

단백질 삼차 구조 예측: 새 구조(new fold) 예측 방법

김 승 연

1. 물리학에 바탕을 둔 방법(physics-based method)

삼차 구조를 모르고 있는 query 단백질이 주어졌을 때, 먼저 비교 모형(comparative modeling) 방법이나 구조 인식(fold recognition) 방법을 이용하여 삼차 구조를 알고 있는 단백질들 중에서 template 단백질을 찾아본다. template 단백질이 결정된 다음에는 template 단백질의 삼차 구조를 이용하여 query 단백질의 삼차 구조를 만들어낸다. 하지만 template 단백질을 전혀 찾을 수 없는 경우도 종종 발생한다. 이런 경우에는 새 구조(new fold) 예측 방법을 사용하여야 한다. 여기에서 new fold 라는 용어는 기존의 구조 라이브러리(fold library)에 없는 삼차 구조의 모양을 의미한다.

new fold 예측 방법에서는 주어진 query 단백질의 아미노산 서열만을 이용하여 query 단백질의 삼차 구조를 예측하여야 한다. new fold 예측 방법의 첫 번째 접근 방법(가장 기본적인 방법)은 주어진 “모든 원자 포텐셜 에너지”(all-atom potential energy 또는 all-atom force field; 예를 들면, AMBER, CHARMM, OPLS)에 분자 동역학(molecular dynamics), 몬테 카를로 방법(Monte Carlo method), 광역 최적화(global optimization) 방법 등을 적용하여 삼차 구조를 예측하는 방법이다. 이 방법은 과학적인 관점에서 보았을 때 가장 훌륭한 방법이지만, 공학적인(또는 실용적인) 측면에서는 효율성이 없는 방법이다.

이 방법은 크게 두 가지의 단점을 가지고 있다. 첫 번째 단점은 삼차 구조를 예측하는데 너무 많은 시간이 소요되기 때문에 아주 작은 크기의 단백질을 제외하고서는 이 방법을 사용할 수가 없다. 두 번째 단점은 현재 사용되고 있는 “모든 원자 포텐셜 에너지”들이 완벽하지 않기 때문에 아주 많은 시간을 투자하더라도 올바른 삼차 구조를 얻을 수 있는지가 보장되지 않는다. 이러한 단점들에도 불구하고 첫 번째 접근 방법은 과학적인 면에서 아주 완벽하기 때문에 이 방법을 이용한 많은 연구가 전세계적으로

진행되고 있다. 최근에 이 방법을 이용하여 20 개의 아미노산으로 구성된 아주 작은 단백질인 Trp-cage(PDB ID: 1L2Y)와 같은 단백질의 삼차 구조를 성공적으로 예측할 수 있었다[1,2].

new fold 예측 방법의 두 번째 접근 방법은 첫 번째 접근 방법과 유사하지만, 첫 번째 접근 방법에서 이용되는 “모든 원자 포텐셜 에너지”를 사용하지 않는다. 대신 “간단한 포텐셜 에너지”(reduced potential energy)를 사용하는데 가장 대표적인 예가 UNRES 포텐셜 에너지이다[3,4,5]. 여기에서 UNRES 는 “UNited-RESidue representation of proteins”를 의미한다. 단백질의 UNRES 표현에서는 아미노산 전체를 두 개의 원자로 표현한다. 첫 번째 원자는 아미노산의 척추(backbone)를 대표하는 원자이고, 두 번째 원자는 각 아미노산의 옆 가지(side chain)를 대표하는 원자이다.

한 아미노산의 척추는 세 개의 원자(질소, 탄소 1, 탄소 2)가 연속적으로 연결되어 이루어져 있고, 이들 원자들에 각각 수소 1, 수소 2, 산소가 결합되어 있다. 프롤린(proline)을 제외한 모든 아미노산은 완전히 똑같은 척추 구조를 가지고 있다. 프롤린의 척추 구조도 다른 아미노산들과 거의 똑같기 때문에 단백질의 UNRES 표현에서는 프롤린과 다른 아미노산들을 구별하지 않고 완전히 같게 취급한다. 척추에서 가장 중요한 원자는 중앙에 위치한 탄소 1 인데, UNRES 표현에서는 이 탄소 1 이 전체 척추를 대표하도록 포텐셜 에너지를 만들었다. 즉 UNRES 표현에서의 첫 번째 원자는 탄소 1 의 위치에 놓여있다. 한편 서로 다른 아미노산들은 서로 다른 모양의 옆 가지를 가지고 있기 때문에 UNRES 표현에서의 두 번째 원자는 옆 가지의 질량 중심(center of mass) 위치에 놓여 있다.

UNRES 표현에 바탕을 둔 new fold 예측 방법은 삼차 구조예측에서 첫 번째 접근 방법(“모든 원자 포텐셜 에너지”에 바탕을 둔 방법)에 비해서 아주 빠르고 상당히 성공[3,4,5]적이지만, 100 개 이상의 아미노산을 가진 단백질의 삼차 구조를 예측하기에는 무리가 있다. 첫 번째 접근 방법과 두 번째 접근 방법은 공통적으로 포텐셜 에너지 함수에 바탕을 두고 있기 때문에, 이 두 방법을 합쳐서 “물리학에 바탕을 둔 방법”(physics-based method)이라고 부른다.

