent with major conformational changes predicted by normal modes. These experimental results validate what coarse-grained NMA models taught us long ago—large-scale motions are insensitive to detailed atomic interactions. A number of excellent tools, such as NOMAD,[69] NORMA,[18] and NMFF,[70] are now available to make it easy and straightforward to apply NMA for interpreting EM density maps.

NMA in Modeling Structures, Ligand-bound Forms, and Complexes

NMA approaches can assist not only in the refinement of experimental structures but also for theoretical modeling. We are currently in a position to improve the quality of homology models by refining the template structures using ENM–NMA. The idea

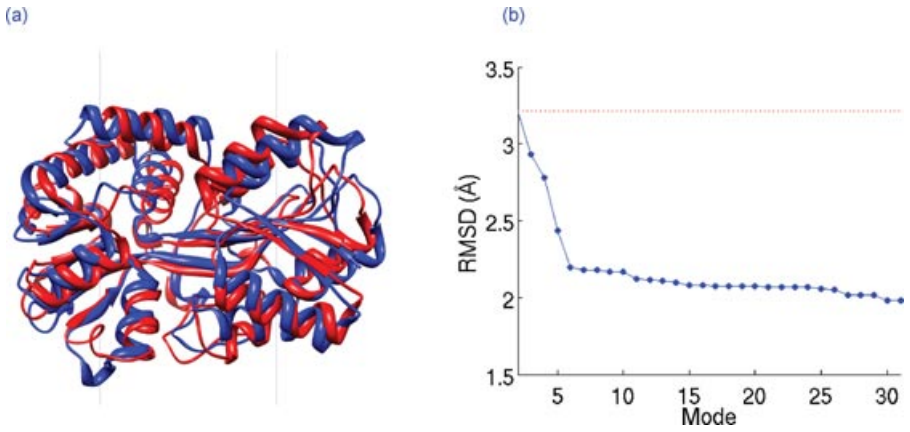

FIGURE 6 | Structural refinement of the bacteria ferric ion-binding protein by deforming a template protein along its normal modes. (a) Superposition of the target and template proteins [Protein Data Bank codes 1xvy (red) and 1qvs (blue), respectively]. The proteins share 37% sequence identity. The root-mean-square deviation (RMSD) between the structures is 3.21 Å. (b) Reduction of the RMSD by reconfigurations of the template along the first 30 normal modes accessible to the template structure, calculated using the anisotropic network model. The red dotted line shows the RMSD between the two structures prior to refinement.

is to restrict the search space to a subspace spanned by a few lower frequency normal modes.[71,72] When the scoring function is put aside, by taking it to be ideal, it has been recently shown that the normal modes perform better than short MD refinement or coarse-grained Monte Carlo sampling for moderate difficulty template refinement.[73] Baker and coworkers[74] have shown, e.g., that a search along dominant principal components of ensembles of structures in a given family assists in improving the accuracy in comparative modeling. The idea is to determine the dominant directions of structural variance upon optimal superimposition of multiple structures (for a given protein or a family of proteins that share similar fold) and evaluate the covariances in the departures of residue/atom coordinates from their mean values, which are organized in a covariance matrix (Eqs. 7 and 10). Eigenvalue decomposition of the latter following Eqs (7) and (10) yields the principal axes (as eigenvectors) of structural deviations. Given that the principal components from experimental structures correlate well with normal modes,[8,71,75] it should be possible to use normal modes instead of principal components for this purpose, and thus to expand the utility of this approach. For more about PCA and the relation between NMA and PCA, the reader is referred to previous work.[8,24,71,75]

The advantage of normal modes is that they can be calculated based on individual template structures and are not dependent on the availability of numerous family representatives that are needed for PCA. Figure 6 demonstrates the potential of this approach for two ferric ion-binding proteins from bacteria that share only 37% sequence identity. It can be seen that movements along the four softest modes starting from a given structure ('template') permit to attain structures significantly closer to the 'target' structure. When combining the first four modes, the root-mean-square deviation (RMSD) may be reduced from 3.2 to 2.1 Å. The use of higher order modes has a marginal effect, if any.

Recently, a combined method that utilizes principal components and normal modes was developed.[76] A recent nice application of a similar methodology is the prediction of open and close conformations of rat heme-free oxygenase,[77] starting from a unique heme-bound X-ray structure, and using templates corresponding to a human and incomplete rat heme-free structures.

