 CMC, vol.8, no.2, pp.67-73, 2008

Application of Component Mode Synthesis to Protein Structure for Dynamic Analysis

Jae In Kim[1], Kilho Eom[2], Moon Kyu Kwak[3] and Sungsoo Na[4]

Abstract: This paper concerns the application of component mode synthesis for biomolecule modeling to understand protein dynamics. As for protein dynamics, eigenvalue problem should be formulated to obtain eigenvalue, eigenvector and thermal fluctuation. To describe the thermal fluctuation of protein, normal mode analysis is introduced and normal modes identify the dynamic behavior of protein very well. Component mode synthesis considers the given complex structure as an assembly of smaller components. The selection of a component may be arbitrary. When the component mode synthesis is applied to formulate the eigenvalue problem of protein structure, we selected a protein which may be composed of two and/or four domains. The domain of protein can be considered as a component. In this sense the number of component is increased as necessary, and the size of each component is decreased for fast calculation. The component mode synthesis widely used in engineering was well applied to understand protein dynamics in present study.

Keyword: Component mode synthesis, normal mode analysis, eigenvalue problem, protein, thermal fluctuation.

[1] Department of Mechanical Engineering, Korea University, Anam-dong, Seongbuk-gu, Seoul 136-713, Korea (e-mail: jay414@ korea.ac.kr)

[2] Corresponding author. Nano-Bio Research Center, Korea Institute of Science & Technology, Hawolgok-dong, Seongbuk-gu, Seoul 136-791, Korea (e-mail: kilhoeom@gmail.com)

[3] Department of Mechanical Engineering, Dongguk University, Pil-dong, Jung-gu, Seoul, 100-715, Korea (e-mail:kwakm@dgu.edu)

[4] Corresponding author. Department of Mechanical Engineering, Korea University, Anam-dong, Seongbuk-gu, Seoul 136-713, Korea (e-mail:nass@korea.ac.kr)

1 Introduction

Traditionally, molecular dynamics (MD) is commonly used for studying nano-scaled structures such as carbon nanotube[Chakrabarty, Tahir (2008)] and even the protein dynamics. Previous researches showed that the molecular dynamics provided some good results for understanding protein behaviors, in particular [Brooks, Karplus, Pettit (1988); McCammon and Harvey (1987)]. However, MD has some limitations such as computational loads and storage space because MD considers all atoms of model protein and implements covalent bonding, hydrogen bonding, and van der Waals interaction, etc. for potential function [McCammon and Harvey (1987)]. Since the potential function for MD is complicated, it takes a long time and needs large size of memory for computing. Moreover, for the biological function of protein structure, it needs time scale in microseconds, while MD produces the solution in time scale of nanoseconds. Somehow it is inconvenient to use MD in studying protein dynamics, which may invoke misleading errors. To overcome these problems, normal mode analysis (NMA) [Xie and Long (2006)] and elastic network model (ENM) [Tirion (1996); Haliloglu, Bahar and Erman (1997); G. Yoon et al (2008)] were introduced. Tirion developed a simple elastic model for protein's conformational fluctuations. Tirion's model gave an insight to introduce Gaussian network model (GNM) by Hailoglu *et al*. GNM regards the protein structure as a simple one-dimensional entropic spring network for alpha carbon residues. GNM introduces two parameters such as a universal force constant and a cut-off radius for implementing a one-dimensional harmonic entropic spring network for protein structures. GNM allowed for

fast computing on the normal modes and fluctuations for protein dynamics. In spite of its fast computation on the normal modes, GNM shows some computational inefficiency of large stiffness matrix for complex proteins. To overcome this inefficiency, many researchers introduced various breakthroughs. In this regard, among them, Ming *et al*. introduced Substructure Synthesis Method for simulating repeated protein structure like F-actin [Ming, Kong, Wu, Ma (2003)].

Eom *et al* introduced dynamic condensation method for dynamic analysis of biomolecules and provided the hierarchical model reduction of protein structure to build the low-resolution structures consisting of the minimal number of atoms for the studies of protein dynamics [Eom, Baek, Ahn and Na (2007)]. In this paper the component mode synthesis was applied to Hemoglobin protein structure associated with GNM to understand protein dynamics, which is not restricted to repeated protein structure. A modified implementation of the boundary conditions at the interface of adjacent biomolecular substructures has been incorporated. The proposed method was applied to model protein successfully, and it is remarkable that the component mode synthesis is capable of reproducing the dynamics of protein structures.

