Proceedings of the Estonian Academy of Sciences,
2020, **69**, 4, 355–360
https://doi.org/10.3176/proc.2020.4.09
Available online at www.eap.ee/proceedings

COMPUTATIONAL BIOLOGY

Mapping the ACE2 binding site on the SARS-CoV-2 spike protein S1: molecular recognition pattern

Aleksei Kuznetsov* and Jaak Järv

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

Received 17 September 2020, accepted 9 October 2020, available online 22 October 2020

Abstract. Coronavirus SARS-CoV-2 enters the host cell via binding with the angiotensin-converting enzyme 2 (ACE2), and here we used computational modelling to study the molecular recognition pattern of this interaction. The fragment of the N-terminal part of the enzyme containing amino acids 19–45 was used as the lead peptide in this study. The structure of this peptide was systematically modified by successive replacement of its amino acids with alanine, serine, glycine, and phenylalanine. Then docking energies were calculated for all these mutant peptides. These docking energies were correlated with physical descriptors, proposed for the modelling of peptide–protein interactions, characterizing hydrophilicity and volume-related properties of amino acid side chains. From these correlations the corresponding specificity factors were obtained for all amino acid positions, and thus the full description of the molecular recognition pattern of the ACE2 α1 domain by the virus S1 protein binding site was obtained.

Key words: SARS-CoV-2 virus spike protein S1; peptide docking; peptide sequence scan with alanine, serine, and phenylalanine; binding site specificity; correlation analysis; COVID-19 therapeutics.

INTRODUCTION

Coronavirus SARS-CoV-2, responsible for the COVID-19 pandemic, enters human cells through the interaction of the virus surface spike protein S1 with the human angiotensin-converting enzyme 2 (ACE2) on the host cell membrane [1,2]. The initial binding step of the virus particle on the surface of the host cell is followed by the fusion of the viral and cell membranes and entry of the viral RNA into the host cell [2–4]. This mechanism of the virus entry suggests that inhibition of the interaction between the S1 protein and ACE2 may be a promising option to combat the infection. A rather straightforward way to achieve this goal is inhibiting one of the binding sites involved in this interaction by using peptides that mimic the counterpart protein. As ACE2 is physiologically important for the functioning of the host cell, inhibition of the virus binding site on this enzyme by using peptides derived from the S-protein structure, as proposed in [5], has understandable shortcomings. However, these do not apply in the case of peptides that mimic the ACE2 structure and inhibit the virus–receptor interaction by binding with the receptor-binding domain (RBD) on the S1 protein [6,7]. Therefore, this study, as well as our previous work [7], was focused on the design of such inhibitory peptides by using computational methods.

It can be suggested that the most straightforward design of effective peptide inhibitors that target the binding site on the S1 protein can be made proceeding from the peptide sequence of the virus binding domain of the ACE2 molecule [7]. The viability of this approach is based on the availability of structural data for proteins S1 and ACE2 and their complex [3,5], published as '6LZG' in the Worldwide Protein Data Bank (PDB) database (www.pdb.org). Using these data, we modelled the structure of the S1 and ACE2 complex as illustrated in Fig. 1.

Initially, these structural data were used for computational analysis of the interaction of RBD on the S1

