

## 소분자 도킹에서의 탐색알고리즘의 현황

정환원 · 조승주\*†

### Recent Development of Search Algorithm on Small Molecule Docking

Hwan Won Chung and Seung Joo Cho\*†

#### Abstract

A ligand-receptor docking program is an indispensable tool in modern pharmaceutical design. An accurate prediction of small molecular docking pose to a receptor is essential in drug design as well as molecular recognition. An effective docking program requires the ability to locate a correct binding pose in a surprisingly complex conformational space. However, there is an inherent difficulty to predict correct binding pose. The odds are more demanding than finding a needle in a haystack. This mainly comes from the flexibility of both ligand and receptor. Because the searching space to consider is so vast, receptor rigidity has been often applied in docking programs. Even nowadays the receptor may not be considered to be fully flexible although there have been some progress in search algorithm. Improving the efficiency of searching algorithm is still in great demand to explore other applications areas with inherently flexible ligand and/or receptor. In addition to classical search algorithms such as molecular dynamics, Monte Carlo, genetic algorithm and simulated annealing, rather recent algorithms such as tabu search, stochastic tunneling, particle swarm optimizations were also found to be effective. A good search algorithm would require a good balance between exploration and exploitation. It would be a good strategy to combine algorithms already developed. This composite algorithms can be more effective than an individual search algorithms.

**Key words** : Small Molecule, Search Algorithm, Docking, Scoring, Induced fit

#### 1. 탐색알고리즘의 중요성

일반적으로 컴퓨터 상에서 분자의 거동을 연구하는 것은 실험을 할 때의 노력과 시간을 줄일 뿐만 아니라, 직접적으로 실험하기 곤란한 상황을 연구하는 데에도 유용하다. 이중 대표적으로 많이 연구 되어온 분야가 신약설계 분야일 것이다. 컴퓨터를 이용한 계산기법의 도입은 소분자들이 실제 어떻게 수용체에 결합하는지에 대한 모델을 보여줌으로서 신약설계에서의 혁신을 가져왔다. 1980년대의 Dock 프로그램을 시작으로<sup>[1]</sup> 소분자와 수용체의 도킹을 다루는 프로그램은 60종이 넘

는 것으로 알려져 있으며<sup>[2]</sup> 자주 쓰이는 프로그램만 나열해도 약 10종 가까이 된다.<sup>[3]</sup> 이러한 프로그램들을 개발한 논문들을 보면 제각기 각 프로그램의 탁월성을 보여주는 결과를 제시하고 있다. 이것은 프로그램을 사용하는 사용자의 입장에서는 상당히 혼란스러운 일이다.<sup>[4-13]</sup> 따라서 어떤 프로그램을 사용하는 것이 가장 좋은 선택이 되는지에 대한 판단을 하는 것은 쉽지 않다. 이러한 판단을 하기 위해서는 프로그램의 개발 과정을 이해하는 것이 필요하다. 일반적으로 도킹프로그램은 대략적으로 3가지의 요소로 나누어서 이해할 수 있다. 먼저 도킹모델을 찾아낼 수 있는 탐색알고리즘이다. 소분자가 수용체에 결합하는 것은 병진운동, 회전운동, 컨포메이션 등의 가능성을 고려할 때, 가능한 모델이 일반적으로 너무나 많아서 모든 가능성을 다 계산하는 것이 불가능하다. 따라서 가능성이 큰 모델을 잘 찾아줄 수 있는 알고리즘이 필요하다. 또 이러한 모델 중에서 어떤 모델이 가장 적당한 것인지를 판별할 수 있는 판별함수가 필요하다. 실질적으로 수용체축에

(Computational Science Center, Future Fusion Technology Division, Korea Institute of Science and Technology, P.O. Box 131, Cheongryang, Seoul 130-650, South Korea)

\*조선대학교 의학과 세포분자 의과학(Department of Cellular and Molecular Medicine, College of Medicine, Chosun University, 375 Seosuk-dong, Dong-gu Gwangju 501-759, Korea)

†Corresponding author: chosj@chosun.ac.kr

(Received : May 22, 2009, Accepted : June 22, 2009)

서의 소분자도킹문제는 정확한 양자역학적 계산으로는 계산의 양이 너무 많아 실질적으로 계산이 불가능하다. 따라서, 대부분의 경우, 자유에너지가 가장 낮은 구조를 판단할 수 있는 비교적 단순화한 함수식을 사용하는데 이것이 판별함수이다.

