Enhancement of SARS-CoV-2 Receptor Binding Domain -CR3022 Human Antibody Binding Affinity via *In silico* Engineering Approach

Fatemeh Sefid[1, 2] , Zahra Payandeh[3] , Bahman Khalesi[4] , Behzad Mansoori[1] , Marzieh Fotovvat[5] , Maryam Touhidinia[6*]

[1]Department of Medical Genetics, Shahid Sadoughi University of Medical Science, Yazd, Iran; [2]Department of Biology, Science and Arts University, Yazd, Iran; [3]Immunology Research Center, Tabriz University of Medical Sciences, Tabriz, Iran; [4]Department of Research and Production of Poultry Viral Vaccine, Razi Vaccine and Serum Research Institute, Agricultural Research Education and Extension Organization, Karaj, Iran; [5]Department of Plant Science, Faculty of Biological Sciences, Kharazmi University, Tehran, Iran; [6]Department of Biology, Faculty of Science, Yazd University, Yazd, Iran

A R T I C L E   I N F O

Original Article

*Correspondence
Email:maryamtohidi90@yahoo.com
Tel: +989132648991
Fax: +983155414636

A B S T R A C T

Introduction: The angiotensin-converting enzyme 2 (ACE2) is the effective primary receptor for SARS-CoV-2. The interaction between ACE2 and the spike protein of the virus is the crucial step for virus entry into the target cells. ACE2 receptor can be blocked by neutralizing antibodies (nAbs) such as CR3022 which targets the virus receptor-binding site. Enhancing the binding affinity between CR3022 and ACE2 would lead to a more efficient blockade of virus entry. **Methods**: In this regard, the amino acids with central roles in the binding affinity of CR3022 antibody to spike protein were substituted. The best mutations to increase the affinity of antibodies were also selected based on protein-protein docking and molecular dynamics simulations. **Result**: The variants 45 (H:30I/G, H:55D/F, H: 103S/Y, L:59T/F, L:98Y/A), 60(H:31T/D, H:55D/E, H:103S/Y, L:59T/D, L:98Y/F), 67(H:30I/G, H:55D/F, H:103S/Y, L:56 W/L, L:59T/Y, L:61E/G), 69(H:31T/D, H:55D/F, H:103S/Y, L:59T/F, L:98Y/A), and 71(H: 31T/D, H:55D/F, H:103S/Y) with respective binding affinities of -167.3, -167.5, -161.6, -173.0, and -169.8 Kcal/mol had higher binding affinities against the RBD of the SARS-CoV2 spike protein compared to the wild-type Ab. **Conclusion**: The engineered antibodies with higher binding affinities against the target protein can improve specificity and sensitivity. Thus, a more successful blockade of the ACE2 is achieved, resulting in a better therapeutic outcome. *In silico* studies can pave the way for designing these engineered molecules avoiding the economic and ethical challenges.

INTRODUCTION

Coronaviruses, a highly diverse family of enveloped positive-sense single-stranded RNA viruses, are generally classified into four major genera: Alpha, Beta, Gamma, and Delta coronavirus. Severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory disorder coronavirus (MERS-CoV) are classified as β-coronaviruses. The novel coronavirus SARS-COV-2, the causative agent of coronavirus disease 2019 (COVID-19), is also in the β-coronaviruses group. This disease initially emerged in China in late 2019, and a global pandemic started in March 2020 (1, 2).

A primary human receptor for 2019-nCoV is angiotensin-converting enzyme 2 (ACE2). As a type I membrane protein, the enzyme is expressed on the surface of different cells such as lung, kidney, intestine, and heart cells. It reduces blood pressure and plays a role in vasoconstriction. Furthermore, decreased ACE2 expression was related to several cardiovascular diseases (3-6). Two separate studies recently reported that ACE2 was involved in the new SARS-CoV-coronavirus entry into the human body (7).

Since the virus entry is mediated by the receptor-binding domain (RBD) of spike (S) glycoprotein, it becomes a suitable target for generating neutralizing antibodies that can prevent virus entry. RBD of 2019-nCoV spike protein has six amino acid substitutions compared to the RBD in SARS spike protein, enabling interaction with ACE2 receptor for SARS-CoV-2 entry (3, 4, 8). Recent studies identified several potent monoclonal antibodies

Sefid et al.

against SARS coronavirus spike protein. Neutralizing antibodies (nAbs) against SARS-CoV-2 produced by experimental vaccination or monoclonal antibody (mAb) technology in passive transfer experiments protected mice from infection by reducing the viral load (9-13).

Among several expressed and purified SARS-CoV-specific antibodies, CR3022 antibody (Ab) targeting RBD showed a stronger neutralizing activity. This antibody in ELISA and Battleford Light Infantry (BLI) assays bound potently to 2019-nCoV RBD (14). The CR3022 Ab could be developed as a candidate therapeutic, alone or in combination with other neutralizing antibodies, to prevent or treat 2019-nCoV infections (14, 15).

This study aimed to improve the binding affinity of neutralizing CR3022 Ab against RBD of S protein via *in silico* protein engineering. Higher binding affinity could improve the diagnostic and therapeutic performance of CR3022 Ab.

MATERIAL AND METHODS

CR3022 complementarity-determining Region Prediction. The Ab binding activity is principally mediated by the complementarity-determining region (CDR). The Paratome web server (http://ofranservices.biu.ac.il/site/services/paratome/) (16) was used to predict the antigen-binding regions (ABRs) of Ab. The amino acid sequence or 3D structure of the protein was submitted to do the analysis. The Paratome was designed by the structural alignment of a non-redundant set of all established Ab-antigen (Ag) complexes in the PDB. DiscoTope (http://tools.iedb.org/discotope) was used to predict discontinuous epitopes from 3D structures of proteins in PDB format.

CR302 conservation of amino acid positions evolution. Estimating the evolutionary conservation of amino acid (AA) positions in protein sequences was assessed by ConSurf server (http://consurf.tau.ac.il) (17). The estimation of AA position was performed based on the phylogenetic relations among homologous sequences. The degree of AA is evolutionary conservation (i.e., its evolutionary rate) and heavily depends on its structural and functional value. Therefore, the importance of each AA position for structure or function was assessed by conservation analysis of AAs' position among the family members. The rate of evolution was calculated based on the evolutionary similarity between the protein and its homologs.

The Consurf parameters to perform the analysis were as follows: the homolog search algorithm was set to HMMER, the E-value cutoff was set to 0.0001, and the proteins database was set to UniRef90. This database clusters sequences and sub-fragments with 11 or more residues that have at least 90% sequence identity (from any organism) into a single UniRef entry, displaying the representative sequence.

CR3022 Interfaces prediction. Potential protein-protein interaction sites were identified by iterative "mapping" of the interaction sites of each structural neighbor involved in a complex to individual residues in the query protein. We used PredUs (18) web server, which is a flexible, interactive, template-based tool. The other tool we used for predicting protein-protein interaction sites was cons-PPISP (consensus Protein-Protein Interaction Site Predictor) [18]. The cons-PPISP uses a consensus neural network method and predicts the residues that will probably form the binding domain for another protein.

The solvent accessibility and deleterious mutations sites were predicted by WESA (https://pipe.rcc.fsu.edu/wesa/) with an accuracy of 80% (19). The decisive prediction for each residue relies on a weighted total of the specific predictions. The final prediction for each residue is based on a weighted sum of the individual predictions. Residues are predicted as buried or exposed with an expected accuracy of 80%. The exposed residues are defined as having a surface area greater than 20% of the maximum area for amino acid type.

The novel method for predicting partner-specific protein interfaces from .pdb files or input sequences was assessed by Extreme Gradient Boosting (xgBoost) (20). This method relies on Interface Prediction of Specific Partner Interactions (BIPSPI) available at http://bipspi.cnb.csic.es/xgbPredApp/. In this model. The .pdb coordinate format was used as the input file format.

CR3022 binding sites and pocket detection. To identify multiscale pockets on the protein surface, we used mathematical morphology via the Grid-based HECOMi finder) program (GHECOM) at http://strcomp.protein.osaka-u.ac.jp/ghecom/ (21). The structure file was fed in the .pdb format to do the analyses. The maximum radius for the large probe was set as 10.0 angstrom. We calculated the non-grid spherical probes (for UCSF DOCK/sievgene) as off.

