Toward Modular Analysis of Supramolecular Protein Assemblies

Jaehoon Kim,[†,⊥] Jin-Gyun Kim,[‡,⊥] Giseok Yun,[†] Phill-Seung Lee,[§] and Do-Nyun Kim[*,†,∥]

[†]Department of Mechanical and Aerospace Engineering, [∥]Institute of Advanced Machines and Design, Seoul National University, Gwanak-ro 1, Gwanak-gu, Seoul 08862, Republic of Korea

[‡]Mechanical Systems Safety Research Division, Korea Institute of Machinery & Materials, Gajeongbuk-ro 156, Yuseong-gu, Daejeon 34103, Republic of Korea

[§]Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology, Daehak-ro 291, Yuseong-gu, Daejeon 34141, Republic of Korea

Ⓢ *Supporting Information*

ABSTRACT: Despite recent advances in molecular simulation technologies, analysis of high-molecular-weight structures is still challenging. Here, we propose an automated model reduction procedure aiming to enable modular analysis of these structures. It employs a component mode synthesis for the reduction of finite element protein models. Reduced models may consist of real biological subunits or artificial partitions whose dynamics is described using the degrees of freedom at the substructural interfaces and a small set of dominant vibrational modes only. Notably, the proper number of dominant modes is automatically determined using a novel estimator for eigenvalue errors without calculating the reference eigensolutions of the full model. The performance of the proposed approach is thoroughly investigated by analyzing 50 representative structures including a crystal structure of GroEL and an electron density map of a ribosome.
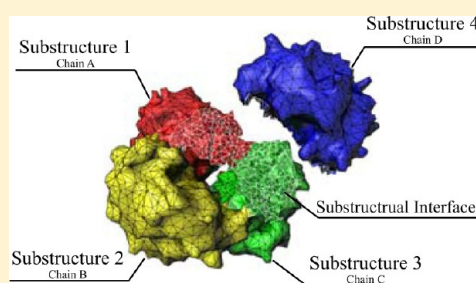
1. INTRODUCTION

Supramolecular proteins and their assemblies participate actively in many cellular functions.[1] While molecular dynamics (MD) simulations provide important insights into the conformational dynamics of these proteins in atomic details,[2] the size of the target molecule and the physical time scale of its motion that can be simulated often limit the applicability of this method.[3] Hence, normal mode analysis (NMA) has been widely used instead, which seeks natural frequencies and corresponding mode shapes of a molecular structure near its equilibrium conformational state by approximating the atomic energy landscape as a simple harmonic potential.[4−7] Its popularity stems from the fact that it is not necessary to calculate the entire number of modes and frequencies in practice because only a small set of low-frequency vibrational modes dominates the biologically important collective functional motions.

Classical atomistic NMA based on full[8] or simplified atomic potentials[4,9−11] is, however, still computationally demanding, as it requires the second derivative of the potential field about an equilibrium state to be calculated, which also needs to be found first using the energy minimization process. For this reason, much simpler, coarse-grained representations of protein structures have been proposed following pioneering works by Gō[9] and Tirion,[12] including the elastic network model (ENM)[12] and its variants[13−16] as well as the finite element (FE) model.[17,18] ENM is arguably the most popular coarse-grained approach, where proteins are modeled as a network of representative atoms, usually α carbons, connected by linear elastic springs to their neighbors within a cutoff distance.[12,19] Despite its

simplicity, ENM has proven to be successful in calculating dynamic properties of protein structures such as low-frequency normal modes, thermal fluctuations in equilibrium,[20,21] and conformational transition pathways.[19] FE model is another popular approach to coarse-graining whose solution accuracy is comparable to that of ENM.[17] It treats proteins as elastic, continuous media enclosed by the molecular surface with homogeneous, isotropic material properties. The FE modeling approach has a distinct advantage over ENM in that it explicitly models the molecular surface, which becomes essential for problems in which the effects of external mechanical forces or the surrounding medium must be taken into account, for example, when simulating the indentation of viral capsids,[22] calculating surface electrostatics,[23] and computing solvent-damping effects,[24] although it requires effort to construct defect-free molecular surfaces unless it is automated.[18,25]

Nevertheless, analysis of high-molecular-weight protein assemblies is challenging even with these coarse-grained modeling approaches. This is partly because we need to model the entire structure of protein assemblies as a whole for analysis without exploiting the intrinsic modularity of their structure. It would be ideal if we could calculate the dynamics of the whole assembly structure using the known properties of individual protein components or by computing them without building a full assembly model. In fact, this type of analysis has been widely used already in many engineering fields in a process termed

component mode synthesis (CMS),[26−30] which is a popular model reduction technique. To illustrate, the structure of an aircraft can be treated as an assembly of the fuselage and wings whose computational models can be constructed independently except at the shared boundaries interfacing the components. These structural components are generally called substructures in CMS. Since their structural information is transmitted to one another through the interfaces only, each component model effectively reduces with dominant substructural modes to its boundaries and therefore the dynamics of the entire aircraft can be calculated using the reduced models of the fuselage and wings, enabling efficient modular analysis and design. For these reasons, CMS is well-suited to build reduced models of supramolecular protein complexes that are the functionally programmed aggregation of smaller biological subunits in nature. In addition, model reduction of proteins to their interacting molecular surfaces would be useful when analyzing, for example, interactions between constituting units.

CMS, although developed for FE models originally, can be applied to any other protein models because it is in principle a matrix projection technique. For instance, CMS has been successfully combined with ENM to develop a hierarchical decomposition and analysis protocol for large protein dynamics.[7,31] Despite its computational efficiency demonstrated in the previous work, ENM-based model reduction methods suffer from the fact that they require the boundary residues (or their representative atoms) between substructures to be chosen, and the selection criterion for this affects the accuracy of the solution.[7] If the chosen number of boundary residues is either too small or large, then the entire assembly becomes more flexible or stiffer than it really is, respectively. Moreover, one drawback of using CMS is that we need to choose the proper number of substructural modes to be used, which determines the reliability of the reduced model. Validating the reduced model via direct comparison of its solution with the reference one is not practical because it requires building and analyzing the full assembly model, rendering it necessary to use an accurate error estimator without calculating the reference full model solution.

To overcome these limitations, here we propose an automated FE-based model reduction method as a first step toward modular analysis of supramolecular protein assemblies. The three-dimensional FE models can be constructed in an unsupervised manner for any available protein structures given in terms of the atomic coordinates (PDB, http://www.pdb.org/) or the electron density maps (EMDB, http://www.emdatabank.org/). The Craig−Bampton (CB) method,[27] which is the most popular CMS method, is used to build a reduced FE model whose substructures can be any components of an assembly, including real biological subunits, user-defined subdomains, and even random partitions. Unlike ENM, the shared boundaries of the substructures are uniquely defined in the FE model without introducing any free parameter. Most importantly, the proposed method employs a novel error estimator[32,33] that evaluates the accuracy of the eigenvalues obtained using the reduced model precisely without prior knowledge on the eigenvalues of the unreduced full model, enabling fine control over the reliability of the reduced model. The entire procedure has been fully automated so that the reduced model is iteratively refined until we obtain the solution within a user-defined error tolerance. We evaluate the proposed method by analyzing 50 structures and present, in particular, the results of two biologically important and extensively investigated molecular machines, GroEL and ribosome, in more detail.

2. METHODS

2.1. Finite Element Model. FE model construction begins with computing the molecular surface of protein structures. When high-resolution atomic coordinates are available, as in Figure 1A, the solvent-excluded surface (Figure 1B), which is
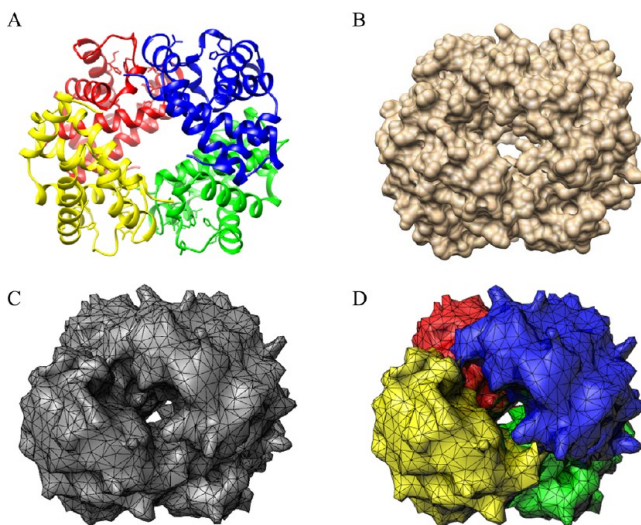

Figure 1. Structure and FE models of hemoglobin. (A) Atomic structure (PDB ID: 1A3N), (B) solvent-excluded surface, (C) original FE model, and (D) partitioned FE model consisting of four substructures. Each substructure corresponds to a chain of hemoglobin. Scale bar represents 10 Å.

conventionally used as the molecular surface, can be computed by rolling a probe over the van der Waals surface of atoms.[17,34] Here, we use a freely available program, MSMS, version 2.6.1[34] (http://mgltools.scripps.edu/packages/MSMS/), with the probe radius of 1.5 Å representing the size of a single water molecule to calculate the triangulated molecular surfaces. For structures provided as electron density maps, the molecular surface is defined as a contour surface whose density level is determined so that the volume enclosed by the contour surface matches with the expected molecular volume.[18] UCSF Chimera, version 1.8.1[35] (http://www.cgl.ucsf.edu/chimera/), is used to obtain the contour surfaces of these structures.

Discretized molecular surfaces obtained using these programs often possess some defects including holes, self-intersecting triangles, isolated fragments, and nonmanifolds that must be repaired before generating the three-dimensional volumetric mesh. This mesh cleanup is accomplished automatically using a carefully designed set of mesh-repairing filters freely available in MeshLab, version 1.3.3[36] (http://meshlab.sourceforge.net/), followed by generation of the three-dimensional volumetric model with four-node tetrahedral elements (Figure 1C) using the commercially available FE analysis program ADINA, version 9.0.4 (ADINA R&D, Inc., Watertown, MA, USA). To build a reduced model later, as described in the next section, each tetrahedron is labeled according to the component to which it belongs, which can be a real biological subunit (Figure 1D), a user-defined subdomain, or a randomly partitioned substructure.

