

Original article:

**ROLE OF CHANGES IN SARS-COV-2 SPIKE PROTEIN IN THE
INTERACTION WITH THE HUMAN ACE2 RECEPTOR:
AN *IN SILICO* ANALYSIS**

Joseph Thomas Ortega¹, Maria Luisa Serrano², Flor Helene Pujol³, Hector Rafael Rangel^{3*}

1 Department of Pharmacology and Cleveland Center for Membrane and Structural Biology, School of Medicine, Case Western Reserve University, Cleveland, OH 44106, USA

2 Unidad de Química Medicinal, Facultad de Farmacia, Universidad Central de Venezuela, Caracas, Venezuela

3 Laboratorio de Virología Molecular, Centro de Microbiología y Biología Celular, Instituto Venezolano de Investigaciones Científicas, Caracas, Venezuela

* **Corresponding author:** Hector Rafael Rangel, Laboratorio de Virología Molecular, Centro de Microbiología y Biología Celular, Instituto Venezolano de Investigaciones Científicas, Caracas, Venezuela; E-mail: hrangel2006@gmail.com

<http://dx.doi.org/10.17179/excli2020-1167>

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>).

ABSTRACT

Many human viral diseases are a consequence of a zoonotic event. Some of the diseases caused by these zoonotic events have affected millions of people around the world, some of which have resulted in high rates of morbidity/mortality in humans. Changes in the viral proteins that function as ligands of the host receptor may promote the spillover between species. The most recent of these zoonotic events that have caused an ongoing epidemic of high magnitude is the Covid-19 epidemics caused by SARS-CoV-2. The aim of this study was to determine the mutation(s) in the sequence of the spike protein of the SARS-CoV-2 that might be favoring human to human transmission. An *in silico* approach was performed, and changes were detected in the S1 subunit of the receptor-binding domain of spike. The observed changes have significant effect on SARS-CoV-2 spike/ACE2 interaction and produce a reduction in the binding energy, compared to the one of the Bat-CoV to this receptor. The data presented in this study suggest a higher affinity of the SARS-Cov-2 spike protein to the human ACE2 receptor, compared to the one of Bat-CoV spike and ACE2. This could be the cause of the rapid viral spread of SARS-CoV-2 in humans.

Keywords: Spike, SARS-CoV-2, ACE2, Coronavirus, outbreak

INTRODUCTION

Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV), Middle Eastern Respiratory Syndrome Coronavirus (MERS-CoV), and the recently identified novel Coronavirus (SARS-CoV-2) belong to the Coronaviridae family, genus Betacoronavirus, that has been related to important epi-

demiological outbreaks. These are enveloped viruses with a positive-sense single-strand RNA of around 32 Kb. The viral particles contain four main structural proteins: the spike, membrane, envelope protein, and nucleocapsid. The spike protein protrudes from the envelope of the virion and plays a pivotal role in the receptor host selectivity and cellu-

lar attachment. Strong scientific evidence showed that SARS and SARS-CoV-2 spike proteins interact with angiotensin-converting enzyme 2 (ACE2) (Chen et al 2020; Wan et al., 2020). Also, other cellular receptors play a secondary role in the viral attachment, as the C-type lectin CD209L, and DC-SIGN binds to SARS-CoV. However, ACE2 appears to be the key functional receptor for the SARS-CoV (Coutard et al., 2020; Satija and Lal, 2007) and probably for SARS-CoV-2 (Walls et al., 2020) The interaction between the viral protein and its cell membrane receptor is a critical step in the replication cycle. Furthermore, the efficiency of viral infection is strongly dependent on this process. Several physicochemical factors are associated with protein-protein interactions. These factors are determined by the nature of residues and the type of chemical interactions occurring between ligand and receptor. Thus, the presence of residues that produce an energetically favored interaction (lower free energy) may drive binding kinetics and finally lead to the fusion event. Based on that, this study aimed to evaluate the energetic profile of the interaction between the SARS-CoV-2 spike protein and the human cell receptor ACE2.

