Preprint to the chapter of the book "Computational Biology: New Research", Jan. 28, 2008.

Coarse-Grained Structural Model of Protein Molecules

Kilho Eom[1], and Sungsoo Na[2]

[1]Nano-Bio Research Center, Korea Institute of Science & Technology (KIST), Seoul 136-791, Republic of Korea

[2]Department of Mechanical Engineering, Korea University, Seoul 136-701, Republic of Korea

Abstract

Understanding the protein mechanics is *a priori* requisite for gaining insight into protein's biological functions, since most protein performs its function through the structural deformation renowned as conformational change. Such conformational change has been computationally delineated by atomistic simulations, albeit the mechanics of large protein structure is computationally inaccessible with atomistic simulation such as molecular dynamics simulation. In a recent decade, normal mode analysis with coarse-grained modeling of protein structures has been a computational alternative to atomistic simulations for understanding large protein mechanics. In this review, we delineate the current state-of-art in coarse-grained modeling of proteins for normal mode analysis. Specifically, the pioneered coarse-grained models such as Go model and elastic network model as well as recently developed coarse-grained elastic network model are summarized and discussed for understanding large protein mechanics.

Keywords: protein mechanics, coarse-grained model, normal mode analysis, Go model, elastic network model

INTRODUCTION

Protein mechanics plays a vital role in the biological function of proteins, since protein performs its biological function through its structural deformation driven by mechanical loading. For instance, motor protein is one of renowned proteins that perform the mechanical function, that is, the transduction of chemical energy into mechanical energy [1]. Specifically, the mechanical function of ATPase motor protein is carried out via its structural change upon ATP binding [2-6]. The chaperonin GroEL-GroES complex functions the assistance of protein folding through rotational motion of its domain upon ATP binding [7, 8]. The giant muscle molecule known as titin performs the mechanical function through structural change from folded structure to unfolded (denatured) structure or vice versa upon mechanical loading or unloading [9-13].

Protein mechanics related to protein's biological function has been computationally studied by atomistic simulations [14] since McCammon *et al.* [15] studied the dynamic behavior of small protein based on molecular dynamics simulation. The thermal fluctuation behavior of small proteins has been well understood by sampling of trajectories obtained from molecular dynamics simulation [16]. Moreover, the mechanical unfolding of a protein such as titin has been well analyzed by molecular dynamics simulation with consideration of mechanical loading applied to the termini of a protein [17-19]. The basic principle of molecular dynamics simulation is to numerically solve the equation of motion, i.e. $m_i \ddot{\mathbf{u}}_i = \mathbf{f}_i$, where m_i is the mass of the i-th atom, $\mathbf{u}_i$ is the displacement field for i-th atom, and $\mathbf{f}_i$ is the force acting on the i-th atom [14, 20]. The computational difficulty in molecular dynamics simulation resides in computation of the force $\mathbf{f}_i$ that is a gradient of an anharmonic potential field prescribed to all atoms. Further, the time step for integrating the equation of motion is typically in the order of femto (10^{-15}) seconds, whereas protein performs the function at much larger time scale from at least nano (10^{-9}) seconds to a few seconds. It has been reported that until now the accessible time scale for molecular dynamics is at most in the order of nano seconds [21]. This indicates that molecular dynamics simulation may be computationally inhibited for large protein mechanics, where large spatial and temporal scale is required.

In recent decades, normal mode analysis (NMA) has been a computational alternative to atomistic simulation such as molecular dynamics for understanding large protein mechanics [14, 22-24]. The principle of NMA is similar to that typically employed for structural mechanics. Specifically, once the stiffness matrix (Hessian matrix) for a structure is constructed, the modal analysis provides the vibration information of such a structure. The stiffness matrix for a protein structure is usually established based on the computation of second gradients of anharmonic potential field prescribed to all atoms. In general, calculation of stiffness matrix is implemented at equilibrium position, which is obtained by minimization of anharmonic potential. This implies that, for large proteins, the computation of stiffness matrix along the minimization process is a computationally expensive process.

Recently, atomistic simulation such as molecular dynamics [15] and NMA [22] with all atoms

has been replaced with a coarse-grained model, where degrees of freedom are enormously reduced. Since the dominant motion of a protein structure is represented by that of a carbon backbone chain [14, 16], the coarse-grained models have been suggested such that protein structure is delineated by α-carbon atoms for protein backbone chain. Moreover, the computational inefficiency usually arises from the complicated anharmonic potential field. Consequently, the simplification of such potential field for a protein structure described by α-carbon atoms is the key issue in the coarse-grained modeling of proteins. Go and coworkers [22] introduced the more simplified potential field for α-carbon atoms such that α-carbon atoms are prescribed by potential field consisting of covalent bonds for consecutive α-carbon atoms and non-bonded interaction (i.e. van der Waal's interaction) for native contacts. The thermal fluctuation behavior of protein structures has been well described by Go model. In a recent decade, Go model has been taken more attention for gaining insight into protein unfolding mechanics. Cieplak and coworkers [25-28] have showed that molecular simulation with Go potential has allowed them to obtain the force-displacement curve for protein unfolding mechanism, quantitatively comparable to experimental data by single-molecule pulling experiments based on atomic force microscopy (AFM) [9]. This may shed light on Go model such that Go potential may be versatile potential field for understanding protein mechanics with computational efficiency.

Inspired by Go model, Tirion suggested the more simple harmonic potential for protein structures [29]. In her model, protein structure is regarded as a harmonic spring network in such a way that α-carbon atoms within the neighborhood is connected by harmonic springs with identical force constant. Tirion's model has revolutionized the protein modeling for understanding protein dynamics relevant to biological function of proteins [30-37]. Such model has reduced computational expense enormously for estimating low-frequency normal modes related to biological function. Moreover, it is very remarkable that low-frequency normal modes from Tirion's model are highly correlated with displacement vector representing the conformational change of proteins [38]. Tirion's model has inspired many researchers for studying protein dynamics and protein mechanics. For example, Wolynes and coworkers [39] studied the energy landscape for protein conformational change based on Tirion's model. Kim *et al* [40] introduced the linear interpolation method based on Tirion's model for describing the conformational change. Brooks and coworkers [41, 42] have studied the conformation change based on iterative method applied to low-frequency normal modes from Tirion's model with a distance constraint for computing the displacement vector related to incremental conformational change. Micheletti and coworkers [43] employed Tirion's model for depicting the thermal denaturation (thermal unfolding) of protein's folded structure. Zheng *et al* [44] have studied the power-stroke mechanism of motor proteins based on Tirion's model. The recent studies by Brooks *et al* [37, 45] and Kim *et al* [46] have showed that low-frequency normal modes of Tirion's model is sufficient to provide the functional mode of viral capsid protein. Recently, Thirumalai and coworkers [47] have shown that low-frequency normal modes are able to describe the allosteric transition of proteins. Bahar and coworkers [48] have provided that allosteric

change of protein structure is well delineated by low-frequency normal modes of Tirion's model. Moreover, they have recently suggested that Tirion's model with Markov method may enable one to understand the allosteric signal transduction corresponding to conformational change [49, 50].

Although Tirion's model has greatly succeeded in studying protein dynamics and mechanics with high computational efficiency, the model reduction scheme (coarse-graining) has been taken into account for large protein structures. The model reduction of Tirion's model is rational since few low-frequency normal modes are necessary for describing the protein dynamics such as conformational change relevant to biological function. Such model reduction has been first attributed to Bahar and coworkers [51] who introduced the coarse-grained structure represented by nodal points whose number is less than total number of α-carbon atoms. In their model, the nodal points within the neighborhood were connected by harmonic spring with identical spring constant. In a recent year, Eom and coworkers [52] have provided the more systematic model reduction method applicable to protein structures. Specifically, they used the model condensation method in a similar spirit to skeletionization method suggested by Rohklin and coworkers [53, 54]. Bahar and coworkers [55] have introduced the Markov method for transformation of original molecular structure into coarse-grained structure. Ma and coworkers [56] have employed the substructure synthesis method, which was broadly used for engineering structural dynamics, for obtaining the low-frequency normal modes relevant to biological function.

