Binding Affinities of Coronaviruses on the Human Receptor ACE2

Bowen Bao

Executive Summary

The SARS-CoV-2 is a new kind of pathogen that caused the 2019 coronavirus pandemic. 2019-nCoV infects the host cells through the interaction between the spike surface glycoprotein (S protein) and the human ACE2 receptor. This paper analyzes the binding energy between ACE2 and (1) the bat coronavirus, (2) the pangolin coronavirus, (3) SARS CoV (2002), (4) SARS CoV-2 (2019). I conducted a sequence alignment of the four coronaviruses to extrapolate the binding sites of the bat and pangolin CoV using the binding sites of SARS CoV 1 and 2. I then find the protein structures of the human receptor ACE2, the four coronaviruses, and their interaction complexes. Since the protein structures of the Pangolin Spike Protein, the Pangolin Spike Protein and ACE2 Interaction, and Bat Spike Protein and ACE2 Interaction are not available, I create those homology models using Swiss-Model. I find the binding energies of the four coronaviruses to the human receptor using PyRosetta. The binding energies are: -8508 (bat coronavirus), -9085 (pangolin coronavirus), -11012 (SARS CoV), -47928 (SARS CoV-2). SARS CoV-2 has the highest binding affinity (lowest/ most stable binding energy) followed by SARS CoV, Pangolin CoV, and Bat CoV. This is consistent with the efficacy of the 2019 coronavirus in that it has the highest binding affinity and the greatest infectivity.

Introduction

2019 coronavirus, known as SARS-CoV-2 is a new kind of pathogen that caused the coronavirus pandemic. The outbreak first identified in Wuhan, China, has spread across the world and was declared as a global pandemic by the World Health Organization (WHO) as of March 11, 2020. The virus was originally believed to come from bats (*Rhinolophus affinis*) but new literature has emerged suggesting that the Malayan pangolin (*Manis javanica*) served as an intermediate host before bringing the bat coronavirus to human hosts.

2019-nCoV infects the host cells through the interaction between the spike surface glycoprotein (S protein) and the human ACE2 receptor. The spike protein binds to the human cell receptor, enters the human cell and begins infection. It is the main component responsible for neutralizing host immune responses and primarily determines the infectivity of the coronavirus. This paper analyzes the binding energy between ACE2 and (1) the bat coronavirus, (2) the pangolin coronavirus, (3) SARS CoV (2002), (4) SARS CoV-2 (2019).

Literature Review

Literature analyzing the crystal structures and sequence alignments of the receptor binding domain of SARS CoV-2 and SARS-CoV have shown that the binding structures are nearly identical in their interaction with ACE2. Sequence comparison between SARS CoV-2 and SARS-CoV show 70% residue similarity in receptor interaction. The interaction pattern between the SARS-CoV, SARS-CoV-2, Bat-CoV are quite similar.

Ortega et al explain that the stronger binding affinities between the different viruses and the human receptor may be due to the greater number of protein-protein contacts. They attained

binding affinities of SARS-CoV-2 (-15.7 Kcal/mol), SARS-CoV (-14.1 Kcal/mol) and determine higher affinity values might be related to the infection kinetics and for its rapid spread[1].

Ou et al showed that the bat CoV (RaTG13) had a much lower binding affinity (KD=1.17mM; ΔG=-17.4kJ/mol) to ACE2 than the pangolin CoV (KD=1.89μM; ΔG=-33.9kJ/mol). While the pangolin CoV was slightly lower than the SARS-CoV-2 prototype strain (KD=14.7nM; ΔG=-46.5kJ/mol)[2].

Methodology

Sequence Analysis

I conducted a sequence alignment of the bat coronavirus, the pangolin coronavirus, SARS CoV, MERS-CoV, and SARS CoV-2. Since the binding structures are similar, I use the binding sites of SARS CoV-2 and SARS CoV as an anchor to extrapolate the binding sites of the bat coronavirus and the pangolin coronavirus[3].

Protein Structure

I find the protein structures of

1) Human ACE2 (PDB ID: 1R42)

Then I find the comparison spike proteins and their interaction complexes

SARS-CoV-2 (2019)

2) SARS-CoV-2 Spike Protein Open (PDB ID: 6VYB)
3) SARS-CoV-2 Spike Protein and ACE2 Interaction (PDB ID: 6LZG)

SARS-CoV (2002)

4) SARS CoV Spike Protein (PDB ID: 5XLR)
5) SARS CoV Spike Protein and ACE2 Interaction (PDB ID: 3D0G)

Bat Coronavirus

6) Bat Coronavirus HKU9 Spike Protein (PDB ID: 5GYQ)
7) Bat Spike Protein and ACE2 Interaction

Pangolin Coronavirus

8) Pangolin Spike Protein
9) Pangolin Spike Protein and ACE2 Interaction

[1] Ortega, Joseph Thomas et al. “Role of changes in SARS-CoV-2 spike protein in the interaction with the human ACE2 receptor: An in silico analysis.” EXCLI journal vol. 19 410-417. 18 Mar. 2020, doi:10.17179/excli2020-1167

[2] Junxian Ou et al. “Emergence of RBD mutations in circulating SARS-CoV-2 strains enhancing the structural stability and human ACE2 receptor affinity of the spike protein.” bioRxiv 2020.03.15.991844; doi: https://doi.org/10.1101/2020.03.15.991844

[3] Using Fig 2 from the following: Lan, J., Ge, J., Yu, J. et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 581, 215–220 (2020). https://doi.org/10.1038/s41586-020-2180-5 as well as Andersen, K.G., Rambaut, A., Lipkin, W.I. et al. The proximal origin of SARS-CoV-2. Nat Med 26, 450–452 (2020). https://doi.org/10.1038/s41591-020-0820-9

Since the protein structures of the Pangolin Spike Protein, Pangolin Spike Protein and ACE2 Interaction, and Bat Spike Protein and ACE2 Interaction are not available on PDB, I create those homology models using Swiss-Model.

