

# A novel peptide analogue of spike glycoprotein shows antiviral properties against SARS-CoV-2: An *in silico* approach through molecular docking, molecular dynamics simulation and MM-PB/GBSA calculations

Kunal Dutta (✉ [Kunal\\_lifesc@mail.vidyasagar.ac.in](mailto:Kunal_lifesc@mail.vidyasagar.ac.in))

Vidyasagar University <https://orcid.org/0000-0002-0818-8787>

Ammar D. Elmezayen

Kadir Has University

Anas Alobaidi

Kadir Has University

Wei Zhu

Huazhong Agricultural University

Olga Morojava

I.N. Blokhina Nizhny Novgorod Research Institute of Epidemiology and Microbiology

Sergey Shityakov

University of Würzburg; China Medical University

Ibrahim Khalifa

Benha University

---

## Research Article

**Keywords:** SARS-CoV-2, spike glycoprotein, machine learning, supported vector machine, antiviral peptide, receptor binding domain, membrane protein

**Posted Date:** December 28th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-32796/v2>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

At the very beginning of the new decade, the COVID-19 pandemic has badly hit modern human societies. SARS-CoV-2, the causative agent of COVID-19 carries dozens of new mutations in its genome. Herein, we made an effort to find new antiviral peptides (AVPs) against SARS-CoV-2. Gladly, with the help of Machine Learning algorithms, and Supported Vector Machine, we have invented three new AVPs against the SARS-CoV-2. Antiviral peptides *viz.*, Seq12, Seq12m, and Seq13m can block the receptor binding domain (RBD) of the SARS-CoV-2, necessary for communication with the angiotensin-converting enzyme 2 (ACE2). In addition, these AVPs retain their antiviral properties, even after the insertion of dozens of new mutations (Rosetta, and FoldX based) in the RBD. Further, Seq12, and Seq12m showed negligible cytotoxicity. Besides, the binding free energy calculated using MM-PB/GBSA method is also in agreement with the molecular docking studies performed using HADDOCK. Furthermore, the molecular interactions between AVPs and the viral membrane protein (M) also showed a thermodynamically favorable interaction, suggesting it could eventually inhibit the viral re-packaging process. In conclusion, this study suggests AVPs *viz.*, Seq12, Seq12m, and Seq13m embrace importance as a potential anti-SARS-CoV-2 therapeutic. These AVPs could also aid virus diagnostic tools in the future.

## Introduction

At present, the entire world is facing challenges in handing the COVID-19 pandemic [1]. Reports are coming out from different countries which suggest repurposing known antiviral drugs against COVID-19 could be fruitful [2-4]. However, none of them have reached a final definitive treatment for COVID-19. Now, multiple countries are in the race to find a successful vaccine against COVID-19 [5]. But SARS-CoV-2, the causative agent of COVID-19, contains dozens of new mutations in its genome. And it is one of the main concerns for all anti-COVID-19 efforts [6]. Moreover, varieties of new regional strains are appearing in a very short time frame [7].

SARS-CoV-2 encodes a spike glycoprotein and for function, they form a trimer. The notorious entry of SARS-CoV-2 inside a human host is primarily mediated via a protein-protein interaction between angiotensin-converting enzyme 2 (ACE2), (host cell), and the receptor binding domain (RBD), located at the spike glycoprotein of the SARS-CoV-2. Therefore, RBD embraces importance as a potential pharmaceutical target.

Herein, we made an effort to find out new antiviral peptides against SARS-CoV-2. In such an effort we utilized Machine Learning and Supported Vector Machine for antiviral peptide predictions. Thereafter, we opt for residue resistant molecular docking, molecular dynamics simulations, and MM-PB/GBSA analysis to characterize the antiviral properties. Best of our knowledge, this is the first report of antiviral peptides against the SARS-CoV-2 derived from the RBD of the SARS-CoV-2.

## Materials And Methods

### *Data sampling and analysis*

Nucleotide sequences of the SARS-CoV-2 were obtained from NCBI viruses ( ). The complete genome sequence of the SARS-CoV-2 strain Wuhan-Hu-1 (GenBank Sequence Accession: MN908947) was used as the query sequence. The viral spike glycoprotein (S-Protein), and membrane protein (M-Protein), (QHD43416.1) were studied using PSI-BLAST. Homology models of the M- and S-Protein were built using i-TASSER and SWISS-MODEL respectively [8, 9] followed by structure validation using PROCHECK [10].

### *Motif search*

The aligned sequences of spike protein resulted from the PSI-BLAST were used to discover motifs using the MEME suite [11, 12]. The primary query included a set of 69 protein sequences, between 16 to 1196 in length (Average length 214.3 amino acids), (Supplementary files) with the following settings: Background: A order-0 background generated from the supplied sequences. Discovery Mode: Classic: optimize the E-value of the motif information content. Site Distribution: Zero or one occurrence (of a contributing motif site) per sequence. Motif Count: for 3 motifs. Motif Width: Between 15 wide and 47 wide (inclusive).

### *Antiviral property prediction*

Identified motifs were checked and validated as antiviral peptide using a Machine Learning algorithm (MLA), and Supported Vector Machine (SVM) *i.e.* Meta-iAVP [13], and AVPred [14] with four prediction model settings *viz.*, AVP motif, sequence alignment, composition analysis, physio-chemical properties (threshold value of 50). The theoretical half-maximum inhibitory concentration (IC<sub>50</sub>) value was calculated using AVP-IC<sub>50</sub>Pred [15].

### *Homology modeling and peptide characterizations*

The predicted antiviral peptide sequences were used to build three-dimensional structures using the PEP-FOLD (v3.5), [16]. The structures obtained were then subject to energy minimization (GROMACS, v5.0) to remove unusual torsions and clashes from their structural geometry [17]. Properties of the predicted antiviral peptides were calculated using Innovagen peptide property calculator [18], ProtParam [19], AlgPred (Mapping of IgE epitopes and PID, MEME/MAST motif, Blast search on allergen representative peptides (ARPs)), [20], ToxinPred (SVM (TrEMBL) + Motif based and SVM (Swiss-Prot) based), [21], HemoPI, (SVM + Motif (HemoPI-2) based), IL4pred (Hybrid (SVM + motif) based ), [22]. AVP-IC<sub>50</sub>Pred (Hybrid Model Features a : Composition (mono-di) + Physico + Secondary structure + Surface accessibility and Features b: Binary (N8/C8) + Physico + Secondary structure + Surface accessibility) was used by selecting a Machine Learning Techniques: Supported Vector Machine (SVMlight) Random Forest (R package), IBk (Weka), KStar (Weka), [15].

### *In silico mutant model of the receptor binding domain*

Crystal structures of the receptor-binding domain (6W41) of the SARS-CoV-2 was obtained from RCSB-PDB [23]. We manually select Chain C of the 6W41 as it is annotated as RBD for ACE2. The *in silico* multi-point mutant models were build using FireProt [24]. The combined mutant model was further refined by manually (based on FoldX, and Rosetta). A detailed structural analysis of the mutant models was carried out using DALI [25], and UCSF Chimera [26].

### *Molecular docking*

HADDOCK-v2.2 was used for molecular docking between AVPs and receptor-binding domain (RBD) of the viral spike glycoprotein, viral membrane protein (M), RNA-dependent-RNA-polymerase (RdRp) [27]. The molecular docking poses were visualized using the previous method [28].

### *Molecular dynamics (MD) simulation*

MD-simulation studies were performed using GROMACS (v2020) with following modifications [29]. The MD-simulation productions of the antiviral peptides were carried out for 100 ns [30]. A detailed method of this MD-simulation using GROMACS (v2020) is described in the supplementary section. Furthermore, MD-simulation studies of the antiviral peptides complex with the wild and mutant type RBD were performed using NAMD software [2]. Input files were generated using the web-based server CHARMM-GUI (<http://www.charmm-gui.org/>). All systems were neutralized with KCl at 0.15 Molar, using the Monte-Carlo ion placing method. Systems were solvated using TIP3 water model, and CHARMM36m force field was used to assign charges. The Energy minimization was performed for 20,000 steps by steepest descent method followed by restrained 5 ns-equilibration at NVT ensemble and unrestrained 100 ns-production at 310 K. RMSD, RMSF, Radius of gyration (Rg), H-bonds, and SASA were analyzed using VMD. MM-PBSA calculations were performed using the Calculation of Free Energy (CaFE) tools [31]. Last 10 ns (the most stable region) were selected from each system and subjected for the MM-PBSA calculations, Internal dielectrostatic was set to 4.0, and the external dielectrostatic was set to 80.0. The Reciprocal of grid spacing was set to 2.0. In addition, for MM-GBSA calculations all-atom MD simulations were performed using the AMBER 16 package with the FF99SB for the protein and peptide molecules [32]. The protein-peptide systems were solvated with the TIP3P water models and neutralized with Na<sup>+</sup> and Cl<sup>-</sup> ions using the tLEaP input script available from the AmberTools. Long-range electrostatic interactions were applied via the particle-mesh Ewald method [33]. The SHAKE algorithm was used to constrain the length of covalent bonds, including hydrogen atoms [34]. The Langevin thermostat was implemented to equilibrate the temperature of the systems at 300 K. A 2.0 fs time step was used for all simulations. The minimization and equilibration protocols were run for 100,000 steps and 200 ps. Finally, 100 ns of classical MD simulations with no constraints were performed for each of the protein-peptide complexes using the MM-GBSA approach. According to the standard protocol, we used the explicit solvation model for all MD simulations and later the implicit solvation as a postprocessing end-state method to calculate binding free energies of molecules in solution utilizing the Python script. Additionally, HawkDock was also used for MM-GBSA calculations per amino acid residues [35]. The results were analyzed using GraphPad Prism 6 (San Diego, CA, USA).

## Characterization of immunogenic properties

Antiviral peptides were checked for their immunogenic properties using the IEDB epitope analysis tool which includes sequential B-cell epitopes [36], T-cell (MHC-II), NetCTL-1.2 [37] and, mapped within the predicted epitopes.

