JCTC Journal of Chemical Theory and Computation

A New Method for Coarse-Grained Elastic Normal-Mode Analysis

Mingyang Lu,[†] Billy Poon,[‡] and Jianpeng Ma*[,†,‡]

Department of Biochemistry and Molecular Biology, Baylor College of Medicine, One Baylor Plaza, BCM-125, Houston, Texas 77030, and Department of Bioengineering, Rice University, Houston, Texas 77005

Received December 4, 2005

Abstract: In this paper, we report a new method for coarse-grained elastic normal-mode analysis. The purpose is to overcome a long-standing problem in the conventional analysis called the tip effect that makes the motional patterns (eigenvectors) of some low-frequency modes irrational. The new method retains the merits of a conventional method such as not requiring lengthy initial energy minimization, which always distorts structures, and also delivers substantially more accurate low-frequency modes with no tip effect for proteins of any size. This improvement of modes is crucial for certain types of applications such as structural refinement or normal-mode-based sampling.

I. Introduction

In biomolecular simulations, normal-mode analysis (NMA) has been playing a very important role in analyzing structural fluctuations around a well-defined molecular conformation. In recent years, elastic normal-mode analysis (eNMA) has been popularized for its simplicity and wide-range applicability.[1−20] The potential function[1] used in the conventional eNMA[5] is a highly coarse-grained elastic network that has a form of

$$V = (\gamma/2)\sum_i\sum_j \sigma_{ij}(|\mathbf{r}_{ij}| - |\mathbf{r}_{ij}^{\,0}|)^2, \quad \sigma_{ij} = \begin{cases} 1 & |\mathbf{r}_{ij}^{\,0}| \le r_c \\ 0 & |\mathbf{r}_{ij}^{\,0}| > r_c \end{cases} \quad (1)$$

where $|\mathbf{r}_{ij}|$ and $|\mathbf{r}_{ij}^{\,0}|$ are the instantaneous and equilibrium values (or initial values from the coordinates) of pairwise distance between the ith and the jth Cα atoms, respectively. The value of Heavidide step function σ_{ij} specifies the effect of cutoff. The absolute value of force constant γ is irrelevant to the calculation and is often set to 1.0. This elastic potential can be applied to either representative atoms from atomic structures such as Cα traces[4,5] or points placed in continuous density maps from low-resolution experiments such as cryogenic electron microscopy (cryo-EM).[6,8] Numerous studies have shown that eNMA is effective in extracting patterns of low-frequency normal modes (for reviews, see refs 21−23). The success of it hinges on a very important fact that, for compact biomolecular structures such as those of globular proteins, the patterns of low-frequency modes, often referred to as deformational modes, are only sensitive to the shape of the molecules, rather than to the detailed atomic structures, which has been quantitatively demonstrated in our recent study.[24]

Another important advantage of coarse-grained eNMA is that it does not require the initial lengthy energy minimization because the minimum of potential function in eq 1 is the current structure. The initial minimization almost always significantly distorts the structure as in NMA based on molecular mechanics force fields such as CHARMM.[25,26]

Despite its enormous success and widespread popularity in application, the current version of eNMA[5] has an inherent weakness, which we refer to as the "tip effect". In systems with structural components, the "tips", protruding out of the main body, e.g., an isolated surface loop or simply a thinner region in the density map; the tip effect can lead to pathological behavior in motions of points near those regions, presumably due to an imbalance of elastic forces among neighboring harmonic oscillators due to lighter packing

* Corresponding author e-mail: jpma@bcm.tmc.edu. Corresponding author mailing address: One Baylor Plaza, BCM-125, Baylor College of Medicine, Houston, TX 77030.

[†] Baylor College of Medicine.

[‡] Rice University.

10.1021/ct050307u CCC: $33.50
Published on Web 03/29/2006

around tip regions. In the modes tinted with the tip effect, the magnitudes of displacement of the points in eigenvectors at or around the tip regions are much larger than those of the rest of the system but with more or less normal-looking eigenvalues. Since the eigenvectors of normal modes are normalized, the abnormally large magnitudes of displacement in the tip regions make the rest of the system have much less movement and sometimes even static. To make matters worse, so far there has been no systematic method to predict which modes have tip effect, and they can occur in even very low-frequency modes. Although on a case-by-case basis, one might be able to bypass such an effect, to our best knowledge, a systematic way is unavailable for overcoming the tip effect. In practice, if one only wants very few lowest-frequency modes for functional interpretation, the tip effect may not be so severe since the modes picked for functional study may not have tip effects. But in cases when a set of continuous low-frequency modes are needed, such as in structural refinement, this could become a major issue because some of the modes in the chosen set will definitely have some degrees of tip effect.

In this paper, we report a modified eNMA for eradicating the tip effect. Our main strategy is to make the overly soft degrees of freedom around the tip regions stiffer. In our recent study,[24] we demonstrated that a Hessian matrix with randomized off-diagonal elements still gives correct low-frequency eigenvectors. The Hamiltonian used in the conventional eNMA and molecular mechanics such as CHARMM[25,26] can be viewed just as two special cases. Therefore, we hypothesize that somewhere between the two extreme cases, one must be able to find a new Hamiltonian in which the degrees of freedom around the tip regions are stiff enough to eradicate the tip effect and yet still retain the merit of not requiring the initial minimization. Such a Hamiltonian should still deliver correct eigenvectors for low-frequency modes.

The results in this paper show that it is indeed possible to find a Hamiltonian so that the normal modes calculated from it not only retain the merits of original eNMA but also do not contain the tip effect. This was true for single polypeptide proteins in their compact native states or extended denatured states. It was also true for multisubunit supramolecular complexes. The overall patterns of eigenvectors of modes for the new Hamiltonian, in comparison with those of conventional eNMA, were also found to be closer to the ones computed by standard molecular mechanics force fields such as CHARMM. We therefore hope that the new eNMA will be a complementary tool to the existing ones in analyzing protein dynamics, especially in certain applications such as structural refinement in which a continuous set of low-frequency modes would be needed.