마지막으로는 이 두 가지의 요소기술을 실제 어떻게 조합하여 프로그램을 만들어 내는가 하는 프로그램 구성에 관한 것이다. 구성의 차이가 많은 프로그램의 다양성을 만들어 낸다. 먼저 소분자를 도킹하는 과정에서 조각모음 방식으로 결합하는 것과 컨포메이션을 계속 변화시키므로서 적합한 결합구조를 찾는 방식이 있다. 예를 들어, Dock 프로그램은 리간드를 작은 부분으로 분해해서 순차적으로 결합시키는 기법을 사용하고, Autodock 프로그램은 리간드 전체를 처음부터 사용하여 컨포메이션을 변화시키는 방법을 사용하고 있다. 어떤 경우에도 탐색알고리즘과 판별함수를 사용해야 한다. 또한 수용체의 결합부위를 표현하는 방식도 결정되어야 할 요소중의 하나이다. 보통 활성자리에서 소분자가 결합할 것으로 예측되는 공간에 대하여 상호작용을 표현할 수 있는 그리드 계산을 한다. 이 3차원 그리드와 소분자와의 상호작용을 계산하는 방식이 탐색의 속도를 많이 향상시킬 수 있기 때문에 널리 사용되고 있다. 최근에는 소분자 뿐만 아니라 수용체의 유연성까지도 고려하는 도킹프로그램이 많이 연구되고 있다. 또한 소분자의 경우에도 유기화학자들이 합성하는 소분자의 범위를 넘어서 펩티드나 다당류 올리고머등의 생물고분자의 도킹에 대한 필요성이 점차 증대하고 있다. 따라서 거대한 탐색공간을 효과적으로 다룰 수 있는 똑똑한 탐색알고리즘의 중요성은 나날이 늘고 있고, 이 핵심기술의 연구개발이 많이 이루어지고 있다.

## 2. 고전적 탐색 알고리즘

### 2.1. MC (Monte Carlo 탐색)

Monte Carlo 탐색 방법은 열역학에서 잘 연구되어 있는 확률의 개념을 도입하여 극소점을 찾아가는 알고리즘이다. 만약 A에서 B로 바뀌는데 B의 에너지가 낮으면 이 변화는 무조건 수용한다. 만약 B로 바뀌는데 에너지가 커지면 다음과 같은 확률로 B로의 변화를 수용한다.

$$P = \exp(-[E_A - E_B]/kT)$$

여기서 P는 확률이고 EA는 A의 에너지, EB는 B의 에너지 k는 볼츠만 상수, T는 시뮬레이션에서 사용하

는 온도이다. 이렇게 하면 이때 얻는 샘플은 대략적으로 볼츠만 분포를 따르게 되는 장점이 있고 따라서 거시적인 물리적 파라메타 계산이 가능해지는 장점이 있다. 탐색기법의 측면에서 보면, 또한 에너지 장벽을 넘어서 다른 곳으로의 탐색이 가능해진다.

### 2.2. MD (Molecular Dynamics) 탐색

실제분자의 거동을 모사하여 분자의 동적인 움직임을 따라가면서 도킹구조를 얻는 방법론이다. 계산에는 시스템의 분자들의 움직임을 계산하기 위한 force field를 사용하게 된다. 뉴턴의 운동법칙을 따르기 때문에 관성, 속도, 질량등에 의해서 움직임이 결정론적으로 정해지게 된다. 포텐셜 장벽을 넘기 어렵고 따라서 샘플링의 관점에서는 시간이 많이 걸려서 적합하지 않으나, 최종적으로 도킹구조를 최적화시킬 때 많이 사용한다.

