Springer

Journal of
Mechanical
Science and
Technology

Journal of Mechanical Science and Technology 22 (2008) 506~513

www.springerlink.com/content/1738-494x

Modified mechanical mass-spring model of biomolecules for understanding dynamics of proteins

Kilho Eom[1], Jeong-Hee Ahn[2], Jae-In Kim[2], Gwon-Chan Yoon[2] and Sungsoo Na[2,*]

[1]*Microsystem Research Center, Korea Institute of Science and Technology, Hawolgok-dong, Sungbuk-gu, Seoul 136-791, Korea*
[2]*Department of Mechanical Engineering, Korea University, Anam-dong, Sungbuk-gu, Seoul 136-713, Korea*

(Manuscript Received June 9, 2007; Revised December 10, 2007; Accepted December 10, 2007)

Abstract

A dynamic model applicable to biomolecular structures for understanding the dynamics and the vibrational behaviors of protein is considered. A mechanical mass-spring model represented by point masses and harmonic springs is presented. The biomolecular structure may be envisioned by a mass and spring system with multi-degrees-of-freedom because dominant atoms in protein may be considered to be point masses, and bonding and non-bonding interactions between atoms of interest and surrounding atoms within some critical distances are implemented by a spring. Furthermore, a model condensation scheme is to be introduced because most proteins have large degree of freedom requiring large computation time and memory, which results in reducing computational cost and maintaining the accurate predictions. From solving the corresponding eigenvalue problem constructed from a multi-degree-of-freedom system, our results show the modified mechanical spring-mass model of a biostructure through a condensation scheme is very successful in predicting the dynamics of molecular structures in terms of thermal fluctuations and eigenmode, etc.

Keywords: Mass-spring model; Biomolecular structure; Multi-degree-of-freedom system; Eigenvalue problem; Model condensation

1. Introduction

In general, a protein performs biological functions related to the dynamic and vibrational functions through molecular structural changes, which may be described mainly by the low-frequency normal modes [1-4]. It is essential to investigate the biological function of proteins because we can understand the mechanism of disease occurrence in the human body as well as develop new medicine through exploring structural conformations highly related to biological function of proteins. In this connection, it should be noticed that the significant connection between biological function and structure change of proteins should be identified. The present study focuses on protein dynamics, at an initial stage, which may eventually provide the information regarding the structural conformation of proteins, for example, structural state change between two equilibrium states such as open and closed states. The structural change of proteins is known to be significantly related to the molecular vibration induced by the thermal energy [1-4], which motivates one to approach the present issue from a mechanical viewpoint.

For understanding the low-frequency modes driven by thermal energy, the normal mode analysis (NMA) has been playing a significant role in understanding the dynamics and the vibrational behaviors of molecular structures [5]. The NMA allows one to interpret the molecular structural change, which may not be analyzed by the molecular dynamics (MD) simulation [6]. It is mainly due to the necessity of large memory capacity to save the trajectories and the large computation time for calculation of the inter-atomic forces from a non-harmonic potential field. Specifically, MD simulation is not applicable to problems requiring large spatial and time scales, such as DNA-protein complex [7].

The major procedure of NMA for the molecular

*Corresponding author. Tel.: +82 2 3290 3370, Fax.: +82 2 926 9290
E-mail address: nass@korea.ac.kr
DOI 10.1007/s12206-007-1202-7

structure is to solve the eigenvalue problem. Unlike the structural dynamic problem, the protein exhibits a very complicated potential field consisting of energies comprising the stretch of chemical bonds, the bending of chemical bonds, the torsion of chemical bonds, van der Waal's interactions (non-bonded interactions), electrostatic interactions, and other physical molecular interaction terms depending on the problem [6]. With the given potential field for proteins, the equilibrium position is evaluated by minimization of potential field, and then the stiffness matrix is calculated from the second derivative of the potential field with respect to the equilibrium position [5]. The computational inefficiency associated with all the previous procedures is ascribed to the complicated potential such that it requires a high computing expense for large proteins in calculating the equilibrium position and stiffness matrix.

