



Mapping of ACE2 binding site on SARS-CoV-2 spike protein S1: docking study with peptides

Aleksei Kuznetsov* and Jaak Järvi

Institute of Chemistry, University of Tartu, Ravila 14a, 50411 Tartu, Estonia

Received 18 May 2020, accepted 2 June 2020, available online 18 June 2020

© 2020 Authors. This is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>).

Abstract. Among different approaches to control the COVID-19 disease, there is clear interest to develop inhibitors which block the virus interaction with the host cells and through this simple mechanism could facilitate developing medication. In this report, interaction of the virus SARS CoV-2 spike protein S1 binding site with potential antiviral peptide ligands is analysed computationally. The peptides are derived from the binding domain of the angiotensin-converting enzyme 2, which is the receptor site for this virus. These calculations reveal that although shortening of these peptides from the N terminus and C terminus reduces their docking energy on the S1 binding site, there is still a number of peptides which effectively bind to the binding site on the SARS CoV-2 spike protein S1, and thus can be used as leads for further optimization of the inhibitory effect. Finally, this may open new perspectives for working out treatments against the virus infection.

Key words: SARS-CoV-2 virus, spike protein S1, binding site mapping, peptide docking, COVID-19 treatments.

1. INTRODUCTION

Infection with SARS-CoV-2 is initiated by entry of the virus RNA into the host cells, and this process is based on interplay between the spike protein on the surface of the viral particle and the angiotensin-converting enzyme 2 (ACE2), which serves as specific cell surface receptor for this virus [1]. Following this interaction, the virus fuses its envelope with the host cell membrane to deliver the nucleocapsid into the target cell and translation of the viral RNA starts [2–4]. This virus entry process involves multiple protein-protein interaction steps, proteolytic modification of the virus proteins, and multiple conformational transitions [3–5]. Therefore, it can be suggested that blockade of some of these interactions may open efficient strategy for reduction of the level of the virus attack and developing therapeutic drugs for COVID-19 disease [6]. Among various possibilities, the goal-directed design of peptides or peptidomimetics as well

as small proteins, which inhibit the spike protein interaction with ACE2, seems to be of particular interest for this task [6–8]. The feasibility of this approach seems to be increased by the fact that the host cell receptor ACE2 possesses higher affinity against S1 binding site in the case of SARS-CoV-2, compared with spike proteins on other previously studied CoV species which use similar cell entry pathway [3]. The structure of participating proteins S1, ACE2, and their complex have been described recently in several papers and are listed in PDB database that makes the testing of potential peptide-based drugs using computational methods, workable. The complex of the spike protein S1 with its receptor (ACE2) can be calculated by molecular dynamics (MD) simulations and is illustrated in Fig. 1.

The possibility of spike protein S1 blockade by fragments of ACE2 has been studied before [7]. This analysis involved four inhibitory peptides derived from structure of ACE2, and interaction of these compounds with the receptor binding domain (RBD) on S1 was analysed using molecular dynamics simulations. For

* Corresponding author, aleksei.kuznetsov@ut.ee

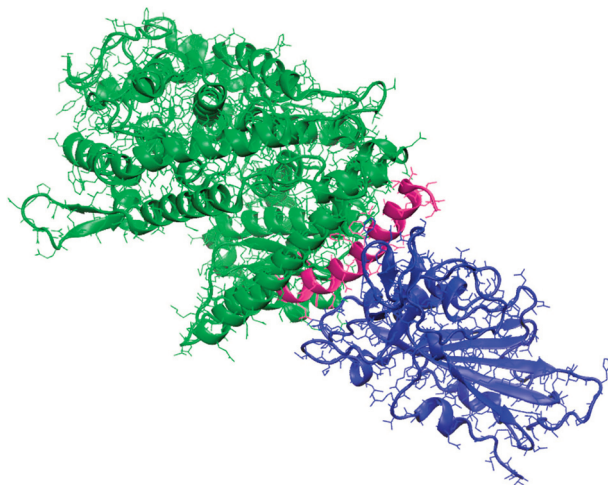


Fig. 1. Cartoon representation of the structure of the SARS-CoV-2 spike protein S1 complex with ACE2, modelled computationally by using data listed in the PDB database (www.pdb.org) as “6LZG”. The spike protein is coloured blue, the ACE2 main body is coloured green and the binding domain $\alpha 1$ is coloured magenta.

