

콜로니 에너지를 이용한 단백질-리간드 결합 문제에서의 엔트로피 효과 계산

(Consideration of the entropic effect in protein-ligand docking using colony energy)

이 주 용, 석 차 옥

서울대학교 화학부

초 록

단백질-리간드 결합 예측은 새로운 신약 선도 물질을 발견하고 최적화 하는데 있어서 중요한 도구로 널리 사용되고 있다. 결합의 정확도는 일반적으로 각 결합 계산에서 사용되는 평가 함수(scoring function)에 따라 결정된다. 평가 함수는 그 함수가 가지고 있는 기본 가정에 따라 force-field based, empirical, knowledge-based의 세 가지로 분류할 수 있다. 이 중에서 force-field based 함수는 물리적인 상호작용을 가장 구체적으로 기술한다. 그러나 현재 제시된 대부분의 force-field based 함수들은 분산력과 정전기적인 힘 등의 에너지만을 고려할 뿐 엔트로피의 영향을 포함하지 않는 단점을 가지고 있다. 본 논문에서는 force-field based 평가 함수를 이용하는 경우 기존의 도킹 프로그램이 생성해 내는 구조 정보를 활용하여, 엔트로피를 고려할 수 있는 방법을 제시한다. 또한 이 방법을 DOCK 평가 함수에 적용시켰을 때 decoy discrimination에서 향상된 결과를 얻어낼 수 있음을 보였다. 이는 더 정확한 도킹 계산이 가능함을 의미한다.

키워드: 단백질-리간드 도킹, 평가 함수, 엔트로피, 콜로니 에너지, DOCK

Abstract

Computational prediction of protein-ligand binding has been widely used as a tool to discover lead compounds for new drugs. Prediction accuracy is determined in part by the scoring function used in docking calculations. Diverse scoring functions are available, and these can be classified into force-field based, empirical, and knowledge-based functions depending upon the basic assumptions made in development. Among these, force-field based functions consider physical interactions the most in detail. However, the force-field based functions have the drawback of not including the entropic effect while considering only the energy contribution such as dispersion or electrostatic forces. In this article, a method to take into account of the entropic effect using the colony energy is suggested when force-field based scoring functions is used by extracting conformational information obtained from the pre-existing docking program. An improved result for decoy discrimination is illustrated when the method is applied to the DOCK scoring function, and this implies that more accurate docking calculation is possible.

Keywords: Protein-ligand docking, scoring function, entropy, colony energy, DOCK

서 론

단백질-리간드 결합을 이해하고 예측하는 것은 생체분자 간의 상호작용과 그 기능에 대한 이해를 위해 필수적일 뿐

아니라, 신약 개발 산업에서의 높은 응용 가능성 때문에 중요한 문제로 인식되고 있다. 계산으로 단백질과 리간드의 결합을 예측하는 방법은 실험구조를 얻을 수 없는 경우나 많은 단백질-리간드 쌍의 결합에 대한 정보를 짧은 시간 내에 얻고자 할 때 쓰일 수 있는 효율적인 도구로 받아들여지고 있다. 특히, 단백질-리간드 도킹 계산을 통한 가상 검색법은 고효율 스크리닝(high throughput screening)과 같은 실험적인 방법에 비해 enrichment factor가 훨씬 높다는 보고도 있으며 (Shoichet, 2004), 실제 많은 제약회사에서 고효율 스크리닝과 함께 상보적으로 사용되고 있다. 현재 다양한

교신저자: 석차옥 (Email: chaok@snu.ac.kr)

This work was supported by the Seoul R&BD Program and a grant from MarineBio21, Ministry of Maritime Affairs and Fisheries, Korea..

도킹 방법이 알려져 있고, 이 중 어느 방법이 좀 더 나은 방법인지에 대한 연구도 계속되고 있지만, 여전히 단백질-리간드 도킹에서 얻어진 결과는 수많은 false positive와 false negative를 포함하고 있으며, 따라서 이들 도킹 방법은 항상 의의 여지가 많다.

