J. of the Chosun Natural Science
Vol. 3, No. 3 (2010) pp. 143 – 147

소분자 도킹에서 탐색공간의 축소 방법

조승주†

Search Space Reduction Techniques in Small Molecular Docking

Seung Joo Cho†

Abstract

Since it is of great importance to know how a ligand binds to a receptor, there have been a lot of efforts to improve the quality of prediction of docking poses. Earlier efforts were focused on improving search algorithm and scoring function in a docking program resulting in a partial improvement with a lot of variations. Although these are basically very important and essential, more tangible improvements came from the reduction of search space. In a normal docking study, the approximate active site is assumed to be known. After defining active site, scoring functions and search algorithms are used to locate the expected binding pose within this search space. A good search algorithm will sample wisely toward the correct binding pose. By careful study of receptor structure, it was possible to prioritize sub-space in the active site using "receptor-based pharmacophores" or "hot spots". In a sense, these techniques reduce the search space from the beginning. Further improvements were made when the bound ligand structure is available, i.e., the searching could be directed by molecular similarity using ligand information. This could be very helpful to increase the accuracy of binding pose. In addition, if the biological activity data is available, docking program could be improved to the level of being useful in affinity prediction for a series of congeneric ligands. Since the number of co-crystal structures is increasing in protein databank, "Ligand-Guided Docking" to reduce the search space would be more important to improve the accuracy of docking pose prediction and the efficiency of virtual screening. Further improvements in this area would be useful to produce more reliable docking programs.

Key words : Structure-based design, Ligand-based design, Docking, Search space, Ligand-guided docking

1. 일반적인 도킹 프로그램 개발 상황

1리간드가 수용체에 결합하는 것은 생물학에서의 가장 중요한 사건이라고 할 수 있다. 특히, 소분자가 수용체에 구체적으로 어떻게 결합하고 있는지를 아는 것은 신약개발에서 대단히 중요한 정보이다. 따라서, 리간드의 결합구조를 예측하는 것은 분자모델링에서 가장 기본적이고 중요한 분야의 하나이다. 또한 이미 많은 프로그램들이 수용체의 유연성을 제한함으로서 탐색공간을 제한하고 있다. 단백질-단백질 결합을 연구하는 데에는 주로 양쪽의 단백질을 유동성이 전혀 없는 강체(剛體)로 생각하고 문제를 해결하는 방식이 주류를 이루고 있다. 이러한 제한은 현실과 맞지 않기 때문에 도킹기술을 실제 문제에 활용하는데 제약이 따른다. 소분자 도킹에서의 경우에서 조차 수용체의 유연성을 고려하는 문제는 아직도 초기단계라고 할 수 있다. 이렇게 탐색공간에 대한 인위적인 제한을 하고 있는 상황임에도 불구하고 아직도 활성부위의 탐색공간은 대단히 커서 현재의 연산능력으로는 충분히 골고루 계산할 수 없다. 지금까지 대다수의 연구들이 평가함수와 탐색알고리즘을 개선하는 쪽에 치중되어 왔다. 이들을 개선하는 연구는 기본적으로 매우 중요한 문제이다.[1] 그러나 다양한 연구를 통해서도 신약개발에서 요구되는 정도의 성능까지 접근하지 못하고 있다. 즉, 현재의 기술은 거대한 탐색공간을 효과적으로 다루지 못하고 있다. 최근의 Andrew R. Leach의 말처럼 이러한 일반적인 방법은 현재 기술적인 한계에 도달한 느낌이다.[2]

조선대학교 의과대학 (Department of Cellular · Molecular Medicine and Research Center for Resistant Cells, College of Medicine, Chosun University, Gwangju 501-759, Korea)

†Corresponding author: chosj@chosun.ac.kr
(Received : September 7, 2010, Revised : September 17, 2010, Accepted : September 27, 2010)

