Biophysics project – Interaction of Sars-CoV2 Spike and human ACE2

Igor Trujnara, Anna Chechenina, Anastasia Sukhorukova

November 2022

1 Introduction

Interaction is a fundamental factor that assures the proper function of any protein in cells. Moreover, in order to infect and operate in the human body, infectious agents have to recognize and attach to some ligands on the cellular surface. These interactions allow a proper recognition of particular target cells using receptor molecules. The receptor-ligand interaction is a dynamic and complex process that requires the residues of the interaction site to have exceptionally high affinity to the target surface.

With the increasing amount of genomic and proteomic data, there arose many ways to model the desired interaction and identify its key parameters. However, there are still many difficulties and unsolved problems. First of all, with the rising amount of structures, it became harder to identify the correct ones. Since all proteins exist in the form of an ensemble of different states depending on outer conditions, even structures obtained for the same protein could vary. The structures in the public databases added from different sources could have different input parameters because there is no unique pipeline for the model construction. One example of this problem is the presence of hydrogens or water molecules in the structure. Some structures in databases could be added without hydrogens, some structures have some additional atoms in the model. Thus, the important part of any interaction investigation is the preprocessing of the raw structures.

Besides problems with raw structure data, another source of problems is an actual interpositioning of the examined structures. Since there are many possible configurations of the surface, it might need a significant computational effort to properly model and predict the actual interaction residues.

All interactions are based on the energies between atoms involved, the overall likelihood of interaction can be estimated by calculating the Gibbs free energy. The interaction model contains many biophysical terms to consider, however, the classical simplified approach that we will be using in our research includes only three types of energy: electrostatic, van der Waals, and solvation energy.

In this study, we examined the model of the SARS-CoV-2 spike protein interaction with its receptor, human angiotensin-converting enzyme 2 (ACE2).

The emergence of the highly pathogenic coronavirus SARS-CoV-2 in Wuhan and its rapid international spread has posed a serious global public health emergency. The situation made the development of therapies to prevent future epidemics of paramount importance. One of the most popular ways to induce an immune response to the virus or its inactivation is based on the specific antibodies that recognize and attach to the particular epitope of one of the viral surface proteins. In the case of SARS-CoV-2, the key protein for the viral life cycle is spike protein. This protein binds to the ACE2 receptor on the surface of the lung, intestine, heart, kidney, and alveolar epithelial type II cells and mediates the viral entry. It was shown that spike protein could attach to some other cellular proteins, but the entry using ACE2 seems to be the primary mechanism [5].

Spike protein is a 180–200kDa protein with an extracellular N-terminus, a transmembrane domain anchored in the viral membrane, and a short intracellular C-terminal segment. Recent crystallographic and electron cryo-microscopic (cryo-EM) studies have provided details into the structure of the SARS-CoV2 S protein, resolved in its free state in both closed and open conformations. The receptor-binding domain of the spike protein (RBD, 319–541 residues) is the one responsible for the interaction with ACE2 receptor in the region of aminopeptidase N [3, 8].

It was shown that the most important residues for the binding are V504, Y505 with Q506, D442, Y495 with F497 and D442, N448 with K444, forming H-bond networks in the structure of the SARS-CoV2 spike protein and re-configuring during after the binding [6]. In another paper it was shown that at the N terminus of 1, Q498, T500, and N501 of the RBD form a network of H-bonds with Y41, Q42, K353, and R357 from ACE2. In the middle part, K417 and Y453 of the RBD interact with D30 and H34 of ACE2, respectively. At the C terminus of 1, Q474 of the RBD is H-bonded to Q24 of ACE2, whereas F486 of the RBD interacts with M82 of ACE2 through van der Waals forces [9]. Also, particular mutations in residues S19P and E329G of ACE2 provide noticeable variations in their intermolecular interactions with the viral spike protein [4].

The RBD region of the SARS-CoV2 is the primary target for the neutralizing antibodies. The residues involved the most in the interaction appear to be the key targets of the antibodies. So the investigation of the essential residues could be an essential background for the antibody design.

