JCTC Journal of Chemical Theory and Computation

Large Protein Dynamics Described by Hierarchical-Component Mode Synthesis

Jae-In Kim, Sungsoo Na,* and Kilho Eom*

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

Received January 15, 2009

Abstract: Protein dynamics has played a pivotal role in understanding the biological function of protein. For investigation of such dynamics, normal-mode analysis (NMA) has been broadly employed with atomistic model and/or coarse-grained models such as elastic network model (ENM). For large protein complexes, NMA with even ENM encounters the expensive computational process such as diagonalization of Hessian (stiffness) matrix. Here, we suggest the hierarchical-component mode synthesis (hCMS), which allows the fast computation of low-frequency normal modes related to conformational change. Specifically, a large protein structure is regarded as a combination of several structural units, for which the eigen-value problem is utilized for obtaining the frequencies and their normal modes for each structural unit, and consequently, such frequencies and normal modes are assembled with geometrical constraint for interface between structural units in order to find the low-frequency normal modes of a large protein complex. It is shown that hCMS is able to provide the normal modes with accuracy, quantitatively comparable to those of original NMA. This implies that hCMS may enable the computationally efficient analysis of large protein dynamics.

1. Introduction

Normal mode analysis (NMA) has enabled one to understand the protein dynamics, related to the biological function of protein, based on the low-frequency normal modes that are usually associated with the conformational change of protein.[1−3] The fundamental of NMA is to solve the eigen-value problem for diagonalization of Hessian (stiffness) matrix for protein structure.[4−6] Here, the stiffness matrix is computed based on the second-derivative of anharmonic potential field with respect to atomistic coordinates at equilibrium state, where potential is globally minimum. Complexity of potential field for protein atomistic structure leads to the computationally expensive process such as energy minimization (to find the equilibrium state) and calculation of stiffness matrix. This has led many research groups to develop the computationally efficient algorithm (or reduced model) to estimate the stiffness matrix and its related low-frequency normal modes computed from NMA with given stiffness matrix.

In a recent decade, there has been an attempt to develop the coarse-grained model for protein structure by reducing the degrees of freedom as well as simplifying the potential field. One of the successful, broadly accepted coarse-grained models is the Go model,[6−9] where only α carbon atoms are taken into account with a simplified potential field composed of a backbone covalent bond stretch and ver der Waal's interaction for native contact. The Go model has successfully predicted the protein dynamics such as conformational fluctuation dynamics[6] as well as protein unfolding mechanics.[7−10] Currently, the Go model can be regarded as a versatile model for the description of protein dynamics and/or mechanics. In a similar spirit, Tirion[11] first suggested a more simplified protein structural model, referred to as an elastic network model (ENM), in such a way that α carbon atoms are only prescribed by the harmonic potential field in the neighborhood. Despite its simplicity, ENM is able to reproduce the low-frequency normal modes and the thermal fluctuation behavior, quantitatively comparable to those estimated by experiments (X-ray crystallography or nuclear magnetic resonance) and/or atomistic simulation.[11] ENM has been broadly employed for gaining insight into the conformational

* Corresponding author e-mail: nass@korea.ac.kr (S.N.) and kilhoeom@korea.ac.kr.

10.1021/ct900027h CCC: $40.75 © 2009 American Chemical Society
Published on Web 05/26/2009

transition upon ligand-binding. For instance, Bahar and co-workers[12−15] reported that conformational change of proteins is well described by a few low-frequency normal modes. Further, several research groups[16−18] developed the model for the description of conformational transition by employing the ENM with constraints for computing the incremental displacement based on normal modes for conformational change at a certain state. Karplus and co-workers[19] reported the plastic network model (PNM) (similarly, mixed network model by Hummer and co-workers[20]) based on ENM by mixing the two potential fields near two distinct equilibrium states to find the pathway for conformational change. Recently, ENM has been employed even for studying protein mechanics such as protein unfolding mechanics. These indicate that ENM becomes a universal model for understanding the protein dynamic and/or mechanics.

However, for a large protein complex, ENM may exhibit the computational inefficiency for studying protein dynamics based on NMA. In order to overcome the computational inefficiency in obtaining the low-frequency normal modes, there have been attempts to introduce the model reduction methods applicable to ENM. For instance, Cui and co-workers[1] have implemented the block normal mode (BNM) analysis to protein structure based on atomistic potential (see Chapter 4 in ref 1). In their work, the protein motion is described in the normal modes of blocks at the residue level as well as the diagonalization scheme of sparse matrix for blocks. Bahar and co-workers[21] have first suggested the coarse-grained elastic network model composed of nodes, much less than the total number of residues, which are connected by entropic springs. Here, they have used the empirical parameters (such as force constant and cutoff distance) to describe the coarse-grained elastic network model. Recently, Eom et al.[22,23] provided the model reduction method, referred to as model condensation, inspired by the skeletonization scheme provided by Rohklin and co-workers.[24] Further, Jernigan and co-workers[25,26] reported the rigid cluster model, which regards protein structure as a combination of rigid domains connected by harmonic springs. Sanejouand and co-workers[27] developed the rotational/translational block (RTB) model, which dictates the protein structure as the rigid blocks (containing more than one residue). Those methods show that the low-resolution structure described by a few degrees of freedom is sufficient to study the protein dynamics such as conformational fluctuation. Recently, Ma and co-worker[28] suggested the coarse-grained network model based on the RTB method. Nonetheless, the quality of low-frequency normal modes is generally degraded as the protein structure is further coarse-grained. This indicates that coarse-graining of protein structure may be sometimes inappropriate for studying the conformational change that is related to low-frequency normal modes.

In this work, we report the hierarchical component mode synthesis (hCMS) for quantitative study on the low-frequency normal modes of a large protein complex. Here, a protein structure is regarded as consisting of structural subdomains, where NMA is implemented, and then such normal-mode information for each subdomain is assembled based on geometric constraint. It is shown that hCMS is capable of fast computation on low-frequency normal modes, quantitatively comparable to those obtained by conventional NMA. This implies that hCMS may enable one to study the large protein dynamics with computational efficiency as well as accuracy.

