biomolecules

MDPI

Article

In Silico Protein Folding Prediction of COVID-19 Mutations and Variants

Sumana Bhowmick †, Tim Jing †, Wei Wang †, Elena Y. Zhang, Frank Zhang and Yanmin Yang *

Department of Neurology and Neurological Sciences, School of Medicine, Stanford University, 1201 Welch Road, MSLS, P259, Stanford, CA 94305, USA

* Correspondence: yyanmin@stanford.edu

† These authors contributed equally to this work.

Abstract: With its fast-paced mutagenesis, the SARS-CoV-2 Omicron variant has threatened many societies worldwide. Strategies for predicting mutagenesis such as the computational prediction of SARS-CoV-2 structural diversity and its interaction with the human receptor will greatly benefit our understanding of the virus and help develop therapeutics against it. We aim to use protein structure prediction algorithms along with molecular docking to study the effects of various mutations in the Receptor Binding Domain (RBD) of the SARS-CoV-2 and its key interactions with the angiotensin-converting enzyme 2 (ACE-2) receptor. The RBD structures of the naturally occurring variants of SARS-CoV-2 were generated from the WUHAN-Hu-1 using the trRosetta algorithm. Docking (HADDOCK) and binding analysis (PRODIGY) between the predicted RBD sequences and ACE-2 highlighted key interactions at the Receptor-Binding Motif (RBM). Further mutagenesis at conserved residues in the Original, Delta, and Omicron variants (P499S and T500R) demonstrated stronger binding and interactions with the ACE-2 receptor. The predicted T500R mutation underwent some preliminary tests in vitro for its binding and transmissibility in cells; the results correlate with the in-silico analysis. In summary, we suggest conserved residues P499 and T500 as potential mutation sites that could increase the binding affinity and yet do not exist in nature. This work demonstrates the use of the trRosetta algorithm to predict protein structure and future mutations at the RBM of SARS-CoV-2, followed by experimental testing for further efficacy verification. It is important to understand the protein structure and folding to help develop potential therapeutics.

Keywords: trRosetta; AlphaFold; SARS-CoV-2; receptor-binding motif; angiotensin-converting enzyme 2 receptor; HADDOCK

check for updates

Citation: Bhowmick, S.; Jing, T.; Wang, W.; Zhang, E.Y.; Zhang, F.; Yang, Y. In Silico Protein Folding Prediction of COVID-19 Mutations and Variants. *Biomolecules* **2022**, *12*, 1665. https://doi.org/10.3390/biom12111665

Academic Editor: Alessandro Paiardini

Received: 29 September 2022
Accepted: 9 November 2022
Published: 10 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The determination of protein 3D structure is crucial to understand its biological function. However, it is time-consuming and expensive. For several decades, the prediction of protein structure directly from sequence information has been an unachievable dream. The advancements in different computational methods to predict protein tertiary structures have now allowed one to study folds and local motifs, molecular folding, evolution, and structure/function relationships in proteins. The BigNet model was the first proposed application of a back propagation neural network (BPNN) to predict the protein 2D and 3D structures [1]. In 1994, Burkhard Rost et al. used evolutionary information as multiple sequence alignments in neural networks as an input which significantly increased the accuracy of the protein prediction [2]. K. T. Simons et al. used simulated annealing and Bayesian scoring function to generate tertiary protein structure based on fragments of unrelated proteins [3]. To date, various methods based on machine learning programs, such as feedforward neural networks, two input neurons, radial basis function neural networks, two neural networks (SPINE-2D), SVM-based contract predictors [4–8], and deep learning programs, such as bidirectional recurrent neural network, residue-residue

coevolution, a balanced network deconvolution (BND), and deep neural network [9–12], have been proposed to find protein tertiary structure. However, proteins usually function through interactions with other proteins or molecules and predicting protein interactions has been even more challenging. For each structure prediction, many short simulations starting from different random sequences are carried out to generate an ensemble of decoy structures that have both favorable local interactions and protein-like global properties [13]. The substantial progress of blind protein prediction using protein ranking refinement and prediction of multimeric complexes has been acknowledged in the Critical Assessment of Structure Prediction (CASP) competition [14–17]. Different techniques have been developed, and benchmarks have been drawn to elucidate the advantages and disadvantages of different methods. For instance, Google's DeepMind developed AlphaFold, a highly ranking method [12]. Another example is the Zhang server, built on iTasser and QUARK [14]. Protein structure prediction using AlphaFold and Rosetta algorithms have been widely used methods and perhaps the most successful methods for de novo protein structure prediction since its inception [12,13,18]. Additionally, several databases and coevolution have been used to predict these interactions [19–21].

