JOURNAL OF BIOMOLECULAR STRUCTURE AND DYNAMICS
https://doi.org/10.1080/07391102.2020.1768150

Taylor & Francis
Taylor & Francis Group

Truncated human angiotensin converting enzyme 2; a potential inhibitor of SARS-CoV-2 spike glycoprotein and potent COVID-19 therapeutic agent

Abdul Basit, Tanveer Ali and Shafiq Ur Rehman

Department of Microbiology and Molecular Genetics, University of the Punjab, Lahore, Pakistan

Communicated by Ramaswamy H. Sarma

ABSTRACT

The current pandemic of Covid-19 caused by SARS-CoV-2 is continued to spread globally and no potential drug or vaccine against it is available. Spike (S) glycoprotein is the structural protein of SARS-CoV-2 located on the envelope surface, involve in interaction with angiotensin converting enzyme 2 (ACE2), a cell surface receptor, followed by entry into the host cell. Thereby, blocking the S glycoprotein through potential inhibitor may interfere its interaction with ACE2 and impede its entry into the host cell. Here, we present a truncated version of human ACE2 (tACE2), comprising the N terminus region of the intact ACE2 from amino acid position 21-119, involved in binding with receptor binding domain (RBD) of SARS-CoV-2. We analyzed the in-silico potential of tACE2 to compete with intact ACE2 for binding with RBD. The protein-protein docking and molecular dynamic simulation showed that tACE2 has higher binding affinity for RBD and form more stabilized complex with RBD than the intact ACE2. Furthermore, prediction of tACE2 soluble expression in *E. coli* makes it a suitable candidate to be targeted for Covid-19 therapeutics. This is the first MD simulation based findings to provide a high affinity protein inhibitor for SARS-CoV-2 S glycoprotein, an important target for drug designing against this unprecedented challenge.

ARTICLE HISTORY
Received 28 April 2020
Accepted 5 May 2020

KEYWORDS
ACE2; SARS-CoV-2; spike glycoprotein; inhibitor; protein-protein docking; MD simulation

1. Introduction

The rapid spread of SARS coronavirus 2 (SARS-CoV-2) demands an immediate public health emergency, and no FDA approved treatment/vaccines are currently available. SARS-CoV-2 Spike (S) protein (1267 amino acids) is essential for virus entry through binding with the host receptor angiotensin converting enzyme II (ACE2) and mediating virus-host membrane fusion (Boopathi et al., 2020; Sarma et al., 2020). The S protein contains two functional domains (S1 and S2). The S1 (residues 14-685) domain performs the function of virion attachment with human ACE2 receptor on epithelial membrane cell surface, followed by its internalization, hence initiating the infection (Hasan et al., 2020). This binding induces certain conformational changes in the S protein, which results the S2 (residue 686-1273) to mediate fusion with cellular membrane. The receptor-binding domain (RBD) of the SARS-CoV-2 S protein are highly conserved and directly involve in binding to human ACE2 (Yuan et al., 2020). Since, ACE2 is not mutated/evolved to recognize S protein of SARS-CoV-2; therefore, using alternative of ACE2 with more binding affinity for S protein than the wild type receptor, may inhibit entry of SARS-CoV-1& −2 into human cells. This strategy can play important role in devising therapeutics of SARS-CoV-2. Several studies have proposed small compounds based inhibitors as therapeutic agents for Covid-19 (Aanouz et al., 2020; Elmezayen et al., 2020; Gupta et al., 2020; Khan et al., 2020; Wahedi et al., 2020). The small compounds based drugs may not efficiently block the entire binding patch of S protein. On the other hand, the peptides based therapeutics can block the entire binding interface (RBD) of S protein (Wan et al., 2020b), as reported for HIV peptide based drug Fuzeona (Jenny-Avital, 2003; Wójcik & Berlicki, 2016). There is growing interest in peptide based therapeutics for Covid-19 treatment (Pant et al., 2020) and approximately 140 peptide based drugs have been evaluated in clinical trials (Fosgerau & Hoffmann, 2015). Peptide based drugs have little side effects and little drug tolerance compared with chemical drugs (Bruno et al., 2013).

In order to block the fusion of SARS-CoV-2 S protein with human cells, a recent study has reported a neck and transmembrane deficient ACE2, called as soluble ACE2 (sACE2), that can block the entry of SARS-CoV-2 into the host cell (Procko, 2020), which is also found safe in healthy human subjects (Haschke et al., 2013) and patients with lung disease (Khan et al., 2017). Recombinant sACE2 is under clinical trials for COVID-19 treatment in Guangdong province of china (Clinicaltrials.gov #NCT04287686). The study proposed that mutations in ACE2 receptors interface may increase S/ACE2 interaction. Another study has proposed a 23 amino acid peptide, derived from ACE2 (amino acid position 21-43), which can bind with SARS-CoV-2 S protein with a low

CONTACT Shafiq Ur Rehman shafiq.mmg@pu.edu.pk Department of Microbiology and Molecular Genetics, University of the Punjab, Lahore, 54590, Pakistan

Supplemental data for this article can be accessed online at https://doi.org/10.1080/07391102.2020.1768150.

nanomolar affinity, and can block the attachment of SARS-CoV-2 to human ACE2 (Zhang et al., 2020).

Since, the binding residues of ACE2 involve in interaction with RBD are located at amino acid position 21-119 (Wan et al., 2020a; Yan et al., 2020), therefore, we hypothesized that this fragment carrying all the binding residues will have better binding affinity for RBD and can hinder the interaction of SARS-CoV-2 with human ACE2, hence blocking its entry into the epithelial cells. We designed a truncated version (tACE2) of ACE2 receptor covering the binding residues and performed protein-protein docking and molecular dynamic simulations to analyze its binding affinity for RBD and complex stability. The tACE2 will compete with wild type human ACE2 receptors for binding to SARS-CoV-2, as they will have more binding affinity for S protein. This will allow all SARS-CoV-2 viral particles to bind strongly with the tACE2, blocking all its available binding sites for the host ACE2 receptors, thus inhibiting its entry into the cell which will be eliminated through body defense mechanisms. We further determined the soluble expression for tACE2 in *E. coli,* a suitable host for bulk production of tACE2.

2. Material and methods

2.1. ACE2 and SARS-CoV-2 spike protein structures and sequence alignment

The PDB structure of ACE2 and RBD of SARS-CoV-2 S glycoprotein (PDB ID: 6m17) was obtained from PDB database. In order to determine the variation in the SARS-CoV-2 S glycoprotein sequence reported from different regions of the globe, 61 S glycoprotein sequences of SARS-CoV-2 including reference sequence (NC_045512, reported from Wuhan, china) were retrieved from NCBI. Multiple sequence alignment of the sequences was performed through MEGA-X. The aligned sequences were then analyzed for amino acid variations.

2.2. Truncation of ACE2 and structure optimization

The PDB structures of ACE2 and RBD were repaired for their missing loops and optimized for energy minimization and amino acid side chain clashes through FoldX (Schymkowitz et al., 2005). Side chains were optimized through FoldX to remove Vander Waals' clashes by mutating residues with bad energy values into new rotamers with energy minimization (Van Durme et al., 2011). The optimized three dimensional (3 D) structures of ACE2 and RBD were used to design truncated ACE2 and studying protein-protein interactions.

Based on protein-protein interactions between ACE2 and RBD shown in ACE2-RBD complex (PDB ID 6m17), a truncated version of ACE2 was produced by removing the C-terminus residues from amino acid position 116-768, leaving a truncated N-terminus fragment tACE2, from 21-119 amino acid position. The first 20 residues of ACE2 is the signal peptide (Huang et al., 2003; Turner & Hooper, 2004), therefore it was also removed. The structure of tACE2 was produced through I-TASSER, which build the model by assembling continues fragments of multiple threading templates, identified through Replica Exchange Monte Carlo (REMC) simulations (Yang et al., 2015).