2.2. Component Mode Synthesis: Craig−Bampton Method. The equation of motion for FE protein models in vacuum can be written as $\mathbf{M}(d^2\mathbf{x}/dt^2) + \mathbf{K}\mathbf{x} = 0$ when no external force is applied, where $\mathbf{M}$ is the mass matrix, $\mathbf{K}$ is the stiffness matrix, $\mathbf{x}$ denotes the displacement vector, and t represents time.

For harmonic motions, the displacement vector can be defined as $\mathbf{x} = \boldsymbol{\varphi} \exp(i\omega t)$ with the imaginary i, where ω represents natural frequency and $\boldsymbol{\varphi}$ is corresponding mode shape. These natural frequencies and normal modes can be obtained by solving the following eigenvalue problem, which is referred to as NMA

$$\mathbf{K}(\boldsymbol{\varphi})_i = \lambda_i \mathbf{M}(\boldsymbol{\varphi})_i \text{ with } i = 1, 2, ..., N \tag{1}$$

where N is the total number of degrees of freedom (DOFs) in the FE model and $\lambda_i = \omega_i^2$.

To construct the reduced FE model, we here employ the CB method, the most popular CMS method. In the CB method, the mass and stiffness matrices and the displacement vector of a FE model consisting of N_s substructures can be rewritten as

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_s & \mathbf{M}_c \\ \mathbf{M}_c^T & \mathbf{M}_b \end{bmatrix}, \mathbf{K} = \begin{bmatrix} \mathbf{K}_s & \mathbf{K}_c \\ \mathbf{K}_c^T & \mathbf{K}_b \end{bmatrix}, \mathbf{x} = \begin{bmatrix} \mathbf{x}_s \\ \mathbf{x}_b \end{bmatrix} \tag{2a}$$

$$\mathbf{M}_s = \begin{bmatrix} \mathbf{M}_s^{(1)} & & & & \\ & \ddots & & \mathbf{0} & \\ & & \mathbf{M}_s^{(k)} & & \\ & \mathbf{0} & & \ddots & \\ & & & & \mathbf{M}_s^{(N_s)} \end{bmatrix},$$

$$\mathbf{K}_s = \begin{bmatrix} \mathbf{K}_s^{(1)} & & & & \\ & \ddots & & \mathbf{0} & \\ & & \mathbf{K}_s^{(k)} & & \\ & \mathbf{0} & & \ddots & \\ & & & & \mathbf{K}_s^{(N_s)} \end{bmatrix}, \mathbf{x}_s = \begin{bmatrix} \mathbf{x}_s^{(1)} \\ \vdots \\ \mathbf{x}_s^{(k)} \\ \vdots \\ \mathbf{x}_s^{(N_s)} \end{bmatrix} \tag{2b}$$

in which $\mathbf{M}_s^{(k)}$ and $\mathbf{K}_s^{(k)}$ are the mass and stiffness matrices of the kth substructure, respectively, and $\mathbf{x}_s^{(k)}$ is the corresponding substructural displacement vector. Subscripts s, b, and c denote substructural, interface boundary, and coupling quantities, respectively. The substructural displacement vector can be expressed as $\mathbf{x}_s = \boldsymbol{\Phi}_s \mathbf{q}_s$, where $\boldsymbol{\Phi}_s$ is a block diagonal matrix storing the substructural normal modes and $\mathbf{q}_s$ denotes the vector of generalized coordinates. $\boldsymbol{\Phi}_s$ can be calculated by solving the eigenvalue problem of each substructure, $\mathbf{K}_s^{(k)}\boldsymbol{\Phi}_s^{(k)} = \mathbf{M}_s^{(k)}\boldsymbol{\Phi}_s^{(k)}\boldsymbol{\Lambda}_s^{(k)}$ ($k = 1, 2, ..., N_s$), where $\boldsymbol{\Lambda}_s^{(k)}$ stores the eigenvalues of the kth substructure on its diagonal. Then, the displacement vector becomes

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_s \\ \mathbf{x}_b \end{bmatrix} = \mathbf{T}_0 \begin{bmatrix} \mathbf{q}_s \\ \mathbf{x}_b \end{bmatrix}, \mathbf{T}_0 = \begin{bmatrix} \boldsymbol{\Phi}_s & -\mathbf{K}_s^{-1}\mathbf{K}_c \\ \mathbf{0} & \mathbf{I}_b \end{bmatrix} \tag{3}$$

where $\mathbf{T}_0$ is the transformation matrix of the CB method and $\mathbf{I}_b$ is an identity matrix of interface boundary. Here, $\boldsymbol{\Phi}_s$ and $\mathbf{q}_s$ can be further decomposed into their dominant and residual modes as

$$\boldsymbol{\Phi}_s = [\boldsymbol{\Phi}_d \quad \boldsymbol{\Phi}_r], \mathbf{q}_s = \begin{bmatrix} \mathbf{q}_d \\ \mathbf{q}_r \end{bmatrix} \tag{4}$$

where subscripts d and r represent dominant and residual quantities, respectively. It is noteworthy that the low-frequency normal modes of substructures usually form the dominant modes and that this number is significantly smaller than the number of residual modes ($N_d \ll N_r$) in practice. By neglecting the residual

modes in eqs 3 and 4, we can approximate the displacement vector and the transformation matrix as

$$\mathbf{x} \approx \overline{\mathbf{x}} = \overline{\mathbf{T}}_0 \begin{bmatrix} \mathbf{q}_d \\ \mathbf{x}_b \end{bmatrix}, \overline{\mathbf{T}}_0 = \begin{bmatrix} \boldsymbol{\Phi}_d & -\mathbf{K}_s^{-1}\mathbf{K}_c \\ \mathbf{0} & \mathbf{I}_b \end{bmatrix} \tag{5}$$

where overbar $(^-)$ denotes approximated quantities throughout the article.

Then, the eigenvalue problem can be written using $\overline{\mathbf{T}}_0$ as

$$\overline{\mathbf{K}}_{CB}(\overline{\boldsymbol{\varphi}}_{CB})_i = \overline{\lambda}_i \overline{\mathbf{M}}_{CB}(\overline{\boldsymbol{\varphi}}_{CB})_i \text{ with } i = 1, 2, ..., N_{CB} \tag{6a}$$

$$\overline{\mathbf{M}}_{CB} = \overline{\mathbf{T}}_0^T \mathbf{M} \overline{\mathbf{T}}_0, \overline{\mathbf{K}}_{CB} = \overline{\mathbf{T}}_0^T \mathbf{K} \overline{\mathbf{T}}_0 \tag{6b}$$

where $\overline{\mathbf{M}}_{CB}$ and $\overline{\mathbf{K}}_{CB}$ are the reduced mass and stiffness matrices, respectively, and $\overline{\lambda}_i$ and $(\overline{\boldsymbol{\varphi}}_{CB})_i$ correspond to the eigenvalue and eigenvector of the reduced system, respectively. N_{CB} is the total number of DOFs in the reduced model, which is much smaller than N because we use the dominant modes only to construct the reduced model ($N_{CB} \ll N$). Finally, the eigenvectors of the original FE model without reduction can be recovered using

$$(\boldsymbol{\varphi})_i \approx (\overline{\boldsymbol{\varphi}})_i = \overline{\mathbf{T}}_0(\overline{\boldsymbol{\varphi}}_{CB})_i \tag{7}$$

satisfying mass-orthonormality and stiffness-orthogonality conditions for the original mass ($\mathbf{M}$) and stiffness ($\mathbf{K}$) matrices.

2.3. Estimation of Eigenvalue Errors. To evaluate the reliability of reduced models, the following equations are generally used

$$\xi_i = \frac{|\overline{\lambda}_i - \lambda_i|}{\lambda_i} \tag{8a}$$

$$\chi_{ij} = (\boldsymbol{\varphi})_i^T \mathbf{M}(\overline{\boldsymbol{\varphi}})_j \tag{8b}$$

in which ξ_i denotes the relative eigenvalue error of the ith mode and χ_{ij} represents the cross-orthogonality check between the ith reference eigenvector and the jth approximated eigenvector, which becomes close to 1 when their correlations are high. The cross-orthogonality check is also well-known as the overlap parameter.[37] Ideally, the accuracy of the reduced model using the CB method can be evaluated exactly using eq 8 if the reference eigensolutions are available. However, this method cannot be practically used because we would like to avoid calculating the reference eigensolutions in any case. Hence, we adopt a recently developed error estimator[32,33] that demonstrates the precise estimation of eigenvalue errors without any knowledge of the reference eigenvalues, which is briefly described below.

