The 2015 World Congress on
Advances in Civil, Environmental, and Materials Research (ACEM15)
Incheon, Korea, August 25-29, 2015

Applying Computational Dynamics method to Protein Elastic Network Model

*Jae In Kim[1)], Kilho Eom[2)] and Sungsoo Na[3)]

[1), 3)] *Department of Mechanical Engineering, Korea University, Seongbuk-gu, Anamdong, Seoul 136-713, Korea*
[2)] *College of Sport Science Sungkyunkwan University, Suwon 440-746, Republic of Korea*
[1)] *jay414@korea.ac.kr*
[2)] *kilhoeom@skku.edu*
[3)] *nass@korea.ac.kr*

ABSTRACT

Normal mode analysis (NMA) is an effective method for analyzing the dynamics of protein structures. Elastic network model (ENM) is a harmonic potential based simple model to describe protein dynamics. ENM based NMA is a widely adopted method due to its simplicity and robustness for various proteins. Despite the simplicity of ENM, still long computation time is needed for supra molecular protein structures. We extended ENM with computational dynamics methods such as condensation (Kim, Jang et al. 2009) and component mode synthesis (Kim, Eom et al. 2008, Kim, Na et al. 2009). These methods allowed to reduce computation time keeping the accuracy of predicting structural measures as similar as ENM.

1. INTRODUCTION

Protein dynamics is essential for protein related to its function. Proteins in living bodies perform its own functions, and these functions are related to binding ligands, transporting a cargo, propagating signals, etc.(Bahar and Rader 2005, Bahar, Lezon et al. 2010) Proteins exhibit its own motions related to their functions. Therefore understanding protein's motions is an important concern as understanding protein's functions. Experimental methods such as X-ray crystallography and NMR, cryo-EM enable to find molecular structure of protein, however they find only single state of protein model. To understand the pathway of conformational change, other methods must be considered. But it is hard to find the molecular detail of conformational change of proteins.

Simulation methods are effective for understanding the conformational change of

---

[1)] Research Professor
[2)] Professor
[3)] Professor

proteins in atomic level. Conventionally molecular dynamics (MD) simulation is applied for understanding protein dynamics.(Izrailev, Stepaniants et al. 1998, Doruker, Atilgan et al. 2000) However, MD considers complicated potential energies and all-atom information of proteins. Moreover, water-box must be added to simulate, which increases the total number of atoms. To overcome this computational problem of large resources, normal mode analysis (NMA) is applied.(Go, Noguti et al. 1983, Cui and Bahar 2006) At first, NMA was performed with all-atom structures. However, Tirion suggested that a simple harmonic potential energy may describe the atomic interactions of protein well.(Tirion 1996) Bahar et.al suggested a model which considers only α-carbon from protein structures with harmonic potential energy, which is elastic network model (ENM).(Haliloglu, Bahar et al. 1997, Atilgan, Durell et al. 2001)

ENM is an efficient method to understand protein dynamics. However in cases considering large proteins, ENM could also require large computational resources. To overcome this problem various reduction methods were introduced.(Ming, Kong et al. 2003, Ming, Kong et al. 2003, Kim, Jernigan et al. 2005)

In structural dynamics view, component mode synthesis (CMS) is a very efficient method maintaining the accuracy of results and reducing computational time. We applied CMS to ENM based protein structures. CMS applied on protein structures described accurate eigenvalues compared to conventional ENM.

2. Elastic network model

As mentioned in introduction, ENM is a simple model to describe protein. ENM describes amino-acid of protein with a single pseudo atom on the position of α-carbon atom.(Haliloglu, Bahar et al. 1997, Atilgan, Durell et al. 2001) This is an essential atom for amino-acid which plays a role as backbone of proteins. Atomic coarse-graining leads to coarse-graining of potential energy.(Tozzini 2005) In contrast to all-atom models, ENM considers simple harmonic potential for interaction between α-carbon atoms. The potential energy is written as following

$$\mathrm{V} = \frac{\gamma}{2}\sum_i^N \sum_j^N \left(r_{ij} - r_{ij}^0\right)^2 \tag{1}$$

here, γ is force constant of harmonic interaction, N is number of atoms, r_{ij} is the distance of two atoms i and j, and superscript 0 stands for the equilibrium conformational state. Force constant is constant for all interactions of atoms. Cut-off distance r_c is introduced to determine the interaction range of atoms. Usually r_c=7~12Å is applied for ENM to describe the fluctuation of protein well.