II. Theory

Internal Coordinate System. The new eNMA is designed to work with Cα traces or a subset of Cα traces. To keep the length of the Cα−Cα pseudobond fixed, the internal coordinate (IC) system[27] was implemented (Figure 1). Although the formalism seems to be more complex than the original eNMA,[1,5] we found it is essential to work in an IC system because it effectively avoids the stretching of the bond length, which is one of the major sources of the tip effect in elastic network potential.

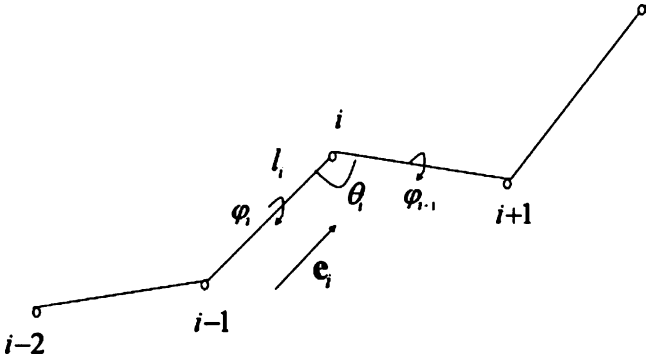

Figure 1. Internal coordinate system used in the new eNMA for eliminating tip effect.

For simplicity, we first describe the case with a single chain. The multichain case can be easily generalized. For a system with N Cα atoms ($i = 1,2,...,N$), the degrees of freedom in IC are $N - 2$ bond angles $\{\theta_i\}$ and $N - 3$ pseudo dihedral angles $\{\varphi_i\}$. Note, the bonds are indexed the same as the second Cα atom, the angles $\{\theta_i\}$ as the middle Cα atom, and the dihedrals as the third Cα atom. The relationship between IC and Cartesian coordinates (CC) is given in the following ways. From IC to CC

$$\mathbf{r}_1 = (0,0,0)^T; \quad \mathbf{e}_2 = (1,0,0)^T; \quad \mathbf{e}_3 = (\cos\theta_2, \sin\theta_2, 0)^T$$

$$\mathbf{e}_{i+1} = \cos\theta_i\,\mathbf{e}_i + \sin\theta_i\cos\varphi_i\frac{(\mathbf{e}_{i-1}\times\mathbf{e}_i)\times\mathbf{e}_i}{||(\mathbf{e}_{i-1}\times\mathbf{e}_i)\times\mathbf{e}_i||} + \sin\theta_i\sin\varphi_i\frac{\mathbf{e}_{i-1}\times\mathbf{e}_i}{||\mathbf{e}_{i-1}\times\mathbf{e}_i||}$$

$$\mathbf{r}_{i+1} = \mathbf{r}_i + l_{i+1}\mathbf{e}_{i+1} \tag{2}$$

where l_i is the ith Cα−Cα pseudo bond length, and $\mathbf{e}_i$ is the directional vector for that bond. The first Cα atom is at the origin, the first bond on the x-axis, and the first bond angle in the plane of XY. From CC to IC, we have

$$\begin{cases} l_i = ||\mathbf{r}_i - \mathbf{r}_{i-1}||; \quad \mathbf{e}_i = \dfrac{\mathbf{r}_i - \mathbf{r}_{i-1}}{l_i}; \quad \theta_i = \cos^{-1}(\mathbf{e}_i\cdot\mathbf{e}_{i+1}) \\ \varphi_i = \begin{cases} \cos^{-1}\left(\left(\dfrac{\mathbf{e}_{i-1}\times\mathbf{e}_i}{||\mathbf{e}_{i-1}\times\mathbf{e}_i||}\right)\cdot\left(\dfrac{\mathbf{e}_i\times\mathbf{e}_{i+1}}{||\mathbf{e}_i\times\mathbf{e}_{i+1}||}\right)\right), \\ \qquad\qquad if \quad \mathbf{e}_{i+1}\cdot(\mathbf{e}_{i-1}\times\mathbf{e}_i) \geq 0 \\ -\cos^{-1}\left(\left(\dfrac{\mathbf{e}_{i-1}\times\mathbf{e}_i}{||\mathbf{e}_{i-1}\times\mathbf{e}_i||}\right)\cdot\left(\dfrac{\mathbf{e}_i\times\mathbf{e}_{i+1}}{||\mathbf{e}_i\times\mathbf{e}_{i+1}||}\right)\right), \\ \qquad\qquad if \quad \mathbf{e}_{i+1}\cdot(\mathbf{e}_{i-1}\times\mathbf{e}_i) < 0 \end{cases} \end{cases} \tag{3}$$

In IC, θ_i is in the range of $[0,\pi]$, and its ordering in i runs as $i = 2,3...,N-1$. Also, φ_i is in the range of $[-\pi,\pi]$, and its ordering runs as $3,4,...,N-1$. For the convenience of further analysis, we lump the ICs into a set $\{\phi_\alpha: \theta_2,\varphi_3,\theta_3,\varphi_4,\theta_4...\varphi_{N-1},\theta_{N-1}\}$, where $\alpha = 1,2,...,2N-5$. The relation of index between CC and IC is

$$i = [\alpha/2] + 2 = \kappa(\alpha)$$
$$\alpha^{\varphi}(i) = 2i - 4(i \geq 3)$$
$$\alpha^{\theta}(i) = 2i - 3(i \geq 2) \tag{4}$$

Modified Elastic Potential Function. We defined a new potential function which has an extra term in addition to the potential function in the conventional eNMA.[1,5] The potential is

$$V = (\gamma/2)\sum_i\sum_j \sigma_{ij}(|\mathbf{r}_{ij}| - |\mathbf{r}_{ij}{}^0|)^2 + (\omega/2)\sum_\alpha(\phi_\alpha - \phi_\alpha{}^0)^2$$

$$\sigma_{ij} = \begin{cases} 1 & |\mathbf{r}_{ij}{}^0| \le r_c \\ 0 & |\mathbf{r}_{ij}{}^0| > r_c \end{cases}; \quad \omega = 3\min(H_{\alpha\alpha}{}^0) \tag{5}$$

where $H_{\alpha\alpha}{}^0$ is the diagonal element of the Hessian matrix of the conventional eNMA potential in IC. Note the summation in the second term goes over all the elements in the set of $\{\phi_\alpha\}$ (all of θ and φ). A scaling factor of 3 is chosen so that only the smallest diagonal term is dramatically changed.

