



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Binder design for targeting SARS-CoV-2 spike protein: An in silico perspective

Ali Etemadi^a, Hamid Reza Moradi^a, Farideh Mohammadian^b,
 Mohammad Hossein Karimi-Jafari^c, Babak Negahdari^a, Yazdan Asgari^a,
 Mohammadali Mazloomi^{a,*}

^a Medical Biotechnology Department, School of Advanced Technologies in Medicine, Tehran University of Medical Sciences (TUMS), Tehran, Iran

^b Department of Medical Biotechnology, Faculty of Advanced Medical Sciences, Tabriz University of Medical Sciences, Tabriz, Iran

^c Department of Bioinformatics, Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran

ARTICLE INFO

Keywords:

COVID-19
 Binder design
 SARS
 ACE2
 Pandemic
 Antiviral medications

ABSTRACT

Introduction: The COVID-19 pandemic is now affecting all people around the world and getting worse. New antiviral medications are desperately needed other than the few approved medications that have shown no promising efficacy so far.

Methods: Here we report three blocking binders for targeting SARS-CoV-2 spike protein to block the interaction between the spike protein on the SARS-CoV-2 and the angiotensin-converting enzyme 2 (ACE2) receptors, responsible for viral homing into the alveolar epithelium type II cells (AECII).

Results: The design process is based on the collected natural scaffolds and using Rosetta interface for designing the binders.

Conclusion: Based on the structural analysis, three binders were selected, and the results showed that they might be promising as new therapeutic targets for blocking COVID-19.

1. Introduction

The number of cases worldwide diagnosed with new Coronavirus (SARS-CoV2) is increasing exponentially (Thakur et al., 2020). There are a lot of interventional clinical trials on SARS-CoV-2 (Courtemanche et al., 2020). Drug repurposing is one of the main approaches in combating the coronavirus scourge (dos Santos, 2020). Antiviral, anti-inflammatory, and supportive agents, ACE2 blockers, and convalescent plasma therapy are being tested. Other new treatments such as photobiomodulation are ongoing but still, there is a few adequate and proven medicine to its treatment (Fekrazad, 2020).

Coronaviruses are zoonotic single-stranded RNA viruses and SARS-CoV-2 belongs to the second group of this family. Their genomic sequences are highly similar (especially between SARS-CoV-1 and SARS-CoV-2) therefore their protein structures have striking resemblance. Spike glycoprotein (S protein) of the virus is the key molecule employed for cell penetration by binding to the receptor-binding domain (RBD) of angiotensin-converting enzyme 2(ACE2). It is the only homing receptor recognized for the virus till date. 83% of the alveolar epithelium type II

cells (AECII) express ACE2, and making the respiratory system the main site of injury. The virus not only uses ACE2 as the receptor but also diminishes its protective role against lung damage (Wan et al., 2020). Also, the presence of AEC2 in other organs results in multi-organ dysfunction (M.-Y. Liu et al., 2020). Reports indicate that 76.5% of the amino acid sequence of the S protein is the same between SARS-CoV-1 and SARS-CoV-2, besides their matching 3D structure of RBD in the spike protein. SARS-CoV-2 shows a higher binding affinity than SARS-CoV-1 which can be an explanatory factor for the fast spread of the virus around the world (Xu et al., 2020).

Owing to the designability and selectivity, protein therapeutics show great promise for the future. As intermediate, small proteins (2–20 kDa) fill the gap between small molecules and antibodies due to their lower molecular weight, these small proteins are highly soluble and stable compared to antibodies (140 kDa). However, they have some pitfalls such as lower serum half-life. Considering the advancement in drug delivery methods, it is possible to overcome these problems in the nearest future (Vazquez-Lombardi et al., 2015). With the advent of high-throughput techniques such as *in-silico* design and experimental

* Corresponding author at: Medical Biotechnology, Tehran University of Medical Sciences (TUMS), Tehran, Iran.

E-mail address: m-mazloomi@tums.ac.ir (M. Mazloomi).

evaluation, it is possible to rapidly produce de novo binder proteins in a massively parallel way. This previously was thought to be impossible (Chevalier et al., 2017). Some inventions are being done on the binding of macrocyclic peptides for inhibiting the entry of the virus (Istifli et al., 2020). On the other end of the drug spectrum (W. Liu et al., 2020), *in-silico* prediction has been utilized to assess binding of the phytochemicals (small molecules) such as RBD and cellular proteases binding monoterpenoids (TMRSS2, Cathepsin B, and Cathepsin L) (Wang et al., 2019).

Considering the global demand, finding a target, independent of any post-translational modification that can be expressed in fast-growing hosts with high scalability would be very beneficial. To date, many biopharmaceuticals have been produced in *Escherichia coli* (*E. coli*) (Rosano and Ceccarelli, 2014). *E. coli* is being used as a host for 30% of the approved protein therapeutics due to its easy scale-up, well-established techniques for manipulation and optimization, various genome-scale metabolic models, etc. Therefore, *E. coli* can be a superior host for the production of antivirals against SARS-CoV2 (Yang et al., 2020).

In this study, we aimed at designing some novel binders to block the

spike glycoprotein (S protein) of the SARS-CoV-2. The design process (Fig. 1) was based on collecting a bunch of scaffolds which are natural proteins for redesigning thereby making them modified binders for targeting the S protein.

2. Methods

2.1. Scaffolds selection and relaxation

In the first step, a list of natural proteins (Supplementary T1) was collected using the advanced search in RCSB (Berman et al., 2000) based on the following criteria: 1) resolution below 2 Å, 2) X-ray diffraction method, 3) must be expressed in *E. coli*, 4) sequence length below 100 residues, 5) monomeric proteins, and 6) no DNAs, RNAs, ligands, and mutations in proteins. These scaffolds are natural proteins and were used as initial proteins for making binders using docking, redesign, and other steps.

To remove all clashes in scaffolds and make them more favorable for Rosetta (Nivón et al., 2013), all selected scaffolds were relaxed using the