## Results And Discussion

SARS-CoV-2 is a novel coronavirus strain and it is new to the scientific communities. Therefore, in-depth monitoring of all aspects of the COVID-19 has been started. For example, signs and symptoms [38], mode of transmission [39], WHO-solidarity trials [40], contact tracing by mobiles apps such as “Arogya Setu” by India [41], CRISPR-Cas based rapid diagnostic of SARS-CoV-2 [42], and also monitoring daily cases by crowd-sourcing from <https://www.covid19india.org/>. Moreover, several reports are coming out as preprints and formal peer-review publications about repurposing known drugs [2] and antiviral drugs against the SARS-CoV-2 [43]. Reports suggesting possible anti-COVID-19 effects of different phytochemicals by *in silico* screening methods [4]. Indeed to address the urgent need for a safe and efficacious vaccine against the SARS-CoV-2 several vibrant initiatives have been started as never before. For example, vaccine manufacturing front-runner come-up with mRNA vaccines [44], viral vector vaccine[45], classical attenuated vaccine *etc.* [46]. However, reports show COVID-19 vaccines are not going to be a silver bullet for the immunization of a community. And, for the long-run alternatives to the traditional therapeutics would be necessary as before [47, 48].

An antiviral peptide is one of such new alternatives [49]. And, antiviral peptides are successful to combat the SARS-CoV, and MERS-CoV [50, 51]. Therefore, with a similar kind of anticipation for the SARS-CoV-2, we have invented three new antiviral peptides against the SARS-CoV-2 (Figure 1 a-c). Initially, the peptide sequences were interpreted as an antiviral peptide using AVPred [14], and Meta-iAVP [13]. The amino acid sequences of the predicted antiviral peptides were then mapped within the spike glycoprotein of the SARS-CoV-2. Figure 1d shows that antiviral peptides namely, Seq12, Seq12m, and Seq13m are in fact analogous peptides of the spike glycoprotein. Fasta sequences of the antiviral peptides are available in the supplementary section, Figure S1.

AVPred is an antiviral peptide prediction server that is based on a few sequence features *viz.*, motifs, alignment, amino acid composition, and physicochemical properties [14]. Finally, the prediction of the antiviral peptide is made during 5-fold cross-validation using Supported Vector Machine (SVM). Antiviral peptides were predicted by the AVPred (based on the physicochemical model) can achieve up to 85% prediction accuracy with 0.70 Matthew's Correlation Coefficient (MCC). However, the experimental validation dataset shows 86% prediction accuracy with 0.71 MCC.

Conversely, Meta-iAVP is based on a novel sequence-based meta-predictor with an effective features representation derived from Machine Learning Algorithms (MLA) and MLA types of features [13]. Interestingly, the use of MLA and MLA types features has increased the overall prediction accuracy, and

MCC of 95.20%, and 0.90 respectively. At present more than 15 peptide-based drugs are in the pipeline of clinical trials [13].

The most worrying concern of an antiviral peptide is its immunogenic profile [52]. A low immunogenic profile is the desired characteristic of an antiviral peptide because a low immunogenic profile reduces the chances of elimination by the host defense system [53]. Therefore, we also investigate the immunogenic profile of the AVPs. Hopefully, none of the AVPs showed immunogenic properties to humans as a host. Furthermore, to intensify the immunogenic profile of the AVPs we have performed epitope prediction using IEDB tools. Soon after, epitope prediction, AVP sequences were mapped within the aligned sequence of the epitopes (only most frequently human alleles were chosen). Results showed that none of the epitopes have significant similarities with the AVPs (Figure S2). A few amino acid residues were apparently similar to the predicted epitopes. However, the binding postures (AVP-RBD) suggest these apparently similar amino acid residues were engaged with the interactions (AVP-RBD). Therefore, the probability of elimination by the host defense system is not considerable.

Furthermore, the AVPs in this present study are capable of inducing an anti-inflammatory cytokine, IL-4 [54]. But, they are not capable to induce IL-10 (a pro-inflammatory cytokine). In addition, AVPs were also non-allergenic as they do not have any known epitope for human IgE [20]. Other physicochemical properties *viz.*, estimated half-life, instability index, water-solubility, theoretical IC<sub>50</sub> values, *etc.* were also summarized in Table 1.

Table 1. Physicochemical and biological properties of the antiviral peptides

Properties	Antiviral peptides		
	Seq12	Seq12m	Seq13m
APVpred <sup>1-5</sup>	AVP <sup>3,5</sup>	AVP <sup>3,4,5</sup>	AVP <sup>4,5</sup>
Meta-iAVP <sup>6</sup>	0.716	0.752	0.85
PROB-HemoPI <sup>7</sup>	0.17	0.17	0.46
Hydrophobicity	-0.23	-0.21	-0.10
Steric hindrance	0.69	0.67	0.65
Solvation	0.38	0.39	0.65
Hydropathicity	-0.78	-0.70	-0.25
Amphiphilicity	0.72	0.62	0.73
Hydrophilicity	0.09	0.03	-0.20
Net hydrogen	47.0	46.0	30.0
Charge (pH = 7)	1.1	0.0	-1.0
Isoelectric point (pI)	8.71	7.10	4.79
Molecular weight <sup>8</sup>	5302.68	5374.75	4657.83
Instability index <sup>9</sup>	37.74 s	46.15 u	53.32 u
Aliphatic index	72.77	72.77	76.10
GRAVY <sup>10</sup>	-0.777	-0.711	-0.0246
Estimated half life <sup>11</sup>	0.8 h	0.8 h	1 h
Water solubility	Good	Good	Poor
Epitope for IgE <sup>12</sup>	Non-allergen\$	Non-allergen	Non-allergen
IL-4 inducer <sup>13</sup>	Non-inducer*	Non-inducer*	Non-Inducer*
IL-10 inducer <sup>14</sup>	Inducer\$\$	Inducer\$\$	Inducer\$\$
IC <sub>50</sub>	39.38 uM	39.31 uM	38.54 uM

1 Hemolytic prediction = PROB score HemoPI was calculated using HemoPI-1/1+motif, (SVM+Motif (HemoPI-2). Mapping of IgE epitopes and PID, MEME/MAST motif, Blast search on allergen representative peptides (ARPs). Half life = mammalian reticulocytes, in vitro. IL-4 inducer prediction are made using SMV method (SMV Threshold = 0.5). \*However, the complete antiviral sequence contain two motif to induce the IL-4 and hydride method (SMV+motif ) score is 1.04, suggesting it is IL-4 inducer peptide. Instability index<sup>9</sup> : s = stable, u = unstable. GRAVY = Grand average of hydropathy.## The half maximal inhibitory concentration (IC<sub>50</sub>) = It a theoretical value calculated using IC<sub>50</sub>Pred. 11 Molecular weight in dalton. IL-10 inducing probability was calculated using RANDOM FOREST probability threshold = 0.5, and SMV method. AVPpred = , Meta-iAVP = .  
<sup>2</sup>Motif, <sup>3</sup>Align, <sup>4</sup>Composition, <sup>5</sup>Physicochemical composition

Recent work by Karoyan et al. [55] suggests hACE2 peptide mimics can block SARS-CoV-2 pulmonary cells infection with high affinity namely by P8, P9 and P10. In addition, we have found that the predicted IC<sub>50</sub> [15] value of those high affinity peptides are 41.38 μM , 41.8 μM, and 42.06 μM respectively by P8, P9 and P10. (Supplementary file). However, anti-SARS-CoV-2 peptides viz., Seq12, Seq12m and Seq13m in this present study have predicted IC<sub>50</sub> of only 39.38 μM, 39.31 μM, and 38.54 μM (Table 1) suggesting it might be promising too. Furthermore, other distinguished scientists have found anti-viral peptides from a

variety of a different source such as, Fibronectin by Beddingfield et al. [56], Aprotinin by Bestle et al. [57], Glycopeptide antibiotic by Zhang et al. [58], mouse  $\beta$ -defensin-4 by Zhao et al. [59], PCSK target motif by Cheng et al.[60], HR2 by Xia et al. [61, 62], HR2 by Zhu et al.[63], heptad repeats 1 and 2 (HR1 and HR2) in the S protein by [64].

However, in this present study antiviral peptides were derived from the spike glycoprotein of the SARS-CoV-2. Besides, previous formal peer-reviewed works have failed to demonstrate how the studied peptides are effective after insertions of mutations in the RBD of the SARS-CoV-2. Here comes the uniqueness of the present study. In this present study molecular docking ( ), MM-PB/GBSA calculations ( ) suggest that the AVPs namely, Seq12, Seq12m, and Seq13m can sustain their properties with the mutant RBD model of the SARS-CoV-2. A recent study on the classification of the circulating strains of novel coronavirus 2019 suggests there are A, B, and C types of coronavirus. A-type coronavirus is ancient/old, B-type is a bit new meaning that it has acquired dozens of new mutations and C-type is the current type, more contagious and has more mutations [65]. The results depicted from our present study suggest Seq12, Seq12m, and Seq13m might be effective to type A, B and C as well.

Antiviral peptides usually rupture the viral capsid, and eventually inhibit the viral replication cycles [66]. As mentioned earlier, the interactions between RBD-ACE2 is essential for viral entry inside the human host [67]. And recently, a research group led by Prof. Qiwei Zhang at the School of Public Health, Southern Medical University, China, has identified the molecular interactions between RBD-ACE2 [68]. Results suggest RBD includes six aromatic amino acid residues *viz.*, TYR449, PHE456, PHE486, TYR489, TYR505, and TYR543. Five polar uncharged amino acid residues ASN487, GLN493, GLN498, THR500, and ASN501. And three non-polar aliphatic amino acid residues, LEU455, GLY496, and GLY502.

We utilized this information in molecular docking studies to zoom-in on the molecular interactions of the AVPs-RBD complexes. However, there is an increasing amount of fear of mutations in the RBD, that would probably make all anti-COVID-19 efforts nil [69]. Indeed, it is very true that if mutations incorporated into the RBD then therapeutics have to evolve accordingly. Therefore, we have constructed a few *in silico* mutant models of the receptor-binding domain (RBDm) of the SARS-CoV-2, such as energy mutant, evolutionary mutant, and combined mutant (Figure 2). The amino acid substitutions were incorporated based on Rosetta and FoldX [24]. And the results are summarized in Table 2. The combined mutant model suggested by the FireProt was further refined by the manual incorporations of amino acid residues based on the best energy substitutions suggested by FoldX, and Rosetta [24]. This combined approach has led to an increase in the free energy of the RBDm from -27.63 kcal/mol (17 mutations) to -32.67 kcal/mol. And among the important amino acid residues of the RBD [67] only ASN487 is conserved (Figure S3).