The protein-protein interaction site was predicted by Meta-PPISP (https://pipe.rcc.fsu.edu/meta-ppisp.html) (22, 23). Several matching methods have been established for calculating protein-protein interaction sites. Meta-PPISP sought to enhance prediction reliability and accuracy by integrating results from separate predictors and reports. The meta-PPISP is developed on three separate web servers, including Promate, cons-PPISP, and PINUP. The meta-PPISP outperforms all three separate servers.

Significant residues selection. Some AAs were selected as significant residues in the CR3022 structure by utilizing the outcome of different software such as Paratome, BIPSPI, Cons PPISP, PredUS, Meta-PPISP, and WESA. These residues were sited in one of three

CDR regions and estimated by the Paratome server. The residues in BIPSPI software with a score above 0.5 have a score above 0.00 in Cons PPISP software and above 4 in GHecom software. In connection with this, PredUS, Meta-PPISP, and WESA predicted residues exploration to pick the major AAs.

These residues, located in one of three CDR regions, were predicted by the Paratome server. Selected residues in BIPSPI software had a score above 0.5, Cons PPISP software a score above 0.00, and GHecom software a score above 4. In this regard, PredUS, Meta-PPISP, and WESA predicted residues research to select the significant amino acids.

SIFT analyses. The Sorting Intolerant From Tolerant (SIFT) server (http://sift.jcvi.org/) was used to find if an AA replacement affects the protein function. The server predicts the conservation degree of AA residues in the sequence alignments that originated from closely linked sequences gathered through PSI-BLAST.

CR3022variants sketching. Seventy-one variants containing mutations in at least one of three ABRs were designed. To design new variants, we replaced the identified essential residues with the tolerable amino acids introduced by the SIFT server (http://blocks.fhcrc.org/sift/SIFT.html). Residues confirmed by bioinformatic analyses were randomly mutated in the suggested variants.

The three-dimensional structure of all submitted variants was determined by structure-based antibody prediction (SAbPred) from the Oxford Protein Informatics Group (OPIG) (24). All 3D models of the mutated antibodies were predicted and assessed for quality by the PROSA server (https://prosa.services.came.sbg.ac.at/prosa.php).

Molecular docking analyses. The 3D structure of each variant and spike antigen provide the inputs for High Ambiguity Driven protein-protein Docking (HADDOCK) (25). HADDOCK is an information-driven flexible docking method for the simulation of biomolecular complexes. To do the docking analyses between the receptor and ligand proteins, we set the H:30I, H:31T, H:33W, H:54G, H:103S, L:56W, L:58S, L:59T, L:61E, and L:98Y as HADDOCK active residues in antibody structure.

Molecular dynamics simulations. Molecular dynamics simulations were performed using the CABS Flex server (http://biocomp.chem.uw.edu.pl/CABSflex2/) for 50 cycles, 50 trajectory frames, and 10 ns, with some additional distance restraints with a global C-alpha restraints weight of 1.0.

RESULTS

CR3022CDR prediction. The Paratome is a bioinformatics tool for the recognition of ABRs in antibodies. Based on the antibody structure under the PDB ID of 6w41, this server predicted three regions as ABRs in the CR3022 heavy chain and three regions as ABRs in the CR3022 light chain. These regions include YGFITYWIG (27-35), WMGIIYPGDSETRY (47-60), and GGSGISTPMDV (98-108), respectively as ABR1, ABR2, and ABR3 in CR3022 heavy chain, and QSVLYSSINKNYLA (27-40), LLIYWASTRES (52-62), and QQYYSTPY (95-102) respectively as ABR1, ABR2, and ABR3 in CR3022 light chain (Table 1).

The predicted discontinuous epitopes (as a chart of DiscoTope score vs. residue ID) are shown in Fig. 1. The predictions with scores above the threshold (red line) are positive (displayed in green), and predictions with scores below the threshold are marked as negative predictions (displayed in orange). In the 3D view, Jmol displays the structure with positive predictions (highlighted in yellow). The side chain of each predicted residue is shown in Figure 1.

CR302 conservation for the evolution of amino acid positions. The nine-color conservation scores are projected onto the three-dimensional structure of the Ab, and FirstGlance displays colored protein structure in Jmol (Fig. 2.). The normal score calculated for each amino acid position was calculated by the Consurf server. The color scale represented by the conservation scores (9 - conserved, 1 - variable) is shown in Table 1.

CR3022 Interface predictions. Potential interfacial residues identified through PredUs are presented in Table 1. Residue 33W in ABRs I of a heavy chain (H chain ABR I), residues W47, I50, R59, and Y60 in ABRs II of a heavy chain (H chain ABR II), residues I102, S103, and D107 in ABRs III of a heavy chain (H chain ABR III), Residue Q27 in ABRs I of a light chain (L chain ABR I), residues L52, Y55, W56 and S58-E61 in ABRs II of a light chain (L chain ABR II), and residues Y97-S99, P101 and Y102 in ABRs III of a light chain are predicted as possible interfacial residues.

The cons-PPISP calculates a score of neural network for every residue. This score estimates if a residue is involved in the protein-protein interaction or not. The interred and not predicted residues will presume as zero scores. A score above zero is considered an interaction residue, and a score below zero, a non-interaction residue (Table 1).

Weighted Ensemble Solvent Accessibility predictor (WESA) identified some AAs as solvent-accessible residues (Table 1). BIPSPI predicted the partner-specific protein-protein interfaces. Residues with a predicted score higher or equal to the precision threshold (0.500) were highlighted in green. The interactive visualization of predicted residues in the Ab structure is shown in Figure 3, and the predicted interface scores for the residues of Ab are listed in Table 1.

Sefid et al.

CR3022 binding sites and pocket detection. GHECOM server discovers five pockets on protein shell by utilizing mathematical morphology. In this respect, it calculates a pocket's score (sum of 1/[Rpocket] /(1/[Rmin]*[vol of shell])) for each residue. A residue in a deeper and bigger pocket has a greater value of pockets. The pockets of small-molecule binding and active locations are greater than the average value; in particular,

the values for the active locations are much greater. This implies that pockets are contributing to the prediction of binding and active sites from protein structures (Fig. 4.) (Table 1) [24]. Pockets contribute to the prediction of binding sites and active sites from protein structures. A residue in a deeper and larger pocket had a larger value. The pockets of small-molecule binding sites and active sites were higher than the average value; specifically, the values for the active sites were much higher.

Table 1. Paratome, predUs, cons-PPISP, GHECOM, WESA, and Consurf server Predictions. Blue, orange, and green show ABR1, ABR2, and ABR3 respectively in the CR3022 heavy chain. Violet, yellow, and red show ABR1, ABR2, and ABR3, respectively, in the CR3022 light chain.

ain: Residue	H:27	H:28	H:29	H:30	H:31	H:32	H:33	H:34	H:35	H:47	H:48	H:49	H:50	H:51	H:52	H:53	H:54
Paratome	Y	G	F	I	T	Y	W	I	G	W	M	G	I	I	Y	P	G
Pred us	Y	G	F	I	T	Y	W	I	G	W	M	G	I	I	Y	P	G
Cons ppisp	0.14	0.124	0	0.081	0.112	0.099	0.068	0	0	0	0	0	0	0	0.063	0	0.085
Consurf	5	3	6	4	1	1	1	4	1	8	5	3	1	7	1	1	1
Ghecom	0.37	0.62	2.76	2.99	0.88	1.32	0	0	0	3.08	1.69	0	0.02	0	0.82	2.83	3.07
WESA	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1
BIPSPI	0.417	0.5276	0.5765	0.6946	1	0.6402	0.6246	0.4596	nan	nan	nan	nan	0.2091	0.3352	0.5082	0.6796	0.6202
Meta-PPISP	P	P	-	P	P	P	P	-	-	-	-	-	-	-	P	-	P

ain: Residue	H:55	H:56	H:57	H:58	H:59	H:60	H:98	H:99	H:100	H:101	H:102	H:103	H:104	H:105	H:106	H:107	H:108
Paratome	D	S	E	T	R	Y	G	G	S	G	I	S	T	P	M	D	V
Pred us	D	S	E	T	R	Y	G	G	S	G	I	S	P	M	T	D	V
Cons ppisp	0.073	0.084	0.084	0.067	0.065	0.09	0	0	0.069	0.044	0.044	0.036	0.04	0	0	0.107	0
Consurf	1	1	1	3	1	8	4	1	1	1	1	1	1	1	2	1	
Ghecom	0.43	0	0	0	1.26	0.02	0	0.1	0.89	0.46	1.61	4.2	3.6	0	0	8.08	6.93
WESA	1	1	1	1	1	0	0	1	1	1	1	1	1	0	0	0	0
BIPSPI	0.4305	0.3976	0.2848	0.3277	0.2426	0.1917	0.6202	0.5643	0.4652	0.2215	0.2111	nan	nan	nan	nan	nan	Nan
Meta-PPISP	N	N	N	N	N	N	-	-	P	P	N	P	P	-	-	N	-