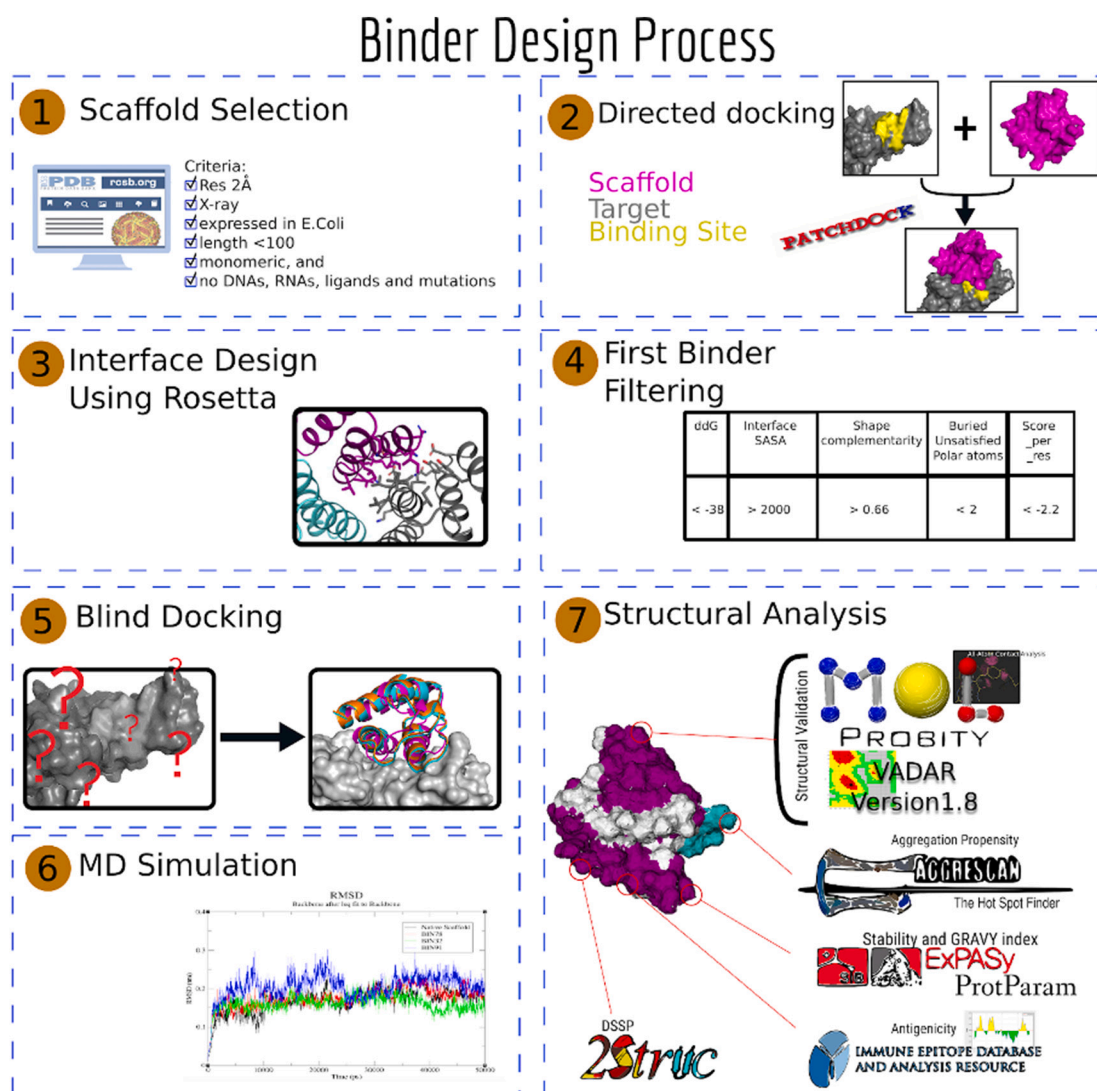


Fig. 1. The figure shows the binder design process for binders design against SARS-coronavirus spike protein. Step 1. Scaffold selection from RCSB. Step 2. Directed docking of all collected scaffolds against the target (SARS-coronavirus spike protein in PDB number 6M0J) using Patchdock. Step 3 and 4. Interface design and Filter designed binders using Rosetta. Step 5. Blind docking of filtered binders against the target using Patchdock and ClusPro to check whether they bind to the target. Step 6. MD simulation of selected binders. Step 7. Structures of all binders were validated by all atom-contact analysis and Ramachandran plots with MolProbity and VADAR respectively and their secondary structures were elucidated by DSSP method using 2Struc. ProtParam was used for evaluating the stability and GRAVY index. IEDB-analysis resources and Aggrescan were utilized respectively for the detection of Antigenicity and Aggregation Hotspots in the sequence of binders.

following script (/path/to/rosetta/main/source/bin/relax.hdf5.linuxgccrelease-database/path/to/rosetta/main/database-IPDBLIST-ignore_unrecognized_res-relax:constrain_relax_to_start_coords-relax:coord_constrain_sidechains-relax:ramp_constraints false-ex1-ex2-use_input_sc-no_his_his_pairE-no_optHfalse-flip_HNQ).

2.2. Initial Docking and Rosetta interface design

We used SARS-CoV-2 spike protein (PDB number 6M0J) (Shang et al., 2020) as a target for docking. Patchdock (Schneidman-Duhovny et al., 2005) was used for docking of all scaffolds against the target. 2000 models per scaffold were generated. Then the FastDesign was used for the interface design of all complexes of scaffolds and the target.

177 designs were selected based on the following scores: ddG < -38, Interface SASA > 2000, Shape complementary > 0.66, Buried Unsatisfied Polar atoms < 2 and score_per_res < -2.2.

2.3. Molecular Dynamics (MD)

Molecular Dynamics (MD) simulation was conducted using the GROMACS package (Berendsen et al., 1995) with an AMBER99SB-ILDN force field and TIP3P water model.

Gromacs was used to analyze the overall RMSD and RMSF after 20 ns MD simulation, and to analyze the protein conformational changes and stability under physiological conditions. 3 trajectories were analyzed for 50 ns, 2 fs per step. Also, MD simulations were done on the original target (6M0J) and promising scaffolds.

A simulation box with at least 1 nm box boundary to protein distance was used for all proteins. Periodic boundary conditions (PBC) and particle mesh Ewald (PME) methods were used for simulations to consider the long-range electrostatic interactions (Darden et al., 1993).

Energy minimization (using the steepest descent method) was done after adding ions (Cl⁻ or Na⁺) to neutralize the total charge, followed by equilibration for 200 ps under the NPT (constant number of particles, pressure, and temperature) ensemble. Position restraints and the temperature were coupled to 310 K using the velocity-rescaling thermostat.

For all the designed binders and also the original scaffold and the target, root-mean-square deviation (RMSD), residue contact maps, the radius of gyration, and minimum distance were analyzed using gmx rms, gmx distance, gmx mdmat, gmx gyrate, and gmx mindist respectively. Based on these criteria the final top 10 designs were selected for 50 ns MD simulation. The gmx distance was used to calculate and measure the distance between selected residues (77–195, 19–244, 22–247, 32–149, 61–219, and 61–235).

The end-state free energy calculations (Average DELTA G binding reported in kcal/mol) of BIN78, BIN32, and BIN91 in the binding states with the SARS-coronavirus spike protein (PDB number 6M0J) were calculated using MMPBSA embedded in AmberTools.

2.4. Docking after binder design

Using patchdock, a type of blind docking was done based on 177 designed binders against the target. For this reason, binding sites were not provided and for each binder, 2000 outputs were checked. The top 10 docked complexes were selected and visually checked to determine which binders were optimal in finding the binding sites of interest. Simultaneously, ClusPro was used to check the binders docking against the target (Kozakov et al., 2017).

Weighted Scores in center representation were used to select the top 10 docked models in ClusPro. Also, the top 10 docked models were selected based on the Patchdock geometric shape complementarity score.

To compare the binding scores of designed binders and SARS coronavirus spike receptor-binding domain against its receptor ACE2, ClusPro was used for docking of the ACE2 and Spike protein and was compared with the binding scores for all binders.

2.5. Structural analysis

All designed binders were examined for their characteristics. The MolProbity, VADAR 1.8 (Willard et al., 2003), and 2Struc (Klose et al., 2010) tools were used for the structural evaluation and DSSP. ExpASY-ProtParam (Gasteiger et al., 2005) was employed for the calculation of stability and the grand average of hydropathy (GRAVY) index. The aggregation potential of the proteins was assessed by Aggrescan (Conchillo-Solé et al., 2007). To forecast antigenic epitopes of the proteins, different algorithms were used from the IEDB web server. Analysis on antigenicity and surface accessibility were performed using Emini methods (Emini et al., 1985) by Kolaskar and Tongaonkar (Kolaskar and Tongaonkar, 1990). Enfuvirtide (T20), an approved antiviral drug, was used for positive control.

Evaluation of structural parameters with MolProbity was performed without any Hydrogen addition and for each PDB file, MolProbity scores were recorded. In VADAR 1.8, the Shrake method (Shrake and Rupley, 1973) was utilized for the values of Van der Waals radii and the definition of polar/nonpolar accessible surface area (ASA) and charged ASA. This method did not consider a uniform radius (1.8 Å) for all side-chain atoms and any hetero-atom attached to carbon was considered polar. Also, the Voronoi procedure was chosen for volume calculation (Richards, 1977). The protein structure coordination files were examined in the main chain, stereo/packing quality, and 3D profile quality indexes (Lüthy et al., 1992). Default thresholds were considered for every IEDB measuring tool (Morris et al., 1992).

Fragment quality application was used in Robetta to pick the best 3- and 9-mer fragments scores and used to predict 3D protein structures based on comparative modeling in Robetta (Kim et al., 2004). PyMol was used for structure visualization (DeLano, 2014).

3. Result

3.1. Docking after binder design

Patchdock and Cluspro were applied for global or blind docking without providing any binding sites. 167 binders from the outputs of RosettaDesign were selected and docked against the target.

Fig. 2 shows the results for all docked designs against the target using ClusPro and Patchdock.

From the top 10 binders selected from both Patchdock and Cluspro, two (3HGL_1_78_2_1_78 or BIN78 and 3HGL_1_45_1_1_32 or BIN32) were found to be the same in both Patchdock and Cluspro outputs and were highly promising. These binders were bound to the same binding site of interest. Also, the binding sites were the same as the site targeted by RosettaDesign. Predicted Pymol RMSD based on the align method for Rosetta binder with BIN78 ClusPro and PatchDock models were 0.263 and 0.393, respectively (Fig. 3A). Both ClusPro and Patchdock predicted that the BIN78 bound to the desired pocket on the target with minimum RMSD.

A closer look at BIN78 (Fig. 3B), revealed that two main helices interacted with the main groove on the target. There were two π - π interactions between the BIN78 and the target: F28 with F202/236 and also H7 with Y251. Ddg scores for BIN32 and BIN78 were -43.998 and -41.691. Table 1 shows some predicted scores for all selected binders.

To check whether our designed binders had enough binding scores, we compared the weighted scores in ClusPro for designed binders BIN78, BIN32, BIN91, and SARS coronavirus spike receptor-binding domain complexed with its receptor (PDB number 6M0J). The weighted scores for BIN78, BIN32, BIN91 were -747.3, -814, and -771.8, respectively. However, this score for the complex of S protein and ACE2 was -719.6.