MOLECULAR SIMULATION: NORMAL MODE ANALYSIS (NMA)

All-atom simulation such as molecular dynamics was first provided by Karplus and coworkers [15]. The potential field V prescribed to protein structure was given by [14]

$$V = \sum_i \frac{K}{2}\left(b_i - b_i^0\right)^2 + \sum_i \frac{C}{2}\left(\theta_i - \theta_i^0\right)^2 + \sum_i D\left[1+\cos\left(n\varphi_i - \delta\right)\right] + \sum_{i,j}\left(\frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^{6}}\right) + \sum_{i,j}\frac{q_i q_j}{\chi r_{ij}} \quad (1)$$

Here, b_i, θ_i, and φ_i are the i-th covalent bond length, bending angle, and dihedral angle (torsional angle), respectively, r_{ij} is the distance between i-th atom and j-th atom, q_i is the charge for i-th atom, and a symbol 0 indicates the equilibrim state. The first term in potential energy represents the stretching energy for covalent bonds, the second term indicates the bending energy, the third term shows the torsional energy, and last two terms provides the non-bonded interactions such as van der Waal's interaction and electrostatic interaction. With the potential energy V given by Eq. 1, molecular dynamics simulation provides the trajectories of position vectors denoted as $\mathbf{x}_i$ for all atoms. The quantity renowned as cross-correlation matrix $\mathbf{L}_{ij}$ shows the thermal fluctuation behavior comparable to experimental quantity such as Debye-Waller factor [16, 57].

$$\mathbf{L}_{ij} = \left\langle \left(\mathbf{x}_i - \mathbf{x}_i^0\right) \cdot \left(\mathbf{x}_j - \mathbf{x}_j^0\right) \right\rangle \quad (2)$$

where $\mathbf{x}_i$ is the position vector for i-th atom, a symbol 0 represents the equilibrium state, and a braket symbol indicates the ensemble average (time average). The diagonal component of cross-correlation

matrix, $\mathbf{L}_{ii}$, is the mean-square fluctuation proportional to Debye-Waller factor (B-factor), i.e. $B_i = (8\pi^2/3)L_{ii}$.

Normal mode analysis (NMA) is referred to as quasi-harmonic analysis [58], since the modal analysis is implemented with harmonic approximation to potential energy V for small displacement.

$$V \approx V_0 + \frac{1}{2}\sum_{i,j} K_{ij}\left(x_i - x_i^0\right)\left(x_j - x_j^0\right) \tag{3}$$

Here, x_i is the generalized coordinates for atoms, and K_{ij} is the Hessian matrix (stiffness matrix) for a protein structure given by $K_{ij} = \partial^2 V/\partial x_i \partial x_j$. Quasi-harmonic analysis (or NMA) is to solve the eigen-value problem such as $K_{ij}v_j = -\omega^2 m_i v_i$, where ω is the natural frequency, v_i is the normal mode corresponding to natural frequency ω, and m_i is the atomic mass for i-th atom. The cross-correlation matrix representing the thermal fluctuation motion can be computed from equilibrium statistical mechanics theory [57, 59].

$$\mathbf{L}_{ij} = \sum_{n=7}^{3N} \frac{k_B T}{m_i \omega_n^{\,2}}\left(v_{i,n} \otimes v_{j,n}\right) \tag{4}$$

where k_B is the Boltzmann's constant, T is the absolute temperature, and the subscript n for natural frequency and normal mode represents the mode number. It should be noted that summation goes from 7 to $3N$, where N is the total number of atoms, since there are six rigid body modes corresponding to zero eigen-modes. Even though there are several different potential fields such as CHARMM and AMBER applicable to protein structures, it was shown that the thermal fluctuation motion and low-frequency normal modes are consistent regardless of details of potential field [58].

GO MODEL

As stated above, the low-frequency normal modes relevant to protein dynamics is insensitive to details of potential field [58]. One may ask which interactions dominate the protein dynamics among various potential fields as mentioned in Eq. 1. Go and coworkers [22] conjectured that short-range interactions may govern the protein dynamics. Moreover, the motion of protein structure is well described by that of backbone chain represented by α-carbon atoms. Go potential is simply represented in the form of [25, 26]

$$V \approx \sum_i \left[\frac{k_1}{2}\left(r_{i,i+1} - r_{i,i+1}^0\right)^2 + \frac{k_2}{4}\left(r_{i,i+1} - r_{i,i+1}^0\right)^4\right] + \sum_{i,j} 4\varepsilon\left(\frac{1}{r_{ij}^{\,6}} - \frac{1}{r_{ij}^{\,12}}\right) \tag{5}$$

Here, $r_{i,j}$ is the distance between i-th and j-th α-carbon atoms, and superscript 0 indicates the equilibrium state. The first summation represents the nonlinear elastic energy for covalent bonds, while the last summation shows the non-bonded interaction for native contact. Native contact is defined in such a way that two α-carbon atoms (i-th and j-th α-carbon atoms) are in the native contact if r_{ij} is less than the certain distance referred to as cut-off distance, d_c, typically given as $d_c \approx 10$ Å.

TIRION'S MODEL: ELASTIC NETWORK MODEL (ENM)

The success of Go model has resulted in the emergence of more simplified model suggested by Tirion

[29]. Specifically, Tirion assumed the harmonic approximation to potential field prescribed to α-carbon atoms. Inspired by Go model, she proposed the harmonic potential field only for native contacts and covalent bonds with identical force constant.

$$V \approx \sum_{i,j} \frac{\gamma}{2}\left(r_{ij} - r_{ij}^0\right)^2 \cdot H\left(r_c - r_{ij}^0\right) \tag{6}$$

where γ is a force constant, r_{ij} is the distance between i-th and j-th α-carbon atoms, superscript 0 indicates the equilibrium state, r_c is the cut-off distance defining a native contact, and $H(x)$ is the Heaviside unit-step function defined as $H(x) = 0$ if $x < 0$; otherwise $H(x) = 1$.

With Tirion's potential, Bahar and coworkers studied the Gaussian dynamics of proteins, which resulted in the emergence of Gaussian network model (GNM) [30, 32]. GNM assumes the isotropic fluctuation, that is, directionality of fluctuation is not taken into account. Even though the motion of proteins is generally anisotropic, the fluctuation behavior such as B-factor is well depicted by GNM. The stiffness matrix for GNM, also referred to as Kirchoff matrix, is given by

$$\Gamma_{ij} = \frac{\partial^2 V}{\partial r_i \partial r_j} = -\gamma\left(1 - \delta_{ij}\right) \cdot H\left(r_c - r_{ij}^0\right) - \delta_{ij} \sum_{k \neq i}^{N} \Gamma_{ik} \tag{7}$$

Here, N is the total number of α-carbon atoms, and δ_{ij} is the Kronecker delta defined as $\delta_{ij} = 1$ if $i = j$; otherwise $\delta_{ij} = 0$. Since isotropic motion is assumed in GNM, GNM is able to only provide the low-frequency normal modes related to mean-square fluctuations. Further, fluctuation information for every residue may provide the insight into the hot spots (residues) which undergo large deformation during the conformational change.