The eigenvalue problem of the original FE model in eq 1 can be rewritten as

$$\frac{1}{\lambda_i}(\boldsymbol{\varphi})_i^T \mathbf{K}(\boldsymbol{\varphi})_i = (\boldsymbol{\varphi})_i^T \mathbf{M}(\boldsymbol{\varphi})_i \tag{9}$$

and the reference eigenvalues and eigenvectors can be expressed using their approximated ones and the corresponding errors as

$$\lambda_i = \overline{\lambda}_i + \delta\lambda_i, (\boldsymbol{\varphi})_i = (\overline{\boldsymbol{\varphi}})_i + (\delta\boldsymbol{\varphi})_i \tag{10}$$

where $\delta\lambda_i$ and $(\delta\boldsymbol{\varphi})_i$ are the eigenvalue and eigenvector errors of the ith mode, respectively. If we consider the effect of residual modes,[32] then the approximated eigenvector $(\overline{\boldsymbol{\varphi}})_i$ can be defined as follows
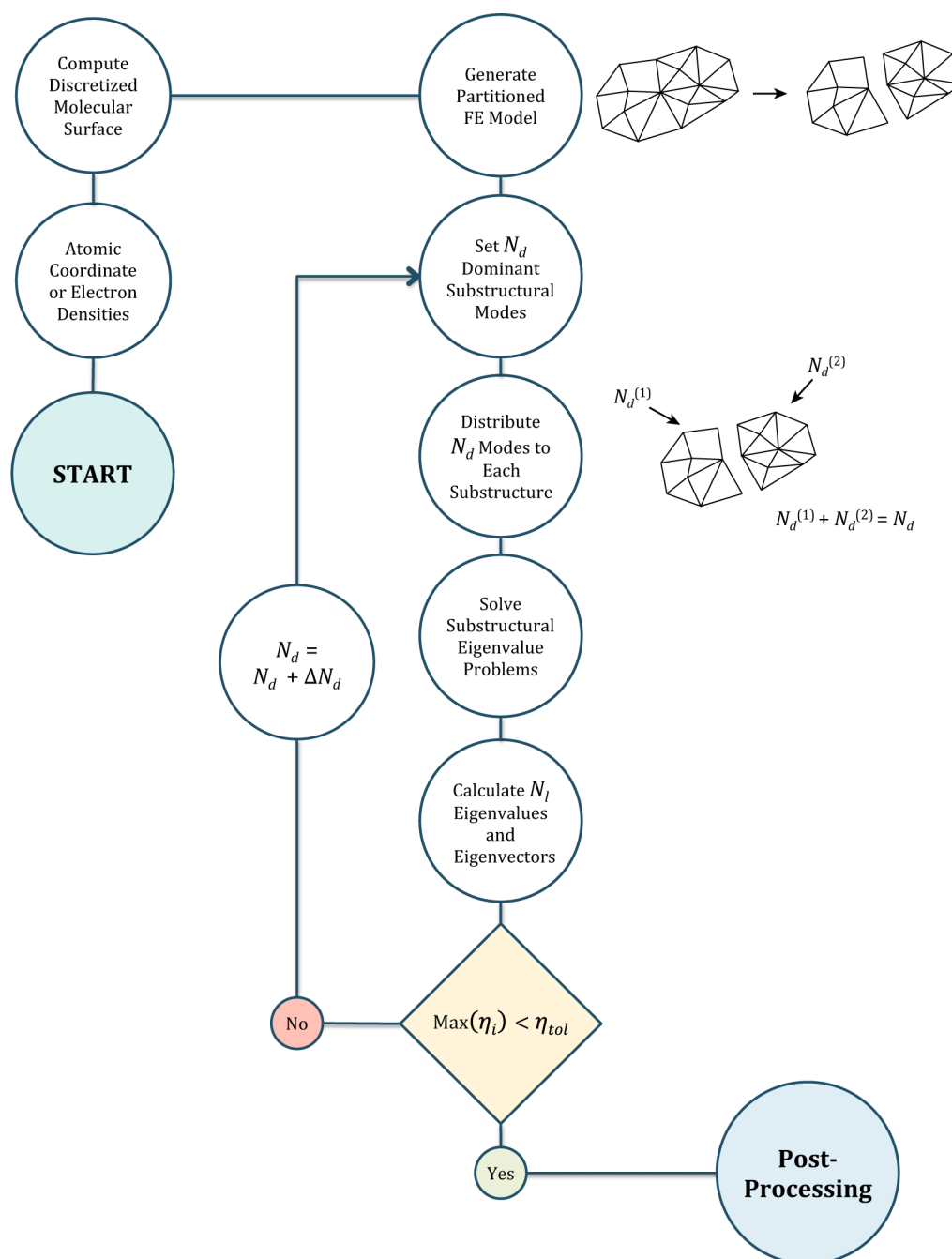
Figure 2. Automated model reduction procedure for FE protein models. In this study, we use $N_l = 200$, $N_d = 20$, $\Delta N_d = 20$, and $\eta_{tol} = 0.1$.

$$(\bar{\boldsymbol{\varphi}})_i = [\bar{\mathbf{T}}_0 + \bar{\mathbf{T}}_r](\bar{\boldsymbol{\varphi}}_{CB})_i,$$

$$\bar{\mathbf{T}}_r = \lambda_i \begin{bmatrix} \mathbf{0} & \mathbf{F}_{rs}[-\mathbf{M}_s\mathbf{K}_s^{-1}\mathbf{K}_c + \mathbf{M}_c] \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{11}$$

with

$$\mathbf{F}_{rs} = \mathbf{K}_s^{-1} - \boldsymbol{\Phi}_d\boldsymbol{\Lambda}_d^{-1}\boldsymbol{\Phi}_d^T \tag{12}$$

where $\bar{\mathbf{T}}_r$ is the additional transformation matrix due to the effect of residual modes and $\mathbf{F}_{rs}$ is called the residual flexibility of substructures. Due to compensation of the residual modes, $(\bar{\boldsymbol{\varphi}})_i$ in eq 11 is much more accurate than the one defined in the original CB method as in eq 7. Then, the relative eigenvalue error becomes

$$\frac{\bar{\lambda}_i}{\lambda_i} - 1 = 2(\bar{\boldsymbol{\varphi}}_{CB})_i^T \bar{\mathbf{T}}_0^T \left[\mathbf{M} - \frac{1}{\lambda_i}\mathbf{K}\right]\bar{\mathbf{T}}_r(\bar{\boldsymbol{\varphi}}_{CB})_i$$

$$+ (\bar{\boldsymbol{\varphi}}_{CB})_i^T \bar{\mathbf{T}}_r^T \left[\mathbf{M} - \frac{1}{\lambda_i}\mathbf{K}\right]\bar{\mathbf{T}}_r(\bar{\boldsymbol{\varphi}}_{CB})_i$$

$$+ (\delta\boldsymbol{\varphi})_i^T \left[\frac{1}{\lambda_i}\mathbf{K} - \mathbf{M}\right](\delta\boldsymbol{\varphi})_i \tag{13}$$

Since $\bar{\lambda}_i$ is always larger than λ_i in the CB method,[38] the left-hand side in eq 13 is the relative eigenvalue error in eq 8a. Consequently, eq 13 is another expression of the relative eigenvalue error. If we assume that the approximated eigenvector $(\bar{\boldsymbol{\varphi}})_i$ is sufficiently close to the reference one $(\boldsymbol{\varphi})_i$, i.e., $(\boldsymbol{\varphi})_i \approx$
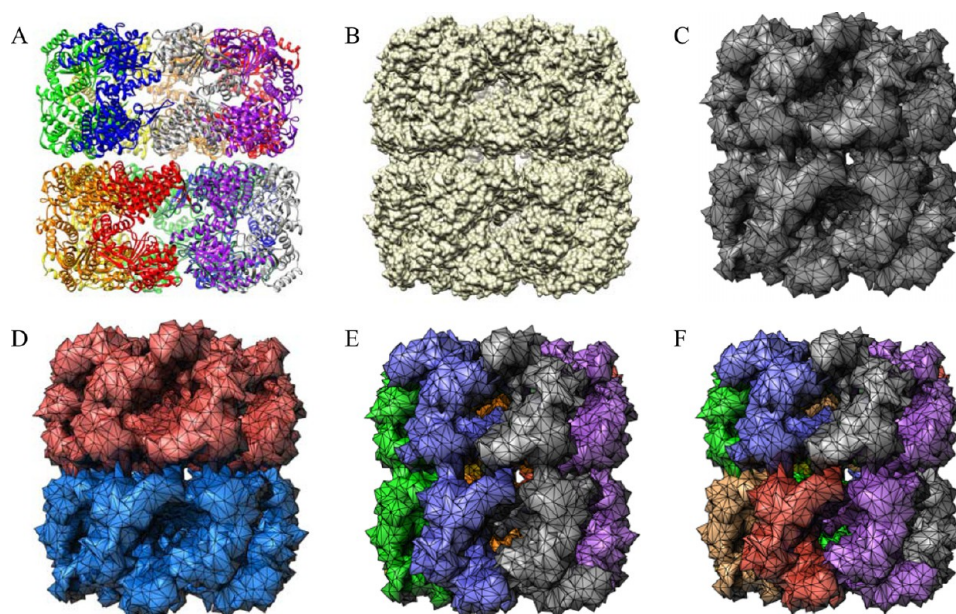
Figure 3. Structure and FE models of bacterial chaperonin GroEL. (A) Atomic structure (PDB ID: 1XCK), (B) solvent-excluded surface, (C) original FE model, and (D−F) partitioned FE models. Each substructure corresponds to one of two rings in the two-substructure model (D), two subunits stacked across the rings in the seven-substructure model (E), and a single subunit in the 14-substructure model (F). Scale bar represents 20 Å.

$(\overline{\boldsymbol{\varphi}})_i$, then the last term of the right-hand side in eq 13 is much smaller than other terms, and it can be negligible. Then, an estimator for the relative eigenvalue error, η_i is defined as

$$
\eta_i = 2(\overline{\boldsymbol{\varphi}}_{CB})_i^T \overline{\mathbf{T}}_0^T \left[\mathbf{M} - \frac{1}{\lambda_i} \mathbf{K} \right] \overline{\mathbf{T}}_r (\overline{\boldsymbol{\varphi}}_{CB})_i
$$
$$
+ (\overline{\boldsymbol{\varphi}}_{CB})_i^T \overline{\mathbf{T}}_r^T \left[\mathbf{M} - \frac{1}{\lambda_i} \mathbf{K} \right] \overline{\mathbf{T}}_r (\overline{\boldsymbol{\varphi}}_{CB})_i
$$

$$(14)$$

with

$$
\overline{\mathbf{T}}_r = \overline{\lambda}_i \begin{bmatrix} \mathbf{0} & \mathbf{F}_{rs}[-\mathbf{M}_s \mathbf{K}_s^{-1} \mathbf{K}_c + \mathbf{M}_c] \\ \mathbf{0} & \mathbf{0} \end{bmatrix}
$$

$$(15)$$

Note that the approximated eigenvalue $\overline{\lambda}_i$ that can be calculated using the reduced model is used in eqs 14 and 15 instead of the reference eigenvalue λ_i in eq 13. Derivation details and other characteristics of the error estimator can be found in refs 32 and 33.

