In Silico Studies for Design and Development of Inhibitor against COVID-19

Akshata R. Pahelkar[1], Pritam V. Bagwe[1], Khushboo Maurya[1], Nikhil R. Pahelkar[2], Shreerang V. Joshi[1], Vikas N. Telvekar[1*]

[1]Department of Pharmaceutical Sciences and Technology, Institute of Chemical Technology, Nathalal Marg, Matunga, Mumbai, 400019, Maharashtra, India; [2]Department of Biosciences and Bioengineering, Indian Institute of Technology Bombay, IIT Area, Powai, Mumbai, 400076, Maharashtra, India

ABSTRACT

Aim: The COVID-19 pandemic as declared by WHO, reported its first case in Wuhan, Hunei province, China. The infections have now been widespread and as of 30th March 2020, 6, 93,224 infection cases and 33,106 deaths have been reported worldwide. This is an international concern related to public health. Several research studies are being carried on around the world to amid the crises. Various treatment and prevention hypothesis regarding vaccine development, repurposing of drugs along with development of new chemical entities are under investigation. COVID-19 is a virus from the coronavirus family of viruses of beta genus. It has similarity with the SARS-CoV which was reported earlier. We plan to propose a scaffold for designing peptide derived specific inhibitor against COVID-19.

Methods: Our strategy involves designing a peptide inhibitor that will interact with RBM of the SARS-CoV-2 and inhibit its entry inside the cell. Computer Aided drug design is one of the tools available which can help in achieving productivity in less time span. We utilised online peptide designing platform called rosetta, Autodock Vena a computer aided drug designing software (Licensed to Mr. Nikhil Pahelkar) and Schrodinger's Drug Discovery Suite (Licensed to ICT, Mumbai).

Result: We propose a peptide based on the sequence of human ACE2 which interacts strongly with RBM of the SARS-CoV-2. Interested readers can utilize the sequence for further modification of the peptide sequence to convert it into a peptidomimetic.

Keywords: Computer aided drug design; COVID-19; Peptides; ACE 2; Spike protein; pandemic.

ABBREVIATIONS

COVID-19: Coronavirus disease 2019; WHO: World Health Organization; PHEIC: Public Health Emergency of International Concern; SARS-CoV-2: Severe Acute Respiratory Syndrome Coronavirus-2; 2019-nCoV: Novel Coronavirus; ACE2: Angiotensin Converting Enzyme 2; RBM: Receptor Binding Motif; RBD: Receptor Binding Domain; S1: Subunit 1; S2: Subunit 2; OPLS: Optimized Potentials for Liquid Simulations; PDB id: Protein Data Bank Identification code; RCSB: Research Collaboratory for Structural Bioinformatics; REU: Residue units

INTRODUCTION

Coronavirus disease 2019 (COVID-19) is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The disease was first identified in 2019 in Wuhan, the capital of Hubei China, and has since spread globally, resulting in the 2019-20 coronavirus pandemic [1,2]. Common symptoms include fever, cough, and shortness of breath. Muscle pain, sputum production, diarrhoea, and sore throat are less common. While the majority of cases result in mild symptoms, some progress to pneumonia and multi-organ failure [3-5].

The COVID-19 virus is said to have acquired a mutation which allows it to have increased affinity towards human ACE2 receptor [6-8]. It also has similar binding affinity for ACE2 receptor of those animals such as cats, ferrets, etc., which share sequence homology with human ACE2 [9]. It is this mutation that is allowing the virus

to be more contagious (Human to human transmission) [10]. It has also been observed that even though there is sequence similarity between SARS-CoV and SARS-CoV-2, the antibodies for SARS-Cov are not active against SARS-CoV-2. This suggests the need of a designed inhibitor which will be structure and conformation specific to bind SARS-CoV-2. Protein-protein interaction inhibitors can be designed by taking into consideration the protein epitopes, the actual sequence that interacts and binds. Whole structure of SARS-CoV-2 can be described as a sequence that has namely two regions, the core region and the spike. The spike region is divided into two further subunits, S1 (Subunit1) which has the active site and the RBD (Receptor Binding Domain) domain, which flanks RBM (Receptor Binding Motif) [11,12]. It is the RBM that interacts with human ACE2 enzyme. S2 (Subunit2) has the poly basic furin cleavage site and O-linked glycans [8]. This site is responsible for its resistance against furin based proteases. It is reported that the O-linked glycans may be responsible for production of a mucin like substance that protects the virus epitopes [8,13-15]. This suggests that an Anti-Biofilm agent/antibiotic active against bacterias that produce capsule in combination with the effective drug could be a promising treatment therapy. Although more studies in this matter would be confirmatory [16,17]. The nine conservative amino acids essential for the interaction were determined from the sequence homology of SARS-CoV and SARS-CoV-2. We have utilised a RCSB deposited PDBid 6M17 as input after preprocessing and minimising the crystal structure for designing a peptide based on human ACE2 sequence. PDBid 6M17 is the crystal structure of SARS-CoV-2 RMD bound to the ACE2.