현재 알려진 도킹 프로그램들에 쓰이는 방법에는 여러 가지 문제점이 많다. 리간드와 결합시 수용체 단백질 구조의 relaxation을 전혀 고려하지 않거나 아주 부분적으로만 고려하는 점, scoring function의 부정확성, conformation sampling의 불완전성 등이 그 예이다. 그렇지만 이는 원칙적으로 정확한 계산이 불가능해서라기보다 계산의 정확성과 시간적 효율성의 타협점을 찾아야 하기 때문에 얻어지는 결과이다. 특히 가상검색의 경우에는 비교적 짧은 시간 내에 수십만, 수백만 개의 리간드를 스크리닝 해내야 한다는 요구 때문에 계산 효율성이 큰 문제가 된다. 따라서 효율적으로 현재 도킹 방법의 문제점을 해결할 수 있는 방법을 찾는 것이 큰 과제이며, 이는 기초적인 여러 도킹 연구, 특히 단백질체 등 오믹스 차원에서의 도킹 연구나 신약 개발을 위한 가상검색 연구의 정확성과 유용성을 높이는데 지대한 공헌을 할 것이다.

단백질-리간드 분자도킹에서 풀어야 할 기술적인 문제는 크게 두 가지로 나눌 수 있다. 첫 번째로는 구조적인 측면에서 단백질과 리간드가 결합할 때 어떤 구조와 방향성을 가지고 있을지를 정확히 예측하는 것이고, 두 번째로는 에너지적인 측면에서 결합 친화도를 정확히 예측하는 것이다. 물론 결합 구조를 정확히 예측하여야 결합 친화도를 정확히 예측할 수 있을 것이므로 두 문제는 서로 밀접한 연관성을 가지고 있다. 특히 두 번째 결합 친화도를 정확히 예측하는 문제는 많은 화합물의 데이터 베이스로부터 특정한 단백질과 강한 상호작용을 갖는 물질의 목록을 뽑아내는 가상검색에 있어서 아주 중요한 문제이다.

도킹의 정확한 예측을 위해서는 먼저 가능한 도킹 구조를 효율적으로 찾아내고 이들 가능한 구조 중 가장 유망한, 즉 에너지적으로 가장 안정한 구조를 찾아내는 것이 필요하다. 가능한 도킹 구조를 검색하는 것은 일종의 global optimization 문제이며, 활발한 연구 대상이 되고 있다. 본 논문에서는 두 번째 단계, 즉 scoring function으로 에너지적으로 안정한 구조를 뽑아내는 단계에 대해 다룬다.

현재 여러 도킹 프로그램에서는 scoring function은 force-field based, empirical, knowledge-based scoring function 등과의 조합이 사용되고 있는데, 이 중 force-field based scoring function은 원자간의 물리적인 상호작용을 가장 구체적으로 기술한다는 장점이 있다. 그렇지만 결합의 안정성이나 결합 친화도와 직접적인 연관을 갖는 “자유에너지”를 에너지만으로 근사한다는 단점이 있다. 자유에너지에는 에너지 외에도 엔트로피가 포함되어 있다. 엔트로피는 에너지와는 달리 ensemble property이므로 정확한 계산을 위해서는 MD나 MC와 같은 detailed balance 또는 microscopic reversi-

bility를 정확히 만족하는 thermodynamic sampling을 하여야 한다. 그러나 보통의 thermodynamic sampling 방법은 importance sampling 방법을 쓰므로 그 자체만으로는 엔트로피나 자유에너지를 계산하는 데는 적합하지 못하며, free energy perturbation이나 thermodynamic integration 등의 특수한 방법이 필요하다. (Encyclopedia of computational chemistry, 2005, Bransdal et al, 2003, Kollman 1993, Beveridge and Dicapua, 1989) 이들 열역학적으로 엄밀한 방법들은 계산 시간이 너무 많이 소요되므로, MMPBSA (Huo et al., 2002, Kollman et al., 2000), LIE (Aqvist et al., 1994, Hansson and Aqvist, 1995, Aqvist, 1996, Smith et al., 1998, Wang et al., 1999) 등의 근사적이나 계산 효율성이 보다 높은 방법들도 개발되어 활발히 사용되고 있다. 그러나 이들 근사적인 방법들도 가상검색 등의 높은 효율성을 요구하는 분자도킹에 쓰이기에는 계산량이 너무 많다.