Table 2. *In silico* mutant models of the receptor binding domain of the SARS-CoV-2

Combined mutant -27.63 kcal/mol (17 mutations)	Amino acid residues (Wild type to mutant type)	FoldX (kcal/mol)	Rosetta (kcal/mol)
	A348P	-1.66	-2.34
	N354E	-0.45	-
	A372T	0.20	-
	S373M	-2.30	-2.82
	T393F	-1.41	-6.03
	N394S	-0.63	-
	S399M	-3.17	-3.09
	R403K	0.48	-
	K417V	0.21	-
	N460K	-0.75	-
	I468L	-0.40	-
	T470Y	-2.10	-2.19
	S477G	-0.03	-
	S494Y	-1.47	-2.12
	G502P	-1.94	-2.61
	V503P	-1.18	-2.72
	H519N	-0.02	-
Energy mutant: -25.41 kcal/mol (8 mutations)	A348P	-1.66	-2.34
	S373M	-2.30	-2.82
	T393F	-1.41	-6.03
	S399M	-3.17	-3.09
	T470Y	-2.10	-2.19
	S494Y	-1.47	-2.12
	G502P	-1.94	-2.61
	V503P	-1.18	-2.72
Evolution mutant: -6.29 kcal/mol (12 mutations)	A348P	-1.66	-1.66
	N354E	-0.45	-0.45
	A372T	0.20	0.20
	T393S	0.47	0.47
	N394S	-0.63	-0.63
	R403K	0.48	0.48
	K417V	0.21	0.21
	N460K	-0.75	-0.75
	I468L	-0.40	-0.40
	T470N	0.06	-0.06
	S477G	-0.03	-0.03
	H519N	-0.02	-0.02

The structural analysis of the receptor-binding domains (RBD) of the SARS-CoV-2 has been summarized in Figure 2. The RRDIs Map analysis shows that RBDm and RBD can be overlapped. Moreover, the RRDIs Map showed that RBD and RBDm had an average distance of 51.74 (SD:1.486, a secreted site of the

map), suggesting they are structurally different (Figure S5). Moreover, the RBD<sub>b/c/e</sub> structurally distinguished from RBD<sub>m</sub>/RBD and they can not be overlapped with either RBD or RBD<sub>m</sub>. The RRDis map showed that the average distance of the RBD<sub>b/c/e</sub> was 28.884 (SD:0.032, a secreted site of the map), suggesting they are structurally close to each other. In addition, the correspondence analysis also suggests RBD<sub>b/c/e</sub> are structural neighbors and, RBD and RBD<sub>m</sub> are structurally distinguished from RBD<sub>b/c/e</sub>. The structural tree (Supplementary Figure S4) showed that RBD<sub>b</sub>, RBD<sub>c</sub>, and RBD<sub>e</sub> are orthologous. Conversely, RBD/RBD<sub>m</sub> are paralogous. Moreover, a comparison of all four mutant types of the RBD by selecting the wild type RBD as a reference, revealed that all models have exactly the same RMSD value of 1.7 except RBD<sub>e</sub> RMSD = 1.5). Besides, RBD<sub>c</sub> was 96% identical with the wild type RBD followed by RBD<sub>m</sub> (94%), RBD<sub>e</sub> (88%), and RBD<sub>b</sub> (84%) and the Dali Z-scores were 25.3, 25.3, 25.2, and 23.9 respectively for RBD<sub>m</sub>, RBD<sub>c</sub>, and RBD<sub>e</sub>, and RBD<sub>b</sub> (Table 3).

Table 3. Structural analysis of different mutant model of the receptor binding domain (RBD) of the SARS-CoV-2

Sl. No.	Mutant models of RBD	RMSD <sup>a</sup>	Percentage of identities <sup>b</sup>	Z-score*
1.	RBD <sub>m</sub>	1.7	94	25.3
2.	RBD <sub>c</sub>	1.7	96	25.3
3.	RBD <sub>e</sub>	1.5	88	25.2
4.	RBD <sub>b</sub>	1.7	84	23.9

<sup>a</sup>RMSD = root mean square deviations, <sup>b</sup>Percentage of identities with the wild type RBD structure, RBD<sub>m</sub> = manual mutant, RBD<sub>c</sub> = combind model, RBD<sub>e</sub> = energy mutant, RBD<sub>b</sub> = evolutionary mutant model of the receptor binding domain. \*Dali Z-score.

Results from molecular docking studies showed that AVPs-RBD had a thermodynamically favorable interaction (Table 4). Moreover, the AVPs were engaged with nearly all-important amino acid residues of the RDB as mentioned in the earlier report [67]. Out of three antiviral peptides of this present study, Seq12 (47 mer) showed an astonishing HADDOCK score of -111.2 kcal/mol followed by Seq13m (81.4 kcal/mol), and Seq12m (76.8 kcal/mol) suggesting AVPs are well-docked with the RBD (Figure 3, S6). In addition, molecular docking studies between RBD<sub>m</sub> and the antiviral peptides (RBD<sub>m</sub>-AVP) summarized in Table 4. Results suggest Seq12, Seq12m, and Seq13m can retain the thermodynamically favorable binding properties with the mutant RBD. Moreover, Seq12 surprisingly showed an improved binding free energy (kcal/mol) in comparison to the wild type RBD (kcal/mol). Further, Seq12, and Seq12m have negligible cytotoxicity (0.17; while 1.0 or higher is considered as toxic and can rapture the human RBCs), (Table 1). These findings combinedly justify the novelty of the AVPs, and its possible therapeutic use in the future.

Table 4. HADDOCK docking properties of the antiviral peptides complex with the receptor binding domain of the SARS-Cov-2

Properties	Systems					
	Seq12 + RBD	Seq12 + RBDm <sup>3</sup>	Seq12m + RBD	Seq12m + RBDm	Seq13m + RBD	Seq13m + RBDm
HADDOCK Score	-111.2	-92.8	-76.8	-76.7	-81.4	-41.7
Cluster size	104	36	13	27	40	14
RMSD <sup>1</sup>	0.5	0.4	2.6	4.7	0.5	3.5
VDW <sup>2</sup> energy	-68.8	-71.3	-71.9	-78.3	-75.2	-63.4
Electrostatic energy	-182.4	-142.0	-145.0	-53.2	-70.3	-73.8
Desolvation energy	-36.9	-36.6	-4.9	-43.1	-20.9	-28.5

<sup>1</sup>RMSD = Root mean square deviation from the overall lowest-energy structure. <sup>2</sup>VDW = Van der Waals energy. <sup>3</sup>RBDm = mutant model of the receptor binding domain of the SARS-CoV-2. Energy units are in kcal/mol.

SARS-CoV-2 has a structural membrane protein called Protein-M [70]. Protein-M is a tri-pass transmembrane protein (Figure 4) which is essential for re-assembling viral structural units into a mature virus [71]. Therefore, Protein-M also embraces importance as a potential therapeutic target. Results from our investigation indicate AVPs are also participating in a thermodynamical favorable interactions with the TM-1 region followed by TM-2 and TM-3 respectively (Figure 4a). In addition, the molecular docking studies between the AVPs and the active site of the viral RNA-dependent RNA polymerase (RdRp) were not thermodynamically favorable interaction (Supplementary section). We speculate that the antiviral peptide Seq12, Seq12m, and Seq13m acts as an anti-SARS-CoV-2 peptide by two possible mechanisms. Firstly, by inhibiting the RBD-ACE2 interaction. And secondly, by binding with M-protein followed by an eventual inhibition of the viral re-assembly/re-packaging.

Results, derived from MD-simulation of the studies of the AVPs are summarised in Figure 5. The root mean square deviation (RMSD) of alpha carbon atoms of all systems are analyzed to detect their stabilities. It is observed from Figure 6 that Seq12m, has the lowest RMSD value than Seq12 and Seq13m respectively. Antiviral peptides namely Seq12, Se12m, and Seq13m have  $\approx 0.274$  nm,  $\approx 0.274$  nm, and  $\approx 0.286$  nm of RMSD values at 1 ns. The highest RMSD values are respectively  $\approx 0.852$  nm,  $\approx 0.908$  nm, and  $\approx 1.074$  nm for Seq12, Seq12m, and Seq13m. A closer look at the RMSD plot suggests Seq12m has the highest fluctuation from 5 ns - 10 ns and gets stabilized throughout the trajectory. However, at the start, RMSD-fluctuation is lower in the case of Seq12 and the highest RMSD-fluctuation is observed between 30 ns - 40 ns. Quite similarly, the RMSD-fluctuation of Seq13m is highest between 20 ns -40 ns. Root means square fluctuation (RMSF) helps to understand the flexibility of each amino acid residue [72]. Seq13m is found to have the highest RMSF from the 5<sup>th</sup> to 7<sup>th</sup> and 25<sup>th</sup> to 30<sup>th</sup> position of the amino acid sequence.

The lower degree of fluctuation with its consistency through the simulation indicates the greater compactness and rigidity of a system [72]. Rg of Seq13m is fluctuated from 10 ns -40 ns and reaches the

highest value of 1.331 at  $\approx 25$  ns. However, Rg values of Seq12 and Seq12m are very similar at the beginning and remain same up to 10 ns. Further, Rg fluctuations are prominent after 10 ns and sustain such dissimilarities throughout the trajectory. The lowest Rg values are  $\approx 1.041$  nm (50 ns),  $\approx 1.380$  nm (before 10 ns), and  $\approx 0.944$  nm (before 10 ns) respectively for Seq12, Seq12m, and Seq13m. In summary systems *i.e.*, Seq12 and Seq12m showed better stability throughout the complete trajectory. The number of inter water-peptide hydrogen bonds in the simulated systems were also compared. The highest number of hydrogen bonds are formed respectively by Seq12 (40 - 45 ns), Seq12m (15 ns), and Seq13m (before 5 ns).

