

## 생체 분자 시뮬레이션 기술

한국과학기술연구원 송진수 · 윤창노

### 1. 서론

생명현상의 규명과 질병의 원인 및 치료를 위한 분자 수준의 생물학적 연구는 생명과학자들 뿐만 아니라 이와 관련된 약학, 생화학, 화학, 생물리학, 수학, 컴퓨터 등의 다양한 분야에 있는 전문가들의 끊임없는 노력과 협력의 결과로 급속한 진전을 보이고 있다.

우리는 보통 분자생물학 연구실에서 유전인자의 염기서열 또는 추출된 단백질의 아미노산 배열을 얻게 되면 먼저 유전자 은행(GenBank)과 단백질 데이터베이스(Protein Data Bank)를 탐색하여 그들의 구조와 기능을 확인하게 되고 그들과 상호작용 또는 결합할 수 있는 리간드를 찾게 되며 얻어진 생물학적인 정보(bioinformation)를 사용하여 생화학적인 메커니즘을 이해하거나 질병의 원인 및 치료의 방법을 강구하게 된다.

이때에 우리는 생체 분자 시뮬레이션(biomolecular simulation) 기술을 사용하여 생화학적 메커니즘을 분자 수준에서 살펴볼 수 있으며, 더 나아가 에너지의 변환에 관여하는 전자의 움직임까지를 추론할 수 있게 해주기 때문에, 보고자 하는 이러한 메커니즘의 확인을 위해 다양한 변이실험, 리간드의 결합력 실험, 단백질 분자의 구조 변화 실험 등을 구상할 수 있게 해 준다.

이러한 생체 분자 시뮬레이션 기술은 컴퓨터를 사용하여 단백질, DNA와 같은 생체 거대 분자들의 삼차원적 구조와 동력학적인 운동변화를 살펴볼 수 있는 장점을 가지고 있으며, 생물정보를 이용한 바이오 컴퓨팅(biocomputing) 부분과 컴퓨터 그래픽스를 기반으로 한 분자 그래픽스(molecular graphics) 부분으로 구성되어 있다. 각 부분은 여러 요소 기술들로

이루어져 있으나 핵심이 되는 중요한 몇 가지 기술을 든다면, DNA 염기서열 또는 단백질 아미노산 배열 데이터베이스 단백질의 3차 구조 계산, 생체 분자간의 결합(molecular docking) 또는 상호작용(molecular interaction) 계산, 에너지 전달을 보이는 양자 생물학적(quantum biology) 계산, 생체 분자 동력학(biomolecular dynamics) 시뮬레이션, 생체 분자 시스템의 컴퓨터 시각화(visualization) 기술 등이다.

여기서는 생물학적 대상물질에 대한 연구를 위하여 생체 분자 구조에 대한 연구 동향과, 이렇게 축조된 구조의 동역학적 실험에 대한 역사, 운동의 형태, 그리고 이론적인 방법을 통하여 우리가 얻을 수 있는 일반적인 정보를 알아 보고 또한, 중요한 분야중의 하나로 인식되고 있는 molecular graphics system이 어떻게 분자 수준에 적용되는가를 간단히 살펴보고자 한다.

### 2. 컴퓨터 기술을 이용한 단백질 구조 연구 동향

컴퓨터를 사용하여서 단백질 구조를 얻는 방법은 크게 comparative modeling(CM), fold recognition (FR), ab initio 등으로 나눌 수 있다. 일반적으로 이러한 세 가지 방법에 의해서 예측된 구조의 정확도는 주어진 서열의 상동성의 정도에 의존한다. 서열 상동성이 30% 이상인 경우에는 주로 CM 방법을 사용하며, 정확도도 세 방법 중 가장 좋다. 서열 상동성이 20~30% 인 경우에는 FR 방법과 ab initio 방법이 사용될 수 있으며 FR 방법이 조금 더 좋은 결과를 얻는다. 서열 상동성이 20% 이하인 경우에도 FR 방법과 ab initio 방법은 서로 경쟁적이나, 이 경우에는 ab initio 방법이 조금 더 좋은 결과를 얻으며, 더 낮은 상동성을 가지는 경우에는 ab initio 방법만 가능하다.