2.4. Automated Model Reduction and Solution Procedure. By combining unsupervised FE model generation and the error estimator for eigenvalues, we develop an automated procedure for FE-based model reduction and NMA of supramolecular protein assemblies (Figure 2). This procedure determines the proper number of substructural modes to be used iteratively based on the estimated eigenvalue errors. Initially, N_d substructural modes in total are distributed to each substructure in proportion to its number of DOFs in order to build a reduced model, which is used to compute N_l lowest normal-mode solutions of the entire structure. Then, the relative eigenvalue errors (η_i) are calculated for each mode and compared with a predefined value of error tolerance (η_{tol}). If all of the relative eigenvalue errors are smaller than the given error tolerance, then we stop the analysis and process the results. Otherwise, we increase the number of substructural modes by ΔN_d and repeat the process until the estimated errors do not exceed the

tolerance. We use $N_l = 200$, $N_d = 20$, $\Delta N_d = 20$, and $\eta_{tol} = 0.1$ in this study.

3. RESULTS AND DISCUSSION

Here, we analyze, using the proposed method, a set of 50 structures including GroEL and ribosome that are actively involved in protein folding[39] and synthesis,[40] respectively.

First, we consider the apo wild-type chaperonin GroEL from *Escherichia coli*, whose structure is determined by X-ray crystallography at 2.9 Å resolution (Protein Data Bank ID: 1XCK).[39] The complex is composed of two coaxial rings, consisting of seven identical subunits each, stacked back to back. Two rings are connected by equatorial domains of each subunit where ATP binds that are connected via intermediate domains to apical domains that contain the binding sites for co-chaperonin GroES.

Next, a eukaryotic ribosome (80S) is analyzed that consists of two biological subunits: the large (60S) and small (40S) ribosomal subunits. Ribosomes produce a polypeptide chain by connecting, in the large subunit, amino acids delivered by transfer RNA (tRNA) in the order that is encoded in messenger RNA (mRNA) and decoded by the small ribosomal subunit. We use the structure obtained using cryo-electron microcopy at a resolution of 5.57 Å (Electron Microscopy Data Bank ID: EMD-2239)[40] to demonstrate the general applicability of the proposed method, although its corresponding crystal structure at a higher resolution is also available. We evaluate the normal modes and the derived results at α carbon atoms for protein residues and phosphorus atoms for RNA residues.

The remaining 48 structures are chosen to cover a broad range of protein size and topology. The number of residues ranges from 76 to 11 568, whereas the molecular weight varies from 8.6 kDa to 2.3 MDa (Table S1). Principal component analysis for the three-dimensional coordinates of these structures confirms the diversity of their shape as well (Figure S2).

3.1. Reduced Models. The molecular surface of GroEL is constructed by computing the solvent-excluded surface of the
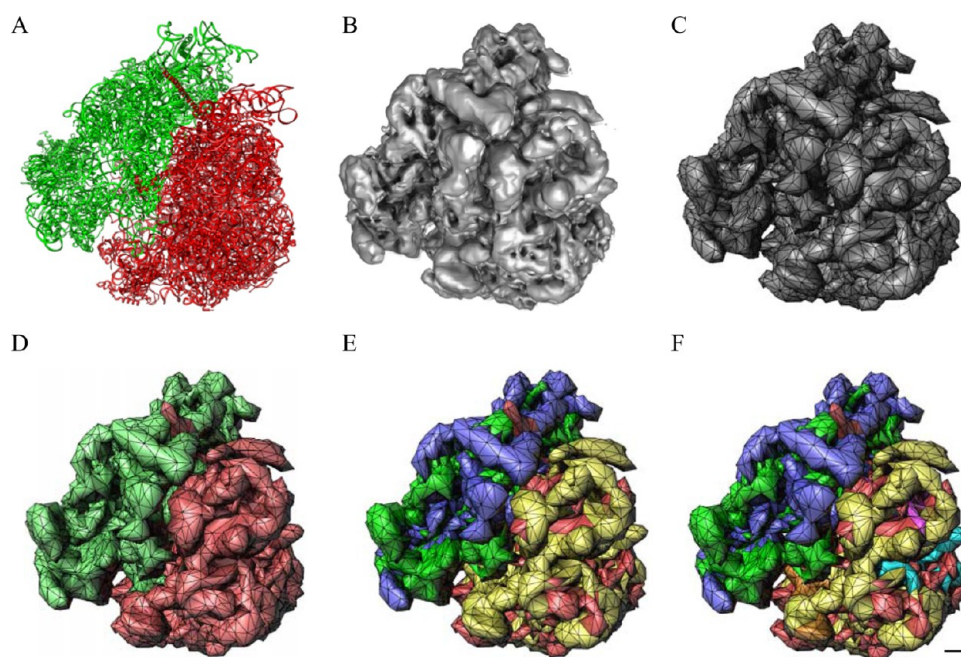
Figure 4. Structure and FE models of a ribosome. (A) Atomic structure (EMDB ID: 2239), (B) contour surface from electron densities, (C) original FE model, and (D−F) partitioned FE models. (D) The two-substructure model consists of 40S (green) and 60S (red) ribosomal subunits that are divided into ribosomal proteins (green and red) and RNAs (blue and yellow) in (E) the four-substructure model. (F) The eight-substructure model is obtained by subdividing RNAs in the four-substructure model into 18S rRNA (blue), 28S rRNA (yellow), 5S rRNA (gray), 5.8S rRNA (cyan), tRNA in the E-site (orange), and short rRNAs (purple). Scale bar represents 20 Å.
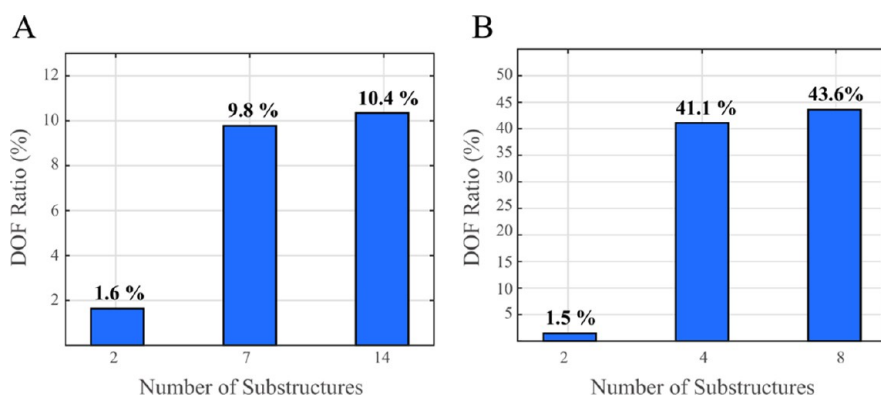

Figure 5. Number of DOFs used in the reduced FE models of (A) GroEL and (B) ribosome with respect to those in the original FE models.

crystal structure (Figure 3A,B). The initially obtained molecular surface consists of 63 028 triangles, which subsequently reduces to 19 954 triangular faces using the quadric edge collapse decimation[41] available in MeshLab. The corresponding three-dimensional volumetric mesh is then generated for the entire structure that is composed of 84 111 tetrahedral elements and 18 665 nodal points using ADINA (Figure 3C). Finally, we build three reduced models of GroEL: the two-substructure model, where the entire structure is partitioned into two rings (Figure 3D), the seven-substructure model, where each substructure consists of two subunits stacked across the rings (Figure 3E), and the 14-substructure model, where each subunit corresponds to one substructure (Figure 3F). The number of interface boundary nodes increases naturally with the number of substructures so that there exist 232, 1757, and 1865 nodes at the interfaces of the two-, seven-, and 14-substructure models, respectively.

The FE-based modeling approach can be applied also to a molecular structure given as electron densities, which, in general, better represents the native conformational states of molecules.

To illustrate, we construct the FE model of a ribosome from its electron density map (Figure 4B), even though its crystal structure at a higher resolution is available as well. Here, we used the contour level of 33 800 to calculate the molecular surface so that the molecular weight of the model becomes the expected one for a ribosome (3305 MDa) assuming an average protein mass density of 1.35 g/cm^3.[42] The final FE model for the entire structure is composed of 62 411 tetrahedral elements and 14 906 nodal points (Figure 4C), which reduces subsequently to the two-substructure model consisting of two ribosomal subunits, 40S and 60S (Figure 4D), the four-substructure model, where each ribosomal subunit is divided into ribosomal proteins and RNAs (Figure 4E), and finally the eight-substructure model obtained by subdividing RNAs in the four-substructure model into 18S rRNA, 28S rRNA, 5S rRNA, 5.8S rRNA, tRNA in the E-site, and short rRNAs[40] (Figure 4F). The resulting interface boundaries of the two-, four-, and eight-substructure models have 498, 6061, and 6430 nodes, respectively. For the remaining 48 structures, three reduced models consisting of two, four, and
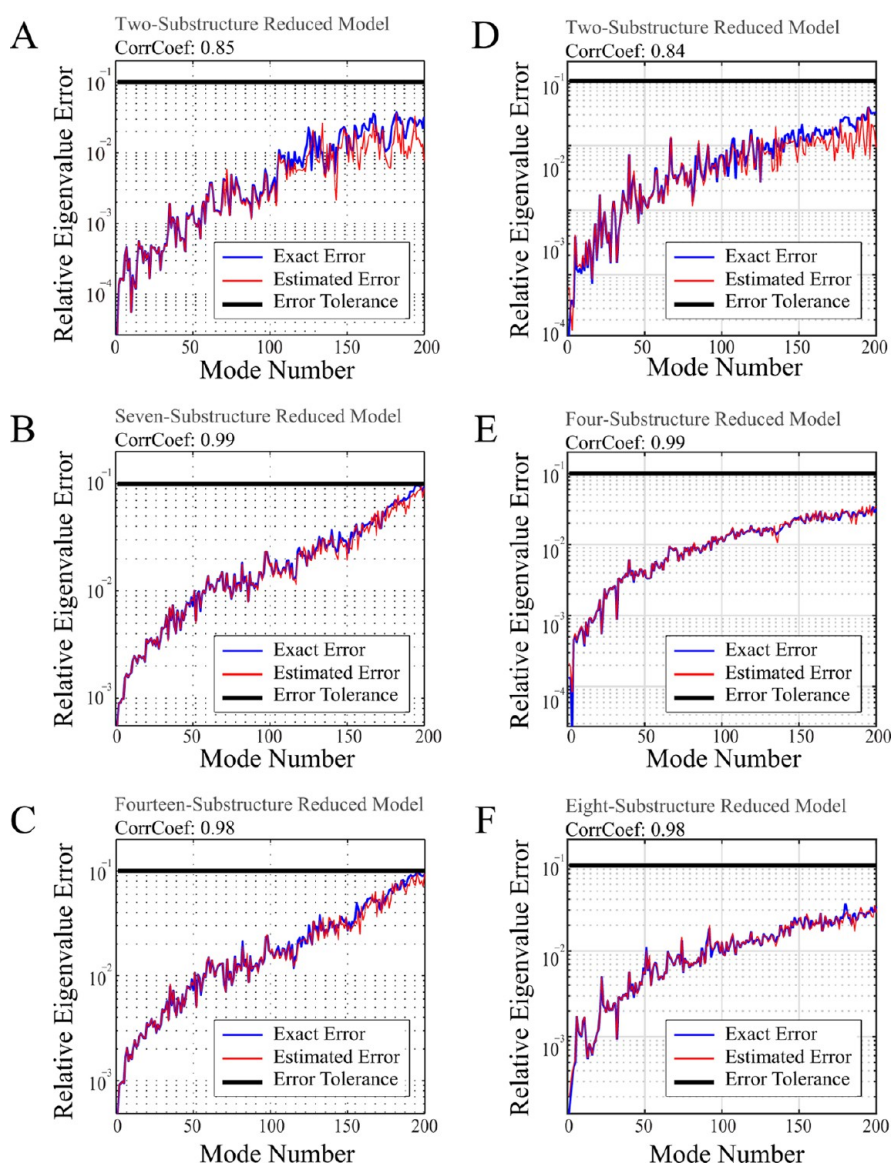
Figure 6. Comparison between the exact and estimated relative eigenvalue errors using the proposed error estimator calculated for the reduced models of (A−C) GroEL and (D−F) ribosome. CorrCoef denotes the correlation coefficient between the exact and estimated relative eigenvalue error curves.