This form of potential function also has its minimum at the current structure, thus no initial minimization is needed. The main purpose of the modified potential is to make the flexible degrees of freedom much stiffer so that the tip effect is eradicated.

Generalized Eigenvalue Problem in IC. The elements of the Hessian matrix in IC are related to those in CC as

$$\begin{cases} H_{\alpha\beta} = \dfrac{\partial^2 V}{\partial\phi_\alpha\partial\phi_\beta} = \sum \dfrac{\partial\mathbf{r}_i}{\partial\phi_\alpha}\cdot\mathbf{h}^{ij}\cdot\dfrac{\partial\mathbf{r}_j}{\partial\phi_\beta} \\ \dfrac{\partial\mathbf{r}_i}{\partial\phi_\alpha} = \dfrac{\partial\mathbf{r}_i}{\partial\theta_{\kappa(\alpha)}} = \begin{cases} (\mathbf{e}_{\kappa(\alpha)} \times \mathbf{e}_{\kappa(\alpha+1)}) \times \mathbf{r}_{\alpha i}, & i > \kappa(\alpha) \\ 0, & i \le \kappa(\alpha) \end{cases} & \phi_\alpha \in \theta, \\ \dfrac{\partial\mathbf{r}_i}{\partial\phi_\alpha} = \dfrac{\partial\mathbf{r}_i}{\partial\theta_{\kappa(\alpha)}} = \begin{cases} \mathbf{e}_{\kappa(\alpha)} \times \mathbf{r}_{\alpha i}, & i > \kappa(\alpha) \\ 0, & i \le \kappa(\alpha) \end{cases} & \phi_\alpha \in \varphi \end{cases} \tag{6}$$

where $\mathbf{r}_{\alpha i} = \mathbf{r}_i - \mathbf{r}_{\kappa(\alpha)}$, and $\mathbf{h}^{ij}$ is the submatrix of Hessian for the $i - j$ pair in CC. The calculation of the Hessian matrix in IC is accelerated from the order of n^4 to n^2. It uses a recursion relationship of matrix manipulation.[28]

$$H_{\alpha\beta} = \chi_\alpha^T \mathbf{R}^{i,j}\chi_\beta, \quad i = \kappa(\alpha), \quad j = \kappa(\beta)$$

$$\chi_\alpha^{(\theta)} = \begin{pmatrix} \mathbf{e}_{\kappa(\alpha)} \times \mathbf{e}_{\kappa(\alpha)+1} \\ (\mathbf{e}_{\kappa(\alpha)} \times \mathbf{e}_{\kappa(\alpha)+1}) \times \mathbf{r}_{\kappa(\alpha)} \end{pmatrix}$$

$$\chi_\alpha^{(\varphi)} = \begin{pmatrix} \mathbf{e}_{\kappa(\alpha)} \\ \mathbf{e}_{\kappa(\alpha)} \times \mathbf{r}_{\kappa(\alpha)} \end{pmatrix}$$

$$\mathbf{R}^{i,j} = \sum_{\substack{i'<i \\ j'\ge j}} \frac{\gamma}{r_{ij}^2}\begin{pmatrix} \mathbf{r}_i \times \mathbf{r}_j \\ \mathbf{r}_{ij} \end{pmatrix}(\mathbf{r}_i \times \mathbf{r}_j \quad \mathbf{r}_{ij}) = \sum_{\substack{i'<i \\ j'\ge j}} S_{i',j'} \tag{7}$$

and

$$\mathbf{R}^{i,j} = \mathbf{R}^{i,j+1} + \mathbf{R}^{i-1,j} - \mathbf{R}^{i-1,j+1} + S_{i-1,j}$$

$$\mathbf{R}^{1,j} = \mathbf{R}^{i,N+1} = 0 \tag{8}$$

recursion order: $(i,j) = (1,N),(1,N-1),...(1,1),$
$(2,N),...(2,2),(3,N),...,(N,N)$

where $\mathbf{R}^{i,j}$ is a 6×6 matrix, and $\chi_\alpha{}^{(\theta)}$, $\chi_\alpha{}^{(\varphi)}$ are 6×1 vectors. When analyzing dynamics, we should eliminate external motion by applying the Eckart Condition[29]

$$\sum_i m_i\mathbf{r}_i = 0; \quad \sum_i m_i\mathbf{r}_i \times \dot{\mathbf{r}}_i = 0 \tag{9}$$

First derivatives can also be calculated accordingly

$$\frac{\partial\mathbf{r}_i}{\partial\phi_\alpha} = \begin{cases} (\mathbf{P}_i\,\mathbf{1})\mathbf{W}_{\kappa(\alpha)}\chi_\alpha & i < \kappa(\alpha) \\ -(\mathbf{P}_i\,\mathbf{1})(\mathbf{1} - \mathbf{W}_{\kappa(\alpha)})\chi_\alpha & i \ge \kappa(\alpha) \end{cases}$$

$$\mathbf{W}_i = \begin{pmatrix} \mathbf{I}^{-1}\mathbf{I}_C & \mathbf{I}^{-1}\mathbf{P}_C^T \\ \mathbf{P}_C/M & M_C/M\mathbf{1} \end{pmatrix} \tag{10}$$

where

$$M_A = \sum_{i'\le i} m_{i'}; \ \mathbf{P}_A = \sum_{i'\le i} m_{i'}\mathbf{P}_{i'}; \ \mathbf{I}_A = \sum_{i'\le i} m_{i'}\mathbf{P}_{i'}^T\mathbf{P}_{i'}$$

$$\mathbf{I} = \sum_{i'} m_{i'}\mathbf{P}_{i'}^T\mathbf{P}_{i'}; \ M = \sum_{i'} m_{i'}$$

$$\mathbf{I}_C = \mathbf{I} - \mathbf{I}_A; \ \mathbf{P}_C = -\mathbf{P}_A; \ M_C = M - M_A$$

$$\mathbf{P}_i = \mathbf{r}_i \times \quad = \begin{pmatrix} 0 & -z_i & y_i \\ z_i & 0 & -x_i \\ -y_i & x_i & 0 \end{pmatrix} \tag{11}$$