## 2.1 비교모델링

CM 방법은 서열 정렬 algorithm을 사용하여 질의 서열과 관계된 서열과 구조를 찾고, 다중 정렬 algorithm[1, 2]을 사용하여 family의 구조와 서열을 정렬하여 구조 template를 얻고 이것을 사용하여 단백질의 구조를 결정하는 방법이다. 좋은 구조의 template를 얻기 위해서는 서열 정렬이 가장 중요하다. 2개의 서열에 대한 정렬은 pairwise alignment, 3개 이상의 서열에 대한 정렬은 multiple alignment라 하고, 상동성 형태에 따라서 global alignment와 local alignment를 구분한다. 서열 정렬은 질의 서열과 상동성이 높은 서열을 알아내어서 서열의 기능을 유추하거나, 관련 있는 서열들 간의 상관관계 구조 등을 예측하기 위해서 사용된다. 두 서열을 비교할 때 서열을 전체적으로 비교하여 최대의 상동성이 나타나도록 정렬하는 경우, 이러한 정렬을 global alignment라 한다. 반면, 두 서열이 어떤 부분의 서열이 높은 상동성을 가지는지를 알기 위해 정렬하는 것을 local alignment라 한다.

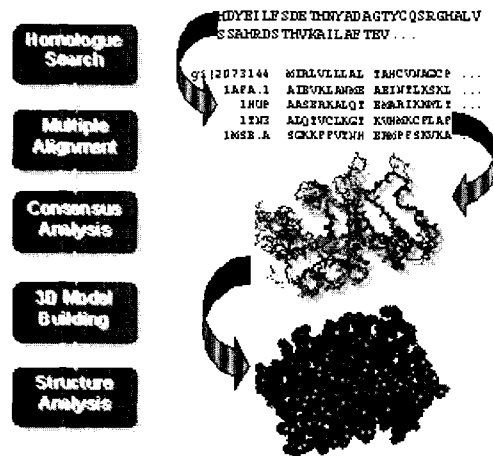


그림 1 일반적으로 사용되는 Comparative Modeling (CM) 방법

## 2.2 구조인식

FR 방법은 질의 서열에 대한 가장 좋은 상동성을 가지는 서열의 상동성 정도가 떨어져서 그 탐색된 구조를 template로 사용할 만큼의 정확도가 나오지 않

으므로, 서열로 얻어진 구조를 사용하여 구조-구조 정렬을 통하여서 template를 찾는 방법[3, 4]을 사용한다.

## 2.3 Ab initio

Ab initio 방법은 CM, FR 방법과 다르게 구조 template을 필요로 하지 않아서, 서열 상동성에 상관 없이 사용 될 수 있다. ab initio 방법은 실제 단백질 구조의 template를 사용하지 않기 때문에 임의의 단백질 구조를 생성하여야 한다. 적은 수의 아미노산으로 이루어진 단백질의 경우에도 단백질이 가질 수 있는 구조의 conformational space의 수는 천문학적인 숫자가 된다. 그래서 ab initio 방법에서는 아미노산을 하나의 원자와 같이 생각하는 unified residue system을 사용하고 이러한 unified-residue 원자로 구조를 표현하는데 lattice model[5]과 off-lattice model을 주로 사용한다. 아래 그림은 lattice model을 사용하여 단백질 구조를 표현한 한 예이다.

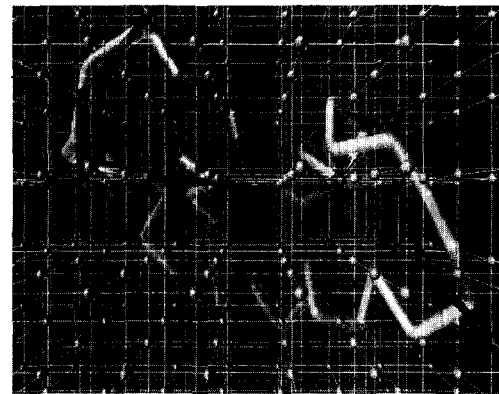


그림 2 lattice model을 사용하여 단백질 구조를 표현한 한 예(회색 구슬들은 격자를 나타내고 빨간 구슬은 격자를 이용하여 단백질의 구조를 나타낸 것이다. 녹색의 사슬이 자연계에 존재 하는 단백질의 실제 구조이다.)

## 3. 분자동역학 모의 실험

분자동역학 모사의 방법은 생체분자들의 이론적 연구를 위한 중요한 기술 중의 하나로서, 이러한 방

법은 분자의 시간에 따른 성질들을 계산한다. 분자동역학 모사는 단백질과 핵산들의 conformational 변화와 fluctuation들에 대한 자세한 정보를 제공하며, 생체 분자들이나 그들의 complex들에 대한 구조, 동역학적, 열역학적 성질들의 연구에 이제 흔하게 사용되고 있다. 이러한 방법은 또한 X-ray 결정구조와 NMR 실험으로부터 구조의 결정에도 사용된다.