In general, Tirion's model is referred to as elastic network model (ENM) [29, 33] since protein structure is represented by harmonic spring network, which takes into account the anisotropy in thermal fluctuation. For simplicity, let us consider only two α-carbon atoms i and j, which are connected by an entropic spring (Gaussian chain) [33, 60-64].

$$V\left(r_{ij}\right) = \frac{\gamma}{2}\left(r_{ij} - r_{ij}^0\right)^2 \tag{8}$$

where $r_{ij} = [(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2]^{1/2}$ with a position vector $\mathbf{r}_i$ for an α-carbon atom i, given by $\mathbf{r}_i = x_i\mathbf{e}_x + y_i\mathbf{e}_y + z_i\mathbf{e}_z$. The stiffness matrix $\mathbf{K}$ for a potential given by Eq. 3 can be easily computed such as

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_{ij} & -\mathbf{K}_{ij} \\ -\mathbf{K}_{ij} & \mathbf{K}_{ij} \end{bmatrix} \tag{9}$$

Here, $\mathbf{K}_{ij}$ is the 3×3 block matrix given by

$$\mathbf{K}_{ij} = \gamma \frac{\left(\mathbf{r}_i - \mathbf{r}_j\right) \otimes \left(\mathbf{r}_i - \mathbf{r}_j\right)}{\left|\mathbf{r}_i - \mathbf{r}_j\right|^2} \tag{10}$$

This indicates that the stiffness matrix for an entropic spring is equivalent to the stiffness matrix for an elastic spring (linear elastic truss) with a spring constant γ. Based on 3×3 block matrix $\mathbf{K}_{ij}$, the stiffness

matrix corresponding to Tirion's potential given by Eq. 6 can be easily computed by assembly of such block matrices. The protein structure described by ENM is suggested in Fig. 1.

COARSE-GRAINED ELASTIC NETWORK MODEL

Coarse-graining of protein structures with few degrees of freedom has been attempted, since protein structure is composed of several rigid domains whose motional behavior is like a rigid-body motion such as rotational motion. In recent years, Jernigan and coworkers [65, 66] suggested that protein structure is represented by complex of rigid bodies corresponding to protein domains. That is, they introduced the Hamiltonian for rigid-body motion of a domain as well as interactions between domains. It was remarkably shown that the dynamic behavior such as conformational change of large protein complex (e.g. GroEL-GroES, viral capsid) has been well illustrated by their coarse-grained model [46, 66]. Bahar and coworkers [51] have taken into account the coarse-graining of ENM based on their physical intuition. Their coarse-grained ENM was established in the same manner to ENM except they rescaled the force constant as well as cut-off distance. It is remarkable that their simple coarse-grained model successfully predicts the thermal fluctuations comparable to original structure as well as experimental data. Furthermore, multi-scale model for proteins has been suggested in such a way that the biologically significant substructures such as binding site are described by refined model such as ENM whereas the rest regions of a protein is described by coarse-grained ENM [67].

The coarse-graining process based on ENM may be systematically implemented by employing the model reduction method typically used in applied mathematics. For instance, Rohklin and coworkers [53, 54] suggested the low-rank approximation to linear algebraic equation, resulting in the reduction of degrees of freedom. They showed that their low-rank approximation, referred to as *skeletionization*, can be directly applicable to electrostatics [68], hydrodynamics [53], and any other applied mathematics problem represented by linear equation [54]. Inspired by *skeletonization* scheme, we have employed the model condensation method to reconstruct the coarse-grained structure, i.e. low-resolution structure, from the original structure, i.e. refined structure (See Fig. 2) [52, 64]. We define the *master residues* as the residues which are taken in the coarse-grained structure, while the *slave residues* are referred to as the residues which are to be eliminated during model condensation. The dynamic motion of a protein structure is governed by harmonic potential V in the form of

$$V = \frac{1}{2}\begin{bmatrix}\mathbf{u}_M & \mathbf{u}_S\end{bmatrix}\begin{bmatrix}\mathbf{K}_M & \mathbf{K}_{MS} \\ \mathbf{K}_{SM} & \mathbf{K}_S\end{bmatrix}\begin{bmatrix}\mathbf{u}_M \\ \mathbf{u}_S\end{bmatrix} \tag{11}$$

where the subscripts M and S indicate the master residues and slave residues, respectively. $\mathbf{K}_M$ represents the harmonic interactions between master residues, $\mathbf{K}_S$ provides the harmonic interactions between slave residues, and $\mathbf{K}_{MS}$ shows the harmonic interactions between master and slave residues. With assumption that slave residues are in equilibrium, the effective stiffness matrix $\mathbf{K}_{eff}$ for coarse-grained ENM described by master residues is computed as follows.

$$\mathbf{K}_{eff} = \boldsymbol{\psi}[\mathbf{K}] = \begin{bmatrix} \mathbf{I}_{3M} & -\mathbf{K}_{MS}\mathbf{K}_{S}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{K}_{M} & \mathbf{K}_{MS} \\ \mathbf{K}_{SM} & \mathbf{K}_{S} \end{bmatrix} \begin{bmatrix} \mathbf{I}_{3M} \\ \mathbf{0} \end{bmatrix} \tag{12}$$

Here, $\boldsymbol{\psi}$ is the linear operator that transfroms the original structure described by stiffness matrix $\mathbf{K}$ to the coarse-grained structure depicted by effective stiffness matrix $\mathbf{K}_{eff}$, and $\mathbf{I}_{3M}$ is the $3N_M \times 3N_M$ identity matrix, where N_M is the total number of master residues.

CONFORMATIONAL FLUCTUATION DYNAMICS

In recent decades, the molecular structures of various proteins have been characterized by experiments based on X-ray crystallography and/or nuclear magnetic resonance (NMR) [20]. Until recently, many experimentalists are attempting to characterize the large protein structures based on X-ray crystallography and NMR, and such protein structures realized by experimentalists are deposited in the protein data bank (PDB; http://www.pdb.org). Characterization of protein structure with such experiments is typically given in terms of Debye-Waller factor (B-factor) representing the mean-square fluctuation of residues driven by thermal energy k_BT. Consequently, the dynamic behavior of proteins based on theoretical models such as molecular model and/or coarse-grained model is typically compared with B-factor obtained by experiments. That is, the conformational fluctuation behavior of proteins plays a role in validating the theoretical models for protein structures.

As shown in Fig. 3, the conformational fluctuation predicted by Tirion's model (ENM) and/or GNM is quantitatively comparable to that obtained by experiments. It is quite remarkable that simple harmonic oscillator network model delineated by two parameters such as force constant and cut-off distance are able to provide the conformational fluctuation of proteins. This remarkable result indicates that native topology (topology of native contacts) plays a dominant role in the conformational fluctuation. Moreover, the comparison of thermal fluctuations predicted by ENM with that by experiments provides the force constant for an entropic spring connecting the native contacts. For instance, F_1-ATPase motor protein (pdb: 1e79) can be represented by GNM with force constant of 0.347 kcal/mol and cut-off distance of 12Å. It should be noted that one has to be cautious in selecting the cut-off distance because the short cut-off distance may generate the unphysical behavior of a structure such as more than six rigid body modes [33]. Further, if one chooses the very large cut-off distance, then the structure is too rigid to fluctuate in the similar pattern to that of real protein.

We take into account the coarse-grained elastic network model and its conformational fluctuation behavior. It is shown that, in Fig. 3, coarse-grained ENM predicts the thermal fluctuation behavior depicted by B-factor qualitatively comparable to that estimated by experiments and/or original structural model. For a protein composed of N α-carbon atoms, the prediction of B-factor based on ENM requires $O(N^3)$ computation, while on the basis of coarse-grained ENM composed of (N/n) α-carbon atoms the calculation of B-factor requires $O(N^3/n^2)$ computation. Coarse-grained ENM reduces the computational cost for predicting thermal fluctuation of proteins by factor of n^2, compared with ENM,

whereas the coarse-grained ENM predicts the thermal fluctuation of proteins quantitatively and qualitatively comparable to that predicted by ENM. The success of coarse-grained ENM in depicting the conformational fluctuation of proteins may be attributed to the fact that protein structure is usually represented by combinations of rigid domains that can be described by few degrees of freedom. This feature of protein structure has been taken into account for establishing the coarse-grained models of proteins. For instance, Jernigan and coworkers [66] provide the rigid cluster model such that protein structure is represented by clusters of rigid bodies with soft springs connecting rigid domains. Further, Tama and coworkers [69] suggested block normal mode analysis, where block matrices were used to describe the rigid blocks of proteins, for delineating the conformational fluctuation of proteins.