P is an operator in matrix form. Then, we have the kinetic matrix

$$T = \frac{1}{2}\sum_{\alpha,\beta}\mathbf{T}_{\alpha\beta}\dot{\phi}_\alpha\dot{\phi}_\beta$$

$$\mathbf{T}_{\alpha\beta} = \sum_i m_i \frac{\partial\mathbf{r}_i}{\partial\phi_\alpha}\cdot\frac{\partial\mathbf{r}_i}{\partial\phi_\beta}$$

$$= \chi_\alpha^T\mathbf{K}_{\alpha\beta}\chi_\beta$$

$$\mathbf{K}_{\alpha\beta} = \begin{pmatrix} \mathbf{P}_A^T \\ M_A\mathbf{1} \end{pmatrix}(\mathbf{P}_C\ M_C\mathbf{1})/M + \begin{pmatrix} \mathbf{I}_A \\ \mathbf{P}_A \end{pmatrix}\mathbf{I}^{-1}(\mathbf{I}_C\ \mathbf{P}_C^T) \tag{12}$$

We end up with solving a generalized eigenvalue problem

$$\mathbf{HV} = \Lambda\mathbf{TV} \tag{13}$$

After obtaining eigenvectors, we can get orthonormal vectors in CC.

$$\Delta\mathbf{r}_i^{(k)} = \sum_\alpha \frac{\partial\mathbf{r}_i}{\partial\phi_\alpha}\Delta\phi_\alpha^{(k)}$$

$$\sum_i \Delta\mathbf{r}_i^{(k)}\cdot\Delta\mathbf{r}_i^{(k')} = \delta_{k,k'} \tag{14}$$

Finally, we define a quantitative localization factor T for the "tip effect"

$$T = \sum_i \left(\frac{||\Delta\mathbf{r}_{i+1} - \Delta\mathbf{r}_i||}{||\mathbf{r}_{i+1} - \mathbf{r}_i||}\right)^3 \tag{15}$$

The larger the T, the more eminent the "tip effect".

Multichain Analysis. To generalize the above method to multichain molecules, we shall virtually connect the last Cα

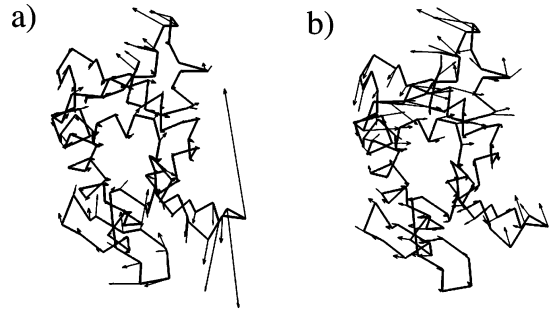

Figure 2. Motional patterns for the fourth mode of lysozyme (PDB code: 3lzt). (a) From the conventional eNMA, the lower-right portion has abnormal motions. (b) From the new eNMA, the motions for lower-right portion is much more realistic.

atom of the proceeding chain to the first Cα atom of the following chain. Hence, we will have 6 additional degrees of freedom for each additional chain. Among them, 5 are internal degrees of freedom and the sixth is for virtual bond length. It is noted that this virtual bond is the only flexible bond. We can make a new convention of order as

$$\{\phi_\alpha:\theta_2,\varphi_3,\theta_3,\varphi_4,\theta_4,...,\varphi_{N_1},\theta_{N_1},l_{N_1+1},\varphi_{N_1+1},\theta_{N_1+1},\varphi_{N_1+2},...\} \quad (16)$$

where N_1 is the number of Cα atoms in the first chain. For the virtual bond that connects the two chains, because

$$\frac{\partial \mathbf{r}_i}{\partial l_{\kappa(\alpha)}} = \mathbf{e}_{\kappa(\alpha)} \; (i > \kappa(\alpha)) \quad (17)$$

we only need to add an additional χ term in calculating matrices:

$$\chi_\alpha^{(l)} = \begin{pmatrix} 0 \\ -\mathbf{e}_{\kappa(\alpha)} \end{pmatrix} \quad (18)$$

Of course, with the additional degrees of freedom, the index order is also changed accordingly.

III. Results

Proteins with Single Polypeptide Chain. Here we first show an example of the tip effect. Lysozyme is a classic protein for NMA as its lowest-frequency mode represents a catalytically important hinge-bending motion.[30,31] Figure 2a shows the tip effect in one of the lowest-frequency modes, the fourth vibrational mode, calculated by the conventional eNMA.[5] The amplitude of motion of a flexible loop at the lower-right corner is abnormally large comparing with the rest of the system. In the new eNMA, the motion of the same region is much more realistic for this low-frequency mode (Figure 2b). Tests on various other proteins showed very similar results.