분자동역학은 1950년도 후반에 Alder와 Wainwright [6]에 의하여 hard sphere들의 상호작용을 연구하기 위하여 처음으로 소개되었다. 그들의 연구로부터 간단한 용매에 대한 많은 관심이 있었으며 1964년 Rahman[7]이 아르곤 용매에 대한 실질적인 포텐셜을 사용함으로써 첫 번째 모사를 수행하게 된다. 첫 번째 단백질 모사는 1977년 BPTI(Bovine Pancreatic Trypsin Inhibitor)에 대한 모사를 McCammon[8]이 수행하였다. 오늘날은 용매화된 단백질들(Protein-DNA complexes, lipid systems, 작은 단백질들의 folding과 Ligand의 결합에 대한 열역학적인 성질들)의 분자동역학 모사를 문헌상 흔하게 찾아볼 수 있다. 분자동역학 모사의 방법들은 상당한 수에 이르렀으며, 특정 문제들에 대한 양자역학과 고전역학이 포함되는 많은 기술들이 존재하여 실제 생물학적 반응들에 적용되고 있다.

대략적인 생체분자의 시간에 따른 운동의 형태는 3가지로 구성되는데, 첫 번째 Local Motion들은  $10^{-15} - 10^{-1}$ s의 시간동안 0.01 - 5 Å의 변화를 보이며 여기에는 Atomic fluctuation들, sidechain 운동들, loop 운동들이 해당된다. 두 번째 Rigid Body Motion들은  $10^{-9}$ -1s의 시간동안 1 - 10 Å의 변화를 보이며 여기에는 Helix 운동들, Domain운동(hinge bending), Subunit 운동들이 해당된다. 그리고 세 번째 Large-Scale 운동들에는  $10^{-7} - 10^4$ s의 시간동안 >5 Å의 변화를 보이며 Helix-Coil 전이, Dissociation/Association, Folding/Unfolding 운동들이 해당된다. 분자동역학을 통한 생물계에서 발생하는 역학적인 과정과 complex의 연구는 단백질의 안정도, conformational 변화들, 단백질 folding, 분자(단백질, 핵산, membrane, 혼합물)들간의 상호작용, 그리고 생물계에서의 Ion transport를 들 수 있다.

이 방법은 생체내에서 분자들의 움직임을 연구하는 이론적인 방법 중의 하나로서, 분자의 운동을 연구하기 위해서는 분자를 구성하는 각 원자 사이의 상호작용을 묘사할 수 있는 실험적인 parameter를 얻

는 것이 중요한데, 이렇게 얻은 parameter를 potential function 또는 force field라 한다. 이들 potential function 또는 force field는 여러 과학자들에 의해 고안되었는데 많이 사용하는 potential function으로는 MM2, MM3, AMBER, CHARMM, ECEPP, CVFF 등이 있다. 이들 potential function들은 각각 다르기 때문에 취급하는 system에 따라 알맞은 것을 골라 사용해야 하기 때문에 나름대로의 장단점을 가지고 있지만 기본꼴에 있어서는 아래의 식과 같이 똑같은 형태를 갖는다.

$$E_{pot} = E_{bond} + E_{bend} + E_{torsion} + E_{nonbonded} \\ = \sum k(b - b_0)^2 + \sum k(\theta - \theta_0)^2 + \sum (1 + \cos(n\phi)) + \sum \left( \frac{1}{r} + \frac{A_{ij}}{r^{12}} + \frac{B_{ij}}{r^6} \right) \quad (1)$$

이와 같은 potential function을 사용하여 분자의 총에너지를 계산할 수 있고 원자의 상호작용 에너지 및 힘을 계산할 수 있게 된다. 분자내의 원자들의 움직임은 Verlet algorithm을 이용하여 trajectory를 계산함으로써 묘사할 수 있게 된다. 즉 시간이 t일 때 그때부터  $\Delta t$  시간이 경과한 후의 한 원자의 위치는 다음과 같이 (2)식처럼 정의된다.

$$r(t + \Delta t) = r(t) + \frac{d}{dt} r(t)\Delta t + \frac{1}{2!} \frac{d^2}{dt^2} r(t)\Delta t^2 + \dots \quad (2)$$

여기서  $\Delta t$  가 매우 작다면 (3)식과 같이 식을 간략화 시킬 수 있다.

$$r(t + \Delta t) = r(t) + \frac{d}{dt} r(t)\Delta t + \frac{1}{2!} \frac{d^2}{dt^2} r(t)\Delta t^2 \quad (3)$$

이 식에다가 Newtonian mechanics를 도입하여 식을 다시 정리하면 (4)식과 같다.