현재 분자도킹에 쓰이는 force-field based scoring function 중에는 용액 상에서 자유롭던 리간드가 수용체에 결합하여 구조적으로 제한됨으로써 감소되는 엔트로피 효과를 고려하기 위해 리간드의 torsion angle의 개수에 비례하는 항을 포함시키는 경우도 있다. 이것은 아주 초보적인 단계의 엔트로피 고려이다. 본 논문에서는 도킹을 위한 scoring function의 효율성을 그대로 지니면서도 통계역학적으로 보다 sound한 근사방법을 도입한다. Colony energy라는 것을 이용하는데 이 방법은 이미 loop prediction에 성공적으로 적용된 예가 있다 (Xiang, 2002). Colony energy 방법에서는 도킹 프로그램에 의해 sample된 여러 구조의 population으로부터 실제 conformational space에 존재하는 population을 추정하는데, 주어진 구조와 비슷한 구조가 더 많이 sample될수록 실제 conformation space에서 이 구조가 대표하는 macrostate의 ensemble이 더 큰 conformational space hypervolume을 차지할 것이라는 reasonable 가정으로부터 출발한다. 실제 도킹 프로그램이 만들어내는 구조를 관찰해보면 비슷한 구조들이 있으며, 이들 population은 많은 경우 다양한 크기의 cluster로 분리가 가능하다. Colony energy를 사용한 방법은 현재 도킹 프로그램이 주는 이러한 정보를 버리지 않고 적절히 활용함으로써 엔트로피 효과를 고려하는 것이다.

본 논문에서는 colony energy를 도킹 방법에 적용시키는 것이 scoring function을 향상시키는데 실질적으로 얼마나 도움을 줄 수 있는지 확인하기 위하여 decoy discrimination의 성공률을 살펴보았다. 도킹 프로그램으로는 DOCK을 사용하였으며, decoy set은 Alan et al.,(2005)이 제안한 set을 사용하였다. 적절한 parametrization을 통해 기존의 scoring function보다 decoy를 더 잘 구별할 수 있었으며, 이는 실제 도킹 문제에 적용되었을 때 false positive와 false negative를 줄여주는 결과를 가져올 것이라는 것을 의미한다. 이 논문에서는 주어진 수용체-리간드 쌍에 대하여 구조적인 decoy를 구별하는 것에 대한 연구만이 포함되어 있다. 서로

다른 수용체-리간드 쌍들에 있어서 결합친화도의 차이를 예측하는 데에도 비슷한 방식으로 엔트로피 효과를 고려하여 예측정확도의 향상을 가져올 수 있을 것으로 예상되며, 이 연구는 현재 진행 중에 있다.

방 법

1. 대상 시스템 선정

본 논문에서 사용된 대상 시스템은 Alan et al.,(2005)의 논문에서 제시된 *geometric decoy* 10세트이다. 여기서 *geometric decoy set*은 단백질 데이터 은행(PDB)에 있는 단백질과 리간드의 결합 구조로부터 가져와서 만들어졌다. 이 논문의 저자들은 단백질과 리간드를 떼어낸 후 도킹 계산 프로그램으로 다시 처리하였을 때 얻어지는 구조 중, 원래의 결정 구조와의 차이가 3Å 이하인 구조는 *native like*라고 보며, 3Å 이상 차이가 나는 구조는 *decoy*라고 본다. Alan et al.,(2005)의 논문에서는 추후에 *scoring function*의 개발과 향상에 쓰일 수 있도록 *decoy*가 *native like* 구조보다 더 좋은 점수를 얻은 시스템들, 즉 *현 scoring function*이 실패한 시스템들을 모아서 보고하였다. 원 논문에서는 결합 계산에 DOCK3.5.54 프로그램을 사용하였다. 제시된 *geometric decoy*들은 5가지의 단백질 당 서로 다른 리간드 각각 2개, 즉 총 10개의 세트를 이루고 있다. 이 *decoy*들은 DOCK3.5.54뿐만 아니라 현재 많이 사용되고 있는 다른 여러 도킹 계산 프로그램들에서도 많은 경우 결정 구조와 다른 결합 구조가 더 선호되는, 다시 말해 성공적이지 못한 경향을 보여주고 있다.