Chain: Residue	L:27	L:28	L:29	L:30	L:31	L:32	L:33	L:34	L:35	L:36	L:37	L:38	L:39	L:40	L:52	L:53	L:54
Paratome	Q	S	V	L	Y	S	S	I	N	K	N	Y	L	A	L	L	I
Pred us	Q	Y	V	S	L	S	S	I	N	K	N	Y	L	A	L	L	I
Cons ppisp	0.02	0.077	0	0.065	0.063	0.058	0.048	0.059	0.072	0.05	0.051	0.039	0	0	0	0	0
Consurf	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	5	7
Ghecom	0.15	0	0.29	1.69	1.07	0	0	0.09	2.41	1.92	3.42	1.53	0.02	0	4.81	2.32	0
WESA	1	1	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0
BIPSPI	0.2704	0.6071	0.4767	0.4071	0.2915	0.3562	0.2482	nan	nan	nan	nan	nan	nan	nan	0.2368	0.2901	0.2331
Meta-PPISP	N	N	-	P	P	P	P	P	P	P	P	P	-	-	-	-	-

Chain: Residue	L:55	L:56	L:57	L:58	L:59	L:60	L:61	L:62	L:95	L:96	L:97	L:98	L:99	L:100	L:101	L:102
Paratome	Y	W	A	S	T	R	E	S	Q	Q	Y	Y	S	T	P	Y
Pred us	Y	W	A	S	T	R	E	S	Q	Q	Y	Y	S	T	P	Y
Cons ppisp	0.053	0.044	0	0.034	0.056	0.062	0.061	0.086	0	0	0	0.039	0.047	0.063	0.098	0
Consurf	3	1	1	1	1	5	1	7	1	1	1	1	1	1	1	1
Ghecom	1.94	3.06	1.25	1.41	0.15	0.04	6.6	3.78	0	0.06	2.31	1.44	2.03	1.86	2.82	3.51
WESA	0	1	0	1	1	1	1	1	0	0	0	1	1	1	0	1
BIPSPI	0.2802	0.2489	0.1975	nan	nan	nan	nan	nan	0.2459	0.2307	0.1949	nan	nan	nan	nan	nan
Meta-PPISP	P	P	-	P	P	N	N	N	-	-	-	P	N	N	N	-

The residues H: S103, H: D107, H: V108, L: L52, and L: E61 had a GHECOM score above 4. The interface residues predicted by Meta-PPISP are shown as residue ID (chain ID): 1:H, 3:H, 25:H, 26:H, 27:H, 28:H, 30:H, 31:H, 32:H, 33:H, 52:H, 53:H, 73:H, 76:H, 96:H, 97:H, 99:H, 100:H, 27C:L, 27D:L, 27E:L, 27F:L, 28:L, 29:L, 30:L, 31:L, 32:L, 49:L, 50:L, 52:L, 53:L, 92:L. P corresponds to a score > 0.34 in meta-PPISP results (P = Positive; N = Negative; - = Buried and not predicted) (Table 1).

Significant residues selection. We selected H:30I, H:31T, H:33W, H:54G, H:103S, L:56W, L:58S, L:59T, L:61E, and L:98Y residues by using the results of various softwares. The H stands for the heavy chain and the L for the light chain. These residues situated in one of three CDR regions were predicted by the Paratome. At least four softwares confirmed the specially selected residues. The cons-PPISP scores above 0.00, BIPSPI scores above 0.5, and GHecom scores above 4 were considered thresholds. In this regard, PredUS, Meta-

PPISP, and WESA predicted research residues to select the significant AAs (Table 2).

SIFT analyses. SIFT predicted the tolerated and detrimental alterations in each position of the submitted sequence. Positions with normalized possibilities lower than 0.05 are expected to be deleterious; those higher than or equal to 0.05 are expected to be tolerated (Table 3).

Table 2. Selected residues regarding various softwares.

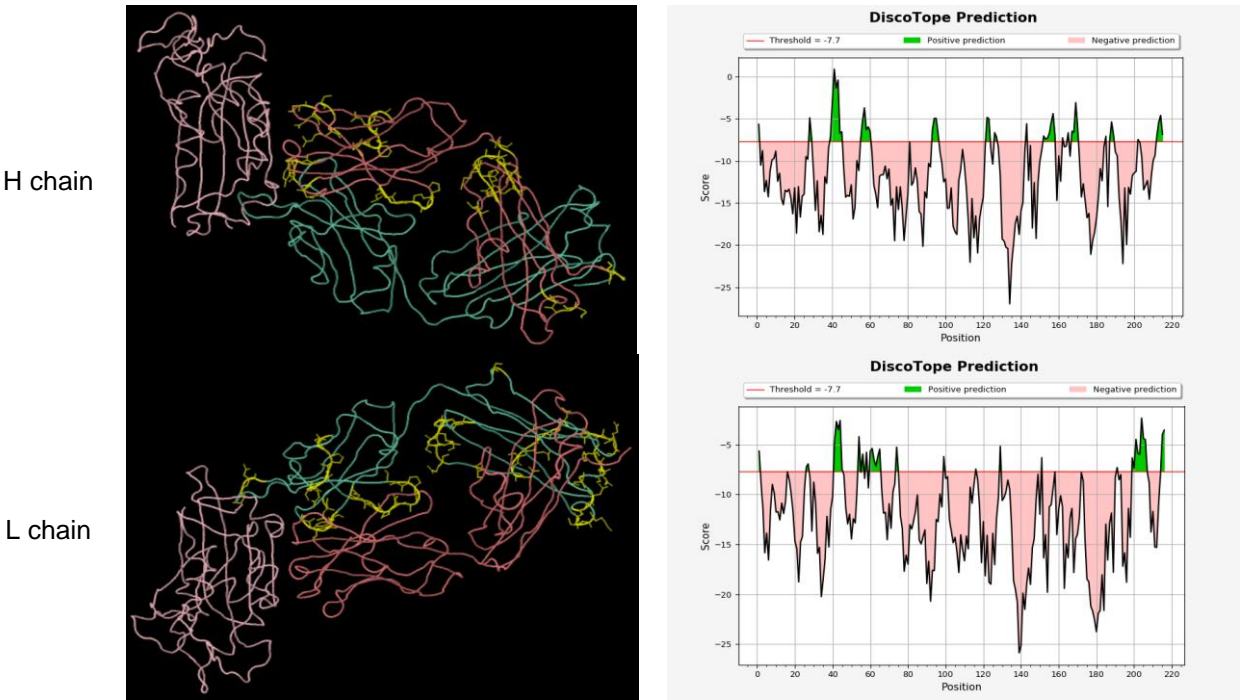|	Cons PPISP	WESA	BIPSPI	Meta-PPISP	PredUs	Ghecom	WESA
H:30I	*	*	*	*			
H:31T	*	*	*	*			
H:33W	*		*	*	*		
H:54G	*			*		*	*
H:103S	*			*	*		*
L:56W	*		*	*			*
L:58S	*			*	*		*
L:59T	*			*	*		*
L:61E	*				*	*	*
L:98Y	*			*	*		*

Fig. 1. The predicted discontinuous epitopes as a chart of DiscoTope score vs. residue id. In the chart, predictions above the threshold (red line) are positive predictions (in green), and predictions below the threshold are negative predictions (in orange). The 3d view uses Jmol to display the structure with positive predictions highlighted in yellow.