In addition, MD-simulation studies of the AVP-RBD complex systems (antiviral peptide + RBD/RBDm) of the 100 ns MD production for prior to the MM-PBSA analyses were summarised in Figure 6. RMSD, and RMSF were stable throughout all MD-trajectory and were most stable in the case of Seq12-RBD/RBDm, and Seq12m-RBD/RBDm complex. However, RMSD values fluctuates in the case of Seq13m-RBD/RBDm complex but these values were within acceptable range of 3nm suggesting extension of the simulation time scale is not required. RMSF value is a measure of the average deviations particular atoms or group of atoms from the initial reference structure [73]. Results showed that, in case of the Seq13m+RBDm, RMSF fluctuation was high from about 138<sup>th</sup> -158<sup>th</sup> position and last twenty five residues of the RBD. Similarly, in the case of Seq13m+RBD the RMSF value highly fluctuates from about 200<sup>th</sup> to rest of the residues of the RBD. However, only two important interacting amino acid residues namely, PHE486 and TYR489 remain within the fluctuating regions (Figure 6b). Besides, the  $\Delta G_{\text{bind}}$  of Seq13m+RBDm and Seq13+RBD were also very good in comparison to the other RBD-AVP complex systems (Table 5) suggesting this little structural deviations does not have negative impacts on binding. The radius of gyration (Rg) around the axis were stable (1.16-1.47 nm) for all complex systems, Besides, Rg of the AVP-RBD complexes were overlapping in the case of all complex systems except Seq12m-RBD. In addition, the solvent accessible surface area (SASA) of the Seq13m-RBD/RBDm complex systems were highest (16425.50 nm<sup>2</sup>), and this value was lowest (13454.30 nm<sup>2</sup>) for the Seq12m-RBDm complex system. Moreover, the number of hydrogen-bonds formed during the MD-simulation run was limited in the case of inter AVP-RBD/RBDm complex (0-10). However, during the entire MD-simulation run more H-bonds (0-99) were formed between protein complex and water.

The results obtained from MM-PB/GBSA analysis are also in agreement with the molecular docking studies (Table 5). Moreover, the binding free energies (MM-GBSA) per amino acid residues (only the top twenty participating amino acid residues) are summarized in Figure 7. It shows that TYR489 was an important amino acid residue of the RBD because TYR489 was a common contributor to the lowest binding free energies of Seq12-RBD (-9.30 kcal/mol), Seq12m-RBD (-6.55kcal/mol), and Seq13m-RBD (-4.91kcal/mol). On the other hand, GLU24 (-9.20 kcal/mol) of Seq12, THR20 (-3.36kcal/mol) of Seq12m, and PRO25 (-6.56 kcal/mol) of Seq13m were critically important for the respective antiviral peptides for the same reasons. Further, GLN498 (-0.22 kcal/mol), and GLN493 (-0.28 kcal/mol) were poor contributors for the antiviral peptide-RBD (AVP-RBD) interactions respectively for Seq12 and Seq12m. However, GLY496 of the RBD poorly contributes to the Seq13m-RBD interaction.

Conversely, Seq12-RBDm, Seq12m-RBDm, and Seq13m-RBDm interactions reveal a different picture (Figure 7). For example, amino acid residues of mutant RBD namely, TYR501 (-6.32 kcal/mol), MET494 (-5.11 kcal/mol), and TYR-450 (-5.24 kcal/mol) contributes lowest binding free energy changes while interacting with Seq12, Seq12m, and Seq13m respectively. It is indeed, noticeable that in all cases (wild and mutant RBD) tyrosine contributes the lowest binding free energy irrespective of its positions in the RBD. The biochemical and biophysical properties of tyrosine residue enable itself for this phenomenon. However, tyrosine has not participated in the top twenty amino acid residues in the case of Seq12m-RBDm. On the contrary, LEU493 was a poor contributor to the binding free energies of Seq12-/ Seq12m-RBDm interactions. However, ASN449 was another poor contributor in the case of Seq13m-RBDm interaction. Overall, the molecular details of AVP-RBD/RBDm interactions calculated using MM-GBSA provide evidence for the fact that AVPs were occupied the important amino acid residues of the RBD which are critical for the RBD-ACE2 interactions [67].

In addition, the MM-PBSA analyses suggest Seq12m-RBDm has the best  $\Delta G_{\text{MMPBSA}}$  (-84.75 kcal/mol), followed by Seq13m-RBDm (-79.30 kcal/mol), and Seq12m-RBDm (-78.72 kcal/mol), *etc.* (Table 5). Moreover, the linear regression fit between predicted  $\Delta G_{\text{PB/GBSA}}$  calculated for 50 ns and predicted  $\Delta G_{\text{PB/GBSA}}$  calculated for 100 ns is summarized in figure 8. It showed that the  $R^2$  of the regression model of the  $\Delta G_{\text{GBSA}}$  was 0.97 ( $p = 0.0003$ ) suggesting  $\Delta G_{\text{bind}}$  are significantly correlated with the model. However, in the case of  $\Delta G_{\text{PBSA}}$  the  $R^2$  was 0.81 ( $p = 0.0149$ ), although all variables were within 95% confidence intervals. This is because we have found that the  $\Delta G_{\text{PBSA}}$  calculated for 50ns and 100ns have the highest difference in the case of Seq13m-RBD ( $\Delta G_{\text{PBSA}}$  calculated for 50 ns was -29.22 kcal/mol and  $\Delta G_{\text{PBSA}}$  calculated for 100 ns was -42.52 kcal/mol), (Table S1).

Table 5. Binding free energy ( $\Delta G_{\text{bind}}$ ) of different AVP-RBD complex systems calculated using MM-GB/PBSA

Systems	MM-GBSA*	MM-GBSA**	MM-PBSA#
Seq12 + RBD	-92.09	-45.37	-49.80
Seq12 + RBDm <sup>§</sup>	-93.31	-53.64	-84.75
Sqe12m + RBD	-50.09	-61.07	-61.84
Sqe12m + RBDm <sup>§</sup>	-75.13	-62.96	-78.72
Seq13m + RBD	-57.96	-40.71	-42.52
Seq13m + RBDm <sup>§</sup>	-53.90	-38.59	-79.30

\* MM-GBSA was calculated using HawkDock.

# MM-PBSA was calculated using CaFE tools with the MD-trajectory generated using NAMD software.

\*\* MM-GBSA was calculated using Abmer 16 package Generalized Born ESURF calculated using 'LCPO' surface areas.

§ Combined mutant model of receptor binding domain is manually optimized to best energy model (RBDm).

Energy units are in kcal/mol.

## Conclusion

In summary, peptide analogs of spike glycoprotein namely, Seq12, Seq12m, and Seq13m showed antiviral properties against the SARS-CoV-2. Molecular docking suggests Seq12, Seq12m, and Seq13m can inhibit RBD-ACE2 interactions. In addition, these antiviral peptides can retain their antiviral properties even after the insertion of dozens of new mutations in the RBD. Furthermore, the AVPs can interfere with viral membrane protein M, and it could eventually inhibit viral re-packaging cycles. MD-simulation studies of the AVP-RBD/RBDm complex of Seq12 and Seq12m showed stable RMSD and RMSF throughout the complete MD-trajectory suggesting a stable interaction. Moreover, the binding free energy calculations using MM-PB/GBSA are also in agreement with the molecular docking studies. Furthermore, AVPs namely, Seq12, and Seq12m showed negligible cytotoxicity and they are non-allergenic for humans. The AVPs in this present study might be a potential therapeutic against the SARS-CoV-2. However, more studies are required for clinical or diagnostic use.

## Declarations

### *Acknowledgements*

Council of Scientific and Industrial Research (CSIR), Govt. of India, New Delhi, India is also sincerely acknowledged by K.D. for Senior Research Fellowship (SRF), sanction letter no. 09/599(0082)/2019-EMR-I.

### *Declaration of interests*

The author declares no competing financial interests.

## References

1. Ting, D.S.W., et al., *Digital technology and COVID-19*. Nature medicine, 2020. **26**(4): p. 459-461.
2. Elmezayen, A.D., et al., *Drug repurposing for coronavirus (COVID-19): in silico screening of known drugs against coronavirus 3CL hydrolase and protease enzymes*. Journal of Biomolecular Structure and Dynamics, 2020(just-accepted): p. 1-12.
3. Dutta, K., et al., *Beclabuvir can inhibit the RNA-dependent RNA polymerase of newly emerged novel coronavirus (SARS-CoV-2)*. 2020.
4. Khalifa, I., et al., *Tannins inhibit SARS-CoV-2 through binding with catalytic dyad residues of 3CLpro: An in silico approach with 19 structural different hydrolysable tannins*. Journal of food biochemistry, 2020. **44**(10): p. e13432.
5. Prompetchara, E., C. Ketloy, and T. Palaga, *Immune responses in COVID-19 and potential vaccines: Lessons learned from SARS and MERS epidemic*. Asian Pac J Allergy Immunol, 2020. **38**(1): p. 1-9.
6. Robson, B., *COVID-19 Coronavirus spike protein analysis for synthetic vaccines, a peptidomimetic antagonist, and therapeutic drugs, and analysis of a proposed Achilles' heel conserved region to minimize probability of escape mutations and drug resistance*. Computers in Biology and Medicine, 2020: p. 103749.
7. Peeri, N.C., et al., *The SARS, MERS and novel coronavirus (COVID-19) epidemics, the newest and biggest global health threats: what lessons have we learned?* International journal of epidemiology, 2020.
8. Zhang, Y., *I-TASSER server for protein 3D structure prediction*. BMC bioinformatics, 2008. **9**(1): p. 40.
9. Guex, N. and M.C. Peitsch, *SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling*. electrophoresis, 1997. **18**(15): p. 2714-2723.
10. Laskowski, R.A., et al., *PROCHECK: a program to check the stereochemical quality of protein structures*. Journal of applied crystallography, 1993. **26**(2): p. 283-291.
11. Bailey, T.L., et al., *MEME SUITE: tools for motif discovery and searching*. Nucleic acids research, 2009. **37**(suppl\_2): p. W202-W208.
12. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs*. Nucleic acids research, 1997. **25**(17): p. 3389-3402.
13. Schaduangrat, N., et al., *Meta-iAVP: A sequence-based meta-predictor for improving the prediction of antiviral peptides using effective feature representation*. International journal of molecular sciences,

