

# 단백질의 동적특성해석을 위한 전산해석기법 연구 Computational Methodology for Biodynamics of Proteins

안정희\*·장효선\*·엄길호\*\*·나성수† \*

Jeonghee AHN, Hyoseon Jang, Kilho EOM and Sungsoo NA

**Key Words** : Protein Dynamics(단백질 동역학), Elastic Network Model(탄성네트워크모델), Model Reduction(모델축소), Conformational Fluctuation(변형거동), Low-Frequency Normal Modes(저차 직교모드)

## ABSTRACT

Understanding the dynamics of proteins is essential to gain insight into biological functions of proteins. The protein dynamics is delineated by conformational fluctuation (i.e. thermal vibration), and thus, thermal vibration of proteins has to be understood. In this paper, a simple mechanical model was considered for understanding protein's dynamics. Specifically, a mechanical vibration model was developed for understanding the large protein dynamics related to biological functions. The mechanical model for large proteins was constructed based on simple elastic model (i.e. Tirion's elastic model) and model reduction methods (dynamic model condensation). The large protein structure was described by minimal degrees of freedom on the basis of model reduction method that allows one to transform the refined structure into the coarse-grained structure. In this model, it is shown that a simple reduced model is able to reproduce the thermal fluctuation behavior of proteins qualitatively comparable to original molecular model. Moreover, the protein's dynamic behavior such as collective dynamics is well depicted by a simple reduced mechanical model. This sheds light on that the model reduction may provide the information about large protein dynamics, and consequently, the biological functions of large proteins.

## 1. 서론

단백질의 생물학적 역할을 규명하기 위해서 동적 특성을 파악하는 것은 필수적이다. 단백질 동역학은 구조적 섭동 에 의해 기술되었다. 즉 단백질의 열적 진동이 파악되어야 하며, 이 논문에서 단백질의 구조적 섭동과 그에 관련된 역할을 이해하기 위해 단순한 기계적 모델이 고찰되었다. 특히 생물학적 역할에 관련된 거대 단백질의 동적 특성을 파악하는 진동모델이 개발되었다.

거대 단백질에 대한 기계적 모델은 간단한 탄성모델과 모델축소기법을 기초로 하여 구성되었다. 거대 단백질 구조는 단백질의 동역학적 특성을 유지하는 범위에서의 최소단위화 구조를 근간으로 하는 모델축소기법을 적용하여 최소의 자유도로 표현되었다. 이 모델에서 간단하게 축소된 모델을 통해 재구성된 열 섭동 형상이 원래의 분자모델과 상당히 유사한 결과를 얻을 수 있었다. 또한 축소된 기계적 모델을 통해 집합적 운동과 같은 단백질의 동적 성질이 잘 묘사되었다. 이러한 모델축소는 거대 단백질 동역학과 생물학적 역할에 대한 정보를 효과적으로 제시할 것으로 사료된다.

## 2. 전산해석을 위한 단백질의 모델링과 시뮬레이션

### 단백질 모델링

Table. 1 과 같이 단백질 모델을 선정하고 Coarse-graining 기법을 적용하여 단백질 구조에서 지배적인 역할을 하고 있는 alpha carbon( $C_\alpha$ )만을 고려하여 Fig. 1b 과 같은 시뮬레이션을 위한 효과적인 모델링을 구성할 수 있다.

Table. 1 Model Proteins

Protein	Pdb code	$N$	$N_d$	$r_c(\text{\AA})$
Retinol binding	1aqb	175	1	12
Dehydrogenase	1ldm	572	4	12
Aploctoferrin	1lfg	858	2	12

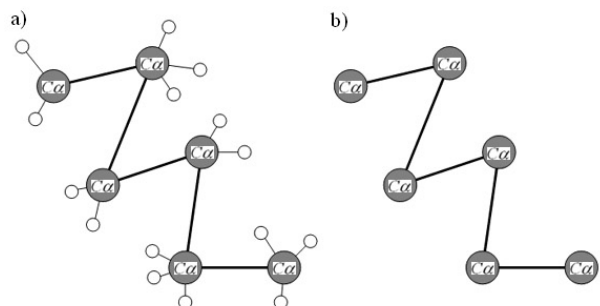


Fig. 1 (a) Simple protein model with all atoms. (b) Simple protein model with  $C_\alpha$  only.