2. Model

2.1. Normal Mode Analysis (NMA) and Elastic Network Model (ENM). NMA assumes that protein motion is described by harmonic motion near equilibrium state.[1,4,5] For a given potential field V for a protein structure, the protein motion is represented in the form of $\mathbf{M}(d^2\mathbf{x}/dt^2) + \mathbf{Kx} = \mathbf{0}$, where $\mathbf{M}$ is the mass matrix (typically, diagonal matrix) and $\mathbf{K}$ is the stiffness (Hessian) matrix given by $\mathbf{K} = \partial_{\mathbf{x}}\partial_{\mathbf{x}}V$, where $\partial_{\mathbf{x}}$ is the gradient with respect to coordinates $\mathbf{x}$. Let $\mathbf{x} = \mathbf{u}\exp[i\omega t]$ with natural frequency ω and its corresponding normal mode $\mathbf{u}$. Then, the protein motion is described by an eigen-value problem as follows: $\mathbf{Ku} = \omega^2\mathbf{Mu}$.

ENM describes the protein structure as the harmonic spring network such that residues within the neighborhood are connected by a harmonic spring with an identical force constant. The potential field V for ENM is in the form of[11,13]

$$V = \frac{\gamma}{2}\sum (R_{ij} - R_{ij}^0)^2 \cdot H(R_c - R_{ij}^0) \tag{1}$$

where R_{ij} is the distance between two residues i and j, R_c is the cutoff distance given as $R_c = \sim 10$ Å, γ is the force constant, $H(x)$ is the Heaviside unit-step function, and superscript 0 indicates the equilibrium state. The potential field V can be also represented in the form of $V = (1/2)\mathbf{v}^T\mathbf{Kv}$, where $\mathbf{v}$ is the displacement vector for all residues, a symbol T represents the transpose of a vector, and $\mathbf{K}$ is the stiffness matrix composed of 3 × 3 block matrices $\mathbf{K}_{ij}$ given by

$$\mathbf{K}_{ij} = -\left[\gamma H(R_c - R_{ij}^0)\frac{(\mathbf{R}_{ij}^0)^T\mathbf{R}_{ij}^0}{(R_{ij}^0)^2}\right](1 - \delta_{ij}) - \delta_{ij}\sum_{l\neq i}\mathbf{K}_{il} \tag{2}$$

Here, $\mathbf{R}_{ij} = \mathbf{R}_j - \mathbf{R}_i$ with $\mathbf{R}_i$ being a position vector for residue i, and δ_{ij} is the Kronecker delta defined as $\delta_{ij} = 1$ if $i = j$; otherwise $\delta_{ij} = 0$.

Statistical mechanics theory allows the computation of correlation matrix $\mathbf{S}$ representing the thermal fluctuation behavior[1,13,29,30]

$$\mathbf{S} = \langle \mathbf{v}^T\mathbf{v}\rangle = \sum_{p=7}^{3N} \frac{k_BT}{\omega_p^2}\mathbf{u}_p^T\mathbf{u}_p \tag{3}$$

where $<A>$ represents the ensemble average (time average) of the quantity A, k_B is the Boltzmann's constant, T is the absolute temperature, and index p indicates the mode index. Here, it should be noted that six zero-normal modes corresponding to rigid body motions are excluded for computing the correlation matrix $\mathbf{S}$. The mean-square fluctuation for residue i is given by $<|\mathbf{R}_i - \mathbf{R}_i{}^0|^2> = \mathbf{S}_{3(i-1)+1,\, 3(i-1)+1} + \mathbf{S}_{3(i-1)+2,\, 3(i-1)+2} + \mathbf{S}_{3(i-1)+3,\, 3(i-1)+3}$.

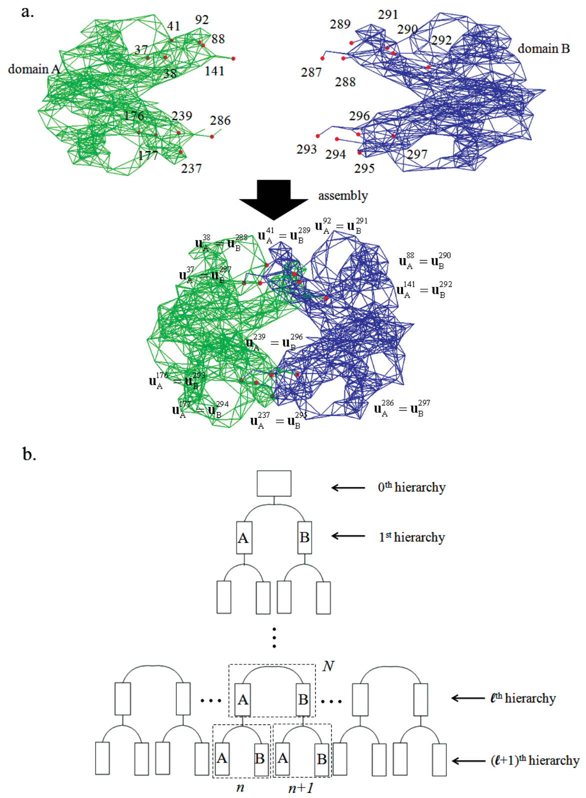
a.
domain A
37
41
92
88
38
141
176
239
286
177
237
289
291
290
292
domain B
287
288
296
293
294
295
297
assembly
$\mathbf{u}_A^{38} = \mathbf{u}_B^{288}$
$\mathbf{u}_A^{41} = \mathbf{u}_B^{289}$
$\mathbf{u}_A^{92} = \mathbf{u}_B^{291}$
$\mathbf{u}_A^{88} = \mathbf{u}_B^{290}$
$\mathbf{u}_A^{37} = \mathbf{u}_B^{287}$
$\mathbf{u}_A^{141} = \mathbf{u}_B^{292}$
$\mathbf{u}_A^{239} = \mathbf{u}_B^{296}$
$\mathbf{u}_A^{176} = \mathbf{u}_B^{293}$
$\mathbf{u}_A^{177} = \mathbf{u}_B^{294}$
$\mathbf{u}_A^{237} = \mathbf{u}_B^{295}$
$\mathbf{u}_A^{286} = \mathbf{u}_B^{297}$
b.
0th hierarchy
A
B
1st hierarchy
N
A
B
ℓth hierarchy
A
B
A
B
(ℓ+1)th hierarchy
n
n+1

Figure 1. (a) Schematic illustration of component mode synthesis (CMS) applied to hemoglobin. Here, hemoglobin is decomposed into 2 subdomains (subdomains *A* and *B*). The red dotted points represent the nodal points (residues) belonging to interface between 2 subdomains. Here, for nodal points at interface, the geometric constraints are imposed such that displacement vectors of residues at interface are continuous. (b) Schematic illustration of hierarchical component mode synthesis (hCMS) which decomposes the protein structure into a hierarchy composed of several subdomains. The eigenvalue problem for the *n*-th subdomain or (*n*+1)-th subdomain in the (*l*+1)-th hierarchy is solved for obtaining the normal modes of the *N*-th subdomain in the *l*-th hierarchy. This process is repeated until one runs into the 0-th hierarchy. In this work, we perform the hierarchical decomposition of protein into subdomains equivalent to protein domains.