Stiffness matrix Γ is evaluated using the cut-off distance condition.

$$\Gamma_{ij} = \begin{cases} -1, & i \neq j \; and \; r_{ij} \leq r_c \\ 0, & i \neq j \; and \; r_{ij} > r_c \\ -\sum_{j,i\neq i} \Gamma_{ij}, & i = j \end{cases} \tag{2}$$

NMA of ENM provides the conformational fluctuations of proteins, whose motion is described by eigenvalue problem

$$\gamma\Gamma_{ij}q_i = \omega^2 q_i \tag{3}$$

where ω is the natural frequency, and q_i is its normal mode.

3. Component Mode Synthesis

To begin with using CMS, we consider the structure as an assemblage of substructures.(Meirovitch 1980, Craig 1981) CMS describe the motion separately over each of the substructure by generating a eigenvalue problem and then assemble the substructures to work together as a single combined structure by using geometric compatibility at the interface of adjacent substructures. Since each substructure is modeled separately, there are redundant coordinates, as atoms shared by two adjacent substructures behave the same motions. The removal of redundant coordinates is carried out during an assembling process in which the constituent substructures are constrained to act as a whole structure. In the process, we calculate the modes of each component using NMA. Substructures are assembled using constraint mass points, and the modes of assembled structure can be composed. Using this methodology, we can approach the protein dynamics in a view point of domains, while maintaining computational accuracy, in terms of thermal fluctuations, and eigensolutions. First, we consider a Hamiltonian for a given protein structure.

$$H = \sum_{s=1}^{N_s} \frac{1}{2}\dot{\mathbf{u}}_s^T M_s \dot{\mathbf{u}}_s + \sum_{s=1}^{N_s} \frac{1}{2}\mathbf{u}_s^T K_s \mathbf{u}_s \tag{4}$$

$\mathbf{u}_s$ is a physical displacement vector for an arbitrary point P on the substructure s, M_s and K_s are mass and stiffness matrices for substructures. Physical displacement vector $\mathbf{u}_s$ is represented as series of space-dependent functions multiplied by time-dependent generalized coordinates as below:

$$\mathbf{u}_s(P,t) = \mathbf{\Psi}_s(P)\xi_s(t) \tag{5}$$

where $\mathbf{\Psi}_s$ may be regarded as assumed mode vector of substructure, and Hamiltonian can be rewritten as

$$H = \frac{1}{2}\begin{bmatrix}\dot{\xi}_1^T & \cdots & \dot{\xi}_{N_s}^T\end{bmatrix}\begin{bmatrix}W_1 & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & W_{N_s}\end{bmatrix}\begin{bmatrix}\dot{\xi}_1\\ \vdots\\ \dot{\xi}_{N_s}\end{bmatrix} + \frac{1}{2}\begin{bmatrix}\xi_1^T & \cdots & \xi_{N_s}^T\end{bmatrix}\begin{bmatrix}E_1 & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & E_{N_s}\end{bmatrix}\begin{bmatrix}\xi_1\\ \vdots\\ \xi_{N_s}\end{bmatrix} \tag{6}$$

$$= \frac{1}{2}\dot{\xi}^T W \dot{\xi} + \frac{1}{2}\xi^T E \xi$$

Where, $\boldsymbol{W}_s$ and $\boldsymbol{E}_s$ are matrices associated with eigenvector —$_s$, respectively.

$$\boldsymbol{W}_s = \mathbf{\Psi}_s^T M_s \mathbf{\Psi}_s \tag{7}$$

and

$$\boldsymbol{E}_s = \boldsymbol{\Psi}_s^T K_s \boldsymbol{\Psi}_s \tag{8}$$

Relating with $—_s$, $\boldsymbol{E}_s$ becomes diagonal matrix of eigenvalues.
W and E are block-diagonal matrices of $\boldsymbol{W}_s$ and $\boldsymbol{E}_s$.

A vector ξ is defined as $\xi = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_{N_s} \end{bmatrix}$ such that the degrees of freedom of a vector ξ is given by n. Vector ξ includes a certain number of redundant values due to constraints of boundaries between adjacent components. Assuming there are n_c constraints and ξ has n dimensions, total dimension is $n_\beta = n - n_c$.
Coordinate transformation between $\xi(t)$ and $\beta(t)$ is related as

$$\xi(t) = \mathbf{B}\beta(t) \tag{9}$$

where $\mathbf{B}$ is a constraint matrix, and $\beta(t)$ is the *n*-dimensional independent generalized coordinate vector. The matrix $\mathbf{B}$ is $n_\beta \times n$ and reflects certain geometric condition at boundary points for which are shared by substructures r and s, because the points has the same displacement as below

$$\mathbf{u}_s = \mathbf{u}_r \qquad s \neq r \tag{10}$$

Moreover, the translational displacements at the interfaces are related to the generalized displacement vector $\xi(t)$. In view of this, one can combine Eq.(10) corresponding to all interfaces into a single constraint equation as