2.3. Protein-protein docking

In order to determine binding affinity of both intact and truncated ACE2 with SARS-CoV-2 S glycoprotein, rigid body protein-protein docking tools; ZDOCK (Pierce et al., 2014), ClusPro (Kozakov et al., 2017), Patchdock (Schneidman-Duhovny et al., 2005) and a flexible protein-protein docking tool, HADDOCK (Van Zundert et al., 2016) were used. The energy function used by ZDOCK is Z score, which is a cumulative of pairwise shape complementarity function with desolvation and electrostatics. The ZDOCK rank the top 10 predicted docking poses on the basis of Z score (Chen et al., 2003). ClusPro uses PIPER's scoring function, which contains terms of shape complementarity, electrostatics, and pairwise potentials applied on the top 1000 conformations produced and ranked on the basis of cluster size. Patchdock uses patchdock score as the energy function which ranked the docked model based on desolvation energy, interface area size and geometric score (Zhang et al., 1997). HADDOCK is a flexible docking method used for docking of protein-protein complexes. HADDOCK drive the docking process by retrieving information from experimentally identified protein complex interfaces. The HADDOCK scoring function consists on combination of various energies and buried surface area. The scoring of the models was performed according to the HADDOCK score. All the generated docking poses of ACE2 and Spike protein were visualized through PyMOl (Schrodinger, 2010).

2.4. Protein-protein binding affinity prediction

Based on the HADDOCK score and the docking RMSD value, the docked complexes of ACE2 and tACE2 with RBD were analyzed for binding affinity ΔG (kcal mol^{-1}) and stability using protein binding energy prediction (PRODIGY) server (Xue et al., 2016). The server predicts the binding affinity and stability on the basis of structural properties of the protein-protein complexes. Stability of the protein-protein complex is measured through dissociation constant K_d (M). The run was performed at different temperatures ranging from 25 to 36°C. The protein-protein docked complex with the minimum RMSD and higher binding affinity was considered for MD simulation to further confirm stability of the complex.

2.5. Molecular dynamic (MD) simulations

MD simulation of the RBD domain in complex with intact ACE2 and tACE2 was performed through GROMACS 5.0.4 (Abraham et al., 2015). Simulation was performed by using charm 36.0 force field and TIP3P cube box as water model. The protein complex in the cubic box was solvated with water molecules to provide an aqueous environment. The system was then neutralized with addition of 3 Na ions followed by energy minimization for removal of conflict

between the atoms. The system was then equilibrated through NVT and NPT at constant temperature (300 K) and pressure (1 bar), respectively. Langevin thermostat was applied to regulate temperature of the system. MD simulation was then run for 20 ns.

2.6. Post translation modifications predictions

In order to determine post translation modifications (PTMs) in ACE2, the protein sequence was submitted to PTM-ssMP server, which combines the submitted sequence and site specific modification profile to predict PTM sites in mammalian protein (Liu et al., 2018). Since, glycosylation is the most abundant and diverse posttranslational modification of proteins, therefore, we further determined the O-glycosylation sites in ACE2 using NetOGlyc 4.0 Server which specifically predict the GalNAc-type O-glycosylation site, unique to Ser and Thr (Steentoft et al., 2013). We further determined the N-glycosylation sites by using NetNGlyc-1.0 Server using a threshold value of 0.5(Gupta et al., 2004).

2.7. Solubility prediction of tACE2 for expression in E. coli

In order to express the tACE2 in *E. coli*, its soluble expressionat 37 °C was determined through CamSol Intrinsic and CamSol Structurally Corrected online solubility prediction tools (Sormanni & Vendruscolo, 2019). CamSol determines the solubility on the basis of amino acid sequence, while CamSol Structurally Corrected tool determines the solubility profile on the basis of the structure, which accounts for amino acid distribution in the structure and their solvent exposure. Both run was performed at pH 7.0. In both methods, the solubility profile scores higher than 1.0 denotes highly soluble regions, while scores lower than −1 indicates poor solubility in *E. coli*.

3. Results and discussion

3.1. SARS-CoV-2 S protein showed sequence conservation in the receptor binding domains

In the current study, we have proposed a truncated version of ACE2 that comprises the binding interface for receptor binding domain (RBD) of SARS-CoV-2 Spike protein. Recently, in-vitro binding assay have confirmed that RBD is mainly responsible for initial binding to ACE2, which further mediate virus entry into the host cell (Lan et al., 2020). Variation in the RBD sequence was analyzed in the SARS-CoV-2 genome reported from various region of the globe so far (Shu et al., 2020). The sequence alignment showed more than 99.99% homology for RBD domain, with only single variation R408I in the SARS-CoV-2 genome reported from India (Figure S1). The rest of the SARS-CoV-2 genome sequences submitted throughout the globe have identical RBD sequence, which indicate that the SARS-CoV-2 RBD is highly conserved globally. Structural elucidation has also found the RBD domain as highly conserved (Lan et al., 2020).

3.2. Designing truncated variant of human ACE2

In order to block the Spike protein attachment to the cell, the ACE2/RBD binding interface comprising residues from position 21-119 of ACE2 was selected as truncated version of ACE2.The structure of tACE2 was built through I-TASEER with C-score 1.22. The C-score value in range −5 to 2 shows correctness of the fold. The high C-score for tACE2 suggest the highly likelihood of the structure. The tACE2 fragment contains almost all binding residues involve in binding with RBD domain of SARS-CoV-2 (Yan et al., 2020), covering two complete helices (Lan et al., 2020). This suggests that rational design of a binder based on this interface with enhanced affinities to RBD may play vital role by blocking the SARS-CoV-2 spike protein interaction with ACE2, thus inhibiting viral entry into the host cell. Previously, peptides based strategies have been employed successfully to inhibit fusion of the SARS-CoV-1 S protein and membrane receptor (Du et al., 2009). Another recent study has reported a 23 amino acid based peptide, a homologue of ACE2 binding interface, which successfully bind with S protein with low nanomolar affinity (Zhang et al., 2020). Since, the binding residues for ACE2 are located at distant location on RBD, thus providing a larger protein binding site, which is difficult for a small size therapeutic peptide to cover the entire binding sites on RBD. However, our proposed tACE2 fragment carrying almost all the binding residues that can block the attachment of RBD with the intact ACE2.

3.3. Protein-protein docking

RBD was docked with intact and truncated ACE2 through HADDOCK, a flexible protein-protein docking tool. The method allows the side-chains and backbone atoms of both the protein and receptor flexible during docking run (Geng et al., 2017). HADDOCK scoring function (HADDOCK score) is a linear combination of non-bonded intermolecular van der Waals (VWs), Coulomb electrostatics energies and empirically derived desolvation energy term (Vangone et al., 2017).The HADDOCK-Score of ACE2 and tACE2 was −111 and −126.6, respectively, (The more negative the better). Similarly the VWs and electrostatic energy of tACE2-RBD complex was also greater than the ACE2-RBD complex, which shows higher binding affinity of tACE2 for RBD than the intact ACE2 (Table 1). The RMSD value of ACE2 and tACE2 in complex with RBD were 0.7 and 0.8, respectively, showing the high likelihood of the docked complexes with native-one (Vangone et al., 2017).

In order to further confirm these docking scores, rigid docking was also performed through Patchdock, Z-dock and ClusPro protein-protein docking tools. The docking results obtained for ACE2 was compared with tACE2 in term of energy functions of each docking tool (Table 2). All the three docking scores are higher for tACE2 than that of the intact ACE2, indicating high affinity of tACE2 for RBD.