In the present study, we aimed to design a scaffold peptide sequence which can be used by interested readers for developing new derivatives. We utilised a complex of SARS-CoV-2 RMD bound to the of ACE2 crystal structure PDBid 6M17 as an input, after processing (processing was performed using Schrodinger's Drug Discovery Suite Software), for rosetta online web server. Rosetta server inspects the interacting sequence residues of the complex (The two interacting domains i.e. of SARS-CoV-2 RMD and ACE2) and based on the epitope sequence complementarity, conformation it gives a short peptide sequence that will have similar or stronger interaction with chain/protein of interest. The output peptide sequence is claimed to interact and modify the native protein-protein interaction. Thus, can serve as the initiation point of further development. The sequence obtained was modified on the basis of binding site analysis of target i.e. SARS-CoV-2 RMD, receptor grid cavity and interacting sequence residues. Modification was such, to examine the effect of increase in chain length and increase in aromatic amino acid residues for higher interaction and bond formation. For increasing the interaction further and binding to a specific residue of SARS-CoV-2 RMD sequence, aromatic amino acids were incorporated, modifying the sequence. The output peptide sequence was 10 residue long, which was increased up to 12 aa residue longer. Derived peptides along with the newly modified were docked to check the binding interaction. Docking studies were performed using Autodock Vena Software. The modifications made showed positive result and higher interaction. Thus, further modifications can be made so as to improve the pharmacokinetic and pharmacodynamic profile of the scaffold peptide to make that drug-able candidate.

The basic idea is that SARS-CoV-2 should have more affinity and thus bind with more interaction energy to these derived peptides. Though this study is an overview, it will surely help researchers in the field with a basic scaffold to proceed with for designing specific peptidomimetics.

MATERIALS AND METHODS

Rosetta online web server was accessed online by Ms. Pritam Bagwe from MacBook Air. Autodock Vena modification and docking operations were performed by Mr. Nikhil Pahelkar on Windows 10 pro platform of Dell laptop-l4JGKR3, Intel core i3 processor and 4GB RAM. Software License was available on the workstations setup in the ICT, Mumbai CADD lab. Workstations have linux CentOS platform, 4 processor speed and 8GB internal with additional external storage capacity.

Protein preparation

Protein preparation was performed by using Protein Preparation Wizard of Maestro interface, Schrödinger software. The PDB Id-6M17 was downloaded and treated to add missing hydrogen, assign proper bond orders, treat metal (breaking bonds to metal and correct the formal charge on metal and neighboring atoms), and to delete water molecules that are more than 5 Å from the heterogeneous groups. The H bonds were displayed. Finally, the protein structures were minimized to the default root-mean-square deviation value of 0.30 Å. The inhibitor structure was minimized using OPLS_3 e force field. Minimization was carried out at pH 7.4. 6Ml7 is a complex of SARS-CoV-2's receptor binding domain and ACE2 and the sodium-dependent neutral amino acid transporter B(O) AT1 in association with ACE2. This complex is present as double homo-mer units, so only one individual chain of the RBD of SARS-CoV-2 binding that of ACE2 was retained. Later the complex was optimized and minimised. The complex PDB downloaded from RCSB is shown in the Figure 1 for reference.
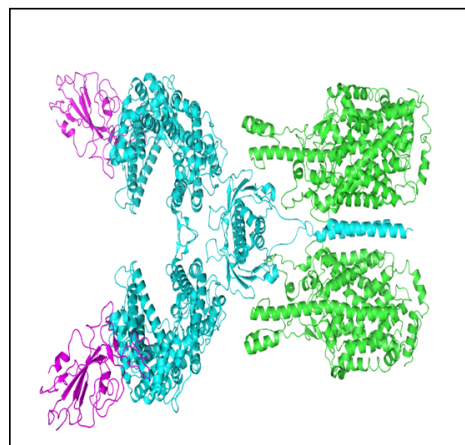

Figure 1: The complex PDBid 6M17 imported from RCSB (Where, spike protein shown in pink color, ACE2 shown in sky blue in color and sodium-dependent neutral amino acid transporter shown in green color).