2.2. Component Mode Synthesis (CMS). For normal-mode analysis of a protein structure, we have employed the component mode synthesis (CMS), which has been broadly utilized in engineering mechanics.[31−33] For a clear description of CMS, we consider the protein structure (e.g., hemoglobin shown in Figure 1(a)) which is decomposed into 2 subdomains. The motion of subdomain *A* is constrained by the subdomain *B*, and vice versa, in such a way that the nodal points (residues) at interface between 2 subdomains are constrained under the continuity of a displacement field. In other words, the interactions between 2 subdomains are prescribed by constraints at interface between 2 subdomains. The correlation between motions of 2 subdomains with using geometric constraints is a generic computational scheme in CMS used in structural dynamics[31−33] rather than using the domain−domain interaction directly. In general, component mode synthesis is implemented such that the number of nodal points (residues) of a subdomain is much larger than that of nodal points belonging to an interface (relevant to geometric constraint). This indicates that block (subdomain) should be selected in such a way that the degrees of freedom related to geometric constraints (i.e., nodal points related to block−block interaction described by constraint) should be much less than that of block.

Now, for convenience, let us describe the motion of protein structure decomposed into 2 subdomains without applying the constraints at this moment. The constraints will be implemented later in the assembly process. The potential energy without constraints is given by

$$V' = \frac{1}{2}(\mathbf{u}_A^T\mathbf{K}_A\mathbf{u}_A + \mathbf{u}_B^T\mathbf{K}_B\mathbf{u}_B) \tag{4}$$

Here, prime indicates that constraints were not implemented at this moment. $\mathbf{K}_i$ and $\mathbf{u}_i$ represent the stiffness matrix and displacement field for subdomain *i* (where $i = A$ or B), respectively. In the similar manner, the kinetic energy without constraints is in the form of

$$T' = \frac{1}{2}(\dot{\mathbf{u}}_A^T\mathbf{M}_A\dot{\mathbf{u}}_A + \dot{\mathbf{u}}_B^T\mathbf{M}_B\dot{\mathbf{u}}_B) \tag{5}$$

where $\mathbf{M}_i$ indicates the mass matrix for subdomains *i* (where $i = A$ or B), and a symbol dot represents the time-derivative. Then, we introduce the linear transformation such that the displacement vector $\mathbf{u}_i$ (where $i = A$ or B) is represented in the form of $\mathbf{u}_i(\mathbf{x}, t) = \Phi_i(\mathbf{x})\cdot\mathbf{v}_i(t)$, where $\Phi_i(\mathbf{x})$ is the matrix whose column vector is the eigenvector of the stiffness matrix $\mathbf{K}_i$. That is, $\Phi_i(\mathbf{x})$ satisfies the eigen-value problem such as $\mathbf{K}_i\Phi_i(\mathbf{x}) = \Phi_i(\mathbf{x})\Lambda_i$, where Λ_i is the diagonal matrix whose component is the eigen-value of $\mathbf{K}_i$. With transformation, the potential energy and the kinetic energy without constraints can be represented in the space spanned by normal modes of each subdomain.

$$V' = \frac{1}{2}[\mathbf{v}_A^T \quad \mathbf{v}_B^T]\begin{bmatrix}\Lambda_A & 0\\ 0 & \Lambda_B\end{bmatrix}\begin{bmatrix}\mathbf{v}_A\\ \mathbf{v}_B\end{bmatrix} \equiv \frac{1}{2}\mathbf{v}^T\Lambda\mathbf{v} \tag{6.a}$$

$$T' = \frac{1}{2}[\dot{\mathbf{v}}_A^T \quad \dot{\mathbf{v}}_B^T]\begin{bmatrix}\Phi_A^T\mathbf{M}_A\Phi_A & 0\\ 0 & \Phi_B^T\mathbf{M}_B\Phi_B\end{bmatrix}\begin{bmatrix}\dot{\mathbf{v}}_A\\ \dot{\mathbf{v}}_B\end{bmatrix} \equiv \frac{1}{2}\dot{\mathbf{v}}^T\mathbf{L}\dot{\mathbf{v}} \tag{6.b}$$

Here, the displacement vector represented in the normal mode space, $\mathbf{v}^T = [\mathbf{v}_A{}^T\ \mathbf{v}_B{}^T]$, has the degrees of freedom larger than the degrees of freedom of a protein, since the constraints are not imposed.

Now, in order to describe the motion of entire protein domains, we have to impose the geometric constraints as shown in Figure 1(a). For instance, nodal points 37 of the domain *A* colored green is the identical nodal point 287 of the domain *B* colored blue, so that the displacement field for such two nodal points should be continuous, i.e. $\mathbf{u}_A{}^{37} = \mathbf{u}_B{}^{287}$. In general, the geometric constraints for interface between two subdomains can be represented in the form of $\mathbf{Pv} = \mathbf{0}$. Since $\mathbf{v}$ has the redundancy because of redundant enumeration of nodal points at the interface belonging to subdomains *A* and *B*, a vector $\mathbf{v}$ can be decomposed into independent variable $\mathbf{w}(t)$ and dependent variable $\mathbf{z}(t)$. The constraint equation, i.e. $\mathbf{Pv} \equiv \mathbf{P}_1\mathbf{w}(t) + \mathbf{P}_2\mathbf{z}(t) = 0$, leads to the relation of

$$\mathbf{B} = \begin{bmatrix} \mathbf{I} \\ -\mathbf{P}_2^{-1}\mathbf{P}_1 \end{bmatrix} \tag{7}$$

where **B** is the constraint matrix. Then, with the application of constraints, the potential energy and the kinetic energy for a protein structure, respectively, become

$$V = \frac{1}{2}\mathbf{w}^T(\mathbf{B}^T\Lambda\mathbf{B})\mathbf{w} \equiv \frac{1}{2}\mathbf{w}^T\mathbf{D}\mathbf{w} \tag{8.a}$$

$$T = \frac{1}{2}\dot{\mathbf{w}}^T(\mathbf{B}^T\mathbf{L}\mathbf{B})\dot{\mathbf{w}} \equiv \frac{1}{2}\dot{\mathbf{w}}^T\mathbf{S}\mathbf{w} \tag{8.b}$$

Here, **D** and **S** are the stiffness matrix and mass matrix, respectively, for a protein structure, represented in the space spanned by the normal modes of subdomains. It should be noted that the potential energy V and the kinetic energy T given by eqs 8.a and 8.b, respectively, are the exact form for potential energy and kinetic energy for a protein structure composed of 2 subdomains, since the geometric constraints that describe the domain−domain interaction are imposed.

