Full Electronic Structure Calculation of Quantum Mechanics in Binding Site of Protease 6LU7

Wenjin Zhou[a] , Xiaodong Wu[b], Linxiang Zhou[c*]

[a] *Department of Computer Science, University of Massachusetts-Lowell,USA*

[b] *Executive Medical Physics Associates, Miami, Florida, USA*

[c] *Department of Physics, Xiamen University, Xiamen, China**

Here we report the binding site (We call it as drug target) of the COVID-19 main protease 3CLpro 6LU7 using Full Electronic Structure Calculation of Quantum Mechanics. The protease 6LU7 has 18038 elecrons , we need to calculate its wave funciton (it is called the molecular orbitor in the chemistry field) and energy level for this huge elecronic system. Then according to three rules to obtain the exact active pocket, active resudes, even active atoms of 6LU7. Finally put active pocket and all candidate small-molecule drugs or peptides group into an artificial intelligence deep learning 3D-CNN program to calculate the binding affinity. The result is that the pocket3 with three active atoms CE3, CZ3 and CH2 of 6LU7 is real design target for small-molecule drugs and the drug Tipranavir is located on the first position with binding affinity about 5.4. Our calculations show that the binding affinity of Remdesivir with 3CLpro 6LU7 is very small, less than 0.5. This just shows that Remdevivir (GS-5734) is the first channel. Because Remdesivir is not a nucleotide analog, it is a prodrug of nucleotide analogs, it has many protective groups around it. When the drug enters the body, these atomic groups will act as enzymes in the cell It is broken down, leaving the active drug and though the drug N3 company with 6LU7, but its binding affinity is only about 4.2, so it could not be a real drug for 6LU7. The quantum mechanics calculation give out such accurate drug target. We have to take up pocket 3 to design inhibitor drug. The drug screening is only the first step in the long and costly pharmaceutical process, but we prove that it can be completely calculation by quantum mechanics. Hopefully, this method can provide an effective and rapid screening method for finding new coronavirus drugs.

*Corresponding author. Email: lxzhouno1@gmail.com

Introduction

on October 28, 2019, at the World's Leading Scientists Forum, the 2013 chemical Nobel Prize winner Michael Levitt suggested that **quantum mechnics may be a better method for pharmaceuticals.**

Indeed, we have to let the drug screening based on the solid basic quantum mechnics calculation.

Though the drug screening is only the first step in the long and costly pharmaceutical process. The discovery and confirmation of drug targets are the basis for screening and discovering active compounds. If there is a deviation in the understanding of drug targets, the entire drug discovery and research work will go astray. But for a long time, the method of drug screening is only to select a thousand drugs every day from a large drug library, only based on the size of free energy. Finally, dozens of drugs are identified, and then experiments are performed. In this classic method, there is no knowledge of the fine microstructure of a drug's protein target.

As today if someone wants to know the spectrum of the hydrogen atom, what should he do to take an experiments or quantum mechanics calculations? Quantum mechanics calculations of the hydrogen spectrum can be as accurate as the number of bits that can be reached with the experimental results.

The 20th century was the century of quantum mechanics, almost all modern technologies were based on quantum mechanics. The basic idea of quantum mechanics is that any microscopic

particle has wave-particle duality,following the Schrödinger wave equation. Quantum mechanics believes that any property of any molecule depends on the wave function of the molecule (called the molecular orbital in the chemical world) and the energy level. Therefor we believe that protein-ligand interactions and protein-protein interactions also should depend on their wave function and energy level. So we can sure that the drug screenng should can be calculated by quantum mechanics. It can substitue classical drug screening method.

However, the protein acts as a biological vitality in its entirety. No matter which amino acid residue it replaces, its vitality cannot be lost, and the vitality can only be changed. This shows that only its full electronic structure is biologically active It is closely related, and by no means a local electronic structure like QM/MM method.

Here the calculation of protein wave functions and energy levels is called electronic structure calculation.