* Corresponding author, aleksei.kuznetsov@ut.ee

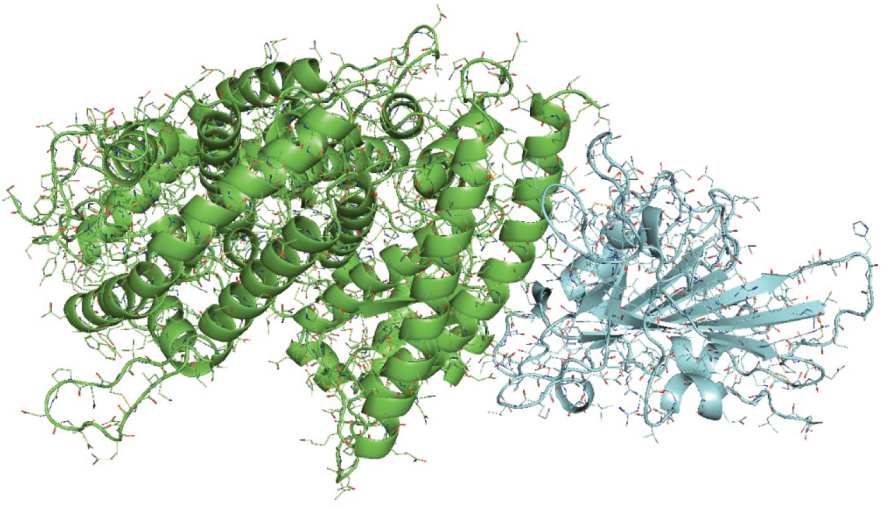

Fig. 1. Computer modelling of the complex formed between the SARS-CoV-2 spike protein S1 (right molecule, cyan) and human angiotensin-converting enzyme 2 (left molecule, green) by using structure data listed at '6LZG' in the PDB database (www.pdb.org).

protein with ACE2 and its fragments [6]. That study demonstrated effective interaction of the peptides derived from the ACE2 structure with the SARS-CoV-2 S1 protein, as already a single α_1-helix domain peptide that contains amino acids 21–55 of the ACE2 N-terminal sequence binds with the S1 protein with almost the same effectiveness as the full-size protein [6]. It was also revealed that this sequence includes 12 amino acid residues of ACE2 that seem to interact with the RBD of SARS-CoV-2, whereas 10 other amino acid residues participating in the virus binding process come from other parts of the ACE2 molecule [6]. Thus, the interaction of the spike protein with its receptor site on the ACE2 molecule is clearly focused on binding the α1 domain with the S1 protein (Fig. 1).

Following [6], in our earlier study [7] we mapped the α1 domain binding site on the S1 protein using computational docking analysis; in that study the peptide sequence 19–45 of the N-terminal part of ACE2 (STIEEQ AKTFLDKFNHEAEDLFYQSS) was truncated from both ends, and the docking of 200 peptide fragments at the binding site on the S1 protein was analysed. We found that the α1 domain sequence can be shortened to a certain extent without significant reduction of the docking energy, which is 'good news' for therapeutic peptide development [8]. This conclusion about the influence of peptide length was confirmed by results published in [9], where the binding of three peptides, also derived from the same ACE2 α1 domain and containing 26, 23, and 20 amino acids, was computationally investigated with the S1 protein.

Thereafter, similar docking analysis of peptides derived from the α1 domain of ACE2 and containing 23 and 19 amino acids was reported [10]. These results demonstrated that alteration of peptide length from 19 to 23 amino acids has practically no effect on the positioning of these compounds in the binding site on the S1 protein, although the binding effectiveness is somewhat reduced by peptide shortening. In conclusion, all these results support the idea that short peptides can effectively bind with the S1 protein and therefore can be used for developing antiviral drugs. To achieve this goal, it seems to be important to improve the effectiveness of peptide binding, first and foremost, through directed modification of the peptide primary structure. Therefore, we continued mapping the S1 binding site for ACE2-derived peptides. In this paper we analyse the molecular recognition pattern of this interaction interface by combining computational docking calculations with methods of quantitative structure–activity analysis.

METHODS

The input files used for modelling an ACE2 and the receptor binding domain of theCoV-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 the docking study were derived proceeding from the α1 domain of the ACE2 protein. The lead peptide sequence STIEEQAKTFLDKFNHEAEDLFYQSSL was systematically modified by gradually altering all amino acids by alanine, serine, glycine, and phenylalanine. Thus, the recognized procedure of alanine scanning [11] was extended to describe binding properties of mutants of serine, glycine, and phenylalanine by using computational data.

The conventional software for ligand docking AutoDock Vina version 1.1.2 [12] and molecular dynamics (MD) simulation GROMACS version 4.6.1 package [13] were used. The mutated peptides were constructed by PyMOL version 1.8.0.4 software package [14]. The graphical representation of protein structures was made by VMD version 1.9 package [15].

