 CMC, vol.9, no.2, pp.137-151, 2009

A 3-D Coarser-Grained Computational Model for Simulating Large Protein Dynamics

Jae-In Kim[1], Hyoseon Jang[2], Jeong-Hee Ahn[3]
Kilho Eom[4] and Sungsoo Na[5]

Abstract: Protein dynamics is essential for gaining insight into biological functions of proteins. Although protein dynamics is well delineated by molecular model, the molecular model is computationally prohibited for simulating large protein structures. In this work, we provide the three-dimensional coarser-grained anisotropic model (CGAM), which is based on model reduction applicable to large protein structures. It is shown that CGAM achieves the fast computation on low-frequency modes, quantitatively comparable to original structural model such as elastic network model (ENM). This indicates that the CGAM by model reduction method enable us to understand the functional motion of large proteins with remarkable computational efficiency.

Keywords: Protein Dynamics, Elastic Network Model, Coarser-Grained Anisotropic Model, Low-Frequency Mode

1 Introduction

Protein dynamics, related to biological function, has been well understood by atomistic model based on molecular dynamics (MD) simulations and/or normal mode analysis (NMA) [Brooks et al (1988); Karplus & Petsko (1990); Cui & Bahar (2005)].

[1] Graduate student, Dept. of Mechanical Eng., Korea University, Anam-dong, Seongbuk-gu, Seoul 136-713, Korea (email: jay414@korea.ac.kr)
[2] Graduate student, Dept. of Mechanical Eng., Korea University, Anam-dong, Seongbuk-gu, Seoul 136-713, Korea (email: hsjang@korea.ac.kr)
[3] Graduate student, Dept. of Mechanical Eng., Korea University, Anam-dong, Seongbuk-gu, Seoul 136-713, Korea (email: shibuya@korea.ac.kr)
[4] Research Professor, Corresponding author, Dept. of Mechanical Eng., Korea University, Anam-dong, Seongbuk-gu, Seoul 136-713, Korea (email:Kilhoeom@gmail.com)
[5] Professor, Corresponding author, Dept. of Mechanical Eng., Korea University, Anam-dong, Seongbuk-gu, Seoul 136-713, Korea (email:nass@korea.ac.kr)

However, atomistic model has the restrictions for gaining insight into large protein dynamics in the current computing capabilities since the current computational state-of-art enables the analysis of protein dynamics in the time-scale of nano-second regime much less than the time-scale relevant to protein's biological function [Cui & Bahar (2005); Philips et al (2002)]. As a consequence, the coarse-grained model has drawn the attention for analyzing the dynamics of large biological and/or chemical structures [Philips et al (2002)]. Especially, for simulating the protein dynamics, elastic network model (ENM) has taken much attention because of its simplicity as well as its capability to capture the dynamic behavior of proteins qualitatively and/or quantitatively comparable to original atomistic model and/or experiments [Tirion (1996); Atilgan et al (2001)]. The ENM describes the protein structure as harmonic spring network system such that the residues represented only by alpha carbons within the neighborhood (cut-off radius) are connected by elastic harmonic springs with identical spring constants [Atilgan et al (2001)]. Such a simple model (ENM) has allowed for depicting the collective dynamics [Van & Cui (2006)], the fluctuation dynamics [Van & Cui (2006)], and the motion along the low-frequency normal modes related to conformational change for various proteins [Tama & Sanejouand (2001); Xu et al (2003); Tobi & Bahar (2005)], which shed light on the energy landscape for conformational transitions [Miyashita et al (2003); Maragakis & Karplus (2005)].

Although the protein structures are represented simply by harmonic spring network in ENM, it is quite remarkable in its capability to capture the dynamic behavior of proteins qualitatively comparable to atomistic model. The success of ENM to capture the protein dynamics may be ascribed to the fact that the collective dynamics of proteins is well described by harmonic and/or quasi-harmonic dynamics. In this connection, Brooks and Karplus showed that the harmonic dynamics is held for describing the collective dynamics of proteins [Brooks & Karplus (1983)]. In a study by Amadei et al., it is shown that the collective dynamics of proteins can be well represented by essential dynamics (quasi-harmonic dynamics) based on the MD trajectories [Amadei et al (1993)]. Moreover, Case and Teeter reported that the details of the potential field prescribed to atomistic structure of protein do not play any role in the collective dynamics of proteins at all [Teeter & Case (1990)]. Recently, Ma suggested that the perturbations of stiffness matrix for protein structure does not induce the change of protein dynamics behavior as long as the topology of protein structure is maintained, implying that protein native topology is responsible for protein dynamics [Lu & Ma (2005)].