2. 알고리듬을 통한 탐색공간의 축소

좋은 알고리듬은 탐색공간의 관점에서 보면 탐색과정에서 탐색의 범위를 효과적으로 축소하여 해를 찾아가는 것으로 정의할 수 있다. Tabu 탐색은 한번 탐색한 공간부근에 다시 접근하는 것을 제한함으로서 탐색의 효율을 증대시키는 방법이고, 국소 최저점 근방에 묶이지 않고 Rugged한 공간을 효과적으로 탐색하기 위한 STUN, 한꺼번에 많은 해를 도입하여 효과적인 해를 찾아가는 방법으로 자연의 경쟁논리를 응용한 GA, 반대로 협동논리를 응용한 PSO등의 알고리듬이 적용되었고 효과적으로 이용되어 왔다. 하지만 알고리듬을 통하여 탐색과정에서 수집한 정보를 가지고 탐색공간을 축소시키는 방법은 한계가 있는 것으로 보인다. 아마도 알고리듬 자체가 주어진 공간에 대하여 학습하는 과정 자체가 이미 많은 시간을 요구하게 될 가능성이 크다. 다음 절에 기술한 것 바와 같이, 오히려 미리 탐색의 범위를 제한하여, 작아진 탐색공간에서 해를 구하는 방법들이 더 효과적인 것으로 보인다.[3]

3. 수용체의 구조정보를 이용하여 탐색공간을 축소하는 방법

앞에서의 탐색알고리듬을 개선하는 경우와 다른 점은 이 경우는 탐색을 시작하기에 앞서서 탐색공간의 물리적인 크기 자체를 축소시키는 방법이다. 리간드와 수용체가 결합하는 물리적인 힘은 주로 정전기력, 수소결합, 소수성 상호작용등에 기인한다고 볼 수 있다

3.1. Cerius의 SBF

Cerius의 SBF(Structure-based Focusing) 프로그램[4]은 수용체의 구조를 잘 분석하여 물리적으로 상호작용에 중요할 것으로 보이는 부위를 3차원적으로 잘 정의한다. 즉, 먼저 활성부위를 규정하고, 상호작용지도를 만든다. 이 지도는 수용체의 구조로부터 얻어지는 가능성이 있는 상호작용에 대한 정보를 총 망라한 것이라고 할 수 있다. 이 지도가 복잡하기 때문에 다음으로는 거리를 기준으로 군집화를 하여 상호작용의 개수를 줄인다. 이렇게 만들어진 수용체구조기반의 pharmacophore를 이용하여 가상탐색에 응용할 수 있다. 이 결과는 탐색공간중에서 가능성이 높은 공간을 정의하고, 이 작용점들을 포함하지 않는 구조를 배제함으로서 구조예측력을 높인 것으로 해석할 수 있다. 일반적으로 가상탐색에서 편리한 feature의 수는 3-5개 정도가 적당한 것으로 알려져 있다. 많은 경우에 생기는 feature의 개수

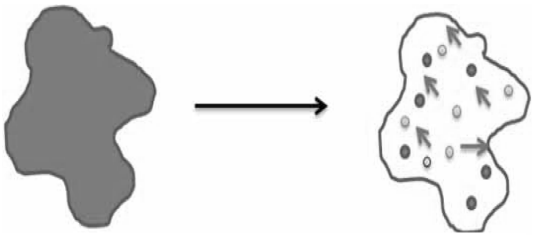

그림 1. 탐색공간의 축소. SBF프로그램은 탐색영역을 수십개 이하의 상호작용점으로 바꾸어서 탐색공간 자체를 크게 축소한다. 이 상호작용점들은 물리적으로 결합에 중요한 수소결합(vector), 소수결합(점) 등을 나타낸다. 이들을 조합으로서 pharmacophore 모델을 만들 수 있다.
Fig. 1. Reduction of Search Space: SBF program reduces the search space by converting the area into dozens of interaction points. These points represents hydrogen bonding or hydrophobic interaction. Combination of these points become a pharmacophore.