The lead peptide was mutated in each position of amino acids with conserved main scaffold of the peptide chain. The best scoring results of peptide positioning were picked for the peptide–S1 complex. The docking energy values were calculated and further processed by using conventional quantitative structure–activity analysis methods. The parameters and the procedure of the MD simulations were described in detail in our previous work [7].

RESULTS AND DISCUSSION

Lead peptide docking

The docking of the lead peptide STIEEQAKTFLDKF NHEAEDLFYQSSL, which includes amino acids 19–45 of the ACE2 α1 domain sequence, with the S1 protein was modelled computationally in our previous study [7]. The structure of this complex is illustrated in Fig. 2. The docking energy –12.6 kcal/mol was obtained from these calculations. The peptide–protein interaction interface includes several amino acids whose involvement in the hydrogen bond formation can be suggested from structural studies. This list includes from 7 amino acids listed in [16] (Q24, D30, E35, E37, D38, Y41, Q42) up to 13 amino acids listed in [3] (S19, Q24, T27, F28, D30, K31, H34, E35, E37, D38, Y41, Q42, L45).

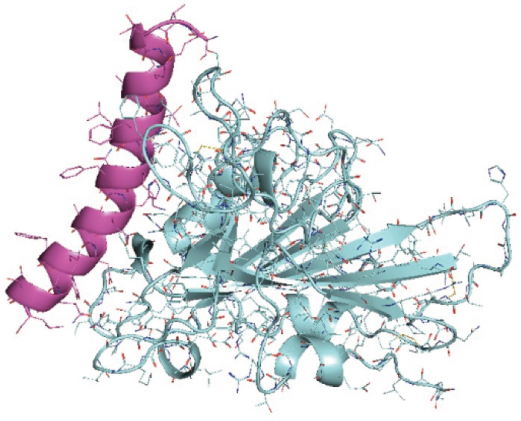

Fig. 2. Computer modelling of the complex between the SARS-CoV-2 spike protein S1 (right molecule, cyan) and the peptide STIEEQAKTFLDKFNHEAEDLFYQSS (left molecule, magenta), corresponding to the sequence 19–45 of the N-terminal α1 domain of the human angiotensin-converting enzyme 2 and serving as the main docking partner for virus binding with the host cell. The complex structure was built computationally by using the molecular dynamics method and structural data listed in the PDB database (www.pdb.org) as '6LZG'.

Recently one more computational analysis of this interaction interface was published [17]; there participation of other 'contact interactions' is mentioned in addition to hydrogen bonds. The list of contact interactions includes van der Waals and hydrophobic bonds as well as π/π and π/cation interactions. Based on [17], formation of hydrogen bonds with participation of amino acids Q24, K31, E35, E37, D38, Y41, and Q42 can be expected in the case of the lead peptide, while contact interactions can be suggested in the case of amino acids Q24, T27, D30, K31, H34, D38, Y41, and Q42. As can be seen, these lists significantly overlap, demonstrating that the actual specificity pattern that governs the interaction interface can be rather complex. This is a good justification of the following analysis.

Computational site-directed mutagenesis of the lead peptide

The contribution of individual amino acid residues into the effectiveness of peptide binding was studied by combining the site-directed mutagenesis method with computational docking analysis as suggested in [11]. Firstly, we made consecutive replacement of all amino acids with alanine, as the methyl group of this amino acid cannot be involved in polar interactions and therefore this method is often used for the determination of 'hot spots' of the protein–peptide interaction interface. Secondly, we also mapped polar and hydrophobic properties of the binding site by scanning the lead peptide with serine, glycine, and phenylalanine. Results of this analysis are summarized in Fig. 3, where the docking energies for alanine, serine, glycine, and phenylalanine mutants of the lead peptide are compared. As can be seen in Fig. 3, in many cases the docking effectiveness of the lead peptide, E_{dock} = –12.6 kcal/mol, is not affected by the replacement of the initial amino acid. This means that these amino acids are probably not involved in the peptide interaction with the S1 protein.