Obtaining peptide based scaffold from Rosseta web server

The minimized complex was used as input on another software website; Rosetta Online Server. We entered the PDB complex as input and mentioned the description of the job and expected number of amino acid residue peptide the software should predict for us. The software Peptide rive [18] identifies a linear peptide segment for a given protein-protein interaction. A sequence segment that contributes the most to binding of the proteins to each other is analysed and can serve as inhibitor for the complex interaction. The peptides that are derived from this sequence

can be modulators of the protein-protein interaction. In current case, the peptides can modulate the Viral Spike and ACE2 interaction and terminate the infection cascade. A list of derived peptides is obtained as output. This is the list of the peptides that contribute most to the interaction of the complex. The present situation generated only one peptide sequence. The software also provides quantitative and visual data about the binding mode and energetics. From the many sequence segments that the software considers, the fragment which constitute a significant portion of the binding energy are considered to be potential candidates for competing with the existing Viral spike and ACE2 complex interaction. This could be the starting sequence for designing and developing peptide-based inhibitors against SARS-CoV-2. Utilizing the software derived sequence as scaffold we designed more five peptide sequences for analyzing binding energy. The modifications were made based on the knowledge of binding site/receptor grid analysis and the major interacting residues. Basically, the aim was to engage those residues of SARS- CoV-2 RBM which interacted with ACE2.

Docking study of peptides

We used the smina docking protocol [19], a fork of AutoDock Vina (v1.1.2) [20], on the selected compounds to quantify their interaction with each of the representative's structures from the generated ensemble. The ligand PDBQT files for docking were generated using the prepared ligand PDBQT script that comes with AutoDock Vina Tools (v1.5.4) [21]. Docking was performed at an exhaustiveness of 8, the search space was defined using the automatic method with smina, and a random seed of 1412428560 was selected. The active site residues mentioned for grid generation were residue number 442, 455, 472, 479, 480, 485, 487, 505 of receptor binding motif of SARS-CoV-2 hits with a binding free energy below -7.0 kcal/mol. Likewise, the interaction with all five ensemble structures were quantified. Compounds with a free binding energy below -6.5 kcal/mol were then subjected to clustering based on the Tanimoto comparison of per-atom molecular fingerprints. For each of the 8 resulting clusters, the best two compounds (one compound in the case of a single structure per cluster) according to the binding energy were selected. The residues present at the interface of the complex are shown in the Figure 2. Majority of the active site residues of both the proteins lie within the complex interface.
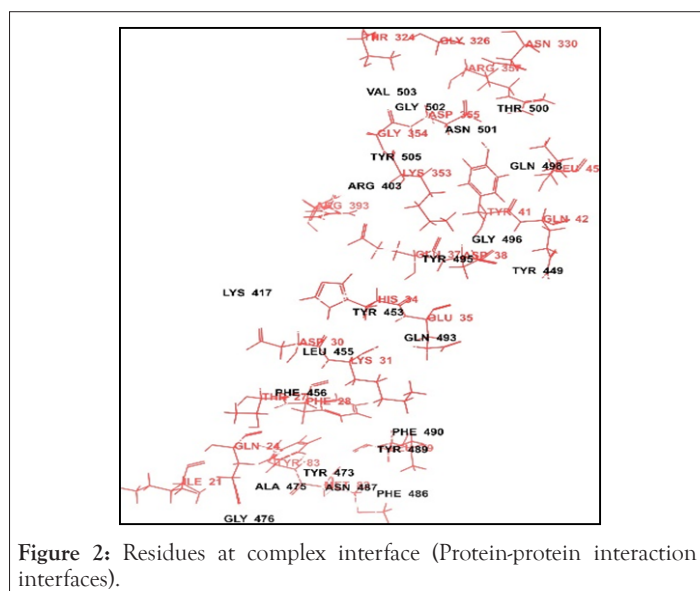

Figure 2: Residues at complex interface (Protein-protein interaction interfaces).