In this respect, the Gaussian network model (GNM) has been developed and become a staple method for understanding protein dynamics and flexibility. The GNM is essentially one-dimensional mechanical mass-spring model (MS model) and a great deal of work and some literature has focused on model reduction methods for improving efficiency for larger macromolecular structures [2, 8]. The MS model assumes that the protein structures can be considered as a collection of C_α atoms connected by harmonic springs such that the C_α atoms, which are selected from each amino acid along the backbone chain, occupy about 10% of all the atoms in proteins, since a protein is a long chain of amino acids connected by polypeptide bonds. A three-dimensional image of the molecular structure of protein (e.g., hemoglobin (Hb); pdb code, 1bbb) is shown in Fig. 1. The number of alpha carbon atoms of Hb represented by point mass in the protein modeling is 572. Despite the simplicity of the model, the numerical results show that the MS model is very successful in predictions on the thermal fluctuations [9], and low-frequency modes, and may also give insight into the dynamic and/or mechanical behavior of proteins. Nevertheless, the MS model may still be computationally inhibitive for supermolecular structures with large degrees of freedom such as viruses. To cope with this problem, a robust model condensation of the MS model is introduced and adopted to modify the MS model, which reduces the degrees of freedom of the MS model so as to decrease the computational cost, while the computational accuracy is maintained.

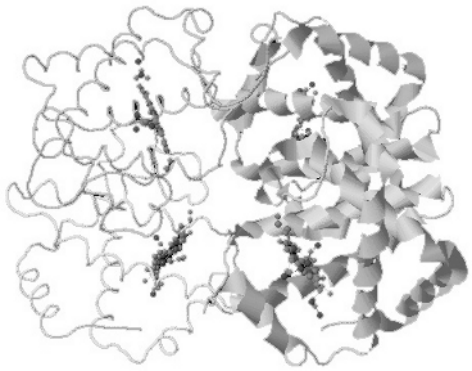
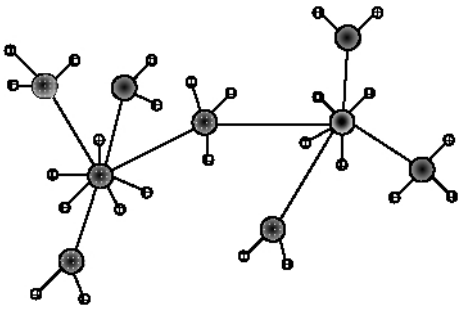
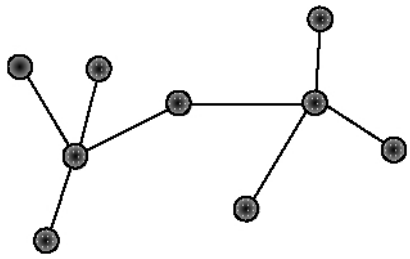


Fig. 1. Secondary structure of model protein: open form of hemoglobin (Hb, 1bbb).



(a) Dominant atoms along the backbone with side chain



(b) Dominant atoms only along backbone

Fig. 2. Schematic diagram of dominant atoms along the backbone.

2. Model derivation and model modification

2.1 Review of mechanical mass-spring model

The MS model regards the protein structure as a simple one-dimensional mass-spring network associated with alpha carbons only. Thus the MS model allows one to construct the equivalent structural dynamic model of proteins such that among all atoms only alpha carbons are taken into account, while the majority of unconsidered atoms are removed (see Fig. 2), and uniform springs of force constant γ connect

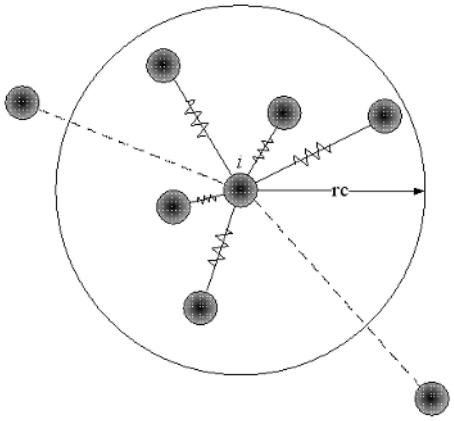


Fig. 3. Schematic diagram illustrating how harmonic spring connects the pairs of alpha carbons in the neighborhood.