On the other hand, however, Fig. 3 also reveals several hot-spot positions as the alanine scan caused a weakening of the peptide binding in the following positions: Q24, T27, F28, D30, K31, E35, D38, Y41, and L45. Importantly, interaction of all these amino acids with the S1 protein has been suggested in structural studies cited above. Therefore, it can be concluded that the computational docking study describes adequately the peptide–protein interactions in the case of the formation of the ACE2–S1 complex.

However, the influence of alanine mutations on the docking energy is rather different along the peptide chain; moreover, these effects are not similar to the changes

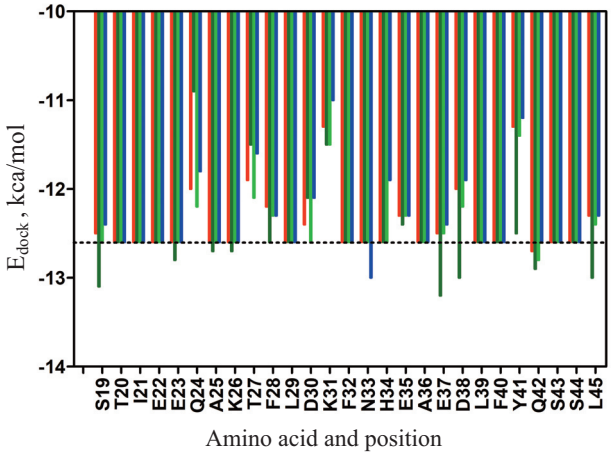

Fig. 3. Influence of the scanning of the lead peptide sequence STIEEQAKTFLDKFNHEAEDLFYQSSL with alanine (red), serine (light green), glycine (blue), and phenylalanine (dark green) on the docking energy E_{dock}. The docking energy of the lead peptide was –12.6 kcal/mol; this level is denoted by the dotted line.

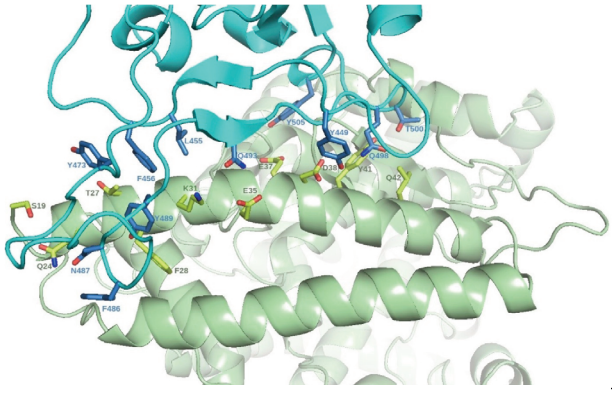

Fig. 4. Mutual positioning of the hot-spot amino acids in the α1 domain of the human angiotensin-converting enzyme 2 (green) and its binding site on the S1 protein (cyan).

caused by other mutations. This means that the interplay of different specificity factors governs, indeed, the binding process, as was suggested in [17].

It is interesting to note that all hot-spot amino acids are located on the same side of the spiral structure of the α1 domain, facing the S1 protein. This situation is illustrated in Fig. 4, where mutual positioning of the hot-spot amino acids in the α1 domain and its binding site on the S1 protein are shown.

Molecular recognition pattern

The molecular recognition pattern of the peptide binding interface can be characterized in terms of structure–activity relationships, assuming that the contribution of each amino acid can be presented as the sum of interactions quantified by certain specificity descriptors [18]. The possibility of encoding these interactions in terms of two orthogonal sets of descriptors, which characterize the volume-related (ϖ) and hydrophilicity-related (η) effects of amino acid side groups [19], simplifies this analysis and opens new perspectives for converting structural data to the property space. These descriptors, listed in Table 1, demonstrate that the hydrophilicity parameter η has a negative value for alkyl and non-polar groups and a positive value for polar and ionic groups, independently of the sign of the net charge on the group. The volume descriptors ϖ vary from –4.04 for glycine, which has the side group of minimal size, up to 4.28 in the case of tryptophan, which is the bulkiest.