CR3022 variants sketching. We offered 71 variants, including mutations in at least one of 3 ABRs. H:30I, H:31T, H:33W, H:54G, H:103S, L:56W, L:58S, L:59T, L:61E, and L:98Y residues established by various programs were mutated in the proposed variants randomly. In addition, H:30I, H:33W, H:57E, H:55D, H:102I, and L:56W were introduced by Meng Yuan *et al*. (26) as interactive residues in the crystal structure of CR3022 in complex with SARS-CoV-2 RBD. In this regard, we also mutated H:57E, H:55D, and H:102I to access a more diverse variant. Mutated sequences were aligned and illustrated (Fig. 5).

SAbPred determined the three-dimensional structure of all submitted variants. The quality estimation of the 3D models by the Prosa web server revealed that the Z-score

of all predicted models was within the range of scores typically found for native proteins of similar size. The Z-score indicates overall model quality. The value is displayed in a plot that contains the z-scores of all experimentally determined protein chains in the current PDB. In this plot, groups of structures from different sources (X-ray, NMR) are distinguished by different colors. The Z-score can be used to check whether the Z-score of the input structure is within the range of scores typically found for native proteins of similar size (Fig. 6).

Protein-protein docking based upon biochemical or biophysical data. HADDOCK server evaluates ligand and receptor integration based on biochemical and/or biophysical data. Table 4 represents the information of

Sefid et al.

variants with the HADDOCK scores that are more than the control. The Van der Waals and electrostatic energy values, in addition to the interred surface between the two

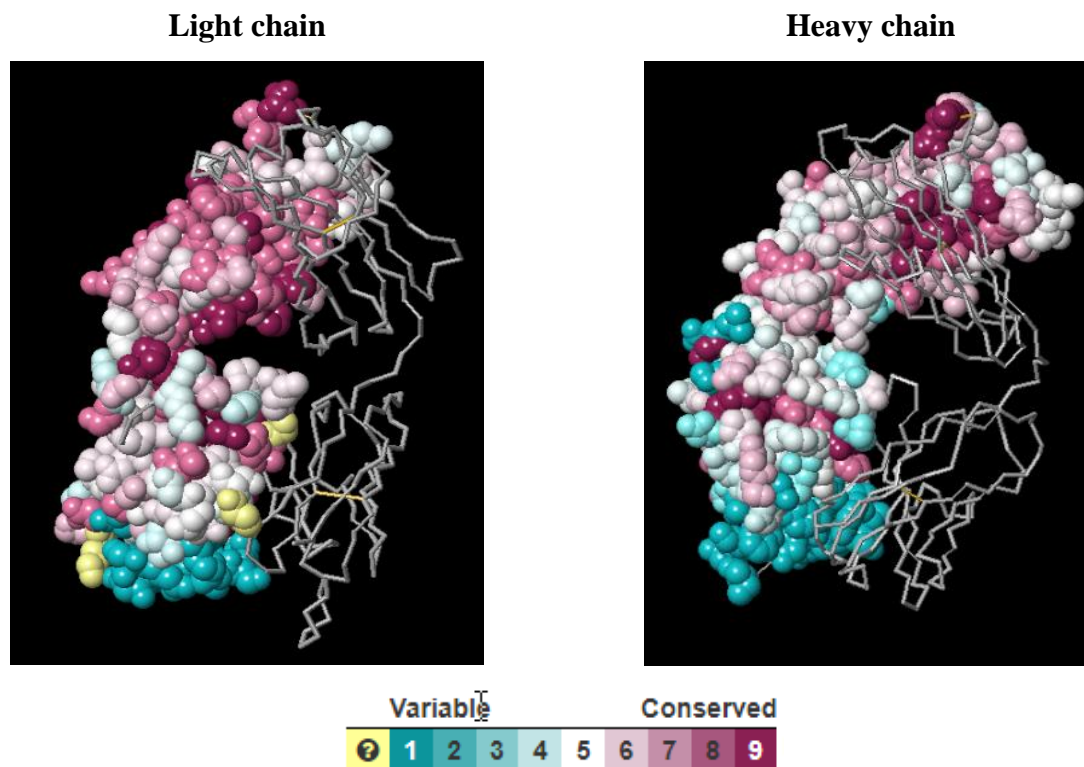complexes, are shown. HADDOCK scores of all variants are shown in Figure 7.

Light chain

Heavy chain

Variable   Conserved

❓ 1 2 3 4 5 6 7 8 9

Fig. 2. CR302 conservation of amino acid positions evolution by the Consurf server. The schematic structure of the colored protein displayed by FirstGlance in Jmol.Conservation scores projected onto the three-dimensional structure of the Ab with nine colors. The light chain analysis is shown on the left, and the heavy chain analysis is on the right.

Molecular dynamics simulations. Molecular dynamics simulations provide an accurate ranking of the potential ligands and binding sites. We obtained ten different models by computationally exhaustive exploratory molecular dynamics simulation using CABS-Flex 2.0. Three-dimensional structures of 10 final models were scrutinized. Visualization of the models showed the structural heterogeneity of the final models. Two options, including the surface and cartoon representation of the trajectory (all final models in superposition), are shown in Figure 8. We selected the first model because of its best structural heterogeneity, optimum free energy, and highly stable configuration. The 'fluctuation plot' provides an interactive 2D plot representing residue-wise fluctuations recorded throughout the simulation. Fluctuations are calculated as RMSF after global superposition. Our complex structure in the residue positions 42, 75, 135, 166, and 205 had a high level of fluctuations (2.44 Å, 2.30 Å, 2.12 Å, 2.30 Å, and 2.04 Å, respectively). The RMSF plot has been depicted in Figure 9 provided by Cabs-Flex 2.0.

DISCUSSION

Protein engineering is a powerful method for developing ideal therapeutic proteins (27). Contemporary bioinformatics techniques are commonly used in protein engineering (28, 29). In diseases like COVID-19, where time is critical, these tools could save a lot of time and effort.

For virus entry into the cells, binding of ACE2 and RBD in spike protein is essential. Therefore, many studies have focused on inhibiting the virus attachment to the binding sites.

Recently Changhai *et al*. (2020) published a recombinant ACE2-Ig with potential applications for the 2019-nCoV diagnosis, prophylaxis, and treatment (30). Other researchers have produced an engineered ACE2 protein using bioinformatics techniques expected to bind RBD with higher affinity while being more thermostable and lacking enzymatic activity (31).
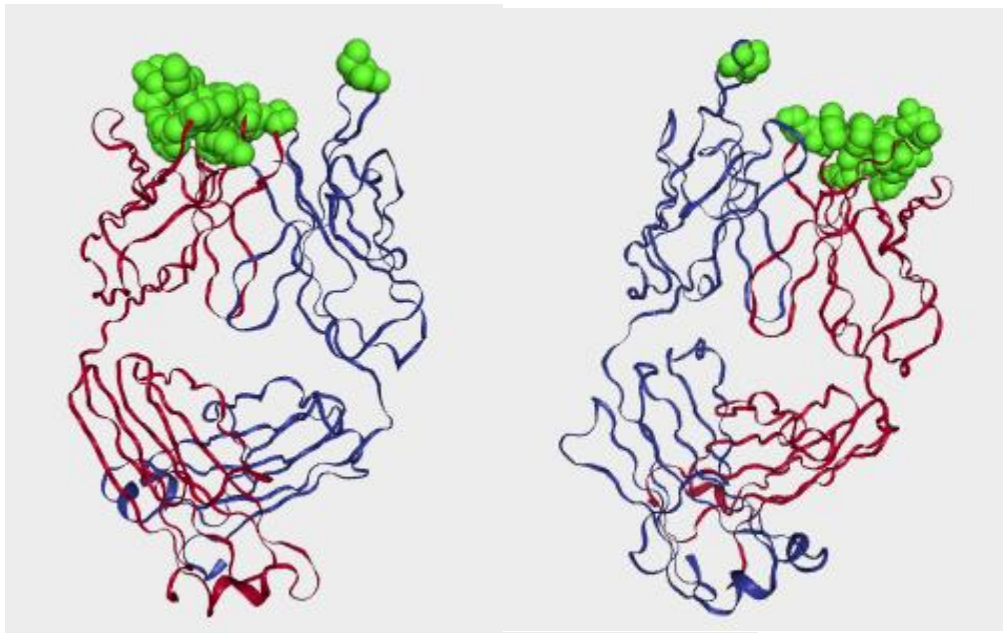
Chain	Residue	Score	Expected precision
H	THR31	6.0353	1
H	ILE30	2.1479	0.6946
H	GLY54	2.0132	0.6796
H	TYR32	1.0832	0.6402
H	TRP33	0.9694	0.6246
H	ASP55	0.7617	0.6202
H	ILE102	0.7564	0.6202
L	ILE34	0.6683	0.6071
H	PHE29	0.4376	0.5765
H	GLY101	0.4356	0.5765
H	SER103	0.3547	0.5643
H	PRO53	0.2697	0.5403
H	GLY28	0.2412	0.5276
H	TYR52	0.1972	0.5082