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{1}{2!} \frac{F}{m} \Delta t^2 \quad (4)$$

where  $\frac{d}{dt} r(t) = v(t)$  and  $\frac{d^2}{dt^2} r(t) = a = \frac{F}{m}$

여기서 원자가 움직이는 초기 속도는 Maxwell-Boltzmann equation을 풀어서 구할 수 있고 원자-원자 상호간의 interaction은 potential function에서 얻을 수 있다. 따라서 실제 분자가 움직이는 모습을 계산해 낼 수 있고 그렇게 해서 구해진 coordinate를 가지고 computer graphics technique을 사용하여 눈으로 확인 가능하다.

### 3.1 주기적 경계조건

분자들간의 수용액에서의 연구를 위해서 분자 주위에 물분자가 존재하도록 하여 계산한다. 이때 물분자를 무한히 많이 고려 할 수 없다. 따라서 수용액 상태에서 solute가 물 전체에 균일하게 퍼져 있고 또한 물분자는 solvent이기 때문에 어떤 부분이든지 유사성을 가진다고 생각할 수 있다. 따라서 solute을 일정한 수의 물로 싸 다음 이를 3차원으로 반복시키면 원래의 solute와 물은 확장된 부분에 의해 둘러싸이게 된다. 이때 실제 움직이는 분자는 real molecule에 해당하는 부분이고 ghost는 real molecule의 반복이고 ghost molecule은 real molecule에 힘을 줄 뿐이며 받지는 않는다. 따라서 box surface 물분자는 edge effect에 의해 물이 증기화 되는 것을 막을 수 있고 solute는 완벽하게 solvation 된 것처럼 느끼게 된다.

### 4. 분자 그래픽스 시스템

오늘날 물질과학과 함께 생명과학 분야에서도 molecular graphics system은 분자 수준에서의 구조 연구에 있어서 매우 중요한 분야로 인식되고 있다. 유

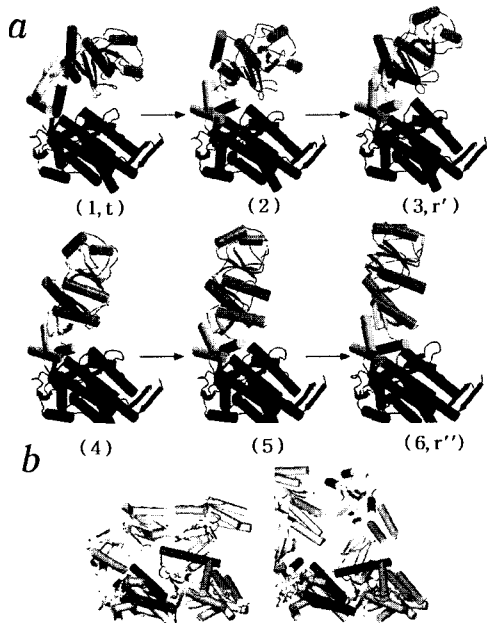


그림 3 GroEL 기능사이클 시뮬레이션에 의한 구조 변화

전공학과 분자생물학에서의 최근의 진보는 개선된 약물, 호르몬, 산업적으로 유용한 효소, 생촉매, 농화학 물질, 다양한 식품 등의 생산을 위하여, 혹은 신물질이나 향상된 단백질의 design을 위하여 상업적 관심이 증가하고 있는 추세이다. 이러한 과정은 분자 구조 정보에 의하여 가속화 되고 있다. 분자간의 상호작용 force들의 계산과 생물질 분자 수준에서의 물리화학적 성질들의 더 많은 정보를 통한 구조적 변화의 추론 등의 3차원적 분자 구조의 연구가 진행되고 있다. 그러나, 컴퓨터 스크린을 통한 생체 분자 구조의 조절은 쉽지 않으며, 특히 생체분자의 docking 계산에서 분자의 형태나 크기를 측정하기는 곤란하다. 이러한 상황에서 생체 구조가 그들의 기능과 관련하여 어떻게 관련되는가의 이해를 위하여 virtual reality system의 user interface를 사용하는 것은 필수적이다.