그 중에서 본 논문의 결과에 제시된 것들은 PNP 단백질과 8IG, IMR 리간드 결합(PDB ID: 1C3X, 1I80), DHFR 단백질과 TABA, TQ3 리간드 결합(PDB ID: 1D8R, 1IA1), AchE 단백질과 FBQ 리간드의 결합(PDB ID: 1HBJ)이다. 사용되지 않은 나머지 5개의 리간드의 경우에는 리간드가 너무 작고 회전할 수 있는 결합이 없어서 한 가지의 최소점만 얻을 수 있었거나 (이 경우 콜로니 에너지를 쓰더라도 결과에 변화가 없다), 원저자의 실수로 인해 잘못된 리간드 구조를 가지고 계산을 한 경우(*personal contact*)로서 본 계산에서는 제외시켰다.

2. 결합 계산

이번 연구에서는 결합 계산을 하기 위해 DOCK5.3 프로그램을 사용하였다. DOCK5.3 프로그램에서는 단백질과 리간드의 모든 수소가 붙어있는 구조의 mol2 포맷을 가지는 파일을 사용한다. 단백질과 리간드 파일은 AMD Manchester CPU 시스템에서 Sybyl7.1 프로그램을 이용하여 처리하였다. 우선 단백질과 리간드가 결합된 PDB 파일에서 물과 금속 이온을 비롯한 다른 결합에 관여하지 않는 원자들은 제외시켰다.

단백질의 경우 Amber99의 전하를 기본으로 사용하였고, 수소를 위치시킨 다음 수소를 제외한 모든 원자들을 고정시킨 후 수소들의 위치를 에너지가 최소화 되도록 처리하였다. 리간드의 경우 Sybyl atom type에 따라서 atom type을 정의하였으며, 전하는 Gasteiger Hückel 방법으로 구하였다. Grid 계산에서는 기본 parameter set을 사용하였고, 결합 계산 시에는 기본 parameter set에 비해서 더 넓게 위상 공간 안에서 최소점을 찾기 위해서 max_orientation을 100으로 증가시켰다. Random seed를 1에서 5000까지 늘리면서 5000개의 독립적인 계산을 한 후 각 계산에서 가장 좋은 점수를 얻은 한 개의 구조를 선택하여 이 5000개의 구조로부터 colony energy를 계산하였다.

3. Colony energy 계산

Colony energy는 다음과 같이 Xiang등의 논문[13]에서 제시된 방법을 적용하여 구하였다. 먼저 i 번째 구조와 동일한 minimum 안에 존재하고 있는 구조들로 이루어진 ensemble의 population을 P_i 라 하면, 이는 현재 sampling에서 얻어진 구조의 개수 M_i 와 i 주변의 구조로부터 예측되는 P_i 에 대한 기여도 $M_{i,neighbor}$ 의 합으로 쓸 수 있다.

$$P_i = M_i + M_{i,neighbor}$$

$$M_{i,neighbor} = \sum_{i \neq j} \alpha_{ij} M_j$$

$$\alpha_{ij} = \exp(-rmsd_{ij}^3 / \beta N)$$

여기서 β 는 scaling factor이며 이 접근법에서 유일한 parameter이며, N 은 리간드를 구성하는 원자들의 개수이다. 이 방법에서는 거리가 가까운 구조가 많이 sample될수록, 그리고 이들의 에너지가 낮을수록 population에 대한 기여도가 커진다. 즉, 5000개의 독립된 결과에서 각 구조에서 다른 구조와의 root mean square deviation을 계산하여 기준이 된 구조와 비슷한 구조가 많은 경우 그 구조가 포함된 local minimum이 위상 공간 상에서 가지는 크기가 크다고 해석할 수 있다. 이는 물리적인 관점에서 엔트로피가 크다는 것으로 개념적으로 해석할 수 있다. 이를 이용한 자유에너지는 다음과 같다.