It is important to mention that the parameters ϖ are well correlated with the conventional scale of the molar refractivity (MR) values ($R^2 = 0.9634$), commonly used for the characterization of the volume-related properties of amino acid side chains [20]. At the same time, the correlation between the hydrophilicity parameters η and the classical hydrophobicity parameters π [21] is weaker ($R^2 = 0.7153$); still these values show obvious similarities.

The systematic scans of peptide binding properties by using alanine, serine, glycine, and phenylalanine together with the amino acid in the parent peptide structure provide five data points for most amino acid positions.

Table 1. Descriptors characterizing volume-related properties and hydrophilicity of amino acid side groups. The values were compiled from [19]

Amino acid	Volume ϖ	Hydrophilicity η
A	−2.90	−1.03
R	2.41	1.31
N	−0.68	0.79
D	−0.92	1.23
C	−1.89	0.15
Q	0.36	1.09
E	0.16	1.28
G	−4.04	0.01
H	0.83	1.15
I	0.51	−1.32
L	0.52	−1.40
K	0.92	1.23
M	0.92	−1.42
F	2.22	−1.47
P	−1.25	−0.64
S	−2.36	0.38
T	−1.19	0.28
W	4.28	−0.18
Y	2.75	−0.18
V	−0.65	−1.27

This is sufficient for structure–activity correlation using the descriptors in Table 1. For example, in the case of mutants Q24A, Q24S, Q24G, and Q24F the E_{dock} values were –12.0, –12.2, –11.8, and –10.9 kcal/mol, respectively. Together with the E_{dock} value –12.6 kcal/mol for Q24, the following correlation was obtained:

$$E_{dock} = (-11.9 \pm 0.22) + (0.06 \pm 0.08)\varpi + (-0.49 \pm 0.21)\eta. \quad (1)$$

This interrelationship describes the influence of the amino acid side group on the docking energy in terms of two specificity factors, ϖ and η. In this case the volume-related effects, characterized by ϖ, play no statistically relevant role, whereas the docking energy is governed by the hydrophilicity (η) of the amino acid side group, quantified by the specificity factor –0.49.

Similar correlations were obtained for all positions of the peptide sequence. Therefore, we suggest that these data characterize the molecular recognition pattern of the binding interface. All results of this analysis are summarized in Fig. 5. It is important to emphasize that the physical meaning of the interactions depends on the selection of the descriptor sets.

Hydrophilicity is revealed in the case of most polar amino acids in the lead peptide (Fig. 5). At the same time the contribution of this specificity factor varies significantly in different positions. On the other hand, there are also volume-related effects (see Fig. 5), and these interactions are also distributed throughout the peptide. In most cases these effects support peptide binding.

Hot spots and the molecular recognition pattern

The location of the hot-spot amino acids in the lead peptide was identified by structural studies as described above as well as by alanine scan as shown in Fig. 2. In summary, this list of amino acids includes S19, Q24, T27, D30, K31, E35, E37, D38, Y41, and L45. As Fig. 5 shows, these positions are indeed important for peptide binding because at least one specificity factor makes a significant contribution in these cases. Therefore, the molecular recognition pattern provides the same information as other approaches. On the other hand, however, the recognition pattern characterizes these interactions also quantitatively.

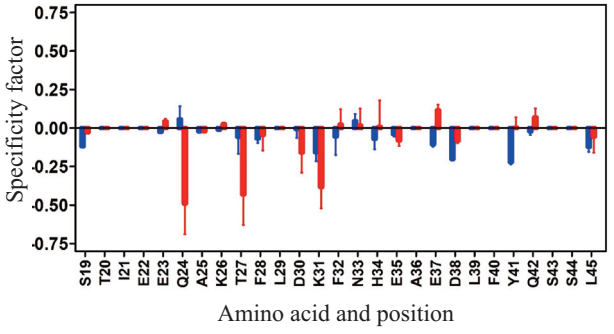

Fig. 5. Intensity factors of hydrophilicity (red bars) and volume effects (blue bars) in different amino acid positions for the binding of the lead peptide, STIEEQAKTFLDKFNHEAEDLFYQSSL, with the spike protein S1 binding site. The specificity factors were calculated by correlating the docking energies for alanine, serine, and phenylalanine mutants of the peptide with the corresponding physical descriptors. Standard errors of the specificity factors are shown.