Our docking results showed that seven residues of ACE2 Glu23, Thr27, Asp30, Glu35, Tyr 83, Asn 330 and Lys 353 of ACE2 interact with RBD residues Lys417, Lys458, Asn487, Tyr

Table 1. Comparison of the cluster statistics of intact and truncated ACE2 docked with RBD through HADDOCK docking tool.

#	Protein	HADDOCK-score*	Cluster size	Docking RMSD	Z-score**	Vander Wall energy	Electrostatic energy
1	ACE2	−111	23	0.7	−1.8	−59.9	−151.2
2	tACE2	−126.6	44	0.8	−2.1	−70.9	−185.2

*The HADDOCK score is defined as: 1.0 Evdw + 0.2 Eelec + 1.0 Edesol + 0.1 EAIR.
**The Z-score produced by HADDOCK indicates standard deviations from the average cluster (the more negative the better).

Table 2. Comparison of rigid body protein-protein docking results of the intact and truncated ACE2 docked with RBD through various docking tools.

#	Protein	Z-dock score	ClusPro score	Patchdock score
1	ACE2	1627	−813.4	16172
2	tACE2	1812	−1083.1	18058

489, Gln493, Tyr495, Gly496, Thr500 and Gly502, respectively, which is almost similar to the binding residues profile of ACE2 interface reported previously (Yan et al., 2020), with additional Thr27 and Glu35 reported by our docking results (Figure 1(A, B)). However, the tACE2 form a different binding residues network than the intact ACE2. Our docking results showed that Ser23, Asn31, Tyr30, Glu36, Gln40, Gln76 and Arg95 of tACE2 are involved in binding with RBD (Figure 1(C, D)). This seems that the truncation has produced the conformational changes in the tACE2-RBD complex which results in exposure of buried binding residues (Basit & Akhtar, 2018), thus facilitate higher binding of tACE2 to the RBD as compared to the native ACE2, which are in agreement with previously reported peptides inhibiting viral attachment with the host cell (Koehler et al., 2013).

3.4. tACE2 demonstrated higher affinity with RBD than wild type ACE2

Since, docking methods are not reliable for predicting binding affinity between protein-protein complexes, due to their simple scoring functions (Ramírez & Caballero, 2016). As binding affinity of protein-protein complex also depends on dissociation constants (K_d), pH and temperature (Kastritis & Bonvin, 2010), while these parameters are not included in the benchmark of docking scoring functions. Therefore, we determined the binding affinity of ACE2 variants for RBD through PRODIGY server, which determine the binding affinity based on structural properties of the protein-protein complexes (Vangone & Bonvin, 2015). The ACE2 and tACE2 complexes showed −10.7 and −12.7ΔG (kcal mol^{-1}) binding affinity for RBD, respectively, at temperatures ranges from 20 to 37 °C, showing higher binding affinity of tACE2 for RBD than the intact ACE2 (Table 3).

Similarly, the dissociation constant K_d value of tACE2-RBD complex was more than three-fold lesser than the intact ACE2-RBD complex, showing that tACE2 is more tightly bound to RBD. The smaller K_d value indicates high stability and strong binding affinity between protein-protein complex (Johnson et al., 2007). The ACE2 variants showed a significant decline in K_d value when temperature was increased from 20 °C to 36 °C, leading to a lower K_d (9.8×10^{-10} M) for tACE2 (higher affinity) than that of intact ACE2 (2.6×10^{-8} M) at 36 °C.This K_d value of tACE2 is lesser than the previously reported K_d value (47 nM) of SBP1 (an ACE2 derived peptide of 23 amino acid) to RBD (Zhang et al., 2020). The optimum stability of the complexes was found at 36°C (Table 3). The dramatic changes of binding kinetics might be caused by reduced stability of ACE2 complex below optimum temperature 36°C (Zhao et al., 2018).

3.5. MD simulation showed stability of tACE2-RBD complex

In order to determine the structural stability and dynamic behavior of intact ACE2-RBD and tACE2-RBD complexes, we performed MD simulation for 20 ns using GROMACS 5.0.4. The docking pose of each complex obtained from HADDOCK with lowest energy was selected for MD simulation run. To investigate structural stability of the complex, RMSD plot of the complex backbone was produced. A uniform RMSD plot signifying structural stability of tACE2-RBD complex. The RMSD value for tACE2 complex was 0.2–0.25 nm, while intact ACE2 showed 0.25–3.0 nm RMSD (Figure 2). The RMSD value of tACE2-RBD complex is lesser than SBP1-RBD complex, reported previously, which is almost 0.8 nm (Zhang et al., 2020), showing higher stability of tACE2-RBD complex.

Root mean square fluctuation (RMSF) was determined to evaluate the residues flexibility of both ACE2 and RBD in the docked complexes. The high RMSF values indicate the mobility of residue side chains in relation to their average position (Kumar et al., 2014).The RMSF plot shows the residues of RBD in tACE2 complex are stable with a few peaks with RMSF more than 0.2 nm (Figure 3(B)), while RBD of ACE2 complex shows many residues with rmsf above 0.35 nm (Figure 3(D)).The residues of tACE2 at position 24, 30, 40, 76 and 95 showed reduction in rmsf value due to creating binding interactions with RBD (Figure 3(A)). The residues involved in binding with other protein, present lower RMSF values, reveal the most stable regions of the complex (Ardalan et al., 2018).Similarly, the residues window of 470–480 of RBD showed higher fluctuation to 0.25 nm, while decrease in fluctuation at the binding residues positions (Figure 3(B)). The most violent fluctuation in the intact ACE2 was observed at C-terminus, which was above 0.7 nm (Figure 3(C)).

The overall RMSF values of both tACE2 and RBD are below 0.2 nm, which indicate that tACE2 complex with RBD is stable, which are in agreement with a previously reported rmsf value 0.4 nm, showing complex stability (Maqsood et al., 2020). The overall trajectories obtained after every 100 ps during a 20 ns MD simulation run, very small backbone deviation for both the intact ACE2 and tACE2 complex was observed (Figure 4). However, the amino acid region 470–489 of RBD has shown backbone fluctuation highlighted as yellow (Figure 4(C)), which we suggest the region of binding site for ACE2. Previously, the amino acid region of the SARS-CoV-2 spike protein (480–488) was also reported as binding region for ACE2 (Ibrahim et al., 2020).