### 2.3. SA (Simulated Annealing)

전산모사에서의 담금질이라고 할 수 있다. 온도를 급격히 올렸다가 갑자기 내리기 때문에 담금질을 연상시키는 방법을 사용하기 때문에 이와 같은 이름이 붙었다. 그림 1에서 처럼 높은 온도에서는 포텐셜 장벽을 뛰어넘을 수 있게 되고 낮은 온도가 되면 확률적으로 전역최적화된 도킹구조를 얻을 가능성이 크다. 위에서의 MD, MC의 약점은 국소최적화성능은 우수하나, 탐색공간이 크고 복잡한 경우에 탐색공간의 다양한 곳을 탐사할 수 있는 탐사성능이 떨어지는 단점이 있다. 따

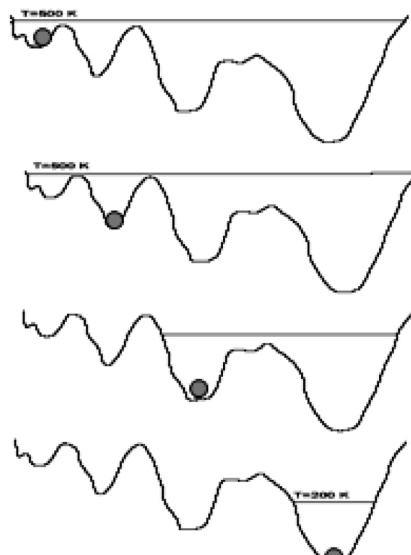


그림 1. 전산모사적 담금질과 전역최적화.

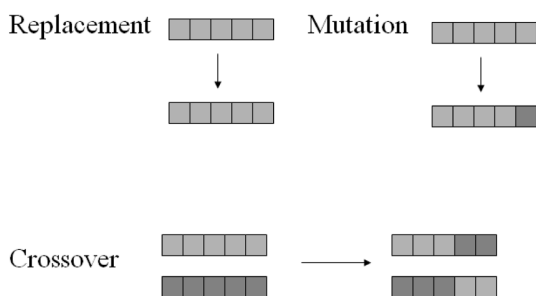


그림 2. GA의 대표적인 변화방법.

라서 MD, MC의 기법과 함께 사용하면 전역 최적화 성능이 크게 향상된다.

### 2.4. GA (Genetic Algorithm)

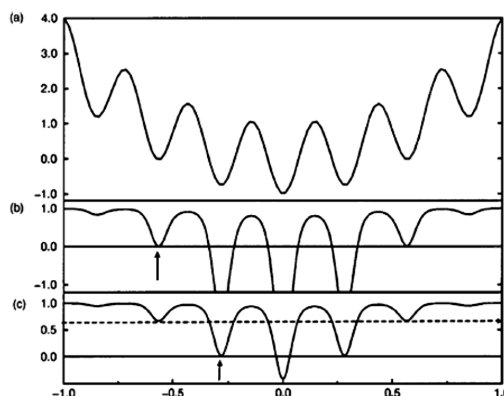
자연선택설을 모방하여 만든 알고리즘이다. 리간드의 병진운동, 회전운동, 컨포메이션의 변화 등을 나타내는 변수들로서 유전자를 구성한다. 이것은 하나의 유전자로 이루어지는 것이 아니라 여러개의 유전자로 이루어진 군집을 이루게 된다.

변화는 이전세대에서 다음세대로 갈 때 유전자에서 나타난다. 단순히 아무변화도 일어나지 않을 수도 있고 돌연변이가 생길수도 있으며, 한 개체의 유전자의 일부가 다른 개체의 교환될 수도 있다. 이러한 변화는 우연히 일어나며 유전자의 변이는 표현형의 차이를 가져오게 된다. 이러한 변화를 조절하는 것이 적합성함수이다. 적합성함수에 의해서 변화가 촉진될 수도 있고 억제될 수도 있다. 유전자가 변이하는 과정에서 직접적인 포텐셜장벽이 없으므로 MC, MD에 비하여 터널링 효과 같은 것을 도입하여 넓은 범위를 탐색하는데 유리하다.