Method

We have created a Quantum Drug Screening (QDS) method[1-4] that combines calculation of the full electron structure of a protein with its hydrophobic pocket and frontier-orbital-perturbation theory. The QDS program uses a divide-and-conquer algorithm approach that recursively breaks down a complex problem into many related sub-problems. These sub-problems are independent of each other and are of the same character as the original problem, so that we can use the solutions of these small sub-problems to obtain the solution of the original, larger problem. First, the whole chain of a protein is divided into many dimers, each containing two

amino acid residues. Next, we perform quantum calculations for the all dimers and construct a large full Hamiltonian matrix using the solution from each dimer. But this large full Hamiltonian matrix is near a diagonalization of blocks, so we can use ENFC theorem in Math to obtain the eigenvalues. In addition, the solution effects must be considered. A few atoms are placed at the end of each dimer to simulate the chemical environment in which the dimer is located.

The system of a protein molecule is so large that the energy levels are extremely compact. Therefore, we need to extend the definition of the frontier orbitals for a protein to be a series of HOMOs and LUMOs, rather than being limited to a single HOMO and LUMO.

For example, The protease 6LU7 has 18038 elecrons and orbital-occupied number NOCC = 9019. It has total 306 residues, so we divided it into 306 dimers. And take 20 orbitals as LUMOs, it is 0.0022 of total orbitals only.

How to determine the active pocket and active atoms? There are three rules[5]: a necessary and sufficient condition for the active residue or active atom of the protein in protein-ligand interaction:

(1) They must be located on LUMOs.

(2) They must fall in a hydrophobic pocket on the surface of the protein, and one active residue is not in two different pockets at the same time.

(3) Their orbital coefficients cannot be zero.

After quantum calculation we will put active pocket and active atoms and all candidate small-molecule drug group into an artificial intelligence deep learning 3D-CNN program[6] to calculate the binding affinity. The training data PDBBind use Chinese Academy of Sciences Shanghai Organic Chemistry graduate School (http://www.pdbbind.org.cn/index.asp)[7].

Results and Discussion

The 3CLpro 6LU7 structure is from ShanghaiTech University, Institute of Pharmacy Jointly Announces High-resolution Crystal Structure of New Pneumonia Coronavirus 3CL Hydrolase(www.rcsb.org/structure/6LU7)[8]. It is a 3CLpro protease. The 3CLpro 6LU7 has 306 residues with 18038 electrons. The function of the COVID-19 main protease 3CLpro 6LU7 is to synthesize proteins through genetic information, to replicate and package a new generation of viruses. Inhibiting it prevents the virus from replicating within the host cell, but it cannot kill existing viruses.

The wave function and Forbidden Band of 6LU7

Its wave function (or call molecular orbital) of a protein only distribute on the several residues, not distribute on the whole protein. Especially, the most frontier orbital local on the those active residues.

That means the distribution of frontier orbital has a kind of intrinsic link with the biological activity of a protein.

There is a wide forbidden band between valence band and conductive band in a protein, which about 10~11 eV.

When there is no exogenous charge carrier, the valence band is full, but the conductive band is empty. Such big forbidden band can not let the heat motion bring electron from the valence band to the conductive band to form a carrier. Therefore, a 6LU7 protein is a insulation when there is no exogenous carrier (electron or hole).