3.2. Web logo and structure prediction

The bite signals or highly frequent residues in all sequences were 30

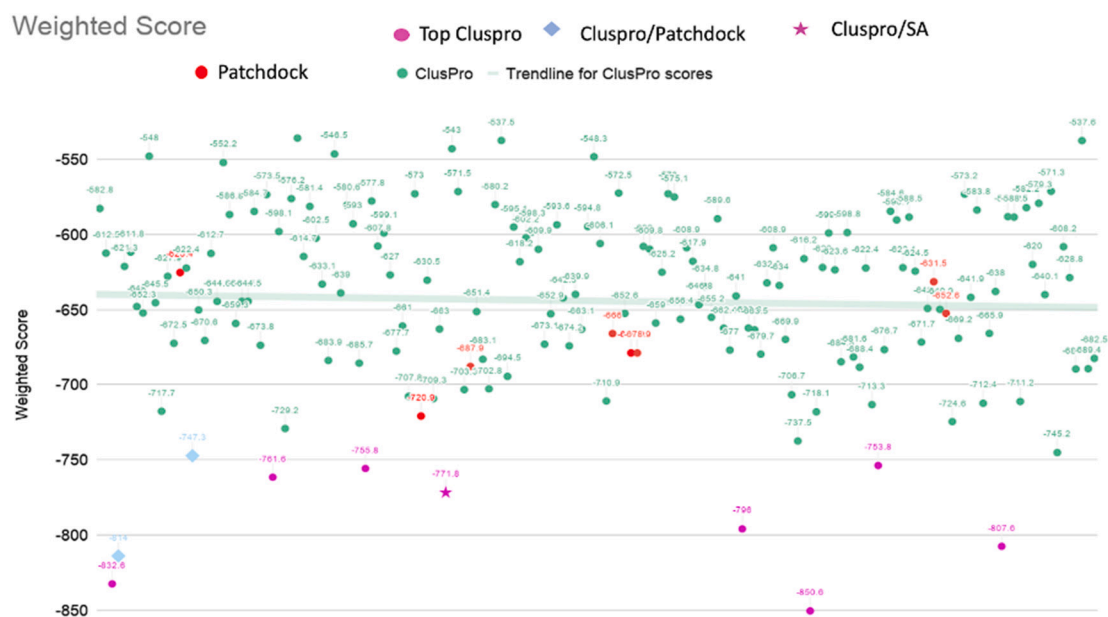


Fig. 2. Weighted core for all docked binders against SARS-coronavirus spike protein (PDB number 6M0J). Figure shows the weighted scores for all docked binders using Cluspro (green), top patchdock (Red). Top cluspro models were shown in Magenta. The models which were selected using both Patchdock and Cluspro also were shown using blue diamonds. Also, the models with high structural analysis and cluspro scores were shown using stars.

residues (out of 78 residues) with 4 having the maximum bite signals: 1, 6, 9, 10, 13, 17, 18, 19, 23, 27, 30, 33, 38, 39, 42, 45, 46, 50, 51, 53, 54, 56, 57, 58, 65, 68, 70, 71, 72 and 75. Fig. 4 shows the Weblogo profile of all selected binders.

Rosetta was used for the comparative modeling of all selected binders. The results of the modeling showed that BIN91 was highly promising in predicting the 3-dimensional structure. The RMSD of BIN32, BIN91, and BIN78 with native scaffolds (3HGL) were 2.81, 1.97, and 2.88 Å, respectively (Fig. 5).

3.3. MD analysis

To investigate the conformational behavior of the binder's models and native scaffold, MD simulation was used during the 50 ns simulation (Fig. 6). The minimum and maximum RMSD score for native scaffolds and the binders were close and the traces for the binders were followed by the traces for native scaffolds, although the RMSD for BIN91 was a little different from the others.

The radius of gyration (Rg) was analyzed to check protein structure compactness and stability of the binders, and also the target and native scaffolds. The Rg values for the binders in the binding form with the target were similar and larger than the target and the native scaffold alone. The contact maps for all three binders also showed the similarity between the binders. Likewise, minimum distances between the atoms in all binders and the native scaffold were analyzed and showed that the binders and native scaffolds were close to each other.

The results for distances between the specified residues (77–195, 19–244, 22–247, 32–149, 61–219, and 61–235) showed that the average distances between these residues were 2.2 nm, 0.9 nm, and 1.2 nm for BIN78, BIN32, and BIN91, respectively.

MMPBSA in AmberTools showed that the binding energy of BIN32, BIN78 and BIN91 in binding states with the SARS-coronavirus spike protein (PDB number 6M0J) were -22.43 , -20.91 and -17.01 , respectively (Average DELTA G binding reported in kcal/mol).

3.4. Structural analysis

Structural analysis was done on all 177 designed proteins. Since there was no perfection, each protein had some pros and cons.

Therefore, by comparing their characteristics, with each other and with the main 3HGL protein, the 10 top proteins with enhanced or comparable parameters were selected (Table 2). Among them, Bin91 had the highest score in the docking experiments. Additionally, two modified proteins, Bin78 and Bin32, had a good performance in both Cluspro and Patchdock sections. For positive control, a Human Immunodeficiency Virus (HIV) entry inhibitor, Enfuvirtide (T20) was used. It has 36 amino acids with a helical secondary structure and a molecular weight of around 4.5 kDa. With the Trademark name of Fuzeon™, Enfuvirtide is considered the first small protein antiviral drug in the market (De Clercq and Li, 2016).

All binders comprised of 78 amino acids. The analysis based on ProtParam indicated that the designed proteins' molecular weight was around 8500 Da. The results from 2Struc indicated 71.8% of the 3HGL is composed of alpha-helix, with 9% as helix-turn, and 19.2% as loops. All three top binders showed an increment in the proportion of helix-turn secondary structure. BIN91 changed to 14.1% and BIN32, 78 to 15.4%. To compensate for this, the alpha-helices decreased to 66.7% for BIN91 and 65.4% for BIN32, and BIN78. Enfuvirtide is composed of 90.9% alpha-helix and 3% helix-turn. Coils constitute the remainder of the structure.

One of the factors employed in classifying these candidates was the MolProbity score which is based on the clash score and geometrical parameters. It was normalized to mirror a crystallographic resolution. The main 3HGL protein had a score of 0.76, and a mild clash of <40 Å, with 6 bad bonds detected. Moreover, this model had a Ramachandran and Rotamer value in the allowed region. The chosen candidate proteins all had lower MolProbity scores which indicated a high quality from a structural standpoint of the model (Fig. 7A). The three top candidates had a MolProbity score of 0.5 which demonstrated high quality in their structure. The MolProbity results suggested that there were no preliminary geometrical problems with the binders, but this does not imply that the overall structure was flawless. To augment the scrutiny on assessing the quality of the structures, VADAR was implemented. It calculated more than 30 structural parameters and presented graphs and tables. In most of the structures, VADAR algorithms detected at least 1 residue that had 1 possible problem in the main chain and Stereo/Packing quality index. This might have resulted from a coordination defect or a systemic problem such as the wrong choice of the standard