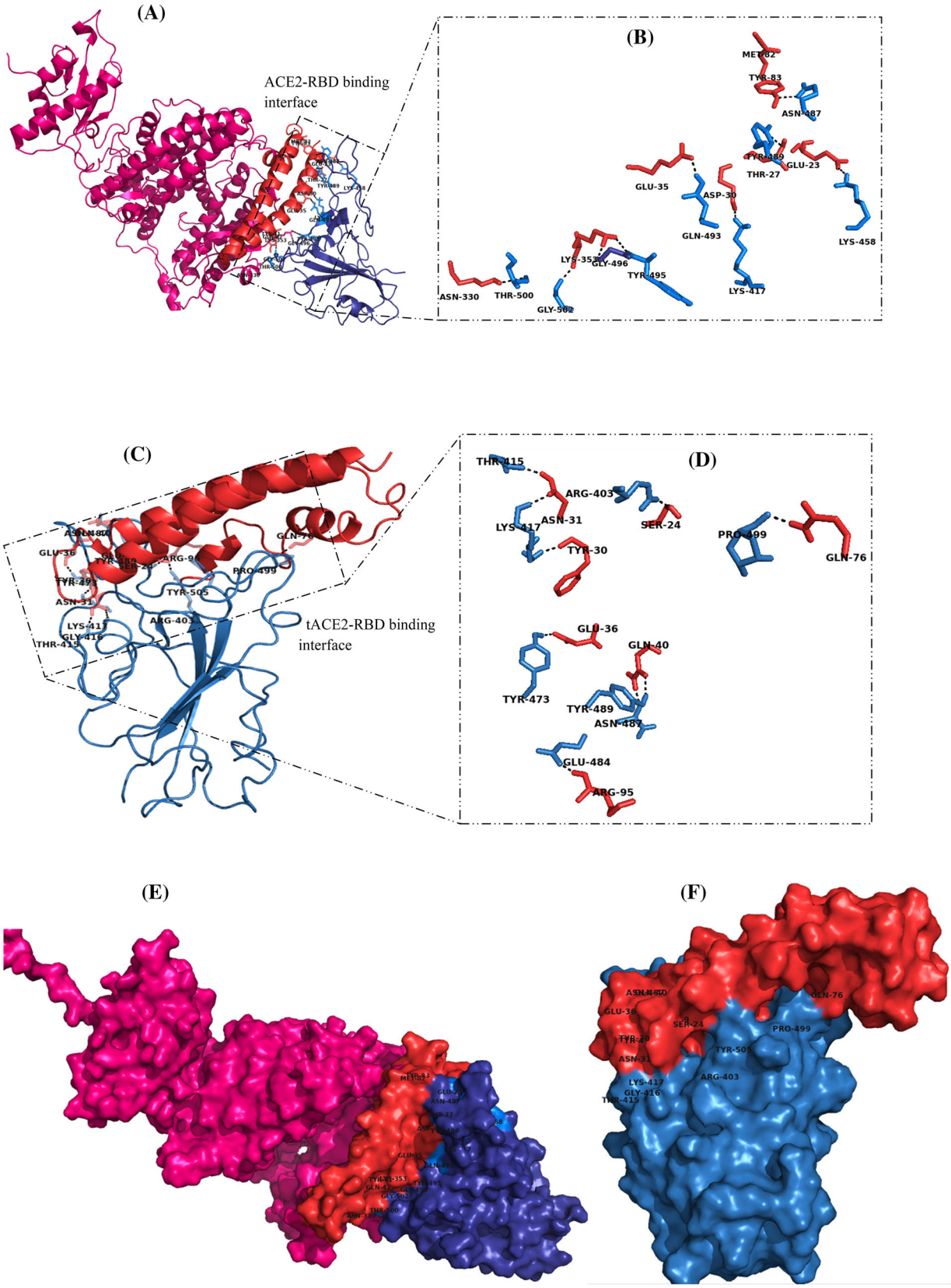

Figure 1. Structural analysis of the **(A)** intact human ACE2 (pink) recognition by RBD (blue) of SARS-CoV-2 S protein. The binding interface of ACE2 comprising two α-helices is shown in red color. **(B)** ACE2-RBD binding interface showing residues of ACE2 (red) involved in interaction with RBD (blue). **(C)** tACE2-RBD complex. **(D)** Binding residues of tACE2 (red) showing interaction with RBD (blue). The interactions are denoted by black dots. **(D)** Surface model of ACE2-RBD complex showing interaction of intact ACE2 (pink), binding interface of ACE2 (red) and RBD (blue) and **(E)** surface model of tACE2-RBD complex showing tACE2(red) and RBD (blue).

Radius of gyration (R g) of both ACE2 complexes describes overall spread of molecule during a 20 ns MD run. A low Rg value indicates better structural integrity and folding behavior (Erva et al., 2016). A slight increase in Rg value of the intact ACE2-RBD complex was observed during first 5 ns of the run, then after no further drifts till end (Figure 5, red line), however, the tACE2-RBD complex was found stable throughout the MD run (Figure 5, violet line), which indicates its structural integrity. Overall, the MD simulation results confirm that tACE2 form a more stabilized complex with RBD and suggest its inhibitory features for SARS-CoV-2 Spike glycoprotein.

Table 3. Binding affinity of intact ACE2 and tACE2 for S protein and their dissociation constant (K_d) at different temperatures.

		Binding affinity ΔG (kcal mol^{-1})		Dissociation constant K_d (M)	
#	Temperature	ACE2	tACE2	ACE2	tACE2
1	20°C	−10.7	−12.7	1.0×10^{-8}	3.2×10^{-10}
2	24°C	−10.7	−12.7	1.3×10^{-8}	4.2×10^{-10}
3	28°C	−10.7	−12.7	1.6×10^{-8}	5.6×10^{-10}
4	32°C	−10.7	−12.7	$2.1 \times 10-8$	7.5×10^{-10}
5	36°C	−10.7	−12.7	2.6×10^{-8}	9.8×10^{-10}
6	37°C	−10.7	−12.7	2.7×10^{-8}	1.0×10^{-9}

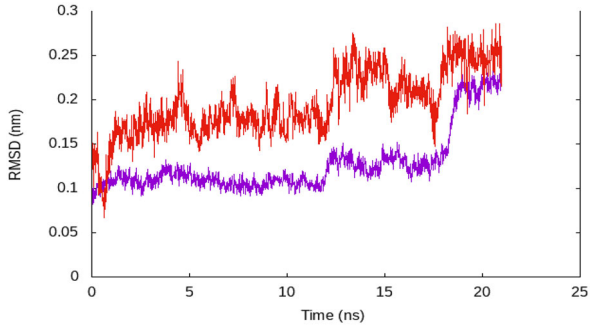

Figure 2. RMSD plot of the ACE2-RBD (red) and tACE2-RBD complex (violet) backbone atoms. The tACE2 complex showing less RMSD value than the intact ACE2, indicating its higher complex stability than the intact ACE2.

3.6. PTMs analysis

Post-translational modifications (PTMs) play important role in protein-protein interactions (Su et al., 2017). Since, experimental methods are high-cost and time-consuming, therefore, it is necessary to theoretically predict PTMs site on protein to be expressed heterologously. PTM-ssMP, which predict PTMs sites on human protein based on local sequence and site specific modification profile (Liu et al., 2018). ACE2 analysis through PTM-ssMP server predicted ubiquitination at position 74 and 304, phosphorylation at 606 and O-glycosylation at 720 residue position. The PTM site at 74 is important for protein degradation and have no role in PPIs (Lecker et al., 2006). In transmembrane proteins, the extracellular domains may only be N-glycosylated (Gupta et al., 2004). However, there was no N-glycosylation and O-glycosylation site predicted for tACE2. These results conclude that there is no PTMs site predicted on tACE2, which is important for protein-protein interactions. Interestingly, an experimental study reported that the lack of glycosylation do not affect the binding of SARS-CoV-1 RBD to human ACE2 (Chakraborti et al., 2005), which strongly support our designed tACE2 fragment, if expressed in *E. coli* may bind efficiently with RBD of SARS-CoV-2 S glycoprotein.