MATERIALS AND METHODS

Sequence analysis

Sequences analyzed were individually retrieved from GenBank (accession numbers are shown in the phylogenetic tree, see Figure 1). The sequence used for the modeling process and construct the spike protein model of SARS-CoV2 were obtained from the protein database and correspond to 6ACC (SARS spike protein) and 6ACD (Bat-spike protein).

Protein modeling

Homology structural models of viral spike protein from SARS-CoV-2 (QHO62112.1) and Bat-CoV (AAZ67052.1) were built by using the tools of the SWISS-MODEL modeling server and the Deep-

View/Swiss-PdbViewer 4.01 software (Arnold et al., 2006). Several models were obtained and the quality of each structure was evaluated. The best model for SARS-CoV-2 spike was obtained using the crystal structure of SARS-CoV spike protein (PDB code 6ACC). On the other hand, for Bat-CoV, the best model was obtained using as template the crystal structure PDB code 6ACD. These models were subjected to further protein structure optimization. Hydrogen atoms were added and the partial charges were assigned for energy refinement. The protein model was embedded in a 100 Å water box. Then, energy minimization was performed while applying constraints to the protein backbone to preserve global folding and optimizing the relative position of the water molecules and protein. The obtained systems underwent MD simulations using NAMD as described by Ortega et al. (2019). All MD simulations described in this study were performed with NAMD 2.12 (Phillips et al., 2005), Vega ZZ 3.1.0.21 (Pedretti et al., 2004). CHARMM force field (Vanommeslaeghe et al., 2010) and Gasteiger charges were used. The obtained structures represent the lowest energy frame of the MD simulations. The quality of the models was established with ProSA (Wiederstein and Sippl, 2007) and PRO-CHECK programs (Laskowski et al., 1993).

Protein-protein docking

Crystal structure for SARS-CoV spike (PDB code 6ACK) and ACE2 (PDB code 1R42) were downloaded from the Protein Data Bank. Also, the homology model for SARS-CoV-2 spike was assayed. Protein preparation was carried out as described above. Then, binding patterns and affinity estimations for the interaction between the viral spike and ACE2 receptor were performed using molecular docking. This process was performed through two steps; first, a blind docking between ligand (spike protein) and receptor (ACE2) was performed using Z-dock software (Pierce et al., 2014). Then, the resulting docking data were processed and analyzed by using the tools of

PRODIGY software (Xue et al., 2016). Finally, results were clustered and analyzed considering binding energies and main interacting residues in each complex.

RESULTS

Homology analysis of the spike proteins of SARS-CoVs and related Bat-CoVs

Phylogenetic analysis of the spike protein sequences of SARS-CoV-2 and Bat-CoVs, SARS-CoV is shown in Figure 1. The results are in agreement with recent reports of an independent introduction of SARS-CoV-2 from a Bat-CoV, different from the spillover which led to the introduction of SARS-CoV, being the Bat-CoV of *Rhinolophus affinis* the probable ancestor of this new virus (Wong et al., 2020). Indeed, the sequences of the whole spike of this Bat-CoV and of SARS-CoV-2 share 97.7% identity (Figure 1). More divergence is found however in the S1 subunit, particularly in the Receptor Binding Domain (RBD) of the different spike proteins. SARS-CoV and Bat-CoV from *Rhinolophus sinicus* (originally signaled as the most closely related virus to SARS-CoV-2)

exhibit several amino acid substitutions and deletions in the RBD compared to SARS-CoV-2. The RBD of Bat-CoV from *Rhinolophus affinis*, although more closely related to the one of SARS-CoV-2, also displayed several amino acid substitutions (Figure 2).