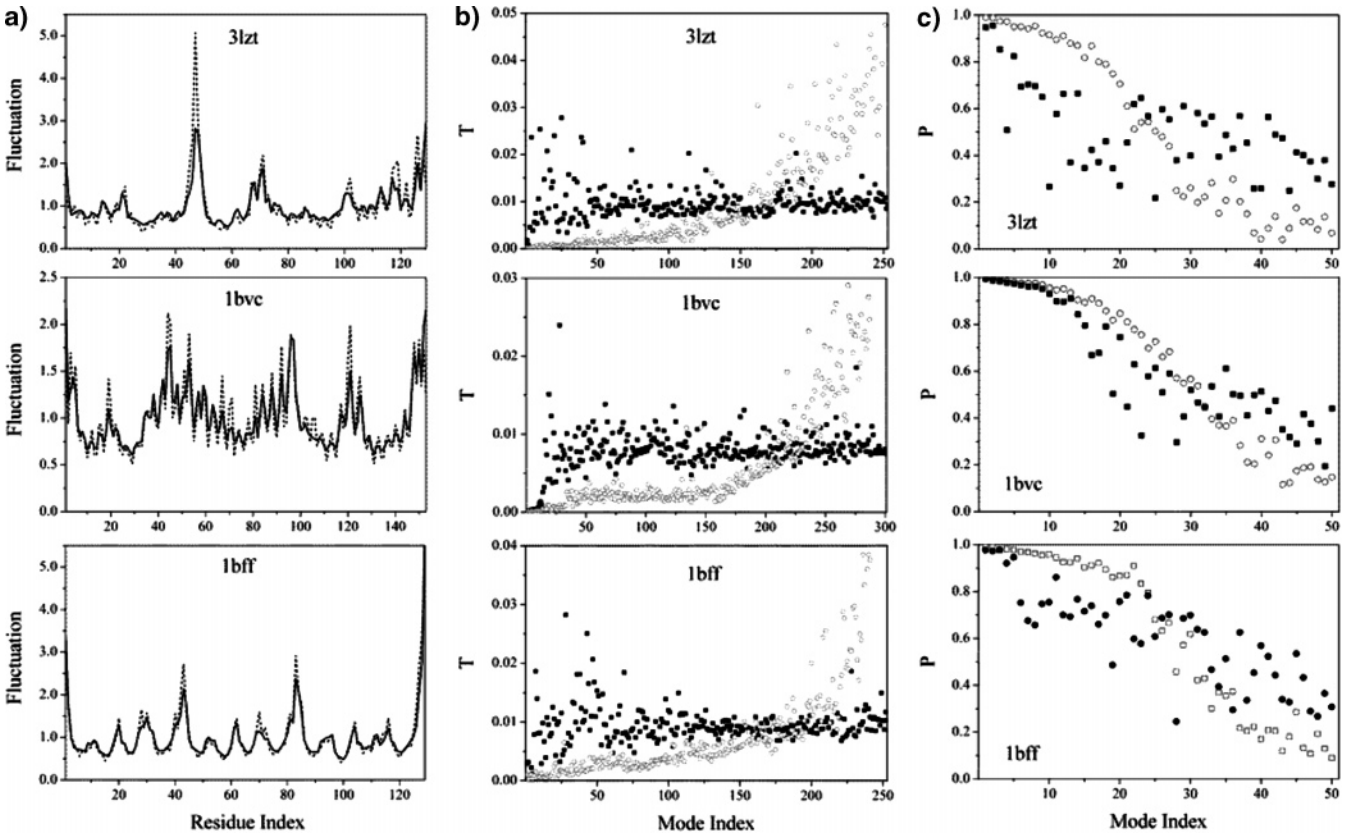

Figure 3. Test of the new eNMA on three different proteins (PDB codes: 3lzt, 1bvc, 1bff). (a) Fluctuation curves calculated by the conventional eNMA (dotted lines) and the new eNMA (solid lines). The curves by the new eNMA are smoother due to the absence of tip effect. (b) Comparison of localization factors (*T* values) for three proteins. The dark squares are for the conventional eNMA and the empty circles are for the new eNMA. The results of the new eNMA are much more rational in low-frequency regime. (c) Results of the projection of an individual mode by one eNMA method onto a subset of 50 low-frequency modes by another method. The empty circles are for case with one new eNMA mode projected onto 50 conventional eNMA modes. The dark squares are just the other way around.