$$\Delta G_i = -RT \ln \left[\sum_i \exp(-\Delta E / RT - rmsd_{ij}^3 / \beta N) \right]$$

결과적으로 colony energy 방법은 유사한 구조를 많이 가지고 있는 local minima의 경우 DOCK 계산에 의해서 얻어진 score보다 더욱 낮은 점수를 얻게 되고, 유사한 구조가 적은 경우에는 원래의 점수를 그대로 유지하게 된다. Colony energy 계산 과정에서 population의 크기에 대한 estimate이 어느 정도 고려되어야 할 것인가를 결정하는 것이 β parameter이며, 이를 결정하는 것이 필요하다. 이번 계산에서는 trial and error의 과정을 통해 $\beta = 0.01$ 의 경우 가장 좋

은 결과를 얻을 수 있었다.

마지막으로 DOCK 프로그램 자체의 버그에 대해 언급하겠다. 5000개의 서로 다른 random seed에 대해서 독립적인 결합 계산을 수행하였을 때, 리간드에 따라서 물리적으로 불가능한 구조를 가지는 결과가 얻어지기도 하였다. 즉, 아래의 (그림1) 과 같이 리간드의 탄소, 질소 혹은 산소 원자들이 서로 간에 1Å 이하로 지나치게 가까워져 있는 경우가 종종 발견된다. 이는 DOCK5.3 버전 프로그램의 내부적인 문제로 보이며(personal contact) 차기 버전에서 개선될 것으로 예상된다.

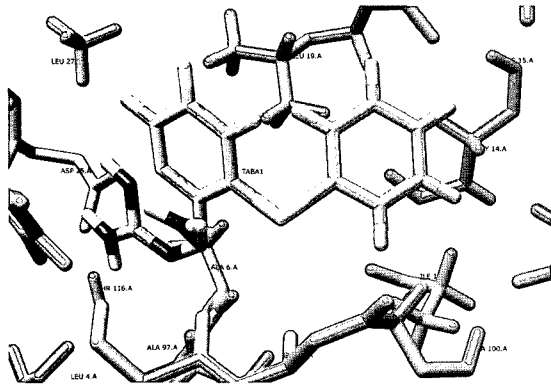


그림 1. DOCK5.3에서 얻어진 물리적으로 불가능한 구조. 벤젠 고리 안에 있는 탄소와 질소간의 거리가 지나치게 짧다.

이와 같이 불가능한 구조들을 post-screen하기 위한 테스트 기준은 다음과 같다: 두 원자 간의 거리가 일반적으로 알려진 van der Waals 반경 값(Bondi, 1964)의 50%이하에 해당하는 쌍이 2개 이상 존재하는 경우 불가능한 구조로 간주한다. DOCK으로부터 얻은 구조들을 post screening을 한 후 이 screening을 통과하지 못한 결과들은 최종적인 콜로니 에너지 계산에 있어서 모두 제외되었다.

결 과

1. DOCK5.3 계산 결과

콜로니 에너지의 영향을 고려하지 않은 DOCK5.3의 계산 결과, 대체로 가장 좋은 score를 가지는 결합된 형태는 결정 구조와 rmsd 3Å 근방에서 얻어졌다. 이는 Graves et al.의 논문[15]에서 DOCK3.5.54를 이용해 얻어진 값에 비해서 대체로 향상된 결과이다.

2. 콜로니 에너지 계산 결과

콜로니 에너지 방법을 적용하여 앞서 나온 결과들을 다

시 평가하였고, 각 리간드에 대한 결과는 (표 1)에 나타나 있다. (그림 2)에서는 5개의 리간드에 대한 결과를 결정 구조와의 rmsd를 x축으로 하고 계산된 에너지 값을 y축으로 하여 나타내었다. 왼쪽에 위치한 그림이 DOCK에서 계산된 에너지이고, 오른쪽은 그 결과를 바탕으로 콜로니 에너지 방법을 적용한 것이다.

