Molecular docking to EGFR tyrosine kinase domain : Structural Validation against Crystal Structures

장준영, Art E. Cho*

세종특별자치시 세종로 2511 고려대학교 세종캠퍼스, 생명정보공학과, 30019

Epidermal growth factor receptor(EGFR)는 HER family에 속하는 tyrosine kinase receptor로서 다양한 하류경로로 신호를 전달하여 세포 증식, 혈관 형성, 세포 사멸을 억제하는 역할을 한다. EGFR이 폐암의 형성에 중요한 역할을 하고 많은 상피세포 종양에서 비정상적으로 활성화됨에 따라 암 치료에 중요한 역할을 하고 있어 EGFR tyrosine kinase inhibitor(TKI)에 관한 많은 연구가 이루어졌다. 위와 같은 약 개발에 있어서 현재 가상 시뮬레이션을 통한 약 후보물질 개발이 진행되고 있다. 특히, Molecular docking 시뮬레이션은 기존의 실험적인 기술(X-ray crystallography, NMR)로는 연구하기가 어려웠던 protein과 ligand간의 상호작용을 예측하여 이에 대한 정보를 제공할 수 있다. 하지만, 우선적으로 Molecular docking 시뮬레이션은 정확한 validation을 기반으로 진행되어야 신뢰할 수 있는 정보를 얻을 수 있다. 따라서 이번 연구에서는 EDISON에서 제공하는 Dock 프로그램과 일반적으로 잘 알려진 Glide, Autodock 프로그램으로 protein data bank(PDB)에서 제공하는 EGFR wild type cocrystal을 redocking하는 방식을 통하여 최상위 rank pose의 RMSD 값을 통한 validation 성능을 비교함으로써 어떤 프로그램이 EGFR과 ligand 간의 결합예측을 하는데 있어서 보다 더 정확한 결과를 낼 수 있는지 알아보고자 하였고 시뮬레이션 결과 Autodock에서 가장 우수한 결과 값을 보여주었다.

핵심어 : Molecular docking, Validation, Edison Dock, Glide, Autodock

서론

Epidermal growth factor receptor(EGFR)은 human epidermal growth factor receptor (HER) family 에 속하는 tyrosine kinase(TK) receptor 로서 [1] 리간드와 결합한 후 이합체를 형성하고, tyrosine kinase domain 의 autophosphorylation 에 의해 활성화된다. 그 후 다양한 하류경로로 신호를 전달하고, 그것에 의하여 세포 증식, 침투, 전이와 혈관 형성에 영향을 주며 세포 사멸을 억제한다.[2]

EGFR 은 폐암의 형성에 중요한 역할을 하는 것으로 밝혀 졌으며 이에 대한 암 치료로 가장 연구가 많이 된 것이 EGFR 의 TK domain 억제제로 현재 시중에 gefitinib 과 erlotinib 이 쓰이고 있다.[3,4] 하지만 약제 내성으로 인하여 약 효능이 떨어짐에 따라 새로운 약 개발 시도가 이루어지고 있다.[4]

위와 같은 신약개발 방법에는 가상으로 약 후보 물질을 선정할 수 있는 docking 시뮬레이션이 있다. Docking 은, 표적 단백질의 활성자리에 약물을 결합시켜 양자역학적 계산 또는 potential 에너지 함수 기반의 결합 에너지를 계산함으로써 적당한 물질을 찾아낼 수 있다.[5] 상용 또는 범용으로 개발된 Autodock, Glide 등 다수의 Docking 프로그램에서 흔히 사용되는 방법은 활성자리를 하나의 Grid 혹은 공간으로 정의하고 활성자리 주변 잔기들과의 Van der Waals 에너지, 정전기적 에너지, steric 에너지 등을 계산하여 단백질 활성자리 안에서의 화학적으로 안정한 구조를 찾아내고 각 에너지의 합산을 통해 Docking score 를 결정하여 물질의 순위를 결정한다.[5] 이렇게 순위가 결정된 화합물을 대상으로 필요한 계산 요소를 추가 계산하여 최종적으로 후보물질을 선정하게 된다.[5] 이와 같은 Docking 시뮬레이션을 통한 신약 후보 물질을 선정하기 위해서는 우선적으로 docking program 의 정확한 결합예측을 통한 validation 이 필요하다.[6]

따라서 본 연구에서는 EDISON 에서 제공하는 DOCK 프로그램과 일반적으로 사용되는 Autodock, Glide 프로그램을 이용하여 여러 개의 EGFR TK co-crystal 을 re-docking 해봄으로써 각 프로그램의 docking scoring 을 비교해 보았다. 이에 따라 어떠한 프로그램이 docking 시뮬레이션을 통한 EGFR 의 신약 후보 물질 개발에 적합한지 알아보고자 한다.