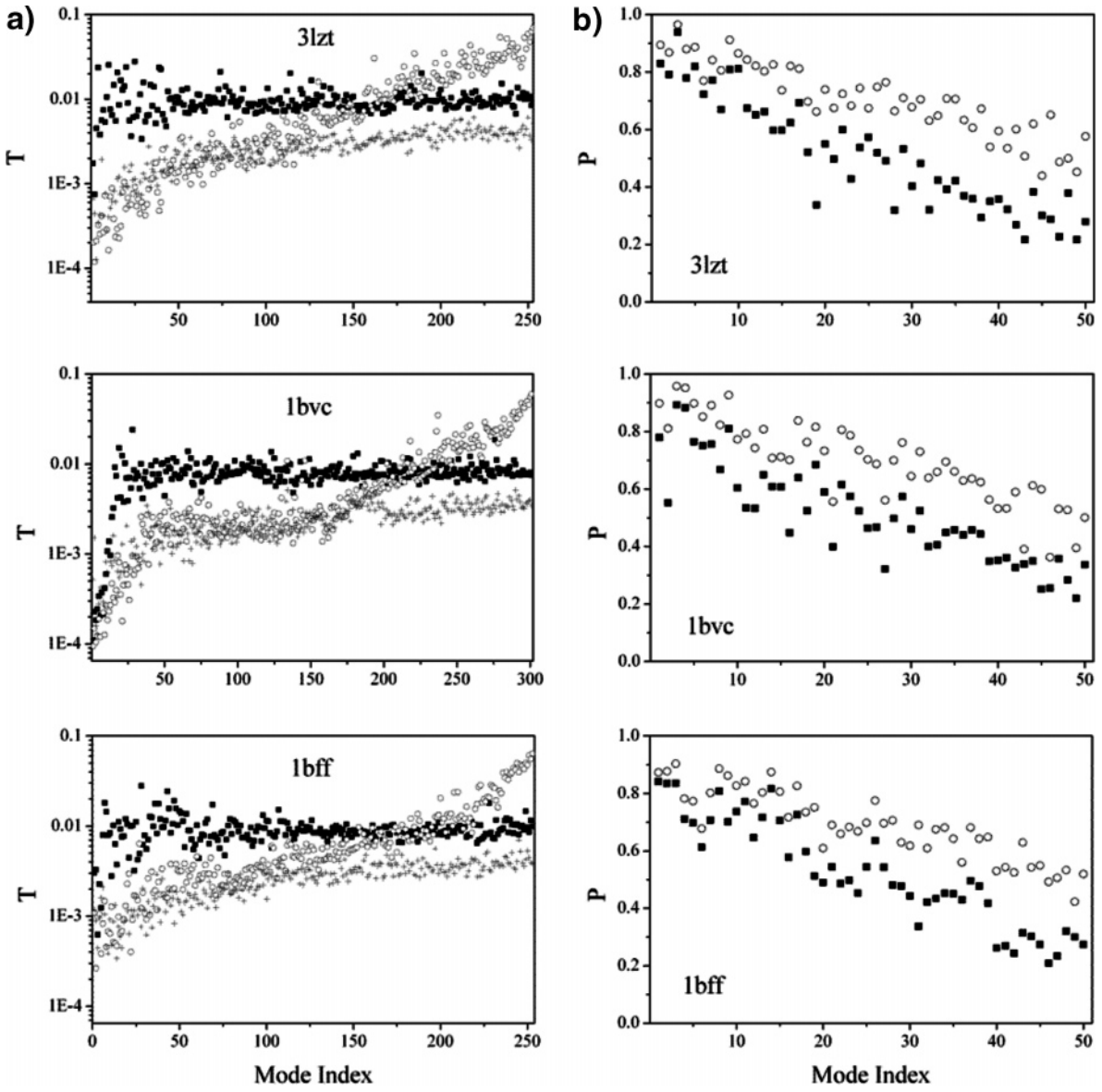

Figure 4. Comparison of modes from the conventional eNMA, the new eNMA, and CHARMM. (a) Localization factors for modes calculated by the conventional eNMA (dark squares), the new eNMA (empty circles), and CHARMM force field (cross). For all three proteins, in the low-frequency modes, the values of localization factor *T* are very much the same between the new eNMA and CHARMM but significantly worse in the conventional eNMA (note the vertical axes are in logarithmic scale). (b) The projection of an individual CHARMM mode (*x*-axis) onto 50 low-frequency mode subspace of the new eNMA (empty circles) and the conventional eNMA (dark squares). The projection coefficients for the new eNMA are systematically larger than those of the conventional eNMA, i.e., the modes from the new eNMA are closer to those of CHARMM modes.

We compared the atomic fluctuation curves computed for three different proteins by the conventional[5] and the new eNMAs in Figure 3a. The curves from the new eNMA (solid lines) are systematically smoother than those from the conventional eNMA as a result of eradication of the tip effect in the eigenvectors by the new eNMA. Figure 3b shows the comparison of the localization factor *T* of low-frequency modes (eq 15) for three proteins (the larger the value of *T* in low-frequency modes, the more localized the motion in that mode, i.e, more eminent the tip effect). In each case, the low-frequency modes from conventional eNMA has a much higher *T* value indicating the contamination of the tip effect. In fact, the *T* value from conventional eNMA stays more or less constant at higher frequency modes because of the divergence of eigenvectors. On the contrary, the *T* value for the modes from the new eNMA is small for low-frequency modes and increases progressively as the frequency increases. This is an expected behavior for normal modes, as the higher the frequency, the more localized motion in the modes. Figure 3c describes the projection of an individual mode by one method onto a subset of 50 low-frequency modes by another method. The purpose was to compare the overall similarity of modes between two methods. The empty circles are the results of using one mode from new eNMA protected onto a subset of 50 modes from conventional eNMA. It is clear that the individual mode of the new eNMA can be reasonably expressed by the subset of low-frequency modes of the conventional eNMA (the lower the frequency of the mode by new eNMA, the better in terms of linear expression). On the contrary, the other

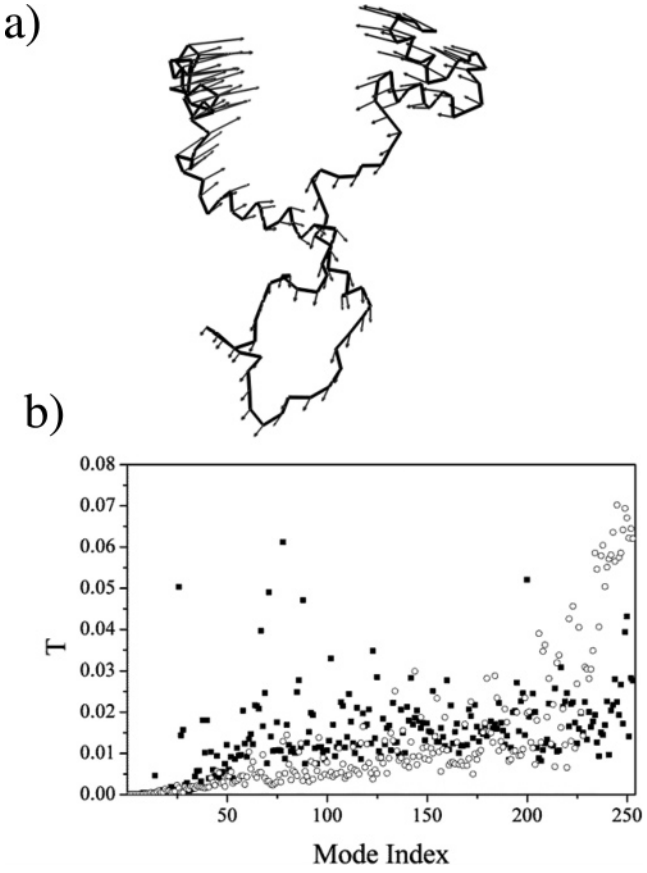

Figure 5. Results on lysozyme (PDB code: 3lzt) with a completed extended conformation. (a) Motional pattern of the first vibrational mode calculated by new eNMA. (b) Tip effect, the solid squares are for conventional eNMA, and the empty circles are for new eNMA.

way around gave much worse results (solid squares). This was because some of the modes of the conventional eNMA are contaminated by the tip effect so that they simply cannot be linearly expressed by a subset of modes of the new eNMA.

Figure 4a shows the comparison of T values for modes computed by the conventional eNMA, the new eNMA, and the CHARMM force field. Clearly the modes by the new eNMA are much closer to those of the CHARMM force field in terms of T values. The curve from the CHARMM force field rises slower than the new eNMA because the CHARMM force field has a much larger mode number. Figure 4b shows the projection of an individual CHARMM mode (x-axis) onto 50 low-frequency mode subspace of the new eNMA (empty circles) and the conventional eNMA (solid squares). The projection coefficients for the new eNMA are systematically larger than those of the conventional eNMA, indicating that the modes of the new eNMA are closer to those of CHARMM. Testing of the method on other proteins led to similar conclusions.