Even though ENM has been capable of analysis on protein dynamics such as fluctuation dynamics (collective dynamics) and/or conformational transitions, the large protein dynamics may not be computationally effectively approached by ENM,

simply because of large degree of freedom for large protein structure. Recently, Drouker et al. reported that conformational fluctuation of large protein structures can be delineated by coarser-grained structural model [Doruker et al (2002); Kurkcuoglu et al (2004); Kurkcuoglu et al (2005)]. Further, Chennubholta and Bahar provided the coarser-grained Gaussian network model based on the Markov statistical method [Chennubhotla et al (2006); Chennubhotla & Bahar (2006)]. Moreover, we previously reported the model condensation method, which establishes the coarser-grained Gaussian (one-dimensional) network model for large protein structure [Eom et al (2007)]. It was shown that collective and fluctuation dynamics of large proteins is well described by coarser-grained isotropic Gaussian (one-dimensional) network model. The success of coarser-grained model in predicting the protein's conformational fluctuation is attributed to the fact that the collective dynamic behavior of a domain can be represented by small degrees of freedom. In this sense, such coarser-grained models may be employed for developing the novel multi-scale model for protein structure in such a way that the regions relevant to biological functions (e.g. ligand-binding site) can be described by the refined molecular model, while the rest of regions are represented by coarser-grained models [Ahmed & Gohlke (2006)]. We also presented that the component mode synthesis widely used in engineering was well applied to understand protein dynamics [Kim et al (2008)]. With the proposed scheme, the domain of protein can be considered as a component and the number of component is increased as necessary, and the size of each component may be decreased for fast calculation.

In this article, we developed the coarser-grained three-dimensional network model, which allows us to gain insight into protein dynamics based on the dynamic model condensation in such a way that the slave residues to be eliminated during model condensation are assumed to be in equilibrium state. It is shown that our coarser-grained three-dimensional model describes anisotropic motion of a residue and possesses the low-frequency modes, quantitatively similar to original molecular model and/or experimental data. This indicates that the proposed model may enable the study of large protein dynamics related to protein conformational change based on the low-frequency normal modes. It implies that the insight into protein conformational change can be gained with coarser-grained model for large proteins.

2 Model proteins

We studied the validation of coarser-grained three-dimensional network model to capture the collective (fluctuation) motion based on the following model proteins: citrate synthase, hemoglobin and kinesin. These model proteins are appropriate for studying in a later stage whether the insight into conformational change can be gained by present model, because of their given two equilibrium states (open and

close forms) provided in protein data bank (http://www.pdb.org). Fig. 1 presents the original structure for model proteins.

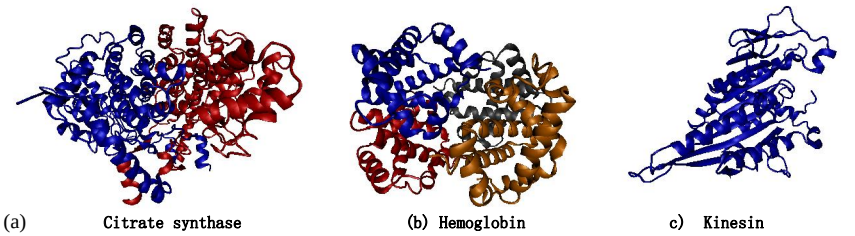


Figure 1: Model proteins

2.1 Elastic Network Model (ENM)

The success of harmonic dynamics for capturing the protein dynamics enabled Tirion to develop the ENM for protein structures [Tirion (1996)]. The ENM regards the protein structure as the harmonic spring network system, in which the neighborhood residues are connected by harmonic springs with identical spring constant [Atilgan et al (2001)]. The potential energy E for ENM is given by

$$E = \frac{1}{2}\sum_{i,j}(u_i - u_j)^* K_{ij}(u_i - u_j) \equiv \frac{1}{2}u^* K u \tag{1}$$

Here u_i is the displacement field for residue i, i.e., $u_i = r_i - r_i^0$ with a position vector r_i for residue i and superscript 0 means the equilibrium state, u is the $3N \times 1$ column vector representing the displacement field for all residues, asterisk indicates the transpose of a vector, K_{ij} is the 3×3 local stiffness matrix for a harmonic spring connecting residues i and j, and $\mathbf{K}$ is the $3N \times 3N$ stiffness matrix for ENM, whose off-diagonal block matrix is given by $-K_{ij}$ and the diagonal block matrix is equal to summation of off-diagonal block matrices.