는 복잡한 활성부위의 경우 수십 개가 되기 때문에 이 방법은 pharmacophore 모델을 설정하는데 어려움이 있다. (가능한 모델의 숫자를 생각하면 천문학적인 숫자가 됨)

3.2. GRID

이보다 좀 더 실용적인 것으로 GRID 프로그램이 있다.[5] 앞서의 SBF의 경우에는 기하학적으로 상호작용을 해석했으나, GRID프로그램은 실제로 상호작용에너지를 좀 더 실제적인 probe를 도입하여 계산한 면이 차이가 있다. 예를 들어 물분자나, carbonyl oxygen같은 실질적인 probe를 이용하여 에너지를 계산함으로서 신약설계에 있어서 응용이 보다 용이한 실질적인 상호작용에너지를 얻는다. 이의 응용으로서 물분자 “hot

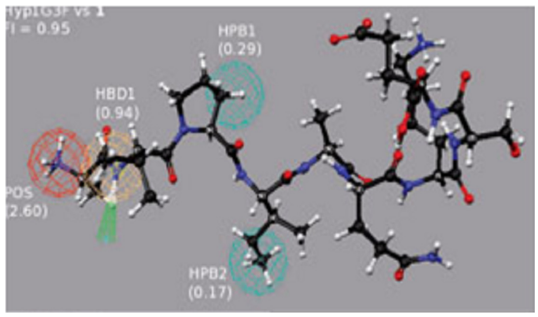

그림 2. GRID 프로그램에서 pharmacophore를 정의한 그림. 괄호속의 숫자가 feature weight임.
Fig. 2. Definition of pharmacophore in GRID program. The numbers in parenthesis are feature weights.

spot"을 계산한 것은 고무적인 일이다. 그 이전에는 구조형성에 중요한 물 분자의 경우 왜 그곳에 물 분자가 있어야 되는지 정량적인 계산이 쉽지 않았다. 앞서의 SBF와 다른 중요한 차이는 SBF의 결과는 모든 feature들이 같은 정도의 중요성을 갖고 있다고 생각할 수 있으나, GRID에서 얻어지는 feature들은 각각의 중요도가 다르다. 즉, 계산된 결합력의 차이에 따라서 중요도를 정할 수 있다. 따라서 feature의 개수를 제한하여 pharmacophore모형을 만들 때, 효과적으로 알맞은 숫자의 pharmacophore를 정의할 수 있다.

4. 결합하고 있는 리간드의 정보를 활용하여 탐색공간을 축소하는 방법 ("Ligand-Guided Docking") 수용체의 구조정보를 이용하여 탐색공간을 축소하는 방법

단백질 데이터베이스에 계속해서 co-crystal의 숫자가 늘고 있기 때문에 리간드의 정보를 활용하여 도킹의 성능을 높이고자 하는 시도도 있다. 이미 대부분의 도킹프로그램은 결합하고 있는 리간드의 정보를 가지고 활성부위의 공간적인 범위를 정하는데 사용하고 있다. 즉, 대부분의 도킹프로그램은 이미 대략적인 활성자리를 알고 있다고 가정함으로서 축소된 탐색공간내에서 정확한 결합구조를 예측하고자 하는 노력을 기울이고 있다. 한 걸음 더 나아가, 보다 적극적으로 결합구조까지 어느 정도 유사할 것으로 생각하고 리간드의 3차구조정보를 활용하는 방법도 있다. 이미 결합하고 있는 소분자와 비슷한 방식으로 결합구조를 유도하기 때문에 탐색의 방향이 유도된다.