RESULTS AND DISCUSSION

Preprocessed, optimised and minimised SARS-CoV-2RBD and ACE2 complex

The PDBid 6M17 after minimization has a complex of SARS-CoV-2RBD and ACE2. The sodium-dependent neutral amino acid transporter B(O)AT1 in the complex is retained since it was found in the bound form to the ACE2 and may have some inherent function. The minimisation was carried out successfully. The Ramachandran plot was checked for accuracy of minimisation. The prepared PDB complex which was used as input for rosetta software is shown in the Figure 3.
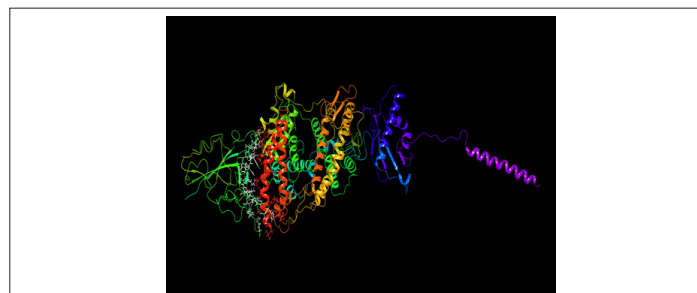

Figure 3: Completely preprocessed and minimized complex structure which was used as input for Rosetta.

Rosetta derived 10 amino acid sequence peptides

The software gives the output information of the peptide derived from a particular chain pair and of a certain length. The Receptor E (RBD of SARS-CoV-2), and Partner B (10 aa derived peptide residue) in PDB format is the output. The Partner B is derived from that sequence of ACE2 which interacts the most with RBD of SARS-CoV-2. Peptide derived is from position 9-18 i.e. residues 29-38 of ACE2, which perfectly match with the active site residues of human ACE2 reported in the literature. The peptide derived is shown in the Figure 4. This is indicated in the Peptide score versus window position diagram in Figure 5. The best linear peptide obtained was LDKFNHEAED, with interface score -6.740 and relative interface score percentage to be 35.17.
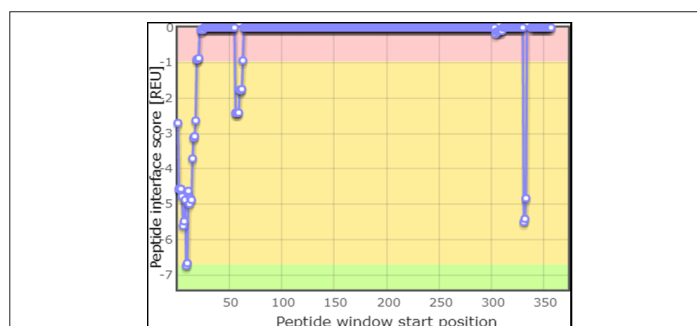

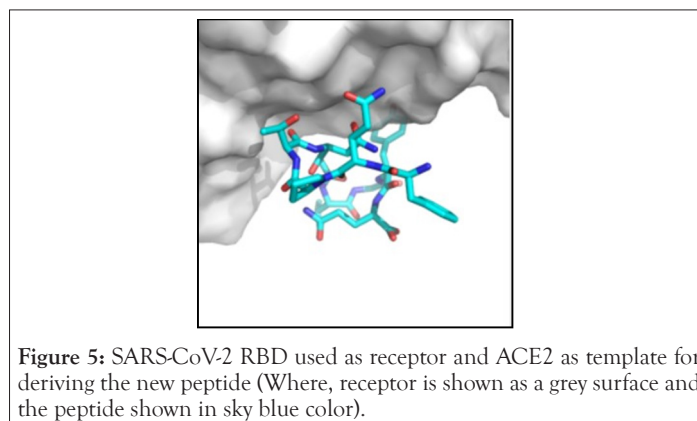**Figure 4:** Peptide score versus window position diagram.

Figure 5: SARS-CoV-2 RBD used as receptor and ACE2 as template for deriving the new peptide (Where, receptor is shown as a grey surface and the peptide shown in sky blue color).