The protein dynamics can be described by the eigenvalue problem: $Kq = \omega^2 q$, where ω is the natural frequency and q is its corresponding normal mode. Statistical mechanics theory enables one to compute the fluctuation matrix $\mathbf{C}$ based on the normal modes.

$$C \equiv \langle u \otimes u \rangle = \sum_{J=1}^{3N} \frac{k_B T}{\omega_j^2} q_j \otimes q_j \tag{2}$$

where angle bracket < > indicates the ensemble average (time average), k_B is the Boltzmann's constant, T is the absolute temperature, and the subscript j represents

the mode index j. It should be noted that the summation in Eq. (2) excludes the six zero-modes corresponding to rigid body motions. The mean-square fluctuation for residue i is in the form of

$$\left\langle |u_i|^2 \right\rangle = \sum_{k=1}^{3} C_{3(i-1)+k,3(i-1)+k} \tag{3}$$

Consequently, the Debye-Waller factor (B-factor) for residue i is

$$B_i = \frac{8\pi^2}{3} \left\langle |u_i|^2 \right\rangle = \frac{8\pi^2}{3} \sum_{k=1}^{3} C_{3(i-1)+k,3(i-1)+k} \tag{4}$$

The correlated motion between residues i and j is well delineated by cross-correlation L_{ij} given by

$$L_{ij} = \frac{\langle u_i \cdot u_j \rangle}{\sqrt{\left\langle |u_i|^2 \right\rangle \left\langle |u_j|^2 \right\rangle}} = \frac{\sum_{k=1}^{3} C_{3(i-1)+k,3(j-1)+k}}{\sqrt{\left(\sum_{p=1}^{3} C_{3(i-1)+p,3(i-1)+p} \right) \left(\sum_{r=1}^{3} C_{3(j-1)+r,3(j-1)+r} \right)}} \tag{5}$$

2.2 Coarser-Grained Elastic Network Model

As stated earlier, the dynamic behavior of large proteins has been approached by the coarser-grained models, in which several residues are represented by a single nodal point. We denote the nodal points for coarser-grained model as the master residues which are taken in coarser-grained structure, while the rest of residues which should be removed during the model condensation (coarse-graining) are denoted as the slave residues. The coarser-graining scheme to select master and slave residues is well described in Doruker et al (2002) and Eom et al (2007). For the coarse-graining of protein structure, the potential energy E given by Eq. (1) can be represented in the form of

$$E = \frac{1}{2} u^* K u = \frac{1}{2} \begin{bmatrix} u_m^* & u_s^* \end{bmatrix} \begin{bmatrix} K_{mm} & K_{ms} \\ K_{sm} & K_{ss} \end{bmatrix} \begin{bmatrix} u_m \\ u_s \end{bmatrix} \tag{6}$$

where the subscript m and s indicate the indices of master residues and slave residues, respectively, and u_m and u_s are the displacement fields for master residues and slave residues, respectively. During model condensation, the slave residues

are assumed to be equilibrium. That is, the slave residues are assumed to have negligible contribution to protein conformational fluctuation.

$$\frac{\partial E}{\partial u_s} = K_{sm}u_m + K_{ss}u_s = 0 \tag{7}$$

From Eqs. (6) and (7), the potential energy E for a coarser-grained network model for protein structure is given by $E = (1/2)u_m^* \tilde{K} u_m$ where $\tilde{K}$ is the stiffness matrix in the form of

$$\tilde{K} = K_{mm} - K_{ms}K_{ss}^{-1}K_{sm} \tag{8}$$

Here, the dimension of $\tilde{K}$ is $3N_m \times 3N_m$, where N_m is the total number of master residues. For a coarser-grained network model of N_m nodal points, where $N_m = (N/n)$ with $n >> 1$, the computational cost to compute the stiffness matrix $\tilde{K}$ for a coarser-grained network model is proportional to computational expense to estimate the inverse of K_{ss} in order of $9N^3(1-1/n)^3 \sim 9N^3$ for $n >> 1$. This indicates that direct model condensation to compute stiffness matrix for a coarser-graining model consisting of very small number of nodal points is computationally unfavorable because of expensive computation on K_{ss}^{-1}. For computational efficiency, we consider the hierarchical model condensation described as following procedures.

