

SBDD를 이용한 best covid19 Mpro inhibitor 선정과 선정된 inhibitor에 대한 비교분석

covid19 단백질 PDB ID: 5rh0
의생명시스템학부 20210387 박채린

1. Introduction

covid19으로 세계가 들썩인 시기가 있었다. Covid19은 severe acute respiratory syndrome coronavirus 2의 감염으로 나타나는 질병을 의미하며, 전 세계적으로 680만 명 이상의 사망을 초래하였다.¹

SARS-CoV-2의 게놈은 약 30,000nt로 구성되어있으며, SARS-CoV-2의 genome은 pp1a, pp1ab 두 개의 중첩되는 polyprotein으로 직접적으로 전사된다. pp1a과 pp1ab는 바이러스의 복제와 전사 기능을 위해 잘리고 processed 되어야 한다.³ 기능을 담당하는 폴리펩타이드들은 광범위한 단백질 분해 과정에 의해 polyprotein으로부터 방출되며, 주로 Mpro(=SARS-CoV-2 main protease = 3C-like protease)에 의해 수행된다. Mpro는 pp1a와 pp1ab8에서의 효소자가분해절단을 시작으로 polyprotein을 최소 11개의 conserved site에서 잘라낸다. 인간에게는 Mpro와 가까이 유사한 homologues가 없기에, Mpro의 바이러스에서의 기능적 중요성을 고려하면 Mpro는 매력적인 항바이러스 약물의 타겟이 된다.²

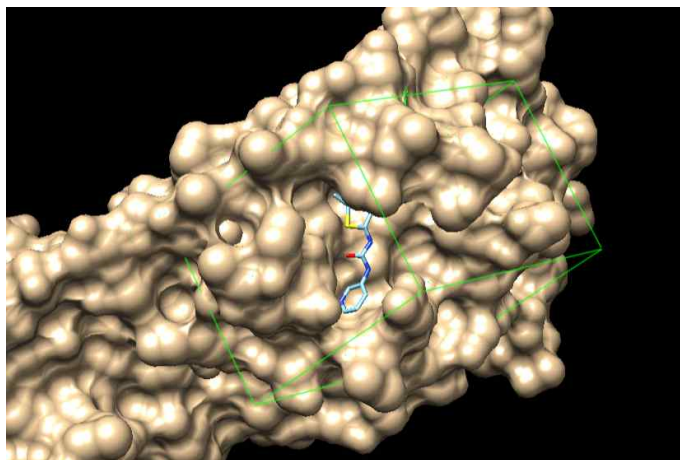
SARS-CoV-2는 주로 S 단백질의 수용체 결합 도메인(RBD)을 통해 숙주 세포 표면의 ACE2와 결합하여 숙주 세포를 감염시킨다. 따라서 단백질 분해 효소에 의한 pp1a/pp1ab의 가수분해는 바이러스 RNA 복제 및 전사 과정에서 없어서는 안 될 핵심 단계이며 COVID-19의 수명 주기에서 중요한 역할을 한다. Mpro는 3개의 도메인을 갖는 시스테인 단백질분해효소로, 전구체 단백질에서 가장 성숙한 절단 사건에 관여한다. Mpro를 억제하면 바이러스 복제를 방지할 수 있으므로 Mpro protease는 중요한 항바이러스 표적이 된다. Mprp에 공유 결합하여 작용하는 억제제는 생체 내에서 약물이 오래 작용하도록 할 수 있지만, 독성 부작용 및 시간을 연장하는 데 도움이 될 수 있지만 독성, 부작용, 표적화에 대한 문제를 가진다. 비공유 억제제는 약한 가역적 결합을 통해 표적 단백질에 경쟁적으로 결합하기 때문에 이러한 위험을 크게 줄일 수 있다. 따라서, 비공유 결합 억제제는 Mpro에 대해 더 높은 선택성을 가질 수 있다. 그러나 Mpro의 catalytic pocket의 아미노산 잔기와 약한 가역적 상호작용을 통해 억제 활성을 나타내기 때문에 약물의 투여 시간을 연장하거나 용량을 늘려야 할 필요가 생길 수 있으며, 이는 약물 내성으로 이루어질 수 있다.³ 따라서 Mpro의 비공유 억제제는 최대한 강하게 bind할 수 있어야 이를 극복할 수 있을 것이다.