DOCK 에너지만을 사용하였을 경우 8ig를 제외한 4개의 리간드 들은 가장 낮은 score를 가지는 conformation이 결정 구조와 이루는 rmsd의 크기가 크게 다르지 않음을 알 수 있다. 그러나, 8ig 리간드의 경우 콜로니 에너지 방법을 적용하지 않았을 경우, 4.23Å의 rmsd 값을 가지나, 콜로니 에너지 방법을 적용하였을 경우에는 rmsd가 2.54Å인 구조가 가장 좋은 score를 보여주고 있다. 이는 콜로니 에너지 방법이 결합 계산 시에도 정확도를 높이는데 기여할 수 있는 가능성이 있음을 보여주고 있다.

표 1. DOCK5.3 에서 가장 낮은 결합에너지를 갖는 구조와 Colony energy를 계산했을 때 얻어진 가장 낮은 결합에너지를 갖는 구조와 결정 구조 사이의 rmsd값

rmsd(Å)	Dock5.3	Colony E
TABA	2.52	2.52
TQ3	3.02	3.13
8IG	4.23	2.54
IMR	2.61	2.61
FBQ	4.72	4.77

또, DOCK의 계산 결과에서는 rmsd가 1~2Å정도 밖에 차이가 나지 않음에도 불구하고, 두 구조간의 에너지 차이가 크게 나는 문제가 종종 일어난다. 위 그림 왼쪽 그림들에서는 비슷한 rmsd 크기를 가지는 구조들의 에너지 분산이 상당히 크다는 것을 확인할 수 있다. 이런 경우, 콜로니 에너지를 적용함으로써 비슷한 구조가 클러스터링 (clustering)이 되는 결과를 보여준다. 그 결과, rmsd와 결합 에너지 간의 비례하는 상관관계가 증가함을 확인할 수 있다. 다시 말해, rmsd가 작을수록 결합 에너지가 더 낮아지고, rmsd가 클수록 결합 에너지가 더 높아지는 경향이 더욱 강하게 나타나고 있다.

물론, 위의 단순한 rmsd - 결합 에너지 다이어그램만으로 완전히 local minimum 들을 클러스터링 해서 분리할 수는 없겠지만, 구조 유사성을 이용한 클러스터링과 동시에 사용된다면, 결합 계산에 있어서 위상 공간 상에 존재하는 최소 점들의 위치와 공간상에서 그 최소점이 차지하는 크기를 더 정확히 확인할 수 있을 것이다.

이는 8ig 리간드의 경우 fbq 리간드보다 콜로니 에너지가 더 향상된 결과를 가져오는 이유 중의 하나라고 생각된다.