Proteins with a Completely Extended Conformation. Since the tip effect often occurs around less-packed regions in protein such as surface loops, we tested the new eNMA on protein conformations that were totally extended. The purpose of doing so was for cases in which one needs to conduct, for example, a Monte Carlo sampling of protein

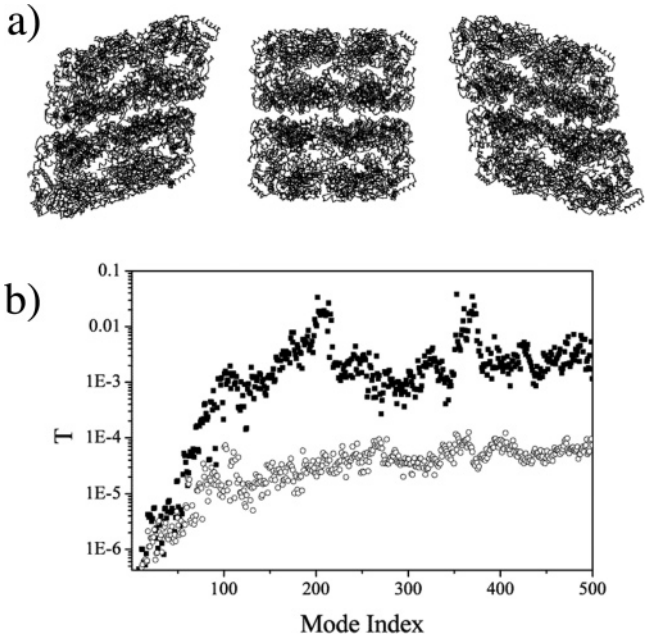

Figure 6. Results on multisubunit supramolecular complex, the molecular chaperonin GroEL. (a) Motional pattern of the second vibrational mode, which is a stretching mode along diagonal line of the molecule. (b) Tip effect, the solid squares are for conventional eNMA, and the empty circles are for new eNMA. Note the vertical axis is made in logarithmic scale.

conformation based on elastic normal modes.[32] Figure 5a shows the motional pattern of the first vibrational mode of the extended lysozyme chain calculated by the new eNMA. Figure 5b shows the T values, the black squares are for conventional eNMA, and the empty circles are for the new eNMA. The tip effect in the low-frequency modes by the conventional eNMA was much more severe.

Proteins with Multiple Polypeptide Chains. We also tested the new method on a multipolypeptide chain complex, the molecular chaperonin GroEL,[33] which is an ATP-driven molecular motor and a supramolecular complex. The molecule has a cylindrical shape formed by two back-to-back stacked 7-fold rings. It is known that GroEL has to undergo huge conformational changes in order to carry out its biological function, which is for facilitating correct folding of unfolded or misfolded polypeptide chains.[34] The system has been a subject of many computational studies.[14,35,36] As an example, Figure 6a shows the motional patterns of a second vibrational mode by the new eNMA. This mode is a stretching mode along a diagonal line of the molecule. The other low-frequency modes are also similar to what was previously observed.[14] For GroEL, however, the tip effect in the conventional eNMA (solid squares) is a few orders of magnitude worse than that in the new eNMA (empty circles). Figure 6b shows the T values for the first 500 vibrational modes; note, the plot was made in the logarithmic scale in the vertical axis.

IV. Concluding Discussion

In this paper, we have reported an improved method for coarse-grained elastic normal-mode analysis. The Hamiltonian used contains the regular term used in conventional

eNMA[1,5] and an additional term that involves pseudo bond angles and dihedrals along the Cα traces. Similar to the Hamiltonian of conventional eNMA, the new one has its minimum at the current structure so that no initial energy minimization is needed. The computational procedure was carried out in the internal coordinate system so that the softest link in elastic potential, the bond length stretching, was avoided. The essential idea of eliminating the tip effect was to increase the stiffness of degrees of freedom even when they are located in regions that are not very densely packed.

We tested our method on proteins with single polypeptide chains in their compact native conformations and extended denatured conformations. We also tested our method on multisubunit supramolecular complexes. Results demonstrated that the new method is capable of easing the tip effect in all cases. It was found that the improvement in the quality of modes was particularly substantial for supramolecular complexes. Therefore, we expect such a method would be a complementary tool to the existing methods. It should be particularly useful in certain special applications in which a set of continuous low-frequency modes are needed such as in structural refinement[37] and in normal mode based sampling.[24]

The scaling factor 3 in eq 5 affects the stiffness of modes. For larger complexes, the value of this factor may be adjusted. In our experience, the value between 3 and 15 is good.

The conventional eNMA was also successfully applied to cryo-EM density maps,[6,8] in which case one does not have the knowledge of chain connectivity. Extension of our current method to cryo-EM density maps requires certain efforts. It may be necessary to artificially construct a pseudochain among all the points used to represent the density maps using methods such as the Traveling Salesman Algorithm. We will report the results in a separate paper.

Acknowledgment. The authors gratefully thank support from the National Institutes of Health (R01-GM067801). B.P. is supported by a predoctoral fellowship from Houston Area Molecular Biophysics Program (HAMBP). M.L. is partially supported by a predoctoral fellowship from W. M. Keck Foundation of the Gulf Coast Consortia through the Keck Center for Computational and Structural Biology. Computer software is available upon request.

References

(1) Tirion, M. M. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* **1996**, *77*, 1905−1908.

(2) Haliloglu, T.; Bahar, I.; Erman, B. Gaussian Dynamics of Folded Proteins. *Phys. Rev. Lett.* **1997**, *79*, 3090−3093.

(3) Bahar, I.; Atilgan, A. R.; Erman, B. Direct evaluation of thermal fluctuations in proteins using a single- parameter harmonic potential. *Fold Des.* **1997**, *2*, 173−181.

(4) Doruker, P.; Jernigan, R. L.; Bahar, I. Dynamic of large proteins through hierarchical levels of coarse-grained structures. *J. Comput. Chem.* **2002**, *23*, 119−127.

(5) Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* **2001**, *80*, 505−515.