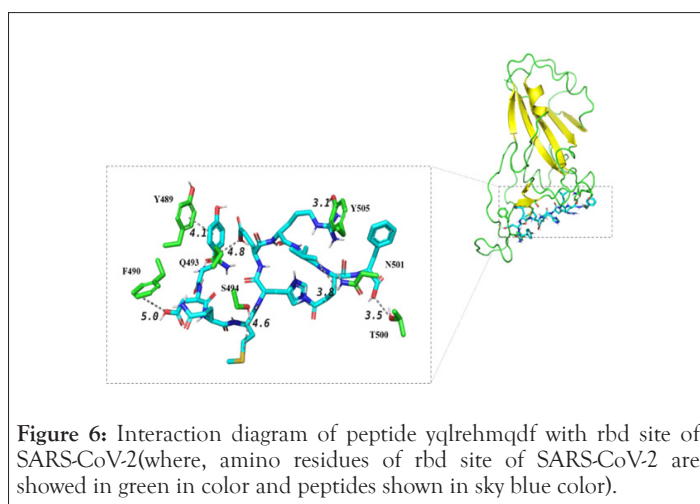The total interface score was found to be -19.16 REU

AutoDock Vina results

We intended to modify the sequence of peptide output from Rosetta with the view of increasing the binding interaction and find the docking score using AutoDock Vina software for designing and development of potential peptide inhibitors for SARS-CoV-2. The list of peptides can be seen in Table 1.

Table 1: Binding energy of docked peptides.

Sr. no	Sequence length	Peptide sequence	Binding energy in kcal/mol
1	10	LDKFNHEAED	-4.7
2	10	QLRWEHYMQD	-4.9
3	10	WQLREHMQDF	-5.1
4	10	YQLREHMQDF	-5.5
5	12	WQLRKFHEMDSF	-5.1
6	12	WQLRKFHEMDSY	-5.1

Preprocessed and minimised pdbid 6m17 of the SARS-CoV-2 and ace2 complex was modified to remove ace2 and sodium ion channel and used to dock 10 amino acid long peptide using autodock vina software. Extended confirmation of LDKFNHEAED binds to the rbd site of SARS-CoV-2 but binding energy was found to be 4.7 kcal/mol. Therefore, to increase binding energy as well as for more interaction between peptide and spike protein of SARS-CoV-2. We modified peptide sequence (knowledge based binding site amino acid analysis) QLRWEHYMQD, WQLREHMQDF, YQLREHMQDF are the new designed peptides containing 2 aromatic amino acids. These peptides were docked against modified spike protein pdbid (6 m17) using auto dock vina software. Among them YQLREHMQDF showed best fit in rbd pocket of SARS-CoV-2 with binding good energy. Interaction diagram of peptide YQLREHMQDF with rbd site of SARS-CoV-2 is shown in the Figure 6.

Figure 6: Interaction diagram of peptide yqlrehmqdf with rbd site of SARS-CoV-2(where, amino residues of rbd site of SARS-CoV-2 are showed in green in color and peptides shown in sky blue color).

Amino residues of rbd site of SARS-CoV-2 are showed in green color and peptide is shown in sky blue color. Inset shows extended conformation of ligand and its interactions with surrounding amino acids of the s1 protein. Ligand showed good pi-pi stacking interaction with y489. Y405, t500 and n501 showed hydrogen bonding with ligand. But, surface diagram showed that ligand had no access to the different pockets and as well as interaction with f486 was absent. Surface diagram of peptide YQLREHMQDF with rbd site of SARS-CoV-2 is shown in the Figure 7.
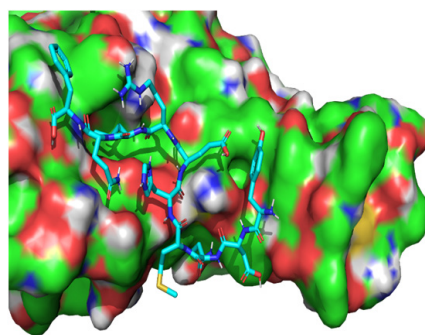

Figure 7: Surface diagram of peptide yqlrehmqdf with rbd site of SARS-CoV-2.