the pairs of alpha carbons in the neighborhood within cutoff distance r_c (see Fig. 3). In MS modeling each point (node or residue) represents alpha carbons. The equilibrium fluctuation of alpha carbons is described by a harmonic potential such as [7]

$$P = \frac{\gamma}{2}\sum_{i,j}^{N}\Gamma_{ij}u_i u_j \tag{1a}$$

where γ is the force constant, u_i is the fluctuation of atom i, and Γ_{ij} is the Kirchhoff matrix (stiffness matrix) representing the interaction between atom i and j, defined as

$$\Gamma_{ij} = -(1-\delta_{ij})H(r_c - \mathrm{r}_{ij}^0) - \delta_{ij}\sum_{k\neq i}^{N}\Gamma_{ik} \tag{1b}$$

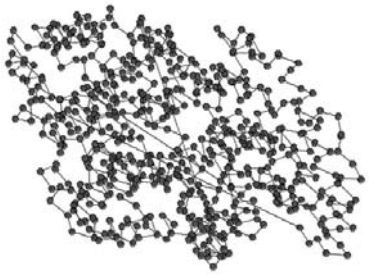
Herein r_{ij} is the distance between alpha carbon atoms i and j; $H(x)$ is the Heaviside unit step function; r_c is the critical cut-off distance, and the superscript 0 indicates the equilibrium position. In addition, δ_{ij} is the Kronecker delta, i.e. $\delta_{ij}=1$ if $i = j$; otherwise $\delta_{ij}=0$. Eq. (1b) is further reduced to the form

$$\Gamma_{ij} = \begin{cases} -1 & \text{if } i\neq j \text{ and } r_{ij} < r_c \\ 0 & \text{if } i\neq j \text{ and } r_{ij} < r_c \\ -\sum\Gamma_{ij} & \text{if } i = j \end{cases} \tag{1c}$$

The eigenvalue problem provides the conformational fluctuation of proteins, which can be described as

$$\gamma\Gamma_{ij}\mathrm{u}_j = \omega^2\mathrm{u}_j \tag{2}$$

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



(a) Original MS model of Hb

(b) Modified MS model of Hb

Fig. 4. Original and Modified Mechanical mass-spring models as the counterpart of Fig. 1.

The force constant γ is assumed to be constant if the distance between the two atoms is within cutoff distance. Specifically, it is assumed that if the distance between two atoms in the neighborhood is less than critical cut-off distance r_c, which is empirically determined as 10~12Å, then the link stiffness is set to a constant, otherwise it is set to zero [10]. In this regard the cut-off distance r_c determines whether two adjacent residues are connected with a spring. For the present case, the cutoff distance during simulation is determined to be 10Å. It should be noted that the damping effect is not considered in the present modeling. A schematic diagram of the both original MS model and modified MS model for simulating the dynamics of hemoglobin is displayed in Fig. 4. Statistical mechanics theory provides the thermal fluctuations around its equilibrium from normal mode analysis as follows [11]:

$$C_{ij} \equiv \left\langle \left(\mathbf{r}_i - \mathbf{r}_i^0\right)\cdot\left(\mathbf{r}_j - \mathbf{r}_j^0\right)\right\rangle = \sum_{p=2}^{N}\frac{kT}{m\omega_p{}^2}\mathrm{u}_i^{(p)}\mathrm{u}_j^{(p)} \tag{3a}$$

where C_{ij} is the correlation matrix, $\langle f\rangle$ indicates the ensemble average or time average of variable f, k is the Boltzmann constant, T is the absolute temperature, ω_p is the natural frequency for the p-th mode, $\mathrm{u}_i{}^{(p)}$ is

the *i*-th component of the eigenvector for the *p*-th mode, *N* is the total number of alpha carbons, and the summation excludes the first mode corresponding to the rigid body motion in the one-dimensional spring network, so index *p* of summation starts with 2. The magnitude of C_{ii} represents the mean-square fluctuation of alpha carbon atoms (often referred to as residue) *i,* which is given by

$$\left\langle \left(\mathbf{r}_i - \mathbf{r}_i^0\right)^2 \right\rangle = C_{ii} \tag{3b}$$