2019. **20**(22): p. 5743.
14. Thakur, N., A. Qureshi, and M. Kumar, *AVPpred: collection and prediction of highly effective antiviral peptides*. Nucleic acids research, 2012. **40**(W1): p. W199-W204.
  15. Qureshi, A., H. Tandon, and M. Kumar, *AVP-IC50Pred: Multiple machine learning techniques-based prediction of peptide antiviral activity in terms of half maximal inhibitory concentration (IC50)*. Peptide Science, 2015. **104**(6): p. 753-763.
  16. Lamiable, A., et al., *PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex*. Nucleic acids research, 2016. **44**(W1): p. W449-W454.
  17. Abraham, M.J., et al., *GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers*. SoftwareX, 2015. **1**: p. 19-25.
  18. Lear, S. and S.L. Cobb, *Pep-Calc. com: a set of web utilities for the calculation of peptide and peptoid properties and automatic mass spectral peak assignment*. Journal of computer-aided molecular design, 2016. **30**(3): p. 271-277.
  19. Portal, E.B.R., *ProtParam tool*, 2011.
  20. Saha, S. and G. Raghava, *AlgPred: prediction of allergenic proteins and mapping of IgE epitopes*. Nucleic acids research, 2006. **34**(suppl\_2): p. W202-W209.
  21. Gupta, S., et al., *In silico approach for predicting toxicity of peptides and proteins*. PloS one, 2013. **8**(9).
  22. Win, T.S., et al., *HemoPred: a web server for predicting the hemolytic activity of peptides*. Future medicinal chemistry, 2017. **9**(3): p. 275-291.
  23. Yuan, M., et al., *A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV*. Science, 2020. **368**(6491): p. 630-633.
  24. Musil, M., et al., *FireProt: web server for automated design of thermostable proteins*. Nucleic acids research, 2017. **45**(W1): p. W393-W399.
  25. Holm, L., *DALI and the persistence of protein shape*. Protein Science, 2020. **29**(1): p. 128-140.
  26. Pettersen, E.F., et al., *UCSF Chimera—a visualization system for exploratory research and analysis*. Journal of computational chemistry, 2004. **25**(13): p. 1605-1612.
  27. Rodrigues, J.P., E. Karaca, and A.M. Bonvin, *Information-driven structural modelling of protein–protein interactions*, in *Molecular modeling of proteins 2015*, Springer. p. 399-424.
  28. Dutta, K., et al., *Effects of secondary carbon supplement on biofilm-mediated biodegradation of naphthalene by mutated naphthalene 1, 2-dioxygenase encoded by Pseudomonas putida strain KD9*. Journal of hazardous materials, 2018. **357**: p. 187-197.
  29. Kadirvel, P., et al., *Molecular Dynamics Simulation Study of Plasmodium falciparum and Escherichia coli SufA: Exploration of Conformational Changes Possibly Involved in Iron-Sulfur Cluster Transfer*. Journal of Biomolecular Structure and Dynamics, 2020(just-accepted): p. 1-16.
  30. Dutta, K., et al., *Paradoxical Bactericidal Effects of Hydrophobic Lung Surfactant Proteins and Their Peptide Mimics Using Liposome Molecular Trojan*. Journal of oleo science, 2018: p. ess18026.

31. Liu, H. and T. Hou, *CaFE: a tool for binding affinity prediction using end-point free energy methods*. *Bioinformatics*, 2016. **32**(14): p. 2216-2218.
32. Case, D., et al., *AMBER 2016 (University of California)*. San Francisco, 2016.
33. Essmann, U., et al., *A smooth particle mesh Ewald method*. *The Journal of chemical physics*, 1995. **103**(19): p. 8577-8593.
34. Miyamoto, S. and P.A. Kollman, *Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models*. *Journal of computational chemistry*, 1992. **13**(8): p. 952-962.
35. Weng, G., et al., *HawkDock: a web server to predict and analyze the protein–protein complex based on computational docking and MM/GBSA*. *Nucleic acids research*, 2019. **47**(W1): p. W322-W330.
36. Vita, R., et al., *The immune epitope database (IEDB) 3.0*. *Nucleic acids research*, 2015. **43**(D1): p. D405-D412.
37. Larsen, M.V., et al., *Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction*. *BMC bioinformatics*, 2007. **8**(1): p. 424.
38. Deeks, J.J., et al., *Diagnosis of SARS-CoV-2 infection and COVID-19: accuracy of signs and symptoms; molecular, antigen, and antibody tests; and routine laboratory markers*. *Cochrane Database of Systematic Reviews*, 2020(4).
39. Klompas, M., M.A. Baker, and C. Rhee, *Airborne transmission of SARS-CoV-2: theoretical considerations and available evidence*. *Jama*, 2020.
40. Pan, H., et al., *Repurposed antiviral drugs for COVID-19; interim WHO SOLIDARITY trial results*. *MedRxiv*, 2020.
41. Kodali, P.B., et al., *How Indians responded to the Arogya Setu app?* *Indian Journal of Public Health*, 2020. **64**(6): p. 228.
42. Azhar, M., et al., *Rapid, field-deployable nucleobase detection and identification using FnCas9*. *bioRxiv*, 2020.
43. Dutta, K., et al., *Beclabuvir can inhibit the RNA-dependent RNA polymerase of newly emerged novel coronavirus (SARS-CoV-2)*. 2020.
44. Mahase, E., *Covid-19: Moderna vaccine is nearly 95% effective, trial involving high risk and elderly people shows*. *BMJ: British Medical Journal (Online)*, 2020. **371**.
45. Graham, S.P., et al., *Evaluation of the immunogenicity of prime-boost vaccination with the replication-deficient viral vectored COVID-19 vaccine candidate ChAdOx1 nCoV-19*. *NPJ vaccines*, 2020. **5**(1): p. 1-6.
46. Amanat, F. and F. Krammer, *SARS-CoV-2 vaccines: status report*. *Immunity*, 2020.
47. Boas, L.C.P.V., et al., *Antiviral peptides as promising therapeutic drugs*. *Cellular and Molecular Life Sciences*, 2019. **76**(18): p. 3525-3542.
48. Mahendran, A.S.K., et al., *The Potential of Antiviral Peptides as COVID-19 Therapeutics*. *Frontiers in pharmacology*, 2020. **11**: p. 1475.

49. Pant, S., et al., *Peptide-like and small-molecule inhibitors against Covid-19*. Journal of Biomolecular Structure and Dynamics, 2020(just-accepted): p. 1-15.
50. Du, L., et al., *MERS-CoV spike protein: a key target for antivirals*. Expert opinion on therapeutic targets, 2017. **21**(2): p. 131-143.
51. Zhao, H., et al., *A novel peptide with potent and broad-spectrum antiviral activities against multiple respiratory viruses*. Scientific reports, 2016. **6**: p. 22008.
52. Tourdot, S., et al., *Chimeric peptides: a new approach to enhancing the immunogenicity of peptides with low MHC class I affinity: application in antiviral vaccination*. The Journal of Immunology, 1997. **159**(5): p. 2391-2398.
53. Mahlapuu, M., C. Björn, and J. Ekblom, *Antimicrobial peptides as therapeutic agents: opportunities and challenges*. Critical reviews in biotechnology, 2020. **40**(7): p. 978-992.
54. Manavalan, B., et al., *AIPpred: sequence-based prediction of anti-inflammatory peptides using random forest*. Frontiers in pharmacology, 2018. **9**: p. 276.
55. Karoyan, P., et al., *hACE2 peptide mimics block SARS-CoV-2 Pulmonary Cells Infection*. bioRxiv, 2020.
56. Beddingfield, B.J., et al., *The integrin binding peptide, ATN-161, as a novel therapy for SARS-CoV-2 infection*. JACC: Basic to Translational Science, 2020.
57. Bestle, D., et al., *TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells*. Life science alliance, 2020. **3**(9).
58. Zhang, G., et al., *Investigation of ACE2 N-terminal fragments binding to SARS-CoV-2 Spike RBD*. bioRxiv, 2020.
59. Zhao, H., et al., *A broad-spectrum virus-and host-targeting peptide against respiratory viruses including influenza virus and SARS-CoV-2*. Nature communications, 2020. **11**(1): p. 1-10.
60. Hoffmann, M., H. Kleine-Weber, and S. Pöhlmann, *A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells*. Molecular Cell, 2020.
61. Xia, S., et al., *A pan-coronavirus fusion inhibitor targeting the HR1 domain of human coronavirus spike*. Science advances, 2019. **5**(4): p. eaav4580.
62. Xia, S., et al., *Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion*. Cell research, 2020. **30**(4): p. 343-355.
63. Zhu, Y., et al., *Design of potent membrane fusion inhibitors against SARS-CoV-2, an emerging coronavirus with high fusogenic activity*. Journal of Virology, 2020.
64. Ling, R., et al., *In silico design of antiviral peptides targeting the spike protein of SARS-CoV-2*. Peptides, 2020. **130**: p. 170328.
65. Forster, P., et al., *Phylogenetic network analysis of SARS-CoV-2 genomes*. Proceedings of the National Academy of Sciences, 2020. **117**(17): p. 9241-9243.
66. Steven, A.C., et al., *The making and breaking of symmetry in virus capsid assembly: glimpses of capsid biology from cryoelectron microscopy*. The FASEB journal, 1997. **11**(10): p. 733-742.

67. Yan, R., et al., *Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2*. Science, 2020. **367**(6485): p. 1444-1448.
68. Lan, J., et al., *Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor*. Nature, 2020: p. 1-6.
69. Abdullahi, I.N., et al., *Implications of SARS-CoV-2 genetic diversity and mutations on pathogenicity of COVID-19 and biomedical interventions*. Journal of Taibah University Medical Sciences, 2020.
70. Luan, J., et al., *Spike protein recognition of mammalian ACE2 predicts the host range and an optimized ACE2 for SARS-CoV-2 infection*. Biochemical and biophysical research communications, 2020.
71. Bianchi, M., et al., *Sars-CoV-2 Envelope and Membrane proteins: differences from closely related proteins linked to cross-species transmission?* Preprints, 2020.
72. Islam, R., et al., *A molecular modeling approach to identify effective antiviral phytochemicals against the main protease of SARS-CoV-2*. Journal of Biomolecular Structure and Dynamics, 2020(just-accepted): p. 1-20.
73. Martínez, L., *Automatic identification of mobile and rigid substructures in molecular dynamics simulations and fractional structural fluctuation analysis*. PloS one, 2015. **10**(3): p. e0119264.

## Supplementary Section

Detailed MD-simulation method.

Table S1. Binding free energy ( $\Delta G_{\text{bind}}$ ) of different AVP-RBD complex systems calculated using MM-GB/PBSA calculated for 50 ns<sup>\*</sup>, #

Figure S1. Fasta sequence of antiviral peptides.

Figure S2. Predicted epitopes for major histocompatibility complex class-II (MHC-II) against the spike glycoprotein. Antiviral peptides (black frame) were mapped within predicted epitopes and presented respectively as Seq12, Seq12m, and Se13m (upwards to downwards).

Figure S3. Mutation profile of the receptor binding domain of SARS-CoV-2.

Figure S4. Structural similarity dendrogram. The dendrogram is derived by average linkage clustering of the structural similarities (based on Dali Z-scores).