CONCLUSION

Here, we present an overview of the foundations, basic approximations, and utility of network models used for delineating protein dynamics. GNM and ANM are physics-based models, whereas the Markov model is information theoretic. The two approaches provide information on different properties: motions in the former case and pathways of communication in the latter. Yet, the average properties predicted by the former such as the mean-square fluctuations in interresidue distances and cross-correlations between residue fluctuations also define information theoretic quantities such as hitting and/or commute times computed for different residue pairs (Eqs. 13 and 14). In a sense, graph/network representation allows us to bridge between dynamic and allosteric signaling

properties. In either case, the native contact topology defines the most cooperative modes of motions, which in turn define the effective resistance to deformation, the efficient communication mechanisms, and the potential structural changes in different forms or environments, as illustrated in three different applications (Figures 2–6).

We would like to emphasize in particular the practical utility of NMA–ENM for gaining a deeper understanding of the conformational space/ensemble accessible to each protein. The advantage of NMA–ENM analyses is that they can be performed based on individual structures and do not depend on the availability of numerous family representatives (that are needed for principal components analyses, for example). This endows these approaches with a significant advantage in filling the gaps in the sequence–structure space in the absence of sufficiently broad structural data for a given family. We anticipate these features of ENM–NMA to be increasingly exploited, especially in the characterization of supramolecular structures and assemblies.

ACKNOWLEDGEMENTS

Support from NIH grant #1R01GM086238-01 is gratefully acknowledged by I.B.

REFERENCES

1. 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.
2. Bahar I, Lezon TR, Bakan A, Shrivastava IH. Normal mode analysis of biomolecular structures: intrinsic motions of membrane proteins. *Chem Rev* 2010, 110:1463–1497.
3. Bahar I, Atilgan AR, Erman B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold Des* 1997, 2:173–181.
4. Haliloglu T, Bahar I, Erman B. Gaussian dynamics of folded proteins. *Phys Rev Lett* 1997, 79:3090–3093.
5. Flory P. Statistical thermodynamics of random networks. *Proc R Soc Lond A* 1976, 351:351–380.
6. Bruschweiler R. Collective protein dynamics and nuclear spin relaxation. *J Chem Phys* 1995, 102:3396–3403.
7. Tirion MM. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys Rev Lett* 1996, 77:1905–1908.
8. 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 U S A* 2009, 106:14349–14354.
9. Tobi D, Bahar I. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc Natl Acad Sci U S A* 2005, 102:18908–18913.
10. James LC, Roversi P, Tawfik DS. Antibody multispecificity mediated by conformational diversity. *Science* 2003, 299:1362–1367.
11. Tokuriki N, Tawfik DS. Protein dynamism and evolvability. *Science* 2009, 324:203–207.
12. Eyal E, Bahar I. Toward a molecular understanding of the anisotropic response of proteins to external forces: insights from elastic network models. *Biophys J* 2008, 94:3424–3435.
13. Sacquin-Mora S, Lavery R. Modeling the mechanical response of proteins to anisotropic deformation. *Chemphyschem* 2009, 10:115–118.
14. Chennubhotla C, Bahar I. Signal propagation in proteins and relation to equilibrium fluctuations. *PLoS Comput Biol* 2007, 3:1716–1726.
15. Cavasotto CN, Kovacs JA, Abagyan RA. Representing receptor flexibility in ligand docking through relevant normal modes. *J Am Chem Soc* 2005, 127:9632–9640.
16. Amadei A, Linssen AB, Berendsen HJ. Essential dynamics of proteins. *Proteins* 1993, 17:412–425.
17. Hinsen K, Reuter N, Navaza J, Stokes DL, Lacapere JJ. Normal mode-based fitting of atomic structure into electron density maps: application to sarcoplasmic reticulum Ca-ATPase. *Biophys J* 2005, 88:818–827.
18. Suhre K, Navaza J, Sanejouand YH. NORMA: a tool for flexible fitting of high-resolution protein structures into low-resolution electron-microscopy-derived density maps. *Acta Crystallogr D Biol Crystallogr* 2006, 62:1098–1100.
19. Tama F, Miyashita O, Brooks CL III. Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis. *J Mol Biol* 2004, 337:985–999.