Structural analysis of Spike-ACE2 complexes

The crystal structures of the spike protein of SARS-CoV and homology models of Bat-CoV (accession number MG772933), Bat-CoV of *Rhinolophus sinicus*, and SARS-CoV-2 interacting with the putative binding domain site in human ACE2 were analyzed. The interaction pattern between the three viral spikes is quite similar. The main region of interaction with the putative cellular receptor counter-part is formed by fifteen residues ordered into a beta-sheet conformation surrounded by two capping loops (Figure 3 and Supplementary Figure 1). Interestingly, sequence comparison between SARS-CoV-2 and SARS-CoV revealed that the residues pre-

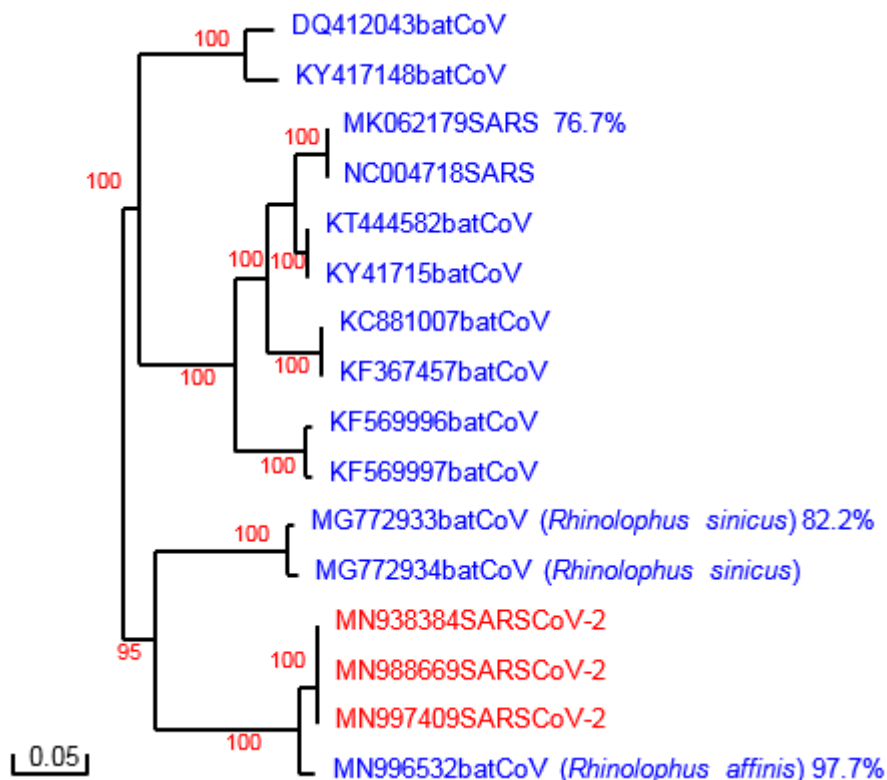


Figure 1: Phylogenetic analysis of SARS-CoV-2 and other coronavirus spike proteins. Phylogenetic tree constructed with Poisson correction and 100 bootstrap replicas. The sequences are named with their accession number. Percent homology with SARS-CoV-2 spike protein is shown for some proteins.

```

MN938384SARSCoV2 318RVQPTESIVRFPNITNLCPFGVEFNATRFASVYAWNRRKRISNCVADYSVLVNSASFSTFKCYGVSPTKLNDL 390
MN996532batCoV 318-----d-----t-----t----- 390
MG772933batCoV 318----q-v-----v--hk-----p----e-tk--d-i---t-f--t-----s--i-- 390
NC004718SARS 318--v-sgdv-----k-p-----e-k-----tf-----a----- 390

MN938384SARSCoV2 391CFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNVNYLYRLEFRKSNLK 463
MN996532batCoV 391-----t-----khi-a-e--f-----a--- 463
MG772933batCoV 391--s---t-l--fs---v-----v-----takq-vg.....-f--sh-stk-- 463
NC004718SARS 391-s-----vk--d-----v-----m--l---tr-i-atst---k--yl-hgk-r 463

MN938384SARSCoV2 464PFFERDISTEIVQAGSTPCNGVEGFNCYPLQYSGFQPTNCGVYQPYRVVVLSFELLHAPATVCGPKK 531
MN996532batCoV 464-----k---qt-l---y--yr---y-d--h-----n----- 531
MG772933batCoV 464----l-sde.....-r.....t-st-d-n-nvple--at-----n-----l 512
NC004718SARS 484-----nvpfspdgk--tppa.l---w--nd---yt-t-i-----n-----l 530
    
```