본글에서는 drug design을 위해 SBDD를 이용하였다. SBDD는 컴퓨터를 이용하여 구조를 기반으로 해 bind site에 가장 잘 bind할 drug을 찾아낸다. Mpro에 대한 더 좋은 inhibitor drug를 design하는 연구를 통해, 최대한 강하게 bind할 수 있는 drug를 찾아 기존의 약물보다 더 부작용이 적고 효율이 좋은 약물을 개발해낼 수 있을 것이다. 이러한 과정을 통해 코로나 바이러스의 확산에 두려워하지 않을 수 있고, 다른 감염병이 등장하더라도 본 연구와 유사한 방법을 사용하여 효율적인 drug를 빠르게 만들어낼 수 있을 것이다.

2. Methods

tool로는 smina, chimera, vmd, caFE, cgenff, pubchem을 이용하였다.

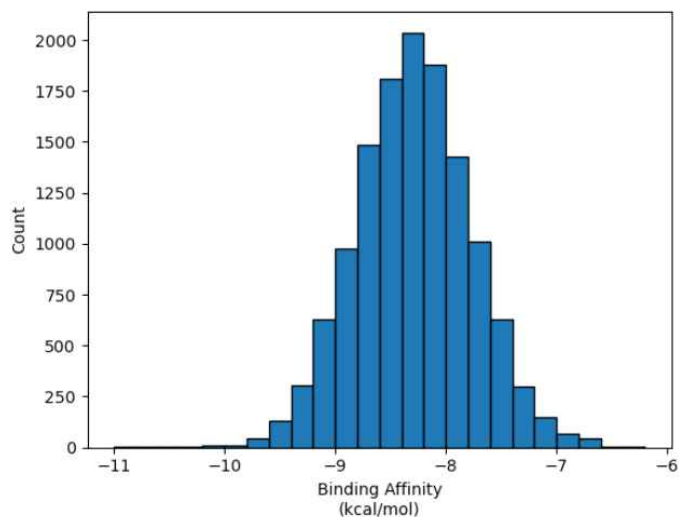
도킹에 사용한 데이터는 Mpro inhibitor가 될 수 있는 분자의 cluster enamine 1, 2, 3, 4 중 enamine 2, 3, 4를 사용하여 총 12950개의 drug를 후보로 이용하였다. enamine은 QSAR를 통해 먼저 도킹이 가능할 것으로 보이는 분자만 걸러낸 ligand이다. drug의 receptor가 될 Mpro는 PDBID: 5rh0 데이터를 이용하였다. 5rh0 complex로부터 receptor와 ligand를 분리하여 Mpro 단백질을 얻었다.



<그림1> Mpro의 binding box

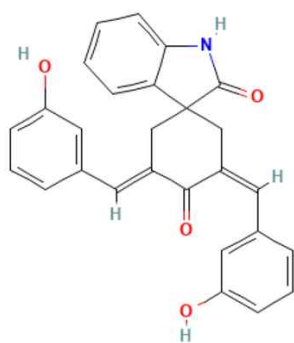
Mpro의 active pocket은 5rh0 데이터에서 기존의 ligand가 붙어있던 곳을 기준으로 근처의 움푹 파여있는 vdw surface일 것이라 가정하였고, 이에 따라 도킹 시에 이용한 binding box는 <그림1>과 같았다. Mpro와 drug의 bind 전과 bind 후의 free energy(글 전체에서 Gibb's free energy를 이용하였음, 단위는 kcal/mol) 변화량이 가장 작은, 즉 가장 음의 방향으로 절댓값이 큰 drug를 smina를 통해 docking하여 찾았다. 얻은 drug의 sdf 파일을 mol2파일로 변환시켜 주었으며, pubchem으로 수소를 붙인 후의 3D 구조가 일치하는지 확인하였다. 후에 chimera를 이용해 net charge가 0이 되게 add charge를 수행하였고, 완성된 파일을 이용하여 추후 작업을 수행하였다. 해당 drug를 Mpro 단백질과 합쳐 complex 파일을 형성해 MD simulation을 18.968ns 수행하여 dcd 결과 파일을 얻었고, 결과 파일을 대략 1000프레임이 되도록 stride 옵션을 통해 잘랐다. Drug만으로 md를 수행한 파일과 complex 파일을 통해 LIE를 수행하여 보다 정확하게 해당 drug가 Mpro에 bind하였을 때의 free energy 변화량이 몇인지를 얻어내었다.