DeepMind developed AlphaFold1, which ranked at the top in CASP13 [22]. Inspired by its successes, Baker lab proposed that limitations in AlphaFold1 could be further improved by using inter-residue orientations as well as distance measurements [23]. "trRosetta uses both residue-residue distances and orientations, which gives richer information on the structure compared to distances only," explained Baker [24]. The components of the method trRosetta include (1) a deep residual-convolutional network which takes an MSA as the input and outputs information on the relative distances and orientations of all residue pairs in the protein and (2) a fast Rosetta model building protocol based on restrained minimization with distance and orientation restraints derived from the network outputs [25]. trRosetta helps predict inter-residue orientations and distances from co-evolutionary data by applying deep knowledge, which significantly improves protein structure prediction [23] from amino acid sequences. One of the key features of trRosetta is that it originally had de novo modeling. However, later automated template detection by HHsearch was included to improve the accuracy. It has been reported that sequence design using trRosetta has a remarkable ability to capture properties of the energy landscape and consider alternative states that can reduce the occupancy of the desired target structure [24,26]. As per Du et al., since its release in 2020, the standalone package has been downloaded by >2000 registered users [23]. The trRosetta models were shown to fit well with the cryo-electron microscopy experimental data [27,28]. They have been intensively used to understand the structure and function of lipid transporters [29], protein function deficiency [30], the *Magnaporthe oryzae* secretome [31], the structural characterization of S59L for efficient treatment for ALS and FTD diseases [32], 3D modeling of antibody domains [33], predicting vaccine construct [34], and more. The results of these studies have encouraged researchers and beyond. However, in CASP14, DeepMind's AlphaFold2 replaced almost all components of AlphaFold1, solving the single-chain-based protein structure prediction problem and creating a milestone [35].

The COVID-19 pandemic has become an international event leading to international devastation, closing, affecting all aspects of normal life. However, the introduction of vaccines in the first half of 2021 caused an evolutionary shift toward immune-escaping variants. The most notable, Omicron BA1, was discovered in South Africa [1,2]. Thus, it is imperative to understand the reasons why certain mutations were able to cause such significant viral changes in pathogenicity, virulence, and transmissibility [3]. Various studies have proposed or undergone in vitro cultivation and examination of the RNA and protein sequences of the viruses, and their results further our understanding of COVID-19 [4]. However, large-scale prophylactic work has not been conducted as the scope of such laboratory experiments is constrained by the time required to culture and mutate individual pathogens when attempting to evaluate for different mutations. However, experimental structures of many proteins are still not available to date. Thus, recently, computational

methods have been gaining in popularity compared to experimental methods. According to previous studies on vaccine efficacy, the Receptor-Binding Domain (RBD), a part of the S1 subunit in the spike (S) (Figure S1) protein of the SARS-CoV-2 virus, interacts directly with the angiotensin-converting enzyme 2 (ACE-2) (EC:3.4.17.23) receptor within the susceptible host. The S glycoprotein facilitates the introduction of viral material into the host cell [36–38]. The domain is located between residues 319–541 of the spike protein. Thus, any potential mutations within the RBD or the S protein may present significantly stronger mechanisms of cell infiltration, increasing the ability of the virus to infect human cells [39]. A thorough examination of the RBD protein structure and potential mutation residues can allow for predictions and, thus, prevention against even more deleterious variants that may arise in the future, saving lives. A novel and promising solution is the implementation of de novo protein folding algorithms to generate a large database of variants, followed by protein–protein interaction analysis to determine the protein function [9], diagnosis [10], and phylogenetic analysis [11].