Therefore, we again modified and designed (Knowledge based binding site amino acid analysis) longer peptide of 12 amino acid sequence which can bind with more extended form and could access more pockets. WQLRKFHEMDSF was the new designed 12 amino acid peptide docked against spike protein of COVID-19 (6 m17) using autodock vina software. Interaction diagram of peptide WQLRKFHEMDSF with rbd site of SARS-CoV-2 is shown in the Figure 8.
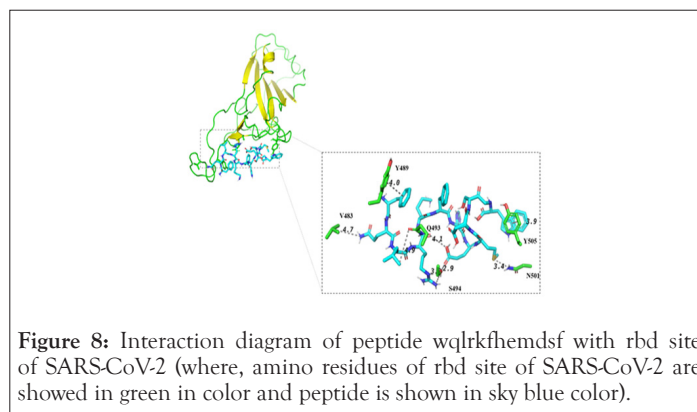

Figure 8: Interaction diagram of peptide wqlrkfhemdsf with rbd site of SARS-CoV-2 (where, amino residues of rbd site of SARS-CoV-2 are showed in green in color and peptide is shown in sky blue color).

Inset WQLRKFHEMDSF showed extended conformation of ligand which has binding energy of 5.1 kcal/mol. This result showed new interaction with 6 m17. It showed pi-pi stacking interaction with y489 and y505. As well as s494 showed hydrogen bond network with the ligand. Apart from this n501, v483 and q493 showed hydrogen bonding with ligand. The surface representation showed that it has more extended conformation compare to YQLREHMQDF as well it accesses more pockets of protein. Charge distribution showed that area of negatively charged residue of the protein has interaction with positive charged amino acid of the ligand. But this peptide also did not show interaction with f486 as well it didn't show interaction with all pockets. Surface diagram of peptide WQLRKFHEMDSF with rbd site of SARS-CoV-2 is shown in Figure 9.
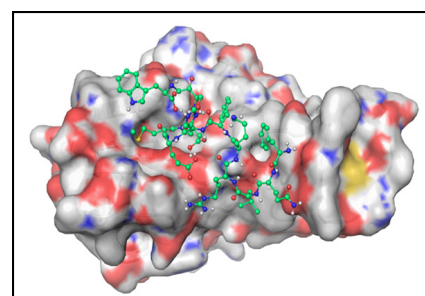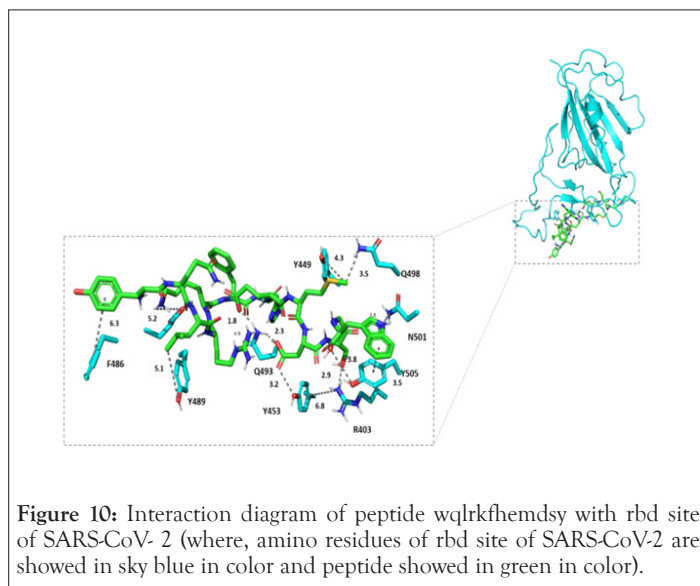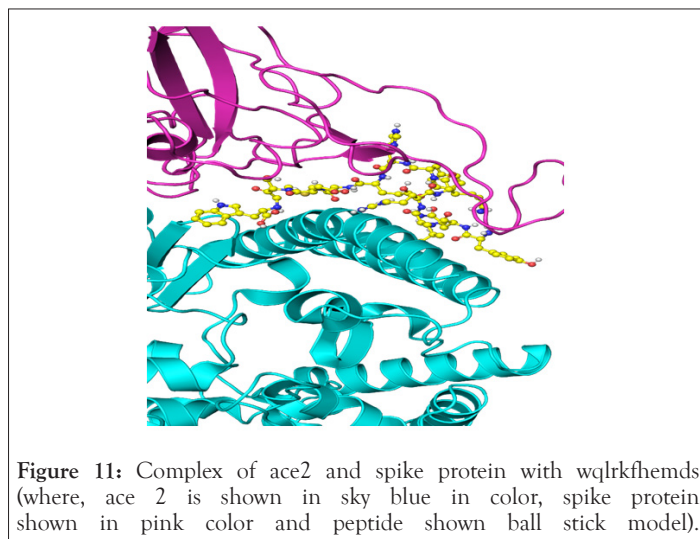