3. Results and Discussion

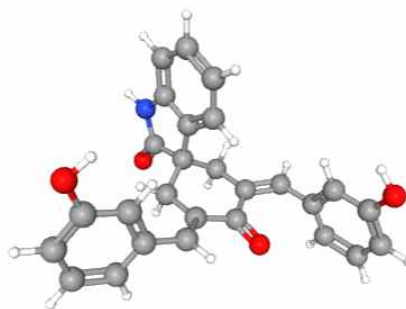


<그림2> 12950개의 분자 도킹 결과 얻은 자유에너지 변화량에 대한 히스토그램

Mpro inhibitor가 될 수 있는 분자의 cluster enamine 2, 3, 4를 사용하여 총 12950개의 drug를 후보로 이용하였고, smina를 통해 얻은 도킹 전과 후의 자유에너지 변화량을 <그림 2> 와 같이 히스토그램으로 나타내었다. 가장 자유에너지 변화량이 작은, 즉 가장 도킹 후에 안정한 complex를 형성한 ligand는 Z569412310로, (2'Z,6'E)-2',6'-bis[(3-hydroxyphenyl)methylidene]spiro[1H-indole-3,4'-cyclohexane]-1', 2-dione 라는 분자이다. 이 분자는 도킹 후에 -10.95406의 자유에너지 변화량을 가졌다.

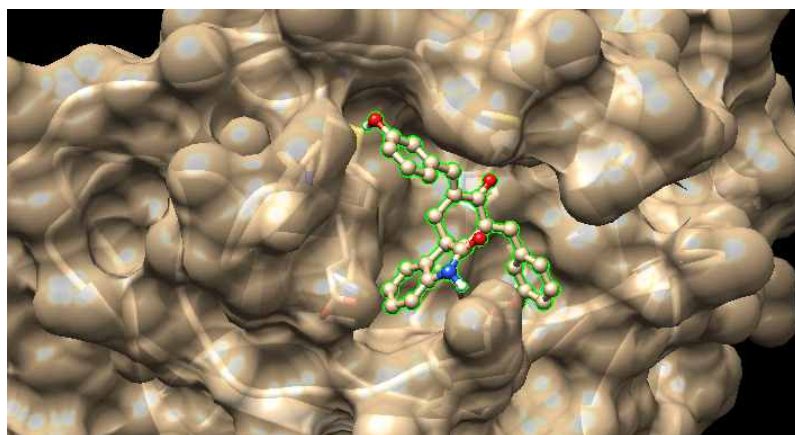


<그림3> Z569412310의 2D 구조



<그림4> Z569412310의 3D 구조

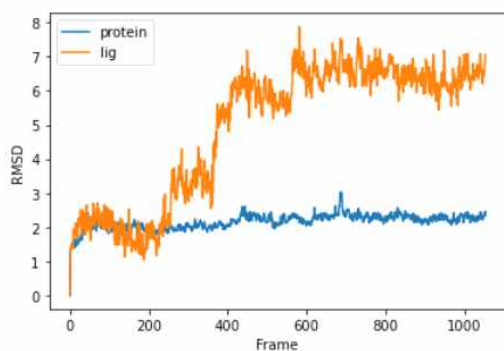
Z569412310는 <그림3>, <그림4>의 구조를 갖는 분자이다.



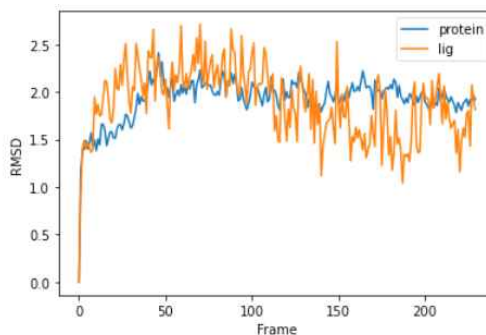
<그림5> Mpro와 Z569412310의 complex

Mpro와 해당 drug가 complex를 형성하면 <그림5>와 같은 모습을 띤다.