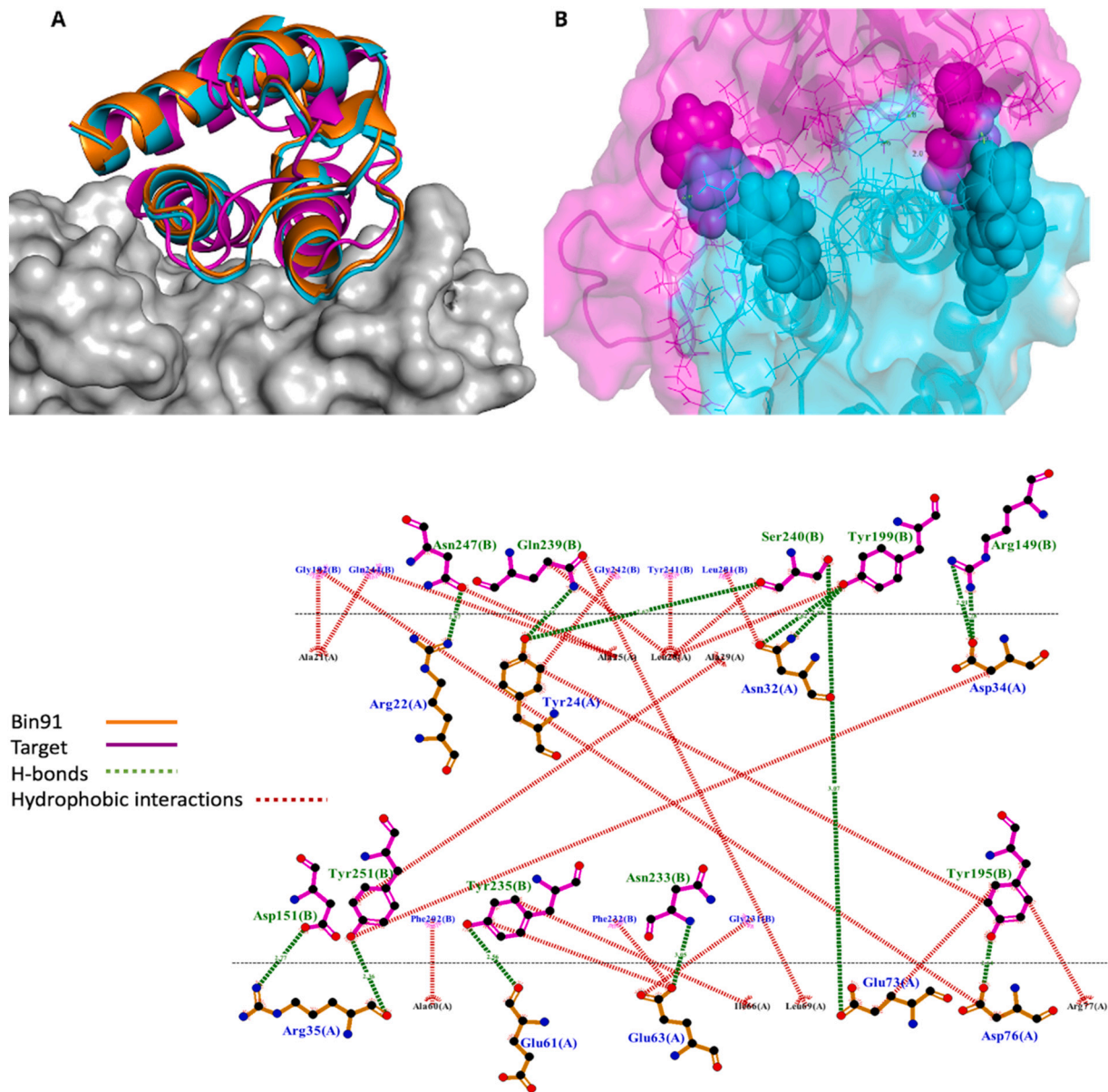


Fig. 3. (A) Superposed models for BIN78 binder. SARS-coronavirus spike protein (PDB number 6M0J) was shown in grey. The figure shows that superposed models for BIN78 binders from Rosetta (Orange), Cluspro (Magenta) and PatchDock (Cyan) were almost identical. (B) Interaction between Bin78 and SARS-coronavirus spike protein (PDB number 6M0J). The BIN78 and the target were shown in Cyan and Magenta, respectively. H bond interactions were shown with a dashed-yellow line. Also, π - π interactions were shown in spheres mode. Lower panel: binding mode and interacting residues are shown for the binder and the target.

Table 1

Rosetta predicted scores for Bin78, Bin32 and Bin91.

Design name	ddG	Interface SASA	Shape complementarity	Buried Unsatisfied Polar atoms	score_per_res	MMPBSA ^a	Van der Waals energy	Electrostatic energy
Bin78	-41.691	2004	0.666	0	-2.381	-20.91	-209.93 kJ/mol	-24.54 kJ/mol
Bin32	-43.998	1966	0.663	1	-2.219	-22.43	-154.74 kJ/mol	-42.71 kJ/mol
Bin91	-37.206	1701	0.673	2	-2.292	-17.01	-164.15 kJ/mol	-93.58 kJ/mol

^a MMPBSA was predicted by AmberTools.

(Richards, 1974). None of the structures in the candidate pool had a 3D profile quality above 0. Inferentially, this implied there was no problem in the local environment, packing, and hydrophobic energy of the structures. Among the three final candidates, BIN78 had a better stereo/packing quality index and overall better performance in this section. Enfuvirtide PDB structure as positive control showed a low MolProbity score and many problems were detected with its 3D profile quality

index. Ramachandran plots produced by VADAR in Fig. 8 indicate that, for BIN32 and BIN91, 92% of ϕ and ψ angles are in the core region, and 7% are in the allowed region. Additionally, 93% of BIN78 ϕ and ψ angles are in the core region, and 5% are located in the allowed region. For Enfuvirtide, 90% of residue angles are in the core region, and only 9% belong to the allowed region. No ϕ and ψ angle outliers were detected in any structure.

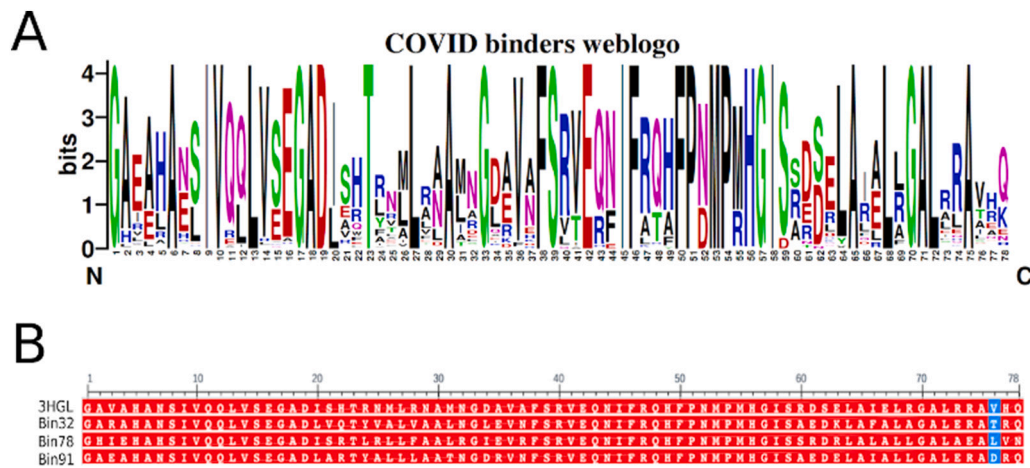


Fig. 4. (A) Weblogo profile of all selected binders. It showed 30 residues with maximum 4 bit signal: 1, 6, 9, 10, 13, 17, 18,19, 23, 27, 30, 33, 38, 39, 42, 45, 46, 50, 51, 53, 54, 56, 57, 58, 65, 68, 70, 71, 72 and 75. These residues were almost unique in all binders. (B) Multiple sequence alignment of the top Binders and 3HGL (produced by COBALT webserver).

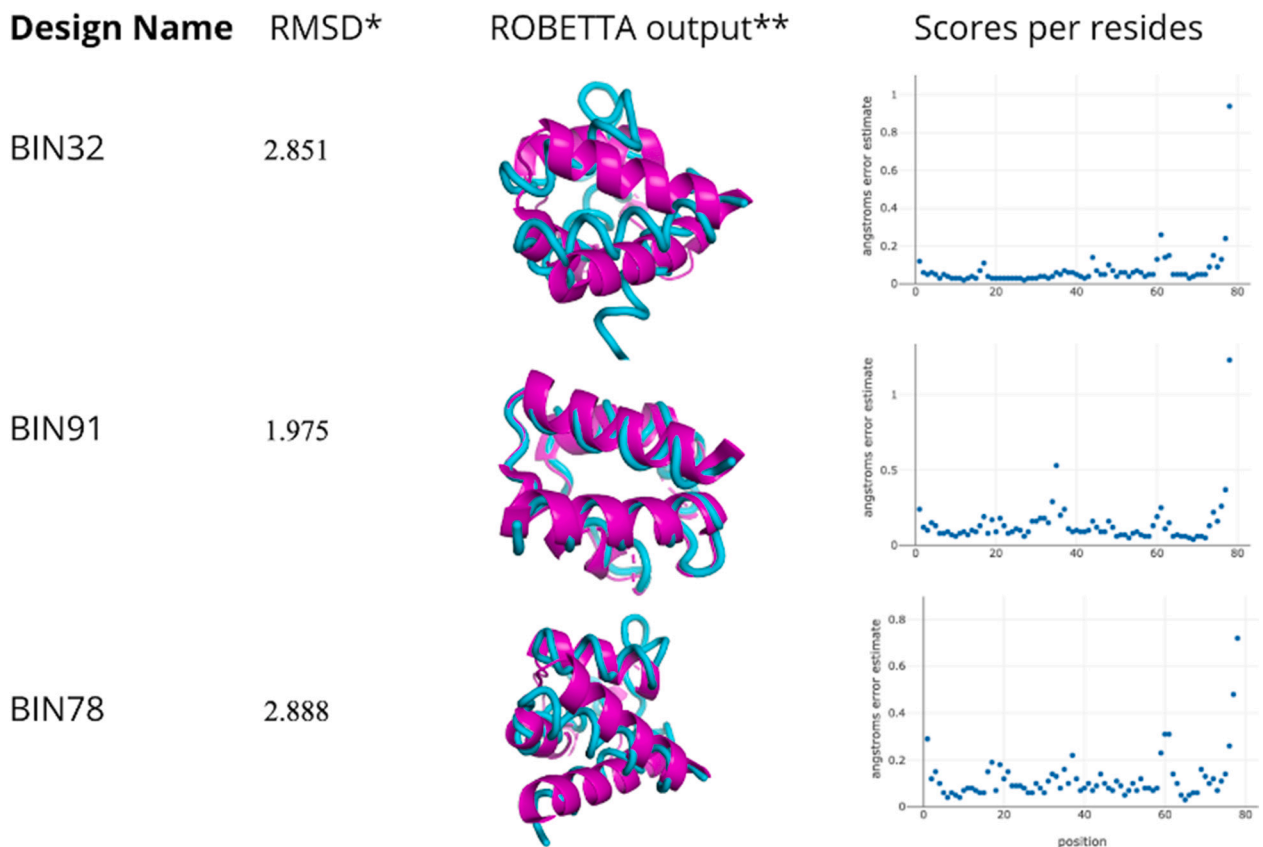


Fig. 5. Comparative modeling using ROBETTA. *RMSD is for aligning the design model on the native scaffold (3HGL). **The target was shown in Magenta and binders were illustrated in Cyan.

