 ICCES, vol.2, no.2, pp.53-59, 2007

Multi-Component Modal Analysis of Protein Structure

G. Yoon[1], K. Bong[2], J. Kim[3], I.H. Ahn[4],
K. Eom[5] and S. Na[6]

Summary

This paper presents multi-component mode methodology applicable to biomolecular structures for understanding the dynamics of proteins. Even though the conventional normal mode analysis has been contributed for analyzing the dynamics and thermal fluctuations of proteins, it frequently encounters with the computational prohibition for large proteins due to memory requirement. To overcome the conventional computational limitations, the drawback motivates one to develop various model reduction methods, which reduces the degrees of freedom of the full model so as to decrease the computational expense, while the computational accuracy is maintained. Our results demonstrate that the multi-component modal analysis applied to the biomolecular structures predicts the dominant eigenmodes and fluctuation of proteins accurately with reducing the computational cost enormously.

Introduction

It is well known that the normal mode analysis (NMA) has been widely employed with great success to gain insights into biological functions of proteins [1]. In general, the protein performs the biological functions through the molecular structural change, which may be described by the low-frequency modes [2-3]. The structural change of proteins is known to be related to the molecular vibration induced by the thermal energy [2-3]. For understanding the low-frequency modes driven by the thermal energy, the NMA has been an effective computational method in computational molecular modeling community, because the NMA has allowed one to interpret the molecular structural change, which may not be analyzed by the Molecular Dynamics (MD) simulation [4] due to the necessity of large memory requirements to save the trajectories and the large number of computation for calculation of the inter-atomic forces from anharmonic potential field.

The basic principle of NMA is to solve the eigenvalue problem for the protein molecular structure. Unlike the structural dynamics problem, the protein exhibits the very complicated potential field consisting of energies representing the stretch of chemical bonds, the bending of chemical bonds, the torsion of chemical bonds,

[1]Graduate Student, Dept. of Mechanical Eng.., Korea Univ., Korea
[2]Graduate Student, Dept. of Mechanical Eng.., Korea Univ., Korea
[3]Graduate Student, Dept. of Mechanical Eng.., Korea Univ., Korea
[4]Graduate Student, Dept. of Mechanical Eng.., Korea Univ., Korea
[5]Researcher, Microsystem Research Center, KIST, Seoul, Korea
[6]Associate Professor, Dept. of Mechanical Eng., Korea Univ., Korea

Van der Waal's interactions (non-bonded interactions), electrostatic interactions, and other physical molecular interaction terms depending on the problem [4]. The computational inefficiency is ascribed to the complicated potential such that it requires the high computing expense for the large proteins in calculating the equilibrium position and stiffness matrix. In this sense, Ma employed the Substructure Synthesis Model (SSM), developed by Meirovitch etc. [5] for solving the structural dynamic problem in aerospace engineering, for molecular vibration of proteins [6].

In the present paper, multi-component modal analysis is applied for a validation of obtaining reliable equilibrium fluctuation, global large amplitude motions, reducing the degrees of freedom in the large molecular structure while keeping the computational accuracy and for investigating the dynamic motion of biostructure, with treating the global motions of a structure as a collection of an assembling of substructures.

Multi-Component Modal Analysis

Let us assume that the particular structure consists of an assemblage of substructures. We will consider a given substructure s and write the total displacement vector $\mathbf{u}_s(x,y,z,t)$ of an arbitrary point $P(x,y,z)$ on the substructure. The kinetic energy associated with the substructure s has the general expression

$$T_s = \frac{1}{2}\int_{D_s} m_s \dot{\mathbf{u}}_s^T \dot{\mathbf{u}}_s dD_s \tag{1}$$

$\mathbf{u}_s$ can be written as

$$\mathbf{u}_s(x,y,z,t) = \mathbf{\Phi}_s(x,y,z)\,\zeta_s(t) \tag{2}$$

Introducing Eq. (2) into (1), we obtain

$$T_s = \tfrac{1}{2}\dot{\zeta}_s^T M_s \dot{\zeta}_s \tag{3}$$

where

$$M_s = \int_{D_s} m_s \mathbf{\Phi}_s^T \mathbf{\Phi}_s dD_s \tag{4}$$

is the mass matrix for the corresponding substructure. Similarly, the potential energy can be written in general form as

$$V_s = \tfrac{1}{2}[\mathbf{u}_s, \mathbf{u}_s] \tag{5}$$

Substitution of Eq. (2) into (5) gives us

$$V_s = \tfrac{1}{2}\zeta_s^T K_s \zeta_s \tag{6}$$

where

$$K_s = [\Phi_s, \Phi_s] \tag{7}$$

is the substructure K matrix..