Figure 9: Surface diagram of peptide wqlrkfhemdsf with rbd site of SARS-CoV-2.

Therefore, n terminus phenylalanine is replaced with tyr. WQLRKFHEMDSY peptide was docked using autodock vina software. It showed binding of the ligand in extended conformation and binding energy was found to be of -5.1 kcal/mol. Interaction diagram of peptide WQLRKFHEMDSY with rbd site of SARS-CoV-2 is shown in the Figure 10.

Figure 10: Interaction diagram of peptide wqlrkfhemdsy with rbd site of SARS-CoV- 2 (where, amino residues of rbd site of SARS-CoV-2 are showed in sky blue in color and peptide showed in green in color).

WQLRKFHEMDSY showed more interaction than other two peptides. Y449, y453, y489 and f486 showed hydrophobic interaction with peptide whereas y505 showed stacking interaction with tryptophan of the peptide. Interestingly residue n501 also showed cation-pi interaction with tryptophan. Q493 showed hydrogen bonding network with aspartate and glutamate of the peptide. Residue y453 was also involved in hydrogen bonding with aspartate. Serine from the peptide inhibitor showed hydrogen bonding with r403. Due to more interaction with target amino acid residue WQLRKFHEMDSY inhibits the binding of spike protein to ace2 receptor as shown in Figure 11.

Figure 11: Complex of ace2 and spike protein with wqlrkfhemds (where, ace 2 is shown in sky blue in color, spike protein shown in pink color and peptide shown ball stick model).

The surface representation showed that peptide has binding interaction with all pockets. Charge distribution showed that negatively charged residues of the protein are stabilised by positive charged residues of the peptide. Surface diagram of peptide WQLRKFHEMDSY with rbd site of SARS-CoV-2 is shown in the Figure 12.

Figure 12: Surface diagram of peptide WQLRKFHEMDSY with RBD site of SARS-CoV-2.
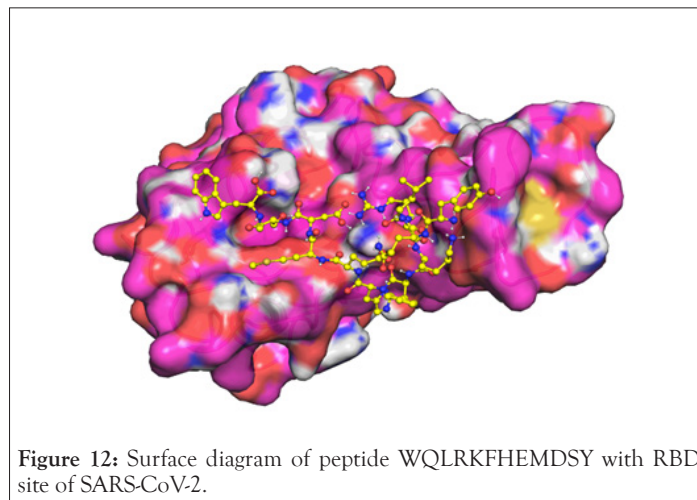
CONCLUSION

The peptide derived from the interface sequence interaction by the software claims to modify the Protein-protein interaction that occurs between SARS-CoV-2 and ACE2. The systematic modification of the basic peptide was performed. The chain length of the peptide was increased so that it has access to all the binding pocket areas of the RBM of SARS-CoV-2 and binds strongly with it thus interrupting its interaction with ACE2 receptor.