Moreover, as the protein structure is more coarse-grained, the magnitude of conformational fluctuation becomes larger, even though the patterns of conformational fluctuation predicted by coarse-grained structure are qualitatively consistent with original structure. This is rational since our coarse-graining scheme reduces the harmonic springs corresponding to slave residues, resulting in the increase of overall compliance of protein structure. This is consistent with a recent work by Bahar and coworkers [51], where they rescaled the force constant in such a way that the force constant for coarse-grained structure is larger than that for original structure. In order for a coarse-grained ENM to predict the conformational fluctuation quantatively comparable to experimental data or original structural model, the force constant should be rescaled in such a way that the overall stiffness of protein structure described by coarse-grained ENM is comparable to that of protein structure depicted by ENM. Fig. 4 shows the thermal fluctuation predicted by coarse-grained ENM with rescaled force constant. It is shown that the conformational fluctuation predicted by coarse-grained ENM is very consistent with experimental data.

LOWEST-FREQUENCY NORMAL MODE

Coarse-grained models such as Go model and ENM are computationally acceptable for computational biology communities, since such models are able to capture the low-frequency normal modes relevant to the biological function of proteins. Such coarse-grained models reduce the degrees of freedom enormously as well as simplify the potential field, and they can provide the meaningful low-frequency normal modes comparable to that computed from atomistic model. This indicates that a *de novo* coarse-grained model for protein structures can be verified based on the comparison of low-frequency normal modes computed from such coarse-grained model with that obtained by conventional models such as atomistic model and/or well-accepted coarse-grained model such as Go model and Tirion's model.

We have validated our coarse-grained ENM by investigating the low-frequency normal modes predicted by coarse-grained ENM. For instance, we consider the lowest-frequency normal modes predicted by both ENM and coarse-grained ENM. As shown in Fig.5, the lowest-frequency normal mode for hemoglobin is well predicted by coarse-grained ENM such that its lowest-frequency normal mode is qualitatively comparable to that obtained from ENM. Specifically, anti-correlated motion between

substructure *A* (residues: 1 ~ 287) and substructure *B* (residues: 288 ~ 428) can be found in both ENM and coarse-grained ENM. This indicates that coarse-grained ENM may provide the lowest-frequency normal mode, related to the functional motion of protein structure, qualitatively comparable to that computed from original structural model such as Go model and ENM. Further, the rescaling of force constant for coarse-grained ENM may not affect the pattern of lowest-frequency normal mode, since the protein topology is only described by cut-off distance. It can be realized that our coarse-graining allows us to predict the functional lowest-frequency normal mode of proteins with reducing the computational cost by factor of n^3.

COLLECTIVE AND CORRELATED MOTION OF PROTEINS

The conformational motion of proteins has been well described as collective motion and/or correlated motion. As shown previously in Fig. 5, the low-frequency normal modes exhibit the collective motion of a protein domain, and such modes depict the correlated motion of protein domains. Before we demonstrate the collective and/or correlated motion predicted by ENM and/or coarse-grained ENM, we review the parameters representing the collective motion and/or correlated motion. The collectivity parameter denoted as κ_i for a given mode index i is defined as [70]

$$\kappa_i = \frac{1}{N_\omega} \exp\left[-\sum_{j=1}^{N_\omega} \left|v_{i,j}\right|^2 \log\left|v_{i,j}\right|^2 \right] \tag{13}$$

where N_ω is the total number of normal modes, $v_{i,j}$ represents the j-th component of normal mode $\mathbf{v}_i$ corresponding to mode index i. The collectivity κ_i is in the range between $1/N_\omega$ and 1, where the value of collectivity close to $1/N_\omega$ represents the localized motion while the value of collectivity close to 1 indicates the collective motion. The correlated motion between residues i and j is well delineated by correlation matrix $\mathbf{C}_{ij}$ defined as [71]

$$\mathbf{C}_{ij} = \frac{\left\langle \left(\mathbf{x}_i - \mathbf{x}_i^0\right) \cdot \left(\mathbf{x}_j - \mathbf{x}_j^0\right) \right\rangle}{\sqrt{\left\langle \left|\mathbf{x}_i - \mathbf{x}_i^0\right|^2 \right\rangle \left\langle \left|\mathbf{x}_j - \mathbf{x}_j^0\right|^2 \right\rangle}} = \frac{\sum_{p=1}^{3} \mathbf{L}_{3(i-1)+p,3(j-1)+p}}{\sqrt{\left(\sum_{p=1}^{3} \mathbf{L}_{3(i-1)+p,3(i-1)+p} \right) \left(\sum_{q=1}^{3} \mathbf{L}_{3(j-1)+q,3(j-1)+q} \right)}} \tag{14}$$

Here, the correlation matrix $\mathbf{C}_{ij}$ in terms of cross-correlation matrix $\mathbf{L}_{ij}$, shown in Eq. 14, is based on ENM whose degrees of freedom are $3N$. The value of $\mathbf{C}_{ij}$ close to -1 shows the anti-correlated motion between residues i and j, whereas the value of $\mathbf{C}_{ij}$ close to 1 indicates the correlated motion between these two residues. When the correlation $\mathbf{C}_{ij}$ is close to zero, the motion of a residue i is uncorrelated with and/or orthogonal to that of a residue j.

For clear understanding of correlated motion described by $\mathbf{C}_{ij}$, let us consider the simple harmonic oscillator embedded in a heat bath with thermal energy k_BT. The potential energy for a harmonic oscillator is represented in the form of $V = (\gamma/2)(u_i - u_j)^2$, where γ is a force constant (spring constant) and u_i is a one-dimensional displacement for a node i (see Fig. 6). The Hessian matrix (stiffness

matrix) for this system can be given by

$$\mathbf{K} = \begin{bmatrix} \gamma & -\gamma \\ -\gamma & \gamma \end{bmatrix} \tag{15}$$

which provides the natural frequencies $\omega_0 = 0$ and $\omega_1 = (2\gamma)^{1/2}$ and their corresponding normal modes $\mathbf{v}_0 = (1, 1)$ and $\mathbf{v}_1 = (1, -1)$. As stated earlier, the zero modes should be excluded for estimating the thermal fluctuation of a system. The non-zero normal mode $\mathbf{v}_1$ for a harmonic oscillator enables us to easily know that the thermal energy drives the anti-correlated motion between two nodal points i and j. This is consitent with the quantity of correlation $\mathbf{C}_{ij}$, i.e. $\mathbf{C}_{ij} = -1$, from the definition of correlation $\mathbf{C}_{ij}$ such as $\mathbf{C}_{ij} = \mathbf{L}_{ij}/(\mathbf{L}_{ii}\mathbf{L}_{jj})^{1/2}$, where cross-correlation $\mathbf{L}_{ij}$ is given by

$$\mathbf{L} = \frac{k_B T}{\omega_1^2} \mathbf{v}_1 \otimes \mathbf{v}_1 = \frac{k_B T}{2\gamma} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \tag{16}$$

Thus, correlation $\mathbf{C}_{ij}$ is a physical parameter describing the correlated motion between these two nodal points.