이론 및 계산방법

Docking 시뮬레이션을 진행하기 위해 protein data bank(PDB)에서 총 15 개의 wild-type EGFR 모델을 구하였다.

maestro 10.2 프로그램을 이용하여 crystal 의 물분자 제거, 수소 첨가, bond order 수정을 해주었고 Prime Refinement 를 사용하여 protein 의 side chain missing atom 을 예측해 주었다. 리간드들은 maestro 프로그램으로 각 프로그램 파일 형식에 맞게 추출하였다. Grid box 의 중심 좌표는 cocrystal ligand 를 기준으로 잡았다.

본 연구에서는 Edison 에서 제공하는 Dock 프로그램과 일반적으로 사용되는 glide, autodock 프로그램을 사용하였다.

Glide. OPLS force field 와 empirical scoring function 인 glide score 를 사용하며 standard precision mode 에서 docking calculation 을 진행하였다.

Autodock. Lamarckian genetic algorithm 과 Amber force field 를 사용하고 리간드에 Gasteiger charge 값을 부여해 주어 docking calculation 을 진행하였다.

Edison dock. Autodock vina 프로그램을 이용하며 Autodock 과 다른 knowledge-based 와 empirical approach 를 융합한 scoring function 을 사용한다.

분석 방법은 re-docking 과정을 거쳐 각 프로그램의 scoring function 에 따라 측정된 리간드의 pose 와 co-crystal 의 리간드 pose 와의 heavy atom 을 기준으로 root mean square deviation (RMSD)값을 측정하였다. RMSD 는 2.0 Å 이하일 때 성공적인 결합 예측으로 간주된다.[6]

$$\mathrm{RMSD}(\mathbf{v},\mathbf{w}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}((v_{ix}-w_{ix})^2+(v_{iy}-w_{iy})^2+(v_{iz}-w_{iz})^2)}$$

Figure1 : RMSD equation

결과 및 토의

Table 1. Docking results : Top rank rmsd values (Å)
Rmsd 값이 2 Å 이상인 결과는 빨간색으로 표기하였다.

PDB	Glide	EDISON Dock	Autodock
1M17	1.48	2.61	1.64
4WKQ	3.10	3.12	1.75
1XKK	3.04	2.60	2.59
2J6M	1.30	1.27	1.42
3POZ	2.90	1.37	1.66
3W32	1.94	1.62	1.48
3W33	1.94	0.62	1.45
3W2S	2.00	0.71	1.03
4JQ7	0.60	6.83	1.09
4JQ8	0.69	0.62	0.94
4JR3	0.68	2.62	1.24
4JRV	2.75	2.90	0.84
5CAV	0.81	1.41	0.92
4ZAU	0.98	5.44	3.17
3BEL	1.07	4.19	1.45
RMSD 평균	1.69	2.53	1.51

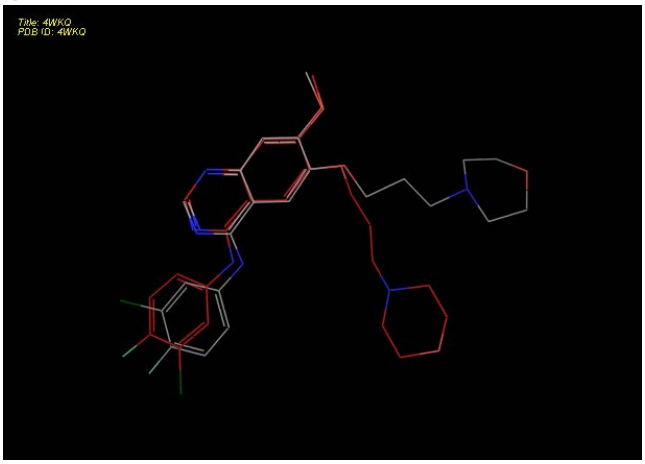

Figure3 : 4WKQ Glide docking result. 회색(cocrystal ligand), 빨간색(Docking pose)

Glide – 4WKQ, 1XKK, 3POZ, 4JRV 모델에서 rmsd 값이 2 를 초과하였고 4WKQ 모델에서 3.10 으로 가장 큰 값을 가졌다(figure3). 나머지 11 개의 모델에서는 모두 rmsd 가 2 이하의 값을 가졌고 가장 낮은 rmsd 값을 가진 모델은 4JQ7 로 0.60 값을 가졌다(figure2). 전체적인 rmsd 평균 값은 1.69 로 측정되었다.