design of these inhibitors, the presence of 15 interactions between amino acids at the ACE2 binding domain and the S1 binding site were taken into consideration [7]. Ten of these amino acids belong to the $\alpha 1$ domain of ACE2, one interacting residue comes from the $\alpha 2$ domain and 4 amino acids belong to the linker region between the $\beta 3$ and $\beta 4$ domains of ACE2. Following this information, four inhibitors were designed in [7], where the shortest peptide was the $\alpha 1$ domain of ACE2, the next included both $\alpha 1$ and $\alpha 2$ domains, and two other compounds were rather large fragments of the protein sequence. Beside these peptides, binding of the whole ACE2 (amino acids 19–615) was studied in comparison [7]. The results of these calculations revealed that the binding energies of the smallest protein fragment, the $\alpha 1$ domain, and the whole protein were rather similar, while binding other peptides containing both $\alpha 1$ and $\alpha 2$ domains, was not so good [7]. Therefore, it may be concluded that the $\alpha 1$ domain of ACE2, consisting of amino acids 21–55, covers alone the main part of the binding effectiveness of this protein with the S1.

Based on this information, we decided to extend this study and investigate binding effectiveness of truncated analogs of the $\alpha 1$ domain with its binding site on the S1 protein using computational docking analysis. The computer-simulated complex, formed between the peptide derived from the $\alpha 1$ domain of ACE2 (amino acids 19–45) and the spike protein S1 was calculated by means of molecular dynamics and is shown in Fig. 2.

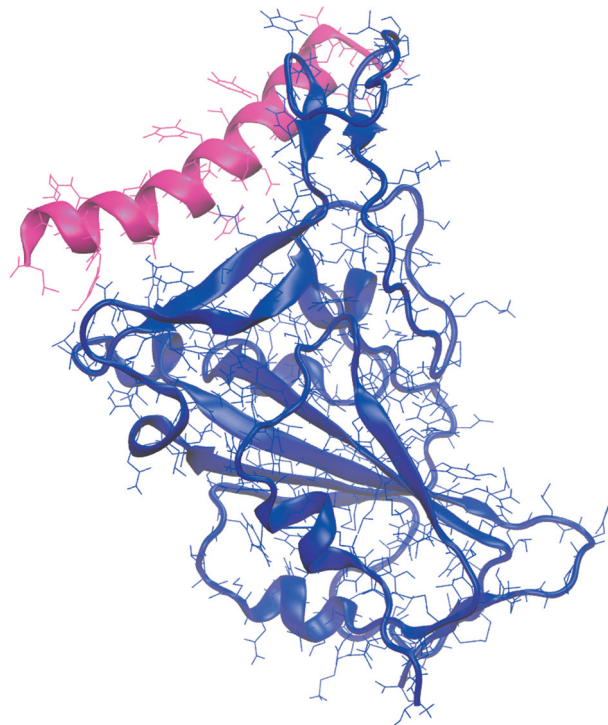


Fig. 2. Cartoon representation of the structure of the SARS-CoV-2 spike protein S1 with bound peptide STIEEQAKTFLDKFNHEAEDLFYQSSL, corresponding to the sequence 19–45 of the receptor protein ACE2. The complex structure was built computationally using molecular dynamics method and structural data listed in the PDB database (www.pdb.org) as “6LZG”. The spike protein is coloured blue and the peptide 19–45 is coloured magenta.

2. METHODS

The input files used for modelling an ACE2 and the receptor binding domain of Cov-2 spike protein S1 (amino acids from 333 to 527) as well as the complex formed between these proteins were built starting from data about the spatial structure of these proteins, obtained by X-ray structure analysis [3,5], and listed as “6LZG” in the PDB database (www.pdb.org).