Fig. 3. BIPSPI interactive visualization of predicted residues in the antibody structure. Residues whose score has an expected precision greater or equal than the precision threshold (0.500) are green. Interface residues prediction by BIPSPI whose score has an expected precision greater or equal than the precision threshold (0.500) are listed below the picture. The light chain is blue, and the heavy chain is red.

Incoming research has shown that human Ab CR3022 binds to SARS-CoV-2 RBD in a cryptic epitope (14). In addition, we found no studies that designed these antibodies against different epitopes of SARS-CoV-2 proteins, i.e., targeting the spike protein and subsequently preventing its binding to human ACE2. Since passive immunization with polyclonal antibodies has been shown to curb outbreaks of the hepatitis A virus and prevent varicella-zoster infection (32), mAb prophylaxis can be an effective way to contain the SARS outbreak. This computational research aimed to advance the binding affinity of CR3022 Ab neutralization to ACE2 (RBD) in the SARS-CoVs2 virus by focusing on AA mutagenesis. We believed that affinity maturation increases the performance of diagnostic antibodies due to enhanced specificity at reduced Ab concentrations. We used site-directed mutagenesis approaches for this order. Site-directed mutagenesis (SDM) methods are used to

Sefid et al.

produce cloned DNAs with changed sequences to investigate the significance of different residues in the structure and function of the proteins (33, 34). Our first goal was to find structurally and functionally essential AAs. We ought to identify sites in the complementarity-determining regions (CDRs) permissive to mutagenesis while maintaining Ag binding. Next, we selected several residues by various methods: 1) predicting the antigen-binding regions of CR3022 Ab using the Paratome web server, 2) predicting the protein-protein interaction sites using ConSurf, PredUs, and Meta-PPISP 3) and finally, partner-specific protein interface prediction by BIPSPI and WESA.
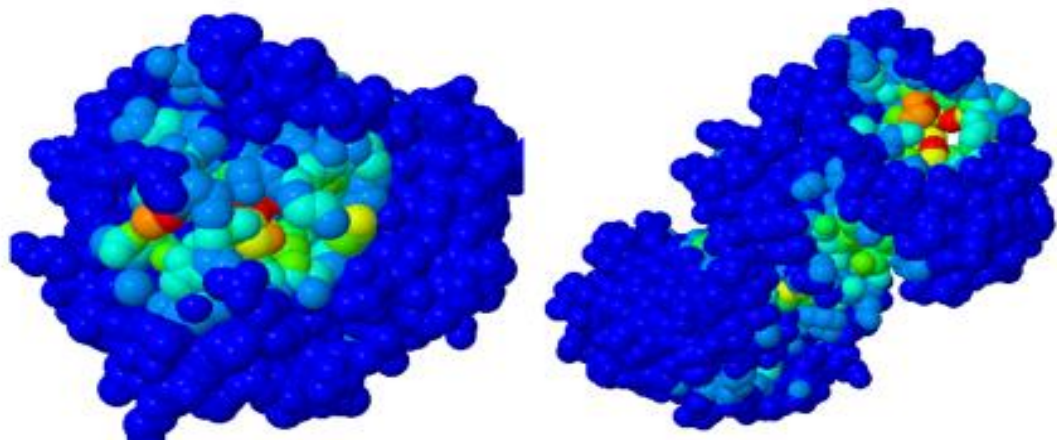

Fig. 4. GHECOM results show Jmol view of a pocket structure based on pockets color (red indicates high pocketness and blue indicates low pocketness). The measure pocketness indicates both the size and depth of a pocket.

Table 3. SIFT results for selected residues

Position	Seq Rep	Predict Tolerated
H:30I	0.89	H p 1 V R q N D G I A k E T S
H:31T	0.89	h i v p 1 q A K R E G N D T S
H:33W	0.89	C p m D E q W K N R G i T S V A H L F Y
H:54G	0.85	c p M W Q E i R K T D A H F V L N S Y G
H:55D	0.84	m h F i P L V r q T A k E G S N D
H:57E	0.81	c F Y m h I P V L G N R Q T D S A K E
H:102I	0.56	c W P D m E K q N G R T S I A V H L F Y
H:103S	0.65	w C m P D I q N G r K h E T V F L S A Y
L:56W	1.00	c p m D Q N W K i E G R T S V H A L F Y
L:58S	1.00	N S
L:59T	0.99	f c Y m h I p v L g D Q N A S R K E T
L:61E	0.98	1 G T V S K Q H D Y P A E
L:98Y	0.93	c W p D m E K Q N G R I T S V A H L F Y

Amino acid (AA) features are in different colors. Black, nonpolar; green, uncharged polar; red, basic; blue, acidic. Initial capitals imply AAs in the alignment; lower case letters arise from prediction.

Table 4. Docking between the wild type and five best mutated CR3022 human antibody variants with RBD Ag.

	Control	variant45	variant60	variant67	variant69	variant71
HADDOCK score	-138.4	-167.3	-167.5	-161.6	-173.0	-169.8
Cluster size	71	40	47	48	32	39
Van der Waals energy	-54.5	-68.9	-64.4	-80.1	-63.8	-63.3
Electrostatic energy	-341.7	-338.4	-466.5	-247.3	-431.1	-435.3
Desolvation energy	-21.9	-35.6	-14.5	-38.8	-27.5	-21.6
Restraints violation energy	63.7	49.1	46.7	67.1	45.4	22.3
Buried Surface Area	1809.7	2242.2	2221.8	2335.3	2136.4	2165.1
Z-Score	-1.7	-1.5	-1.2	-2.2	-1.7	-1.6

BIPSPI predictions from two input structural models (Ab and Ag) were performed using PREDICT from the structural data option. By employing the results of different softwares, we selected H:30I, H:31T, H:33W, H:54G, H:103S, L:56W, L:58S, L:59T, L:61E, and L:98Y residues. Based on the prediction by the Paratome web server, these residues were placed in one of the CDR regions. At least four softwares confirmed the specially selected residues. A cons-PPISP score above 0.00, BIPSPI score above 0.5, and GHecom score above four were considered thresholds. In this regard, PredUS, Meta-PPISP, and WESA predicted residues to choose the major AAs. Finally, the SIFT server was used to predict whether an AA switch interferes with the protein act. Various studies have used PREDICT and SIFT servers to make the necessary predictions (28, 29, 35).

Based on these results, we substituted the AAs. Then, the docking analyses were done between the structure of engineered antigens and the receptor structure. Molecular docking plays a significant role in targeted medicine designing. It is the only theoretical method that explicitly models physical interactions between proteins.

HADDOCK 2.1 performs the docking of CR3022 Ab variants and SARS-CoV 2 spike RBD, totally flexible, utilizing molecular dynamics simulations. Docking between wild-type and mutant antibodies indicated that the engineered mutations had strengthened the binding affinity between mutated antibodies to the receptor molecule. One of the criteria for the comparison of docking results is the Gibbs free energy (ΔG). The smaller ΔG has higher stability and binding affinity.