Furthermore, the B factor of a residue *i* experimentally measurable by, e.g., X-ray crystallography can be compared to the mean square fluctuations of residues. The B factor of residue *i* also can be calculated according to [9]

$$B_i = \frac{8\pi^2}{3}\left\langle \left(\mathbf{r}_i - \mathbf{r}_i^0\right)^2 \right\rangle = \frac{8\pi^2}{3} C_{ii} \tag{3c}$$

which is directly compared to the corresponding B factor, also available in the pdb files stored in the Protein Data Bank.

2.2 Modified mass-spring model

The eigenvalue problem constitutes the most time consuming part of the computations for large proteins. In this connection, we implemented model condensation of the original MS model for model protein Hb. The potential energy and kinetic energy for the molecular system can be written, respectively, in matrix form as:

$$P = \frac{1}{2}\mathbf{q}^T\mathbf{K}\mathbf{q}\,, \quad K = \frac{1}{2}\dot{\mathbf{q}}^T\mathbf{M}\dot{\mathbf{q}} \tag{4a, b}$$

Herein, **K** and **M** are stiffness and mass matrix of the system and **q** is a displacement vector. The protein molecular structure described by MS model may be partitioned into two groups, groups *A* and *B,* such that group *A,* called the master substructure, represents the dominant subset of the MS model, while group *B,* called the slave substructure, indicates the subset which is supposed to be eliminated for model modification. For the present study, we preserve *N*/*n* (*n* = 2, 4, 8, 16, *etc*.) residues (atoms), called master substructures, while we refer to *N* (1-1/*n*) residues to be discarded, called slave substructures, through model modification. When we select *n* = 2 for master substructure, the atoms associated with the master substructure residues correspond to 2, 4, 6, ⋯ , etc., while the rest of the residues including 1, 3, 5, ⋯ etc., denote the slave structure. With *n* = 3, the master structure includes residues such as 3, 6, 9, ⋯ , etc., and the other atoms are slave structure. In a similar fashion, one can construct a modified structure retaining (*N*/*n*) residues uniformly throughout the atoms of structure once *n* is determined. In this respect, the stiffness matrix and mass matrix can be partitioned accordingly as follows: [12]

$$P = \frac{1}{2}\begin{bmatrix}\mathbf{q}_A\\ \mathbf{q}_B\end{bmatrix}^T\begin{bmatrix}\mathbf{K}_{AA} & \mathbf{K}_{AB}\\ \mathbf{K}_{BA} & \mathbf{K}_{BB}\end{bmatrix}\begin{bmatrix}\mathbf{q}_A\\ \mathbf{q}_B\end{bmatrix}$$

$$K = \frac{1}{2}\begin{bmatrix}\dot{\mathbf{q}}_A\\ \dot{\mathbf{q}}_B\end{bmatrix}^T\begin{bmatrix}\mathbf{M}_{AA} & \mathbf{M}_{AB}\\ \mathbf{M}_{BA} & \mathbf{M}_{BB}\end{bmatrix}\begin{bmatrix}\dot{\mathbf{q}}_A\\ \dot{\mathbf{q}}_B\end{bmatrix} \tag{5a, b}$$

where $\mathbf{K}_{AA}$ represents the stiffness of substructure *A*, $\mathbf{K}_{AB}$ indicates the interaction between substructures *A* and *B*, and so on. Herein $\mathbf{q}_B$ represents displacement of substructure *B*, while $\mathbf{q}_A$ means displacement of substructure *A*.