The peptides used for docking study were derived from the $\alpha 1$ domain of ACE2 protein and the lead sequence STIEEQAKTFLDKFNHEAEDLFYQSSL was systematically truncated from both ends to produce 200 different peptides. The final list of studied peptides is shown in Table 1.

The conventional software for ligand docking AutoDock Vina version 1.1.2 [9] and MD simulation GROMACS version 4.6.1 package [10] were used. The best scoring results of peptide positioning were picked for peptide-S1

Table 1. Docking energy (kcal/mole) of peptides with ACE2 binding site on S1 spike protein of SARS-CoV-2. The peptides were derived from the sequence STIEEQAKTFLDKFNHEAEDLFYQSSL (amino acids 19–45 of the ACE2 binding domain α 1), through gradual shortening of the peptide from both ends. The first and the last amino acid of each peptide and its number in the sequence are listed

	S	T	I	E	E	Q	A	K	T	F	L	D	K	F	N	H	E	A	E	D	
	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	
K	-8.1																				
T	-8.1	-8.1																			
F	-8.8	-8.7	-8.6																		
L	-9.1	-9.0	-8.9	-8.6																	
D	-9.1	-9.1	-8.9	-8.8	-8.4																
K	-9.5	-9.9	-9.7	-9.7	-9.2	-9.1															
F	-9.1	-10.0	-9.8	-9.1	-9.0	-8.7	-8.7														
N	-9.1	-10.0	-9.8	-9.1	-9.0	-8.5	-7.8	-8.1													
H	-9.7	-10.1	-10.5	-9.7	-9.5	-9.1	-8.0	-8.3	-7.7												
E	-0.1	-9.8	-9.7	-9.7	-9.7	-9.0	-8.3	-8.0	-7.8	-8.0											
A	-10.4	-10.2	-9.7	-10.1	-9.7	-9.2	-8.4	-8.3	-8.1	-8.1	-7.8										
E	-10.4	-10.1	-11.0	-10.2	-9.9	-9.8	-9.3	-9.0	-9.0	-9.1	-8.6	-7.9									
D	-10.6	-10.6	-11.1	-10.6	-10.6	-10.2	-9.9	-9.7	-9.6	-9.6	-9.1	-8.9	-7.9								
L	-10.5	-10.7	-11.2	-10.4	-10.2	-10.1	-8.9	-8.8	-8.8	-9.4	-8.7	-8.9	-8.4	-8.1							
F	-11.0	-10.9	-11.2	-10.5	-10.3	-10.2	-10.2	-10.2	-10.2	-10.2	-9.5	-9.1	-9.1	-8.7	-8.2						
Y	-12.1	-11.7	-11.7	-11.7	-11.7	-11.5	-10.3	-10.3	-10.3	-10.3	-9.6	-9.4	-9.4	-9.4	-8.8	-8.4					
Q	-12.2	-11.8	-11.8	-11.7	-11.7	-11.6	-10.4	-10.3	-10.3	-10.3	-10.0	-9.3	-9.2	-8.9	-9.3	-9.0	-9.0				
S	-12.2	-11.8	-11.8	-11.7	-11.7	-11.6	-10.6	-10.6	-10.6	-10.6	-9.9	-9.9	-9.6	-9.3	-8.9	-8.4	-8.8	-8.0			
S	-12.2	-11.8	-11.8	-11.7	-11.7	-11.6	-10.7	-10.6	-10.7	-10.7	-10.0	-10.2	-9.8	-9.6	-9.1	-8.5	-8.9	-8.2	-7.7		
L	-12.6	-12.2	-12.2	-12.1	-12.1	-12.0	-10.8	-10.8	-10.9	-11.4	-10.3	-10.3	-9.8	-9.6	-9.2	-8.6	-8.9	-8.4	-8.3	-8.8	

complex. The docking energy values were calculated and listed in Table 1. The protonation of the protein was handled automatically by the GROMACS `pdb2gmx` tool, which uses the most common protonation state for amino-acid residue at pH 7. The constrained bond lengths and angles, torsion angles, atomic point charges, and van der Waals parameters for the molecules were assigned using the GROMOS 53A6 force field parameter set [11]. The complex structure was solvated in a cubic box with side length of 5 nm and filled by the SPC water model for solvent [12]. System was neutralized by adding Na^+ and Cl^- ions. At first, the system was allowed to relax to reach equilibrium state: the cell was first stirred in the NVT ensemble and after that kept in the NPT ensemble with constant pressure (1 atm) and temperature (300 °K), using a time step of 2 fs. The temperature was controlled using the modified Berendsen thermostat algorithm [13] and pressure was controlled using the Parrinello–Rahman method [14].