The LUMOs (the lowest unoccupied molecular orbitals) of 6LU7

The system of a protein molecule is so large that the energy levels are extremely compact. For such systems, we need to extend the definition of the frontier orbitals to be a series of HOMOs and LUMOs, not merely a single HOMO or LUMO. Thus, the first 10–20 occupied and unoccupied molecular orbitals were selected. The protein-ligand interaction only occurs between the LUMOs of protein and the HOMO of its ligand, not between the HOMOs of protein and the LUMO of its ligand.

```
================================================================
No.            Energy Band   Position of Wave Function
               (eV)          (residue)      (occ_percent)
----------------------------------------------------------------
9015(20)       0.10864       SER_267          1.000
9016(21)       1.43488       TRP_218          0.998
9017(22)       1.69699       TRP_207          0.999
9018(23)       1.70670       PHE_223          0.995
9019(24)       1.74113       TRP_31           0.999
9020(25)       1.81587       TYR_209          0.999
9021(26)       1.81615       PHE_219          0.995
9022(27)       1.83430       PHE_219          0.994
9023(28)       1.83874       PHE_223          0.992
9024(29)       1.87378       PHE_159          0.996
9025(30)       1.96893       PHE_181          0.798
                             TYR_182          0.201
9026(31)       1.97417       PHE_8            0.996
9027(32)       1.97983       TYR_54           0.999
9028(33)       1.98462       PHE_159          0.999
9029(34)       1.98868       PHE_181          0.213
                             TYR_182          0.787
9030(35)       1.99246       PHE_185          0.997
9031(36)       1.99961       TYR_161          0.998
9032(37)       2.00847       TYR_237          0.997
9033(38)       2.01931       PHE_66           0.991
9034(39)       2.03968       TYR_37           0.999
9035(40)       2.07164       PHE_181          0.998
================================================================
```

Figure 1. LUMOs of 6LU7

Residues List Composed of Pockets (only list larger four pockets)

Pocket 1: Thr24 Thr25 Thr26 Leu27 His41 Cys44 Thr45 Ser46 Met49 Pro52 Tyr54 Phe140 Leu141 Asn142 Gly143 Ser144 Cys145 His163 His164 Met165 Glu166 Leu167 Pro168 Hes172 Asp187 Arg188 Gln189 Thr190 Gln192

Pocket 2: Gln107 Pro108 Gly109 Gln110 Pro132 Ile200 Thr201 Val202 Asn203 Glu240 Pro241 Leu242 Asp245 His246 Ile249 Thr292 Pro293

Pocket3: Glu14 Gly15 Cys16 Met17 Val18 **Trp31** Ala70 Gly71 Val73 Leu75 Asn95 Pro96 Lys97

Pocket4: Phe8 Gln110 Thr111 Gln127 Asn151 Ile152 Asp153 Thr292 Phe294 Asp295

1. The above residue with bold is active residue.

2. There are no active residues in pocket1 and pocket2.

3. Although pocket 1 is much larger than the other pockets, it has no active residue. And pocket 3 and pocket 4 are both small, **which tells an important fact that in the screening of small molecules of this protein, drugs must be small.**

POC:	N_mth	Area_sa	Area_ms	Vol_sa	Vol_ms
POC: 1	1	352.018	554.707	**323.734**	937.892
POC: 2	1	103.528	230.380	**68.474**	286.176
POC: 3	1	71.131	123.005	**49.770**	182.987
POC: 4	1	78.590	144.487	**47.681**	195.010

Though pocket3 and pocket4 have active residues, but we can see from LUMOs, the activity of Trp31 in pocket3 is much greater then Phe8 in pocket4 because the active residue Trp31 is located in 5th position of LUMOs, but Phe8 in 12th of LUMOs. So, **the pocket3 is real design target for small-molecule drugs instead of looking at the entire 6LU7 protease as a target. This is the difference between quantum method and classical method. Only quantum mechanics calculation give out such accurate target! We have to take up pocket 3 to design inhibitor drug.**