Thus, the aim of our research was to verify results on the protein interaction between ACE2 and spike protein of the SARS-CoV2 and identify the most important residues for this biding using the Gibbs energy method. We identified the most valuable residues and compared them to results obtained in other papers.

2 Methods

2.1 Structure preparation

The structure of the Spike-ACE2 complex, obtained by X-ray crystallography, was fetched from RCSB PDB. The structure was analyzed and fixed using the `biobb_structure_checking` Python module [2]. The structure was checked for alternative locations, abnormal amide interactions, side chain chirality, side chain clashes and backbone breaks. Water molecules, small ligands and metal ions were removed. Hydrogen atoms were removed and re-added to ensure integrity.

2.2 Preliminary exploration

The structure was analyzed visually by using PyMOL [7]. Interfaces were assessed by using the program's distance measurement option. Multiple options were assessed until a suitable distance was found. An offset was added to the result to ensure complete inclusion of both interfaces.

2.3 Interface determination

The rest of the analysis was performed using `Biopython` [1] in Google Colab. The first part of the analysis was determining the residues that are part of the interface. This was done by measuring minimum distance from each residue to a residue in the other chain, and checking each residue against the previously determined cutoff distance.

2.4 Energy analysis

Energy analysis was performed in several steps. Extra information on atoms was added from an atom data library, obtained externally. An energy analysis function was created, which accounts for residue solvation, atom electrostatic interaction and atom van der Waals interactions. With this function, energy was assessed for the entire structure, as well as for the interfaces.

Subsequently, Ala-scanning was performed. To this end, a modified version of the energy analysis function was used, which calculates interaction energies for non-alanine atoms only. The function was used to determine numerical influence of each residue on the interaction. Certain residues were determined to be potentially significant using an arbitrary threshold.

3 Results

First, we have identified the residues in ACE2 and Spike that contributed significantly to the overall interaction energy of the Spike-ACE2 complex. This was done initially by selecting residues in Spike within 6Å of Spike and vice versa.

This distance was first determined by means of trial and error making various selections in Pymol, to try and select all the residues inside the interaction site.

The interaction set we found consisted of 45 residues from ACE2 and 38 residues from Spike, located within less than 6Å of each other (see Supplementary, Table 1).

We proceeded to compute the total energy of the interaction between all the residues in the two proteins, as well as the total energy of the interaction inside the interaction set only. By subtracting the two we got a difference of -15.03J, which we deemed small enough to settle for the 6Å distance and use this interaction set as our best estimate in the next steps.

By performing a simplified Ala-scanning procedure on the interaction set, we were able to identify the residues in Spike and ACE2 proteins that made the most significant contribution to the interaction of the two proteins, namely Lys353 in ACE2 and Tyr505 in Spike, as shown in Figure 1.