## 3. 최근의 탐색 알고리즘 개발 동향

### 3.1. STUN (Stochastic Tunneling)

분자도킹을 위해서는 다양한 통계적 최적화 기법이 사용된다. simulated annealing의 발전된 형태인 stochastic tunneling(이하, STUN)을 사용하여 분자도킹 문제를 푸는 방식이다. 기존 탐색 방법에 대한 새로운 통계적 최적화 방법 (STUN)은 Effective Potential을 사용하여 최저 해를 기준으로 적합도(fitness)를 scaling 하므로 국소해(local solution)에서 벗어나지 못하는 freezing이 발생하지 않는 특징이 있다. 또한, 제어 온도가 simulated annealing과 같이 한번 변경된 다음 고



Wenzel, Hamacher, Phys. Rev. Lett. 82, 3003 (1999)

그림 3. STUN기법이 effective potential 및 온도 조정.

정되는 것이 아니라 탐색 상태에 따라서 적용하여 지능적으로 변경된다. 즉, 온도가 낮은 경우에 국소 탐색을 진행하고, 계속 해의 품질이 나아지지 않으면 Heating을 통해 전역 탐색을 수행할 수 있도록 한다. Heating과 Cooling이 일정한 평균치를 기준으로 주기적으로 나타나서 국소 탐색과 전역 탐색이 균형을 이루어 진행되는 것이다.

$$f_{STUN}(x) = 1 - e^{-\gamma(f(x) - f_0)}$$

이러한 특징은 탐색 성능이 기존 autodock의 LGA (Lamarckian GA)에 비해 뛰어날 수 있고, 이것은 타겟 단백질의 다수의 아미노산 잔기에 대한 flexibility를 고려하는 도킹에서 중요한 특징이 된다. 이 STUN기법은 기존의 방법에 비해서 문제 공간이 rugged한 경우, 즉, 다수의 국소 최적점을 갖는 경우에 뛰어난 결과를 보이고 분자도킹 역시 이러한 경우에 해당한다.

### 3.2. PSO (Particle Swarm Optimization)

PSO(Particle Swarm Optimization) 기법은 flexible docking method 개발에서 최근에 크게 가능성을 인정받고 있다. PSO기법은 여타 방법(Simulated Annealing, Genetic Algorithm등등)에 비해 연속적이고 실수 값을 갖는 함수 최적화에 뛰어나기 때문에 소분자 도킹 문제의 해결에 효과적이라고 할 수 있다. PSO기법은 유전 알고리즘과 같은 개체군 기반의 방법이지만, 기본 개념은 경쟁이 아닌 협업에 바탕을 두고 있다. 따라서, 유전 연산자를 갖지 않으며, 대신 자신만의 경험과 속도를 가지고 움직이는 Particle들을 이용한다.

### 3.3. Tabu 탐색

Tabu 탐색은 이웃해 탐색 방법의 하나로서 각 단계 별로 가장 바람직한 해를 찾아서 이동해 가는 방법이다. 탐색 단계에서 해의 품질이 현재보다 좋지 않아도 이동하므로, 지역 최적화에 머무르지 않는다. 특히, 일종의 단기기억인 타부조건을 도입하여 매 단계에서 이동할 수 있는 움직임에서 최근의 방문하였던 해를 배제하므로, 탐색 과정에서의 순환을 방지할 수가 있다. 타부 탐색은 주로 이산적인 최적화 문제의 해결에 많이 사용되었지만, 최근에 분자도킹을 위해서 응용하려는 시도가 기타 최적화방법과 융합되어 시도되고 있다.