As shown in Fig. 7, we consider the collectivity parameters κ_i calculated from both ENM and coarse-grained ENM. It is quite remarkable that coarse-grained ENM is able to reproduce the collectivity parameters corresponding to low-frequency normal mode quantitatively comparable to that estimated from ENM. This indicates that collective motion of proteins can be well depicted by coarse-grained structure represented by few degrees of freedom. This may be attributed to the fact that protein consists of several rigid domains that can be described by few degrees of freedom, and that the collective motion arises from the low-frequency functional modes. However, the coarse-grained ENM cannot predict the collectivity for high-frequency normal modes. Specifically, as shown in Fig. 7, the high-frequency normal modes are related to localized motion which cannot be predicted from coarse-grained ENM. This indicates that localized modes (high-frequency modes) of protein can be only estimated from refined molecular model. Fig. 8 shows the correlation matrix $\mathbf{C}_{ij}$ evaluated from ENM and coarse-grained ENM. It is remarkably found that the collective motion of each domain is well described by both ENM and coarse-grained ENM. Further, coarse-grained ENM provides the correlation between domains, qualitatively comparable to the correlation predicted by ENM. However, coarse-grained ENM overestimates the quantity of correlation between domains. This may be ascribed to our coarse-graining scheme, that is, reduction of harmonic springs corresponding to slave residues, leading to overestimation of overall flexibility and its corresponding correlated motion between domains.

CONFORMATIONAL TRANSITION

Conformational change of a protein is quite related to the biological function of a protein. Atomistic simulation such as targeted MD simulation has been employed for understanding conformational change of very small proteins. Remarkably, NMA has been an alternative to MD simulation, since the low-frequency normal modes at equilibrium state are able to well describe the conformational change of

proteins. This NMA is referred to as principal component analysis (PCA) that diagonalizes the Hessian matrix (stiffness matrix) [72].

Since it was shown that low-frequency normal modes are independent of details of potential field [58] but depend on the topology of protein structures [73], ENM was broadly employed for understanding the conformational change of proteins. Tama and coworkers [37, 38] showed that low-frequency normal modes obtained from ENM are highly correlated with a vector representing the conformational change between two equilibrium states. Bahar and coworkers [74] showed that the conformational change from tense form to relaxed form for hemoglobin is driven by entropic effect described by low-frequency normal modes from ENM. Brooks and coworkers [41, 67] predicted the conformational change depicted by low-frequency normal modes with a perturbation of Tirion's potential that incorporates the distance constraint. Kim *et al* [40] suggested the linear interpolation between two conformations with constraint that the intermediate conformation distant from the interpolated coordinate is determined by minimization of harmonic potential. Wolynes and coworkers [39] provided the nonlinear elastic energy landscape for conformational change of proteins based on low-frequency normal modes from ENM. Further, Karplus and coworkers [75] employed the same methodology suggested by Wolynes and coworkers for describing the conformational change of a motor protein. Further, Kidera and coworkers [76] used the linear response theory with Tirion's model for depicting the conformational change of proteins.

For delineating the correlation between low-frequency normal modes and conformational change, the parameters referred to as overlap I_k and/or cumulative involvement S_k are defined such as

$$I_k = \frac{\left|\left(\mathbf{r}_{open} - \mathbf{r}_{close}\right) \cdot \mathbf{v}_k\right|}{\left|\left(\mathbf{r}_{open} - \mathbf{r}_{close}\right)\right|} \tag{17.a}$$

and $$S_k = \sum_{p=1}^{k} I_p^{\,2} \tag{17.b}$$

Here, $\mathbf{r}_{open}$ and $\mathbf{r}_{close}$ represent the position vectors corresponding to open form and close form, respectively, and $\mathbf{v}_k$ is the k-th normal mode. I_k indicates the correlation between k-th normal mode and conformational change, and S_k is a quantity representing the contribution from low-frequency normal modes (from first mode to k-th mode) to the conformational change. Fig. 9 shows the overlap and cumulative involvement predicted by ENM and coarse-grained ENM. It is remarkable that both models predict that the conformational change is highly correlated with a few low-frequency normal modes. This is consistent with a recent finding that low-frequency normal modes are sufficient to represent the conformational change of a protein.

CONCLUSION

In this article, we review the current state-of-art in coarse-graining of protein molecules for understanding

their dynamics relevant to biological functions. The coarse-graining procedure is usually acceptable as long as the protein topology related to its dynamics is sufficiently delineated by such coarse-grained models. We briefly overviewed the broadly accepted coarse-grained models such as Go model and Tirion's model (ENM), which enables one to gain insight into protein dynamics such as conformational fluctuation and conformational change related to the biological function. Moreover, a recently developed coarse-grained ENM models are taken into account and it is shown that such coarse-grained ENM may allow one to achieve the fast computation on low-frequency normal modes related to biological function. It is provided that the possibility of coarse-graining for a protein structure is attributed to the fact that protein structure is usually composed of several rigid domains that can be described by few degrees of freedom. As previously shown, both ENM and coarse-grained ENM predicts the low-frequency normal modes and the thermal fluctuation quantitatively similar to that obtained by experiments. Further, both ENM and coarse-grained model such as rigid cluster model predict the conformational transitions between two conformations. However, it has to be validated whether coarse-grained ENM is acceptable or not for prediction of conformational change. To our knowledge, this issue has not been well considered except a recent work by Brooks and coworkers [67] who employed mixed ENM for understanding conformational change.

As stated above, coarse-grained models have been successful for studying the conformational dynamics of proteins. However, since some proteins such as titin perform the mechanical function, the protein unfolding behavior has to be well understood for insight into the mechanical function. Atomistic simulation is still restricted to small proteins, leading to consideration of coarse-grained models. A recent study by Cieplak and coworkers [28] suggested the molecular model based on Go potential under the mechanical loading. It is very remarkable that their model based on Go potential allows them to predict the force-displacement relation under the mechanical loading, comparable to the results of AFM experiments. Moreover, McCammon and coworkers [77] employed Tirion's potential with mechanical loading applied to the termini of a protein. It was remarkably shown that even Tirion's model is acceptable for gaining insight into mechanical unfolding of proteins. Recently, Rief and coworker [78] revisited Tirion's model with bond-breaking model for protein unfolding mechanics. It is remarkably found that their elastic bond network model [78] allowed Rief and coworker to predict the probability distribution of rupture force, quantitatively comparable to AFM experimental data [79].

In summary, the coarse-grained models such as Go model and Tirion model have been reviewed for protein dynamics relevant to biological function. Moreover, such models can be extended for understanding mechanical unfolding of protein structure. In conclusion, coarse-grained models such as Tirion's model may be versatile for understanding the large protein dynamics and/or large protein unfolding mechanics.

ACKNOWLEDGEMENT

This work was supported in part by Nano-Bio Research Center at KIST (to K.E.) and LG YONAM FOUNDATION, Basic Research Program of the Korea Science & Engineering Foundation (KOSEF) under grant No. R01-2007-000-10497-0, and Korea Ministry of Science & Technology under grant No. R-11-2007-028-00000-0 (to S.N.).

REFERENCES

[1] Kolomeisky, A.B. and M.E. Fisher, *Molecular Motors: A Theorist's Perspective.* Annu. Rev. Phys. Chem., 2007. **58**: 675.

[2] Duncan, T.M., V.V. Bulygin, Y. Zhou, M.L. Hutcheon, and R.L. Cross, *Rotation of subunits during catalysis by Escherichia coli F1-ATPase.* Proc. Natl. Acad. Sci. USA, 1995. **92**: 10964.

[3] Kinoshita, K., R. Yasuda, K. Noji, S. Ishiwata, and M. Yoshida, *F1-ATPase: a rotary motor made of a single molecule.* Cell, 1998. **93**: 21.

[4] Noji, H., R. Yasuda, M. Yoshida, and K. Kinosita, *Direct observation of the rotation of F1-ATPase.* Nature, 1997. **386**: 299.

[5] Sabbert, D., S. Engelbrecht, and W. Junge, *Functional and idling rotatory motion within F1-ATPase.* Proc. Natl. Acad. Sci. USA, 1997. **94**: 4401.

[6] Yasuda, R., H. Noji, K. Kinosita, and M. Yoshida, *F1-ATPase is a highly efficient molecular motor that rotates with discrete 120^o steps.* Cell, 1998. **93**: 1117.