Select the (N/ℓ) master residues and the $N(1-1/\ell)$ slave residues, where $1 < \ell << n$.

Partition the stiffness matrix K in the form of (6).

Compute the stiffness matrix $\tilde{K}$for a coarser-grained model consisting of (N/ℓ) nodal points according to Eq. (8).

Set the coarser-grained structure as the initial structure for the further model condensation, i.e. $N \leftarrow N/\ell$and $K \leftarrow \tilde{K}$.

Repeat the steps (i)-(iv) until one obtains the coarser-grained structure composed of (N/n)nodal points and its corresponding stiffness matrix$\tilde{K}$.

Once the stiffness matrix $\tilde{K}$ is computed for a coarser-grained model, then the fluctuation and/or collective dynamics are can be easily understood from Eqs. (2)-(4).

3 Results

3.1 Fluctuation Dynamics

We consider the thermal fluctuation dynamics of model proteins based on elastic network model (ENM), referred to as original structural model, and coarser-grained anisotropic model (CGAM). As shown in Fig. 2 proteins are modeled as

harmonic spring networks referred to as elastic network model for original structure and coarser-grained network model for coarser-grained structure.

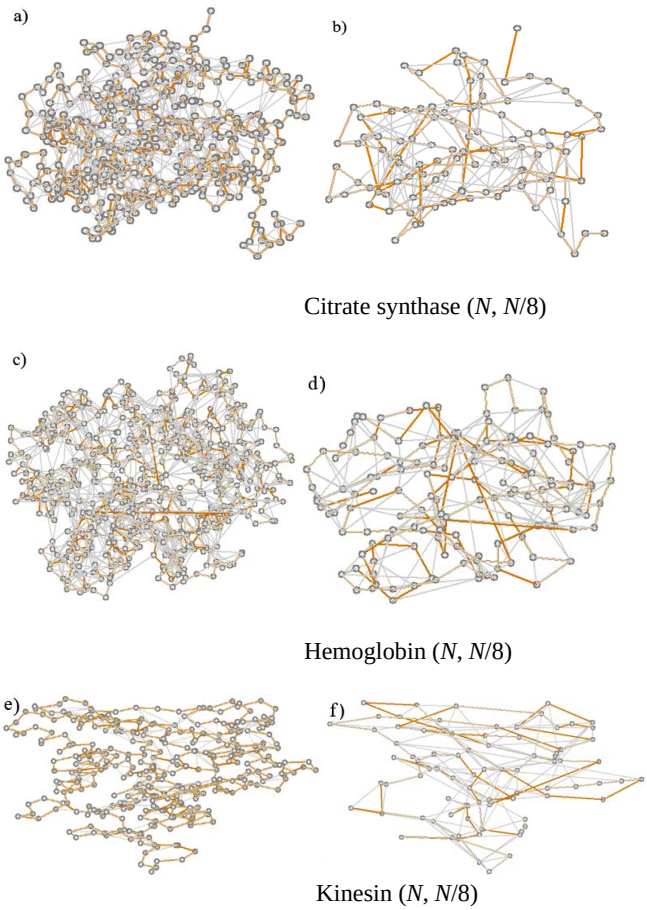
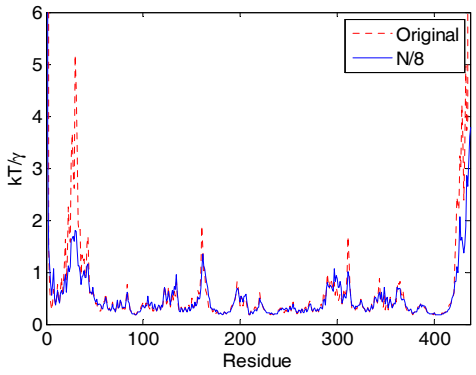