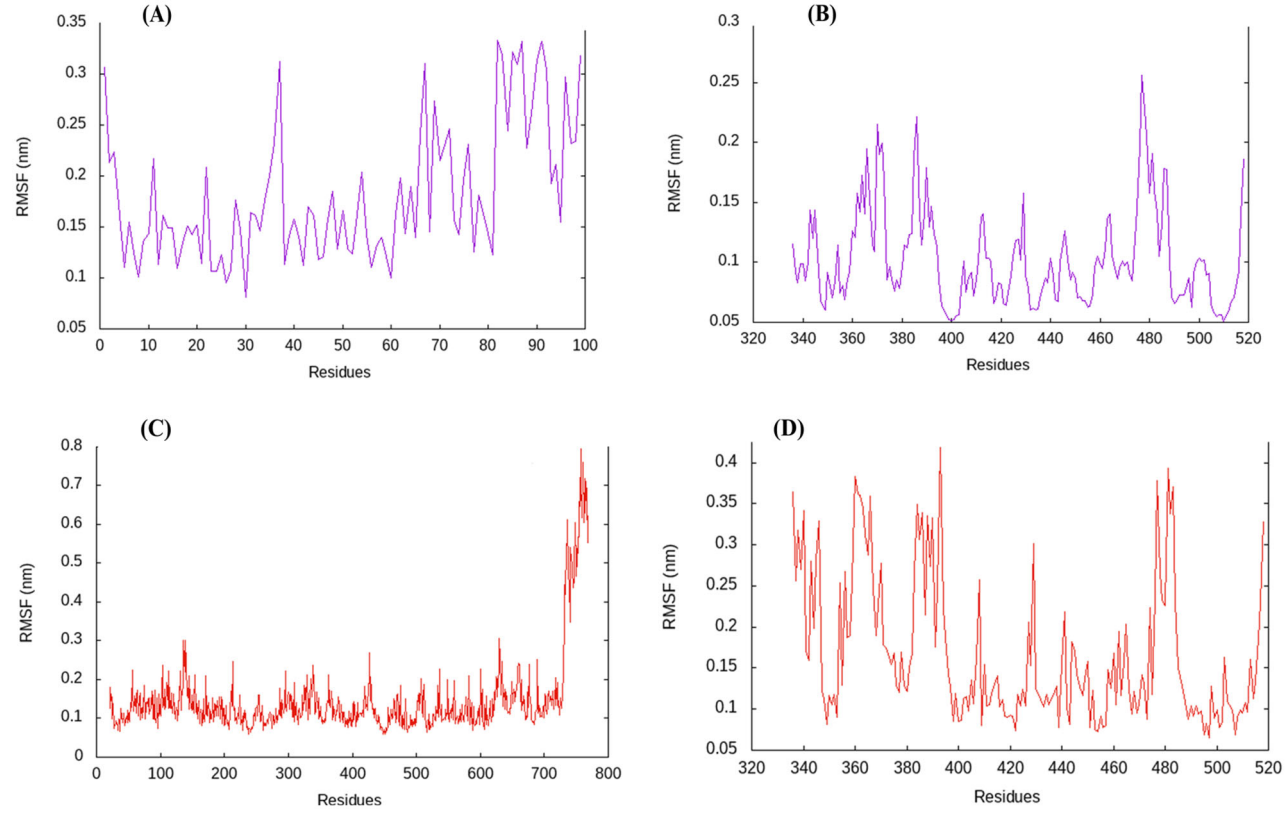

Figure 3. RMS fluctuation of residues side chains of **(A)** tACE2, **(B)** RBD in complex with tACE2, **(C)** Intact ACE2 and **(D)** RBD in complex with ACE2.

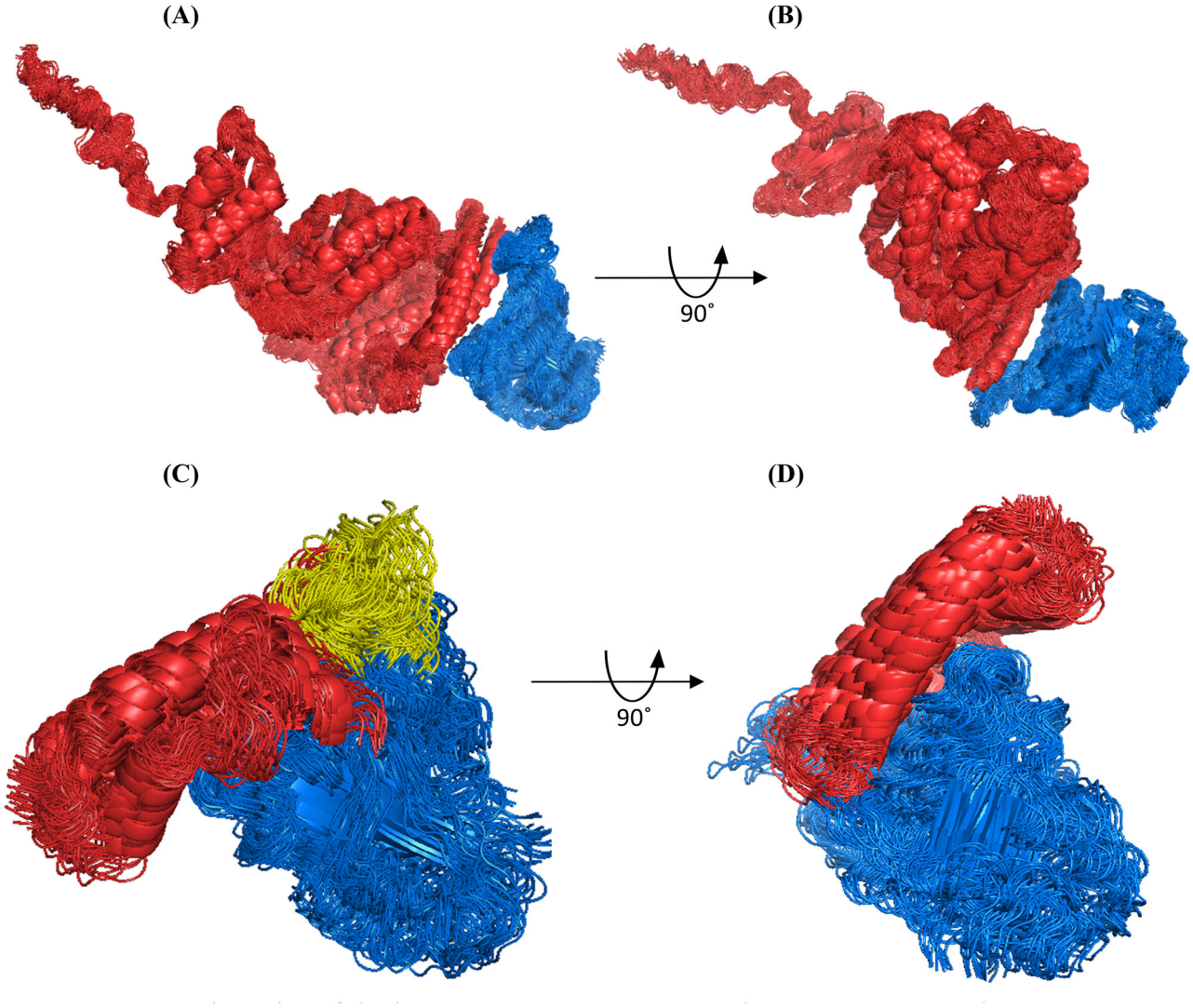

Figure 4. PDB trajectories of the intact ACE2-RBD **(A&B)** and tACE2-RBD complex **(C&D)**, showing the backbone fluctuation of protein structure after a 20 ns MD simulation run. A slight fluctuation can be seen in the RBD of the tACE2 complex at position 471–489, highlighted yellow. ACE2 variants and RBD are shown by red and blue colors, respectively.