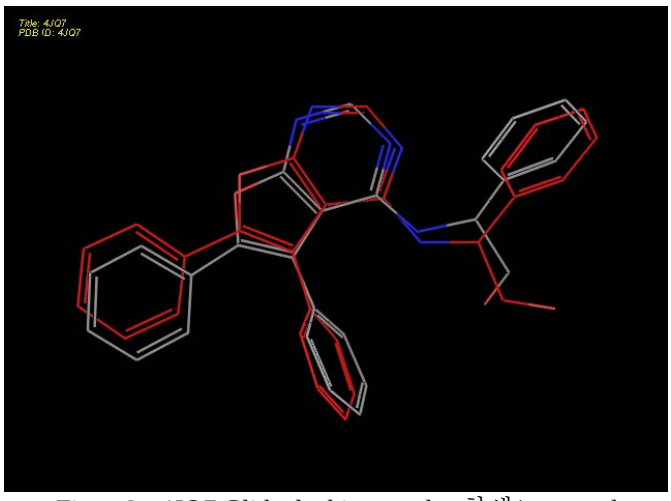

Figure2 : 4JQ7 Glide docking result. 회색(cocrystal ligand), 빨간색(Docking pose)

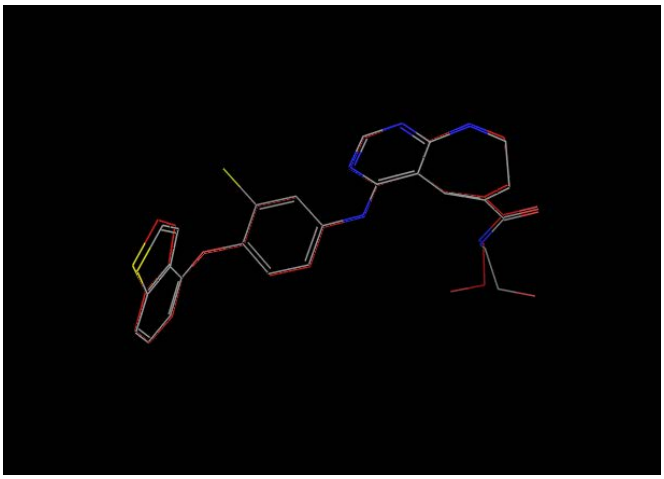

Figure4 : 3W33 Edison Dock docking result. 회색(cocrystal ligand), 빨간색(Docking pose)

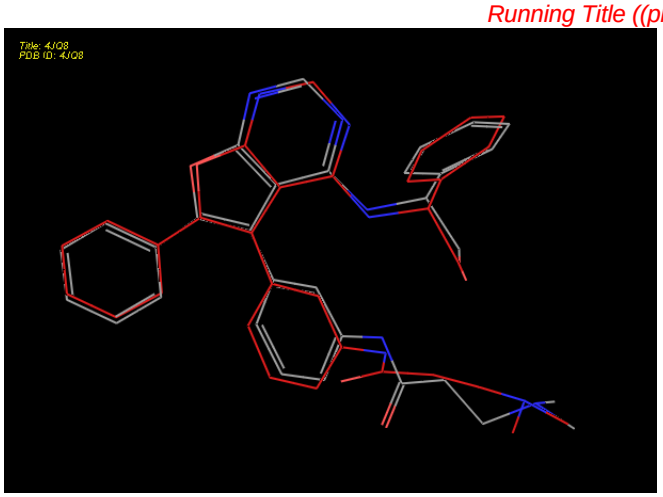

Figure5 : 4JQ8 Edison_Dock docking result. 회색(cocrystal ligand), 빨간색(Docking pose)

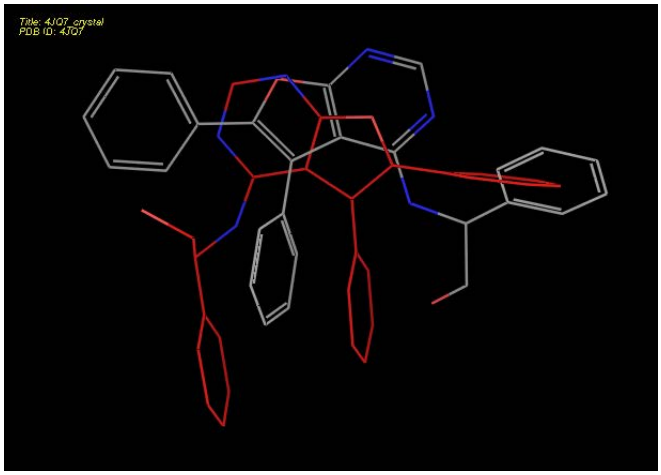

Figure6 : 4JQ7 Edison_Dock docking result. 회색(cocrystal ligand), 빨간색(Docking pose)

Edison dock – 1M17, 4WKQ, 1XKK, 4JQ7, 4JR3, 4JRV, 4ZAU, 3BEL 모델에서 rmsd 값이 2 를 초과하였고 4JQ7 모델에서 6.83 으로 가장 큰 값을 가졌다(figure6). 나머지 7 개의 모델에서는 모두 rmsd 가 2 이하의 값을 가졌고 3W33(figure4), 4JQ8(figure5) 모델에서 공통으로 rmsd 최저 값 0.62 를 가졌다. 전체적인 rmsd 평균 값은 2.53 으로 측정되었다.