In this research, CABS-Flex 2.0 software was deployed for the flexible molecular dynamic simulation of our docked complexes. CABS-Flex could present the stable arrangement of the antigen-antibody complex. The highest RMSF value indicates more fluctuations within the complex structure during the simulation process. Fluctuations in the structure of the antigen-antibody complex show its high flexibility and a potential structure of the complex. As per our findings, the complex had many fluctuations in chain A; the highest amplitude was residue 42 of ~2.44Å.

Based on our results, the variants 45, 60, 67, 69, and 71 antibodies had smaller bonding energy, and their stabilities had also increased compared to the wild type. Based on obtained scores, some mutated variants, including variant 45 (-167.3 Kcal/mol), variant 60 (-167.5 Kcal/mol), variant 67 (-161.6 Kcal/mol), variant 69 (-173.0 Kcal/mol), and variant 71 (-169.8 Kcal/mol) had higher affinity for binding to SARS-CoV2 RBD compared to the wild type Ab. The buried area between the two complexes in these variants was more positive than the natural state, meaning that these mutations have modified and increased the Ab binding properties relative to wild-type antibodies. Consistent with our study, *in silico* analysis of the interaction between viral S protein and ACE2 receptors showed two neutralizing mouse antibodies, F26G19, and D12, against SARS-CoV2 designed using simulation and antibody-antigen docking could bind to SARS-CoV S protein with high affinities [36]. Also, investigating the affinity maturation of CR3022 convalescent antibody towards RBD of SARS-CoV-2 S protein, alongside in silico designed antibodies, showed that three candidates antibodies, SAM1, SAM2 (post-MDS), and SAM3 (pre-MDS), had higher binding affinities than their counterparts and CR3022 antibody. Besides better binding affinity, they also demonstrated greater target specificity towards SARS-CoV-2 S RBD due to blocking the human ACE-2 receptor binding site, as predicted in the study (37).

Sefid et al.

| Residue	H:30	H:31	H:33	H:54	H:55	H:57	H:102	H:103	L:56	L:58	L:59	L:61	L:98
CONTROL	I	T	W	G	D	E	I	S	W	S	T	E	Y
Variant1	H	T	D	G	D	E	I	S	W	S	T	E	Y
Variant2	D	N	W	G	D	E	I	S	W	S	T	E	Y
Variant3	G	A	E	G	D	E	I	S	W	S	T	E	Y
Variant4	I	R	K	G	D	E	I	S	W	S	T	E	Y
Variant5	I	T	W	W	F	E	I	S	W	S	T	E	Y
Variant6	I	T	W	R	D	N	I	S	W	S	T	E	Y
Variant7	I	T	W	G	F	Y	I	S	W	S	T	E	Y
Variant8	I	T	W	W	G	L	I	S	W	S	T	E	Y
Variant9	I	T	W	G	D	E	W	T	W	S	T	E	Y
Variant10	A	T	W	M	D	E	W	S	W	S	T	E	Y
Variant11	I	E	W	G	K	E	I	Y	W	S	T	E	Y
Variant12	I	T	Q	G	D	G	I	Y	W	S	T	E	Y
Variant13	S	T	N	G	P	E	K	S	W	S	T	E	Y
Variant14	I	E	W	W	D	N	I	K	W	S	T	E	Y
Variant15	V	N	W	D	E	E	T	D	W	S	T	E	Y
Variant16	T	T	S	F	D	S	I	S	W	S	T	E	Y
Variant17	I	A	T	W	D	T	I	S	W	S	T	E	Y
Variant18	I	T	W	G	A	F	W	T	W	S	T	E	Y
Variant19	T	S	F	F	S	F	I	S	W	S	T	E	Y
Variant20	T	K	D	G	D	E	W	T	W	S	T	E	Y
Variant21	I	T	W	Y	A	S	Y	F	W	S	T	E	Y
Variant22	S	A	Y	M	S	Y	W	Y	W	S	T	E	Y
Variant23	E	D	F	Y	S	Y	S	N	W	S	T	E	Y
Variant24	V	A	R	G	D	E	Y	F	W	S	T	E	Y
Variant25	I	T	W	M	A	F	F	W	W	S	T	E	Y
Variant26	H	K	W	W	F	E	S	F	W	S	T	E	Y
Variant27	I	T	W	G	D	E	I	S	S	N	Y	E	Y
Variant28	I	T	W	G	D	E	I	S	F	S	S	Y	Y
Variant29	I	T	W	G	D	E	I	S	W	N	T	A	F
Variant30	I	T	W	G	D	E	I	S	F	S	Q	Y	V
Variant31	I	T	W	G	D	E	I	S	E	S	T	S	F
Variant32	I	T	W	G	D	E	I	S	G	N	A	H	A
Variant33	V	T	S	G	A	E	I	Y	K	S	Y	E	F
Variant34	I	A	W	G	F	E	I	A	W	N	T	Q	W
Variant35	S	A	F	G	D	E	I	S	W	S	T	E	V
Variant36	I	S	F	G	D	E	I	S	L	S	Y	G	Y
Variant37	I	T	W	W	A	S	I	S	W	S	T	T	D
Variant38	I	T	W	W	D	P	I	D	W	S	Y	Y	Y
Variant39	I	T	W	W	V	Q	I	S	W	N	S	A	Y
Variant40	I	T	W	G	D	E	W	Y	W	S	T	E	N
Variant41	I	T	W	G	D	E	R	T	W	S	L	D	H
Variant42	A	T	W	W	D	E	W	S	W	N	T	E	W
Variant43	I	A	W	G	S	E	I	N	W	S	A	E	Y
Variant44	I	T	T	G	D	L	I	S	F	S	T	G	Y
Variant45	G	T	W	G	F	E	I	Y	W	S	F	E	A
Variant46	I	T	N	G	D	L	I	D	W	S	T	Y	Y
Variant47	I	A	W	G	I	E	M	S	W	S	R	E	H
Variant48	I	A	W	G	D	F	I	S	Y	S	T	E	W
Variant49	S	T	W	G	F	E	W	S	W	N	T	E	Y
Variant50	I	D	W	G	D	L	I	F	W	S	R	E	Y
Variant51	I	A	W	W	D	E	W	S	R	S	T	E	Y
Variant52	I	T	N	G	D	S	I	T	W	S	T	Y	Y
Variant53	I	A	W	G	S	E	A	S	W	S	T	E	F
Variant54	A	S	W	W	F	E	I	S	W	N	R	E	Y
Variant55	I	E	W	W	D	R	I	V	F	S	T	D	Q
Variant56	I	T	A	G	D	L	I	S	F	S	T	A	G
Variant57	I	K	W	G	P	F	N	S	H	S	T	E	H
Variant58	A	S	W	G	D	D	A	Y	W	N	T	E	Y
Variant59	I	T	V	G	D	R	I	Y	E	S	T	Q	Y
Variant60	I	D	W	G	E	E	I	Y	W	S	D	E	F
Variant61	I	T	A	G	D	D	S	S	F	S	L	E	Y
Variant62	I	T	W	G	D	E	I	S	W	S	F	E	A
Variant63	G	T	W	G	F	E	I	Y	W	S	T	E	Y
Variant64	I	T	W	G	D	E	I	S	L	S	Y	G	Y
Variant65	I	S	F	G	D	E	I	S	W	S	T	E	Y
Variant66	I	S	F	G	D	E	I	S	W	S	F	E	A
Variant67	G	T	W	G	F	E	I	Y	L	S	Y	G	Y
Variant68	G	T	W	G	F	E	I	Y	W	S	D	E	F
Variant69	I	D	W	G	E	E	I	Y	W	S	F	E	A
Variant70	I	T	W	G	D	E	I	S	W	S	D	E	F
Variant71	I	D	W	G	E	E	I	Y	W	S	T	E	Y