2 Gaussian Network Model

Gaussian network model is a simple mass-spring model for investigating fluctuation dynamics of protein. The position of nodes for the specific obtained by experiment was deposited in protein data bank and the springs connecting the nodes are representative of the bonded and non-bonded interactions between residues. The nodes which are inside interaction range called cutoff radius r_c are connected by spring. The force constant is taken to be uniform for all springs. The cutoff radius is taken as 7Å in present study.

The Gaussian network model assumes that the protein is fluctuating about the equilibrium state. The potential field of folded protein for GNM is

$$P = \frac{\gamma}{2}\sum_{i,j}^{N}\left(\mathrm{r}_{ij} - r_{ij}^{0}\right)^2 H\left(r_c - r_{ij}^{0}\right) \tag{1}$$

where $H\left(r_c - r_{ij}^{0}\right)$ is the Heaviside step function, i.e. $H\left(r_c - r_{ij}^{0}\right) = 1$ if $r_c \geq r_{ij}$, otherwise $H\left(r_c - r_{ij}^{0}\right) = 0$. Herein,r_{ij} is the distance between residues i and j and $\gamma$is the spring constant. The potential energy can be approximated by a harmonic expansion, which results in [Haliloglu, Bahar, Erman, (1997)]

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

where $u_i$is the fluctuation of residue i, and Γ_{ij} is the ij-th element of connectivity matrix implying the interaction between residue i and j, defined as [Qiang, Bahar (2006)]

$$\Gamma_{ij} = \begin{cases} -1 & \text{if } i \neq j \text{ and } r_{ij} \leq r_c \\ 0 & \text{if } i \neq j \text{ and } r_{ij} > r_c \\ -\sum\limits_{j,i\neq i}\Gamma_{ij} & \text{if } i = j \end{cases} \tag{2b}$$

The fluctuation of proteins can be described by eigenvalue problem

$$\gamma\Gamma_{ij}e_j = \omega^2 e_i \tag{2c}$$

where ω is the natural frequency and e_i is the normal mode.

In this context, the fluctuation matrix Q is defined from statistical mechanics theory [Haliloglu, Bahar, Erman, (1997)] as

$$Q_{ij} = \left\langle \left(r_i - \left\langle r_i^0 \right\rangle\right)\cdot\left(r_j - \left\langle r_j^0 \right\rangle\right)\right\rangle = \sum_{n=2}^{N}\frac{kT}{\gamma\omega_n^2}e_i^{(n)}e_j^{(n)} \tag{3a}$$

Here $\langle\rangle$ denotes ensemble average, k is the Boltzmann constant, T is the absolute temperature, ω_n is the natural frequency for the n-th mode, N is the total number of residues and summation excludes the one zero-mode associated with rigid body motions. The mean square fluctuation of residue is given by

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

3 Component Mode Synthesis

In this paper, we used component mode synthesis for protein analysis, which was well defined in structural engineering [Hale, Meirovitch (1980)]. Component mode synthesis (CMS) was introduced for solving very large structural dynamics problems, where the structure consists of several natural components, for example, fuselage, wings of airplanes or the space shuttle orbiter and its payloads. The basic concept of component mode synthesis is to consider the structure as an assemblage of components. CMS describe the motion separately over each of the domain, referred to as component by generating a component eigenvalue problem and then constrain the components to work together as a single original structure by enforcing geometric compatibility at the interface of adjacent components [Craig, R (1981); Meirovitch, L. (1980)]. Since each component is modeled separately, there are redundant coordinates, as atoms shared by two adjacent components behave the same motions. The removal of redundant coordinates is carried out during an assembling process in which the constituent components are constrained to act as a whole structure. The efficiency of CMS lies in the fact that the scheme only deals with an eigenvalue problem for a much smaller component compared with original structure. The way to choose substructure is arbitrary, though, however, in the present paper, the substructure is chosen on the base of domain (substructure) of the target proteins, which consists of several domains depending on model protein. 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. In this respect, let us consider a given component s and write the displacement vector u_s of an arbitrary point P on the component [Refer to Fig. 1]. Physical displacement vector is represented as series of space-dependent functions multiplied by time-dependent generalized coordinates as below:

$$u_s(P,t) = \Psi_s(P) q_s(t) \tag{4}$$

where Ψ_s may be regarded as assumed eigenvector of specific component, s and $q_s(t)$ is generalized coordinate.