Figure 2: Protein Structures: Original model (left) and coarser-grained model (right)

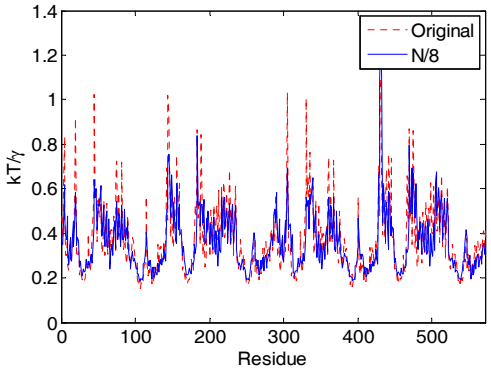
The coarser-grained structure, shown in Fig. 2, for model protein allows the reduction of computational cost, by order of $9n^2$, for calculating the thermal fluctuation such as Debye-Waller factor, where n is the level of coarse-graining such that a coarser-graining model consists of (N/n)nodal points with N denoted as original number of residues. Fig. 3 represents the Debye-Waller factor, denoted as $B_i = (8\pi^2/3)\langle u_i \rangle^2 = (8\pi^2/3)\left[\left\langle u_{i,x}^2 \right\rangle + \left\langle u_{i,y}^2 \right\rangle + \left\langle u_{i,z}^2 \right\rangle\right]$ where $u_{i,x}, u_{i,y}$ and $u_{i,z}$is

the displacement driven by thermal fluctuation for residue i along the x-axis, y-axis, and z-axis, respectively, for model proteins based on ENM and CGAM, where CGAM consists of $N/8$ nodal points. Here N is the total number of residues in original protein structures. It is quite remarkable that anisotropic fluctuation behavior of model proteins is well delineated by CGAM, quantitatively comparable to fluctuation behavior obtained by original structural model (i.e. ENM). The capability of CGAM to capture the anisotropic thermal fluctuation behavior of model proteins may be attributed to the rigidity of protein domains. Specifically, a number of proteins consist of several rigid domains, which can be described by small number of nodal points. This allows several research groups to introduce the various coarser-grained model [Doruker et al (2002); Eom et al (2007)] for gaining insight into thermal fluctuation of proteins. Our CGAM is constructed in the similar spirit to our previous one-dimensional coarser-grained model [Eom et al (2007)], which was extended to 3-dimensional coarser-grained model to capture anisotropic fluctuation motion. As stated earlier, since our coarser-graining scheme is based on the elimination of entropic springs connected to slave residues, which leads to increase in the overall flexibility of protein domains, the rescaling of spring constant is mandatory for maintaining the overall flexibility of protein structure (in Fig. 3, the spring constant is rescaled).

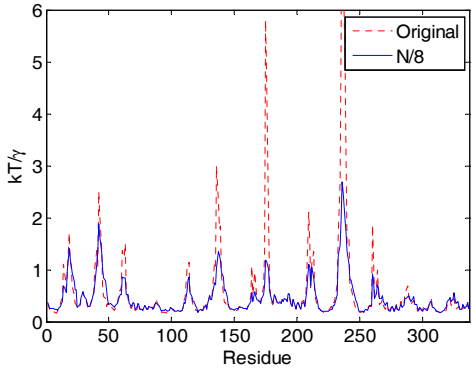
The correlated motion for model proteins is taken into account with both ENM and CGAM. The correlated motion of proteins is generally ascribed to rigid domain, implying that the correlated motion may be described by the coarser-grained structural model (CGAM). Fig. 4 displays the cross-correlation map L_{ij}, representing the correlation of motions between two residues (nodal points) i and j, for model proteins with use of both ENM and CGAM. It is remarkably shown that coarse-graining structure is able to reproduce the correlated motion between sub-domains, quantitatively comparable to that obtained from original structure. This indicates that our CGAM may is the robust model for gaining insight into protein dynamics. The robustness of CGAM may depend on the level of coarse-graining, n. Since the correlated motion of proteins cannot be represented by a couple of dominant low-frequency normal modes[8], the CGAM may be robust for understanding the correlated motion as long as the CGAM is able to provide the sufficient number of low-frequency normal modes, qualitatively similar to those of original structural model (e.g. ENM). In our case, we chose the coarse-graining level n as n=8 for CGAM. As shown in Fig. 4, the quantitative similarity of L_{ij} between ENM and CGAM suggests that our CGAM is very robust enough to provide the meaningful information of correlated motions of model proteins.