In the eigenvalue problem Eq. (6) associated with present model modification, the mass matrix generally neglects point masses associated with slave residues. The fact that certain diagonal entries in the mass matrix are zero is an indication that the corresponding displacements (slave residues) are not significant to the solution and can be eliminated during the eigenvalue problem formulation. Its net result is to reduce the order of the eigenvalue problem, which enables one to partition the eigenvalue problem as follows: [12]

$$\begin{bmatrix}\mathbf{K}_{mm} & \mathbf{K}_{ms}\\ \mathbf{K}_{sm} & \mathbf{K}_{ss}\end{bmatrix}\begin{bmatrix}\mathbf{q}_m\\ \mathbf{q}_s\end{bmatrix} = \omega^2\begin{bmatrix}\mathbf{M}_{mm} & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}\mathbf{q}_m\\ \mathbf{q}_s\end{bmatrix} \tag{6}$$

Herein $\mathbf{K}_{mm}$ represents the stiffness of substructure including master residues, $\mathbf{K}_{ss}$ denotes stiffness of substructure associated with slave residues, while $\mathbf{K}_{ms}$ indicates the interaction between master substructure and slave substructure. Furthermore, $\mathbf{q}_m$ and $\mathbf{q}_s$ represent master displacement and slave displacement, respectively. Eq. (6) can be separated into the two equations

$$\begin{aligned}&\mathbf{K}_{mm}\mathbf{q}_m + \mathbf{K}_{ms}\mathbf{q}_s = \omega^2\mathbf{M}_{mm}\mathbf{q}_m\\ &\mathbf{K}_{sm}\mathbf{q}_m + \mathbf{K}_{ss}\mathbf{q}_s = \mathbf{0}\end{aligned} \tag{7a, b}$$

Solving Eq. (7 b) for $\mathbf{q}_s$, one obtains

$$\mathbf{q}_s = -\mathbf{K}_{ss}^{-1}\mathbf{K}_{sm}\mathbf{q}_m \tag{8}$$

Substituting Eq. (8) into Eq. (7a), one obtains the condensed eigenvalue problem as follows:

$$\mathbf{K}_1\mathbf{q}_m = \omega^2\mathbf{M}_1\mathbf{q}_m \tag{9}$$

where

$$\mathbf{K}_1 = \mathbf{K}_{mm} - \mathbf{K}_{ms}\mathbf{K}_{ss}^{-1}\mathbf{K}_{sm}, \quad \mathbf{M}_1 = \mathbf{M}_{mm} \tag{10a, b}$$

It should be noticed that the eigenvalue problem for the original structure is defined as

$$\mathbf{Kq} = \omega^2\mathbf{Mq} \tag{11}$$

where **K**, **M**, **q** are defined in Eq. (4)

Following the above steps allows the modification of the MS model on the basis of elimination of slave residues not taken in the modified model.

Both the eigenvalue and eigenvector for the modified MS model of a protein can be very easily computed from Eq. (4) through Eq. (10) which can be directly compared with the eigensolutions obtained from direct normal mode analysis of the original MS model. We compared the eigensolution of the original MS model with one of the modified model with the reduced degrees of freedom, *N*/16 (*n*=16) in the present study. Furthermore, for the comparison of fluctuation of the original MS model and the modified structure in the case of the degrees of freedom, *N*/16, the model modification was implemented in a hierarchical manner such that we reduce the degrees of freedom from *N* (*N* = degrees of freedom for original MS model) to *N*/*n*, (*n* = 2, 4, 8, 16, *etc*.). Specifically, the procedure is as follows: (i) partition the stiffness matrix as in Eq. (6), (ii) reconstruct the effective stiffness matrix for master residues, (iii) set the condensed model as initial structure for further modification, (iv) the steps (i)- (iii) are repeated 16 times for the present case because in this example the original MS model can be grouped in terms of 16 substructures such as (1, *n*+1, 2*n*+1, ⋯), (2, *n*+2, 2*n*+2, ⋯), ⋯ (*n*, 2*n*, 3*n*, ⋯) in order to compare the mean square fluctuation for every residue between the original structure and the modified one. It should be noticed that when we take one of the 16 substructures as master structure, the rest of them are regarded as slave structures in each routine. In this way we can solve eigenvalue problems of large molecular structure with 1/16th smaller size of system matrix; otherwise, the problem may not be solved due to memory limitation of the computational device.