eight arbitrary partitions are constructed for each structure using a freely available mesh-partitioning program, METIS, version 5.1.0[43] (http://glaros.dtc.umn.edu/gkhome/metis/metis/overview/).

Reduced models are constructed for all of the substructural FE models with the aim of providing the 200 lowest normal modes accurately within the error tolerance of 10%. In general, the DOF reduction rate decreases with the number of substructures due to the increase of boundary nodes (Figure S3). Reduced models of GroEL use 1.6% (the two-substructure model), 9.8% (the seven-substructure model), and 10.4% (the 14-substructure model) of DOFs of the original model, whereas 1.5% (the two-substructure model), 41.1% (the four-substructure model), and 43.6% (the eight-substructure model) of DOFs remain in ribosome's reduced models (Figure 5). Relatively large numbers of interface DOFs in the ribosome's reduced models are due to highly complex RNA conformations that lead to widely spread interface regions. Note that interface reduction techniques[44−47] may be employed to reduce the interface DOFs without compromising solution accuracy, although this is beyond the scope of this study.

3.2. Performance of the Error Estimator. Using a sufficiently large, but not too large, number of modes for each substructure is important to obtain an accurate solution efficiently. As described earlier, we develop an automated procedure to achieve a target precision of a normal-mode solution in an iterative manner using an accurate error estimator. We evaluate the performance of the error estimator used in this study by comparing the estimated relative errors of eigenvalues (η_i) with the exact ones (ξ_i) for the results of GroEL and ribosome.

For GroEL and ribosome, the estimated relative eigenvalue errors are surprisingly well-matched with the exact ones over the frequency range with correlation coefficients greater than 0.8 (Figure 6). In particular, estimated and exact error curves are nearly identical in the low-frequency range. This property renders the proposed estimator useful when a small set of lowest normal modes is of interest and importance, which is the case in many applications, including calculating thermal fluctuations in equilibrium, predicting conformational transition pathways, and
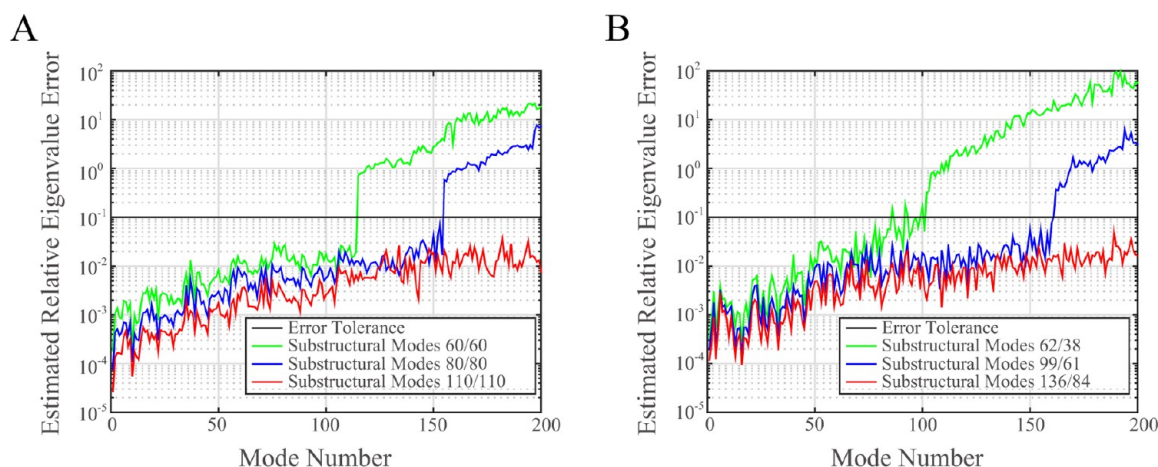
A

B

Figure 7. Convergence of eigensolutions for the two-substructure reduced model of (A) GroEL and (B) ribosome. The number of dominant substructural modes increases incrementally until the estimated relative eigenvalue errors become smaller than the error tolerance.
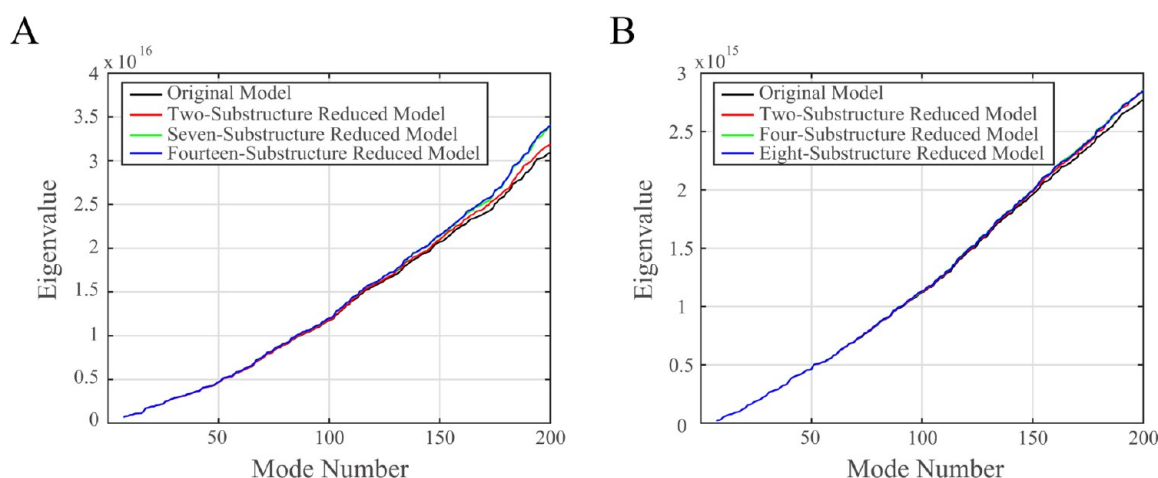
A

B

Figure 8. Lowest 200 eigenvalues excluding rigid-body modes obtained using the original and reduced FE models for (A) GroEL and (B) ribosome.

fitting a high-resolution crystal structure flexibly into a lower-resolution electron microscopy structure.

Accurate error estimation without knowing or calculating the eigenvalues of the original, unreduced protein model enables us to develop an automated procedure that iteratively determines a sufficient number of modes for each substructure, checks the solution accuracy of the reduced models, and performs analysis without manual intervention until the solution converges within the error tolerance. Usually, eigenvalue errors jump at a certain mode number when insufficient numbers of modes are used for reduced models, after which point the solution becomes unreliable (Figure 7). The proposed procedure systematically increases the number of substructural modes until the estimated errors are below the desired error tolerance within the range of eigenvalues of interest. For example, 110 and 110 substructural modes are necessary for the two-substructure GroEL model to obtain a solution satisfying the error tolerance of 0.1, whereas 136 and 84 substructural modes are required for the ribosome model. It is noteworthy that the low-frequency eigenvalues are still accurate even when significantly smaller numbers of modes are used, suggesting that we may use a higher error tolerance in practice because these low-frequency modes are the primary contributors.

Results for the other structures show that the estimated eigenvalue errors are not dependent on the molecular weight, the

molecular shape, or the number of substructures, which demonstrates the applicability of the proposed error estimation scheme to any protein structure (Figures S1 and S2).

3.3. Eigensolutions. As already noted, eigenvalues and eigenvectors are computed for each reduced model so that the maximum relative error in eigenvalues does not exceed 0.1. Eigenvalues increase almost linearly at low frequencies corresponding to non-rigid-body normal modes (Figure 8). Eigenvalues of the reduced models are almost identical to those of the original model at the lowest normal modes below mode 120, and they begin to deviate from them slowly thereafter. The maximum relative eigenvalue errors of the two-substructure models for GroEL and ribosome are 0.036 and 0.040, respectively, which are lower than the predefined error tolerance, set to 0.1 here.