Equilibrated simulations were performed on the systems for 10 ns. Following the setting up of parameters, the MD simulations were conducted for each protein state. Newton's equations of motion were integrated with the leap-frog algorithm with a time step of 2 fs. For each subgroup, the temperature (300 °K) and isotropic pressure (1 atm) were kept constant using the Berendsen coupling algorithm [13] with time constants $\tau_T = 0.1$ ps and $\tau_P = 1$ ps, respectively.

Interactions within the short-range cutoff of 1.2 nm were evaluated at every time step. Long-range electrostatic interactions were calculated using the particle-mesh Ewald (PME) summation method [15] with a 6th order interpolation polynomial and a 1.2 nm cutoff that were updated every 10 steps together with the pair-list. Bond lengths were constrained using the LINear Constraint Solver (LINCS) algorithm [16] for the protein. The set-ups were then simulated for 30 ns.

3. RESULTS AND DISCUSSION

Peptide series were derived from the structure of the ACE2 $\alpha 1$ domain of ACE2 [3,5]. Starting from the peptide sequence STIEEQAKTFLDKFNHEAEDLFYQSSL, located in this domain and containing amino acids 19–45, the peptide ligands of different lengths were derived by truncating this sequence systematically from both ends until octapeptide length. The list of all 200 studied peptides and their docking energy values with the S1 binding site are presented in Table 1. It can be seen that the peptide docking energy depends on peptide structure, and firsthand, the influence of peptide length stands out. This dependence is illustrated in Fig. 3, presenting the docking energy landscape for the used set of compounds.