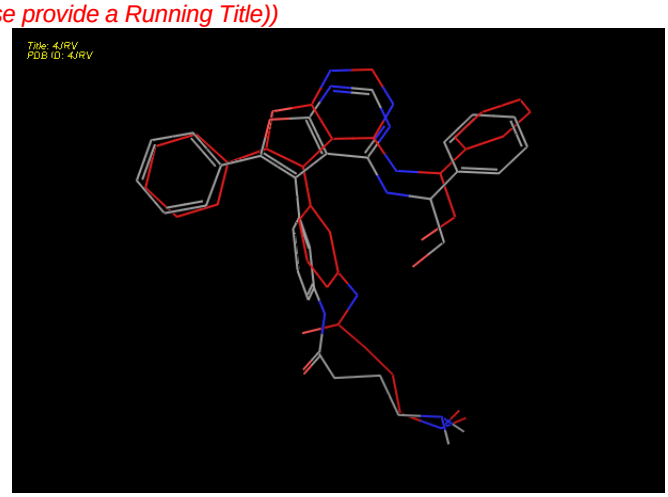

Figure7 : 4JRV Autodock docking result. 회색(cocrystal ligand), 빨간색(Docking pose)

Figure8 : 4JAU Autodock docking result. 회색(cocrystal ligand), 빨간색(Docking pose)

Autodock – 1XKK, 4ZAU 모델에서 rmsd 값이 2 를 초과하였고 4ZAU(figure8) 모델에서 3.17 로 가장 큰 값을 가졌다. 나머지 13 개의 모델에서는 모두 rmsd 가 2 이하의 값을 가졌고 4JRV(figure7), 모델에서 rmsd 최저 값 0.84 를 가졌다. 전체적인 rmsd 평균 값은 1.51 으로 측정되었다.

결론

본 연구에서는 EGFR 을 타겟으로 Autodock, Glide, Edison dock 총 3 개의 프로그램을 이용하여 docking 시뮬레이션을 진행하였다. Docking 결과 Autodock 프로그램에서 가장 우수한 validation 값을 가졌다.

Glide 에서도 Autodock 못지 않게 우수한 validation 값을 가졌으나 Edison dock 에서는 전체 평균 rmsd 값이 2.53 인 것을 볼 때 EGFR 에 대해 정확한 결합 예측 계산을 해주지는 못한 것으로 판단된다. 비록 Autodock 프로그램이 가장 우수한 결과를 가졌지만 1XKK, 4ZAU 에 대해서는 정확한 결합 예측 계산을 해주지는 못하였다. 또한 Autodock 에서는 4ZAU 에 대한 rmsd 는 3.17 로 좋지 못하나 Glide 에서는 rmsd 값이 0.98 로 좋은 결합 예측을 보여 주었다. 이번 연구에서 보여 주는 것과 같이 비록 Autodock 이 EGFR 에 대한 validation 이 평균적으로 좋을 지라도 부분적으로 한계점을 보여주었다. 따라서 이러한 pose prediction 에 보다 더 정확한 scoring function 개발이 필요해 보이며 위와 같은 한계점을 극복한다면 신약 개발 연구에 큰 도움이 될 것으로 기대해 본다.

감사의 말. 이 연구는 EDISION 과제를 통해 지원받아 수행되었습니다.

참고문헌

1. Cohen S, Ushiro H, Stoscheck C, Chinkers M. A native 170,000 epidermal growth factor receptor-kinase complex from shed plasma membrane vesicles. J Biol Chem 1982; 257: 1523-31.
2. Normanno N, De Luca A, Bianco C, *et al*. Epidermal growth factor receptor (EGFR) signaling in cancer. Gene 2006; 366: 2-16.
3. Scagliotti GV, Selvaggi G, Novello S, Hirsch FR. The biology of epidermal growth factor receptor in lung cancer. Clin Cancer Res 2004; 10: 4227-32.
4. Pao W, Miller VA, Politi KA, et al. Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. PLoS Med 2005;2:e73.
5. 이재영, 최지원, 신재민, 노경태. 컴퓨터를 이용한 신약개발, 향후 기술 전망. 정보과학회지 2013.
6. Mark Agostino, Cassandra Jene, Tristan Boyle, Paul A. Ramsland, Elizabeth Yuriev. Molecular Docking of Carbohydrate Ligands to Antibodies: Structural Validation against Crystal Structures. J. Chem. Inf. Model. 2009; 49: 2749-2760