Considering the original structure containing m substructures, s=1,2,...,m we need to introduce the following NxN block diagonal matrices in order to integrate individual substructures into assembled structure

$$M_d = block - diag[M_s] \quad K_d = block - diag[K_s] \tag{8}$$

Herein we have disjoint vector

$$\zeta_d(t) = \left[\zeta_1^T(t)\,\zeta_2^T(t)\cdots\zeta_m^T(t)\right]^T \tag{9}$$

assuming this vector has dimension m and there are c constraints, that the number of independent generalized coordinates is $n = m - c$, where n is the number of degree of freedom of the system. Denoting $\zeta(t)$ the n-dimensional independent generalized coordinate vector, we can write the relation between $\zeta_d(t)$ and $\zeta(t)$ in the matrix form

$$\zeta_d(t) = C\zeta(t) \tag{10}$$

where C is an $m \times n$ transformation matrix, depending on the nature of the constraints. With Eq.(10), we can write Rayleigh quotient

$$R = \zeta^T K\zeta / \zeta^T M\zeta \tag{11}$$

where

$$K = C^T K_d C, \quad M = C^T M_d C \tag{12}$$

are the $n \times n$ stiffness and mass matrices for the assembled structure.

From the Rayleigh-Ritz principle (11), rendering R stationary leads to an eigenvalue problem

$$KU = MUA \tag{13}$$

where U is an $n \times n$ modal matrix, and A is the diagonal matrix of the eigenvalue for the assembled structure. By using $U' = \Phi_d CU$, where Φ_d is defined in similar way with Eq. (9) we can obtain U', whose column vectors give the atomic displacements in the corresponding modes of the assembled structure. These atomic displacements can be compared with the eigenvectors from NMA of the assembled structure.

For a simple one dimensional example illustrating multi-component modal analysis (MCMA), 400-mass point chain is shown in Fig. 1. The figure represents

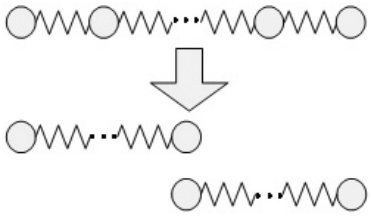

Figure 1: 400-mass point chain connected by springs

atom chains connected by spring on straight line in the equilibrium configuration with same distance apart.

The eigenvalues obtained from standard NMA in original 400-mass point was compared with results using multi-component modal analysis in Fig. 2. The multi-component modal analysis at present performed two different configuration mode synthesis after fusing two substructures; one is each with 200 modes(200/200) and the other is 219 modes in total with two different mode synthesis (200/20). The eigenvalues from multi-component modal analysis fusing each with 200-mass point match those from obtained using NMA with full 400-mass chain. Fig. 3 displays enlarged plot only with first 40 modes. Both synthesis methods have excellent results for the lowest frequency modes before first 40modes.

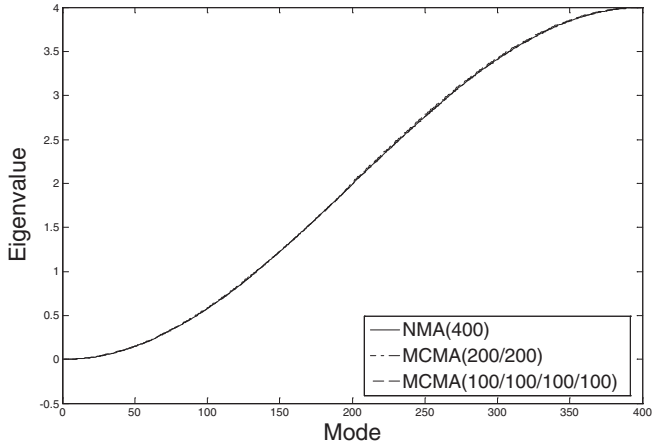

Figure 2: Eigenvalues for the fused two different synthesis

Results

We consider hemoglobin (Hb) for modeling protein structure using mechanical mass-spring model of which number of dominant atoms called Cα is 574. The