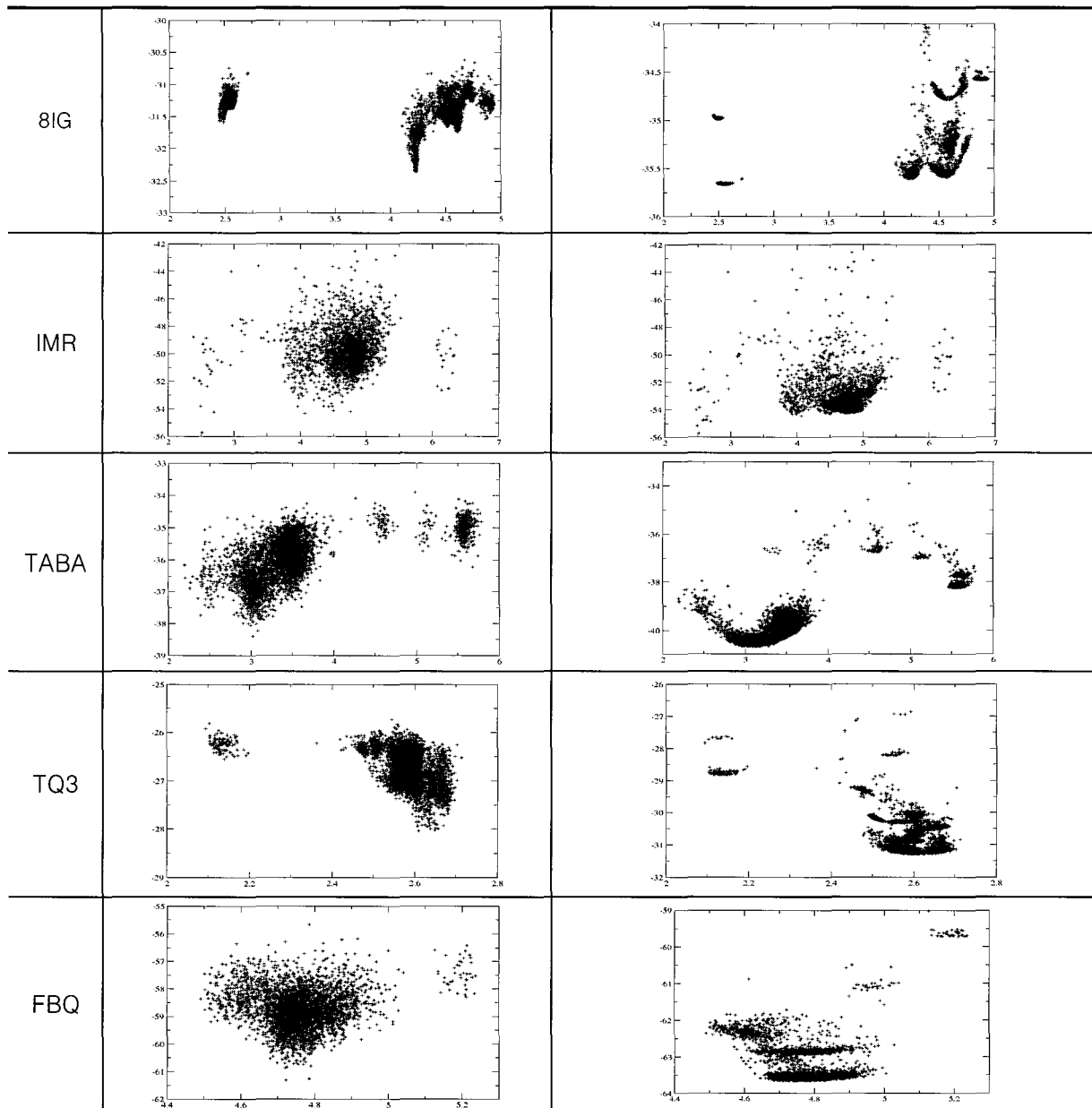


그림 2. 결합에너지 (왼쪽 : 에너지, 오른쪽 : colony 에너지) 대 결정 구조로부터의 rmsd 값

rmsd라는 수치 자체가 모든 원자들의 변화를 결과적으로 하나의 숫자로 환원 시킨 것이기 때문에, 리간드가 실제로 어떤 부분이 어느 정도 변하는지를 정확하게 고려하기가 어렵다. 하지만 이러한 효과는 작고 회전할 수 있는 가지(branch)의 수가 많지 않은, 단단한 분자(rigid molecule)의

경우, 분자의 내부적인 움직임이 거의 없다고 가정할 수 있으므로 리간드 구조간의 rmsd의 크기가 곧 각 구조간의 차이를 매우 정확하게 반영한다고 할 수 있다.

결론

본 논문에서는 이미 널리 사용되고 있는 결합 계산 프로그램들이 x-ray 결정 구조를 제대로 재현해 내지 못하는 *geometric decoy* 들에 대해서 콜로니 에너지를 이용하여 엔트로피 효과를 고려했을 경우, 어느 정도 결과의 정확성을 향상시킬 수 있는지 살펴보았다.

계산 결과, 8ig 리간드와 PNP 단백질 간의 결합에서 엔트로피 효과를 고려하지 않았을 경우에는 예측된 구조가 결정 구조와 rmsd 값이 4Å 이상 차이가 났으나, 콜로니 에너지 방법을 적용하였을 경우에는 2.53Å의 rmsd 값을 가지는 구조가 가장 낮은 에너지를 가지도록 결과가 향상 되었다. 다른 *geometric decoy* 들에 대해서는 단순히 에너지만을 평가 함수로 쓰는 경우와 대체로 비슷한 결과를 주었으며, 전체적으로 rmsd와 에너지 간의 상관관계가 더 강해지는 긍정적인 결과를 얻을 수 있었다.