a) Citrate synthase

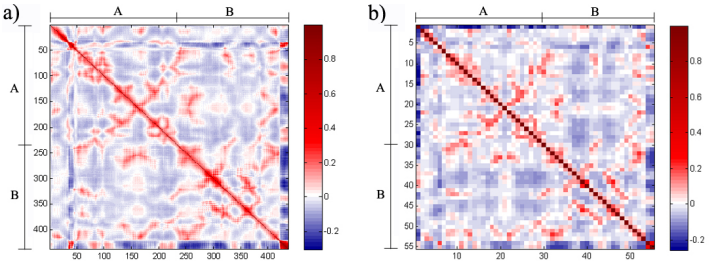


b) Hemoglobin

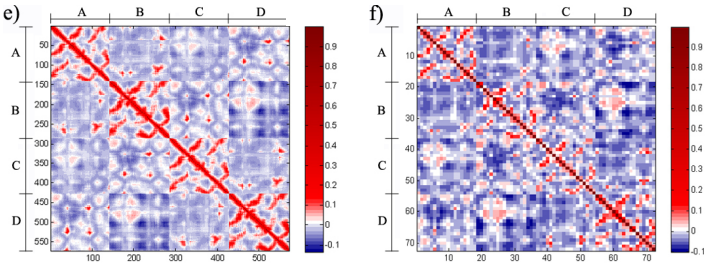


c) Kinesin

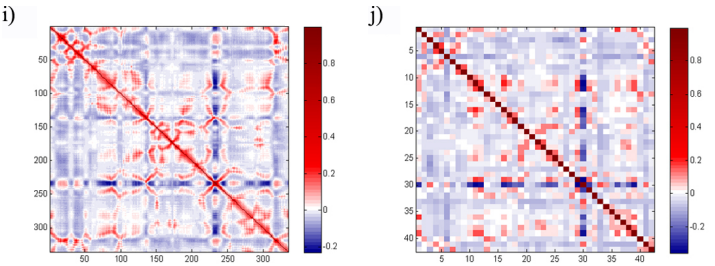
Figure 3: Debye-Waller factors (B-factors) for model proteins such as (a) citrate synthase, (b) hemoglobin, and (c) kinesin.



a) Citrate synthase



b) Hemoglobin



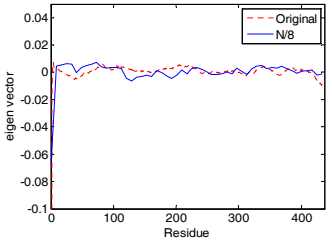
c) Kinesin

Figure 4: Cross-correlation for (i) original structure (left) and (ii) coarse-graining structure (right)

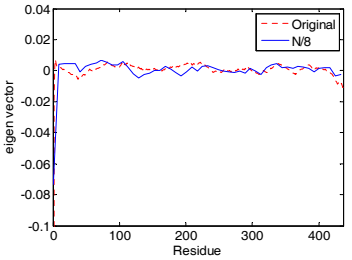
3.2 Low-Frequency modes

Since the low-frequency modes are associated with not only the collective dynamics but also the structural deformation related to biological functions of proteins, the low frequency modes for coarse-grained models for three model proteins are taken into account for validity of our coarse-grained model. Lowest-frequency modes, obtained from original structure as well as coarse-grained structure, in each direction for model proteins are shown in Fig. 5.