Fig. 5. Illustration of mutations sequences
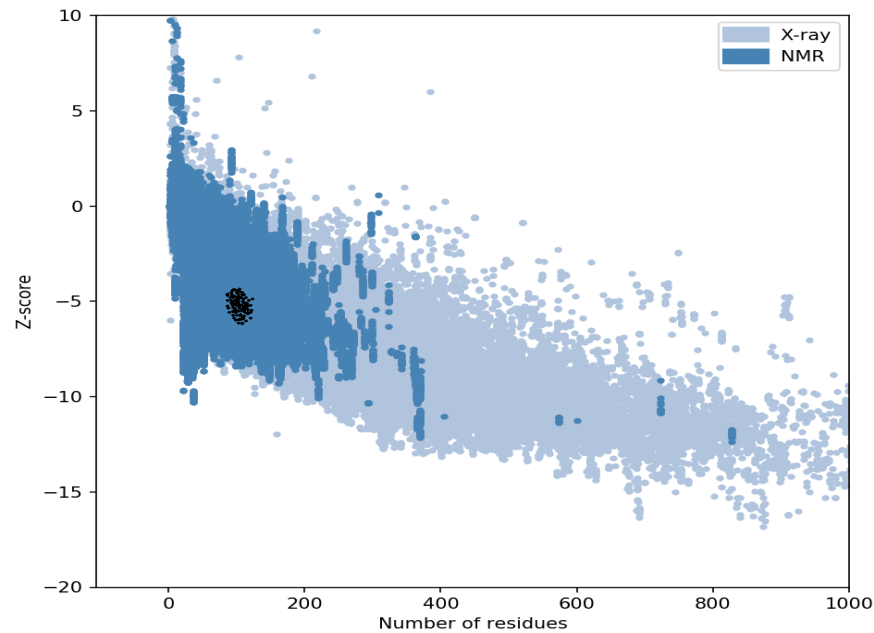
Fig. 6. Overall model quality of variant structures**.** Black dots represent the Z-Score. The Z-score indicates overall model quality. Its value is displayed in a plot that contains the Z-Score of all experimentally determined protein chains in the current PDB. In this plot, groups of structures from different sources (X-ray, NMR) are distinguished by different colors. Z-Score can be used to check whether the Z-Score of the input structure is within the range of scores typically found for native proteins of similar size.
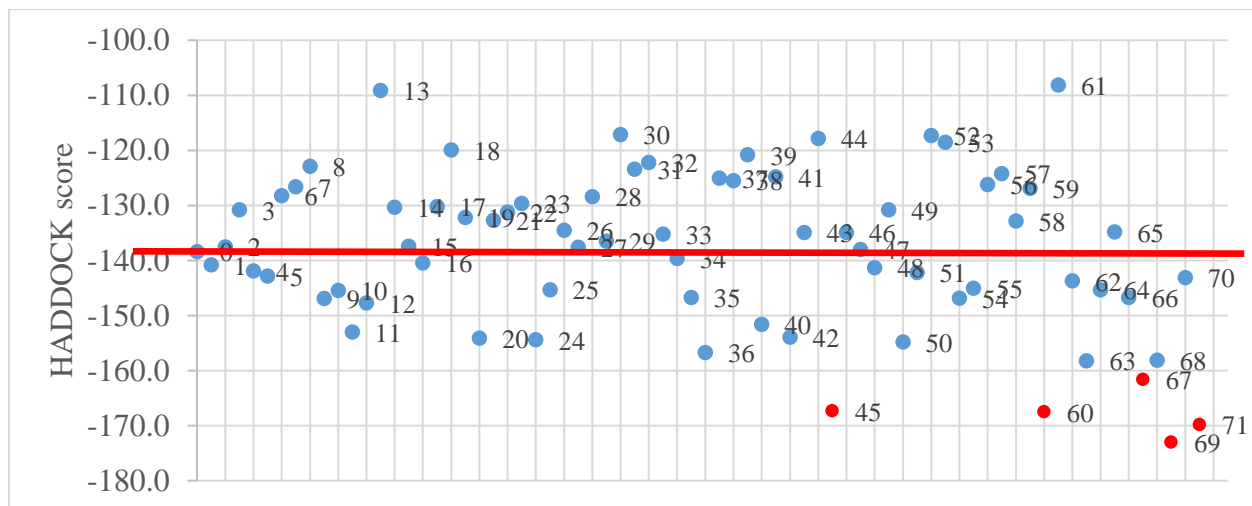
Fig. 7. HADDOCK Score of all variants. The variants with scores below were predicted to have an enhanced affinity toward the CR3022 human Ab. The 5-best mutated CR3022 human Ab variants are colored in red on the graph.

Our results suggest that the CR3022 Ab could be developed as a candidate therapeutic alone or combined with other neutralizing antibodies to prevent or treat nCoV-2019 infections. Optimizing the CR3022 properties to boost their binding characteristics and affinity maturation to avoid cross-reactivity and selecting suitable antibodies by adding new mutations to create the desired Ab is advantageous.
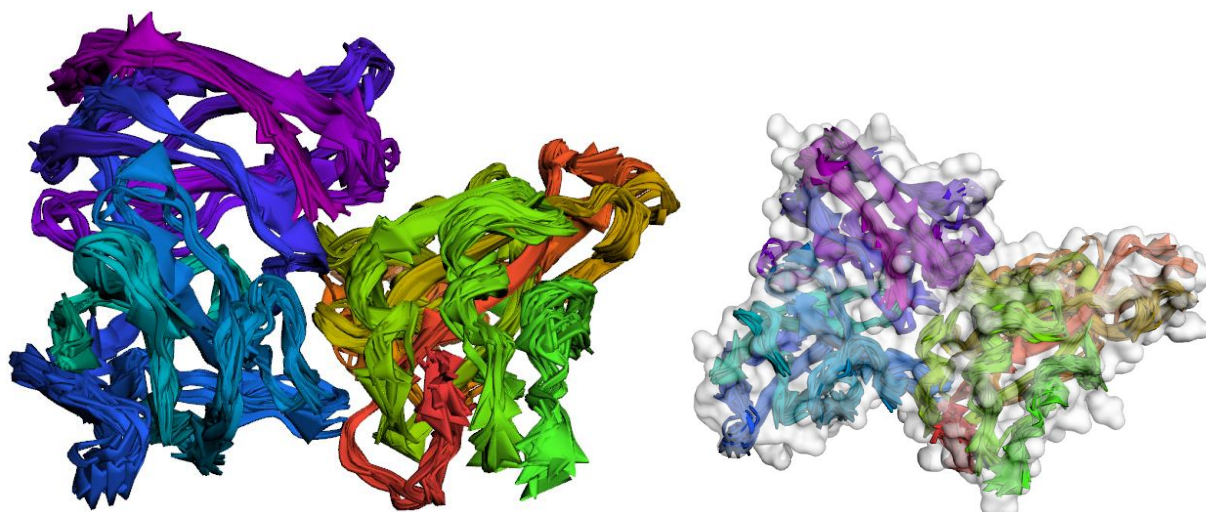
ACKNOWLEDGMENT

Fig. 8. Two viewing options such as surface and cartoon representation of the trajectory (all final models in superposition)
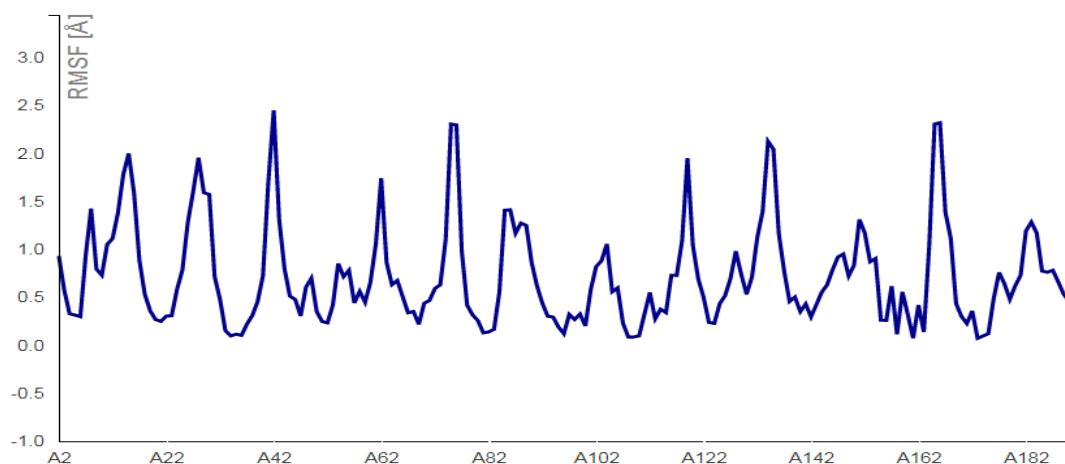

Fig. 9. The Root Mean Square Fluctuation (RMSF) plot of the protein complex.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest associated with this manuscript.