[7] Ranson, N.A., D.K. Clare, G.W. Farr, D. Houldershaw, A.L. Horwich, and H.R. Saibil, *Allosteric signaling of ATP hydrolysis in GroEL-GroES complexes.* Nat. Struct. Mol. Biol., 2006. **13**: 147.

[8] Keskin, O., I. Bahar, D. Flatow, D.G. Covell, and R.L. Jernigan, *Molecular mechanisms of chaperonin GroEL-GroES function.* Biochemistry, 2002. **41**: 491.

[9] Marszalek, P.E., H. Lu, H.B. Li, M. Carrion-Vazquez, A.F. Oberhauser, K. Schulten, and J.M. Fernandez, *Mechanical unfolding intermediates in titin modules.* Nature, 1999. **402**: 100.

[10] Carrion-Vazquez, M., A.F. Oberhauser, T.E. Fisher, P.E. Marszalek, H. Li, and J.M. Fernandez, *Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering.* Prog. Biophys. Mol. Biol., 2000. **74**: 63.

[11] Li, H., A.F. Oberhauser, S.B. Fowler, J. Clarke, and J.M. Fernandez, *Atomic force microscopy reveals the mechanical design of a modular protein.* Proc. Natl. Acad. Sci. USA, 2000. **97**: 6527.

[12] Oberhauser, A.F., P.E. Marszalek, H.P. Erickson, and J.M. Fernandez, *The molecular elasticity of the extracellular matrix protein tenascin.* Nature, 1998. **393**: 181.

[13] Schafer, L.V., E.M. Muller, H.E. Gaub, and H. Grubmuller, *Elastic Properties of Photoswitchable Azobenzene Polymers from Molecular Dynamics Simulations.* Angew. Chem. Int. Ed., 2007. **46**: 2232.

[14] McCammon, J.A. and S. Harvey, *Dynamics of proteins and nucleic acids*. 1987, Cambridge: Cambridge University Press.

[15] McCammon, J.A., B.R. Gelin, and M. Karplus, *Dynamics of folded proteins.* Nature, 1977. **267**: 585.

[16] Amadei, A., A.B.M. Linssen, and H.J.C. Berendsen, *Essential Dynamics of Proteins.* Proteins: Struct. Funct. Genet., 1993. **17**: 412.

[17] Lu, H., B. Isralewitz, A. Krammer, V. Vogel, and K. Schulten, *Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation.* Biophys. J., 1998. **75**: 662.

[18] Lu, H. and K. Schulten, *Steered molecular dynamics simulations of force-induced protein domain unfolding.* Proteins, 1999. **35**: 453.

[19] Sotomayor, M. and K. Schulten, *Single-Molecule Experiments in Vitro and in Silico.* Science, 2007. **316**: 1144.

[20] Brooks, C.L., M. Karplus, and B.M. Pettit, Adv. Chem. Phys., 1988. **71**: 1.

[21] Elber, R., *Long-timescale simulation methods.* Curr. Opin. Struct. Biol., 2005. **15**: 151.

[22] Hayward, S. and N. Go, *Collective Variable Description of Native Protein Dynamics.* Annu. Rev. Phys. Chem., 1995. **46**: 223.

[23] Cui, Q., G.H. Li, J.P. Ma, and M. Karplus, *A normal mode analysis of structural plasticity in the biomolecular motor F-1-ATPase.* J. Mol. Biol., 2004. **340**: 345.

[24] Ma, J.P., *Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes.* Structure, 2005. **13**: 373.

[25] Cieplak, M., T.X. Hoang, and M.O. Robbins, *Thermal folding and mechanical unfolding pathways of protein secondary structures.* Proteins: Struct. Funct. Genet., 2002. **49**: 104.

[26] Cieplak, M., T.X. Hoang, and M.O. Robbins, *Folding and stretching in a Go-like model of titin.* Proteins: Struct. Funct. Genet., 2002. **49**: 114.

[27] Cieplak, M., T.X. Hoang, and M.O. Robbins, *Thermal effects in stretching of Go-like models of titin and secondary structures.* Proteins: Struct. Funct. Bioinfo., 2004. **56**: 285.

[28] Marek, C., P. Annalisa, and H. Trinh Xuan, *Mechanical properties of the domains of titin in a Go-like model.* J. Chem. Phys., 2005. **122**: 054906.

[29] Tirion, M.M., *Large amplitude elastic motions in proteins from a single-parameter, atomic analysis.* Phys. Rev. Lett., 1996. **77**: 1905.

[30] Haliloglu, T., I. Bahar, and B. Erman, *Gaussian dynamics of folded proteins.* Phys. Rev. Lett., 1997. **79**: 3090.

[31] Bahar, I., A.R. Atilgan, M.C. Demirel, and B. Erman, *Vibrational dynamics of folded proteins: Significance of slow and fast motions in relation to function and stability.* Phys. Rev. Lett., 1998. **80**: 2733.

[32] Bahar, I., B. Erman, R.L. Jernigan, A.R. Atilgan, and D.G. Covell, *Collective motions in HIV-1 reverse transcriptase: Examination of flexibility and enzyme function.* J. Mol. Biol., 1999. **285**: 1023.

[33] Atilgan, A.R., S.R. Durell, R.L. Jernigan, M.C. Demirel, O. Keskin, and I. Bahar, *Anisotropy of fluctuation dynamics of proteins with an elastic network model.* Biophys. J., 2001. **80**: 505.

[34] Bahar, I. and A.J. Rader, *Coarse-grained normal mode analysis in structural biology.* Curr. Opin. Struct.

Biol., 2005. **15**: 586.

[35] Tozzini, V., *Coarse-grained models for proteins.* Curr. Opin. Struct. Biol., 2005. **15**: 144.

[36] Cui, Q. and I. Bahar, *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems*. 2005: CRC Press.

[37] Tama, F. and C.L. Brooks, *Symmetry, form, and shape: Guiding principles for robustness in macromolecular machines.* Annu. Rev. Biophys. Biomol. Struct., 2006. **35**: 115.

[38] Tama, F. and Y.H. Sanejouand, *Conformational change of proteins arising from normal mode calculations.* Protein Eng., 2001. **14**: 1.

[39] Miyashita, O., J.N. Onuchic, and P.G. Wolynes, *Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins.* Proc. Natl. Acad. Sci. USA., 2003. **100**: 12570.

[40] Kim, M.K., W. Li, B.A. Shapiro, and G.S. Chirikjian, *A comparison between elastic network interpolation and MD simulation of 16S ribosomal RNA.* J. Biomol. Struct. Dyn., 2003. **21**: 395.

[41] Zheng, W.J. and B.R. Brooks, *Normal-modes-based prediction of protein conformational changes guided by distance constraints.* Biophys. J., 2005. **88**: 3109.

[42] Zheng, W.J. and B.R. Brooks, *Modeling protein conformational changes by iterative fitting of distance constraints using reoriented normal modes.* Biophys. J., 2006. **90**: 4327.

[43] Micheletti, C., J.R. Banavar, and A. Maritan, *Conformations of Proteins in Equilibrium.* Phys. Rev. Lett., 2001. **87**: 088102.

[44] Zheng, W.J. and S. Doniach, *A comparative study of motor-protein motions by using a simple elastic-network model.* Proc. Natl. Acad. Sci. USA., 2003. **100**: 13253.

[45] Tama, F. and C.L. Brooks, *Diversity and Identity of Mechanical Properties of Icosahedral Viral Capsids Studied with Elastic Network Normal Mode Analysis.* J. Mol. Biol., 2005. **345**: 299.

[46] Kim, M.K., R.L. Jernigan, and G.S. Chirikjian, *An elastic network model of HK97 capsid maturation.* J. Struct. Biol., 2003. **143**: 107.