In this study computational docking energy values were used for analysing the peptide–protein interaction interface; therefore, the effects of entropy on the recognition pattern were not analysed. Irrespective of this drawback, we believe that the presented results may still be useful for further optimization of the peptide structure and will enable improving its binding properties. Most interestingly, already the scan results reveal some point mutations that increased the peptide binding effectiveness (Fig. 3).

CONCLUSIONS

Docking of 108 peptides derived from the ACE2 binding domain sequence (19–45) STIEEQAKTFLDKFNHEAE DLFYQSSL with the receptor binding site of the CoV-2 virus S1 protein was analysed and the molecular recognition pattern of the peptide–protein interaction interface was mapped quantitatively. The studied peptides were obtained by systematic scanning of the lead peptide sequence with alanine, serine, glycine, and phenylalanine. The results revealed that replacement of amino acids in the lead peptide reduced its binding effectiveness in certain critical positions (Q24, T27, D30, K31, E35, E37, D38, Y41, and L45), which agree with the sites where the formation of hydrogen bonds and a salt bridge can be predicted proceeding from structural data. The scanning results were used for correlation analysis of the influence of the amino acid side chain structure on peptide binding, and the contributions of distinct interactions were quantified at each amino acid position. The set of orthogonal descriptors that characterize volume-related and hydrophilic properties of the amino acid side groups was used for this analysis. The set of these correlations was used to characterize quantitatively the molecular recognition pattern of the peptide binding site on the CoV-2 virus S1 protein.

ACKNOWLEDGEMENTS

This work was financially supported by QanikDX OÜ, Estonia, registration No. 4523084 (grant LLTKT20014). Computational analysis was carried out in the High-Performance Computing Center of the University of Tartu.