4.1. SG-Dock(Similarity-Guided Dock)과 SP-Dock (Similarity-Penalized Dock)1 Cerius의 SBF

단백질 데이터베이스에 계속해서 co-crystal의 숫자 Fradera 등이 소분자 도킹 프로그램인Dock기반에 MIMIC 프로그램을 모듈화하여 만든 프로그램이다.[6] 리간드의 3차 구조정보를 도킹에 응용한 논문은 이 논문이 아마 처음인 듯하다 (2000년). 이 논문 제목에서는 "Similarity-Driven"이라는 말을 사용하였다. 기술적인 면에서는 맞지만 내용상으로는 "Ligand Structure-Guided"가 더 적당한 제목이 될 것 같다. 즉, 평가함수(TRB)를 기존의 Dock의 평가함수인 DRB에 3차 유사함수인 SAB의 곱으로서 새로운 평가함수를 정의하였다.

$$TRB = DRB \cdot SAB \quad (1)$$

이들은 이 평가함수를 도킹구조를 유도(SG-Dock)하는 것과 도킹의 결과 구조를 평가(SP-Dock)하는 데에 모두 사용하였다. SP-Dock은 역시 가상탐색에서도 우수한 결과를 얻을 수 있었다. Docking score는 여기서는 설명하지 않고, 3차 유사도 계산법을 설명하겠다.

Field에 기초한 3차 유사도 (ZAB)는 다음과 같이 정의할 수 있다.

$$Z_{AB} = \int F_A(r)F_B(r)dr \quad (2)$$

여기서 $F_A(r)$과 $F_B(r)$는 각각 참고분자(reference molecule) A와 목적분자(target molecule) B의 molecular filed이다.

$$S_{AB}(t,\theta,\tau_B) = \frac{Z_{AB}(t,\theta,\tau_B)}{(Z_{AA}Z_{BB}(\tau_B))^{1/z}} \quad (3)$$

SAB는 분자의 겹침에 의존하는데 또한 conformation에도 의존하게 된다. 따라서 겹침, 즉, 유사도는 분자의 병진운동 t, 회전운동 θ, 그리고 목적분자의 conformation의 자유도 τ_B(회전이 가능한 결합의 갯수)에 의존한다. 이 경우 유사도 SAB는 conformation의 변화가 가능한 B 분자의 A로의 중첩을 계산하는데 이용한다. 유사도는 계산에 사용하는 field의 성격에 따라, 0 ~ 1, 또는 -1 ~ 1의 값을 갖는다. steric field와 electrostatic field는 2:1의 비율이 좋다는 연구결과가 있다.

이 방법을 32개의 구조정보가 있는 Thrombine 억제제에 적용하였다. 알고 있는 구조가 32개나 되므로 어떤 구조정보를 기준으로 해야 되는지가 문제가 된다. 그림 3에 보면 대단히 다양한 구조가 존재하고 있다. 이때, docking결과의 리간드 의존성을 알아내기 위하

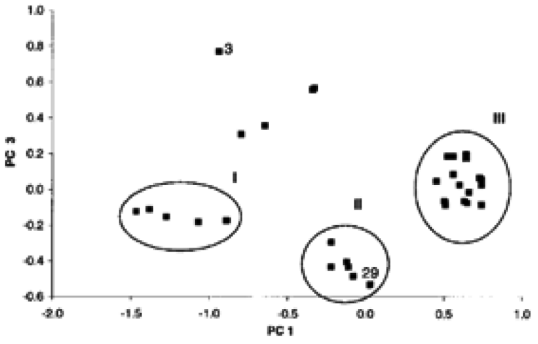

그림 3. 32개의 thrombine 억제제의 주성분 분석 및, 군집분석 결과. Group III에 반이상의 화합물이 있어서 이 구조는 SG-Dock과 SP-Dock에서 제외되었다.
Fig. 3. The results of principal component analysis followed by cluster analysis over 32 thrombine inhibitors: Since there are more than half of the compounds in Group III, these structures were excluded in SG-Dock and SP-Dock.