(6) Ming, D.; Kong, Y.; Lambert, M.; Huang, Z.; Ma, J. How to Describe Protein Motion Without Amino-acid Sequence and Atomic Coordinates. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 8620−8625.

(7) Ming, D.; Kong, Y.; Wu, Y.; Ma, J. Substructure Synthesis Method for Simulating Large Molecular Complexes. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 104−109.

(8) Tama, F.; Wriggers, W.; Brooks, C. L. Exploring Global Distortions of Biological Macromolecules and Assemblies from Low-resolution Structural Information and Elastic Network Theory. *J. Mol. Biol.* **2002**, *321*, 297−305.

(9) Hinsen, K. Analysis of domain motions by approximate normal mode calculations. *Proteins* **1998**, *33*, 417−429.

(10) Ming, D.; Kong, Y.; Wu, Y.; Ma, J. Simulation of F-actin Filaments of Several Microns. *Biophys. J.* **2003**, *85*, 27−35.

(11) Ming, D.; Kong, Y.; Wakil, S. J.; Brink, J.; Ma, J. Domain Movements in Human Fatty Acid Synthase by Quantized Elastic Deformational Model. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 7895−7899.

(12) Kong, Y.; Ming, D.; Wu, Y.; Stoops, J. K.; Zhou, Z. H.; Ma, J. Conformational Flexibility of Pyruvate Dehydrogenase Complexes: A Computational Analysis by Quantized Elastic Deformational Model. *J. Mol. Biol.* **2003**, *330*, 129−135.

(13) Beuron, F.; Flynn, T. C.; Ma, J.; Kondo, H.; Zhang, X.; Freemont, P. S. Motions and Negative Cooperativity between p97 Domains revealed by Cryoelectron Microscopy and Quantized Elastic Deformational Model. *J. Mol. Biol.* **2003**, *327*, 619−629.

(14) Keskin, O.; Bahar, I.; Flatow, D.; Covell, D. G.; Jernigan, R. L. Molecular Mechanisms of Chaperonin GroEL-GroES Function. *Biochemistry* **2002**, *41*, 491−501.

(15) Wang, Y.; Rader, A. J.; Bahar, I.; Jernigan, R. L. Global ribosome motions revealed with elastic network model. *J. Struct. Biol.* **2004**, *147*, 302−314.

(16) Xu, C.; Tobi, D.; Bahar, I. Allosteric changes in protein structure computed by a simple mechanical model: hemoglobin T ↔ R2 transition. *J. Mol. Biol.* **2003**, *333*, 153−168.

(17) Tama, F.; Valle, M.; Frank, J.; Brooks, C. L., III Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9319−9323.

(18) Tama, F.; Brooks, C. L., III The Mechanism and Pathway of pH Induced Swelling in Cowpea Chlorotic Mottle Virus. *J. Mol. Biol.* **2002**, *318*, 733−747.

(19) Chacon, P.; Tama, F.; Wriggers, W. Mega-dalton biomolecular motion captured from electron microscopy reconstructions. *J. Mol. Biol.* **2003**, *326*, 485−492.

(20) Miyashita, O.; Onuchic, J. N.; Wolynes, P. G. Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12570−12575.

(21) Ma, J. New advances in normal mode analysis of supermolecular complexes and applications to structural refinement. *Curr. Protein Pept. Sci.* **2004**, *5*, 119−123.

(22) Ma, J. Usefulness and Limitations of Normal Mode Analysis in Modeling Dynamics of Biomolecular Complexes. *Structure* **2005**, *13*, 373−380.

(23) Bahar, I.; Rader, A. J. Coarse-grained normal mode analysis in structural biology. *Curr. Opin. Struct. Biol.* **2005**, *15*, 586−592.

(24) Lu, M.; Ma, J. The Role of Shape in Determining Molecular Motions. *Biophys. J.* **2005**, *89*, 2395−2401.

(25) MacKerell, A. D.; Bashford, Jr., D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, I., W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem.* **1998**, *B102*, 3586−3616.

(26) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.

(27) Kamiya, K.; Sugawara, Y.; Umeyama, H. Algorithm for normal-mode analysis with general internal coordinates. *J. Comput. Chem.* **2003**, *24*, 826−841.

(28) Noguti, T.; Go, N. A method of rapid calculation of a second derivative matrix of conformational energy for large molecules. *J. Phys. Soc. Jpn.* **1983**, *52*, 3685−3690.

(29) Eckart, C. Some studies concerning rotating axes and polyatomic molecules. *Phys. Rev.* **1935**, *47*, 552.

(30) McCammon, J. A.; Gelin, B. R.; Karplus, M.; Wolynes, P. G. The hinge-bending mode in lysozyme. *Nature* **1976**, *262*, 325−326.

(31) Brooks, B.; Karplus, M. Normal modes for specific motions of macromolecules: Application to the hinge-bending mode of lysozyme. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 4995−4999.

(32) Wu, Y.; Tian, X.; Lu, M.; Chen, M.; Wang, Q.; Ma, J. Folding of small helical proteins assisted by small-angle X-ray scattering profiles. *Structure (Camb.)* **2005**, *13*, 1587−1597.

(33) Xu, Z.; Sigler, P. B. GroEL/GroES: Structure and function of a two-stroke folding machine. *J. Struct. Biol.* **1998**, *124*, 129−141.

(34) Sigler, P. B.; Xu, Z.; Rye, H. S.; Burston, S. G.; Fenton, W. A.; Horwich, A. L. Structure and function in GroEL-mediated protein folding. *Anuu. Rev. Biochem.* **1998**, *67*, 581−608.

(35) Ma, J.; Karplus, M. The allosteric mechanism of the chaperonin GroEL: A dynamic analysis. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 8502−8507.

(36) Ma, J.; Sigler, P. B.; Xu, Z.; Karplus, M. A dynamic model for the allosteric mechanism of GroEL. *J. Mol. Biol.* **2000**, *302*, 303−313.

(37) Wu, Y.; Ma, J. Refinement of F-actin Model Against Fibre Diffraction Data by Long-range Normal Modes. *Biophys. J.* **2004**, *86*, 116−124.

CT050307U