3. Numerical results and discussions

We considered hemoglobin (Hb) and motor proteins as model proteins for verification of modeling biomolecules by using mechanical the mass-spring model and employed experimental protein data deposited in the Protein Data Bank (www.protein.org). In particular, both PDB code, 1 bbb for closed form of Hb and PDB code 1e79 for F_1-ATPase motor protein are taken into account for evaluation of modified protein model. In Fig. 5, the verification of the original full MS model was performed by comparing the B factor of MS model with that of the experimental data. The results show that the full MS model is moderately reliable since it predicts the thermal fluctuation behavior of model proteins qualitatively comparable to that of experimental data. In this respect, the MS model allows us to replicate the thermal fluctuation comparable to experimental data obtained by X-ray crystallography as well as to understand the dynamics through the normal modes. The abscissa in x-axis label indicates each atom (residue) in the protein. The robustness of the modified MS model was also validated in two ways; one was done by comparing the eigenvalues of original full MS model to that of modified MS model in Fig. 6, and the other is that the pattern of B factors of modified MS model was compared to those of the experimental B factor by X-ray crystallography in Fig. 7. It is noted that even though the modified MS model permits us to reduce the computation on eigenvalue problems, it is quite remarkable in that both results of Figs. 6 and 7 demonstrate the feasibility of the modified MS model to identify the dynamics of the protein structure. According to Fig. 8, the results show that the hierarchically modified MS model, whose degree of freedom is *N*/16, predicts the fluctuation behavior generated by the original MS model. Specifically, the amplitude of fluctuation for the modified MS model is higher than initial MS model because the model modification is based on the removal of harmonic springs in the initial MS model so as to soften the molecular structure when the model modification is implemented. Nevertheless, the amplitude, representing the amount of mean square fluctuation, may not be significant for understanding the dynamic behavior of MS model, since the MS model assumes the appropriate single force constant parameter γ which is obtained from the data fitting to experimental data. In order to have comparable predictions of thermal fluctuation, for the modified MS model, we need to fix the single force constant parameter accordingly. Results presented in Fig. 9 represent that the modified MS model exhibits

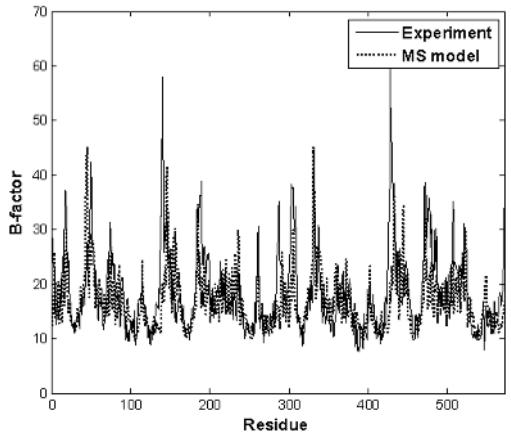


Fig. 5. Comparison of B factors of protein (Hb) obtained by experimental and theoretical results.

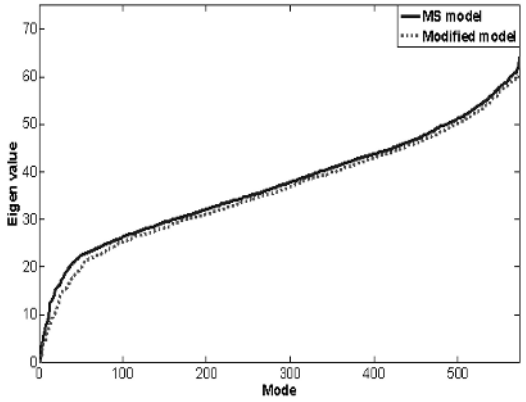
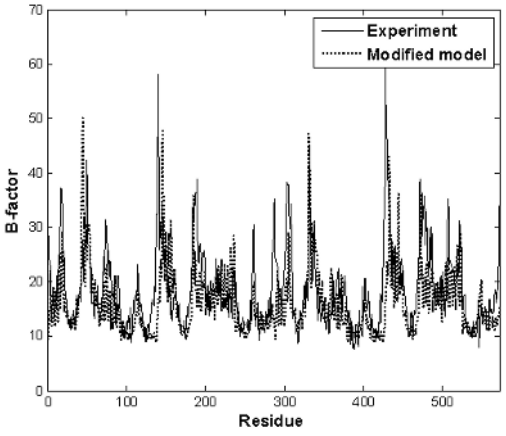
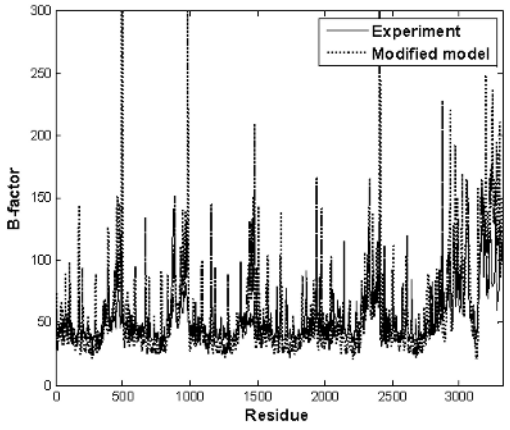