$$\mathbf{T}\xi = 0 \tag{11}$$

Then dividing the vector $\xi(t)$ into vector $\beta(t)$ of independent variables and a vector $\mathbf{p}$ of dependent variables and partitioning the matrix T as follows:

$$\mathbf{T} = \left[\mathbf{T}_1 \vdots \mathbf{T}_2\right] \tag{12}$$

Eq.(11) can be rewritten as

$$\mathbf{T}_1\beta + \mathbf{T}_2\mathbf{p} = 0 \tag{13}$$

which yields

$$\mathbf{p} = -\mathbf{T}_2^{-1}\mathbf{T}_1\beta \tag{14}$$

Eq.(14) permits one to write a relation Eq.(9) between the $\beta(t)$ of independent generalized coordinates for the full structure and the vector ξ.

$$\mathbf{B} = \begin{bmatrix} I \\ \cdots \\ -\mathrm{T}_2^{-1}\mathrm{T}_1 \end{bmatrix} \tag{15}$$

Where I is a unit matrix of order n_β.
The constraint condition allows one to integrate the components to assembled structure. Now we can rewrite the Hamiltonian for assembled protein structure.

$$H = \frac{1}{2}\dot{\beta}^T Q\dot{\beta} + \frac{1}{2}\beta^T O\beta \tag{16}$$

where

$$\boldsymbol{Q} = \mathbf{B}^T \boldsymbol{W} \mathbf{B}, \quad \boldsymbol{O} = \mathbf{B}^T \boldsymbol{E} \mathbf{B} \tag{17}$$

are $n_\beta \times n_\beta$ matrices for the assembled structure.

Using these Q and O matrices leads to eigenvalue problem

$$O\mathbf{U} = Q\mathbf{U}\Lambda \tag{18}$$

$\mathbf{U}$ is $n_\beta \times n_\beta$ modal matrix and Λ is diagonal matrix of eigenvalues of reconstructed structure. Here we use the relation

$$\boldsymbol{\Phi} = \varphi \mathbf{B} \mathbf{U} \tag{19}$$

Here φ is defined in similar way with ξ, which obtains $n \times n_\beta$ matrix **...**, whose column vectors give the atomic displacement of each modes of the reconstructed structure. This atomic displacement can be compared with eigenvectors of NMA of the original structure.

4. Results

4.1 Low frequency normal modes

First we consider the lowest frequency mode of the protein structure, which is known as it plays a role in protein dynamics. In figure 1, we show the lowest frequency mode for both original ENM structure and structure from CMS. Result is plotted for F0 ATPase motor protein (PDB code: 1c17)

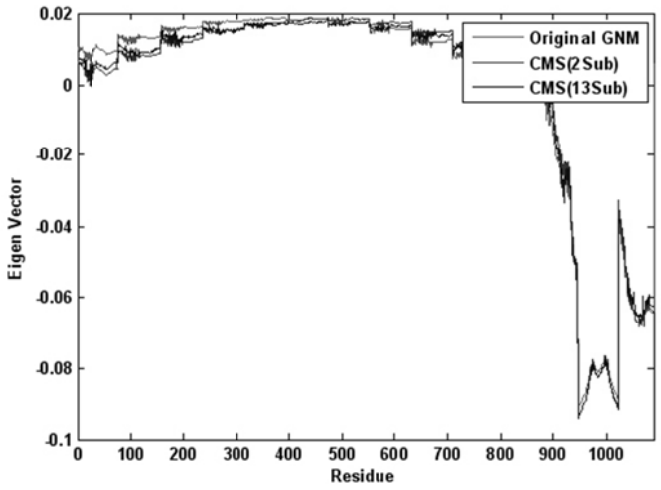

Figure 1. The lowest frequency mode of target proteins from the component mode synthesis structure and original ENM structure: F0 ATPase motor protein (PDB code: 1c17)

4.2 Mode Correlation

Mode Correlation shows direct compare of normal modes of two methods quantitatively. Here, we compared normal modes from original ENM structure and CMS structure in figure 2. Calculated values vary from 0 to 1, and smaller value shows bright colors and large values shows dark colors. For the results, 50 lowest frequency modes are shown for each target protein, and they are showing very similar modes for the

lowest 10 modes. However, we have shown some factors like mean square fluctuation and cross correlation which use all modes of each structures showing very good result. Despite of unequalness of higher modes, other facts showed good results. This indicates that the low frequency modes plays an important role for dynamics of protein structures.