The normal-mode analysis of a protein structure can be, thus, represented by the following eigen-value problem such as $\mathbf{DU} = \mathbf{SU}\Omega$, where **U** is the modal matrix and Ω is the diagonal matrix whose component is the eigen-value of a protein structure. In order to describe the protein dynamics with respect to normal modes, the modal matrix **U** has to be transformed into matrix **Z**, whose column vectors represent the normal modes, such as $\mathbf{Z} = \Phi\mathbf{BU}$, where Φ is the matrix given by $\Phi^T = [\Phi_A{}^T\ \Phi_B{}^T]$.

As stated above, the component mode synthesis for a protein with two domains is straightforward and easy to be implemented. However, a large protein complex has several rigid domains, so that we need to consider component mode synthesis with several subdomains. For such a large protein, we introduce the hierarchical component mode synthesis (hCMS), which adopts the component mode synthesis in a hierarchical manner. As shown in Figure 1(b), we divide the protein structure into subdomains in a hierarchical manner. Let us denote the index l representing the index of hierarchy and the index n to indicate the index of subdomain in the l-th hierarchy. The process for hCMS is given as follows:

(i) Define the stiffness matrix $\mathbf{K}_n{}^{(l+1)}$ and the mass matrix $\mathbf{M}_n{}^{(l+1)}$ for the n-th subdomain in the (l+1)-th hierarchy.

(ii) Solve the eigen-value problem with using $\mathbf{K}_n{}^{(l+1)}$ in order to obtain the normal modes of subdomains to construct the matrices such as $\Phi_{A,n}{}^{(l+1)}$ and $\Phi_{B,n}{}^{(l+1)}$.

(iii) Convert the stiffness matrix $\mathbf{K}_n{}^{(l+1)}$ using normal modes $\Phi_{A,n}{}^{(l+1)}$ and $\Phi_{B,n}{}^{(l+1)}$, that is, find the matrix $\mathbf{D}_n{}^{(l+1)}$. Similarly, transform the mass matrix $\mathbf{M}_n{}^{(l+1)}$ to obtain the matrix $\mathbf{S}_n{}^{(l+1)}$.

(iv) From the eigen-value problem $\mathbf{D}_n{}^{(l+1)}\mathbf{U}_n{}^{(l+1)} = \mathbf{S}_n{}^{(l+1)}\mathbf{U}_n{}^{(l+1)}\mathbf{\Omega}_n{}^{(l+1)}$, the normal mode $\mathbf{Z}_n{}^l$ for the n-th subdomain in the l-th hierarchy can be obtained.

(v) Repeat the process (i)−(iv) until the normal modes $\mathbf{Z}_n{}^l$ for every subdomain in the l-th hierarchy are obtained.

(vi) Set the normal modes $\mathbf{Z}_n{}^l$ as eigen-modes $\Phi_{A,N}{}^l$ for the subdomain A belonging to the N-th subdomain in the l-th hierarchy. Similarly, the normal modes $\mathbf{Z}_{n+1}{}^l$ is set to the eigen-modes $\Phi_{B,N}{}^l$.

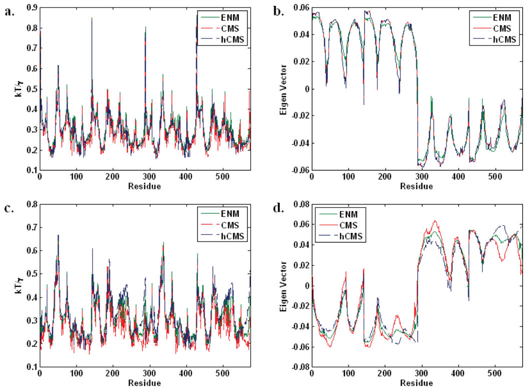

Figure 2. (a) Debye−Waller temperature factors obtained from ENM, CMS (consisting of two subdomains), and hCMS (composed of four subdomains) for hemoglobin in close form (pdb: 1a3n), (b) lowest-frequency normal mode (excluding the zero modes corresponding to rigid body motions) obtained from ENM, CMS, and hCMS for hemoglobin (pdb: 1a3n), (c) Debye−Waller temperature factor of hemoglobin in close form (pdb: 1bbb) described by ENM with cutoff radius of $R_c = 7$ Å, CMS based on such an ENM, and hCMS, and (d) lowest-frequency normal mode for hemoglobin (pdb: 1bbb) obtained from such an ENM, CMS, and hCMS.

(vii) Repeat the process (i)−(vi) until one runs into the 0-th hierarchy.

Here, it should be noted that we perform the hierarchical decomposition of a protein structure until each subdomain has the appropriate degrees of freedom (see Results and Discussion).

Results and Discussion

We considered the model proteins such as hemoglobin (in open and close forms), citrate synthase, and motor protein F_0-ATPase. These proteins have several subdomains, i.e. 2 subdomains for citrate synthase, 4 subdomains for hemoglobin, and 13 subdomains for F_0-ATPase, so that CMS (or hCMS) is applicable to understanding the dynamics of such proteins as well as low-frequency normal modes relevant to conformational change.

Conformational Dynamics of Hemoglobin. We consider the hemoglobin, which is a good model protein that is well described by NMA and ENM. Hemoglobin consists of 4 subdomains (chains) such as α_1, β_1, α_2, and β_2 chains. In our study, we take into account two types of CMS for a description of hemoglobin dynamics. Specifically, we first consider the CMS, which regards the protein structures as a combination of 2 subdomains. We also take into account hCMS to describe the hemoglobin structure as two subdomains, each of which is composed of α and β chains.