Fig. 6. Comparison of eigenvalue of the original MS model and modified model.

similar characteristics of first eigenmode for collective motions to that of initial MS model. According to the fact that Hb is known to have collective or similar motion of four domains (monomers) of hemoglobin, the low-frequency mode from the modified MS model is able to describe the collective motion of four domains of hemoglobin [Xu, 2003]. That is, hemoglobin is considered to have four domains: domain 1 (residue index (alpha carbon index): 1-141) and domain 2 (residue index: 142-286), referred to as substructure A, exhibit the correlated motion, and also domain 3 (residue index: 287-427) and domain 4 (residue index: 428-572), referred to as substructure B, have a correlated motion. This collective behavior of the low-frequency mode is well replicated by the solution of the condensed eigenvalue problem of the modified MS model. Further, the modified MS model



(a) Protein (Hb)



(b) Motor protein

Fig. 7. Comparison of B-factors reconstructed from the modified model and experimental results.

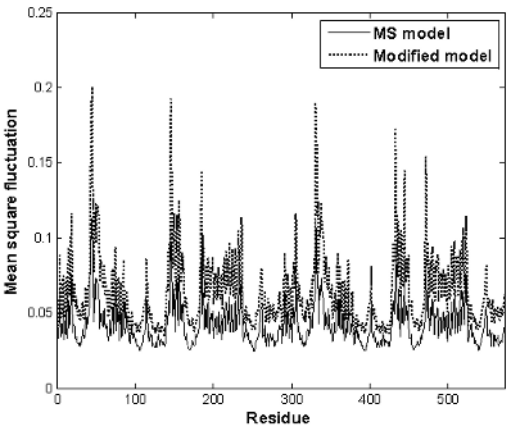
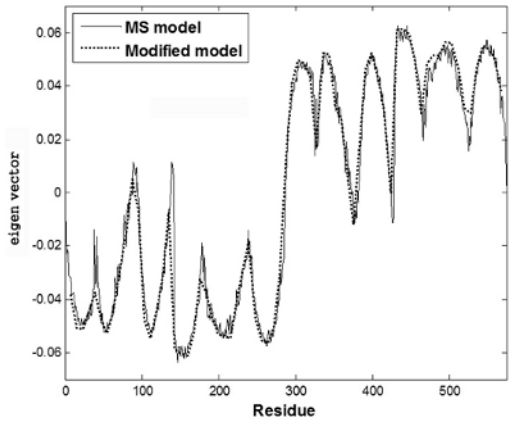
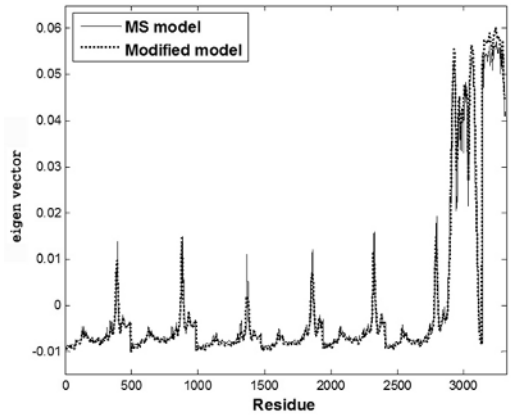


Fig. 8. Comparison of mean square fluctuations of protein (Hb) generated by original MS model and modified model.