The binding energies and interaction of the peptide inhibitor with protein were analyzed. It was found that 12 amino acid residue long peptide inhibitor showed more interaction than 10 amino acid residue long peptide. Among 12 amino acid peptides, WQLRKFHEMDSY showed more interaction and extended conformation which can interact with all pockets present on the surface. Therefore, WQLRKFHEMDSY can be used as peptide inhibitor against spike protein of the SARS-CoV-2 with further design and formulation development. This modified peptide sequence can act as scaffold for development of specific peptidomimetics with good pharmacokinetic and pharmacodynamics profile. *In vitro* and *in vivo* studies can be performed using both 12 amino acid peptides because WQLRKFHEMDSF also showed good interaction. Thus, both WQLRKFHEMDSY and WQLRKFHEMDSF can be used for further investigational studies. The results will validate the *in-silico* analysis and pave a way for further modification.

CONFLICT OF INTEREST

All authors have contributed equally for the article. The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

REFERENCES

1. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. Nat Med. 2020; 26(4):450-452.

2. Adhikari SP, Meng S, Wu Y, Mao Y, Ye R, Wang Q, et al. A Literature review of 2019 novel coronavirus during the early outbreak period: epidemiology, causes, clinical manifestation and diagnosis, prevention and control. Infect Dis Poverty. 2020; 9(1):1-12.

3. WMCH, Wuhan, Municipal health and health commission's briefing on the current pneumonia epidemic situation in our city. 2020.

4. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in wuhan, china, of novel coronavirus infected pneumonia. N Engl J Med. 2020; 382(13):1199-1207.

5. CDC 2019 novel coronavirus, Wuhan, China. 2019.

6. WHO. Novel coronavirus-china. 2020.

7. Gorbalenya AE, Baker SC, Baric RS, De Groot RJ, Drosten C, Gulyaeva AA, et al. The species severe acute respiratory syndrome-related coronavirus: Classifying 2019-nCOV and naming it SARS-CoV-2. Nat Microbiol. 2020; 5(4):536-544.

8. 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; 181(2):281-292.e6.

9. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science. 2020; 367(6483):1260-1263.

10. Liu J, Cao R, Xu M, Wang X, Zhang H, Hu H, et al. Hydroxychloroquine, a less toxic derivative of chloroquine, is effective in inhibiting SARS-CoV-2 infection *in vitro*. Cell Discov. 2020; 6(1):1-4.

11. Bagdonaite I, Wandall HH. Global aspects of viral glycosylation. Glycobiology. 2018; 28(7):443-467.

12. Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol. 2019; 17(3):181-192.

13. Almazán F, Sola I, Zuñiga S, Marquez-Jurado S, Morales L, Becares M, et al. Coronavirus reverse genetic systems: Infectious clones and replicons. Virus Res. 2014; 189:262-270.

14. Wan Y, Shang J, Graham R, Baric RS, Li F. receptor recognition by novel coronavirus from wuhan: An analysis based on decade-long structural studies of SARS. J Virol. 2020; 94(7):1-9.

15. Huang L. Novel peptide inhibitors of angiotensin-converting enzyme 2. J Biol Chem. 2003; 278(18):15532-15540.

16. Sedan Y, Marcu O, Lyskov S, Schueler-Furman O. Peptiderive server: Derive peptide inhibitors from protein-protein interactions. Nucleic Acids Res. 2016; 44(1):536-541.

17. London N, Raveh B, Movshovitz-Attias D, Schueler-Furman O. Can self-inhibitory peptides be derived from the interfaces of globular protein-protein interactions? Proteins: Struct Funct Bioinf. 2010; 78(15):3140-3149.

18. Lyskov S, Chou F, Conchúir S, Der B, Drew K, Kuroda D et al. Serverification of molecular modeling applications: The rosetta online server that includes everyone (ROSIE). PLoS ONE. 2013; 8(5):e63906.

19. Koes DR, Baumgartner MP, Camacho CJ. Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. J Chem Inf Model. 2013; 53(8):1893-1904.

20. Brooks BR. Auto dock vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J comp chem. 2009; 30:1545-1614.

21. Morris G, Huey R, Lindstrom W, Sanner M, Belew R, Goodsell D et al. Auto dock 4 and auto dock tools 4: Automated docking with selective receptor flexibility. J Comp Chem. 2009; 30(16):2785-2791.