First, let us take into account the mean-square fluctuation (MSF) of hemoglobin, obtained by ENM, CMS (hemoglobin composed of 2 subdomains), and hCMS (hemoglobin consisting of 4 subdomains). At this moment, we employed the hemoglobin in close form (pdb: 1a3n), in which a ligand is bounded. Here, the force constant for ENM is given as $\gamma = 0.886$ kcal/mol·Å^2 with setting $R_c = 7$ Å by comparing the Debye−Waller factor (B-factor). Figure 2(a) shows the conformational fluctuation behavior predicted by ENM, CMS, and hCMS. This

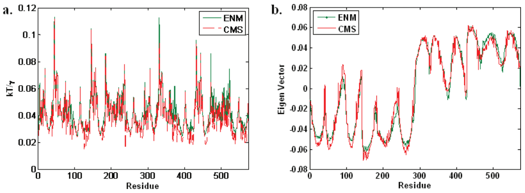

Figure 3. (a) Debye−Waller temperature factor for hemoglobin (pdb: 1bbb) described by ENM with cutoff radius of R_c = 12 Å, and CMS based on such an ENM, and (b) lowest-frequency normal modes computed from such an ENM and CMS.

indicates the robustness of CMS or hCMS for analyzing the conformational fluctuation behavior. More specifically, in order to validate the robustness of CMS or hCMS, we have also considered the lowest-frequency normal mode (with excluding the zero normal modes corresponding to rigid body motion), which is highly related to the conformational change of protein. As shown in Figure 2(b), normal modes obtained from ENM and CMS (or hCMS) are almost identical to each other with a correlation of $r > 0.99$. This indicates that low-frequency normal mode is well dictated by CMS (or hCMS). However, for a hemoglobin in open form (pdb: 1bbb) described by ENM with using $R_c = 7$ Å, the low-frequency normal mode (or B-factor) anticipated from CMS (or hCMS) is similar but not exactly identical to that of ENM (see Figure 2(c) and (d)). Even though the constraint condition is changed (i.e., different definition of atoms at interface), the quality of normal mode or B-factor computed from CMS (or hCMS) has not been improved (not shown). This may be attributed to a protein structure such that the structural feature of hemoglobin in an open form may not be well depicted by ENM with $R_c = 7$ Å. For validation of this conjecture, we describe the hemoglobin by using ENM with $R_c = 12$ Å and its related CMS (or hCMS). It is shown that the lowest-frequency normal modes computed from ENM and CMS (or hCMS) are almost identical to each other, based on protein topology dictated by $R_c = 12$ Å (see Figure 3(a) and (b)).

For further investigation of the quality of normal modes estimated from CMS (or hCMS), we have introduced the parameter, referred to as *overlap*,[34] defined as $\chi_{ij} = \mathbf{v}_i^{ENM}\cdot\mathbf{v}_j^{CMS}$, where $\mathbf{v}_i^{ENM}$ and $\mathbf{v}_j^{CMS}$ are the *i*-th and *j*-th normal modes obtained from ENM and CMS, respectively. The quantity χ_{ij} close to 1 indicates the high correlation (similarity) between the *i*-th normal mode obtained from ENM and the *j*-th normal mode computed from CMS, while such a quantity close to zero represents that two normal modes are rarely correlated. We have shown the density map of *overlap* for hemoglobin in both forms (see Figure 4). It is remarkable that low-frequency normal modes, which play a role in conformational change, obtained from CMS are highly correlated to those estimated from ENM. This suggests the robustness of CMS (or hCMS) in the prediction of low-frequency normal modes that are typically involved in conformational change. However, the correlation between high-frequency normal modes obtained from ENM and CMS is decreased. This provides that high-frequency normal modes, related to localized motion (e.g., fast motion of receptor due to ligand−receptor binding), may not be anticipated from CMS (or hCMS). Further, for quantifying the correlation between normal modes computed from ENM and CMS, we have adopted the quantity such as *spanning coefficient*,[34] defined as $\delta_i = \sum_{j=1}^{M}\chi_{ij}$. The spanning coefficient indicates that a normal mode of CMS can be spanned by *M* normal modes of ENM. Here, we set $M = 50$, which means that the spanning coefficient indicates how much a normal mode of CMS can be represented by 50 low-frequency normal modes of ENM. As shown in Figure 4(c) and (d), low-frequency normal modes (up to ∼40 normal modes) of CMS can be well dictated by the space spanned by 50 low-frequency normal modes of ENM. This implies that, for hemoglobin, low-frequency normal modes (up to ∼40 low-frequency modes) can be well predicted by CMS (or hCMS).

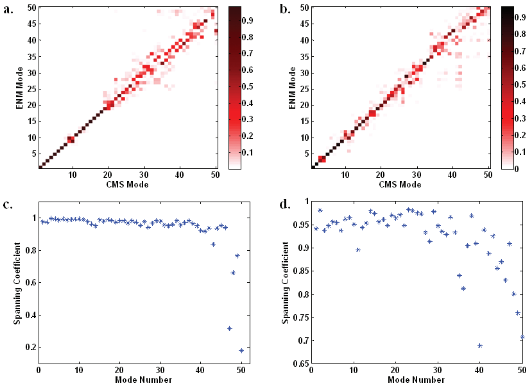

Figure 4. Overlap between normal modes obtained from ENM and CMS for hemoglobin in (a) close form (pdb: 1a3n) and (b) open form (pdb: 1bbb). Spanning coefficient between normal modes computed from ENM and CMS for hemoglobin in (c) close form and (d) open form. It is shown that low-frequency normal modes estimated from CMS are highly correlated with those from ENM.

Conformational Dynamics of Model Proteins. We have considered several model proteins such as citrate synthase in open, and close forms, and F0-ATPase motor protein. We compared the B-factors of model proteins obtained from CMS (or hCMS) with those estimated from ENM. Similar to the case of hemoglobin, the B-factors are well reproduced by CMS (or hCMS), indicating the robustness of CMS for predicting the fluctuation behavior. For further validation, we have taken into account the lowest-frequency normal modes of model proteins obtained from ENM as well as CMS (or hCMS). It is shown that low-frequency normal modes anticipated from CMS (hCMS) are almost identical to those evaluated from ENM. With similar analyses on *overlap* and *spanning coefficient*, the low-frequency normal modes of model proteins related to their biological function (e.g., conformational change) can be well dictated by CMS (hCMS). For details, see the Supporting Information.

Effect of Constraints on Protein Dynamics Described by CMS. Constraint equation is essential in CMS (or hCMS) in order to convert the stiffness matrix (and/or the mass matrix) into that spanned by normal modes of subdomains. It is assumed that the boundary nodal points were selected in such a way that two residues belonging to two different subdomains are boundary nodal points at the interface between two subdomains if the distance between two such residues is within a certain distance, referred to as search

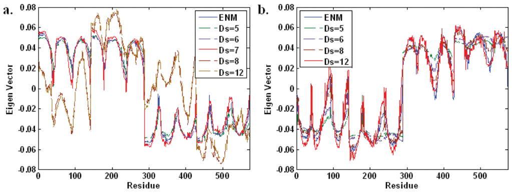

Figure 5. Lowest-frequency normal modes obtained from CMS with different types of constraint (i.e., different search distances D_S) for hemoglobin in (a) close form (pdb: 1a3n) and (b) open form (pdb: 1bbb). It is shown that the constraint should be selected such that $D_S \sim R_c$, indicating that overall stiffness should be maintained when constraint is determined.