The kinetic energy of individual component can be defined using the generalized coordinates as

$$T_s = \frac{1}{2} \dot{q}_s^T M_s \dot{q}_s \tag{5}$$

where M_s is component mass matrix.

Similarly, we also have potential energy as

$$V_s = \frac{1}{2} q_s^T K_s q_s \tag{6}$$

where K_s is a component stiffness matrix.

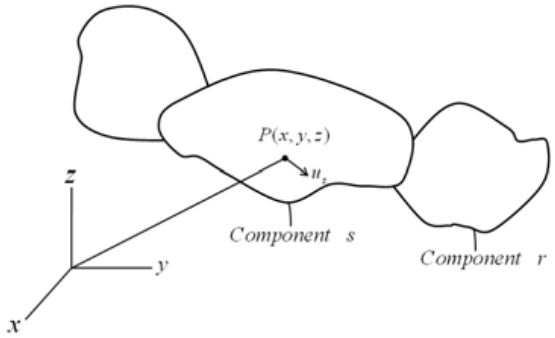


Figure 1: Configuration of a structure with components

By means of Lagrange's equation [Meirovitch, L. (1980)], which can be written as

$$\frac{d}{dt}\left(\frac{\partial T_s}{\partial \dot{q}_s}\right) - \frac{\partial T_s}{\partial q_s} + \frac{\partial V_s}{\partial q_s} = Q \tag{7}$$

component equations of motion can be obtained after introducing Eq. (5) and (6) into Eq. (7).

$$M_s \ddot{q}_s(t) + K_s q_s = Q \tag{8}$$

Herein, generalized force Q is assumed to be zero.

The assembling process can be invoked by introducing the matrices

$$M_d = \begin{bmatrix} M_1 & & & \\ & M_2 & & \\ & & \ddots & \\ & & & M_s \end{bmatrix} \tag{9a}$$

$$K_d = \begin{bmatrix} K_1 & & & \\ & K_2 & & \\ & & \ddots & \\ & & & K_s \end{bmatrix} \tag{9b}$$

The corresponding equations of motion for assembled components are written as

$$M_d\ddot{q}(t) + K_d q(t) = 0 \tag{10}$$

The displacement vector $q(t)$ includes a certain number of redundant values due to constraints of boundaries between adjacent components. Assuming that there are N_c constraints, and $q(t)$ has dimension M, then the number of independent generalized coordinates is $n = M - N_c$. At this point, coordinate transformation between $q(t)$ and $\eta(t)$ is related as

$$q(t) = B\eta(t) \tag{11}$$

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

$$u_r = u_s \tag{12}$$

Moreover, the translational displacements at the interfaces are related to the generalized displacement vector $q(t)$ In view of this, one can combine Eq. (12) corresponding to all interfaces into a single constraint equation as

$$Tq = 0 \tag{13}$$

in which T is a $N_c \times M$ matrix. Then dividing the vector $q(t)$ into vector $\eta(t)$ of independent variables and a vector p of dependent variables and partitioning the matrix T as follows:

$$T = [T_1 \vdots T_2] \tag{14}$$

Eq. (13) can be rewritten as

$$T_1\eta + T_2 p = 0 \tag{15}$$

which yields

$$p = -T_2^{-1}T_1\eta \tag{16}$$

Eq. (16) permits one to write a relation Eq. (11) between the η(t) of independent generalized coordinates for the full structure and the vector q.

$$B = \begin{bmatrix} I \\ \cdots \\ -T_2^{-1}T_1 \end{bmatrix} \tag{17}$$

where I is a unit matrix of order n.