A calculated parameter in ProtParam was the GRAVY score. Kyte & Doolittle calculated the GRAVY score for different soluble and membrane proteins. For this purpose, they summed up all the amino acid hydrophathy scores in a sequence and divided the result by the number of residues. It can be inferred from their data that there is a high probability for proteins with a negative value of GRAVY score to be soluble, and proteins with a positive value to be membranous (Kyte and Doolittle, 1982). As shown in Fig. 7C, the GRAVY index value of a large number of proteins is at the boundary and above 0. The GRAVY score of

top binders and controls are represented in Table 2. BIN91 with a GRAVY score of -0.168 is likely the most soluble protein among the top 3 candidates. The other parameter that was taken from ProtParam was the instability index. Grusprad et al. proposed that the stability and instability of a protein depend on the arrangement of amino acids in a specific order. It was shown that there are some dipeptides (the smallest unit of order) in the protein's sequence which defines the stability (Feng, 2020). The instability index of all designed proteins is plotted in Fig. 7D. Although it has been stated that proteins with an instability index below

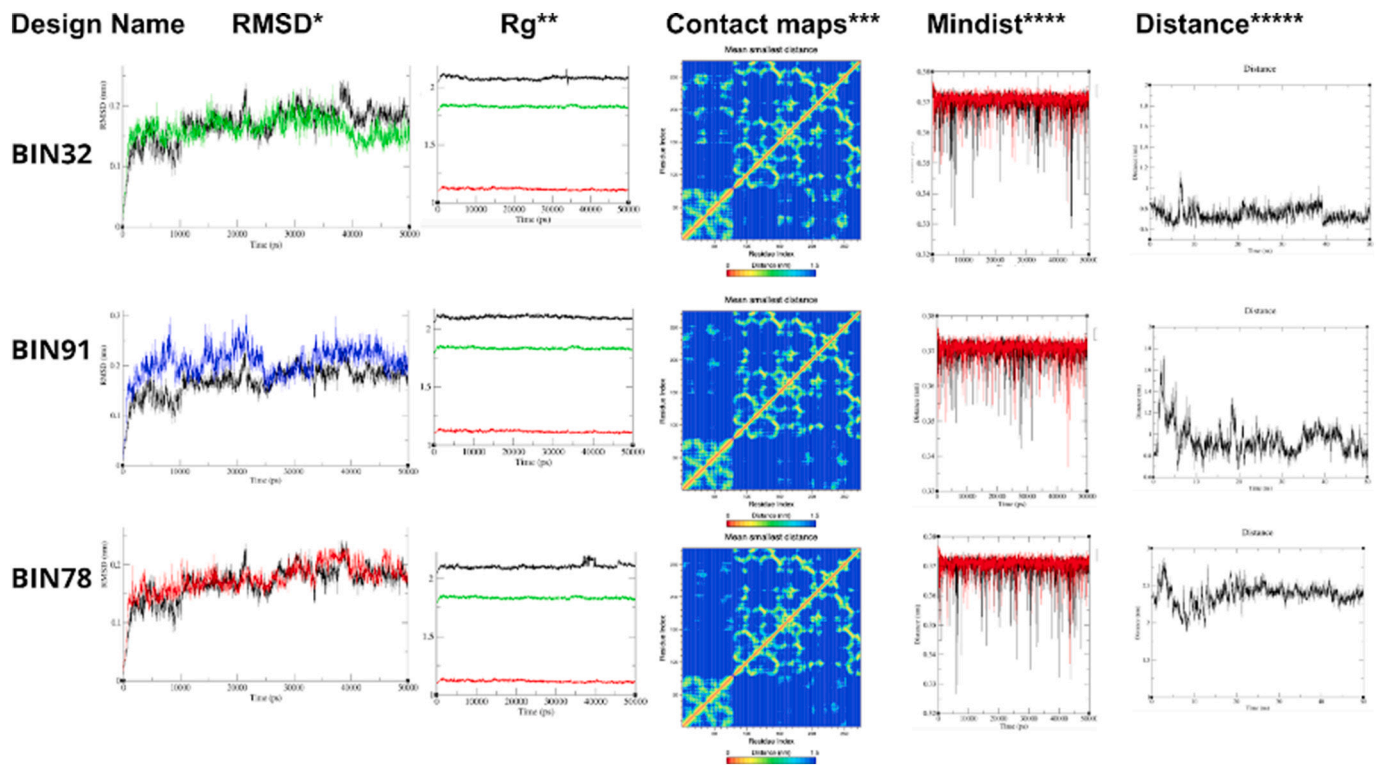


Fig. 6. Molecular dynamic simulation analysis. *For RMSD the traces for the native scaffold(3HGL), BIN32, BIN91, and BIN78 were shown in Black, green, blue and red, respectively. **For Radius of gyration (Rg), the binders, the target, and native scaffold were colored in black, green and red, respectively. ***Contacts maps show that there are some common area between the target (big upper square) and the binders (small lower square) which are showing the interaction between the target and the binders. ****For minimum distance (mindist) the red and black traces are corresponding to the native scaffold and the binders. *****The results for distances between the specified residues showed that the average distances between these residues were 2.2 nm, 0.9 and 1.2 for BIN78, BIN32 and BIN91, respectively.

Table 2

Analysis results from MolProbity, VADAR, ProtParam, IEDB, and Aggrescan for the main 3HGL protein and 10 top Binder candidates based on structural analysis and two additional binders based on Docking results. Enfuvirtide is used as the positive control.

Design name	MolProbity			VADAR		ProtParam		Aggrescan		
	Ramachandran favored (Goal > 98%)	Favored Rotamer (Goal > 98%)	MolProbity score	Stereo/packing quality index (α)	Main chain possible problem (β)	Instability index (γ)	GRAVY index (δ)	Na4vSS	THASr	NnHS
3HGL	98.68%	98.41%	0.76	2	2V,1A	59.03	-0.321	-16.2	0.01	1.282
3HGL089_37	98.68%	98.41%	0.5	2	2V,1O	49.19	-0.423	-16.9	0.048	2.564
3HGL089_114	100%	100%	0.5	2	3V,1O	57.88	-0.274	-12.2	0.058	2.564
3HGL178_1_126	100%	100%	0.5	2	3V	32	-0.141	-11.4	0.071	2.564
3HGL178_1_64	100%	100%	0.5	2	4V,1O	36.82	-0.205	-13.9	0.062	2.564
3HGL178_2_61	100%	100%	0.5	2	2V	29.58	-0.146	-10.1	0.087	2.564
3HGL178_2_88	100%	100%	0.5	2	2V	35.68	-0.163	-11.1	0.073	2.564
3HGL192_59	100%	98.36%	0.5	1	None	48.93	-0.153	-12.3	0.091	2.564
3HGL209_2_88	100%	100%	0.5	1	3V	52.53	-0.201	-11.2	0.025	2.564
3HGL214_144	98.68%	100%	0.5	2	2V,1O,1A	47.85	-0.431	-18.5	0.035	1.282
3HGL178_1_91 (Bin91)	100%	100%	0.5	2	3V	31.22	-0.168	-13.2	0.095	2.564
3HGL045_1_32 (Bin32)	100%	100%	0.5	3	4V,1O	30.96	0.021	-4	0.116	3.846
3HGL078_2_78 (bin78)	100%	100%	0.5	1	2V	58.28	0.119	1.7	0.13	2.564
Enfuvirtide (+control)	90.32%	89.66%	1.5	3	3V	62.65	-0.875	-20.5	0.174	5.556

(α) Indicates Possible Problems in the Protein by assigning a score to these three criteria: Torsion angle, Omega angle and Van der Waals radii.