먼저 Mpro와 drug의 complex가 평형에 도달하여 md simulation이 수행되었는지를 확인하기 위해 RMSD 그래프를 그려주었다. 파이썬의 MDAnalysis package를 이용하였다.



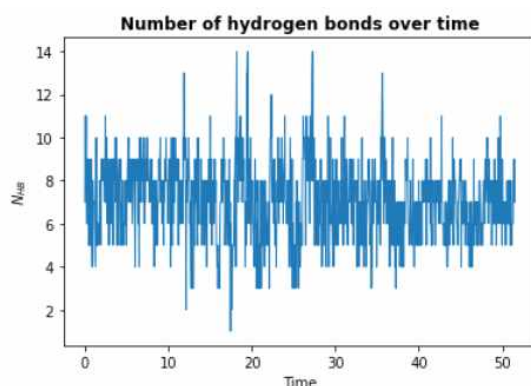
<그림6> 전체 frame에 대한 RMSD



<그림7> 230번 frame까지의 RMSD

18 stride option을 적용한 파일을 이용한 <그림6>과 <그림7>에서는 1frame이 $1000\text{fs}(\text{dcdfreq}) * 18(\text{stride}) = 18000\text{fs} = 18\text{ps}$ 를 의미한다.

<그림7>을 통해 200frame까지는, 즉 3.60ns 즈음까지는 ligand의 RMSD값이 protein과 비슷하며 변화량의 크기가 1Å 내외에서 움직이므로 ligand와 protein이 결합한 상태로 평형에 도달하였다고 추측할 수 있다. 하지만 3.60ns이후로 ligand의 RMSD값이 크게 증가하다가 600 frame 이후부터 일정한 값을 갖게 되는데, 600 frame 이후부터는 ligand와 protein이 결합하지 않고 떨어진 상태로, 또는 5Å 정도의 거리에서 vdw 등의 상호작용을 하는 상태로 평형에 도달하였을 것이라고 추측할 수 있다.



<그림9> 시간에 따른 complex의 수소결합 수 변화

600 frame 이후부터는 ligand와 protein이 결합하지 않고 떨어졌다고 가정한다면 일정 시간 이후부터는 수소결합의 수가 줄어들어야한다. 하지만 hydrogen bond를 형성하는 hydrogen과 acceptor를 각각 (resname LIG and around 10A not resname LIG)로 설정하고 시간에 따라 형성된 수소결합의 수에 대한 그래프를 보았을 때에는, 수소결합의 개수가 크게 변화하는 구간이 없었고 대체로 비슷한 수를 유지하는 결과가 나타났다. 따라서 수소결합의 개수가 ligand와 Mpro 사이의 수소결합이 아닌, ligand끼리, 또는 Mpro끼리의 수소결합이 count 되었거나, drug와 Mpro가 5A의 거리에서도 수소결합을 아직 맺고 있을 것이라 생각해볼 수 있다. 하지만 수소결합은 길어도 3.3A 거리까지만 상호작용이 가능하기에 ligand와 Mpro 사이의 수소결합이 아닌, ligand끼리, 또는 Mpro끼리의 수소결합이 count 되었을 것일 가능성이 더 높다.

```

1 -----
2 Title          Frames          Mean          SD
3 -----
4 Bound:
5 Vdw:           1054          -55.0140         3.9054
6 Elec:           1054          -43.8032         6.8590
7 -----
8 Free:
9 Vdw:           1054          -25.8797         8.5826
10 Elec:          1054          -37.9673        13.9300
11 -----
12 Delta:
13 Vdw:                   -29.1343
14 Elec:                   -5.8359
15 -----
16 Final: (alpha=0.18, beta=0.06, gamma=0.0)
17 Total:                   -5.5943
18 -----
19 * All energy values are in kcal/mol

```

<그림8> LIE 수행 결과

얻은 drug와 Mpro가 결합하였을 때의 정확한 free energy 변화량을 구하기 위하여 ligand only, complex 파일을 형성하였고 md simulation을 수행하여 dcd 파일을 얻은 뒤 LIE를 수행해 <그림8>과 같은 결과를 얻었다. 결과적으로, Mpro와 Z569412310가 결합하였을 때의 binding free energy는 -5.5943으로 smina를 통해 도킹하여 얻었을 때의 값보다는 그 크기가 줄어든 것을 확인할 수 있었다.