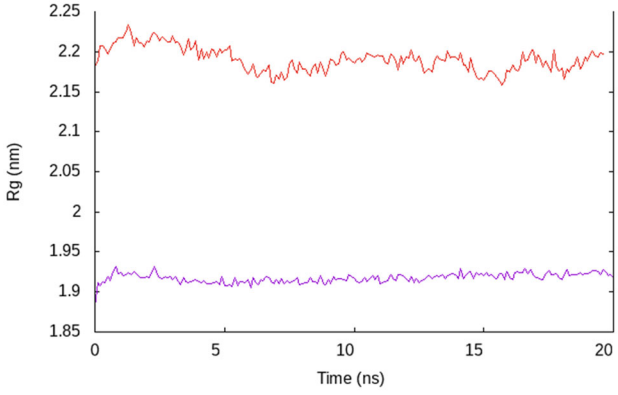

Figure 5. Radius of gyration (Rg) plot of ACE2-RBD (red) and tACE2-RBD complex (violet).

3.7. tACE2 predicted soluble expression in E. coli

Since, there was no PTM site predicted in tACE2, therefore, *E. coli* would be an ideal host for its large scale expression. *E. coli* is the easiest, quickest, and cheapest expression host with a fully known genome, most widely used for hetrologous expression of recombinant protein (Basit et al., 2019). Since, ACE2 is eukaryotic protein; therefore, its expression in its native form in *E. coli* will be uncertain, as most of the eukaryotic protein showed insoluble expression in *E. coli*, which need to be refolded in-vitro (Basit et al., 2018), which is costly and time consuming. That's why, the protein that express in soluble form in *E. coli* are referred as "low hanging fruit", as their bulk production is cost effective and easy to recover (Maqsood et al., 2020). Both sequence and structure based solubility prediction tool using CamSol software predicted expression of intact ACE2 in a completely insoluble form in *E. coli* with intrinsic solubility score −1.027 and complete soluble expression of tACE2 with a solubility score of 1.23. The software generate solubility profile with one score per residue, where regions with scores higher than 1 denote highly soluble regions, while scores lower than −1 showing poorly soluble ones (Figure 6(A, B)). These results

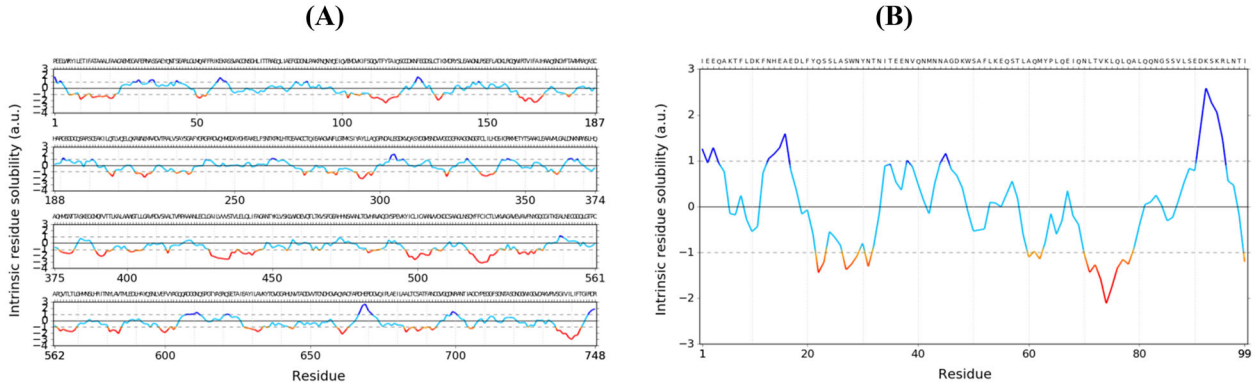

Figure 6. Solubility prediction profile score of **(A)** intact ACE2 showing most of the residues in the window are below 0 (cyan) and −1 (red), while very few residues are above 1 (blue), favors insoluble expression for intact ACE2. **(B)** a large number of residues are between −1 and 1 (cyan), while a window of 7 residues at C-terminal are above 1 (blue), favoring total solubility score above 0, hence soluble expression for tACE2.

propose *E. coli* as a suitable host for soluble expression of tACE2 using pET28a (+) as an expression vector, which favors single step purification.

4. Conclusion

Structure-based rational design of inhibitory protein with enhanced affinities to the SARS-CoV-2 Spike glycoprotein may facilitate development of potential therapeutics. In this study, we have designed a truncated version of human angiotensin converting enzyme 2 as a potential inhibitor of spike glycoprotein. The truncated protein tACE2 was extensively studied through protein-protein docking and MD simulation for binding to RBD of SARS-CoV-2 spike glycoprotein. We found that tACE2 can bind to RBD with a higher binding affinity and form more stabilized complex than the intact ACE2. In addition, the tACE2 sequence predicted soluble expression in *E. coli*, which makes it an easy target for rapid production at large scale for SARS-CoV-2 prevention. We believe that this study narrow down the region of interaction between SARS-CoV-2 S glycoprotein and human ACE2 and paves the way to further enhance the binding affinity between tACE2 and SARS-CoV-2 S glycoprotein through rational design. This will open a new path to COVID-19 treatment.

Disclosure statement

The authors declare no conflict of interest.

References

Aanouz, I., Belhassan, A., El Khatabi, K., Lakhlifi, T., El Idrissi, M., & Bouachrine, M. (2020). Moroccan medicinal plants as inhibitors of COVID-19: Computational investigations. *Journal of Biomolecular Structure and Dynamics*, 1–12. https://doi.org/10.1080/07391102.2020.1758790

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001

Ardalan, N., Mirzaie, S., Sepahi, A. A., & Khavari-Nejad, R. A. (2018). Novel mutant of Escherichia coli asparaginase II to reduction of the glutaminase activity in treatment of acute lymphocytic leukemia by molecular dynamics simulations and QM-MM studies. *Medical Hypotheses*, *112*, 7–17. https://doi.org/10.1016/j.mehy.2018.01.004

Basit, A., & Akhtar, M. W. (2018). Truncation of the processive Cel5A of Thermotoga maritima results in soluble expression and several fold increase in activity. *Biotechnology and Bioengineering*, *115*(7), 1675–1684. https://doi.org/10.1002/bit.26602

Basit, A., Asghar, F., Sadaf, S., & Akhtar, M. W. (2018). Health improvement of human hair and their reshaping using recombinant keratin K31. *Biotechnology Reports (Amsterdam, Netherlands)*, *20*, e00288. https://doi.org/10.1016/j.btre.2018.e00288

Basit, A., Tajwar, R., Sadaf, S., Zhang, Y., & Akhtar, M. W. (2019). Improvement in activity of cellulase Cel12A of Thermotoga neapolitana by error prone PCR. *Journal of Biotechnology*, *306*, 118–124. https://doi.org/10.1016/j.jbiotec.2019.09.011

Boopathi, S., Poma, A. B., & Kolandaivel, P. (2020). Novel 2019 coronavirus structure, mechanism of action, antiviral drug promises and rule out against its treatment. *Journal of Biomolecular Structure and Dynamics*, 1–10. https://doi.org/10.1080/07391102.2020.1758788

Bruno, B. J., Miller, G. D., & Lim, C. S. (2013). Basics and recent advances in peptide and protein drug delivery. *Therapeutic Delivery*, *4*(11), 1443–1467. https://doi.org/10.4155/tde.13.104

Chakraborti, S., Prabakaran, P., Xiao, X., & Dimitrov, D. S. (2005). The SARS coronavirus S glycoprotein receptor binding domain: Fine mapping and functional characterization. *Virology Journal*, *2*(1), 73.