So far, we can see that 6LU7 has two active pockets: pocket3 and pocket4. The pocket3 has three active atoms: CE3, CZ3 and CH2. The pocket 2 has two active atoms: CE2 and CZ.
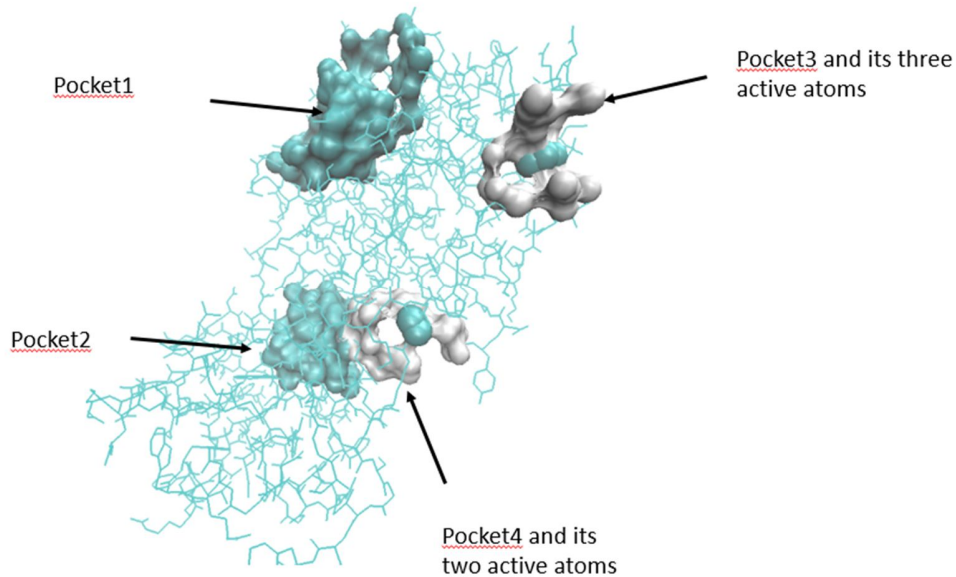

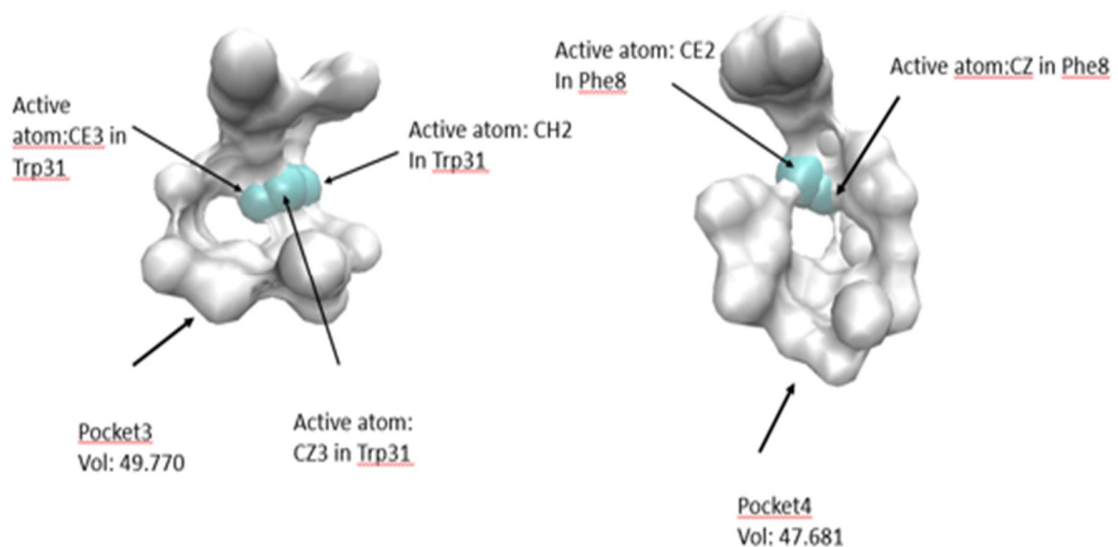Figure 2. Pockets and active atoms of 6LU7

Figure 3. Pocket3 , pocket4 and their active atoms

Run 3D-CNN to find binding affinity

Input pocket3 with residue Trp31 and its three active atoms in 6LU7 and 27 Nucleoside Analogues into 3D-CNN program to find the binding affinity. A nucleoside analog is a structurally modified nucleoside acid. This modification is usually performed on a five-carbon sugar.