The constraint condition allows one to integrate the components to assembled structure. Introducing Eq. (11) into Eq. (10) and premultiplying by B^T, the resulting equation of motion is expressed as

$$M\ddot{\eta}(t) + K\eta(t) = 0 \tag{18}$$

where

$$M = B^T M_d B, \quad K = B^T K_d B \tag{19}$$

Constraints

Using the component mode synthesis, geometric compatibility condition or constraint has a dominant role on the method. Geometric compatibility condition is already shown in Eq. 12, however, proper modification should be applied to the protein structure. While the component mode synthesis was applied in engineering parts, there is a single constraint point between two substructures and/or components. However, protein structure is very complex structure based on connectivity between several mass points in boundary. For the Hemoglobin structure, 36 pairs (72 constraint points) of constraint points based on $r_c = 6$ are found. Since the protein structure is complicated there are some mass points in one domain at boundaries which are connected with two or more other mass points in other domain. If we consider all of these constraints, formulation of boundary constrains should be modified such that during the process of Eq.19, while assembling the components, the mass matrix should be identity matrix. In Fig.1 (c), 23 pairs of constraint points are described in green circle and squares. Circle

and squares are the connected constraint pairs, so total of 46 constraint points are shown in Fig. 1 (c). Four colors of the protein structure, cyan, magenta, yellow, and key is describing the different domain of Hemoglobin, which are considered as components.

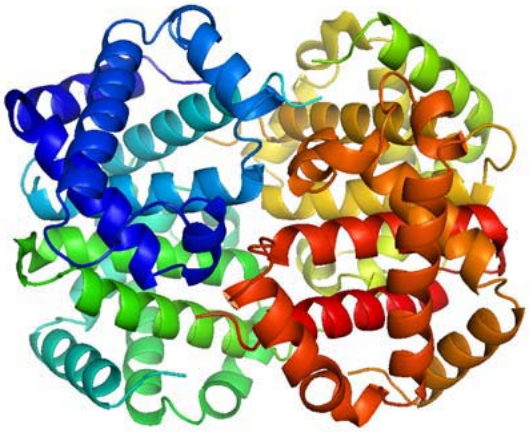
4 Model Protein

Hemoglobin was chosen as model protein in present study. Hemoglobin transports oxygen from the lungs to the rest of the body, such as to the muscles. Hemoglobin has two stable states and the closed form of Hemoglobin (pdb code: 1a3n) was selected as model protein. The structure has 572 dominant residues, and those residues were considered mass points for Gaussian network model.

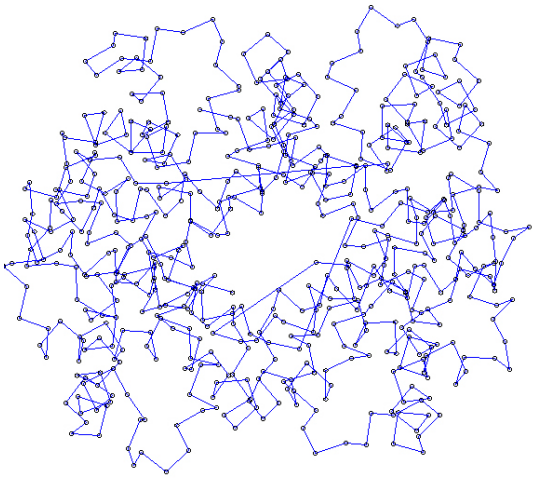
5 Simulation Results

Fig.2 displays the Hemoglobin model. Fig.2 (a) shows the molecular structure of Hemoglobin, and Fig. 2 (b) represents the mass-spring model of dominant atoms in GNM. Hemoglobin consists of 4 domains such as A, B, C and D. Both A and C domains possess 141 residues, respectively, while each of B and D domains includes 145 residues. For the present study, the two groups of simulations are conducted; first considering the Hemoglobin structure as two components, and secondly Hemoglobin as four components. For the one simulation, A and B domain were the first component, while C and D domain were the second component. For the other simulation, A, B, C and D were taken as individual component. For the former simulation, 11 mass points (residues) were used for the constraint condition. And total of 23 mass points were used for the constraint condition in the later simulation. Fig. 2 (c) displays constraint points at boundaries between adjacent components.