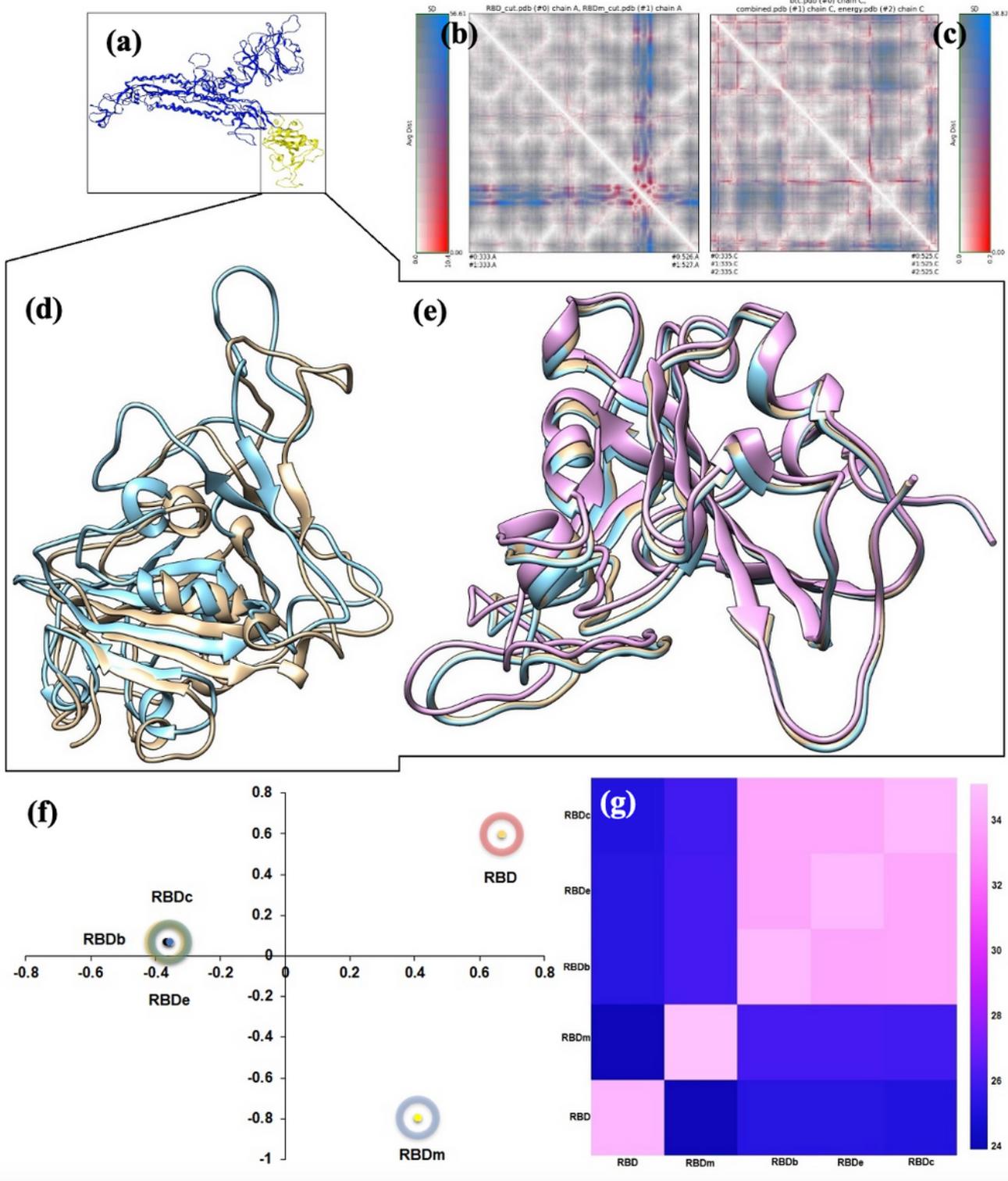
Figure S5. RRDis Map of different mutant models of the RBD of the SARS-CoV-2. (a) RBD and RBDm overlapping, (b) RRDis Map, (c) RBDb/c/e overlapping, (d) RRDis Map

Figure S6. HADDOCK scores and the i-I-RMSD of the docked clusters. (a) HADDOCK clusters of antiviral peptides with the receptor-binding domain of the SARS-CoV-2, (i) Seq12 + RBD, (ii) Seq12m + RBD, (iii) Seq13m + RBD, (b) HADDOCK clusters with antiviral peptides with mutant receptor binding domain of the



**Figure 1**

Peptide sequence mapping and three-dimensional structure of the AVPs. (a) Seq12, (b) Seq12m, (c) seq13m, (d) sequence mapping of the AVPs with the spike glycoprotein of the SARS-CoV-2. amino acid sequences were aligned using multiple sequence alignment with high accuracy and high throughput. \*common amino acid residues.



**Figure 2**

Comparative structural analysis of the receptor-binding domain (RBD) of the SARS-CoV-2. (a) Wild type spike protein of SARS-CoV-2, RBD is highlighted in yellow (b) RRDis map of wild type and manual mutant model of the RBD (c) RRDis map of energy mutant, evolutionary mutant, combine mutant (d) overlapping three-dimensional structures of wild type RBD (gold) and manual mutant model of RBD (cyan), (e) overlapping three-dimensional structures of different mutant models of RBD, energy mutant (pink), evolutionary mutant (gold), combine mutant (cyan), (f) correspondence analysis of different structural models (based on Dali Z-scores), (g) heatmap of different mutant models of the RBD. \* Corresponding analysis is a multidimensional scaling method. It positions data with the most similar structural neighborhoods near each other.

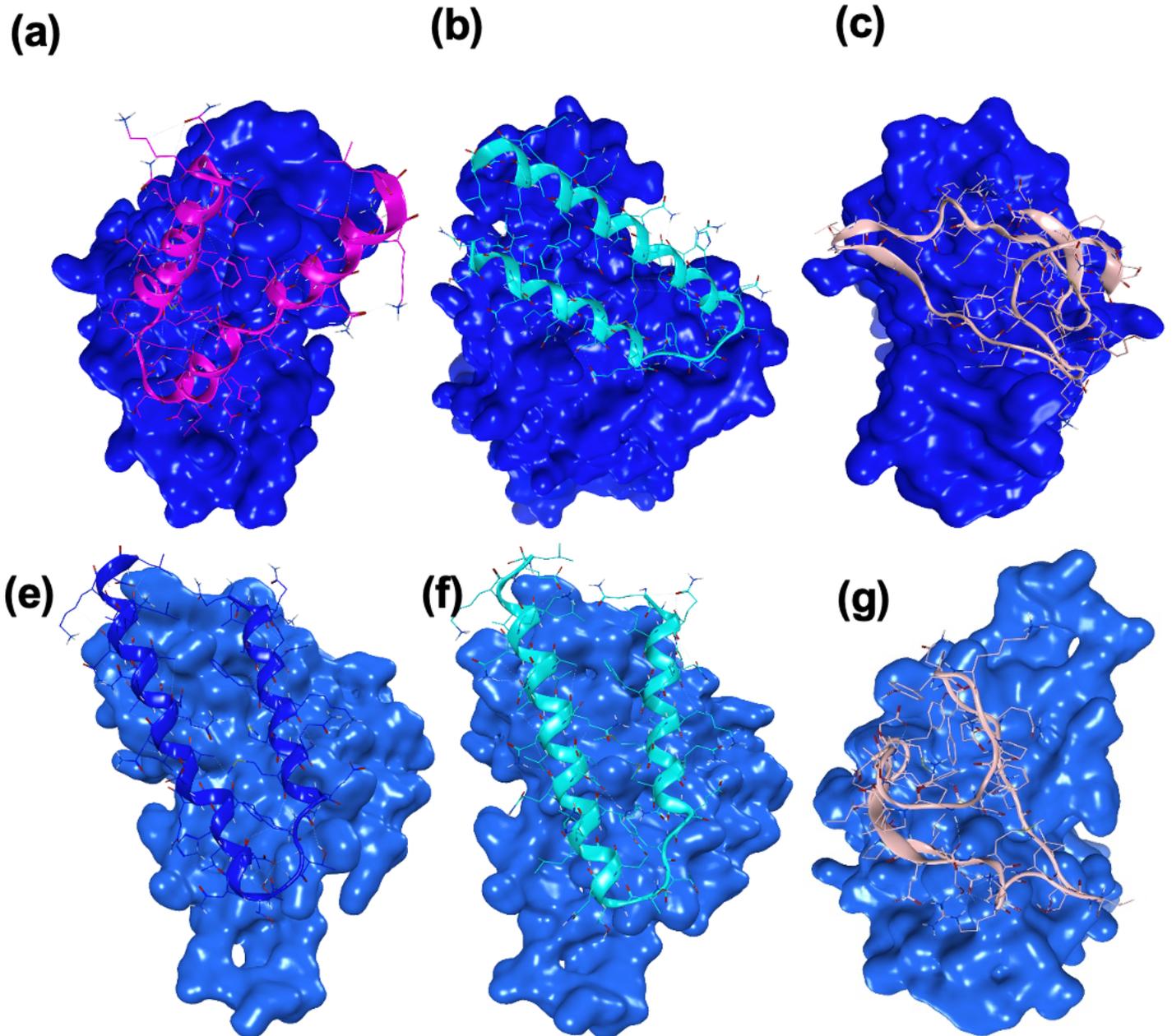
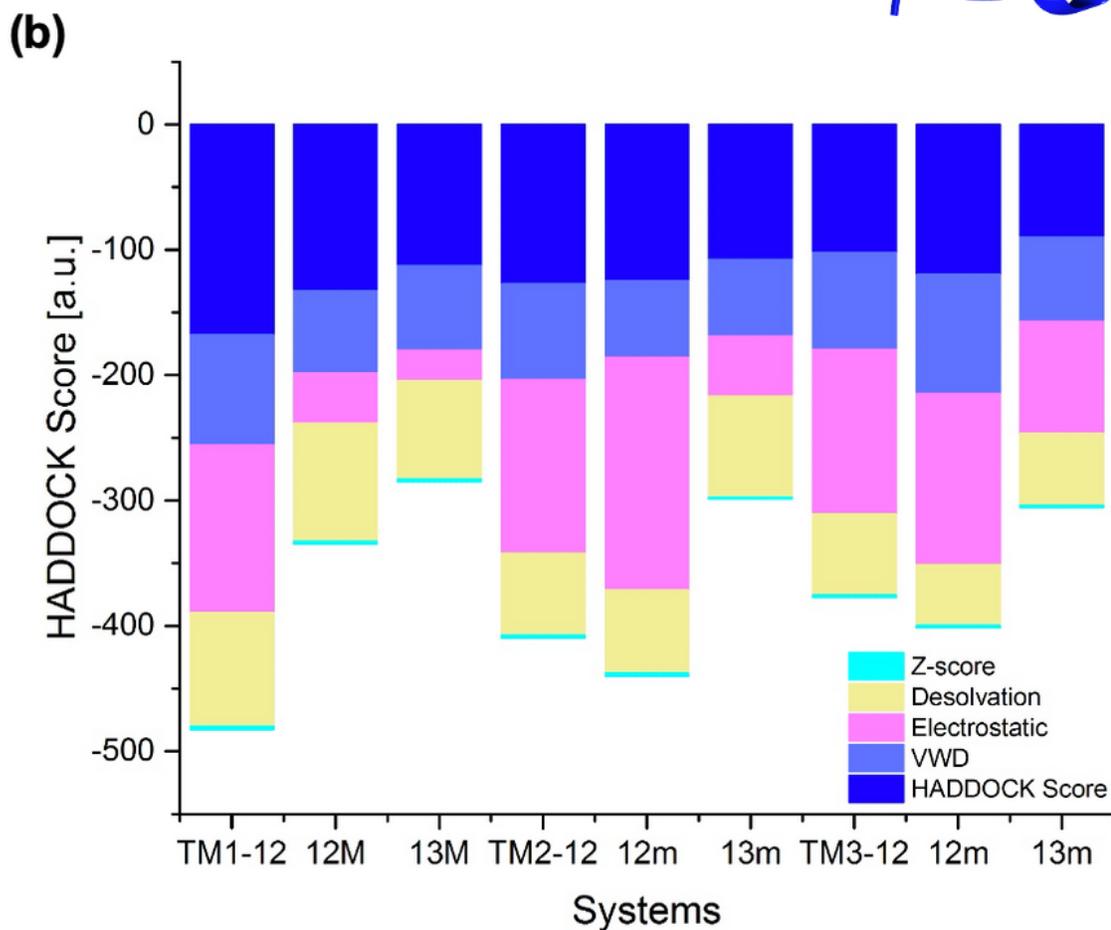
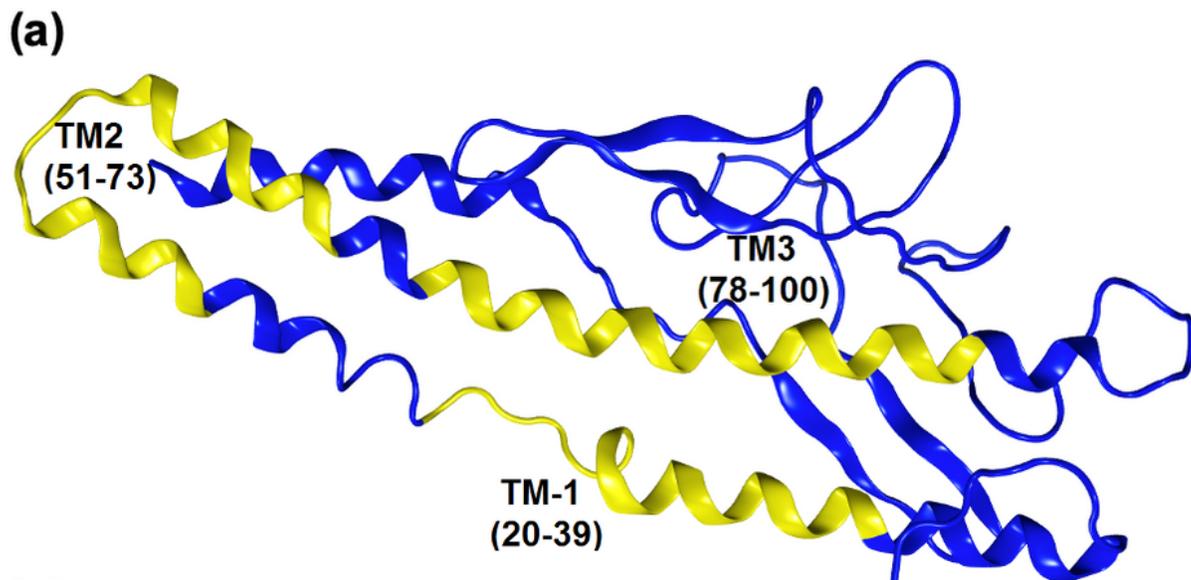