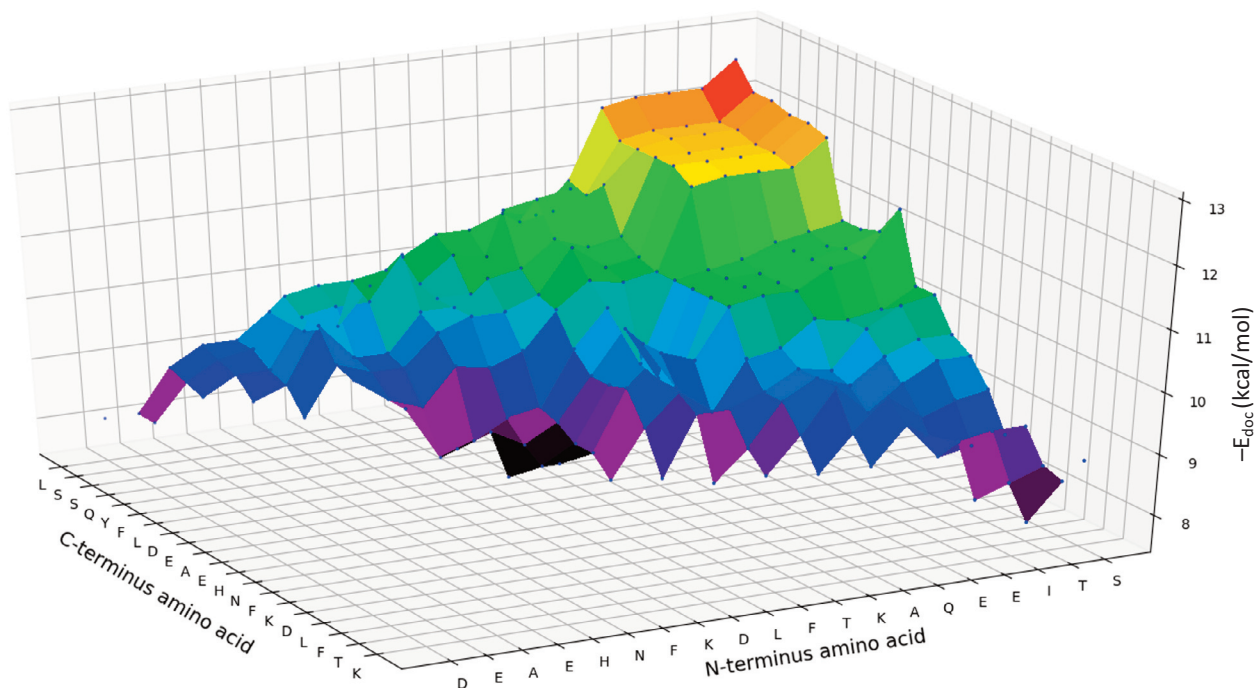


Fig. 3. Docking energy landscape for binding of peptides with the receptor binding site on SARS-CoV-2 spike protein S1. Peptides were derived from the ACE2 sequence 19–45, STIEEQAKTFLDKFNHEAEDLFYQSSL, through systematic truncation from the N-terminus and C-terminus, and the end amino acids of peptides are shown in the plot.

Although for getting systematic structure-affinity relationships, more extensive variations of the structure of the peptides must be made. Still, some binding site properties can be assessed by grouping the obtained results into subseries. For example, starting from the longest sequence studied (amino acids 19–45), two subseries can be designed. The first subseries includes peptides, which sequence starts from amino acid 19 and is truncated at the C-end. This group of peptides, further denoted as series 19-N, includes also peptides 19–44, 19–43 and so on, up to the octapeptide 19–26. In the same way, the longest peptide can be systematically shortened from the N-terminus, and this subseries, denoted as N-45, includes sequences 20–45, 21–45 and so on, up to the peptide containing amino acids 38–45. Such data grouping simplifies their meaningful analysis, as can be seen in Fig. 4, where the values of E_{doc} are compared for peptide subseries N-45 and 19-N, and these results can be discussed in terms of the binding site “topography”.

It is important to mention that the amino acids in positions below 19 are not fixed in the protein structure and do not participate in the formation of the binding domain $\alpha 1$. Therefore, extension of the peptide sequence in this direction was not made also in this study. On the other end, however, amino acids in this non-structured area of the protein N-terminus have practically no influence on the peptide docking energy, as the E_{doc} vs N plot levels off at $N < 24$ (subseries N-45). Interestingly,

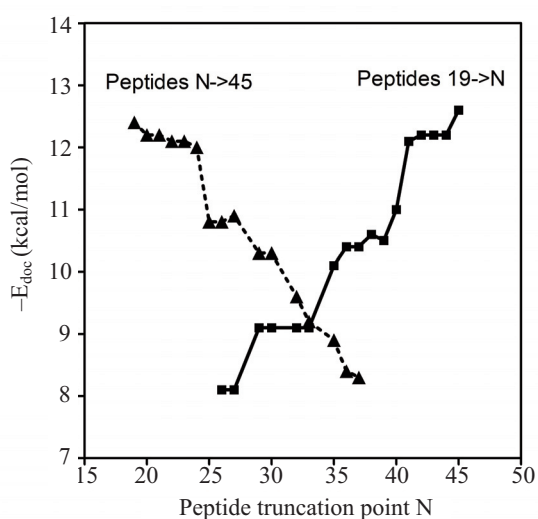


Fig. 4. Dependence of docking energy on peptide length in the case of two subseries, derived from the sequence STIEEQAKTFLDKFNHEAEDLFYQSSL (amino acids 19–45 in ACE2 structure) by systematic truncation of the N-terminus and C-terminus of the peptide. The two series are denoted as N->45 and 19->N, where N stands for the last number of the amino acid in the peptide.

this conclusion seems to be in line with our understanding about the interaction interface between ACE2 and S1, proposed in [2], showing that amino acid 24 is the first in the ACE2 sequence, which interacts with the binding site on S1. Therefore, these results of the binding site mapping agree with data about the interaction interface. The E_{doc} vs N plot levels also off in the case of peptides belonging to subseries 19-N, where the truncation is made at the C-terminus, while the amino acid in position 45 seems to be involved in peptide binding, as was also found in [3].