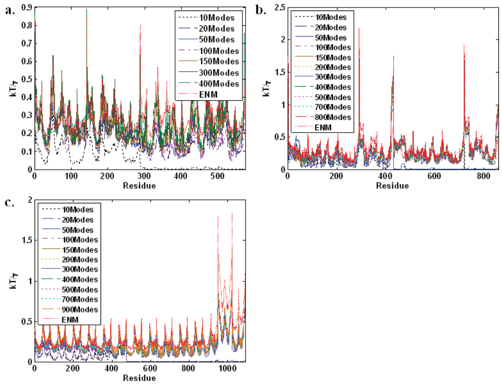

Figure 6. Debye−Waller temperature factors obtained from CMS in reduced space G* spanned by some normal modes (i.e., 10 ∼ O(N) normal modes, where N is the total degrees of freedom) for (a) hemoglobin (pdb: 1a3n), (b) citrate synthase (pdb: 5csc), and (c) F0-ATPase motor protein (pdb: 1c17). It is shown that, at least, more than 50 normal modes should be used for space G* in CMS.

distance D_S. If D_S is very small, then two subdomains will be less constrained so that the whole protein structure will be more flexible than it is. On the other hand, if D_S is very large, then two subdomains are so constrained that the protein structure is more rigid than it is. Figure 5 depicts the low-frequency normal modes of hemoglobin obtained from CMS with different constraints D_S. Here, the structure of hemoglobin in close form (pdb: 1a3n) is described by ENM with $R_c = 7$ Å, and D_S is varying from 5 Å to 12 Å. For $D_S = 5$ Å the number of constraints is 4, whereas for $D_S = 12$ Å the total number of constraints is 25. As shown in Figure 5(a), as long as D_S is in the range of 5 Å−7 Å, the lowest-frequency normal mode obtained from CMS is quantitatively comparable to that computed from ENM. In the case of hemoglobin in open form (pdb: 1bbb), we describe its structure as a Gaussian Network Model (GNM) with a cutoff distance of $R_c = 12$ Å. For such a case, the CMS with use of $D_S = 12$ Å provides the lowest-frequency normal mode quantitatively comparable to that estimated from GNM (Figure 5(b)), while the functional low-frequency normal mode cannot be dictated by the CMS with $D_S < 12$ Å. These two examples suggest that the search distance D_S for constraint should be chosen such that D_S is quantitatively comparable to the cutoff distance R_c used in ENM (or GNM). This may be attributed to the fact that, if $D_S \sim R_c$, the overall stiffness of a protein described by CMS is close to that dictated by the original structure. It is implied that the constraint should be selected as long as the constraint does not affect the overall stiffness of the protein structure responsible for conformational fluctuation dynamics.

Conformational Fluctuation Dictated by Normal Modes of CMS. The key of CMS is to transform the stiffness matrix **K** in the Cartesian coordinate into that represented in the space *G* spanned by normal modes of subdomains. This leads the matrix **S** (i.e., stiffness matrix represented in the space *G*) to be a diagonal matrix, and, consequently, it improves the computational efficiency to obtain the natural frequencies and their related normal modes for a protein. Here, we have further considered the space, in which **S** is represented, spanned by some normal modes (from lowest-frequency mode to a certain frequency normal mode) rather than all normal modes of subdomains. Here, such a space is denoted as *G**. This representation in *G** will enhance the computation of functional low-frequency normal modes of proteins as well as their conformational fluctuation. Figure 6 shows the Debye−Waller temperature factors of model proteins obtained from both ENM and CMS which employs the different number of normal modes spanning the space *G** for **S**. It is shown that, at least, more than 50 normal modes should be utilized in CMS in order to have the physically meaningful conformational fluctuation information. In order to quantify the correlation of normal modes between ENM and CMS with the use of a different number of normal modes, we have introduced the correlation parameter *r* defined as[35]

$$r = \frac{\sum_{i=1}^{N} (B_i^{ENM} - \langle B^{ENM}\rangle)(B_i^{CMS} - \langle B^{CMS}\rangle)}{\sqrt{\sum_{i=1}^{N} (B_i^{ENM} - \langle B^{ENM}\rangle)^2 \sum_{j=1}^{N} (B_j^{CMS} - \langle B^{CMS}\rangle)^2}} \tag{9}$$

Here, B_i^{ENM} and B_i^{CMS} indicate the Debye−Waller temperature factor for residue *i* obtained from ENM and CMS, respectively, *N* is the total number of residues, and angle brackets < > represent the average given by $\langle B\rangle = (1/N)\sum_{j=1}^{N} B_j$. A value of the correlation parameter *r* close to 1 indicates that the B-factor obtained from CMS is highly correlated to that from ENM, while a value of *r* approaching 0 indicates the uncorrelation between B-factors obtained from ENM and CMS, and the value of *r* close to −1 shows the anticorrelation between B-factors computed from ENM and CMS. As shown in Figure 7, it is shown that, at least, more than 50 normal modes spanning *G** should be employed in CMS in order to gain the B-factors with a correlation of >80% compared to that obtained from ENM. It indicates that, if one utilizes the 50 normal modes spanning the space *G** in CMS, one is able to estimate the Debye−Waller temperature factor comparable to ENM. In other words, 50 normal modes employed in CMS are sufficient to represent the protein dynamics with computational efficiency.

Degree of Hierarchical Decomposition. We have performed the hierarchical component mode synthesis (hCMS) to protein structure until each hierarchical subdomain is identical to the protein domain. For instance, the hCMS has

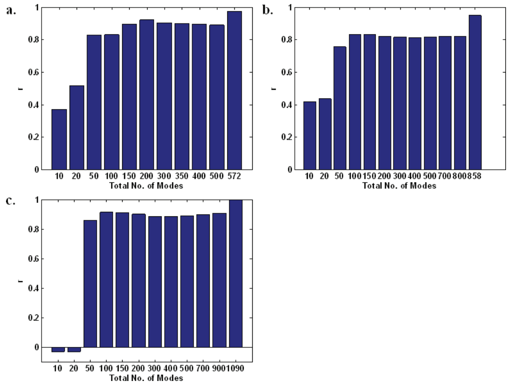

Figure 7. Correlation coefficient between Debye–Waller temperature factors computed from ENM and CMS with different reduced space G* for (a) hemoglobin (pdb: 1a3n), (b) citrate synthase (pdb: 5csc), and (c) F0-ATPase motor protein (pdb: 1c17). It is shown that 50 normal modes spanning space G* in CMS provides the Debye–Waller temperature factor quantitatively comparable to that computed from original structure (ENM) with a correlation of >80%.