분자동역학 모사를 위한 다양한 virtual reality device들은 분자들의 상호작용과 display를 위하여 environment[9]들로 사용되며, 병렬 컴퓨터들이 역학적인 생체분자의 물리, 화학적 성질들을 모사하기 위하여 사용되고 있다. 수많은 data set들을 간단하게 표현함으로써 visibility를 향상시키거나 실제 생체분자의 모사를 가능하게 하는 연구도 함께 수행되고 있다. 이러한 예로는 atomic force vector들을 분자 수준에서의 force로 사용함으로써 계산 시간의 상당한 감소를 이루어 거대분자들간의 상호작용 및 단백질과 inhibitor 또는 약물 개발 연구에 적용되고 있다[10].

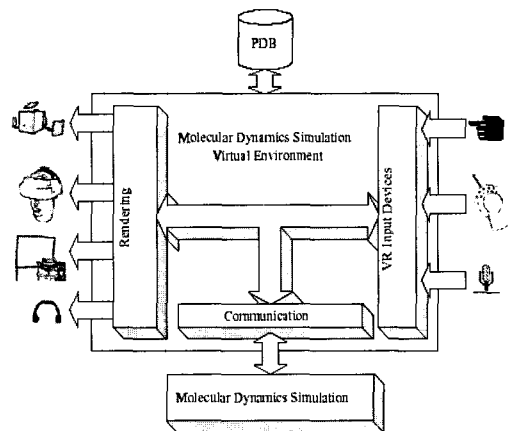


그림 4 분자동역학 시뮬레이션을 위한 가상환경

**참고문헌**

[1] S. B. Needleman and C. D. Wunsch, "A general method applicable to the search for similarities in the amino acid sequences of two proteins," J. Mol. Biol. Vol. 48, pp. 443-453, 1970.

[2] T. F. Smith and M. S. Waterman, "Identification of common molecular subsequences," J. Mol. Biol. Vol. 147, pp. 195-197, 1981.

[3] M. J. Sipple et. al., "ProSup : a refined tool for protein structure alignment," Protein Engineering. Vol. 13, pp. 745-752, 2000.

[4] J. Jung and B. Lee, "Protein structure alignment using environmental profiles," Protein Engineering, Vol. 13, pp. 535-543, 2000.

[5] J. Skolnick and A. Kolinski, "Simulations of the Folding of a Globular Protein," Science, Vol. 250, pp. 112 -1125, 1990.

[6] Alder, B. J. and Wainwright, T. E., J. Chem. Phys., Vol. 27, pp. 1208, 1957.

[7] Rahman, A., Phys. Rev., Vol. A136, pp. 405, 1964.

[8] McCammon, J. A., Gelin, B. R., and Karplus, M., Nature, Vol. 267, pp. 585, 1977.

[9] Ma, J. P., Sigler, P.B., Xu, Z. H., and Karplus, M. A, "Dynamic model for the allosteric mechanism of GroEL," J. Mol. Biol. Vol. 302, pp. 303-313, 2000.

[10] Zhuming, Ai., and Torsten Frohlich. "Molecular Dynamics Simulation in Virtual Environments," EUROGRAPHICS, Vol. 17, pp. 3, 1998.

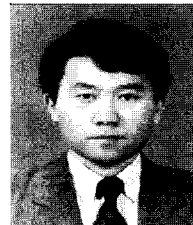
[11] C. N., Yoon, M. H., Chi, H. Ko., and J. Park., "Applying Virtual Realty to Molecular Graphics System," J. of Comput. Sci. & Technol, Vol. 11, No. 5, pp. 507-511, 1996.

**송진수**



1994. 3~1996. 2 충남대학교 자연과학대학원, 화학석사  
 1996. 2~1999. 5 한국생명공학연구원 세포스위치단백질구조연구단 연구원  
 1999. 5~2002. 11 (주)에이프로젠 연구원  
 2003. 1~현재 한국과학기술연구원 생체대사연구센터 연구원  
 E-mail : jssong@kist.re.kr

**윤창노**



1976~1980 연세대학교 화학과(학사)  
 1980~1982 한국과학기술원 화학과(석사)  
 1982~1985 한국과학기술원 화학과(박사)  
 1985~1993 한국과학기술연구원 도핑콘트롤센터 선임연구원  
 1990~1991 코넬대 화학과 Postdoc  
 1993~현재 한국과학기술연구원 생체대사연구센터 책임연구원  
 E-mail : cody@kist.re.kr

● **정보보호연구회 하계학술대회** ●

- 일 자 : 2003년 8월 20일
- 장 소 : 아주대학교
- 주 최 : 정보보호연구회
- 논문접수마감 : 2003년 7월 25일
- 접수처 : 아주대 예홍진 교수

E-mail : hjyeh@adang.ajou.ac.kr

경북대 이형목 교수

E-mail : hmlee@inforsec.knu.ac.kr