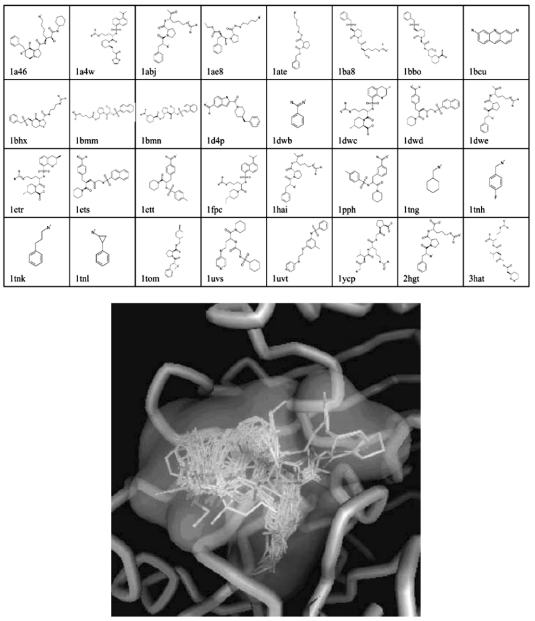

그림 4. 32개 화합물의 2차구조와 수용체 안의 결합구조.
Fig. 4. 2D structures and Docked structures of 32 compounds.

여, 리간드 자체의 유사도 검색을 하였다. 모두 32개 이므로 32 × 32이 유사도 계산을 한 다음 주성분 분석을 하였다. 3개의 주성분으로 79%의 변화도를 설명할 수 있었다. 이중 구조가 확연히 다른 5개를 제외하면 3개의 군집으로 분류되는데 이중 가장 많은 숫자가 있는 group 3의 경우 (50% 이상)는 처음부터 제외되었다. 구조분석은 group I과 II에서 각각 1개씩 대표를 뽑아서 수행하였다. 두 경우 모두 SG-Dock의 경우가 더 구조적으로 정확한 결과를 얻었다 (RMSD가 대략 평균적으로 1Å정도 개선됨.)

실제 구조를 보아도 리간드는 상당히 다양하지만 (그림 4) 실제 3차결합 구조가 오히려 단순해 보인다. 이 시뮬레이션에 알게 된 것은 리간드의 3차정보를 이용할 때, 당연하지만 리간드의 구조에 대한 결과의 의존성이다. Fradera등의 저자들은 리간드 자체의 구조를 이용하는 것 보다, pharmacophore를 이용하는 것이 리간드 자체의 구조를 이용하는 것이 (특정한 리간드에 대한 의존성을 줄임으로서) 더 나을 것이라고 제안하였다.

4.2. DOCKER

Wu등이 개발한 SDOCKER는 MD(분자동력학)기반의 CHARMM을 도킹프로그램으로 개조한 CDOCKER 프로그램에 앞서서와 마찬가지로 MIMIC프로그램을

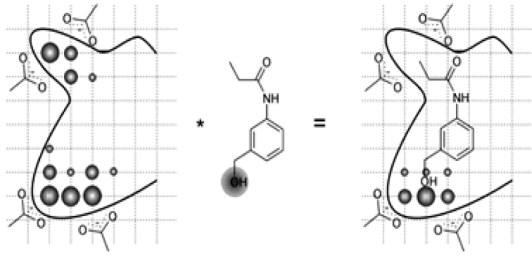

그림 5. 수용체로부터 온 grid에 리간드에서온 weight 정보를 고려하여 새로운 grid를 만듬.
Fig. 5. Newly computed grid by considering the potential from the receptor.

모듈화하여 만든 것이다.[7] 이것은 MD기반의 도킹프로그램이므로 리간드의 유사도에 힘상수(ksim force constant)를 곱하여 리간드가 고정되어 있는 리간드 템플렛으로 끌려가게 조정해 놓았다. 여기서 음의 부호는 A와 B의 분자가 겹치는 방향으로 이끌리게 됨을 나타낸다.

$$\text{Esim} = -\text{ksimSAB} \tag{4}$$

프로그램이 MD에 기반한 점 이외에는 앞에서의 경우와 유사하다. Thrombin, CDK2, HIV-1 protease등의 억제제를 사용하여 실험해 보았다. 구조의 정확도는 최소한 3개의 경우 모두 10%이상 증가하였다.