(β) O - indicates possible problem with Omega angle, V - indicates possible problem with fractional volume, A - indicates possible problem with fractional ASA.

(γ) Instability index below 40 is considered as stable.

(δ) GRAVY index below 0 are more likely globular above 0 are more likely membranous.

40 are considered to be stable, the main 3HGL protein had an instability index of 59.03..

Another important aspect of choosing these candidates is the

aggregation potential of the molecules. For this purpose, the normalized parameters such as global aggregation propensity (Na4vSS), Total Hotspot Area per residue (THASr), and the number of Hotspot areas

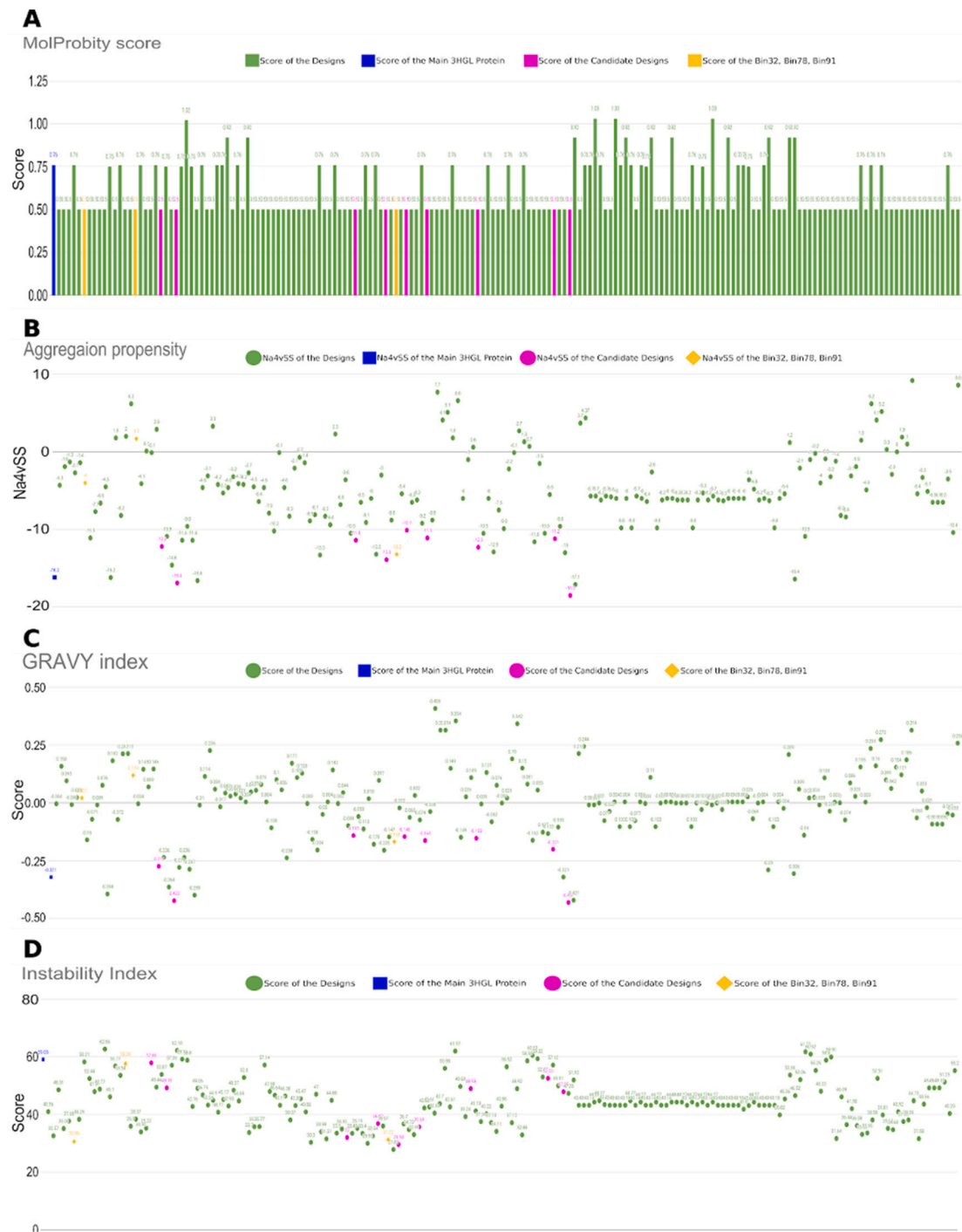


Fig. 7. Showing predicted scores for all binders, the resulted top ten binders are shown in magenta and BIN32, 78, 91 are in yellow (A) MolProbity scores from MolProbity webserver (B) Na4vSS which is obtained from Aggrescan (C, D) GRAVY and Instability index from ExPASy-ProtParam.

(NnHS) were chosen. It permitted direct comparison between binders and the positive control. The main 3HGL had -16 Na4vSS, which was one of the lowest among the proteins indicating its low aggregating potential (Fig. 7B). Considering the data evaluated from the analysis of datasets by the original article, proteins with less than -5.18 for Na4vSS, 0.9 THSAr, and 3.35 NnHS parameters remained soluble when they are over-expressed in *E. coli*. Hence, the designed protein pool was filtered by these criteria (Kolaskar and Tongaonkar, 1990). The top ten selected protein designs all had a value below -10 for Na4vSS, except BIN32 and BIN78 for which a different selection procedure was performed (Table 2). BIN97 and BIN78 had one, while BIN32 had two

additional numbers of hotspots in the same range of sequence compared to the main 3HGL protein (Fig. 8). Compared to Enfuvirtide, all three top designs had an advantage in THSAr and NnHS, but when Na4vSS was considered, Enfuvirtide had a low value (-20), indicating a low aggregation propensity of this small protein.

The final major selection parameter was the antigenicity of the proteins. The average, minimum, and maximum scores based on Kolaskar and Tongaonkar and Emini methods are presented in Table 3. Kolaskar and Tongaonkar with 75% accuracy, is a semi-empirical method for locating the antigenic determinants in the sequence. It uses the physicochemical properties of residues and their frequencies in