Chen, R., Li, L., & Weng, Z. (2003). ZDOCK: An initial-stage protein-docking algorithm . *Proteins: Structure, Function, and Bioinformatics*, *52*(1), 80–87. https://doi.org/10.1002/prot.10389

Du, L., He, Y., Zhou, Y., Liu, S., Zheng, B.-J., & Jiang, S. (2009). The spike protein of SARS-CoV—a target for vaccine and therapeutic development. *Nature Reviews Microbiology*, *7*(3), 226–236. https://doi.org/10.1038/nrmicro2090

Elmezayen, A. D., Al-Obaidi, A., Şahin, A. T., & Yelekçi, K. (2020). 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*, 1–13. https://doi.org/10.1080/07391102.2020.1758791

Erva, R. R., Rajulapati, S. B., Durthi, C. P., Bhatia, M., & Pola, M. (2016). Molecular dynamic simulations of Escherichia coli L-asparaginase to illuminate its role in deamination of asparagine and glutamine residues. *3 Biotech*, *6*(1), 2https://doi.org/10.1007/s13205-015-0339-9

Fosgerau, K., & Hoffmann, T. (2015). Peptide therapeutics: Current status and future directions. *Drug Discovery Today*, *20*(1), 122–128. https://doi.org/10.1016/j.drudis.2014.10.003

Geng, C., Narasimhan, S., Rodrigues, J. P., & Bonvin, A. M. (2017). Information-driven, ensemble flexible peptide docking using HADDOCK. In *Modeling Peptide-Protein Interactions* (pp. 109–138). Springer.

Gupta, M. K., Vemula, S., Donde, R., Gouda, G., Behera, L., & Vadde, R. (2020). In-silico approaches to detect inhibitors of the human severe acute respiratory syndrome coronavirus envelope protein ion channel.

Journal of Biomolecular Structure and Dynamics, 1–11. https://doi.org/10.1080/07391102.2020.1751300

Gupta, R., Jung, E., & Brunak, S. (2004). *NetNGlyc 1.0 Server*. Center for Biological Sequence Analysis, Technical University of Denmark. http://www.cbs.dtudk/services/NetNGlyc.

Hasan, A., Paray, B. A., Hussain, A., Qadir, F. A., Attar, F., Aziz, F. M., Sharifi, M., Derakhshankhah, H., Rasti, B., Mehrabi, M., Shahpasand, K., Saboury, A. A., & Falahati, M. (2020). A review on the cleavage priming of the spike protein on coronavirus by angiotensin-converting enzyme-2 and furin. *Journal of Biomolecular Structure and Dynamics*, 1–9. https://doi.org/10.1080/07391102.2020.1754293

Haschke, M., Schuster, M., Poglitsch, M., Loibner, H., Salzberg, M., Bruggisser, M., Penninger, J., & Krähenbühl, S. (2013). Pharmacokinetics and pharmacodynamics of recombinant human angiotensin-converting enzyme 2 in healthy human subjects. *Clinical Pharmacokinetics*, *52*(9), 783–792. https://doi.org/10.1007/s40262-013-0072-7

Huang, L., Sexton, D. J., Skogerson, K., Devlin, M., Smith, R., Sanyal, I., Parry, T., Kent, R., Enright, J., Wu, Q-l., Conley, G., DeOliveira, D., Morganelli, L., Ducar, M., Wescott, C. R., & Ladner, R. C. (2003). Novel peptide inhibitors of angiotensin-converting enzyme 2. *Journal of Biological Chemistry*, *278*(18), 15532–15540. https://doi.org/10.1074/jbc.M212934200

Ibrahim, I. M., Abdelmalek, D. H., Elshahat, M. E., & Elfiky, A. A. (2020). COVID-19 spike-host cell receptor GRP78 binding site prediction. *Journal of Infection*, *80*(5), 554–562. https://doi.org/10.1016/j.jinf.2020.02.026

Jenny-Avital, E. R. (2003). Enfuvirtide, an HIV-1 fusion inhibitor. *The New England Journal of Medicine*, *349*(18), 1770.

Johnson, R. J., McCoy, J. G., Bingman, C. A., Phillips, G. N., Jr,., & Raines, R. T. (2007). Inhibition of human pancreatic ribonuclease by the human ribonuclease inhibitor protein. *Journal of Molecular Biology*, *368*(2), 434–449. https://doi.org/10.1016/j.jmb.2007.02.005

Kastritis, P. L., & Bonvin, A. M. (2010). Are scoring functions in protein − - protein docking ready to predict interactomes? Clues from a novel binding affinity benchmark. *Journal of Proteome Research*, *9*(5), 2216–2225. https://doi.org/10.1021/pr9009854

Khan, A., Benthin, C., Zeno, B., Albertson, T. E., Boyd, J., Christie, J. D., Hall, R., Poirier, G., Ronco, J. J., Tidswell, M., Hardes, K., Powley, W. M., Wright, T. J., Siederer, S. K., Fairman, D. A., Lipson, D. A., Bayliffe, A. I., & Lazaar, A. L. (2017). A pilot clinical trial of recombinant human angiotensin-converting enzyme 2 in acute respiratory distress syndrome. *Critical Care (London, England)*, *21*(1), 234. https://doi.org/10.1186/s13054-017-1823-x

Khan, S. A., Zia, K., Ashraf, S., Uddin, R., & Ul-Haq, Z. (2020). Identification of chymotrypsin-like protease inhibitors of SARS-CoV-2 via integrated computational approach. *Journal of Biomolecular Structure and Dynamics*, 1–10. https://doi.org/10.1080/07391102.2020.1751298

Koehler, J. W., Smith, J. M., Ripoll, D. R., Spik, K. W., Taylor, S. L., Badger, C. V., Grant, R. J., Ogg, M. M., Wallqvist, A., Guttieri, M. C., Garry, R. F., & Schmaljohn, C. S. (2013). A fusion-inhibiting peptide against Rift Valley fever virus inhibits multiple, diverse viruses. *PLoS Neglected Tropical Diseases*, *7*(9), e2430. https://doi.org/10.1371/journal.pntd.0002430

Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., Beglov, D., & Vajda, S. (2017). The ClusPro web server for protein–protein docking. *Nature Protocols*, *12*(2), 255–278. https://doi.org/10.1038/nprot.2016.169

Kumar, K., Anbarasu, A., & Ramaiah, S. (2014). Molecular docking and molecular dynamics studies on β-lactamases and penicillin binding proteins. *Molecular Biosystems*, *10*(4), 891–900. https://doi.org/10.1039/C3MB70537D

Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., & Zhang, L. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, 1–9.

Lecker, S. H., Goldberg, A. L., & Mitch, W. E. (2006). Protein degradation by the ubiquitin-proteasome pathway in normal and disease states. *Journal of the American Society of Nephrology: JASN*, *17*(7), 1807–1819. https://doi.org/10.1681/ASN.2006010083

Liu, Y., Wang, M., Xi, J., Luo, F., & Li, A. (2018). PTM-ssMP: A web server for predicting different types of post-translational modification sites using novel site-specific modification profile. *International Journal of Biological Sciences*, *14*(8), 946–956. https://doi.org/10.7150/ijbs.24121

Maqsood, B., Basit, A., Khurshid, M., & Bashir, Q. (2020). Characterization of a thermostable, allosteric L-asparaginase from Anoxybacillus flavithermus. *International Journal of Biological Macromolecules*, *152*, 584–592. https://doi.org/10.1016/j.ijbiomac.2020.02.246

Pant, S., Singh, M., Ravichandiran, V., Murty, U. S. N., & Srivastava, H. K. (2020). Peptide-like and small-molecule inhibitors against Covid-19. *Journal of Biomolecular Structure and Dynamics*, 1–15. https://doi.org/10.1080/07391102.2020.1757510

Pierce, B. G., Wiehe, K., Hwang, H., Kim, B.-H., Vreven, T., & Weng, Z. (2014). ZDOCK server: interactive docking prediction of protein–protein complexes and symmetric multimers. *Bioinformatics (Oxford, England)*, *30*(12), 1771–1773. https://doi.org/10.1093/bioinformatics/btu097

Procko, E. (2020). The sequence of human ACE2 is suboptimal for binding the S spike protein of SARS coronavirus 2. *bioRxiv*, 1–14.