## 4. 결 론

분자도킹에서 리셉터의 구조가 flexible하게 변하는 것을 고려하는 induced fit방법에 대한 시도가 중요하게 대두되고 있고, 이를 위해서는 높은 탐색 성능이 반드시 필요하다. 그러나 비교적 단순한 경우인 rigid 도킹의 경우를 고려해 볼 때에도, 리간드의 병진운동, 회전운동 그리고 리간드 자체의 conformation을 고려하면, configuration은 1014개를 계산할 필요가 있다. 따라서 탐색공간 전체를 다 계산하는 것은 실질적으로 가능하지 않다. 오직 가능한 configuration의 극히 일부를 계산해야 하기 때문에 계산적인 비용과 탐색공간의 크기가 적절하게 되어야 한다. 탐색 성능은 해를 찾는데 소요되는 시간과 발견된 최종 해의 품질에 관련해서 정의될 수 있다. 궁극적으로, 짧은 시간에 좋은 품질의 해를 찾는 것이 탐색알고리즘의 목표하라 정의할 수 있다. 단백질-단백질 상호작용, 단백질-DNA 상호작용뿐만 아니라, 다당류분자, DNA조각, 펩티드 같은 상당히 복잡한 리간드의 도킹문제들은 본질적으로 리간드와 수용체의 구조적인 유동성을 생각하지 않으면 풀리지 않는 문제이고 점차 연구의 방향이 이렇게 복잡한 경우의 도킹을 연구하는 쪽으로 이동하고 있다. 이렇게 리간드와 수용체의 유동성 때문에 엄청나게 광범위한 문제 공간에서의 보다 효과적인 도킹구조를 얻어내기 위해서는 효율적인 탐색알고리즘이 필수적이고, 전역탐색과 국소최적화를 적절히 안배 함으로서 달성이 가능할 것이다.

## 감사의 글

이 논문은 2009년도 정부(교육과학기술부)의 재원으로 한국과학재단의 지원을 받아 수행된 연구임(2009-0076553)

조선자연과학논문집 제2권 제2호, 2009

## 참고문헌

- [1] R. D. Taylor, P. J. Jewsbury, and J. W. Essex, "A review of protein-small molecule docking methods", *J. Comput. Aided. Mol. Des.* 16, p. 151, 2002.
- [2] N. Moitessier, P. Englebienne, D. Lee, J. Lawandi, and C. R. Corbeil, "Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go", *British Journal of Pharmacology* 153, ps7, 2008.
- [3] R. Wang, Y. Lu, and S. Wang, "Comparative Evaluation of 11 Scoring Functions for Molecular Docking", *J. Med. Chem.* 46, p. 2287, 2003.
- [4] E. Perola, W. P. Walters, and P. S. Charifson, "A Detailed Comparison of Current Docking and Scoring Methods on Systems of Pharmaceutical Relevance", *Proteins. Struct. Funct. Bioinf.* 56, p. 235, 2004.
- [5] T. S. Gasch and M. Stahl, "Scoring functions for protein-ligand interactions: a critical perspective", *Drug. Discov. Today.* 1, p. 231, 2004.
- [6] U. Rester, "Dock around the Clock-Current Status of Small Molecule Docking and Scoring", *QSAR Comb. Sci.* 7, p. 605, 2006.
- [7] E. M. Krovat, T. Steindl, and T. Langer, "Recent Advances in Docking and Scoring", *Current Computer-Aided Drug Design* 1, p. 93, 2005.
- [8] H. Chen, P. D. Lyne, F. Giordanetto, T. Lovell, and J. Li, "On Evaluating Molecular-Docking Methods for Pose Prediction and Enrichment Factors", *J. Chem. Inf. Model.* 46, p. 401, 2006.
- [9] H. Alonso, A. A. Bliznyuk, and J. E. Gready, "Combining Docking and Molecular Dynamic Simulations in Drug Design", *Medicinal Research Reviews* 26, p. 531, 2006.
- [10] H. Chen, P. D. Lyne, F. Giordanetto, T. Lovell, and J. Li, "On Evaluating Molecular-Docking Methods for Pose Prediction and Enrichment Factors", *J. Chem. Inf. Model.* 46, p. 401, 2006.
- [11] M. K. Gilson and H. X. Zhou, "Calculation of Protein-Ligand Binding Affinities", *Annu. Rev. Biophys. Biomol. Struct.* 36, p. 21, 2007.
- [12] M. J. Hartshorn, M. L. Verdonk, G. Chessari, S. C. Brewerton, W. T. M. Mooij, P. N. Mortenson, and C. W. Murray, "Diverse, High-Quality Test Set for the Validation of Protein-Ligand Docking Performance", *J. Med. Chem.* 50, p. 726, 2007.
- [13] M. L. Verdonk, P. N. Mortenson, R. J. Hall, M. J. Hartshorn, and C. W. Murray, "Protein-Ligand Docking against Non-Native Protein Conformers", *J. Chem. Inf. Model.* 48, p. 2214, 2008.