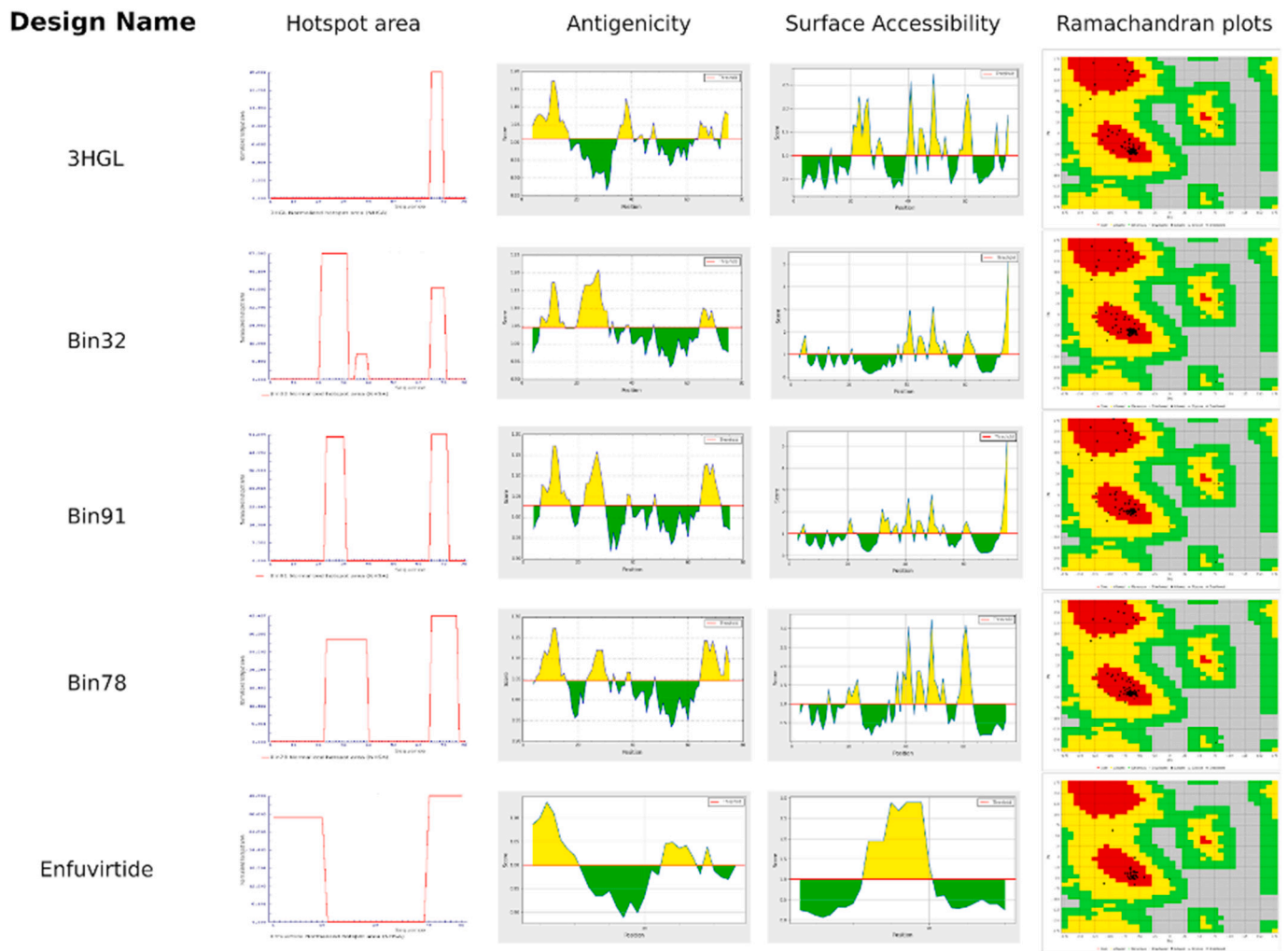


Fig. 8. Comparative analysis of hotspot area, antigenicity, accessibility of surface and Ramachandran plots for the top Binders and the controls. Hotspot areas represent the regions which are prone to aggregation. The green areas in Antigenicity and Surface accessibility plots represent sequences which are not Antigenic and these regions are not located at the surface. Any sequence above the threshold which is colored in yellow, indicates a positive result for Antigenicity and surface accessibility. These plots are obtained via Aggrescan, IEDB and VADAR web servers.

Table 3

IEDB measuring tools scores. Average, minimum and maximum scores of Kolaskar & Tongaonkar and Emini for assessing the Antigenicity and Surface accessibility respectively is shown here.

Design name	Kolaskar & Tongaonkar			Emini		
	Minimum	Average	Maximum	Minimum	Average	Maximum
3HGL	0.865	1.01	1.173	0.262	1	2.722
3HGL178_1_91 (Bin91)	0.918	1.028	1.173	0.112	1	5.394
3HGL178_1_32 (Bin32)	0.934	1.046	1.208	0.147	1	5.262
3HGL178_1_78 (Bin78)	0.934	1.046	1.173	0.154	1	3.233
Enfuvirtide (+control)	0.888	0.998	1.133	0.061	1	2.889

experimentally determined epitopes (Emini et al., 1985). As Fig. 8 shows, after the modifications, all three top binders retained the first position of antigenic determinant which starts from around the 7th to 15th amino acid position. They also had another position adjacent to the first one, which starts from around the 21st to the 30th residue. All three top binders also had a third spot which was above the threshold, but it was not predicted to be antigenic in BIN78 by the webserver's algorithm. The calculation was based on the surface accessibility scale on a product instead of addition within the window. The Emini method used the formula $S_n = (n + 4 + i) (0.37)^{-6}$ which gave the probability for a hexapeptide sequence to be on the surface. S_n greater than 1.0 shows a

higher probability for a sequence to reside on the surface of the protein (Shrake and Rupley, 1973). Results from Emini demonstrated that most parts of the antigenic sequences were not exposed at the surface. BIN91 had the least average (1.028) score between the three top binders, while the other two both had an average score of 1.046. The two ends of the Enfuvirtide sequence were the most antigenic parts but the Emini method did not consider them as exposed to the surface (Fig. 8).

4. Discussion

To date, no drugs have been approved for the treatment of the

COVID-19. Remdesivir is the sole antiviral drug that is recommended for patients with mild disease. However, it is on limited supply (Dang and Dang, 2020), and the attainment of herd immunity by vaccination may take at least a year while the infection rate of SARS-CoV-2 keeps increasing rapidly. Despite this, the effective range of the vaccines remains inconclusive based on some reports (Feng, 2020).

AvrPtoB a *Pseudomonas syringae* pv tomato multi-domain effector protein, which plays a role in interrupting the immune responses of tomato plants by inhibiting the programmed cell death and promotes disease (Feng and Cheng, 2020). To elucidate the crystal structure, Dong expressed a portion of this protein (AvrPtoB121–205 Or 3HGL) that was stable and sufficient for the interaction with its target in *E. coli* (Oh and Martin, 2011). In this study, 3HGL was used as a scaffold with the interface of its surfaces redesigned using RosettaDesign (Dong et al., 2009). This method was applied by different groups for making diagnostic and therapeutic proteins (Liu and Chen, 2016). Targeting the S protein with inhibitors is one of the prominent ways of developing therapeutics (Willis et al., 2015). Not only do they inhibit the attachment of RBD to ACE2 and halts the entry of the virus, but they also preserve the ACE2. Consequently, they can function as regulators of the renin-angiotensin-aldosterone system, and protect the lung from injury by their protease function against angiotensin (Ang) 1. By attaching to the products of the ACE2, Ang 1–7, and Mas pathway they exert their anti-inflammatory, and anti-proliferative effects. Furthermore, they increase oxygenation and reduce the hypersensitivity of the airway and downregulate the apoptosis of alveolar epithelial cells. However, ACE2 is not the only enzyme that participates in the production of angiotensin. A week later, the level of this peptide gets back to normal, but in the short run, it could be quite damaging (Samavati and Uhal, 2020).

The results of ClusPro docking scored for selected binders and also spike protein against the receptor ACE2, showed that our binders are promising and might be useful for targeting the virus as it bound to the target stronger than the spike protein.

As can be seen from Fig. 5B, the substitution of caline (residue 76 in 3HGL) with aspartate (in BIN91) or threonine (in BIN32) increased the stability of the protein which may partly be due to the arginine in the 77th position of the sequence. In BIN78 substitution of valine with leucine (a methyl added as the bridge in side chain) did not change the instability index (58.28) considerably. Indicating the limited half-life of this binder. Guruprasad's method, however, does not consider the higher-order factors that may affect the results (Guruprasad et al., 1990). Buried charges may impact the folding of proteins and their stability because it is not so energetically favorable to bury a charge inside the protein structure.

There are some regions of the binders that are considered antigenic based on their sequence. Apart from the structural standpoint, other factors may contribute to the immunogenicity of the proteins like administration dose, excipient formulation, host proteins, patients' immune/genetic background, and the aggregation of the proteins (van Beers and Bardor, 2012). The Aggrescan webserver predicted that BIN91 had the least aggregating propensity, but these predictions were based on the sequence and the folded protein may take a structure that increases or decreases the aggregation potential. Hiding some of the regions during the folding may vary the characteristics of the protein. That notwithstanding, it is possible to address such challenges with drug delivery methods (Bodier-Montagutelli et al., 2018).

5. Conclusion

Since SARS-CoV-2 inflicts a respiratory disease, the strategy of inhaled protein therapeutics could be of help. However, there are several questions to consider for this purpose such as formulation, stability, and the characterization of proteins in the droplets.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.genrep.2021.101452>.

Abbreviations

ACE2	Angiotensin-converting enzyme 2
RBD	Receptor Binding Domain
S protein	spike glycoprotein
AECII	alveolar epithelium type II cells
RMSD	root-mean-square deviation
RMSF	root-mean-square fluctuation
NPT	constant number of particles, pressure, and temperature
PBC	periodic boundary conditions
PME	particle mesh Ewald
MD	Molecular Dynamics
GRAVY index	grand average of hydropathy index
VADAR	Volume Area Dihedral Angle Reporter
IEDB	Immune-Epitope Database
ASA	accessible surface area
DSSP	Dictionary of Secondary Structure of Proteins
Na4vSS	Normalized amino acid aggregation-propensity value, window average Sequence Sum
THSAr	Total Hot Spot Area per residue
NnHS	Normalized number of Hotspots

CRedit authorship contribution statement

Ali Etemadi: Conceptualization, Data curation, Formal analysis. **Hamid Reza Moradi:** Resources, Software, Writing – review & editing. **Farideh Mohammadian:** Resources, Software. **Mohammad Hossein Karimi-Jafari:** Resources, Software. **Babak Negahdari:** Funding acquisition, Writing – original draft, Writing – review & editing. **Yazdan Asgari:** Writing – original draft, Writing – review & editing. **Mohammadali Mazloomi:** Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

None.

Acknowledgements

Declared none.

Ethics approval and consent to participate

Not applicable.

Human and animal rights

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All relevant data and materials are provided within the manuscript.

Funding

Tehran University of Medical Sciences, Grant/Award Number: 99-2-148-49874.