20. Brooks B, Karplus M. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc Natl Acad Sci U S A* 1983, 80:6571–6575.
21. Go N, Noguti T, Nishikawa T. Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc Natl Acad Sci U S A* 1983, 80:3696–3700.
22. Levitt M, Sander C, Stern PS. Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. *J Mol Biol* 1985, 181:423–447.
23. Bahar I, Rader AJ. Coarse-grained normal mode analysis in structural biology. *Curr Opin Struct Biol* 2005, 15:586–592.
24. Cui Q. Bahar IE. *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems*. Boca Raton: Chapman & Hall/CRC; 2006.
25. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, et al. The Protein Data Bank. *Acta Crystallogr D Biol Crystallogr* 2002, 58:899–907.
26. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys J* 2001, 80:505–515.
27. Doruker P, Atilgan AR, Bahar I. Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: application to alpha-amylase inhibitor. *Proteins* 2000, 40:512–524.
28. Eyal E, Yang LW, Bahar I. Anisotropic network model: systematic evaluation and a new web interface. *Bioinformatics* 2006, 22:2619–2627.
29. Bahar I, Atilgan AR, Demirel MC, Erman B. Vibrational dynamics of folded proteins: significance of slow and fast motions in relation to function and stability. *Phys Rev Lett* 1998, 80:2733–2736.
30. Demirel MC, Atilgan AR, Jernigan RL, Erman B, Bahar I. Identification of kinetically hot residues in proteins. *Protein Sci* 1998, 7:2522–2532.
31. Haliloglu T, Keskin O, Ma B, Nussinov R. How similar are protein folding and protein binding nuclei? Examination of vibrational motions of energy hot spots and conserved residues. *Biophys J* 2005, 88:1552–1559.
32. Rader AJ, Bahar I. Folding core predictions from network models of proteins. *Polymer* 2004, 45:659–668.
33. 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.
34. Oberhauser AF, Carrion-Vazquez M. Mechanical biochemistry of proteins one molecule at a time. *J Biol Chem* 2008, 283:6617–6621.
35. Sulkowska JI, Cieplak M. Stretching to understand proteins – a survey of the protein data bank. *Biophys J* 2008, 94:6–13.
36. Grater F, Shen J, Jiang H, Gautel M, Grubmuller H. Mechanically induced titin kinase activation studied by force-probe molecular dynamics simulations. *Biophys J* 2005, 88:790–804.
37. Lee EH, Gao M, Pinotsis N, Wilmanns M, Schulten K. Mechanical strength of the titin Z1Z2–telethonin complex. *Structure* 2006, 14:497–509.
38. Lu H, Schulten K. The key event in force-induced unfolding of titin's immunoglobulin domains. *Biophys J* 2006, 79:51–65.
39. Lu H, Krammer A, Isralewitz B, Vogel V, Schulten K. Computer modeling of force-induced titin domain unfolding. *Adv Exp Med Biol* 2000, 481:143–160.
40. Pabon G, Amzel LM. Mechanism of titin unfolding by force: insight from quasi-equilibrium molecular dynamics calculations. *Biophys J* 2006, 91:467–472.
41. Dietz H, Rief M. Elastic bond network model for protein unfolding mechanics. *Phys Rev Lett* 2008, 100:098101.
42. Sulkowska JI, Kloczkowski A, Sen TZ, Cieplak M, Jernigan RL. Predicting the order in which contacts are broken during single molecule protein stretching experiments. *Proteins* 2008, 71:45–60.
43. Su JG, Li CH, Hao R, Chen WZ, Wang CX. Protein unfolding behavior studied by elastic network model. *Biophys J* 2008, 94:4586–4596.
44. Brockwell DJ, Paci E, Zinober RC, Beddard GS, Olmsted PD, Smith DA, Perham RN, Radford SE. Pulling geometry defines the mechanical resistance of a beta-sheet protein. *Nat Struct Biol* 2003, 10:731–737.
45. Carrion-Vazquez M, Li H, Lu H, Marszalek PE, Oberhauser AF, Fernandez JM. The mechanical stability of ubiquitin is linkage dependent. *Nat Struct Biol* 2003, 10:738–743.
46. Dietz H, Berkemeier F, Bertz M, Rief M. Anisotropic deformation response of single protein molecules. *Proc Natl Acad Sci U S A* 2006, 103:12724–12728.
47. West DK, Brockwell DJ, Olmsted PD, Radford SE, Paci E. Mechanical resistance of proteins explained using simple molecular models. *Biophys J* 2006, 90:287–297.
48. Dutta A, Bahar I. Metal-binding residues are distinguished by their lower mobilities and efficient signal propagation properties. *Structure* 2010, 18:1140–1148.
49. Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, Pupko T, Ben-Tal N. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res* 2005, 33:W299–W302.
50. Balabin IA, Yang W, Beratan DN. Coarse-grained modeling of allosteric regulation in protein receptors. *Proc Natl Acad Sci U S A* 2009, 106:14253–14258.
51. Chennubhotla C, Yang Z, Bahar I. Coupling between global dynamics and signal transduction pathways: a mechanism of allostery for chaperonin GroEL. *Mol Biosyst* 2008, 4:287–292.