The accuracy of the eigenvalues computed using the reduced models indicates that the calculated eigenvectors are accurate as well regardless of the level of model reduction. The reduced models can reproduce biologically important functional motions of proteins as the original, unreduced models predict. For example, GroEL together with GroES shows a highly dynamic reaction cycle to assist protein folding where significant conformational changes are involved in each functional step. Upon ATP binding, GroEL exhibits a large structural change, particularly in the apical domain required to bring in a partially
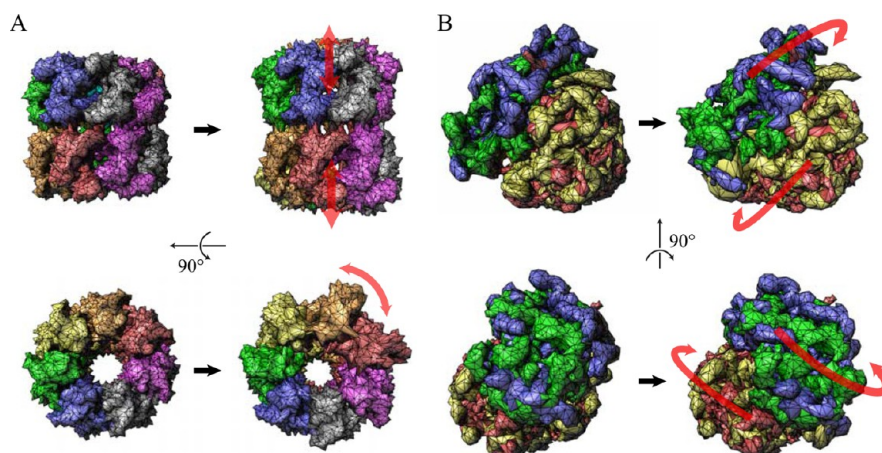
Figure 9. Representative normal modes of (A) GroEL (mode 10 obtained using the 14-substructure model) and (B) ribosome (mode 2 obtained using the four-substructure model).
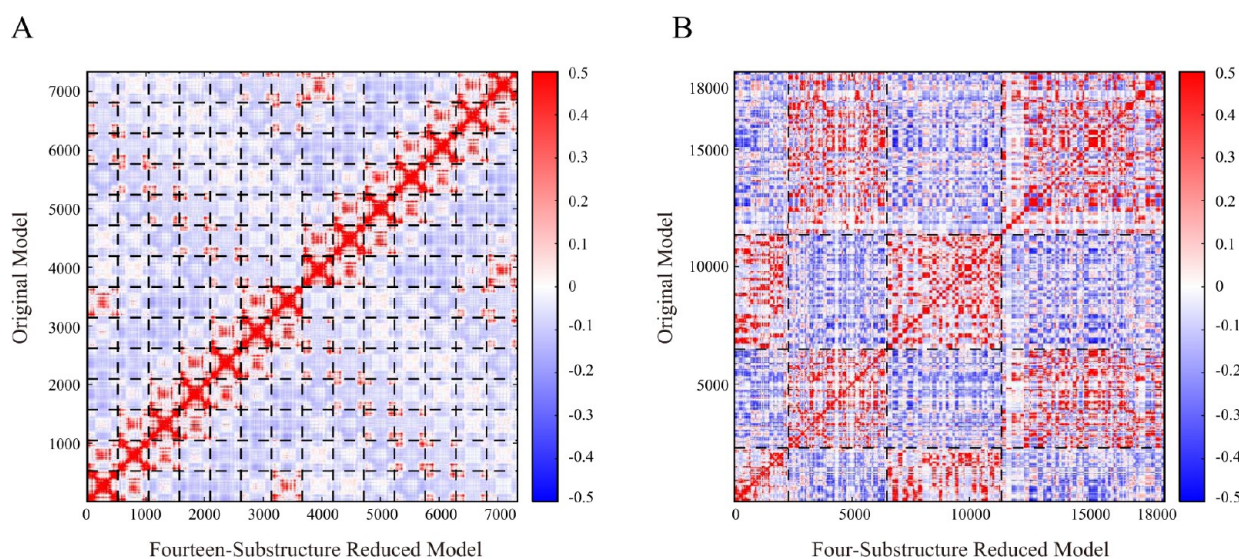

Figure 10. Cross-correlation maps of the original FE model (upper-left triangle) and the reduced FE model (lower-right triangle) of (A) GroEL and (B) ribosome. Each axis represents residue numbers.

folded peptide chain and bind GroES.[48,49] This functional movement appears in the tenth mode of the original model that is also well-reproduced using all of the reduced models (Figures 9A and S4A). For ribosomes, the ratchet-like rotation of the small 40S subunit relative to the large 60S subunit is a well-known functional motion important to translocation of tRNAs.[50−52] The reduced models as well as the original model predict this ratchet-like movement well in their second mode (Figures 9B and S4B).

Next, we calculate the cross-correlation maps to further investigate the accuracy of eigensolutions obtained using the reduced models. The cross-correlation map contains the correlation coefficients between thermal fluctuations of residues measured at α carbon positions. The correlation coefficient between residues i and j is given as $C_{ij} = \langle \Delta \mathbf{r}_i^T \Delta \mathbf{r}_j \rangle / (\langle \Delta \mathbf{r}_i^T \Delta \mathbf{r}_i \rangle \langle \Delta \mathbf{r}_j^T \Delta \mathbf{r}_j \rangle)^{1/2}$. $\Delta \mathbf{r}_i$ and $\Delta \mathbf{r}_j$ represent the fluctuation vectors from the mean α carbon positions and $\langle \Delta \mathbf{r}_i^T \Delta \mathbf{r}_j \rangle = k_B T \sum_k (\boldsymbol{\varphi}_{ik}^T \boldsymbol{\varphi}_{jk} / \lambda_k)$, where $\boldsymbol{\varphi}_{ik}$ and $\boldsymbol{\varphi}_{jk}$ are the eigenvectors of residues i and j, respectively, corresponding to mode k, k_B denotes the Boltzmann constant, and T is the temperature, set to be 300 K here.

In the cross-correlation maps of GroEL and ribosome (Figure 10), the upper-left triangle represents the cross-correlation values

obtained from the eigenvectors of the original structure, whereas the lower-right triangle is used to store those of the reduced models. Correlations between residues in their dynamic motion obtained using the reduced models are not distinguishable from those obtained using the original model for both GroEL and ribosome (Figures 10 and S5). We can observe, in the map of GroEL, a highly correlated group of 14 clusters on the diagonal, each of which corresponds to a GroEL subunit (Figure 10A). This result indicates that the conformational dynamics of GroEL can be well-described as relative motions between the subunits about their interfaces. In addition, these clusters show the highest positive correlations with their nearest neighbors in the same ring, whereas the highest negative correlations are shown between the subunits in the opposite rings, implying positive cooperativity within a ring and negative cooperativity between rings.[53,54] Similarly, a highly correlated group of four clusters is observed in the map of ribosome (Figure 10B) corresponding to ribosomal proteins and RNAs in the large (60S) and small (40S) subunits. Clusters in the same subunit are positively correlated, whereas those in the different subunits are negatively correlated, as expected from the ratchet-like rotation between the subunits.
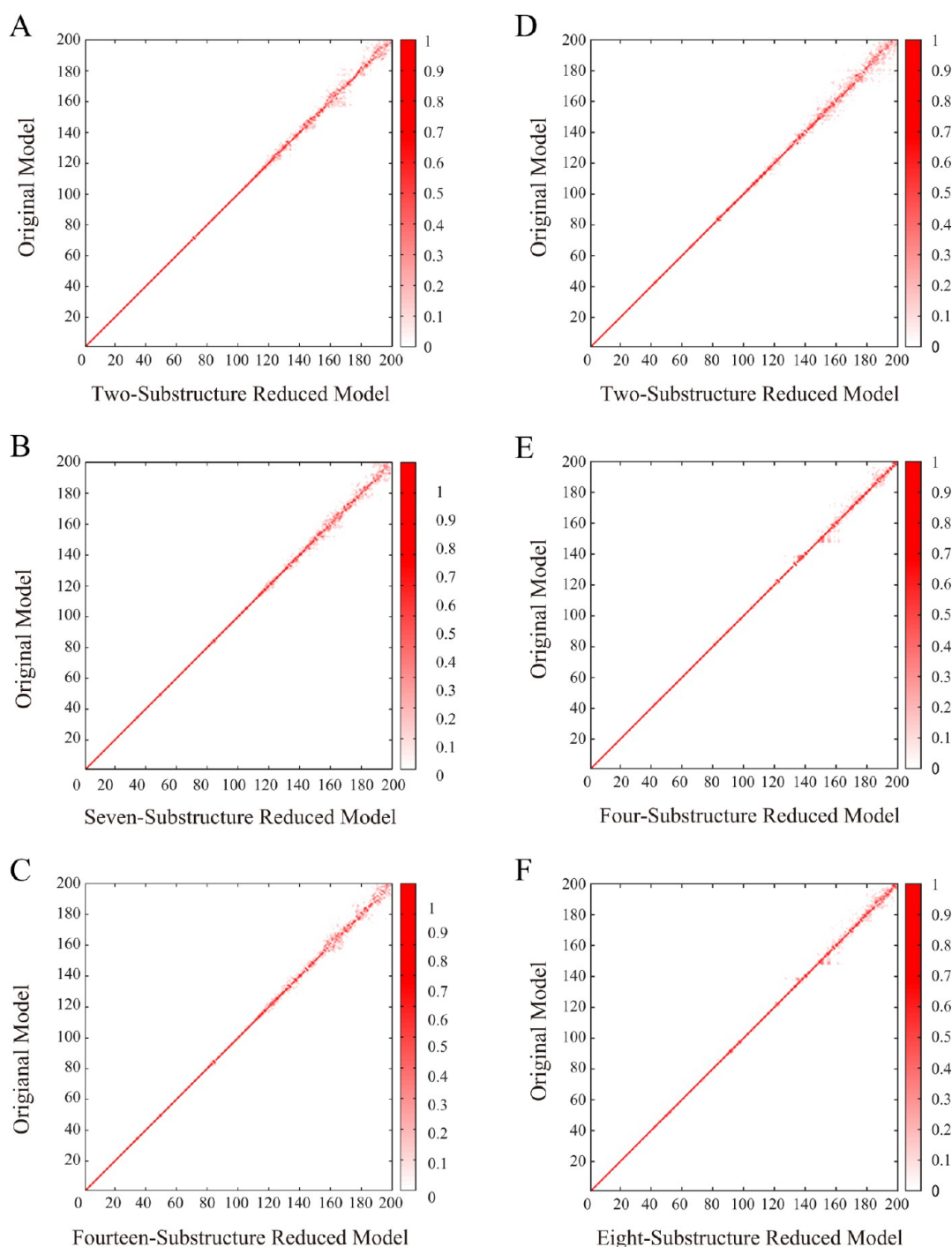
Figure 11. Mode overlap between the eigenvectors obtained using the original and reduced FE models of (A−C) GroEL and (D−F) ribosome. Each axis represents mode numbers.