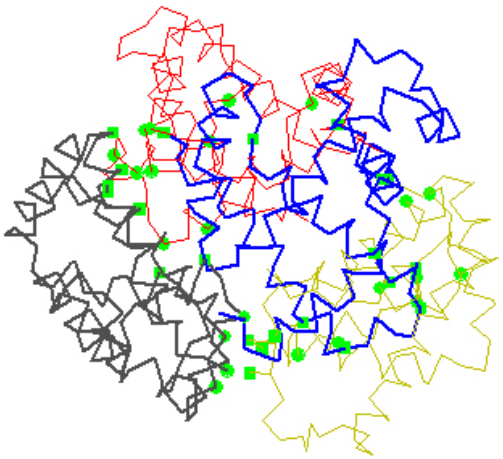
The eigenvalue are compared by both original GNM and component mode synthesis based model in Fig. 3. The results compared the original GNM structure with 2 component based structure and 4 components based structure using component mode synthesis.



(a) Molecular structure



(b) Mass- spring model



(c) Constraint points (dotted points) and four substructures (different colors)

Figure 2: Hemoglobin model

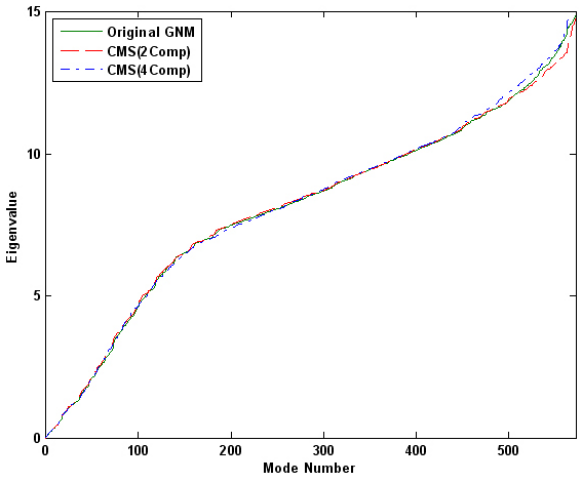


Figure 3: Comparison of eigenvalues of original GNM with ones of component mode synthesis based model

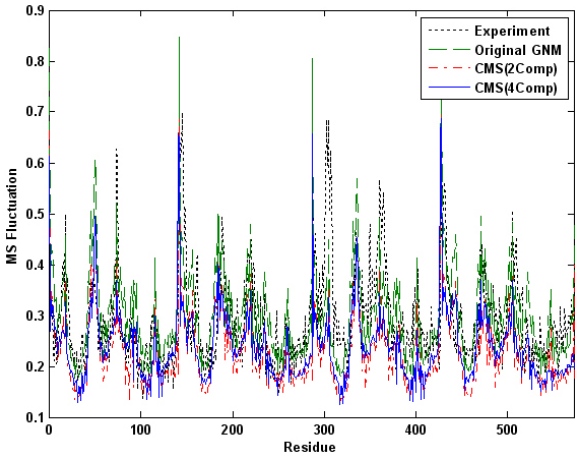


Figure 5: Comparison of mean square fluctuation of X-ray crystallography, GNM modeling and component mode synthesis.

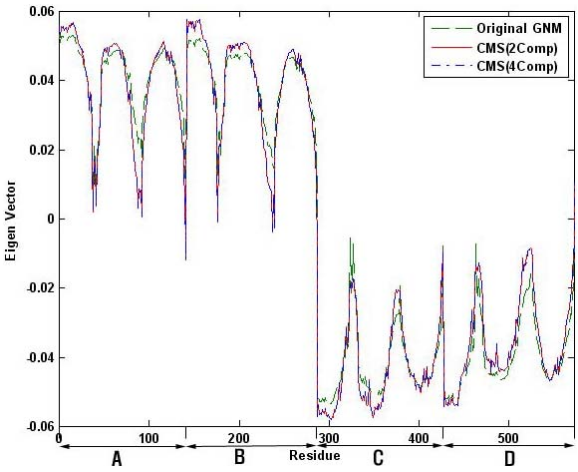


Figure 4: Comparison of lowest frequency mode for hemoglobin structure

It is shown that the eigenvalues from three different configurations are showing incredibly similar results, especially for the lowest modes, even some variance in the high frequency modes though. Since low frequency modes play dominant role on protein dynamics, the component mode synthesis provides quantitatively comparable results to original GNM.