2. 조각 맞추기(fragment assembly) 방법

예측 속도가 너무 느리고 크기가 큰 단백질에 적용할 수 없는 physics-based method 의 단점을 극복하기 위해서 최근에 개발된 방법이 조각 맞추기(fragment assembly) 방법이다[6,7,8,9]. 조각 맞추기 방법은 과학적인 측면에서 볼 때 불완전하고 근거가 빈약한 방법이지만, 실용적인(또는 공학적인) 측면에서는 아주 성공적인 훌륭한 방법이다. 조각 맞추기 방법은 조각 라이브러리(fragment library)라는 데이터베이스를 이용하기 때문에, 생물정보학(bioinformatics)적 요소를 가지고 있다. 동시에 조각들을 맞추어서 query 단백질의 삼차 구조를 만들 때는 physics-based method 와 유사한 방법을 사용하기 때문에, 조각 맞추기 방법은 물리학적 요소도 가지고 있다. 즉 조각 맞추기 방법은 생물정보학과 물리학이 혼합된 방법이다. 현재 조각 맞추기 방법은 new fold 예측 분야에 이용되는 모든 방법들 중에서 가장 뛰어나며, 구조 인식(fold recognition) 분야의 문제들 중에서도 아주 어려운 문제를 푸는데 탁월한 성과를 보여주고 있다[10].

Protein Data Bank(PDB)에 있는 모든 단백질들을 크기가 일정(예를 들면 3, 9, 또는 15 개의 아미노산)한 조각들(fragments)로 나눌 수 있다. 그리고 이들 조각들에 두 번째 글인 “단백질 구조 예측의 기초: BLAST 와 프로파일(profile)”에서 언급한 PSI-BLAST 로부터 결정된 “위치에 의존하는 점수 행렬”(PSSM)을 결합해준다. 즉 PSSM 정보를 담고 있는 단백질 조각들의 모음이 fragment library 이다. 현재 프로파일(PSSM) 데이터베이스를 만드는 작업을 진행 중이기 때문에(<http://www.cheric.org>), 앞으로는 어느 분이나 fragment library 를 손쉽게 만들 수 있을 것이다. 조각 맞추기 방법의 기본 철학은 “new fold 를 가진 단백질일지라도 이 단백질의 부분부분의 모양은 이미 PDB 에 존재하고 있는 어느 단백질의 한 부분과 동일하거나 유사하다”라는 가정이다.

조각 맞추기 방법에서는 query 단백질이 주어지면 먼저 PSI-BLAST 를 이용하여 query 단백질의 PSSM 을 만든다. PSSM 을 얻었으면 query 단백질의 PSSM 을 크기가 일정한 조각들로 나눈 다음, 각각의 조각들의 PSSM 에 대해서 fragment library 에 있는 모든 조각들의 PSSM 과 비교하여 유사한 PSSM 을 가진 조각들을 fragment library 로부터 추려낸다. fragment library 에서 추려진 조각들이 query 단백질의 한 부분을 형성하는 후보가 된다. 후보들 중에서 마지막으로 선택되는 조각은 포텐셜 에너지에 의해서

선택된다. 조각 맞추기 방법에서는 조각 안에 있는 원자들 사이의 상호작용(interaction)을 고려할 필요가 없고, 조각들 사이의 상호작용만 고려하면 되기 때문에 포텐셜 에너지를 빠르게 계산할 수 있다.

손쉽게 사용할 수 있는 조각 맞추기 서버로는 robetta 가 있다. 이 서버는 웹 사이트 “<http://robetta.bakerlab.org/>”에서 사용할 수 있다.

[참고문헌]

- [1] C. Simmerling, B. Strockbine, and A. E. Roitberg, All-atom structure prediction and folding simulations of a stable protein, *J. Am. Chem. Soc.* **124** (2002) 11258–11259.
- [2] P. Carnevali, G. Toth, G. Toubassi, and S. N. Meshkat, Fast protein structure prediction using Monte Carlo simulations with modal moves, *J. Am. Chem. Soc.* **125** (2003) 14244–14245.
- [3] S.-Y. Kim, J. Lee, and J. Lee, Folding of small proteins using a single continuous potential, *Journal of Chemical Physics* **120** (2004) 8271–8276.
- [4] S.-Y. Kim, J. Lee, and J. Lee, Folding simulations of small proteins, *Biophysical Chemistry* **115** (2005) 195–200.
- [5] S. Ołdziej, C. Czaplewski, A. Liwo, M. Chinchio, M. Nancias, J. A. Vila, M. Khalili, Y. A. Arnautova, A. Jagielska, M. Makowski, H. D. Schafroth, R. Kaźmierkiewicz, D. R. Ripoll, J. Pillardy, J. A. Saunders, Y. K. Kang, K. D. Gibson, and H. A. Scheraga, Physics-based protein-structure prediction using a hierarchical protocol based on the UNRES force field: assessment in two blind tests, *Proc. Natl. Acad. Sci. U.S.A.* **102** (2005) in press.
- [6] K. T. Simons, C. Kooperberg, E. Huang, and D. Baker, Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions, *J. Mol. Biol.*

268 (1997) 209–225.

- [7] J. Lee, S.-Y. Kim, K. Joo, I. Kim, and J. Lee, Prediction of protein tertiary structure using PROFESY, a novel method based on fragment assembly and conformational space annealing, *Proteins: Structure, Function, and Bioinformatics* **56** (2004) 704–714.
- [8] J. Lee, S.-Y. Kim, and J. Lee, Protein structure prediction based on fragment assembly and parameter optimization, *Biophysical Chemistry* **115** (2005) 209–214.
- [9] J. Lee, S.-Y. Kim, and J. Lee, Protein structure prediction based on fragment assembly and the beta-strand pairing energy function, *Journal of the Korean Physical Society* **46** (2005) 707–712.
- [10] *Proteins: Structure, Function, and Bioinformatics* **53**, Number S6 (2003) 333–595.