REFERENCES

1. Shang, J., Wan, Y., Luo, C., Ye, G., Geng, Q., Auerbach, A., and Li, F. Cell entry mechanisms of SARS-CoV-2. *PNAS*, 2020, **117**, 11727–11734. https://doi.org/10.1073/pnas.2003138117
2. Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A., and Li, F. Structural basis of receptor recognition by SARS-CoV-2. *Nature,* 2020, **581**, 221–224. https://doi.org/10.1038/s41586-020-2179-y
3. 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. https://doi.org/10.1016/j.cell.2020.03.045
4. 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. https://doi.org/10.1126/science.abb2762
5. Peter, E. and Schug, A. The inhibitory effect of a Corona virus spike protein fragment with ACE2. *bioRxiv* 2020.06.03.132506. https://doi.org/10.1101/2020.06.03.132506
6. Han, Y. and Král, P. Computational design of ACE2-based peptide inhibitors of SARS-CoV-2. *ASC Nano*, 2020, 5143–5147. https://doi.org/10.1021/acsnano.0c02857
7. Kuznetsov, A. and Järv, J. Mapping of ACE2 binding site on SARS-CoV-2 spike protein S1: docking study with peptides. *Proc. Estonian Acad. Sci*., 2020, **69**, 228–234. https://doi.org/10.3176/proc.2020.3.06
8. VanPatten, S., He, M., Altiti, A., Cheng, K. F., Ghanem, M. H., and Al-Abed, Y. Evidence supporting the use of peptides and peptidomimetics as potential SARS-CoV-2 (COVID-19) therapeutics. *Future Med. Chem.*, 2020. Advance online publication. https://doi.org/10.4155/fmc-2020-0180
9. Barh, D., Tiwari, S., Andrade, B., Giovanetti, M., Costa, E., Kumavath, R., Ghosh, P., Góes-Neto, A., Alcantara, L., and Azevedo, V. Potential chimeric peptides to block the SARS-CoV-2 spike receptor-binding domain. *F1000Research*, 2020, **9**, 576. https://doi.org/10.12688/f1000research.24074.1
10. Baig, M., Alagumuthu, M., Rajpoot, S., and Saqib, U. Identification of a potential peptide inhibitor of SARS-CoV-2 targeting its entry into the host cells. *Drugs R&D*, 2020, **20**, 161–169. https://doi.org/10.1007/s40268-020-00312-5
11. Kortemme, T., Kim, D. E., and Baker, D. Computational alanine scanning of protein-protein interfaces. *Science's STKE*, 2004, **2004**, pl2. https://doi.org/10.1126/stke.2192004pl2
12. Trott, O. and Olson, A. 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. https://doi.org/10.1002/jcc.21334
13. 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
14. PyMOL. The PyMOL Molecular Graphics System. Version 1.8.0.4. Schrödinger, LLC. https://www.pymol.org
15. Humphrey, W., Dalke, A., and Schulten, K. VMD: visual molecular dynamics. *J. Molec. Graphics*, 1996, **14**(1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5
16. 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. https://doi.org/10.1038/s41586-020-2180-5
17. Laurini, E., Marson, D., Aulic, S., Fermeglia, M., and Pricl, S. Computational alanine scanning and structural analysis of the SARS-CoV-2 spike protein/angiotensin-converting enzyme 2 complex. *ACS Nano*, 2020, **14**, 11821–11830. https://doi.org/10.1021/acsnano.0c04674
18. Sneath, P. Relations between chemical structure and biological activity in peptides. *J. Theor. Biol*., 1966, **12**, 157–195. https://doi.org/10.1016/0022-5193(66)90112-3
19. Barley, M., Turner, N., and Goodacre, R. Improved descriptors for the quantitative structure–activity relationship modeling of peptides and proteins. *J. Chem. Inf. Model.*, 2018, **58**, 234–243. https://doi.org/10.1021/acs.jcim.7b00488
20. Järv, J. and Ragrarsson, U. Linear free energy relationships in cAMP-dependent protein kinase reactions with synthetic substrates. *Bioorg. Chem*., 1991, **19**, 77–87. https://doi.org/10.1016/0045-2068(91)90045-Q
21. Hansch, C. Recent advances in biochemical QSAR. In *Correlation Analysis in Chemistry* (Chapman, N. B. and Shorter, J., eds). Springer, Boston, MA, 1978, 397–438. https://doi.org/10.1007/978-1-4615-8831-3_9

ACE2 sidumistsenter viiruse SARS-CoV-2 teravikvalgul S1: sidumiskoha spetsiifilisuse mustri uuring

Aleksei Kuznetsov ja Jaak Järv

SARS-CoV-2 teravikvalgul S1 asuv sidumistsenter seostub raku pinnal asuva retseptorvalguga ACE2, millele järgneb viiruse RNA tungimine peremeesrakku. Seetõttu pakub huvi selle sidumistsenri omaduste detailsem uurimine, milleks käesolevas töös konstrueeriti retseptorvalgu struktuurist lähtudes peptiidi analoogid ja arvutati nende jaoks dokkimisenergia väärtused. Saadud tulemustest lähtudes kirjeldati S1 teravikvalgu sidumistsentri spetsiifilisuse mustrit. Uuritavad ühendid saadi lähtepeptiidi järjestuse STIEEQAKTFLDKFNHEAEDLFYQSSL skaneerimisel alaniini, seriini, glütsiini ja fenüülalaniiniga. Arvutustulemused näitavad, millised aminohapped vastutavad viiruse ja retseptorvalgu seostumise eest ning need punktmuteerimise andmed on kooskõlas ACE2–S1 kompleksi struktuuri uuringute tulemustega. Kasutatud aminohapete külgrühmade struktuuri ja nende mõju vahelise sõltuvuse analüüs lubab aga sidumiskoha spetiifilisust ka kvantitatiivselt kirjeldada. Saadud andmetest lähtudes saab otsida võimalusi peptiidide sidumise efektiivsuse parandamiseks, mis on oluline viiruse toimet inhibeerivate peptiidide loomise seisukohast.