The high accuracy in the cross-correlation maps obtained using the reduced models indicates that the lowest normal modes, dominant contributors to dynamic correlations, are predicted precisely. This is confirmed by mode overlap for both the GroEL and ribosome models regardless of the number of substructures used in the analysis. Mode overlap between the ith eigenvector of the original model (φ_i) and the jth eigenvector of the reduced model ($\bar{\varphi}_j$) is defined as $P_{i,j} = \varphi_i \cdot \bar{\varphi}_j / |\varphi_i||\bar{\varphi}_j|$. In a low-frequency range, the nonzero mode overlap values appear diagonally (Figure 11). This means that the eigenvectors obtained using the reduced models are almost identical to

those obtained using the original model. Slight spreads in the mode overlap matrices are observed only at the high-frequency modes and increase a little with the number of substructures.

It is not surprising that the derived properties from normal-mode solutions can be accurately predicted as well because we can calculate the eigenvalues and eigenvectors with high precision, which is also tunable. For example, the equilibrium thermal fluctuation is such a property that provides fundamental insight into the conformational dynamics and that is usually investigated by calculating the root-mean-square fluctuation (RMSF) amplitudes at the residue level, given as $(\langle \Delta \mathbf{r}_i^T \Delta \mathbf{r}_i \rangle)^{1/2} =$
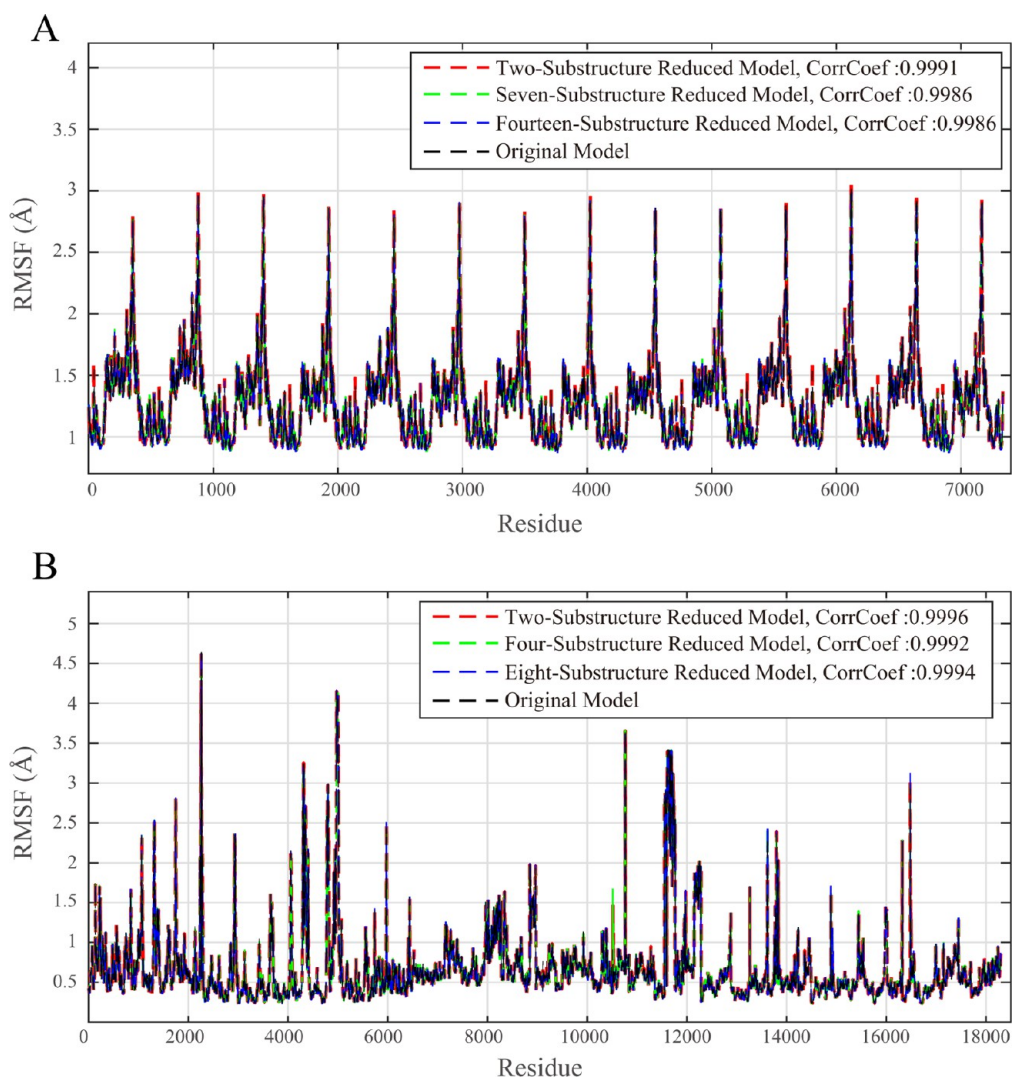
J

Figure 12. Comparison of RMSF profiles computed using the original and reduced FE models for (A) GroEL and (B) ribosome. CorrCoef denotes the correlation coefficient between RMSF profiles obtained using the original and reduced models.
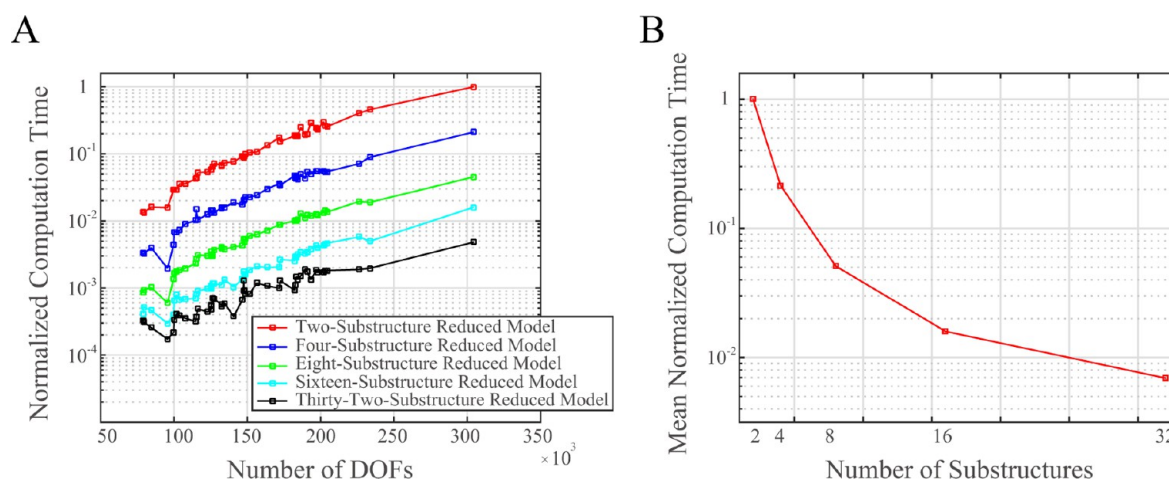

Figure 13. Normalized computation time to calculate 200 lowest normal modes using the reduced FE models as a function of the number of (A) DOFs and (B) substructures.

$(k_B T \sum_k (\varphi_{ik}^T \varphi_{ik})/\lambda_k)^{1/2}$ for residue i. Here, we calculate RMSFs using 200 lowest normal modes at the α carbon positions of residues for both GroEL and ribosome. As expected, nearly

identical RMSF profiles are obtained using all of the reduced models with correlation coefficients greater than 0.99 (Figure 12). It is worthwhile to mention that FE models provide

eigensolutions that are comparable to those obtained using more commonly used ENMs.[17] To illustrate, RMSFs computed using the FE models for protein structures considered in this study show high correlations (>0.8) with those calculated using Gaussian network and anisotropy network models (Figures S6–S8).

Naturally the accuracy of eigensolutions obtained using the reduced model decreases with the level of error tolerance that we set. We tested error tolerances of 20, 30, and 50% in addition to the default 10% and observed deterioration in the computed eigenvalues and eigenvectors with these increased tolerances, as expected. In particular, increasingly wider spreads at higher frequencies in the mode overlap matrix appear, suggesting that the predicted normal modes at higher frequencies are becoming less accurate with the elevation of the tolerance (Figures S9 and S10). Nonetheless, the low-frequency normal modes are almost insensitive to the change in the tolerance. As a result, RMSF profiles obtained using various error tolerance levels remain almost identical to one another (Figure S11). Therefore, significantly reduced substructural models might be used in practice as far as low-frequency normal modes of the entire structure are concerned.

3.4. Computational Efficiency. Here, we investigate the proposed method in terms of computation time. We construct two-, four-, eight-, 16-, and 32-substructure reduced models with random partitions for every protein structure in this study and measure the computation time to calculate 200 lowest eigenvalues and eigenvectors with N_d = 200. While the computation time obviously increases with the number of DOFs of the model, the rate of increase is independent of the level of model reduction (Figure 13A). More importantly, the computation time is significantly decreasing with the number of substructures used in the reduced model (Figure 13B). For example, using a 32-substructure reduced model is almost 100 times faster than using a two-substructure model, indicating the efficiency of CMS methods with a so-called divide-and-conquer strategy. Although these results are obtained using randomly partitioned reduced models for convenience, the same results will be obtained even when we use the reduced models partitioned with biologically relevant subunits. Furthermore, it is important to note that the type of substructures (biological subunit or random partition) hardly affects the eigensolutions, as demonstrated for GroEL and ribosome (Figures S14 and S15). Hence, the use of random partitions will be an attractive and desirable option to build a reduced model in practice. For instance, each substructure in the two-substructure reduced models of GroEL and ribosome can be further divided into random partitions with additional but negligibly small meshing efforts, which will reduce the computation time significantly.