[47] Zheng, W.J., B.R. Brooks, and D. Thirumalai, *Low-frequency normal modes that describe allosteric transitions in biological nanomachines are robust to sequence variations.* Proc. Natl. Acad. Sci. USA., 2006. **103**: 7664.

[48] Tobi, D. and I. Bahar, *Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state.* Proc. Natl. Acad. Sci. USA., 2005. **102**: 18908.

[49] Chennubhotla, C. and I. Bahar, *Markov propagation of allosteric effects in biomolecular systems: application to GroEL-GroES.* Mol. Syst. Biol., 2006. **2**: Article No 36.

[50] Chennubhotla, C. and I. Bahar, *Signal propagation in proteins and relation to equilibrium fluctuations.* PLOS Computat. Biol., 2007. **3**: 1716.

[51] Doruker, P., R.L. Jernigan, and I. Bahar, *Dynamics of large proteins through hierarchical levels of coarse-grained structures.* J. Comput. Chem., 2002. **23**: 119.

[52] Eom, K., S.-C. Baek, J.-H. Ahn, and S. Na, *Coarse-graining of protein structures for normal mode studies.*

J. Comput. Chem., 2007. **28**: 1400.

[53] Cheng, H., Z. Gimbutas, P.G. Martinsson, and V. Rokhlin, *On the compression of low rank matrices.* SIAM J. Sci. Comput., 2005. **26**: 1389.

[54] Liberty, E., F. Woolfe, P.-G. Martinsson, V. Rokhlin, and M. Tygert, *Randomized algorithms for the low-rank approximation of matrices.* Proc. Natl. Acad. Sci. USA., 2007. **104**: 20167.

[55] Chennubhotla, C. and I. Bahar, *Markov methods for hierarchical coarse-graining of large protein dynamics*, in *Lecture Notes in Computer Science*. 2006. p. 379.

[56] Ming, D., Y. Kong, Y. Wu, and J. Ma, *Substructure synthesis method for simulating large molecular complexes.* Proc. Natl. Acad. Sci., 2003. **100**: 104.

[57] Chandler, D., *Introduction to modern statistical mechanics*. 1987: Oxford University Press.

[58] Teeter, M.M. and D.A. Case, *Harmonic and quasiharmonic descriptions of crambin.* J. Phys. Chem., 1990. **94**: 8091.

[59] Weiner, J.H., *Statistical mechanics of elasticity*. 1983: Dover publication.

[60] Doi, M. and S.F. Edwards, *The Theory of Polymer Dynamics*. 1986, New York: Oxford University Press.

[61] Makarov, D.E. and G.J. Rodin, *Configurational entropy and mechanical properties of cross-linked polymer chains: Implications for protein and RNA folding.* Phys. Rev. E., 2002. **66**: 011908.

[62] Eom, K., P.C. Li, D.E. Makarov, and G.J. Rodin, *Relationship between the Mechanical Properties and Topology of Cross-Linked Polymer Molecules: Parallel Strands Maximize the Strength of Model Polymers and Protein Domains.* J. Phys. Chem. B, 2003. **107**: 8730.

[63] Eom, K., D.E. Makarov, and G.J. Rodin, *Theoretical studies of the kinetics of mechanical unfolding of cross-linked polymer chains and their implications for single-molecule pulling experiments.* Phys. Rev. E., 2005. **71**: 021904.

[64] Eom, K., J.H. Ahn, S.C. Baek, J.I. Kim, and S. Na, *Robust reduction method for biomolecules modeling.* CMC-Computers Materials & Continua, 2007. **6**: 35.

[65] Kurkcuoglu, O., R.L. Jernigan, and P. Doruker, *Mixed levels of coarse-graining of large proteins using elastic network model succeeds in extracting the slowest motions.* Polymer, 2004. **45**: 649.

[66] Kim, M.K., R.L. Jernigan, and G.S. Chirikjian, *Rigid-cluster models of conformational transitions in macromolecular machines and assemblies.* Biophys. J., 2005. **89**: 43.

[67] Zheng, W., B.R. Brooks, and G. Hummer, *Protein conformational transitions explored by mixed elastic network models.* Proteins: Struct. Funct. Bioinfo., 2007. **69**: 43.

[68] Martinsson, P.G., *Fast evaluation of electro-static interactions in multi-phase dielectric media.* J. Comput. Phys., 2006. **211**: 289.

[69] Tama, F., F.X. Gadea, O. Marques, and Y.H. Sanejouand, *Building-block approach for determining low-frequency normal modes of macromolecules.* Proteins: Struct. Funct. Genet., 2000. **41**: 1.

[70] Lienin, S.F. and R. Bruschweiler, *Characterization of collective and anisotropic reorientational protein dynamics.* Phys. Rev. Lett., 2000. **84**: 5439.

[71] Van Wynsberghe, A.W. and Q. Cui, *Interpreting correlated motions using normal mode analysis.* Structure, 2006. **14**: 1647.

[72] Lou, H. and R.I. Cukier, *Molecular Dynamics of Apo-Adenylate Kinase: A Principal Component Analysis.* J. Phys. Chem. B, 2006. **110**: 12796.

[73] Lu, M.Y. and J.P. Ma, *The role of shape in determining molecular motions.* Biophys. J., 2005. **89**: 2395.

[74] Xu, C.Y., D. Tobi, and I. Bahar, *Allosteric changes in protein structure computed by a simple mechanical model: Hemoglobin T <-> R2 transition.* J. Mol. Biol., 2003. **333**: 153.

[75] Maragakis, P. and M. Karplus, *Large amplitude conformational change in proteins explored with a plastic network model: Adenylate kinase.* J. Mol. Biol., 2005. **352**: 807.

[76] Ikeguchi, M., J. Ueno, M. Sato, and A. Kidera, *Protein structural change upon ligand binding: Linear response theory.* Phys. Rev. Lett., 2005. **94**.

[77] Shen, T., L.S. Canino, and J.A. McCammon, *Unfolding Proteins under External Forces: A Solvable Model under the Self-Consistent Pair Contact Probability Approximation.* Phys. Rev. Lett., 2002. **89**: 068103.

[78] Dietz, H. and M. Rief, *An elastic bond network model for protein unfolding mechanics*, unpublished.

[79] Dietz, H., F. Berkemeier, M. Bertz, and M. Rief, *Anisotropic deformation response of single protein molecules.* Proc. Natl. Acad. Sci. USA., 2006. **103**: 12724.

Figure Captions

Fig. 1. Model protein, i.e. citrate synthase (pdb: 4cts) described by (a) molecular structure and (b) elastic network model

Fig. 2. Molecular structure of a model protein (citrate synthase) delineated by (a) elastic network model and (b) coarse-grained elastic network model

Fig. 3. B-factor of a motor protein (pdb: 1e79) predicted by elastic network model and coarse-grained elastic network model in comparison with experimental data

Fig. 4. Comparison between experimental data and B-factor predicted by coarse-grained elastic network model with rescaled force constant

Fig. 5. Lowest-frequency normal mode for a hemoglobin computed from elastic network model and coarse-grained elastic network model

Fig. 6. Schematic of a one-dimensional harmonic oscillator that undergoes the thermal fluctuation

Fig. 7. Collectivity parameter κ_i for a hemoglobin (pdb: 1a3n) estimated from elastic network model and coarse-grained elastic network model

Fig. 8. Correlation matrix $\mathbf{C}_{ij}$ for a motor protein (pdb: 1e79) evaluated by (a) elastic network model and (b) coarse-grained elastic network model

Fig. 9. Overlap I_k and cumulative involvement S_k for citrate synthase computed from elastic network model and coarse-grained elastic network model. Blue color represents the calculation based on elastic network model, whereas red color indicates the computation based on coarse-grained elastic network model. A bar graph shows the square of overlap, and dotted line presents the cumulative involvement.