† 교신저자; 고려대학교 기계공학과

E-mail : nass@korea.ac.kr

Tel : (02) 3290-3854, Fax : (02) 6008-3855

\* 고려대학교 기계공학과

\*\* 한국과학기술연구원(KIST)

### 탄성네트워크모델

단백질구조에서 지배적인 역할을 하고 있는  $C_\alpha$  에 대해서 기계공학적 질량, 스프링을 이용하여 포텐셜 에너지(E)를 구할 수 있고, 다음과 같이 정의된다.

$$E = \sum_i \sum_j \frac{\gamma}{2} (r_{ij} - r_{ij}^0)^2 H(r_c - r_{ij}^0) \quad (1)$$

여기서,  $r_{ij}$ 는 i, j 번째의  $C_\alpha$  간의 거리,  $\gamma$ 는 스프링 상수,  $r_c$  (cutoff radius)는 10~12Å,  $H(r)$ 은 Heaviside step function 이고, 이를 통하여 탄성네트워크의 1 차원 모델인 GNM(Gaussian network model)의 포텐셜 에너지를 구할 수 있다.

$$E = \frac{\gamma}{2} \sum_i \sum_j \Gamma_{ij} u_i u_j \quad (2)$$

여기서,  $u_i$ 은 i 번째  $C_\alpha$ 의 fluctuation,  $\Gamma_{ij}$ 는 다음과 같이 정의된다.

$$\Gamma_{ij} = -(1 - \delta_{ij}) H(r_c - r_{ij}^0) - \delta_{ij} \sum_{k \neq i} \Gamma_{ik} \quad (3)$$

$\delta_{ij}$ 는 Kronecker delta 이다. 구조가 변하면서 생기는 거동(fluctuation)은 고유치 문제에 의하여 결정된다.

$$\mathcal{H} \Gamma_{ij} q_j = \omega^2 q_i \quad (4)$$

$\omega$ 는 고유진동수,  $q_i$ 는 직교 모드(normal mode)이며 이를 통하여 거동 행렬(fluctuation matrix)  $Q_{ij}$ 가 다음과 같이 구성된다.

$$Q_{ij} \equiv \langle (r_i - \langle r_i \rangle) \cdot (r_j - \langle r_j \rangle) \rangle \\ = \sum_{n=2}^N \frac{kT}{\rho \omega_n^2} q_i^{(n)} q_j^{(n)} \quad (5)$$

여기서,  $\langle \rangle$ 는 전체모드의 평균값, n은 n 번째의 모드, 그리고 한 개의 zero 모드가 강체 모드에 의하여 존재한다.

평형상태에서 열에너지로 인한 i 번째  $C_\alpha$ 의 거동은

다음과 같고,

$$\langle (\Delta r_i)^2 \rangle = Q_{ii} \quad (6)$$

Debye-Waller factor(B-factor)은 다음과 같이 정의된다.

$$B_i = \frac{8\pi^2}{3} Q_{ii} \quad (7)$$

각  $C_\alpha$  사이의 움직임에 따른 상관관계는 다음과 같이 정의된다.

$$C_{ij} \equiv \frac{\langle \Delta r_i \cdot \Delta r_j \rangle}{\sqrt{\langle (\Delta r_i)^2 \rangle \langle (\Delta r_j)^2 \rangle}} = \frac{Q_{ij}}{\sqrt{Q_{ii} Q_{jj}}} \quad (8)$$

$C_{ij}$  값이 1에 가까우면 correlated 모드, -1에 가까울수록 anti-correlated 모드이다. 그리고 직교모드에서 집합성(Collectivity)이 다음과 같이 정의된다.

$$\kappa_i = \frac{1}{N} \exp \left[ - \sum_{j=1}^N |q_j^{(i)}|^2 \log |q_j^{(i)}|^2 \right] \quad (9)$$

위의 식에서  $\kappa_i$  값의 범위는  $1/N < \kappa_i < 1$ ,  $\kappa_i$ 는 전체움직임을 좌우하는 확률을 의미한다.