(a) Hb obtained



(b) Motor protein obtained

Fig. 9. First eigenvector by original MS model and modified model.

can reproduce the dynamic behavior of low-frequency mode that the motion of substructure A is anti-correlated to that of substructure B. That is, the direction of the fluctuation of the substructure A is opposite to that of the substructure B [See Fig. 9(a)]. Those results suggest that the both the initial MS model and modified MS model enable one to obtain accurate predictions on the dynamics of proteins. A similar conclusion as the case of Hb is also applied for motor proteins [see Fig. 9(b)]

4. Conclusions

This paper describes a novel computational technique for understanding protein dynamics using the simple mechanical mass-spring model of biomolecular structures. Even though it is simple, it successfully demonstrates identification of the dynamics of biomolecules such as thermal fluctuations and eigenmodes with reduction of the number of degrees of freedom by as much as a factor of 16, while maintaining fidelity to the original model. Moreover, the modified dynamic MS model exhibits reliable predictions of equilibrium fluctuations and vibrational behaviors of proteins and may allow us to understand the dynamic behavior of protein structure within reasonable time on a PC. This methodology may be further applicable to supermolecules to predict their dynamics and vibrational behaviors closely related to structural conformation of proteins, which hardly can be obtained by experiment and/or other conventional approaches such as molecular dynamics and normal mode analysis due to computational memory and time.

Acknowledgments

The support by Basic Research Program of the Korea Science & Engineering Foundation (KOSEF) under grant No. R01-2007-000-10497-0 is appreciated. S. Na acknowledges the financial support of both Korea University and LG Foundation during a sabbatical year at The University of Texas at Austin. Last but not least, this work also supported by the Korea Science and Engineering Foundation grant funded by the Korea Government (MOST) (No. R11-2007-028-00000-0).

References

[1] Q. Cui, G. Li, J. Ma and M. Karplus, A normal mode analysis of structural plasticity in the biomolecular motor F (1)-ATPase, *J. Mol. Biol*. 340 (2) (2004) 345-372.

[2] F. Tama and Y. H. Sanejound, Conformation change of proteins arising from normal mode calculations, *Protein Eng.* 14 (2001) 1-6.

[3] J. Ma and M. Karplus, Ligand-Induced conformational changes in ras p21: a normal mode and energy minimization analysis, *J. Mol. Biol.* 274 (1997) 114-131.

[4] F. Tama, F. X. Gadea, O. Marques and Y. H. Sanejouand, Building block approach for determining low-frequency normal modes of macromolecules, *Proteins: Structure, Function, and Genetics* 41 (1) (2000) 1-7.

[5] J. H. Weiner, Statistical Mechanics of Elasticity, Dover Publications, INC., 2nd ed. (2002).
[6] M. P. Allen and D. J. Tildesley, Computer Simulation of Liquids, Oxford University Press (1989).
[7] M. M. Tirion, Large amplitude elastic motions in proteins from a single-parameter atomic analysis, *Phys. Rev. Lett*. 77 (1996) 1905-1908.
[8] D. Ming, Y. Kong, Y. Wu and J. Ma, Simulation of F-Actin filaments of several microns, *Biophysical Journal* 85 (2003) 27-35.
[9] B. Erman, The gaussian network model: precise prediction of residue fluctuations and application to binding problems, *Biophysical J*. (91) (2006) 3589-3599.
[10] J. Trylska, V. Tozzini and J. A. McCammon, Exploring global motions and correlations in the ribosome, *Biophys. J.* 89 (2005) 1455-1463.
[11] T. Haliloglu, I. Bahar and B. Erman, Gaussian dynamics of folded proteins, *Phys. Rev. Lett.* 79 (1997) 3090-3093.
[12] L. Meirovitch, Computational methods in structural dynamics, Sijthoff and Noordhoff, (1980).