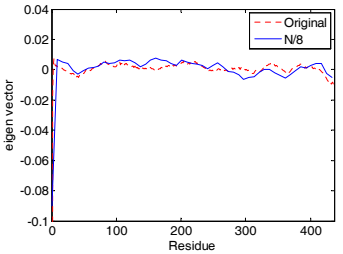
It is remarkable that, the low-frequency modes for the original structure can be also



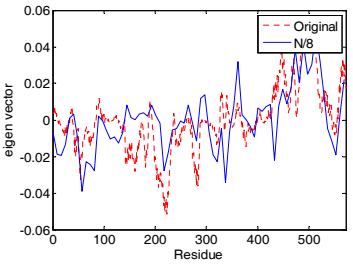
(a) X-component of 1st low frequency eigenvector of citrate synthase



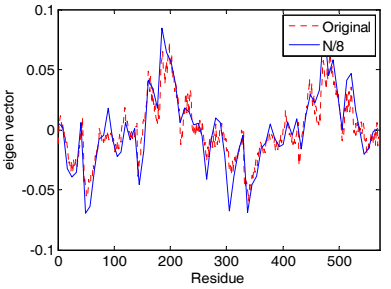
(b) Y-component of 1st low frequency eigenvector of citrate synthase



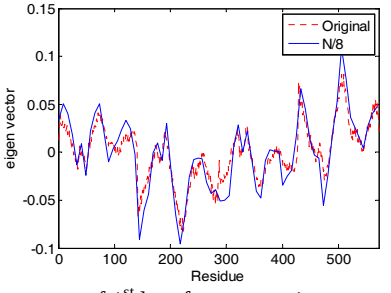
(c) Z-component of 1st low frequency eigenvector of citrate synthase



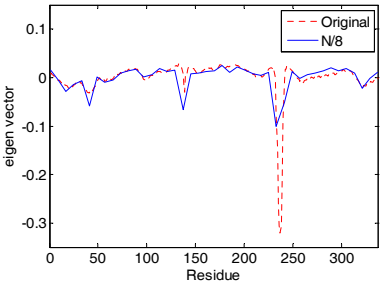
(d) X-component of 1st low frequency eigenvector of hemoglobin



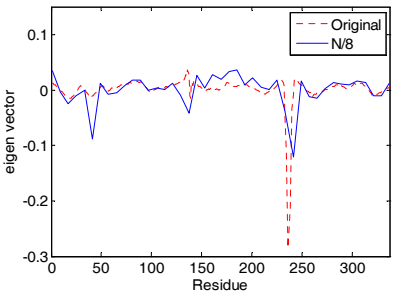
(e) Y-component of 1st low frequency eigenvector of hemoglobin



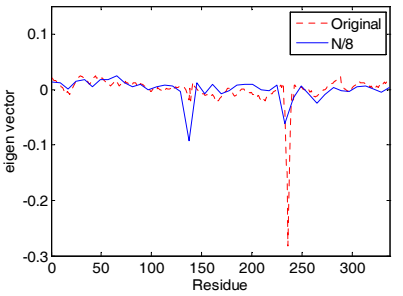
(f) Z-component of 1st low frequency eigenvector of hemoglobin



(g) X-component of 1st low frequency eigenvector of kinesin



(h) Y-component of 1st low frequency eigenvector of kinesin



(i) Z-component of 1st low frequency eigenvector of kinesin

Figure 5: Lowest-frequency normal modes, obtained from original structure as well as coarser-grained structure, in each direction for model proteins such as (a) citrate synthase, (b) hemoglobin, and (c) kinesin.

captured by the coarse-graining structure. This indicates that our proposed model does not affect the characteristics of low-frequency modes for proteins. That is, our coarse-grained model may allow for the fast computation on low-frequency modes related to structural changes of proteins. The results also imply that our CGAM is robust model reduction scheme, which allows for gaining insight into functional low-frequency modes to make a contribution to thermal fluctuation for proteins.

4 Conclusions

Elastic network model has allowed for gaining insight into protein dynamics related to biological functions of proteins. However, such a model exhibits the computational limitations in computing the low-frequency modes, and thermal fluctuations associated with protein structural changes for large protein structures. In present work, we provided the coarser-graining anisotropic model applicable to

large protein structures. It was shown that coarse-grained models by model reduction scheme enable us to understand protein dynamics with remarkable computational efficiency.

Acknowledgement: This work was supported by Basic research program of the Korea Science & Engineering Foundation (No. R01-2007-000-10497-0; R11-2007-028-03002)

References

Amadei, A.; Lissen, A.; Berendesen, H. (1993): Essential dynamics of proteins, *Proteins: Structure, Function, and Genetics* **17**(4): pp. 412-425.

Atilgan, A.R., et al. (2001): Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model, *Biophys. J*. **80**(1) pp. 505-515.