Figure 2: Receptor Binding Domain of the spike protein sequence alignment of SARS-CoV-2 and other related Coronaviruses. Sequence alignment for the interacting domain of SARS-CoV-2 (MN938384), Bat-CoV (MN996532 and MG772933) and SARS-CoV (NC004718). The key amino acids described for the interaction with ACE2 are shown in red, and in blue others amino acid related with the interaction in SARS-CoV2. (Lines (-) = same amino acid, dots (.) =deletion)

sent in the receptor-interacting motive are highly conserved with 70 % identity, sharing nine residues between both viruses. In the SARS-CoV RBD are present residues that allowed the interspecies infection, known as Y442, L472, N479, D480, and T487 (Lu et al., 2015). However, in SARS-CoV-2, slight modification of some residues could improve the interaction with the human cellular receptor: L455, F486, Q493, and N501. In SARS-CoV, two main residues (479 and 487) have been associated to the recognition of the human ACE2 receptor (Lu et al., 2015). These residues suffered a punctual mutation from civet to human, K479N and S487T (Li, 2013). In the SARS-CoV-2, the residues corresponding to N479 correspond to Q493 and T487 to N501. These changes in the SARS-CoV-2 represent energetically favorable changes for the interaction with the receptor. The local environment present in the ACE2 receptor allows these mutations to produce a significant number of electrostatic stabilizing interactions (Table 1). Furthermore, as mentioned previously, the presence of the two capping loops in the binding do-

main is likely producing a stabilization effect over the interaction with the cellular receptor. Our models showed that these capping loops appear in both human-infecting viruses but are absent in the bat virus. The data showed here strongly suggest that these capping loops produce an increase in the electrostatic interactions between the spike protein and the cellular receptor. In SARS-CoV, the residues present in these capping loops showing direct interaction with the receptor are R426, S432, T433, Y436, P462, D463, S472, and N473 and in SARS-CoV-2 are V445, Y449, Y473, Q474, A475, E484, G485, F486, and N487. The counter-pairs located in the ACE2 receptor are shown in Table 1. Altogether, the higher number of protein-protein contacts (Table 2) and the longer capping loops could explain the increase in binding affinities in SARS-CoV-2 (-15.7 Kcal/mol) in comparison with SARS-CoV (-14.1 Kcal/mol) (Table 3). Thus, these loops could play an important role together with the punctual mutations being an interesting clue to determine the host receptor specificity for the viral spike protein.