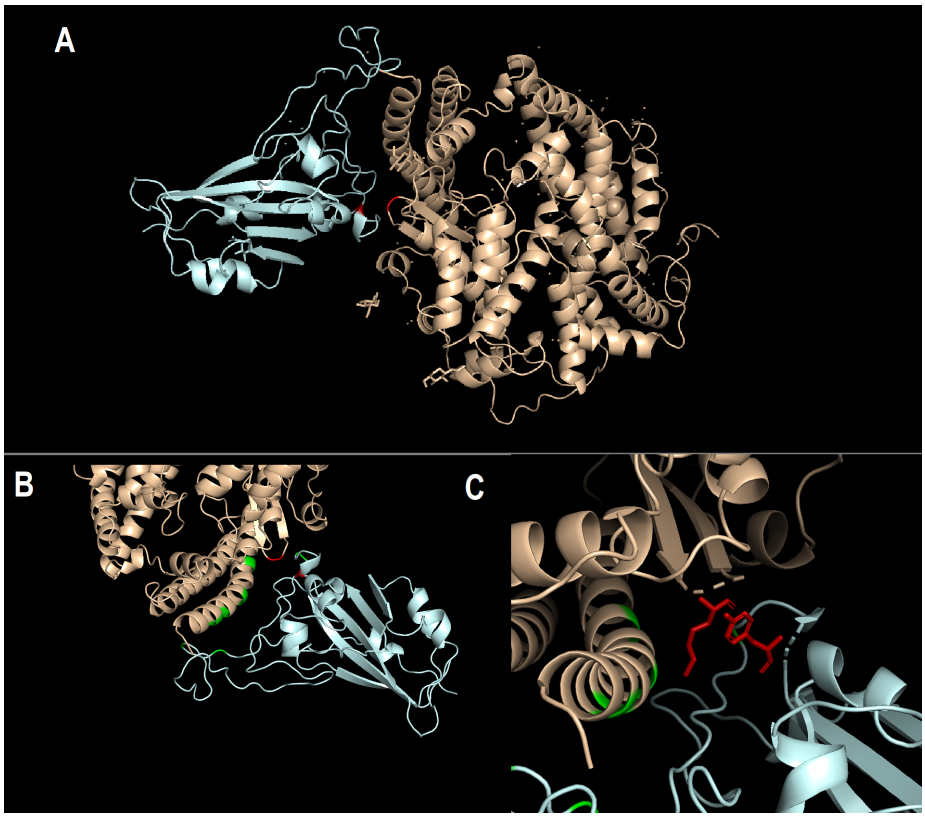

Figure 1: The SARS-CoV-2 spike protein (orange) and ACE-2 (cyan) complex with the most important residues for the interactions selected.
A - LYS353 on the chain A and TYR505 on the chain E have the highest impact in the interaction (red);
B - top-10 important residues for the interaction (A/LYS353 and E/TYR505 in red, green for the others, see Results);
C - zoomed interaction between A/LYS353 and E/TYR505;

4 Discussion

4.1 Application

As described above, we have succeeded in identifying two strong candidates of residues on which the interaction of the Spike-ACE2 complex could rely. Hence, we could argue that if these two residues are in fact crucial for the interaction, disrupting either one of them should have important consequences on the complex formation or function. Even though this is just limited evidence, it can be useful in the context of Spike inhibition, as this is the crucial element

of preventing spread of the virus in the human body.

4.2 Pharmacophore

With the information obtained from this investigation, preliminary information can be inferred on the possible pharmacophore against the Spike. Given that the Spike relies on hydroxylic residues for complex formation, the inhibitor would need to be of sufficient size (to provide coverage and maximize affinity to the long binding region in the protein) and have high hydrogen bond formation capacity (to bind strongly to the key residues). Additionally, as aromatic side groups in the form of phenylalanine are present in the binding site, the inhibitor could have a phenyl group in a similar orientation to create stacking interactions.

4.3 Further research

Using the material that was developed during this practical, we could proceed to analyse the interaction energies for Spike and other human receptors, in particular those similar to ACE2. The analysis should still focus on residues located in close proximity to the receptor's usual active site, and the part of Spike with the highest affinity to it. It could hence be determined whether the two particular residues that were identified in this exercise continue being relevant in other interactions. If so, it could further be argued that the residues indeed are of high importance in the formation of complexes between Spike and human proteins.

Such information could be useful in the context of drug development. For example, identifying residues that are important for the formation of a receptor complexes could be useful for finding efficient inhibition strategies to prevent the formation of Spike complexes – and thus, infections leading to COVID-19. Moreover, knowing whether the key residues conserve their importance when Spike binds to other human proteins (if there are such) could also provide insight into the number of different complexes whose formation would be prevented by blocking (that is, it could give us some information on the specificity of a potential inhibition strategy).

References

[1] P. J. Cock et al. "Biopython: Freely available python tools for Computational Molecular Biology and Bioinformatics". In: *Bioinformatics* 25.11 (2009), pp. 1422–1423. DOI: 10.1093/bioinformatics/btp163.