Bahar, I.; Cui, Q. (2005): Normal Mode Analysis: Theory and Applications to Biological And Chemical Systems, *Chapman & Hall/CRC Mathematical & Computational Biology.*

Brooks, C.L.; Karplus, M.; Pettit, B. M. (1988): Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamic, *Chem. Phys*., **71** p. 12-49.

Brooks, B.; Karplus, M. (1983): Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor, *Proceedings of the National Academy of Sciences of the United States of America* **80**(21) pp. 6571-6575.

Chennubhotla, C.; Bahar, I. (2006): Markov Methods for Hierarchical Coarse-Graining of Large Protein Dynamics, *In Lecture Notes in Computer Science* **3909** pp. 379-393.

Chennubhotla, C.; Bahar, I. (2006): Markov propagation of allosteric effects in biomolecular systems: application to GroEL-GroES, *Mol Syst Biol*. **2**.

Doruker, P.; Jernigan, R.; Bahar, I. (2002): Dynamics of large proteins through hierarchical levels of coarse-grained structures, *Journal of Computational Chemistry* **23**(1): pp. 119-127.

Elber, R. (2005): Long-timescale simulation method,. *Current Opinion in Structural Biology* **15**(2) pp. 151-156.

Eom, K.; Baek, S.; Ahn, J.; Na, S. (2007): Robust Reduction Method for Biomoluecues Modeling, *CMC: Computers, Materials & Continua* 6(1), pp. 35-42.

Karplus, M.; Petsko, G.A. (1990): Molecular dynamics simulations in biolog,. *Nature*, **347**(6294) pp. 631-639.

Kim, J.; Eom, K.; Kwak, M.; Na, S. (2008): Application of Component Mode Synthesis to Protein Structure for Dynamic Analysis, *CMC:Computers, Materials*

& Continua 8(2), pp.67-73.

Kurkcuoglu, O.; Jernigan, R.; Doruker, P. (2004): Mixed levels of coarse-graining of large proteins using elastic network model succeeds in extracting the slowest motions, *Polymer*. **45**(2) pp. 649-657.

Kurkcuoglu, O.; Jernigan, R.; Doruker, P. (2005): Collective Dynamics of Large Proteins from Mixed Coarse-Grained Elastic Network Model, *QSAR & Combinatorial Science* **24**(4) pp. 443-448.

Lu, M.; Ma, J. (2005): The Role of Shape in Determining Molecular Motions, *Biophys. J.* **89**(4) pp. 2395-2401.

Maragakis, P.; Karplus, M. (2005): Large Amplitude Conformational Change in Proteins Explored with a Plastic Network Model: Adenylate Kinase, *Journal of Molecular Biology* **352**(4) pp. 807-822.

Miyashita, O.; Onuchic, J.N.; Wolynes, P.G. (2003): Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins, *Proceedings of the National Academy of Sciences of the United States of America* **100**(22) pp. 12570-12575.

Phillips, R.; Dittrich, M.; Schulten, K. (2002): Quasicontinuum Representations Of Atomic-Scale Mechanics: From Proteins to Dislocations. *Annual Review of Materials Research*. **32**(1) pp. 219-233.

Tama, F.; Sanejouand, Y.H. (2001): Conformational change of proteins arising from normal mode calculations, *Protein Eng*. **14**(1) pp. 1-6.

Teeter, M.M.; Case, D.A. (1990): Harmonic and quasiharmonic descriptions of Crambin, *The Journal of Physical Chemistry* **94**(21) pp. 8091-8097.

Tirion, M.M. (1996): Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Physical Review Letters* **77**(9) pp. 1905.

Tobi, D.; Bahar, I. (2005): Structural changes involved in protein binding correlatewith intrinsic motions of proteins in the unbound state, *Proceedings of the National Academy of Sciences of the United States of America* **102**(52) pp. 18908-18913.

Van Wynsberghe, A.W.; Cui, Q. (2006): Interpreting Correlated Motions Using Normal Mode Analysis, *Structure* **14**(11) pp. 1647-1653.

Xu, C.; Tobi, D.; Bahar, I. (2003): Allosteric Changes in Protein Structure Computed by a Simple Mechanical Model: Hemoglobin T<–>R2 Transition. *J. of Molecular Biology* **333**(1) pp. 153-168.