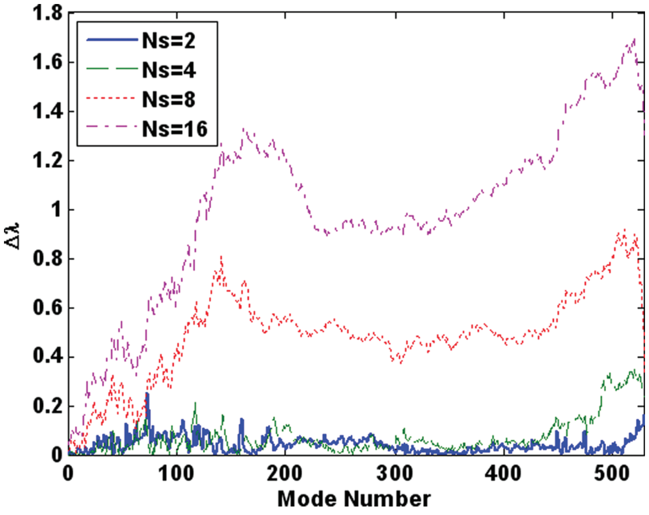

Figure 8. The difference between eigen-values, for hemoglobin, obtained from normal-mode analysis (NMA) and hierarchical component mode synthesis (hCMS) with different hierarchial decomposition. Here, $\Delta\lambda = |\ddot{\lambda}^{NMA} - \ddot{\lambda}^{hCMS}|$, where $\ddot{\lambda}^{NMA}$ and $\ddot{\lambda}^{hCMS}$ represent the eigen-value obtained from NMA and hCMS, respectively. When a hemoglobin is decomposed into 2 or 4 subdomains, the difference of eigen-values, $\Delta\lambda$, is insignificant. However, if a hemoglobin is decomposed into more subdomains, the difference of eigen-values, $\Delta\lambda$, becomes larger. This indicates that hCMS can be implemented until protein strucure is decomposed into protein domains.

been applied to hemoglobin such that the hemoglobin structure is decomposed to 4 subdomains. In order to investigate the available degree of hierarchy for hCMS, we have considered the hemoglobin (composed of 4 protein domains) with different hierarchies — (i) 2 subdomains, (ii) 4 subdomains, (iii) 8 subdomains, and (iv) 16 subdomains. Figure 8 shows the difference between eigen-values, for hemoglobin, obtained from NMA and hCMS with different hierarchies. It is shown that, as long as hemoglobin is decomposed into 2 or 4 subdomains, the difference between eigen-values obtained from NMA and hCMS is insignificant.

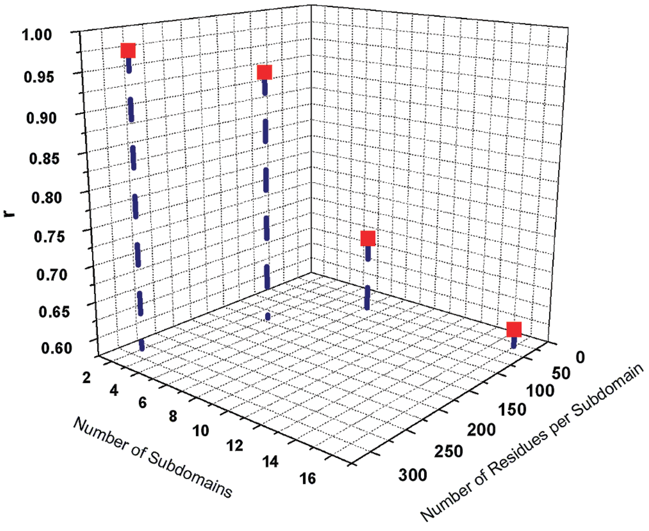

Figure 9. Correlation between thermal fluctations, for hemoglobin, obtained from normal-mode analysis (NMA) and hierarchical component mode synthesis (hCMS). It is shown that, as long as a hemoglobin is split into 2 or 4 subdomains, the fluctuation behavior (Debye–Waller B factor) obtained from hCMS is quantitatively comparable to that from NMA with correlation of >90%. However, if a hemoglobin is decomposed into many subdomains (e.g., 16 subdomains), the fluctuation behavior estimated from hCMS deviates from that predicted from NMA with correlation of <70%. This indicates that our hCMS has to be implemented until a protein structure is decomposed at protein domain level rather than residue level.

On the other hand, as hemoglobin is decomposed into smaller subdomains (i.e., smaller than protein domain), the eigen-values obtained from hCMS are deviated from those obtained from NMA. This indicates that hCMS with decomposition of the protein structure into many subdomains would not provide a good prediction of conformational fluctuation. Specifically, Figure 9 shows the correlation between B-factors, for hemoglobin, obtained from normal-mode analysis (NMA) and hCMS with given hierarchies. It is shown that the hCMS with 2 or 4 subdomains predicts the thermal fluctuation behavior with correlation of >90% to NMA. However, once hemoglobin is decomposed into more than 4 subdomains, then the correlation between B-factors obtained from NMA and hCMS is <80%. As the hemoglobin is decomposed into more subdomains (much larger than number of protein domains), the worse correlation between B-factors obtained from NMA and hCMS is obtained. For instance, if hemoglobin is divided into 16 subdomains (composed of ~25 residues), then the correlation between B-factors obtained from NMA and hCMS is r = ~60%. This indicates that our hCMS has to be implemented such that a protein structure can be decomposed into the subdomains at the protein domain level. This is attributed to the fact that, if protein is decomposed into many subdomains composed of a small number of residues, then the degrees of freedom related to geometric constraint is equivalent to degrees of freedom of subdomains, which leads to the overconstraint of the subdomain. This implies that our

Table 1. Computational Time for Estimating Normal Modes

	model protein		
	F0-ATPase (pdb:1c17)	citrate synthase (pdb:5csc)	citrate synthase (pdb: 6csc)
GNM	63.37 s	27.6 s	25.82 s
CMS ($N_s = 2$, $D_s = 5$ Å)	22.19 s	9.77 s	NA
CMS ($N_s = 2$, $D_s = 6$ Å)	22.76 s	9.93 s	10.33 s
CMS ($N_s = 2$, $D_s = 7$ Å)	23.73 s	10.09 s	10.61 s

hCMS can be implemented at the protein domain level rather than the residue level.