### 3. 효율적인 전산해석을 위한

#### 단백질구조의 축소 기법

거대 단백질을 해석할 경우, 계산에 따른 많은 어려움이 있다. 이러한 문제점을 해결하기 위하여 단백질 구조를  $N/n$ 으로 축소화 하고, n번 반복하여 전체시스템 N을 재구성 할 수 있음을 밝혀냈다. 전체 시스템으로부터 추출된  $N/n$ 은 master 그룹이 되고, 나머지  $N(1-1/n)$ 은 slave 그룹이 된다. 따라서 master와 slave 그룹으로 이루어진 전체 시스템의 포텐셜 에너지는 다음과 같이 정의될 수 있다.

$$E = \frac{\gamma}{2} u^* \Gamma u = \frac{\gamma}{2} [u_1 \ u_2] \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \quad (10)$$

여기서, 첨자 1 과 2 는 각각 master 와 slave 를 지칭한다.  $\Gamma_{11}$  는 master 집합체 안에 들어있는  $C_\alpha$  사이의 pair-wise interaction 을 의미하고,  $\Gamma_{12}$  는 master 와 slave 집합체의  $C_\alpha$  사이의 pair-wise interaction 을 의미한다. slave  $C_\alpha$  들의 거동(fluctuation)은 시스템에 고려되지 않으므로 다음과 같이 평형상태로 가정한다.

$$\frac{\partial E}{\partial u_2} = \Gamma_{21}u_1 + \Gamma_{22}u_2 = 0 \quad (11)$$

따라서,  $C_\alpha$  들의 거동(fluctuation)은 오직 master 그룹의  $C_\alpha$  만으로 다음과 같이 정의할 수 있다.

$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} I_m \\ -\Gamma_{12}\Gamma_{22}^{-1} \end{bmatrix} u_1 \quad (12)$$

$I_m$  은  $m \times m$  항등 행렬이고,  $m$  은 master 안의  $C_\alpha$  의 개수이다. 식 (4), (10), (12)를 통하여  $\tilde{\Gamma}$  를 재구성 할 수 있다.

$$\gamma \tilde{\Gamma}_{ij} q_j = \omega^2 q_i \quad (13)$$

$$\tilde{\Gamma} \equiv L[\Gamma] = \begin{bmatrix} I_m & -\Gamma_{12}\Gamma_{22}^{-1} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix} \begin{bmatrix} I_m \\ 0 \end{bmatrix} \quad (14)$$

Coarse-grained 구조의 동역학적 특성인 거동(fluctuation), cross-correlation, 집합성(collectivity)는 식 (5)-(9), (13), (14)를 이용하여 구할 수 있다.

## 4. 결과

### B-factor

단백질의 거동은 X-ray crystallography, 전자현미경을 통하여 얻을 수도 있지만 간단한 탄성 네트워크모델로 시뮬레이션을 하여 얻을 수도 있다. 그림. 2 는 단백질 모델에 따른 실험 치와 시뮬레이션 결과를 비교하였다. 실험결과, 일반적인 GNM 방법 그리고 단백질 모델축소기법(mass condensation)을 비교하였을 때, 같은 경향성을 파악할 수 있었고 따라서 단백질 구조에

탄성네트워크 모델을 이용한 모델축소기법을 적용시킬 수 있음을 보였다.