The binding site mapping reveals that effectively binding peptides may be shorter than the full $\alpha 1$ domain of ACE2. On the other hand, the peptides having docking energy $E_{\text{doc}} \leq -12$ kcal/mol must have at least 22 amino acids. This finding leaves plenty of room for a more systematic peptide design, and supports feasibility of the idea about using the peptide ligands as medications for COVID-19 disease, as administration of short peptides appears to be more acceptable compared with large peptides or proteins, including the soluble ACE2 itself [17].

Interaction of ACE2 with S1 of SARS-CoV-2 can be described by the apparent dissociation constants K_d , and these values ranging between 94.6 nM [3] and 14.7 nM [4], were recently published for this interaction. These dissociation constants correspond to the Gibbs free energy values -9.6 kcal/mol and -10.4 kcal/mol for this protein-peptide interaction, respectively. Although these values remain within the range of docking energies listed in Table 1, these parameters cannot be directly compared, as the free energy includes contribution of entropy that can play significant role in the process of protein-protein or protein-peptide interaction [18]. However, still the nanomolar range for K_d values can be predicted for the most potent peptides listed in Table 1. This prediction is supported by the fact that the binding energy of the $\alpha 1$ domain of ACE2 and the whole protein, calculated in [7], demonstrated similar binding effectiveness of the peptide and the protein. Therefore, the design of peptide inhibitors for hindering the CoV-2 interaction with the host cells and entering the viral RNA into cell through interaction with the ACE2 binding domain seems to be a promising approach, especially if shorter peptides can be used in drugs and their binding effectiveness can be further optimized.

4. CONCLUSIONS

Docking of 200 peptides, derived from the ACE2 binding domain sequence, which is responsible for CoV-2 spike protein S1 binding with the host cell surface, has been studied for mapping of the protein-protein interaction interface. The results demonstrate that the peptide binding

effectiveness depends on the peptide length and certain optimal structure can be found for this interaction. We suggest that further optimization of this interaction is possible through systematic variation of the amino acid composition of the peptides and application of quantitative structure-activity analysis together with docking analysis and molecular dynamics calculations, and this study could be important for designing antiviral drugs to inhibit CoV-2 activity.

ACKNOWLEDGEMENTS

This work was financially supported by QanikDX OÜ, Estonia (grant LLTKT20014). Computational analysis was carried out in the High Performance Computing Center of University of Tartu. The publication costs of this article were partially covered by the Estonian Academy of Sciences.