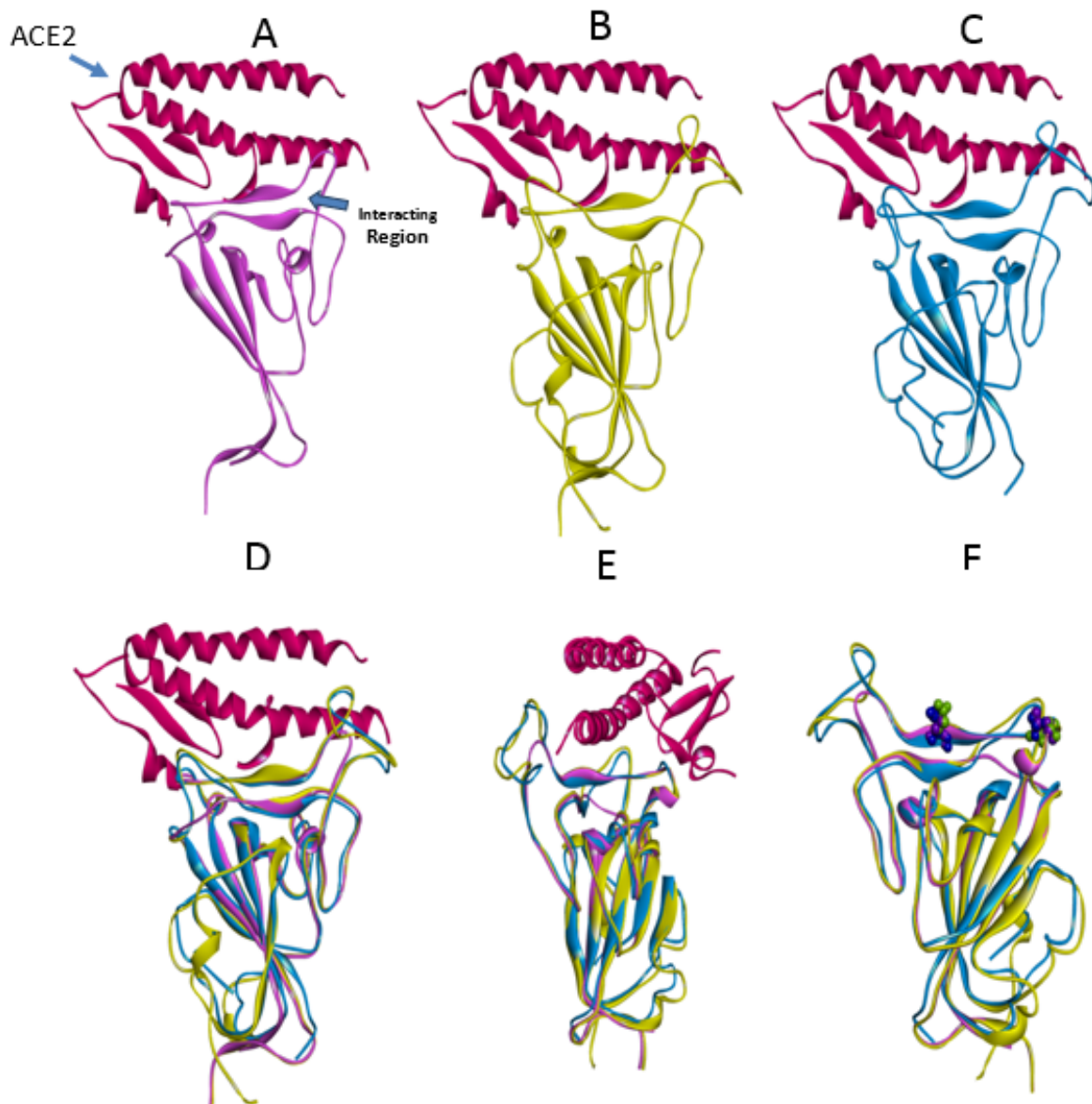


Figure 3: Coronavirus spike proteins. The spike proteins in complex with the RBD of ACE2 (dark pink) are shown **A)** Bat-CoV, **B)** SARS-CoV, and **C)** SARS-CoV-2. A comparison between the three spike proteins are shown in **D)** and a 45 degree turn is also shown in **E)**. The location of the main residues mutated in SARS-CoV (position 479 and 487) and SARS-CoV-2 are shown in **F)** (green and blue).

DISCUSSION

Several amino acid substitutions in RBD were identified in the SARS-CoV-2 RBD compared to Bat-SARS-CoVs and SARS-CoV. Mutations in the spike protein could change the tropism of a virus, including new hosts or increasing viral pathogenesis (Shang et al., 2020). HIV infection is a good model of change in viral tropism by mutations in the envelope proteins. These mutations

switch co-receptor use (CCR5 to CXCR4) increasing the viral pathogenesis (Mosier 2009). Interestingly, our data showed that these changes are not only related to the ability of interaction with the human ACE2 receptor but also for improving this receptor recognition. The presence of two loops around the RBD of SARS-CoV-2 might be promoting the interaction with the ACE2 receptor, improving the binding to this recep-

tor by increasing the number of atoms involved (Tables 1 and 2). The amino acid substitutions and the longer capping loops could explain the increase in binding affini-

ties in SARS-CoV-2 compared to SARS-CoV (Table 1). Higher affinity values might be related to the dynamic of infection and the rapid spread observed for this virus.