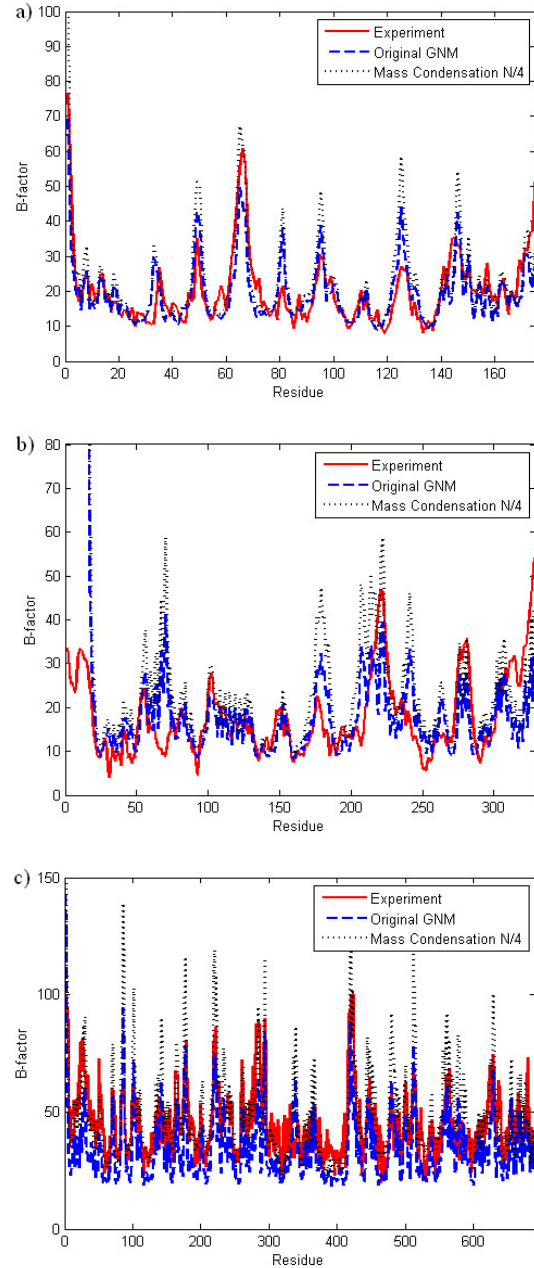


Fig 2. Comparison of B-factor obtained from the coarse-grained structure retaining N/4 alpha carbons with experimental B-factor by X-ray crystallography: (a) Retinol Binding (pdb code: 1aqb), (b) Dehydrogenase (pdb code: 1ldm), (c) Aploactoferrin (pdb code: 1lfg)

### Lowest-Frequency Normal Mode

본 연구에서 저차 모드(low-frequency mode)에서 단백질의 동적 특성이 가장 활발하게 일어나는 것을 볼 수 있었다. 그림 3 을 통하여 시뮬레이션에 따른 저차 모드를 비교한 것이다. 결과에서 알 수 있듯이 Coarse-graining 이 되는 동안에도 저주파 모드의 특성이 보존되는 것을 알

수 있다.