Ramírez, D., & Caballero, J. (2016). Is it reliable to use common molecular docking methods for comparing the binding affinities of enantiomer pairs for their protein target? *International Journal of Molecular Sciences*, *17*(4), 525. https://doi.org/10.3390/ijms17040525

Sarma, P., Sekhar, N., Prajapat, M., Avti, P., Kaur, H., Kumar, S., Singh, S., Kumar, H., Prakash, A., Dhibar, D. P., & Medhi, B. (2020). In-silico homology assisted identification of inhibitor of RNA binding against 2019-nCoV N-protein (N terminal domain). *Journal of Biomolecular Structure and Dynamics*, 1–11. https://doi.org/10.1080/07391102.2020.1753580

Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., & Wolfson, H. J. (2005). PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Research*, *33*(Web Server issue), W363–W367. https://doi.org/10.1093/nar/gki481

Schrodinger, L. (2010). The PyMOL molecular graphics system. Version 1.5.

Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., & Serrano, L. (2005). The FoldX web server: An online force field. *Nucleic Acids Research*, *33*(Web Server issue), W382–W388. https://doi.org/10.1093/nar/gki387

Shu, C., Huang, X., Huang, T., Chen, L., Yao, B., Zhou, J., & Deng, C. (2020). Potential inhibitors for targeting Mpro and Spike of SARS-CoV-2 based on sequence and structural pharmacology analysis. *STEMedicine*, *1*(2), e41–e41. https://doi.org/10.37175/stemedicine.v1i2.41

Sormanni, P., & Vendruscolo, M. (2019). Protein solubility predictions using the CamSol method in the study of protein homeostasis. *Cold Spring Harbor Perspectives in Biology*, *11*(12), a033845. https://doi.org/10.1101/cshperspect.a033845

Steentoft, C., Vakhrushev, S. Y., Joshi, H. J., Kong, Y., Vester-Christensen, M. B., Schjoldager, K. T.-B G., Lavrsen, K., Dabelsteen, S., Pedersen, N. B., Marcos-Silva, L., Gupta, R., Bennett, E. P., Mandel, U., Brunak, S., Wandall, H. H., Levery, S. B., & Clausen, H. (2013). Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *The EMBO Journal*, *32*(10), 1478–1488. https://doi.org/10.1038/emboj.2013.79

Su, M.-G., Weng, J. T.-Y., Hsu, J. B.-K., Huang, K.-Y., Chi, Y.-H., & Lee, T.-Y. (2017). Investigation and identification of functional post-translational modification sites associated with drug binding and protein-protein interactions. *BMC Systems Biology*, *11*(Suppl 7), 132–132. https://doi.org/10.1186/s12918-017-0506-1

Turner, A. J., & Hooper, N. M. (2004). 84 - Angiotensin-converting enzyme 2. In A. J. Barrett, N. D. Rawlings, & J. F. Woessner (Eds.), *Handbook of proteolytic enzymes* (2nd ed., pp. 349–351). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-079611-3.50092-6

Van Durme, J., Delgado, J., Stricher, F., Serrano, L., Schymkowitz, J., & Rousseau, F. (2011). A graphical interface for the FoldX forcefield. *Bioinformatics (Oxford, England)*, *27*(12), 1711–1712. https://doi.org/10.1093/bioinformatics/btr254

Van Zundert, G., Rodrigues, J., Trellet, M., Schmitz, C., Kastritis, P., Karaca, E., Melquiond, A., van Dijk, M., De Vries, S., & Bonvin, A. (2016). The HADDOCK2. 2 web server: User-friendly integrative modeling of

biomolecular complexes. *Journal of Molecular Biology*, *428*(4), 720–725. https://doi.org/10.1016/j.jmb.2015.09.014

Vangone, A., & Bonvin, A. M. (2015). Contacts-based prediction of binding affinity in protein-protein complexes. *elife*, *4*, e07454. https://doi.org/10.7554/eLife.07454

Vangone, A., Rodrigues, J. P. G. L. M., Xue, L. C., van Zundert, G. C. P., Geng, C., Kurkcuoglu, Z., Nellen, M., Narasimhan, S., Karaca, E., van Dijk, M., Melquiond, A. S. J., Visscher, K. M., Trellet, M., Kastritis, P. L., & Bonvin, A. M. J. J. (2017). Sense and simplicity in HADDOCK scoring: Lessons from CASP-CAPRI round 1. *Proteins: Structure, Function, and Bioinformatics*, *85*(3), 417–423. https://doi.org/10.1002/prot.25198

Wahedi, H. M., Ahmad, S., & Abbasi, S. W. (2020). Stilbene-based natural compounds as promising drug candidates against COVID-19. *Journal of Biomolecular Structure and Dynamics*, 1–16. https://doi.org/10.1080/07391102.2020.1762743

Wan, Y., Shang, J., Graham, R., Baric, R. S., & Li, F. (2020a). Receptor recognition by the novel Coronavirus from Wuhan: An analysis based on decade-long structural studies of SARS Coronavirus. *Journal of Virology*, *94*(7), 1–9. https://doi.org/10.1128/JVI.00127-20

Wan, Y., Shang, J., Graham, R., Baric, R. S., & Li, F. (2020b). Receptor recognition by the novel coronavirus from Wuhan: An analysis based on decade-long structural studies of SARS coronavirus. *Journal of Virology*, *94*(7), 1–9. https://doi.org/10.1128/JVI.00127-20

Wójcik, P., & Berlicki, Ł. (2016). Peptide-based inhibitors of protein–protein interactions. *Bioorganic & Medicinal Chemistry Letters*, *26*(3), 707–713. https://doi.org/10.1016/j.bmcl.2015.12.084

Xue, L. C., Rodrigues, J. P., Kastritis, P. L., Bonvin, A. M., & Vangone, A. (2016). PRODIGY: A web server for predicting the binding affinity of protein-protein complexes. *Bioinformatics (Oxford, England)*, *32*(23), 3676–3678. https://doi.org/10.1093/bioinformatics/btw514

Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., & Zhou, Q. (2020). Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science (New York, N.Y.)*, *367*(6485), 1444–1448. https://doi.org/10.1126/science.abb2762

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2015). The I-TASSER Suite: Protein structure and function prediction. *Nature Methods*, *12*(1), 7–8. https://doi.org/10.1038/nmeth.3213

Yuan, M., Wu, N. C., Zhu, X., Lee, C.-C D., So, R. T., Lv, H., Mok, C. K., & Wilson, I. A. (2020). A highly conserved cryptic epitope in the receptor-binding domains of SARS-CoV-2 and SARS-CoV. *Science*, *368*(6491), 630–633. https://doi.org/10.1126/science.abb7269

Zhang, C., Vasmatzis, G., Cornette, J. L., & DeLisi, C. (1997). Determination of atomic desolvation energies from the structures of crystallized proteins. *Journal of Molecular Biology*, *267*(3), 707–726. https://doi.org/10.1006/jmbi.1996.0859

Zhang, G., Pomplun, S., Loftis, A. R., Loas, A., & Pentelute, B. L. (2020). The first-in-class peptide binder to the SARS-CoV-2 spike protein. *bioRxiv*, 1–15. https://doi.org/10.1101/2020.03.19.999318

Zhao, J., Kong, Y., Zhang, F., & Linhardt, R. J. (2018). Impact of temperature on heparin and protein interactions. *Biochemistry & Physiology*, *7*(2), 1–5.