이는 구조간의 유사성을 이용해서 엔트로피를 고려하는 방법이 결합 계산의 정확성을 높이는 데 기여할 수 있는 가능성을 보여주었다고 할 수 있으며, 특히 8ig 리간드의 경우와 같이 리간드가 작고, 회전 가능한 가지의 개수가 적을 때, 더 잘 적용될 수 있을 것이라 예상된다.

그러나 가능성뿐만 아니라, 콜로니 에너지 방법의 한계에 대해서도 고려해야 할 측면이 있다. 콜로니 에너지 방법은 계산 과정에 임의의 parameter가 들어간다. 본 논문에서 사용된 parameter는 직접 다양한 범위의 parameter를 적용시켜 본 후, 계산에 사용된 5가지 시스템 모두에 대해서 상관관계를 가장 잘 보여주는 값으로 결정하였다. 앞으로 더 많은 시스템에 적용시키기 위해서는 이 값이 다양한 시스템에 대해서도 좋은 결과를 살펴보고, 이 parameter를 가능한 한 많은 시스템에 대해서 training하여 최적화하는 것이 필요할 것이다.

참고문헌

- [1] Shoichet BK. (2004) Virtual screening of chemical libraries. *Nature*, 432:862-3.
- [2] Encyclopedia of computational chemistry, vol 2, (2005) John Wiley & Sons, Inc.
- [3] Brandsdal BO, Österberg F, Almlöf M, Feierberg I, Luzhkov VB, Åqvist J. (2003) Free energy calculations and ligand binding. *Adv Prot Chem* 66: 123-58.
- [4] Kollman PA. (1993) Free energy calculations: Applications to chemical and biochemical phenomena. *Chem Rev* 93:2395 - 2417.
- [5] Beveridge DL, Dicapua FM. (1989) Free energy via molecular simulations: application to chemical and biochemical system. *Annu Rev Biophys Biophys Chem*, 18: 431 - 92.
- [6] Huo S, Wang J, Cieplak P, Kollman PA, Kuntz ID. (2002) Molecular dynamics and free energy analysis of Cathepsin D inhibitor interactions: Insight into structure based ligand design. *J. Med. Chem.*, 45:1412-9.
- [7] Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE. (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.*, 33:889-97.
- [8] Aqvist J, Medina C, Samuelsson JE. (1994) A new method for predicting binding affinity in computer aided drug design. *Protein Eng.*, 7: 385-91.
- [9] Hansson T, Aqvist J. (1995) Estimation of binding free energies for HIV proteinase inhibitors by molecular dynamics simulations. *Protein Eng.* 8: 1137-44.
- [10] Aqvist J. (1996) Calculation of absolute binding free energies for charged ligands and effects of long range electrostatic interactions. *J. Comput. Chem.*, 17: 1587-97.
- [11] Smith RH Jr., Jorgensen WL, Tirado Rives J, Lamb ML, Janssen PA, Michejda CJ, Smith MBK (1998) Prediction of binding affinities for TIBO inhibitors of HIV 1 reverse transcriptase using Monte Carlo simulations in a linear response method. *J. Med. Chem.*, 41: 5272-86.
- [12] Wang W, Wang J, Kollman PA. (1999) What determines the van der Waals coefficient beta in the LIE (linear interaction energy) method to estimate binding free energies using molecular dynamics simulations? *Proteins*, 34:395-402.
- [13] Xiang Z, Soto CS, Honig B. (2002) Evaluating conformational free energies: The colony energy and its application to the problem of loop prediction. *PNAS*, 99:7432-7
- [14] DOCK5.3. http://dock.compbio.ucsf.edu/DOCK_5/index.htm.
- [15] Alan PG, Ruth B, Shoichet BK. (2005) Decoys for docking. *J. Med. Chem.*, 48:3714 - 28.
- [16] Sybyl, version 7.1. Tripos Inc., <http://www.tripos.com>.
- [17] Bondi A. (1964) van der Waals volume and radii. *J. Phys. Chem.*, 68:441-51.