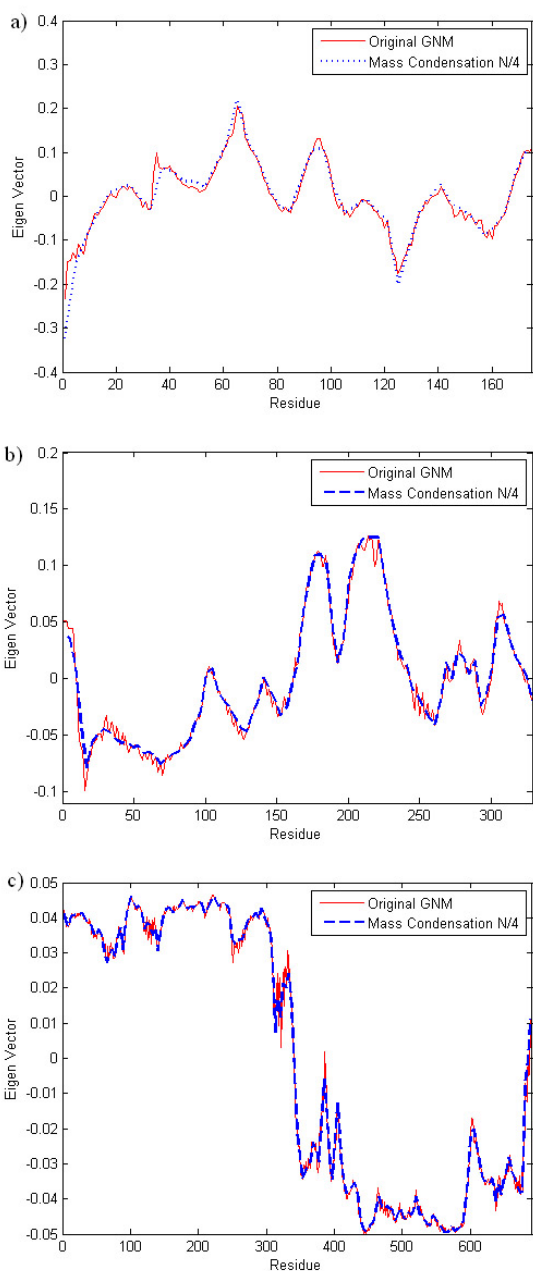


Fig 3. Lowest-frequency normal mode obtained from the original structure and the coarse-grained structures for model proteins: (a) Retinol Binding (pdb code: 1aqb), (b) Dehydrogenase (pdb code: 1ldm), (c) Aploactoferrin (pdb code: 1lfg)

## 5. 결론

본 논문에서는 GNM 을 이용하여 단백질의 동적 특성을 해석하는 방법이 소개되었다. 거대단백질 의 동적 해석시의 어려움을 해결하기 위하여 채택 된 단백질 구조축소기법을 적용한 시뮬레이션 결과 단백질 본래의 동적 특성이 유지되는 것을 알 수

있었다. 따라서, 구조해석의 어려움이 있는 거대단백질을 해석 할 때 계산의 효율성을 높이기 위한 하나의 방법으로 Low-Frequency Normal Mode 를 기반으로 단백질구조축소기법을 적용 할 수 있다.

## 6. 후기

This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MOST) (No.R11-2007—028-00000-0). S. Na also acknowledge the support y Basic Research Program of the Korea Science & Engineering Foundation (KOSEF) under grant No. R01-2007-000-10497-0.

## 참고문헌

- (1) Kilho Eom, S.-C. B. J.-H. A. S. N. (2007). "Coarse-graining of protein structures for the normal mode studies." *Journal of Computational Chemistry* 28(8): 1400-1410.
- (2) Brooks, C.L.; Karplus, M.; Pettit, B. M. *Adv Chem Phys*, 71, 1.
- (3) McCammon, J. A.; Harvey, S. *Dynamics of proteins and nucleic acids*; Cambridge University Press: Cambridge, 1987.
- (4) Elber, Q.; Bahar, I. *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems*; CRC Press: Boca Raton, FL, 2005.
- (5) Elber, R. *Curr Opin Struct Biol* 2005, 15, 151.
- (6) Hayward, S.; Go, N. *Annu Rev Phys Chem* 1995, 46, 223.
- (7) Hinsen, K. *Proteins: Struct Func Genet* 1998, 33, 417.
- (8) Cui, Q.; Li, G. H.; Ma, J. P.; Karplus, M. *J Mol Biol* 2004, 340, 345.
- (9) Valadie, H.; Lacapre, J. J.; Sanejouand, Y. H.; Etchebest, C. *J Mol Biol* 2003, 332, 657.
- (10) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins: Struct Funct Genet* 1993, 17, 412