[2] Josep Lluís Gelpí and Pau Andrio. *Bioexcel/biobb_structure_checking*. Sept. 2022. URL: https://github.com/bioexcel/biobb_structure_checking.

[3] Yuan Huang et al. "Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19". In: *Acta Pharmacologica Sinica* 41.9 (Aug. 2020), pp. 1141–1149. DOI: 10.1038/s41401-020-0485-4. URL: https://doi.org/10.1038/s41401-020-0485-4.

[4] Mushtaq Hussain et al. "Structural variations in human ACE2 may influence its binding with SARS-CoV-2 spike protein". In: *Journal of Medical Virology* 92.9 (Apr. 2020), pp. 1580–1586. DOI: 10.1002/jmv.25832. URL: https://doi.org/10.1002/jmv.25832.

[5] Cody B. Jackson et al. "Mechanisms of SARS-CoV-2 entry into cells". In: *Nature Reviews Molecular Cell Biology* 23.1 (Oct. 2021), pp. 3–20. DOI: 10.1038/s41580-021-00418-x. URL: https://doi.org/10.1038/s41580-021-00418-x.

[6] Tomer Meirson, David Bomze, and Gal Markel. "Structural basis of SARS-CoV-2 spike protein induced by ACE2". In: *Bioinformatics* 37.7 (Aug. 2020). Ed. by Arne Elofsson, pp. 929–936. DOI: 10.1093/bioinformatics/btaa744. URL: https://doi.org/10.1093/bioinformatics/btaa744.

[7] Schrödinger, LLC. "The PyMOL Molecular Graphics System, Version 1.8". Nov. 2015.

[8] Arzu Uyar and Alex Dickson. "Perturbation of ACE2 Structural Ensembles by SARS-CoV-2 Spike Protein Binding". In: *Journal of Chemical Theory and Computation* 17.9 (Aug. 2021), pp. 5896–5906. DOI: 10.1021/acs.jctc.1c00325. URL: https://doi.org/10.1021/acs.jctc.1c00325.

[9] Renhong Yan et al. "Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2". In: *Science* 367.6485 (Mar. 2020), pp. 1444–1448. DOI: 10.1126/science.abb2762. URL: https://doi.org/10.1126/science.abb2762.

4.4 Supplementary

Table 1: Residues on the 6Å distance in the interaction site of ACE2 protein and a SARS-CoV2 spike protein with the highest impact to the interaction

height**Index**	**Res_name**	**Chain**	**Energy**	**Res_pos**
0	SER	A	1.1655	19
1	GLU	A	1.4012	35
2	ARG	E	1.5712	403
3	TYR	E	1.6531	473
4	GLY	E	1.7546	476
5	GLY	E	1.8192	446
6	LEU	A	2.0759	45
7	LYS	E	2.6448	417
8	ARG	A	2.6662	357
9	GLU	A	2.8824	37
10	GLY	E	2.9874	496
11	ASN	A	3.2671	330
12	LEU	A	3.5324	79
13	ALA	E	4.5867	475
14	TYR	E	4.7823	453
15	MET	A	5.0191	82
16	PHE	A	5.0934	28
17	GLY	A	5.1975	354
18	GLN	E	5.5384	493
19	ASP	A	5.8212	355
20	ASP	A	5.9824	38
21	ASP	A	6.4902	30
22	GLY	E	6.6234	502
23	TYR	E	6.8364	449
24	GLN	A	7.2045	42
25	LEU	E	7.8129	455
26	TYR	A	8.4827	83
27	GLN	A	9.2644	24
28	ASN	E	9.6545	487
29	ASN	E	9.6969	501
30	GLN	E	9.7374	498
31	PHE	E	9.8606	456
32	THR	A	10.0728	27
33	HIE	A	10.8366	34
34	LYS	A	11.5697	31
35	TYR	E	12.311	489
36	TYR	A	12.436	41
37	PHE	E	13.422	486
38	THR	E	14.0061	500
39	TYR	E	16.6163	505
40	LYS	A	21.6895	353