Table 1: Residues involved in the Interaction between viral spike and ACE2 (SARS)

SARS-CoV spike	ACE2	SARS-CoV-2 spike	ACE2
R426	T27, A25	R403	I21,E23,Q24
S432	E329	G404	Q24
T433	L79	D405	Q24, A25, K26, T27
Y436	T324, Y83, Y41	E406	T27
Y440	S19	R408	F28
TY442	K353, Y41	Q409	P28
L443	Q24	K417	L29, D30, K31
P462	E329, E35	I418	K31
D463	D355	Y421	K31
L472	Q24, Q325, Q42	V445	k31, F32
N473	R357, N330, D355	Y449	N33
Y475	T27, T324, Y41	Y453	N33, H34
N479	R393, N330	L455	H34, E35
K353	Y481	F456	E35, A36, E37
G482	D38, Q42	Y473	D38
F483	E329	Q474	D38
Y484	k353, F28, T27, Y41	A475	D38
T485	G352, G354	E484	D38
T486	G354, L45, K353, K31	G485	L39
T487	G354, G326, H34, L45, K353	F486	L39, F40,Y41
G488	D38, Q325, D355	N487	Y41, Q42, S43, S44
I489	Q42, Q325	C488	L45
Y491	S19, M89, Y83	Y489	L45, A46, N49, E56, E57
D38	Q492	F490	N58, Q60
		L492	Q60
		Q493	Q60, N61
		S494	N61, N64, A65
		Y495	A65, K68
		G496	K68
		Q496	K68, F72, E75, Q76
		P499	Q76, T78
		T500	L79, A80
		N501	A80, Q81, M82, Y83
		G502	Y83, L100
		V503	E329
		G504	N330, L351, G352
		Y505	k353, G354, D355, R357

Table 2: Number of protein-protein contacts (PPC) between CoV spikes and ACE2

Number of Contacts (PPCs)	Bat-CoV	SARS-CoV	SARS-CoV-2
PPCs charged-charged	4	5	13
PPCs charged-polar	11	14	13
PPCs charged-apolar	19	28	38
PPCs polar-polar	5	6	14
PPCs polar-apolar	19	32	33
PPCs apolar-apolar	23	32	32
Total PPCs	81	117	143

Table 3: Binding affinity (ΔG) and dissociation constant (Kd) predicted values for the interaction between viral spike and ACE2 receptor

Protein-protein complex (Viral spike/ACE2)	ΔG (kcal/mol)	Kd (M) at 25.0 °C
Bat-CoV	-8.2	1.5×10^{-6}
SARS-CoV	-14.1	1.1×10^{-10}
SARS-CoV-2	-15.7	5×10^{-12}

The origin of the SARS-CoV-2 has been not fully elucidated. While this study was in course, another study of Wong et al. (2020), showed a high similarity at protein level in the RBD among the coronaviruses isolated from the recent outbreak (SARS-CoV-2), those isolated from pangolin and *Rhinolophus affinis* (RaTG13). The authors also suggest that Pangolin might be the intermediate host, with a 98 % identity with the human virus, at the receptor binding motif, between the bat and human. The spike model of RaTG13 is quite similar to that obtained from SARS-CoV and SARS-CoV-2 and the loops in the RBD are also present (data not shown). The protein sequence of the receptor binding motif, has 5 important amino acids. When comparing the sequence of SARS-CoV-2 with that of the isolated viruses of pangolin and *Rhinolophus affinis*, 1 and 4 differences are observed respectively in the amino acids considered key for the union with ACE2 (Yan et al., 2020; Wong et al., 2020). These differences should mean slightly less favorable binding energies between