REFERENCES

- Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., Zheng, X.-S., Zhao, K., Chen, Q.-J., Deng, F., Liu, L.-L., Yan, B., Zhan, F.-X., Wang, Y.-Y., Xiao, G.-F., and Shi, Z.-L. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 2020, **579**, 270–273. <http://doi.org/10.1038/s41586-020-2012-7>
- Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., and Zhou, Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science*, 2020, **367**, 1444–1448. <http://doi.org/10.1126/science.abb2762>
- Wang, Q., Zhang, Y., Wu, L., Niu, S., Song, C., Zhang, Z., Lu, G., Qiao, C., Hu, Y., Yuen, K. Y., Wang, Q., Zhou, H., Yan, J., and Qi, J. Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell*, 2020, **181**, 894–904. <http://doi.org/10.1016/j.cell.2020.03.045>
- Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., Graham, B. S., and McLellan, J. S. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*, 2020, **367**, 1260–1263. <http://doi.org/10.1126/science.abb2507>
- Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., and Wang, X. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, 2020, **581**, 215–220. <http://doi.org/10.1038/s41586-020-2180-5>
- Wu, C., Liu, Y., Yang, Y., Zhang, P., Zhong, W., Wang, Y., Wang, Q., Xu, Y., Li, M., Li, X., Zheng, M., Chen, L., and Li, H. Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharm. Sin. B*, 2020, **10**(5), 566–788. <http://doi.org/10.1016/j.apsb.2020.02.008>
- Han, Y. and Král, P. Computational design of ACE2-based peptide inhibitors of SARS-CoV-2. *ACS Nano*, 2020, **14**(4), 5143–5147. <http://doi.org/10.1021/acsnano.0c02857>
- Li, Y., Zhang, Z., Yang, L., Lian, X., Xie, Y., Li, S., Xin, S., Cao, P., and Lu, J. The MERS-CoV receptor DPP4 as a candidate binding target of the SARS-CoV-2 spike. *iScience*, 2020, **23**, 101160. <http://doi.org/10.1016/j.isci.2020.101160>
- Trott, O. and Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.*, 2010, **31**(2), 455–461. <http://doi.org/10.1002/jcc.21334>. PMID: 19499576; PMCID: PMC3041641
- Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, 2008, **4**(3), 435–447. <https://doi.org/10.1021/ct700301q>
- Oostenbrink, C., Villa, A., Mark, A. E., and van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.*, 2004, **25**(13), 1656–1676. <http://doi.org/10.1002/jcc.20090>
- Berendsen, H. J. C., Grigera, J. R., and Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.*, 1987, **91**(24), 6269–6271. <https://doi.org/10.1021/j100308a038>
- Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 1984, **81**(8), 3684–3690. <https://doi.org/10.1063/1.448118>
- Parrinello, M. and Rahman, A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.*, 1981, **52**(12), 7182–7190. <http://doi.org/10.1063/1.328693>
- Darden, T., York, D., and Pedersen, L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, 1993, **98**(12), 10089. <http://doi.org/10.1063/1.464397>
- Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comp. Chem.*, 1998, **18**(12), 1463–1472. [https://doi.org/10.1002/\(SICI\)1096-987X\(199709\)18:12<1463::AID-JCC4>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H)
- Monteil, V., Kwon, H., Prado, P., Hagelkrüys, A., Wimmer, R. A., Stahl, M., Leopoldi, A., Garreta, E., Hurtado Del Pozo, C., Prosper, F., Romero, J. P., Wirnsberger, G., Zhang, H., Slutsky, A. S., Conder, R., Montserrat, N., Mirazimi, A., and Penninger, J. M. Inhibition of SARS-CoV-2 infections in engineered human tissues using clinical-grade soluble human ACE2. *Cell*, 2020, **181**(4), 905–913. <http://doi.org/10.1016/j.cell.2020.04.004>
- Sun, Z., Yan, Y. N., Yang, M., and Zhang, J. Z. H. Interaction entropy for protein-protein binding. *J. Chem. Phys.*, 2017, **146**(12), 124124. <http://doi.org/10.1063/1.4978893>

ACE2 sidumistsenter viiruse SARS-CoV-2 teravikvalgul S1: peptiidide dokkimise uuring

Aleksei Kuznetsov ja Jaak Järv

Arendusfaasis olevate erinevate COVID-19 ravivõimaluste hulgas eristub selge huvi inhibiitorite väljatöötamise vastu, mis blokeerivad viiruse tungimist peremeesrakku ja hõlbustavad selle lihtsa mehhanismi kaudu haiguse ravi. Sellest ideest lähtudes on käesolevas töös uuritud SARS-CoV-2 teravikvalgul S1 asuva sidumistsentri omadusi, sest selle kaudu seostub see viirus raku pinnal asuva retseptorvalguga ACE2, millele järgneb viiruse RNA tungimine peremeesrakku. Uuringuks vajalikud peptiidid (200 ühendit) konstrueeriti ACE2 struktuurist lähtudes ja nende jaoks arvutati dokkimise energia, mis kirjeldab nende seostumist viiruse teravikvalguga S1. Need arvutused näitavad, et juba uuritud peptiidide hulgas esineb mitmeid ühendeid, mis efektiivselt seonduvad teravikvalguga ja seega takistavad viiruse seondumist rakuga. Saadud andmetest lähtudes on võimalik jätkata uurimistööd peptiidide sidumise efektiivsuse edasiseks suurendamiseks, pidades silmas viirusinfektsioonivastaste ravimite loomise perspektiivi.