Protein Modeling

Pangolin Spike Protein

I create this homology model by taking the Fasta file of the Pangolin coronavirus (Ascension No: QIQ54048.1) from NCBI[4]. I remove the 'X's (unknown sequences) from the fasta sequence and use SARS-CoV-2 Spike Protein (PDB ID: 6VYB) as the template. The PDB of Model 1 is downloaded.

Pangolin Spike Protein and ACE2 Interaction

I used the fasta file of the Pangolin coronavirus (Ascension No: QIQ54048.1) from NCBI after removing the 'X's and the fasta file of ACE2 (PDB ID: 1R42) from PDB. I use the SARS-CoV-2 Spike Protein and ACE2 Interaction (PDB ID: 6LZG) as the template.

Bat Spike Protein and ACE2 Interaction

I used the fasta sequence of the Bat Coronavirus HKU9 Spike Protein (PDB ID: 5GYQ) and the fasta sequence of ACE2 (PDB ID: 1R42) from PDB. I use the SARS-CoV-2 Spike Protein and ACE2 Interaction (PDB ID: 6LZG) as the template.

Results

Using PyRosetta, I calculate the binding energy using this formula:

$$E_{binding} = E_{complex} - E_{spike\ protein} - E_{ACE2}$$

The results are the following:

	Energy Score			
Virus	**Complex**	**Spike Protein**	**ACE2**	**Binding**
SARS CoV-2 (2019)	3841.36	44248.99	7521.35	-47928.98
SARS CoV (2002)	3909.14	7400.42	7521.35	-11012.64
Pangolin Coronavirus	437.89	2001.86	7521.35	-9085.32
Bat Coronavirus	574.74	1562.01	7521.35	-8508.62

Since higher affinity values are related to the higher infectivity of the virus, we see that the SARS CoV-2 has the highest binding affinity (lowest/ most stable binding energy) followed by SARS Cov, Pangolin CoV, and Bat CoV. These results are consistent with those found by Ou et al.

Note: The binding energies are calculated using the original protein structures and not the cleaned PDB files because the cleaned versions would have removed too many residues and provided inaccurate results.

4 https://www.ncbi.nlm.nih.gov/protein/1824829172

Discussion

Since research on the role of pangolins has only recently emerged, most available literature studies the crystal structures of the pangolin coronavirus and its interaction mechanisms. Very little research has been done studying the binding affinity of the pangolin coronavirus and its relativity to other strands of coronavirus. This paper contributes to the discussion of the role of pangolins in SARS CoV-2 and gives an introductory analysis on relationship between the binding affinities and virus infectivity. Further research can use this method of utilizing PyRosetta to study the binding affinities of the individual interaction residues as opposed to just the protein in whole.

References

Swiss Model

- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46(W1), W296-W303 (2018).
- Guex, N., Peitsch, M.C., Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. Electrophoresis 30, S162-S173 (2009).
- Bienert, S., Waterhouse, A., de Beer, T.A.P., Tauriello, G., Studer, G., Bordoli, L., Schwede, T. The SWISS-MODEL Repository - new features and functionality. Nucleic Acids Res. 45, D313-D319 (2017).
- Studer, G., Rempfer, C., Waterhouse, A.M., Gumienny, G., Haas, J., Schwede, T. QMEANDisCo - distance constraints applied on model quality estimation. Bioinformatics 36, 1765-1771 (2020).
- Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L., Schwede, T. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. Scientific Reports 7 (2017).

Sequence Analysis EScript

- Robert, X. and Gouet, P. (2014) "Deciphering key features in protein structures with the new ENDscript server". *Nucl. Acids Res*. **42**(W1), W320-W324 - doi: 10.1093/nar/gku316 (**freely accessible online**).

Literature Reviews:

- Lan, J., Ge, J., Yu, J. *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581,** 215–220 (2020). https://doi.org/10.1038/s41586-020-2180-5
- Andersen, K.G., Rambaut, A., Lipkin, W.I. et al. The proximal origin of SARS-CoV-2. Nat Med 26, 450–452 (2020). https://doi.org/10.1038/s41591-020-0820-9
- Ortega, Joseph Thomas et al. “Role of changes in SARS-CoV-2 spike protein in the interaction with the human ACE2 receptor: An in silico analysis.” EXCLI journal vol. 19 410-417. 18 Mar. 2020, doi:10.17179/excli2020-1167

- Junxian Ou et al. “Emergence of RBD mutations in circulating SARS-CoV-2 strains enhancing the structural stability and human ACE2 receptor affinity of the spike protein.” bioRxiv 2020.03.15.991844; doi: https://doi.org/10.1101/2020.03.15.991844