52. Diamond R. On the use of normal modes in thermal parameter refinement: theory and application to the bovine pancreatic trypsin inhibitor. *Acta Crystallogr A* 1990, 46:425–435.
53. Kidera A, Go N. Refinement of protein dynamic structure: normal mode refinement. *Proc Natl Acad Sci U S A* 1990, 87:3718–3722.
54. Kidera A, Inaka K, Matsushima M, Go N. Normal mode refinement: crystallographic refinement of protein dynamic structure. II. Application to human lysozyme. *J Mol Biol* 1992, 225:477–486.
55. Kidera A, Inaka K, Matsushima M, Go N. Normal mode refinement: crystallographic refinement of protein dynamic structure applied to human lysozyme. *Biopolymers* 1992, 32:315–319.
56. Chen X, Poon BK, Dousis A, Wang Q, Ma J. Normal-mode refinement of anisotropic thermal parameters for potassium channel KcsA at 3.2 Å crystallographic resolution. *Structure* 2007, 15:955–962.
57. Delarue M, Dumas P. On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models. *Proc Natl Acad Sci U S A* 2004, 101:6957–6962.
58. Poon BK, Chen X, Lu M, Vyas NK, Quiocho FA, Wang Q, Ma J. Normal mode refinement of anisotropic thermal parameters for a supramolecular complex at 3.42-Å crystallographic resolution. *Proc Natl Acad Sci U S A* 2007, 104:7869–7874.
59. Ni F, Poon BK, Wang Q, Ma J. Application of normal-mode refinement to X-ray crystal structures at the lower resolution limit. *Acta Crystallogr D Biol Crystallogr* 2009, 65:633–643.
60. Chen X, Lu M, Poon BK, Wang Q, Ma J. Structural improvement of unliganded simian immunodeficiency virus gp120 core by normal-mode-based X-ray crystallographic refinement. *Acta Crystallogr D Biol Crystallogr* 2009, 65:339–347.
61. Song G, Jernigan RL. vGNM: a better model for understanding the dynamics of proteins in crystals. *J Mol Biol* 2007, 369:880–893.
62. Suhre K, Sanejouand YH. On the potential of normal-mode analysis for solving difficult molecular-replacement problems. *Acta Crystallogr D Biol Crystallogr* 2004, 60:796–799.
63. Falke S, Tama F, Brooks CL, III, Gogol EP, Fisher MT. The 13 angstroms structure of a chaperonin GroEL–protein substrate complex by cryo-electron microscopy. *J Mol Biol* 2005, 348:219–230.
64. Mitra K, Schaffitzel C, Shaikh T, Tama F, Jenni S, Brooks CL 3rd, Ban N, Frank J. Structure of the *E. coli* protein-conducting channel bound to a translating ribosome. *Nature* 2005, 438:318–324.
65. Ming D, Kong Y, Lambert MA, Huang Z, Ma J. How to describe protein motion without amino acid sequence and atomic coordinates. *Proc Natl Acad Sci U S A* 2002, 99:8620–8625.
66. Tama F, Wriggers W, Brooks CL III. Exploring global distortions of biological macromolecules and assemblies from low-resolution structural information and elastic network theory. *J Mol Biol* 2002, 321:297–305.
67. Brink J, Ludtke SJ, Kong Y, Wakil SJ, Ma J, Chiu W. Experimental verification of conformational variation of human fatty acid synthase as predicted by normal mode analysis. *Structure* 2004, 12:185–191.
68. Chacon P, Tama F, Wriggers W. Mega-dalton biomolecular motion captured from electron microscopy reconstructions. *J Mol Biol* 2003, 326:485–492.
69. Lindahl E, Azuara C, Koehl P, Delarue M. NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. *Nucleic Acids Res* 2006, 34:W52–W56.
70. Tama F, Miyashita O, Brooks CL III. Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM. *J Struct Biol* 2004, 147:315–326.
71. Leo-Macias A, Lopez-Romero P, Lupyan D, Zerbino D, Ortiz AR. Core deformations in protein families: a physical perspective. *Biophys Chem* 2005, 115:125–128.
72. Leo-Macias A, Lopez-Romero P, Lupyan D, Zerbino D, Ortiz AR. An analysis of core deformations in protein superfamilies. *Biophys J* 2005, 88:1291–1299.
73. Stumpff-Kane AW, Maksimiak K, Lee MS, Feig M. Sampling of near-native protein conformations during protein structure refinement using a coarse-grained model, normal modes, and molecular dynamics simulations. *Proteins* 2008, 70:1345–1356.
74. Qian B, Ortiz AR, Baker D. Improvement of comparative model accuracy by free-energy optimization along principal components of natural structural variation. *Proc Natl Acad Sci U S A* 2004, 101:15346–15351.
75. Yang LW, 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.
76. Han R, Leo-Macias A, Zerbino D, Bastolla U, Contreras-Moreira B, Ortiz AR. An efficient conformational sampling method for homology modeling. *Proteins* 2008, 71:175–188.
77. Marechal JD, Perahia D. Use of normal modes for structural modeling of proteins: the case study of rat heme oxygenase 1. *Eur Biophys J* 2008, 37:1157–1165.