4.3. AFMoC.2 DOCKER

CoMFA의 경우에는 리간드를 가지고 field를 만드는데 여기에서는 수용체의 좌표를 이용하여 field를 만든다. 여기에 원래 도킹에서 사용하는 grid와 CoMFA와 같은 방식으로 계산한 weight을 합쳐서 새로운 grid를 정의할 수 있다.[8]

5. 탐색공간 축소 연구의 개발 방향

알고리듬을 개선하여 정확한 결합구조를 예측하는 문제는 어느 정도 한계에 다다른 것 같다. 오히려 리간드와 수용체의 구조정보를 보다 잘 활용하는 것이 더 효율적인 것으로 보인다. SBF와 유사한 프로그램으로서 LigandScot와[9] FlexX-Pharm이[10] 있다. LigandScout 역시 구조기반의 pharmacophore 모델링 프로그램이다. FlexX-Pharm의 경우도 비슷한데 사용자가 중요한 feature를 정의할 수 있도록 되어있다. 그러나 이러한 문제에 관한 도킹프로그램은 몇 개의 연구에 제한되어 있다. 특히, 리간드의 구조정보를 활용하는 문제, 리간드와 수용체의 상호작용을 활용하는 문제들은 아직도

더 연구되어야 할 부분이 많은 미개척 분야로 보인다. 그리고 특정 리간드에 대한 의존성 같은 것들도 좀 더 객관적인 방법으로 접근할 수 있는 여지가 많이 있다.

감사의 글

이 연구는 교육과학기술부・한국연구재단 지원 산하 내성세포연구센터지원(R13-2003-009)으로 수행되었음.

참고문헌

[1] S. J. Cho and H. W. Chung “Recent Development of Scoring Functions on Small Molecular Docking”, Journal of the Chosun Natural Science, Vol. 3, p. 49, 2010.

[2] A. R. Leach, B. K. Shoichet and C. E. Perishoff “Docking and Scoring”, J. Med. Chem., Vol. 49, p. 5851, 2006.

[3] S. J. Cho and H. W. Chung “Recent Development of Search Algorithm on Small Molecule Docking”, Journal of the Chosun Natural Science, Vol. 2, p. 1, 2009.

[4] A. M. Hoffren, C. M. Murray and R. D. Hoffmann “Structure-based focusing using pharmacophores derived from the active site of 17 -hydroxysteroid dehydrogenase”, Curr. Pharm. Des., Vol. 77, p. 547, 2001.

[5] F. Ortuso, T. Langer and S. Alcaro “GBPM: GRID-based pharmacophore model: concept and application studies to protein-protein recognition”, bioinformatics, Vol. 22, p1449, 2006.

[6] X. Fradera, R. M. A. Knegtel and J. Mestres “Similarity-Driven Flexible Ligand Docking”, PROTEINS: Structure, Function, and Genetics, Vol. 40, p. 623, 2000.

[7] G. Wu and M. Vieth “SDOCKER: A Method Utilizing Existing X-ray Structures To Improve Docking Accuracy”, Vol. 47, p. 3142, 2004.

[8] H. Gohlke and G. Klebe “DrugScore Meets CoMFA: Adaptation of Fields for Molecular Comparison (AFMoC) or How to Tailor Knowldege-Based Pair-Potentials to a Particular Protein”, J. Med. Chem., Vol. 45, p. 4153, 2002.

[9] G. Wolber and T. Langer “LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters”, Chem. Inf. Mod., Vol. 45, p. 160, 2005.

[10] S. A. Hindel, M. Rarey, C. Buning and T. Lengaue “Flexible Docking under Pharmacophore type constraints”, J. Comput. Aided Mol. Des., Vol. 16, p. 129, 2002.