these viruses with ACE2 compared to the SARS-CoV-2, shown in this study. Thus, the loops observed in the spike protein of SARS-CoV-2 could play an important role together with the amino acid substitutions, being an interesting clue to determine the host receptor specificity for the viral spike protein. Altogether, structural changes and residues composition in the viral spike protein could be associated with increased infection kinetics and viral spreading. Comparative studies to determine the impact *in vitro* of the mutation and loops in RBD of SARS-CoV and SARS-CoV-2 are required in order to predict possible zoonotic event in the future.

REFERENCES

- Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. *Bioinformatics*. 2006;22:195-201.
- Chen Y, Guo Y, Pan Y, Zhao ZJ. Structure analysis of the receptor binding of 2019-nCoV. *Biochem Biophys Res Commun*. 2020;17:S0006-291X(20)30339-9.

- Coutard B, Valle C, de Lamballerie X, Canard B, Seidah NG, Decroly E. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res.* 2020;10(176):104742.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK - a program to check the stereochemical quality of protein structures. *J Appl Crystallogr.* 1993;26:283–91.
- Li F. Receptor recognition and cross-species infections of SARS coronavirus. *Antiviral Res.* 2013;100:246-54.
- Lu G, Wang Q, Gao GF. Bat-to-human: spike features determining 'host jump' of coronaviruses SARS-CoV, MERS-CoV, and beyond. *Trends Microbiol.* 2015;23:468-78.
- Mosier DE. How HIV changes its tropism: evolution and adaptation? *Curr Opin HIV AIDS.* 2009;4:125-30.
- Ortega JT, Serrano ML, Suárez AI, Baptista J, Pujol FH, Cavallaro LV, et al. Antiviral activity of flavonoids present in aerial parts of *Marcetia taxifolia* against Hepatitis B virus, Poliovirus, and Herpes simplex virus in vitro. *EXCLI J.* 2019;18:1037-48.
- Pedretti A, Villa L, Vistoli G. VEGA - An open platform to develop chemo-bio-informatics applications, using plug-in architecture and script programming. *J Comput Aided Mol Des.* 2004;18:167–73.
- Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. *J Comput Chem.* 2005;26:1781–802.
- Pierce BG, Wiehe K, Hwang H, Kim B-H, Vreven T, Weng Z. ZDOCK server: interactive docking prediction of protein–protein complexes and symmetric multimers. *Bioinformatics.* 2014;30:1771–3.
- Satija N, Lal SK. The molecular biology of SARS coronavirus. *Ann NY Acad Sci.* 2007;1102:26-38.
- Shang J, Wan Y, Liu C, Yount B, Gully K, Yang Y, et al. Structure of mouse coronavirus spike protein complexed with receptor reveals mechanism for viral entry. *PLoS Pathog.* 2020;16(3):e1008392.
- Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem.* 2010;31:671–90.
- Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell.* 2020;6:S0092-8674(20)30262-2.
- Wan Y, Shang J, Graham R, Baric RS, Li L. Receptor recognition by novel coronavirus from Wuhan: An analysis based on decade-long structural studies of SARS. *J Virol.* 2020; epub ahead of print.
- Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 2007; 35:W407–10.
- Wong MC, Javornik Creegen SJ, Ajami NJ, Petrosino JF. Evidence of recombination in coronaviruses implicating pangolin origins of nCoV-2019. *bioRxiv.* 2020; <https://doi.org/10.1101/2020.02.07.939207>
- Xue LC, Rodrigues JP, Kastriitis PL, Bonvin AM, Vangone A. PRODIGY: a web server for predicting the binding affinity of protein-protein complexes. *Bioinformatics.* 2016;32:3676-8.
- Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE2. *Science.* 2020; epub ahead of print.