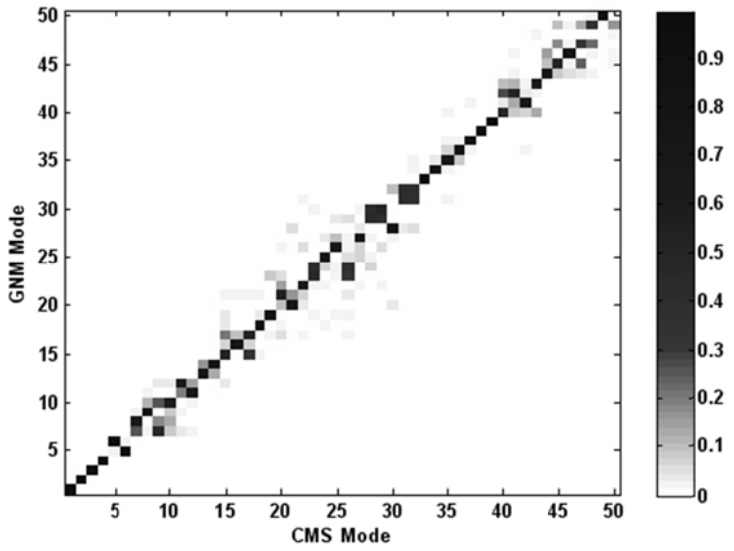

Figure 2. Mode Correlation of target proteins from the component mode synthesis structure and original ENM structure: F0 ATPase motor protein (PDB code: 1c17)

5. CONCLUSIONS

In this study, we approached the thermal fluctuations for proteins with CMS, which considers the structure as an assembly of a number of substructures. We divided the original protein structure to substructures refer to domain of original protein structure, and applied CMS to assemble the structure. During the process, stiffness matrix was diagonalized, and high frequency modes were reduced by the number of constraints. Also, CMS was possible to predict with not only 2 substructures but even with 13 substructures for F0 ATPase motor protein. Selected proteins were not restricted to repeated structure as actin filaments(Ming, Kong et al. 2003, Ming, Kong et al. 2003), but CMS was able to assemble the substructures in different configurations. The results were compared with the simulation data from original ENM. Original ENM structure was showing a very good result for simulating thermal fluctuation of protein structures. And the simulation result from applying CMS to protein structures showed coincidence with the original ENM structure. CMS has some benefits due to reducing high frequency normal modes and simplifying stiffness matrix into diagonal term, and this may show a way for simulating large protein structures which cannot be done by original ENM.

REFERENCES

Atilgan, A. R., S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin and I. Bahar (2001). "Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model." Biophys. J. **80**(1): 505-515.
Bahar, I., T. R. Lezon, L.-W. Yang and E. Eyal (2010). "Global Dynamics of Proteins: Bridging Between Structure and Function." Annual Review of Biophysics **39**(1).
Bahar, I. and A. J. Rader (2005). "Coarse-grained normal mode analysis in structural biology." Current Opinion in Structural Biology **15**(5): 586-592.
Craig, R. J. (1981). Structural Dynamics An introduction to computer methods, John Wiley & Sons.
Cui, Q. and I. Bahar (2006). Normal Mode Analysis : Theory and applications to biological and chemical systems, Chapman & Hall/CRC.
Doruker, P., A. R. Atilgan and I. Bahar (2000). "Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: Application to alpha-amylase inhibitor." Proteins: Structure, Function, and Genetics **40**(3): 512-524.
Go, N., T. Noguti and T. Nishikawa (1983). "Dynamics of a small globular protein in terms of low-frequency vibrational modes." Proceedings of the National Academy of Sciences of the United States of America **80**(12): 3696-3700.
Haliloglu, T., I. Bahar and B. Erman (1997). "Gaussian Dynamics of Folded Proteins." Physical Review Letters **79**(16): 3090.
Izrailev, S., S. Stepaniants, B. Isralewitz, D. Kosztin, H. Lu, F. Molnar, W. Wriggers and K. Schulten, Eds. (1998). Computational Molecular Dynamics: Challenges, Methods, Ideas. Computational Science and Engineering. Berlin, Springer-Verlag.
Kim, M. K., R. L. Jernigan and G. S. Chirikjian (2005). "Rigid-Cluster Models of Conformational Transitions in Macromolecular Machines and Assemblies." Biophys. J. **89**(1): 43-55.
Meirovitch, L. (1980). Computational methods in structural dynamics, Sijthoff and Noordhoff.
Ming, D., Y. Kong, Y. Wu and J. Ma (2003). "Simulation of F-Actin Filaments of Several Microns." Biophys. J. **85**(1): 27-35.
Ming, D., Y. Kong, Y. Wu and J. Ma (2003). "Substructure synthesis method for simulating large molecular complexes." Proceedings of the National Academy of Sciences **100**(1): 104-109.
Tirion, M. M. (1996). "Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis." Physical Review Letters **77**(9): 1905.
Tozzini, V. (2005). "Coarse-grained models for proteins." Current Opinion in Structural Biology **15**(2): 144-150.