The primary low-frequency normal mode that is generally renowned to play a role in protein dynamics such as conformation change was considered to prove the robustness of component mode synthesis. In Fig. 4 the primary low-frequency normal mode obtained by both original GNM and component mode synthesis. It is shown that the characteristics of lowest-frequency normal mode are well preserved in the proposed method. It is also observed that A domain and C domain has the opposite motions and same trend for B and D domain.

The mean-square fluctuations of model proteins are compared by both GNM and component mode synthesis in Fig. 5. It is remarkable that component mode synthesis provides the mean-square fluctuation qualitatively comparable to the one obtained by both GNM and experiment one, even though component allows one to reduce the computational burden on the mean-square fluctuation.

6 Conclusion

The present paper shows that it is very successful for predicting thermal fluctuation of protein structure via component mode synthesis. The component mode synthesis may suggest further high frequency mode reduction which might be applicable to the large protein structures that are hardly approachable with traditional method such as normal mode analysis. This method may provide hierarchical substructure model allow to construct low-frequency based model which is dominant for

protein fluctuations. For further study, component mode synthesis may enable to study dynamics of biological supra large molecular structures. Specially, the dynamics which are not accessible by conventional NMA may be possible to approach with using component mode synthesis.

Acknowledgement: This work was supported by Basic Research Program of the Korea Science and Engineering Foundation (KOSEF) under grant No. R01-2007-000-10497-0 and R11-2007-028-00000-0.

References

Arnab Chakrabarty; Tahir cagin (2008): Computational Studies on Mechanical and Thermal Properties of Carbon Nanotube Based Nanostructures , *CMC: Computers, Materials & Continua*, Vol. 7, No. 3. pp. 167-190.

Brooks, C.L.; Karplus, M.; Pettit, B.M. (1988): A Theoretical Perspective of Dynamics, Structure, and Thermodynamics, *Advances in Chemical Physics*, 71, 1.

Craig, R. Jr. (1981): *Structural Dynamics An introduction to computer methods*, John Wiley & Sons.

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.

Doruker, P.; Jernigan, R.L.; Bahar, I. (2002): Dynamics of large proteins though hierarchical levels of coarse-grained structures, *J. Comput. Chem*., 23(1), 119-127.

Eom, K.; Baek, S.; Ahn, J.; Na, S. (2007): Coarse-Graining of protein structures for the normal mode analysis, *J. of Computational Chemistry*, 28(8), pp. 1400-1410.

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

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.

Haliloglu, T.; Bahar, I.; Erman, B. (1997): Gaussian dynamics of folded protein, *Phys. Rev. Lett*., 79, pp.3090-3093.

McCammon, J. A.; Harvey, S. (1987): *Dynamics of proteins and nucleic acids*, Cambridge University Press, Cambridge.

Meirovitch, L. (1980): *Computational methods in structural dynamics*, Sijthoff and Noordhoff.

Ming, D.; Kong, Y. ; Wu, Y. ; Ma, J. (2003): Substructure synthesis method for simulating large molecular complexes, *Proceedings of the National Academy of Sciences*, 100(1), pp. 104-109.

Qiang Cui; Ivet Bahar (2006): *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems*, Chapman & Hall.

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

Tirion, M. M. (1996): Large amplitude elastic motions in proteins from a single-parameter atomic analysis, *Phys. Rev. Lett*., 77, pp.1905-1905.

Weiner, J. H. (2002): *Statistical Mechanics of Elasticity*, 2^{nd} ed. Dover Publications, INC.

Xie, G.Q.; Long, S.Y. (2006): Elastic vibration behaviors of carbon nanotubes based on micropolar mechanics, *CMC: Computers, Materials & Continua*, pp.11-20.

Yoon, G.; Park, H; Na, S.; Eom, K. (2008): Mesoscopic model for mechanical characterization of biological protein materials, *J. of Computational Chemistry*, advanced online, DOI: 10.1002/jcc21107.