Figure 3

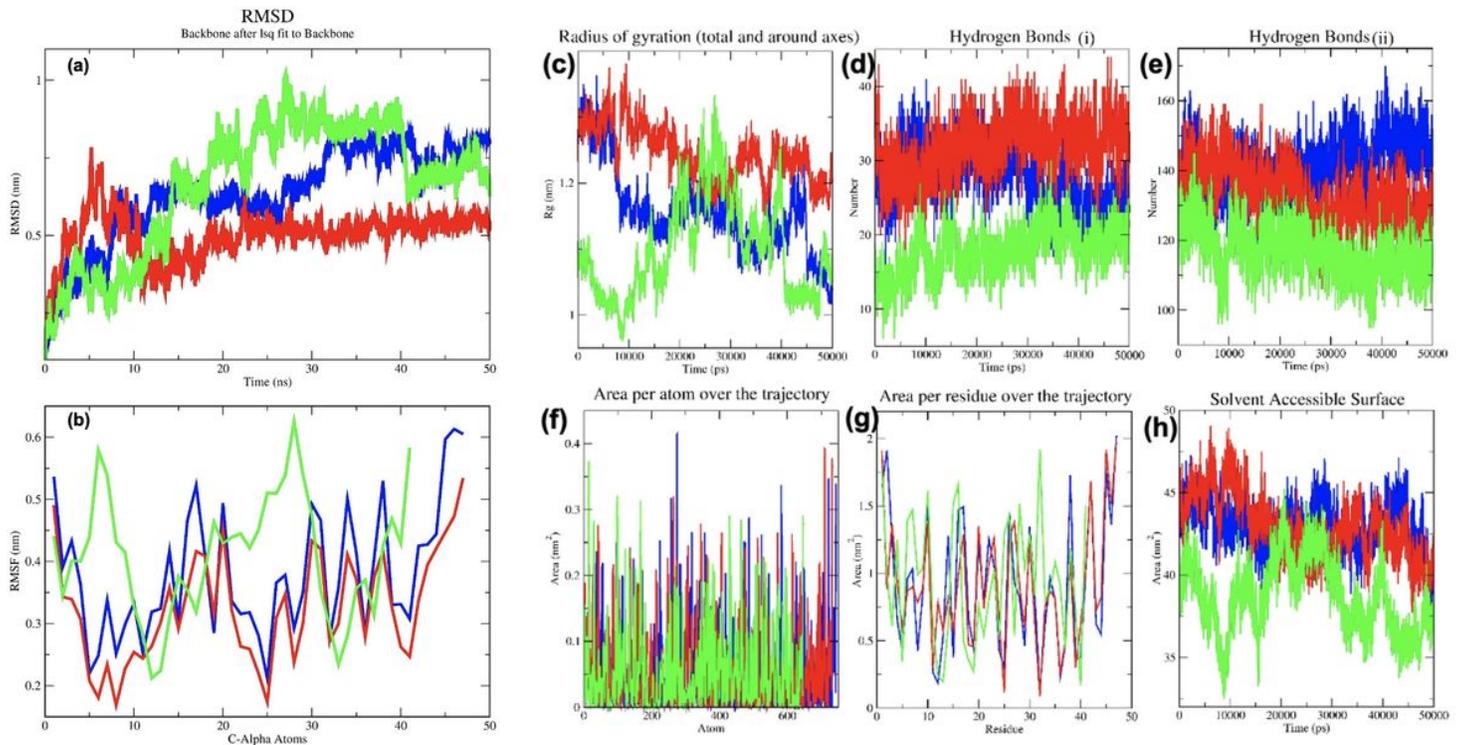
Binding posture of the AVPs with the receptor binding domain of the SARS-CoV-2. wild type receptor binding domain complex with (a) Seq12, (b) Seq12m, (c) Sqe13m, mutant receptor binding domain complex with (d) Seq12, (e) Seq12m, (f) Sqe13m.



**Figure 4**

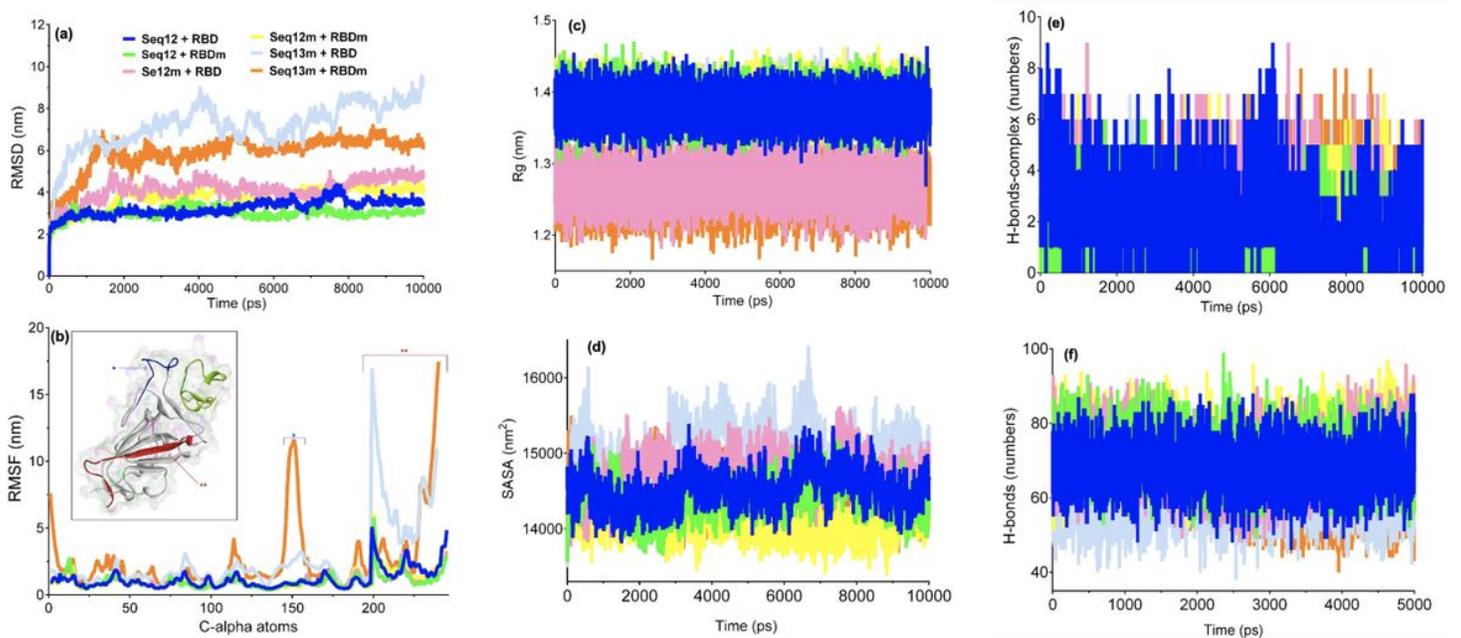
Membrane protein (M) of the SARS-CoV-2. (a) predicted transmembrane domains of the M-protein are highlighted in yellow, (b) analyses of molecular docking results between the AVPs and the

transmembrane domains.