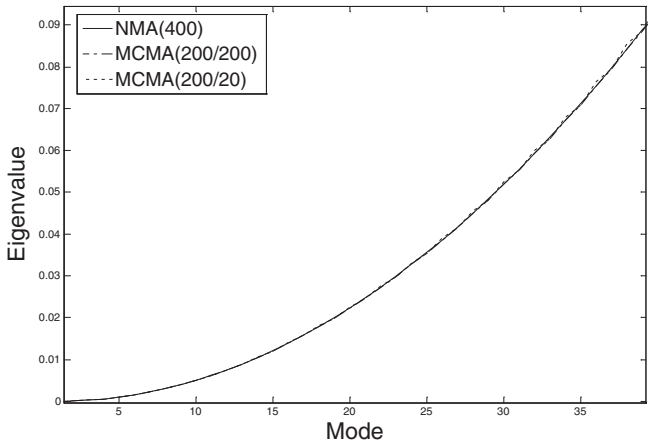

Figure 3: Enlarged version of first 40 modes

schematic diagram of hemoglobin is shown in Fig. 4. Specifically, Hb is known to have collective motion of 4 domains (monomers) of hemoglobin [7]. That is, the monomer 1 (residue number: 1-141) and monomer 2 (residue number: 142-286), referred to as substructure A, exhibit the correlated motion, and also the monomer 3 (residue number: 287-427) and monomer 4 (428-572), referred to as substructure B, have the correlated motion.

Figure 4: Structure of hemoglobin

From the given structure, the Hb may be considered either two substructures or four substructures in order to simulate two different synthesis based on multi-component modal analysis. We draw the similar conclusions in the protein structure as the previous sample problem (mass chain) concerning eigenvalues both in Figs. 5 and 6. Results presented in Fig. 7 show that the MCMA exhibits the similar characteristics of fluctuation for collective or similar motions to that of original full model.

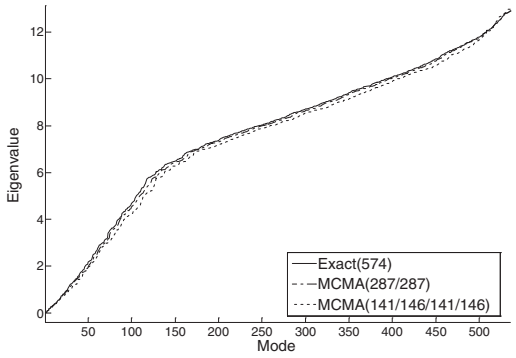

Figure 5: Eigenvalues for the two different synthesis

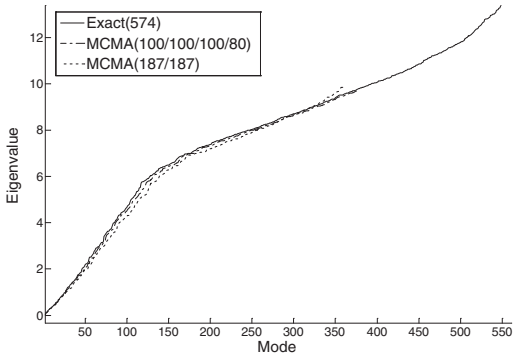

Figure 6: Eigenvalues obtained using low frequency modes

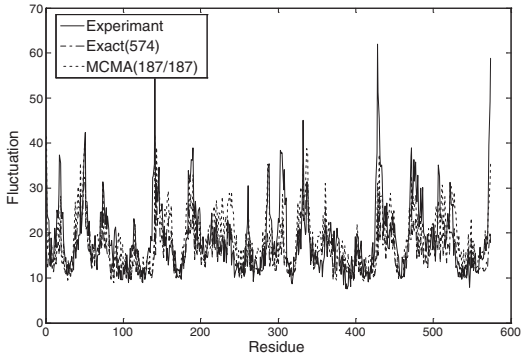

Figure 7: Comparisons of mean square fluctuations of protein (Hb) generated by full model and multi-component modal analysis

References

1. Weiner, J. H., (2002) *Statistical Mechanics of Elasticity*, 2^{nd} ed. Dover Publications, INC.
2. Cui, Q., Li, G., Ma, J., Karplus, M., (2004) A normal mode analysis of structural plasticity in the biomolecular motor F(1)-ATPase, *J. Mol. Biol.*, **340** 345
3. F. Tama, Y. H. Sanejound, (2001),"Conformation change of proteins arising from normal mode calculations," *Protein Eng.*, **14** pp.1-6
4. Allen, M.P., Tildesley, D.J., (1989) *Computer Simulation of Liquids*, Oxford University Press
5. Ming, D., Kong, Y., Wu, Y., Ma, J., (2003), "Substructure synthesis method for simulating large molecular complexes," 104, *Proc. Natl. Acad. Sci. USA*, **100**
6. Hale, A.L., Meirovitch, L., (1980)," A general substructure synthesis method for the dynamic simulation of complex structures," *Journal of Sound and Vibration* **69**(2) pp.209-326
7. Xu, C., Tobi, D., Bahar, I., 2003, "Allosteric changes in protein structure computed by a simple mechanical model: Hemoglobin T↔R2 transition," *J. Mol. Biol.* **333,** 153