These 27 Nucleoside Analogues are some from ShanghaiTech University

(http://www.cas.cn/syky/202001/t20200125_4732909.shtml), some from the RCSB protein bank (www.rcsb.org)[9] as follows:

Abacavir   Atazanavir   Carfilzonvib   Chloroquine

Conformer3D_CID_479503   Conformer3D_CID_4124851

Darunavir   DB00822   DB05448   DB09010   DB12129

Desoxycytidin   Desoxythymidin   Emtricitabine   Favipiravir

GC376(3CL-pro-1)   Indinavir   Lamivudine   Lopinavir

N3   Raltegravir   Remdesivir   Ribavirin   Ritonavir

Saquinavir   Tipranavir   Zidovudine

The results on the binding affinity as Figure 4:

	A	B
1	name	binding affinity
2	Tipranavir	5.4369173
3	Chloroquine	5.296913
4	Emtricitabine	5.2732882
5	DB12129	5.2558193
6	Zidovudine	5.1494913
7	Conformer3D_CID_4124851	5.147705
8	Conformer3D_CID_479503	5.103711
9	DB05448	5.0905895
10	DB09010	5.067651
11	Abacavir	5.065818
12	Indinavir	4.998952
13	Raltegravir	4.9055414
14	Darunavir	4.900685
15	Lopinavir	4.8285456
16	Desoxythymidin	4.7767634
17	DB00822	4.710844
18	Lamivudine	4.7105045
19	Desoxycytidin	4.691762
20	Atazanavir	4.6150427
21	Remdesivir	4.607145
22	Saquinavir	4.583124
23	Carfilzonvib	4.5791874
24	Ritonavir	4.521985
25	Favipiravir	4.51707
26	GC376	4.470148
27	Ribavirin	4.281646
28	N3	4.224579

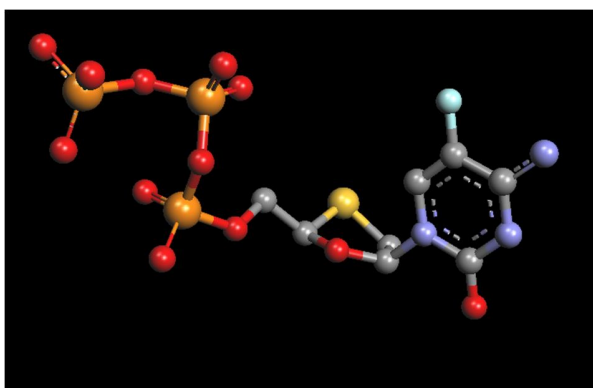Figure 4. The binding affinity of Nucleoside Analogues

We can see that four nucleoside analogues have strongest binding affinities:
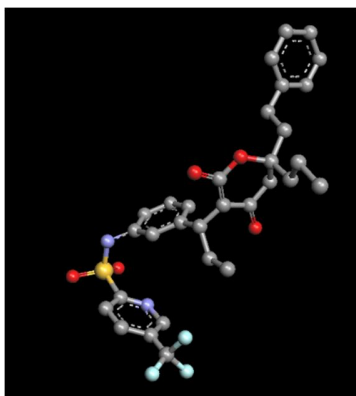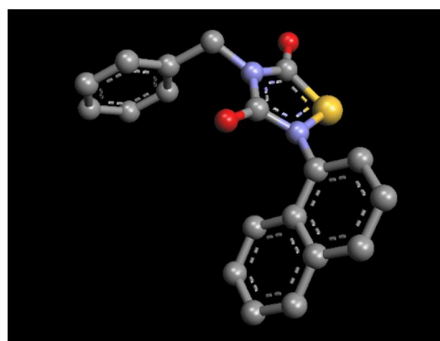
Tipranavir
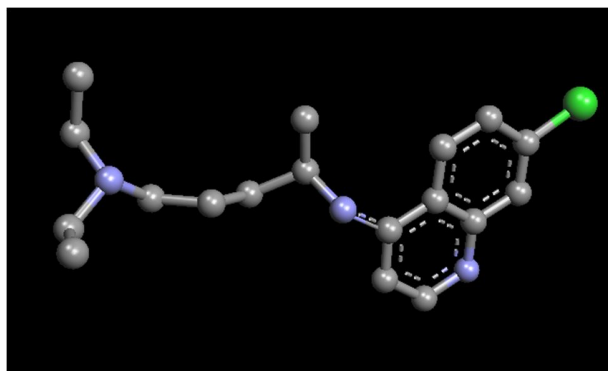
Chloroquine

Emtricitabine

DB12129

We have to be aware of is a paper published by NEJM and Science considered that maybe, Remdesivir by Gilead company is very good drug for cureing coronavirus. On the other hand, our calculations show that the binding affinity of Remdesivir with 3CLpro 6LU7 is very small, less than 0.5. This just shows that Remdevivir (GS-5734) is the first channel. Because Remdesivir is not a nucleotide analog, it is a prodrug of nucleotide analogs, it has many protective groups around it. When the drug enters the body, these atomic groups will act as enzymes in the cell It is broken down, leaving the active drug.