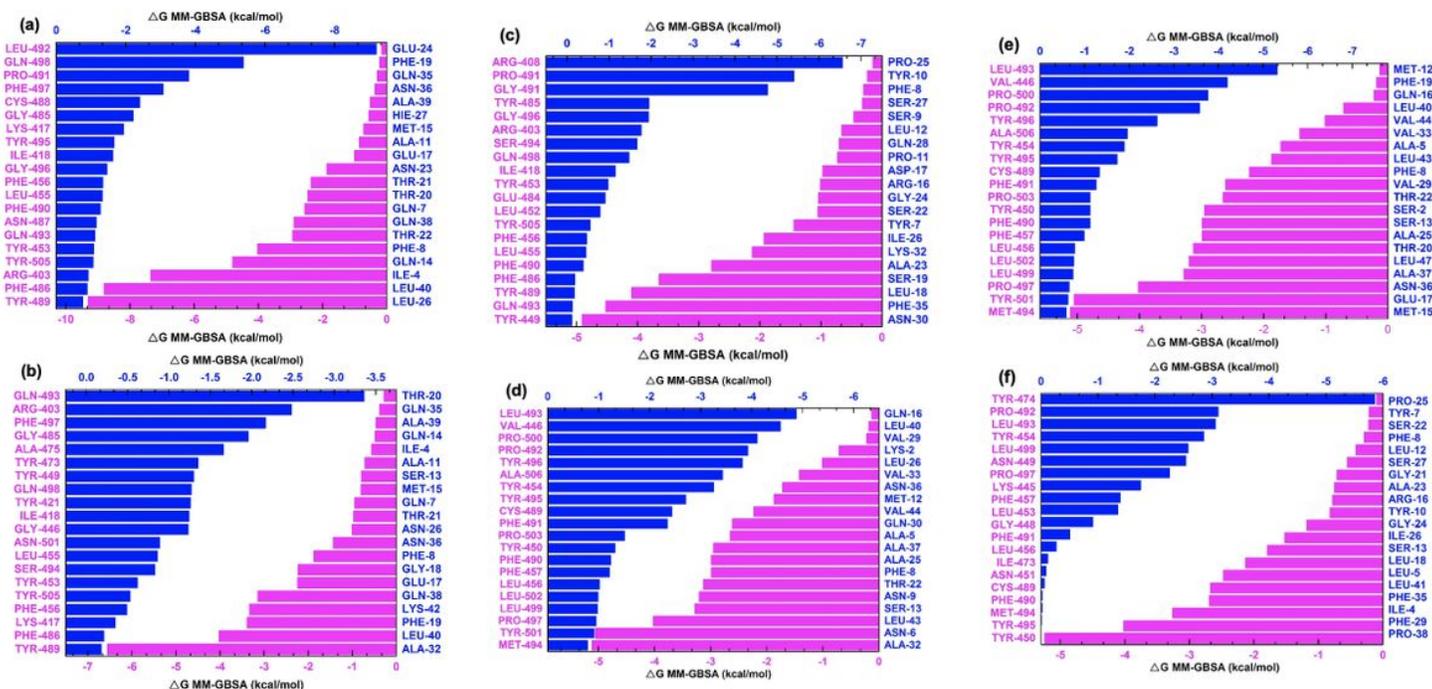
**Figure 5**

Comparisons of MD-Simulation results of the AVPs. (a) RMSD, (b) RMSF, (c) radius of gyration (Rg), (d) inter-peptide H-bond count, (e) water-peptide H-bond count, (f) area per atom over the trajectory, (g) area per residues over the trajectory, (h) SASA. Seq12 (blue), Seq12m (red) and Seq13m (green).



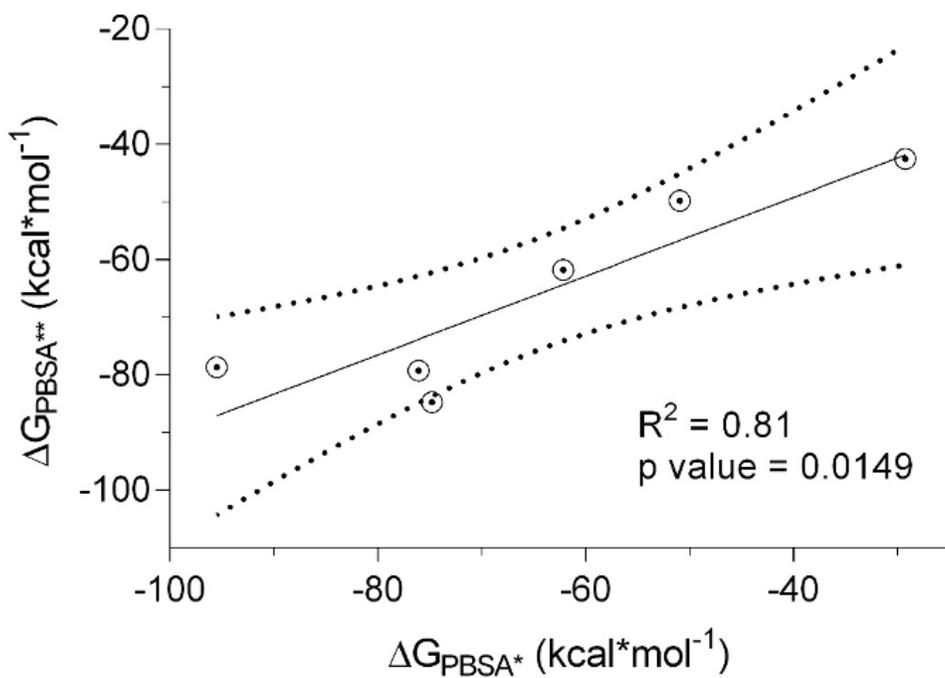
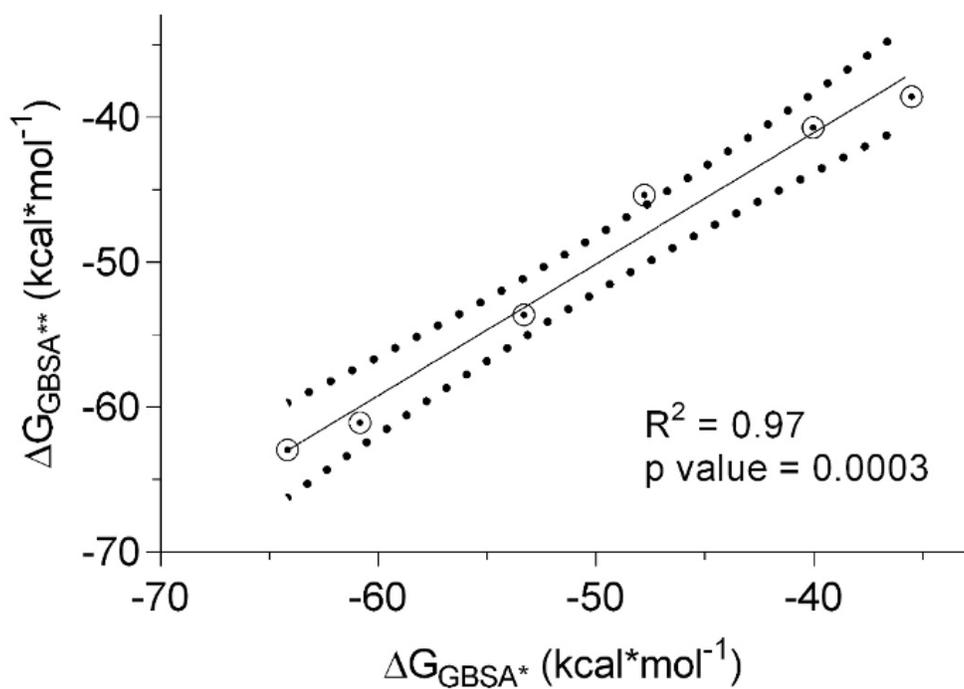
**Figure 6**

Comparisons of the MD-Simulation results of the AVPs complexes with the wild type and mutant type receptor binding domain of the SARS-CoV-2. (a) RMSD, (b) RMSF, high fluctuating amino acid residues of the Seq13m+RBD/m, 138th -158th position (\*), (blue helices), last 25 amino acid residues (red beta sheet), (\*\*) of the RBDm, Seq13m (green) (c) radius of gyration (Rg) around the axis, (d) SASA, (e) total number of H-bond counts of the complex, (f) total numbers of H-bonds with water by restrained 5000 ps-equilibration at NVT ensemble and unrestrained 100 ns-production at 310 K for MM-PBSA calculations.



**Figure 7**

Distribution of binding free energies (MM-GBSA) per amino acid residues. (a) Seq12 + RBD, (b) Seq12m + RBD, (c) Seq13m + RBD, (d) Seq12 + RBDm, (e) Seq12m + RBDm, (f) Seq13m + RBDm. Only top twenty participating amino acid are presented according to the ascending / descending order of energies (MM-GBSA, kcal/mol).



**Figure 8**

Linear regression fit (solid line) between predicted  $\Delta\text{GPB/GBSA}$  calculated for 50 ns (\*) and predicted  $\Delta\text{GPB/GBSA}$  calculated for 100 ns (\*\*). The dotted line around the solid line indicates 95% confidence intervals.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [GraphicalAbstract.png](#)
- [SupplementaryMaterials.docx](#)