| Title | Frames | Mean | SD |
|---|--------|----------|---------|
| Bound: | | | |
| Vdw: | 230 | -56.7029 | 3.6712 |
| Elec: | 230 | -41.9092 | 6.4466 |
| Free: | | | |
| Vdw: | 230 | -30.0062 | 7.7634 |
| Elec: | 230 | -42.7086 | 13.0454 |
| Delta: | | | |
| Vdw: | | -26.6967 | |
| Elec: | | 0.7995 | |
| Final: (alpha=0.18, beta=0.06, gamma=0.0) | | | |
| Total: | | -4.7574 | |
| * All energy values are in kcal/mol | | | |

<그림9> 230 frame까지만 잘라 수행한 LIE의 결과

<그림6>과 <그림7>에서의 결과를 바탕으로 230 frame까지를 기준으로 하여 LIE를 수행하면 drug와 Mpro가 잘 붙어있는 상태에서 LIE가 수행되기 때문에 binding free energy가 더 낮은 값이 나올 것이라 기대하였고, LIE를 수행해보았지만, -4.7574로 오히려 크기가 더 작은 free energy가 나오게 되어, 5Å 거리만큼 떨어진 complex가 drug와 Mpro가 가까이 있는 complex보다 조금 더 안정하다는 결과를 도출해내었다.

[illegible]

<그림10> 5rh0 기존 ligand를 Mpro에 smina로 도킹한 결과

5rh0 데이터에서 기존에 Mpro와 결합하던 N-(5-methylthiophen-2-yl)-N'-pyridin-3-ylurea와 도킹을 통해 얻은 Z569412310 drug의 binding free energy와 형성되는 complex의 안정성

을 비교하기 위해서, N-(5-methylthiophen-2-yl)-N'-pyridin-3-ylurea를 Mpro에 동일한 binding box를 이용하여 도킹하였다. 그 때의 binding free energy는 -5.69216이었다. 따라서 Z569412310는 -10.95406의 자유에너지 변화량을 가졌기 때문에, smina의 도킹 결과에 따르면 Z569412310가 기존의 5rh0에서의 ligand보다 더 안정한 complex를 형성하여 더 binding affinity가 강한 drug임을 결론지을 수 있다. 하지만 LIE 수행 결과를 Z569412310의 실제 binding free energy라고 가정하면 오히려 5rh0에서 결합하던 ligand보다 불안정한 complex를 형성함이 되기 때문에, N-(5-methylthiophen-2-yl)-N'-pyridin-3-ylurea를 이용하여 MD를 진행하고 LIE를 수행하여 LIE의 결과끼리 비교한다면 어느 ligand가 가장 Mpro에 잘 bind하여 그 기능을 잘 inhibit할 수 있는지를 비교하여 알아낼 수 있을 것이다.

이번 연구에서 반성할 점은 smina를 돌릴 때 파일을 잘못 넣어 complex의 md simulation을 20ns를 채워 수행하지 못한 점과, 기존 5rh0 ligand 또한 md simulation을 통해 LIE를 통해 binding free energy를 구하였다면 유의미한 비교를 진행할 수 있었을 것이라는 점이다. 또한 enamine cluster 1에 포함된 ligand들은 도킹을 수행하지 못하여, 이 속에 Z569412310보다 더 좋은 drug가 발견되었을 수 있다는 가능성이 있다.

4. Conclusion

| binding free E (kcal/mol) | 5rh0 기존 ligand | 새로운 best ligand | 더 좋은 리간드 |
|---------------------------|----------------|-----------------|----------|
| smina로 docking | -5.69216 | -10.95406 | 기존<신규 |
| MD 후 LIE | ? | -5.5943 | ? |

Smina를 이용한 도킹 결과에 따르면, 5rh0 기존 ligand보다 Z569412310가 Mpro에 더 강하게 bind하여 더 안정한 complex를 형성하므로, Mpro inhibitor로서 더 좋은 기능을 수행할 것으로 기대된다.

Z569412310와 Mpro가 5Å 거리만큼 떨어진 complex가 Z569412310와 Mpro가 매우 가까이 있는 complex보다 조금 더 안정하다.

5. Reference

1. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10113999/>
2. <https://pubmed.ncbi.nlm.nih.gov/32272481/>
3. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10333551/>