(a) (b)

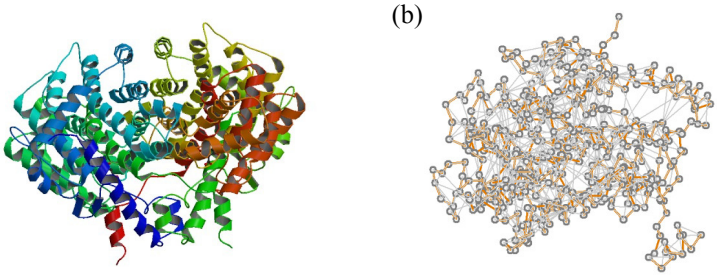

Fig. 1.

(a) (b)

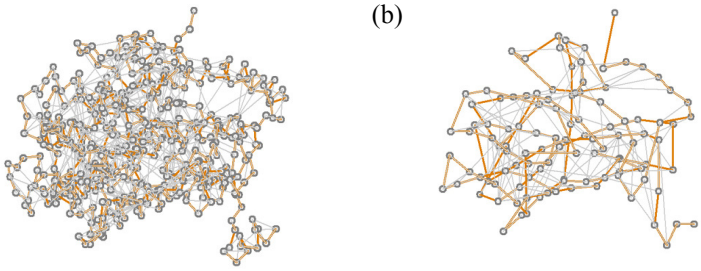

Fig. 2.

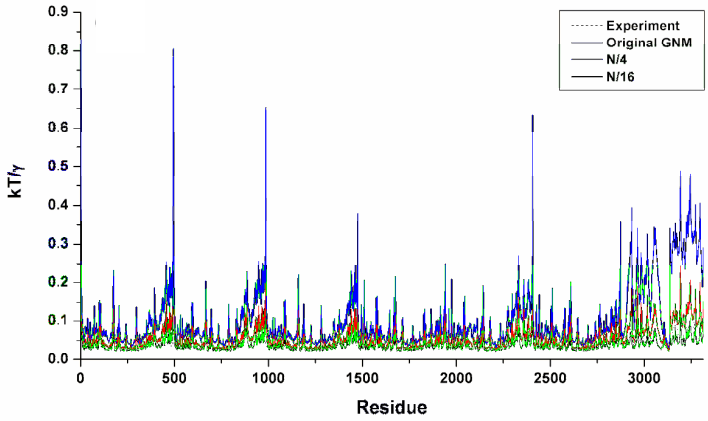

Fig. 3.

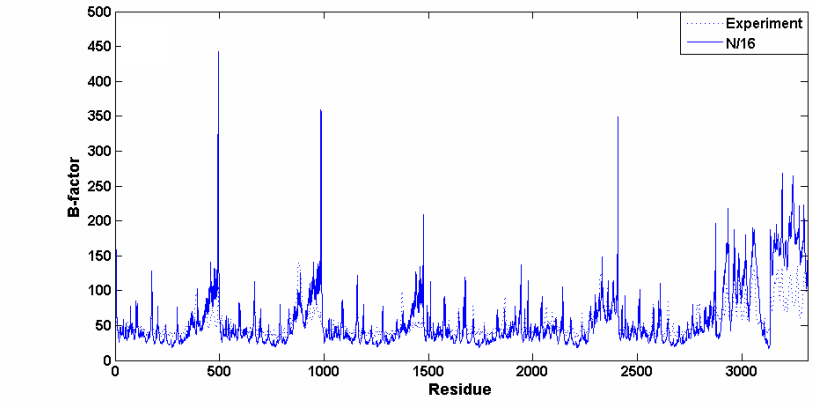

Fig. 4.

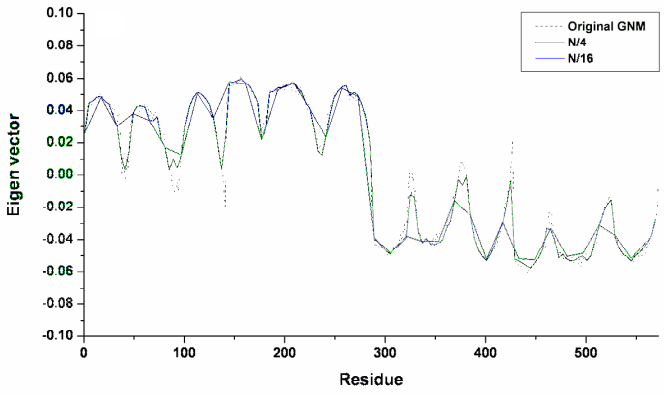

Fig. 5.

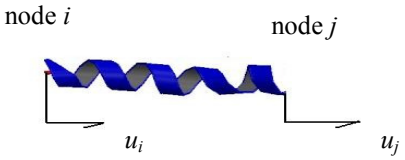

Fig. 6.

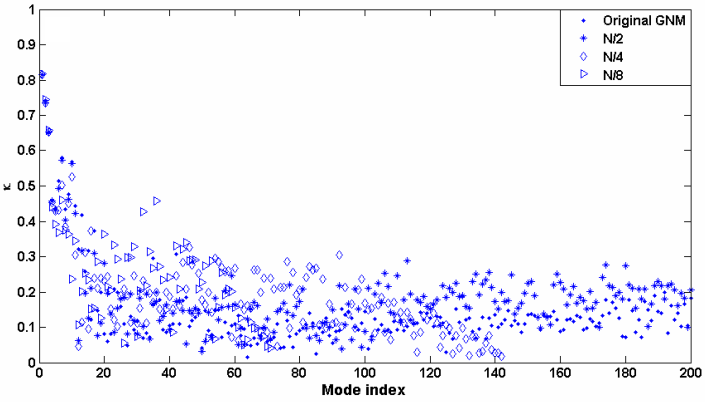

Fig. 7.

(a) (b)

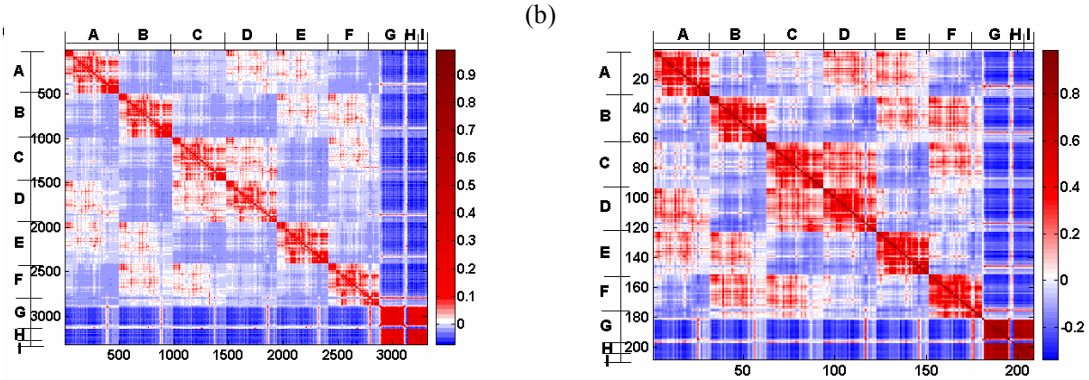

Fig. 8.

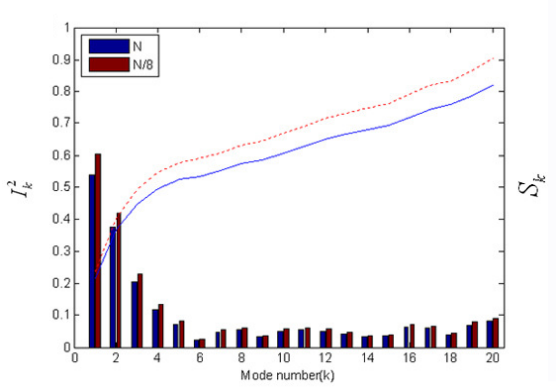

Fig. 9.