REFERENCES

1. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020; 395 (10223): 497-506.

2. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol. 2020; 5: 536-44.

3. Donoghue M, Hsieh F, Baronas E, Godbout K, Gosselin M, Stagliano N, et al. A novel angiotensin-converting enzyme–related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1-9. Circ Res. 2000; 87 (5): e1-9.

4. Boehm M, Nabel EG. Angiotensin-converting enzyme 2—a new cardiac regulator. N Engl J Med. 2002; 347 (22): 1795-7.

5. Zhang H, Kang Z, Gong H, Xu D, Wang J, Li Z, et al. The digestive system is a potential route of 2019-nCov infection: a bioinformatics analysis based on single-cell transcriptomes. BioRxiv. 2020.

6. Zhao Y, Zhao Z, Wang Y, Zhou Y, Ma Y, Zuo W. Single-Cell RNA Expression Profiling of ACE2, the Receptor of SARS-CoV-2. Am J Respir Crit Care Med. 2020; 202 (5): 756–9.

7. Zhang Q, Xiang R, Huo S, Zhou Y, Jiang S, Wang Q, et al. Molecular mechanism of interaction between SARS-CoV-2 and host cells and interventional therapy. Signal Transduct Target Ther. 2021; 6 (1): 233.

8. Ge X-Y, Li J-L, Yang X-L, Chmura AA, Zhu G, Epstein JH, et al. Isolation and characterization of a bat SARS-like

coronavirus that uses the ACE2 receptor. Nature. 2013; 503 (7477): 535-8.

9. Der Sarkissian S, Grobe JL, Yuan L, Narielwala DR, Walter GA, Katovich MJ, et al. Cardiac overexpression of angiotensin converting enzyme 2 protects the heart from ischemia-induced pathophysiology. Hypertension. 2008; 51 (3): 712-8.

10. Coleman CM, Liu YV, Mu H, Taylor JK, Massare M, Flyer DC, et al. Purified coronavirus spike protein nanoparticles induce coronavirus neutralizing antibodies in mice. Vaccine. 2014; 32 (26): 3169-74.

11. Chen W-H, Chag SM, Poongavanam MV, Biter AB, Ewere EA, Rezende W, et al. Optimization of the production process and characterization of the yeast-expressed SARS-CoV recombinant receptor-binding domain (RBD219-N1), a SARS vaccine candidate. J Pharm Sci. 2017; 106 (8): 1961-70.

12. Pang J, Wang MX, Ang IYH, Tan SHX, Lewis RF, Chen JI-P, et al. Potential rapid diagnostics, vaccine and therapeutics for 2019 novel coronavirus (2019-nCoV): a systematic review. J Clin Med. 2020; 9 (3): 623.

13. Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P, et al. Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. Cell Host Microbe. 2020; 27 (3): 325-8.

14. Tian X, Li C, Huang A, Xia S, Lu S, Shi Z, et al. Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human monoclonal antibody. Emerg Microbes Infect. 2020; 9 (1): 382-5.

15. Ter Meulen J, Van Den Brink EN, Poon LL, Marissen WE, Leung CS, Cox F, et al. Human monoclonal antibody combination against SARS coronavirus: synergy and coverage of escape mutants. PLoS Med. 2006; 3 (7): e237.

16. Kunik V, Ashkenazi S, Ofran Y. Paratome: an online tool for systematic identification of antigen-binding regions in antibodies based on sequence or structure. Nucleic Acids Res. 2012; 40 (W1): W521-4.

17. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. 2016; 44 (W1): W344-50.

18. Zhang QC, Deng L, Fisher M, Guan J, Honig B, Petrey D. PredUs: a web server for predicting protein interfaces using structural neighbors. Nucleic Acids Res. 2011; 39 (Web Server issue): W283-7.

19. Fernández-Recio J. Prediction of protein binding sites and hot spots. Wiley Interdiscip Rev Comput Mol Sci. 2011; 1 (5): 680-98.

20. Sanchez-Garcia R, Sorzano COS, Carazo JM, Segura J. BIPSPI: a method for the prediction of partner-specific protein–protein interfaces. Bioinformatics. 2019; 35 (3): 470-7.

21. Kawabata T. Detection of multiscale pockets on protein surfaces using mathematical morphology. Proteins: Struct Funct Bioinf. 2010; 78 (5): 1195-211.

22. Qin S, Zhou H-X. meta-PPISP: a meta web server for protein-protein interaction site prediction. Bioinformatics. 2007; 23 (24): 3386-7.

23. Zhou H-X, Qin S. Interaction-site prediction for protein complexes: a critical assessment. Bioinformatics. 2007; 23 (17): 2203-9.

24. Dunbar J, Krawczyk K, Leem J, Marks C, Nowak J, Regep C, et al. SAbPred: a structure-based antibody prediction server. Nucleic Acids Res. 2016; 44 (Web Server issue): W474-8.

25. De Vries SJ, Van Dijk M, Bonvin AM. The HADDOCK web server for data-driven biomolecular docking. Nat Protoc. 2010; 5 (5): 883-97.

26. Yuan M, Wu NC, Zhu X, Lee C-CD, So RT, Lv H, et al. A highly conserved cryptic epitope in the receptor-binding domains of SARS-CoV-2 and SARS-CoV. Science. 2020; 368 (6491): 630-3.

27. Jones DS, Silverman AP, Cochran JR. Developing therapeutic proteins by engineering ligand–receptor interactions. Trends Biotechnol. 2008; 26 (9): 498-505.

28. Payandeh Z, Rajabibazl M, Mortazavi Y, Rahimpour A, Taromchi AH. Ofatumumab monoclonal antibody affinity maturation through *in silico* modeling. Iran Biomed J. 2018; 22 (3): 180-92.

29. Payandeh Z, Rajabibazl M, Mortazavi Y, Rahimpour A, Taromchi AH, Dastmalchi S. Affinity maturation and characterization of the ofatumumab monoclonal antibody. J Cell Biochem. 2019; 120 (1): 940-50.

30. Lei C, Fu W, Qian K, Li T, Zhang S, Ding M, et al. Potent neutralization of 2019 novel coronavirus by recombinant ACE2-Ig. Biorxiv. 2020.

31. Alhenc-Gelas F, Drueke TB. Blockade of SARS-CoV-2 infection by recombinant soluble ACE2. Kidney Int. 2020; 97 (6): 1091-3.

32. Keller MA, Stiehm ER. Passive immunity in prevention and treatment of infectious diseases. Clin Microbiol Rev. 2000; 13 (4): 602-14.

33. Antikainen NM, Martin SF. Altering protein specificity: techniques and applications. Bioorg Med Chem. 2005; 13 (8): 2701-16.

34. Edelheit O, Hanukoglu A, Hanukoglu I. Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies. BMC Biotechnol. 2009; 9 (1): 61-8.

35. Eyvazi S, Payandeh Z, Khalili S, Sefid F, Ouladsahebmadarek E. Bevacizumab Antibody Affinity Maturation to Improve Ovarian Cancer Immunotherapy: *In Silico* Approach. Int J Pept Res Ther. 2019; 25 (4): 1417-30.

36. Park T, Lee S-Y, Kim S, Kim MJ, Kim HG, Jun S, et al. Spike protein binding prediction with neutralizing antibodies of SARS-CoV-2. BioRxiv. 2020.

37. Saini S, Agarwal M, Pradhan A, Pareek S, Singh A, Dhawan G, et al. Affinity maturation of cross-reactive CR3022 antibody against the receptor-binding domain of SARS-CoV-2 via *in silico* site-directed mutagenesis. ResearchSquare. 2021.

Sefid et al.

Cite this article: ───

Sefid F, Payandeh Z, Khalesi B, Mansoori B, Touhidinia M. Enhancement of SARS-CoV-2 Receptor Binding Domain - CR3022 Human Antibody Binding Affinity via *In silico* Engineering Approach. J Med Microbiol Infect Dis, 2021; 9 (3): 156-169. DOI: 10.52547/JoMMID.9.3.156