References

- Berendsen, H.J., van der Spoel, D., van Drunen, R., 1995. GROMACS: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* 91 (1–3), 43–56.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., et al., 2000. The protein data bank. *Nucleic Acids Res.* 28 (1), 235–242.
- Bodier-Montagutelli, E., Mayor, A., Vecellio, L., Respaud, R., Heuzé-Vourc'h, N., 2018. Designing inhaled protein therapeutics for topical lung delivery: what are the next steps? Taylor & Francis.
- Chevalier, A., Silva, D.-A., Rocklin, G.J., Hicks, D.R., Vergara, R., Murapa, P., et al., 2017. Massively parallel de novo protein design for targeted therapeutics. *Nature* 550 (7674), 74–79.
- Conchillo-Solé, O., de Groot, N.S., Avilés, F.X., Vendrell, J., Daura, X., Ventura, S., 2007. AGGRESKAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. *BMC Bioinformatics* 8 (1), 65.
- Courtemanche, C., Garuccio, J., Le, A., Pinkston, J., Yelowitz, A., 2020. Strong social distancing measures in the United States reduced the COVID-19 growth rate: study evaluates the impact of social distancing measures on the growth rate of confirmed COVID-19 cases across the United States. *Health Affairs*. <https://doi.org/10.1377/hlthaff.2020.00608>.
- Dang, A., Dang, S., 2020. Hydroxychloroquine and Remdesivir in COVID-19: A Critical Analysis of Recent Events.
- Darden, T., York, D., Pedersen, L., 1993. Particle mesh Ewald: an $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* 98 (12), 10089–10092.
- De Clercq, E., Li, G., 2016. Approved antiviral drugs over the past 50 years. *Clin. Microbiol. Rev.* 29 (3), 695–747.
- DeLano, W., 2014. The PyMOL Molecular Graphics System, Version 1.8. Schrödinger LLC. There is no corresponding record for this reference [Google Scholar].
- Dong, J., Xiao, F., Fan, F., Gu, L., Cang, H., Martin, G.B., et al., 2009. Crystal structure of the complex between *Pseudomonas* effector AvrPtoB and the tomato Pto kinase reveals both a shared and a unique interface compared with AvrPto-Pto. *The Plant Cell*. 21 (6), 1846–1859.
- dos Santos, W.G., 2020. Natural history of COVID-19 and current knowledge on treatment therapeutic options. *Biomed. Pharmacother.* 129, 110493. <https://doi.org/10.1016/j.biopha.2020.110493>. In this issue.
- Emini, E.A., Hughes, J.V., Perlow, D., Boger, J., 1985. Induction of hepatitis a virus-neutralizing antibody by a virus-specific synthetic peptide. *J. Virol.* 55 (3), 836–839.
- Fekrazad, R., 2020. Photobiomodulation and antiviral photodynamic therapy as a possible novel approach in COVID-19 management. Mary Ann Liebert, Inc., Publishers, 140 Huguenot Street, 3rd Floor New.
- Feng, E., 2020. Some Recovered Coronavirus Patients In Wuhan Are Testing Positive Again. NPR Goats and Soda. Mar 27.
- Feng, E., Cheng, A., 2020. Mystery in Wuhan: recovered coronavirus patients test negative then positive. In: National Public Radio: News and Analysis.
- Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein identification and analysis tools on the ExPASy server. In: *The proteomics protocols handbook*. Springer, pp. 571–607.
- Guruprasad, K., Reddy, B.B., Pandit, M.W., 1990. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Eng. Des. Sel.* 4 (2), 155–161.
- Istifi, E.S., Tepe, A.Ş., Sarikürkcü, C., Tepe, B., 2020. Interaction of certain monoterpene hydrocarbons with the receptor binding domain of 2019 novel coronavirus (2019-nCoV), transmembrane serine protease 2 (TMPRSS2), cathepsin B, and cathepsin L (CatB/L) and their pharmacokinetic properties. *Turk. J. Biol.* 44 (3), 242.
- Kim, D.E., Chivian, D., Baker, D., 2004. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* 32 (suppl_2), W526–W531.
- Klose, D., Wallace, B.A., Janes, R.W., 2010. 2Struc: the secondary structure server. *Bioinformatics* 26 (20), 2624–2625.
- Kolaskar, A., Tongaonkar, P.C., 1990. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett.* 276 (1–2), 172–174.
- Kozakov, D., Hall, D.R., Xia, B., Porter, K.A., Padjhony, D., Yueh, C., et al., 2017. The ClusPro web server for protein–protein docking. *Nat. Protoc.* 12 (2), 255.
- Kyte, J., Doolittle, R.F., 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157 (1), 105–132.
- Liu, H., Chen, Q., 2016. Computational protein design for given backbone: recent progresses in general method-related aspects. *Curr. Opin. Struct. Biol.* 39, 89–95.
- Liu, M.-Y., Zheng, B., Zhang, Y., Li, J.-P., 2020. Role and mechanism of angiotensin-converting enzyme 2 in acute lung injury in coronavirus disease 2019. In: *Chronic Diseases and Translational Medicine*.
- Liu, W., Morse, J.S., Lalonde, T., Xu, S., 2020. Learning from the past: possible urgent prevention and treatment options for severe acute respiratory infections caused by 2019-nCoV. *Chembiochem* 21 (5), 730–738. <https://doi.org/10.1002/cbic.202000047>.
- Lüthy, R., Bowie, J.U., Eisenberg, D., 1992. Assessment of protein models with three-dimensional profiles. *Nature* 356 (6364), 83–85.
- Morris, A.L., MacArthur, M.W., Hutchinson, E.G., Thornton, J.M., 1992. Stereochemical quality of protein structure coordinates. *Proteins: Struct., Funct., Bioinf.* 12 (4), 345–364.
- Nivón, L.G., Moretti, R., Baker, D., 2013. A pareto-optimal refinement method for protein design scaffolds. *PLoS One* 8 (4), e59004.
- Oh, C.-S., Martin, G.B., 2011. Effector-triggered immunity mediated by the Pto kinase. *Trends Plant Sci.* 16 (3), 132–140.
- Richards, F.M., 1974. The interpretation of protein structures: total volume, group volume distributions and packing density. *J. Mol. Biol.* 82 (1), 1–14.
- Richards, F.M., 1977. Areas, volumes, packing, and protein structure. *Annu. Rev. Biophys. Bioeng.* 6 (1), 151–176.
- Rosano, G.L., Ceccarelli, E.A., 2014. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front. Microbiol.* 5, 172.
- Samavati, L., Uhal, B.D., 2020. ACE2, much more than just a receptor for SARS-CoV-2. *Front. Cell. Infect. Microbiol.* 10, 317.
- Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., Wolfson, H.J., 2005. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* 33 (suppl_2), W363–W7.
- Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., et al., 2020. Structural basis of receptor recognition by SARS-CoV-2. *Nature* 581 (7807), 221–224.
- Shrake, A., Rupley, J.A., 1973. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.* 79 (2), 351–371.
- Thakur, S., Chauhan, V., Galwankar, S., Kelkar, D., Vedhagiri, K., Aggarwal, P., et al., 2020. Covid-19 testing strategy of India-current status and the way forward. *J. Global Infect. Dis.* 12 (2), 44.
- van Beers, M.M., Bardor, M., 2012. Minimizing immunogenicity of biopharmaceuticals by controlling critical quality attributes of proteins. *Biotechnol. J.* 7 (12), 1473–1484.
- Vazquez-Lombardi, R., Phan, T.G., Zimmermann, C., Lowe, D., Jermutus, L., Christ, D., 2015. Challenges and opportunities for non-antibody scaffold drugs. *Drug Discov. Today* 20 (10), 1271–1283.
- Wan, Y., Shang, J., Graham, R., Baric, R.S., Li, F., 2020. Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. *J. Virol.* 94 (7).
- Wang, X.S., Chen, P.H.C., Hampton, J.T., Tharp, J.M., Reed, C.A., Das, S.K., et al., 2019. A genetically encoded, phage-displayed cyclic-peptide library. *Angew. Chem. Int. Ed.* 58 (44), 15904–15909.
- Willard, L., Ranjan, A., Zhang, H., Monzavi, H., Boyko, R.F., Sykes, B.D., et al., 2003. VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res.* 31 (13), 3316–3319.
- Willis, J.R., Sapparapu, G., Murrell, S., Julien, J.-P., Singh, V., King, H.G., et al., 2015. Redesigned HIV antibodies exhibit enhanced neutralizing potency and breadth. *J. Clin. Invest.* 125 (6), 2523–2531.
- Xu, X., Chen, P., Wang, J., Feng, J., Zhou, H., Li, X., et al., 2020. Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. *Sci. China Life Sci.* 63 (3), 457–460.
- Yang, D., Park, S.Y., Park, Y.S., Eun, H., Lee, S.Y., 2020. Metabolic engineering of *Escherichia coli* for natural product biosynthesis. *Trends Biotechnol.* 38 (7), 745–765. <https://doi.org/10.1016/j.tibtech.2019.11.007>.