Computational Cost for hCMS. In order to compare the computational speed of hCMS with that of general NMA, we have measured the computation time to estimate the thermal fluctuation of model proteins, composed of >1000 residues, based on hCMS and NMA. As shown in Table 1, hCMS with different constraints exhibits the faster computation on low-frequency normal modes and conformational fluctuation than general NMA by a factor of ~2. This indicates that our hCMS enhances the computational time to estimate the conformational fluctuation by a factor of ~2 and that our hCMS can be applicable to iterative NMA of a large protein complex for understanding their conformational transition. The fast computation of fluctuation dynamics using CMS is attributed to fact that the representation of the stiffness matrix in the normal modes of subdomains enhances the computation in solving the eigen-value problem.[31−33]

Conclusion

We demonstrate the application of component mode synthesis (CMS) or hierarchical CMS (hCMS) for the conformational dynamics of a large protein complex. We have shown that hCMS enables the computationally efficient estimation of functional low-frequency normal modes, the Debye−Waller temperature factor, and correlated motion. The key of hCMS (or CMS) is to represent the stiffness matrix in the space G spanned by normal modes of subdomains. Moreover, it is shown that the reduced space G^* allows us to depict the large protein dynamics with enhanced computational efficacy. This computationally efficient hCMS may improve the computational estimation of conformational transition between two equilibrium states, which is usually computed from iterative normal-mode analysis with certain constraints.[16,17] In the long run, such hCMS will enable the understanding of the functional motion of large protein complexes as well as their energy landscape for conformational transition described by iterative normal-mode analysis.

Acknowledgment. This work was supported by KOSEF (Grant No. R11-2007-028-03002) and KRF (Grant No. KRF-2008-314-000012).

Supporting Information Available: Results for conformational dynamics of model proteins with hierarchical component mode synthesis. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Cui, Q.; Bahar, I. *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems*; CRC Press: 2005.

(2) Bahar, I.; Rader, A. J. *Curr. Opin. Struct. Biol.* **2005**, *15*, 586–592.

(3) Tama, F.; Brooks, C. L. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 115–133.

(4) Brooks, B.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 6571–6575.

(5) Janezic, D.; Venable, R. M.; Brooks, B. R. *J. Comput. Chem.* **1995**, *16*, 1554–1566.

(6) Hayward, S.; Go, N. *Annu. Rev. Phys. Chem.* **1995**, *46*, 223–250.

(7) Cieplak, M.; Hoang, T. X.; Robbins, M. O. *Proteins: Struct., Funct., Genet.* **2002**, *49*, 114–124.

(8) Cieplak, M.; Hoang, T. X.; Robbins, M. O. *Proteins: Struct., Funct., Genet.* **2002**, *49*, 104–113.

(9) Sulkowska, J. I.; Cieplak, M. *Biophys. J.* **2008**, *95*, 3174–3191.

(10) Yoon, G.; Park, H.-J.; Na, S.; Eom, K. *J. Comput. Chem.* **2009**, *30*, 873–880.

(11) Tirion, M. M. *Phys. Rev. Lett.* **1996**, *77*, 1905–1908.

(12) Bahar, I.; Atilgan, A. R.; Demirel, M. C.; Erman, B. *Phys. Rev. Lett.* **1998**, *80*, 2733–2736.

(13) Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. *Biophys. J.* **2001**, *80*, 505–515.

(14) Xu, C. Y.; Tobi, D.; Bahar, I. *J. Mol. Biol.* **2003**, *333*, 153–168.

(15) Tobi, D.; Bahar, I. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 18908–18913.

(16) Miyashita, O.; Onuchic, J. N.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12570–12575.

(17) Zheng, W. J.; Brooks, B. R. *Biophys. J.* **2005**, *88*, 3109–3117.

(18) Whitford, P. C.; Miyashita, O.; Levy, Y.; Onuchic, J. N. *J. Mol. Biol.* **2007**, *366*, 1661–1671.

(19) Maragakis, P.; Karplus, M. *J. Mol. Biol.* **2005**, *352*, 807–822.

(20) Zheng, W.; Brooks, B. R.; Hummer, G. *Proteins: Struct., Funct., Bioinf.* **2007**, *69*, 43–57.

(21) Doruker, P.; Jernigan, R. L.; Bahar, I. *J. Comput. Chem.* **2002**, *23*, 119–127.

(22) Eom, K.; Ahn, J. H.; Baek, S. C.; Kim, J. I.; Na, S. *CMC: Comput. Mater. Continua* **2007**, *6*, 35–42.

(23) Eom, K.; Baek, S.-C.; Ahn, J.-H.; Na, S. *J. Comput. Chem.* **2007**, *28*, 1400–1410.

(24) Cheng, H.; Gimbutas, Z.; Martinsson, P. G.; Rokhlin, V.; SIAM, J. *Sci. Comput.* **2005**, *26*, 1389–1404.

(25) Kurkcuoglu, O.; Jernigan, R. L.; Doruker, P. *Polymer* **2004**, *45*, 649–657.

(26) Kim, M. K.; Jernigan, R. L.; Chirikjian, G. S. *Biophys. J.* **2005**, *89*, 43–55.

(27) Tama, F.; Gadea, F. X.; Marques, O.; Sanejouand, Y. H. *Proteins: Struct., Funct., Genet.* **2000**, *41*, 1–7.

(28) Lu, M.; Ma, J. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 15358–15363.

(29) Weiner, J. H. *Statistical mechanics of elasticity*; Dover Publication: 1983.

(30) Chandler, D. *Introduction to modern statistical mechanics*; Oxford University Press: 1987.

(31) Meirovitch, L. *Computational methods in structural dynamics*; SIJTHOFF & NOORDHOFF: Rockville, Maryland, USA, 1980.

(32) Bhat, R. B. *J. Sound Vib.* **1985**, *101*, 271–272.

(33) Sung, S. H.; Nefske, D. J. *AIAA J* **1986**, *24*, 1021–1026.

(34) Van Wynsberghe, A. W.; Cui, Q. *Biophys. J.* **2005**, *89*, 2939–2949.

(35) Kondrashov, D. A.; Cui, Q.; Phillips, G. N., Jr. *Biophys. J.* **2006**, *91*, 2760–2767.

CT900027H