Note that, in this work, we use the CB method because it is the most popular and highly verified CMS method. However, there exist many other CMS methods[29,30,55−57] as alternatives to the CB method that offer ample opportunities for us to improve the proposed method even further. Nevertheless, this requires the development of an accurate and efficient error estimator corresponding to the alternative CMS method to be used, which is an essential prerequisite for automated procedures.

4. CONCLUSIONS

We present our effort toward modular analysis of supramolecular protein assemblies by developing an unsupervised model reduction procedure for FE-based protein models. The dynamics of each constituent substructure is described only using a small number of dominant vibrational modes and the boundary DOFs shared by neighboring substructures, enabled by employing the CB method and a powerful estimator of eigenvalue errors. Results for a comprehensive set of structures demonstrate the excellent performance of the proposed method with tunable accuracy, which is also applicable to any other modeling approach, such as ENM. Furthermore, our method is expected to be useful in many other problems, including, for example, protein−protein interactions, where individual proteins can be reduced to the interacting boundary models, protein−solvent interactions, where solvent-excluded protein surfaces may include the effect of a surrounding water box via similar model reduction, and analysis of viral capsids, where their unique symmetries can be further utilized.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00329.

List of analyzed structures (Table S1); normalized principal components (Table S2); RMSD between the exact relative eigenvalue errors and the estimated ones (Figure S1); diversity in the three-dimensional shape of the analyzed structures (Figure S2); number of DOFs in the original and reduced FE models (Figure S3); residue-level overlap between the eigenvectors obtained using the original FE model and the reduced FE models for GroEL and ribosome (Figure S4); cross-correlation maps of the original FE model and the reduced FE models for GroEL and ribosome (Figure S5); comparison of RMSF profiles computed using the original FE model, Gaussian network model, and anisotropic network model for GroEL (Figure S6) and ribosome (Figure S7); correlation coefficients between RMSFs computed using FE models and GNM/ANM for the remaining 48 structures except GroEL and ribosome (Figure S8); mode overlap between the eigenvectors obtained using the original FE model and the seven-substructure reduced FE model at various error tolerance levels calculated for GroEL (Figure S9) and ribosome (Figure S10); comparison of RMSF profiles computed using the original FE model and the reduced FE model at various error tolerance levels for GroEL and ribosome (Figure S11); reduced models and the eigenvalue errors of structures at various molecular weights (Figure S12) and shapes (Figure S13); comparison of RMSF profiles of GroEL (Figure S14) and ribosome (Figure S15) computed using the reduced models with random partitions and with biological subunits (PDF).

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: dnkim@snu.ac.kr.

Author Contributions

⊥J.K. and J.-G.K. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Bahar, I.; Lezon, T. R.; Yang, L.-W.; Eyal, E. *Annu. Rev. Biophys.* **2010**, *39*, 23−42.

(2) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646−652.

(3) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. *Science* **2010**, *330*, 341−346.

(4) Cui, Q.; Bahar, I. *Normal Mode Analysis: Theory and Applications to Biological Andchemical Systems*; Chapman and Hall/CRC: Boca Raton, FL, 2005.

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

(6) Janežič, D.; Venable, R. M.; Brooks, B. R. *J. Comput. Chem.* **1995**, *16*, 1554−1566.

(7) Kim, J.-I.; Na, S.; Eom, K. *J. Chem. Theory Comput.* **2009**, *5*, 1931−1939.

(8) Levitt, M.; Sander, C.; Stern, P. S. *J. Mol. Biol.* **1985**, *181*, 423−447.

(9) Ueda, Y.; Taketomi, H.; Gō, N. *Biopolymers* **1978**, *17*, 1531−1548.

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

(11) Tozzini, V. *Curr. Opin. Struct. Biol.* **2005**, *15*, 144−150.

(12) Tirion, M. M. *Phys. Rev. Lett.* **1996**, *77*, 1905.

(13) Bahar, I.; Atilgan, A. R.; Erman, B. *Folding Des.* **1997**, *2*, 173−181.

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

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

(16) Nogales-Cadenas, R.; Jonic, S.; Tama, F.; Arteni, A. A.; Tabas-Madrid, D.; Vázquez, M.; Pascual-Montano, A.; Sorzano, C. O. S. *Nucleic Acids Res.* **2013**, *41*, W363−367.

(17) Bathe, M. *Proteins: Struct., Funct., Genet.* **2008**, *70*, 1595−1609.

(18) Kim, D.-N.; Nguyen, C.-T.; Bathe, M. *J. Struct. Biol.* **2011**, *173*, 261−270.

(19) Kim, M. K.; Chirikjian, G. S.; Jernigan, R. L. *J. Mol. Graphics Modell.* **2002**, *21*, 151−160.

(20) Wang, Y.; Rader, A. J.; Bahar, I.; Jernigan, R. L. *J. Struct. Biol.* **2004**, *147*, 302−314.

(21) Rader, A.; Vlad, D. H.; Bahar, I. *Structure* **2005**, *13*, 413−421.

(22) Gibbons, M. M.; Klug, W. S. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2007**, *75*, 031901.

(23) Sakalli, I.; Schöberl, J.; Knapp, E. W. *J. Chem. Theory Comput.* **2014**, *10*, 5095−5112.

(24) Sedeh, R. S. *Contributions to the analysis of proteins*. Ph.D. Thesis, Massachusetts Institute of Technology, 2011.

(25) Kim, D.-N.; Altschuler, J.; Strong, C.; McGill, G.; Bathe, M. *Nucleic Acids Res.* **2011**, *39*, D451−455.

(26) Hurty, W. C. *AIAA J.* **1965**, *3*, 678−685.

(27) Bampton, M.; Craig, R. R., Jr. *AIAA J.* **1968**, *6*, 1313−1319.

(28) MacNeal, R. H. *Comput. Struct.* **1971**, *1*, 581−601.

(29) Kim, J.-G.; Boo, S.-H.; Lee, P.-S. *Comput. Method Appl. M.* **2015**, *287*, 90−111.

(30) Kim, J.-G.; Lee, P.-S. *Int. J. Numer. Meth. Engng.* **2015**, *103*, 79−93.

(31) Eom, K.; Ahn, J.; Baek, S.; Kim, J.; Na, S. *CMC Comput. Mater. Con.* **2007**, *6*, 35−42.

(32) Kim, J.-G.; Lee, K.-H.; Lee, P.-S. *Comput. Struct.* **2014**, *139*, 54−64.

(33) Kim, J.-G.; Lee, P.-S. *Comput. Method Appl. M.* **2014**, *278*, 1−19.

(34) Sanner, M. F.; Olson, A. J.; Spehner, J. C. *Biopolymers* **1996**, *38*, 305−320.

(35) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605−1612.

(36) Cignoni, P.; Callieri, M.; Corsini, M.; Dellepiane, M.; Ganovelli, F.; Ranzuglia, G. Meshlab: an open-source mesh processing tool. In *Eurographics Italian Chapter Conference*: Salereno, Italy, 2008; Scarano, V., Chiara, R. D., Erra, U. , Eds.

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

(38) Elssel, K.; Voss, H. *SIAM J. Matrix. Anal. Appl.* **2006**, *28*, 386−397.

(39) Bartolucci, C.; Lamba, D.; Grazulis, S.; Manakova, E.; Heumann, H. *J. Mol. Biol.* **2005**, *354*, 940−951.

(40) Hashem, Y.; des Georges, A.; Fu, J.; Buss, S. N.; Jossinet, F.; Jobe, A.; Zhang, Q.; Liao, H. Y.; Grassucci, R. A.; Bajaj, C.; Westhof, E.; Madison-Antenucci, S.; Frank, J. *Nature* **2013**, *494*, 385−389.

(41) Heckbert, P. S.; Garland, M. *Computational Geometry* **1999**, *14*, 49−65.

(42) Fischer, H.; Polikarpov, I.; Craievich, A. F. *Protein Sci.* **2004**, *13*, 2825−2828.

(43) Karypis, G.; Kumar, V. *SIAM J. Sci. Comput.* **1998**, *20*, 359−392.

(44) Castanier, M. P.; Tan, Y.-C.; Pierre, C. *AIAA J.* **2001**, *39*, 1182−1187.

(45) Junge, M.; Brunner, D.; Becker, J.; Gaul, L. *Int. J. Numer. Meth. Engng.* **2009**, *77*, 1731−1752.

(46) Bourquin, F.; d'Hennezel, F. *Comput. Method Appl. M.* **1992**, *97*, 49−76.

(47) Markovic, D.; Park, K. C.; Ibrahimbegovic, A. *Int. J. Numer. Meth. Engng.* **2007**, *70*, 163−180.

(48) Ranson, N. A.; Farr, G. W.; Roseman, A. M.; Gowen, B.; Fenton, W. A.; Horwich, A. L.; Saibil, H. R. *Cell* **2001**, *107*, 869−879.

(49) Skjaerven, L.; Martinez, A.; Reuter, N. *Proteins: Struct., Funct., Genet.* **2011**, *79*, 232−243.

(50) Tama, F.; Valle, M.; Frank, J.; Brooks, C. L. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 9319−9323.

(51) Frank, J.; Agrawal, R. K. *Nature* **2000**, *406*, 318−322.

(52) Agirrezabala, X.; Lei, J.; Brunelle, J. L.; Ortiz-Meoz, R. F.; Green, R.; Frank, J. *Mol. Cell* **2008**, *32*, 190−197.

(53) Ma, J.; Karplus, M. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 8502−8507.

(54) Cui, Q.; Karplus, M. *Protein Sci.* **2008**, *17*, 1295−1307.

(55) Bennighof, J. K.; Lehoucq, R. B. *SIAM J. Sci. Comput.* **2004**, *25*, 2084−2106.

(56) Rixen, D. J. *J. Comput. Appl. Math.* **2004**, *168*, 383−391.

(57) Park, K.; Park, Y. H. *AIAA J.* **2004**, *42*, 1236−1245.