At a press conference of the Joint Prevention and Control Mechanism of the State Council held on February 17, Sun Yanrong, deputy director of the Biological Center of the Ministry of Science and Technology, stated that it can be clear that chloroquine phosphate has a certain effect in the treatment of new crown pneumonia. (http://www.jjckb.cn/2020-02/19/c_138796714.htm).

According to Figure 4, we can see that the drug Chloroquine is located on the second position, so it is indeed a good drug. But the drug Tipranavir is located on the first position.

Finally, we take smaller drug DB12129 to do docking using autoDock program[10], the results are as follows:

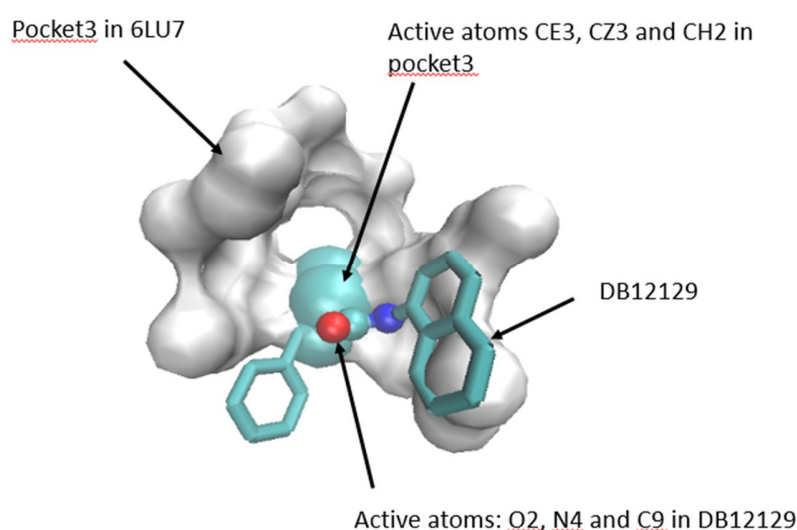DB12129 has three active atoms O2,N4 and C9 on its HOMO.

Figure 5. Docking between 6LU7 and DB12129

Conclusion

This quantum drug screening method based on quantum mechanical calculations of full-electronic structures of proteins provides directly access to the fine structure of active pockets, active residues and even active atoms of the protease 3CLpro 6LU7 . Just as today it is necessary to know that the spectrum of hydrogen atoms can be calculated by quantum mechanics, not by experiment. This method make drug screening based on solid basic science of quantum mechanics. The drug screening as the first step in long and costly pharmaceutical process can be calculate by full-electronic quantum mechanics, also not by experiment.

Acknowledgement

References

[1] Zhou, L. X., & Shan, J. W. (2008). Protein computer simulation.

Xiamen: Xiamen University.

[2] Ladik, J. *et al.*, *Chem. Phys.*, 108:203 (1986).

[3] Ye, Y. J., *J. Math. Chem.*, 14:121 (1993).

[4] Tian et al., Nucleic Acids Res. 2018. PMID: 29860391 DOI: 10.1093/nar/gky473.

[5] Wenjin Zhou, Allison M. Rossetto, Xiaodong Pang & Linxiang Zhou "Computational full electron structure study of biological activity in Cyclophilin A", at Accepted online: 12 Aug 2015.Published online: 22 Sep 2015.

[6] https://gitlab.com/cheminfIBB.

[7] http://www.pdbbind.org.cn/index.asp.

[8] http://www.shanghaitech.edu.cn/.

[9] https://www.rcsb.org/.

[10] http://autodock.scripps.edu/.

[11] http://www.biodiscover.com/news/celebrity/736118.html.