In this study, we use de novo or ab initio prediction of the 3D structures of the receptor-binding sequences of COVID-19 using trRosetta algorithm, a separate algorithm from AlphaFold v2. Subsequently, refinement of predicted complexes is conducted using HADDOCK, an information-driven flexible docking approach; and PRODIGY, a binding affinity descriptor based only on structural properties of a protein–protein complex between the RBD mutations and the ACE-2 receptor. We predicted the 3D structures of different naturally occurring variants of COVID-19 and compared them with the existing structure from an experimental PDB structure. This provided insight into variations of the RBD and its interactions with the ACE-2 receptor in relation to infiltration and infectivity, leading us to the residues of the Receptor-Binding Motif (RBM). We further predicted potential mutations at specific conserved RBM residues of different naturally occurring variants using trRosetta. We also predicted the mutations' interactions with the ACE-2 receptor using HADDOCK and PRODIGY. Our study revealed certain mutations yet to be in existence with the potential to have higher binding affinity to the ACE-2 receptor. Additionally, the efficacy of the proposed mutation was further assessed using experimental testing. This study introduces a pipeline methodology that can be used in future studies of COVID-19 and as a guide for strategies to combat new/old diseases.

2. Materials and Methods

2.1. Sequence Identification

The spike protein sequence for the Original Wuhan strain of COVID-19 (Gene ID: 43740568, Sequence: NC_045512.2) was retrieved from the NCBI database (https://www.ncbi.nlm.nih.gov/ accessed on 23 June 2021) in FASTA format. In addition, we generated the sequences of the Alpha, Beta, Gamma, Delta, and Omicron variants by manually editing the original Wuhan strain with the RBD residue mutations provided by the CDC (https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html, accessed on 23 June 2021) (Table S1).

2.2. De Novo Protein Folding and Structural Analysis

The SARS-CoV-2 Receptor-Binding Domain (RBD), residues 319–541, of each variant, were selected for de novo protein folding. We initially used both transform-restrained Rosetta (trRosetta) (https://yanglab.nankai.edu.cn/trRosetta/ accessed on 23 June 2021) and AlphaFold v2.1.0 de novo protein folding algorithms, respectively, for the sequences of the six variants (Original, Alpha, Beta, Gamma, Delta, and Omicron) to construct a 3D computer model of the protein. The multi-sequence alignment (MSA) inputs were conducted using the HH-suite3 script (https://github.com/soedinglab/hh-suite, accessed on 23 June 2021) with operational instructions available on an open-source GitHub repository [13]. The HH-suite3 script uses the UniRef30 database (https://colabfold.mmseqs.com/ accessed on 23 June 2021) of MSAs to search with the target sequence using HHblits with all subsequent instructions detailed in the GitHub ReadMe [14]. The multi-sequence alignment features

were then uploaded into the trRosetta online portal, where MSA files can be uploaded to generate the final PDB model with default parameters [15].

The AlphaFold algorithm AlphaFold v2.1.0 has been released as open source by Deep Mind, with a version available for use on a Colab notebook [12]. The sequences were directly run as per the directions on the Colab notebook (https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb, accessed on 23 June 2021) with enough computational power, such as through utilizing Colab Pro's access to faster GPUs and TPUs. The AlphaFold algorithm produces Protein Data Bank (PDB) files, which is a way to encode a 3D computer protein model.

2.3. Visualizing and Preliminary Analysis

The models generated from trRosetta and AlphaFold v2.1.0 for the naturally occurring variants were then superimposed and aligned using the PyMol software (https://pymol.org/2/ accessed on 23 June 2021) [16] against a laboratory model of the RBD–ACE-2 complex (PDB: 6M0J) retrieved from the Protein Data Bank (https://www.rcsb.org, accessed on 23 June 2021). We utilized the PyMol super command, which super aligns two protein selections. Super does a sequence-independent structure-based dynamic programming alignment (unlike the align command) followed by a series of refinement cycles intended to improve the fit by eliminating pairing with high relative variability [17]. The super command is more reliable than the align command for proteins with low sequence similarity, generating a numerical Root-Mean-Square-Deviation (RMSD) value for quantification.

2.4. Docking Analysis

The receptor-binding domains were then computationally docked with the ACE-2 receptor using High Ambiguity Driven Biomolecular DOCKing (HADDOCK 2.4) [17] (https://wenmr.science.uu.nl/haddock2.4/ accessed on 23 June 2021). All the sequences were divided into two sets; Molecule 1 was the RBDs produced from trRosetta, and Molecule 2 was the ACE-2 receptor from PDB: 6M0J. HADDOCK 2.4, an experimentally based docking program, was used to dock Molecule 1 to Molecule 2 uploaded respectively with a default setting. To launch the docking process, we selected the active residues for Molecule 1, roughly 151-186 aa, which was closest to the ACE-2 receptor (residues 1-48 aa) of PDB: 6M0J's interaction interface that is corroborated by literature [19]. Furthermore, the PRODIGY (protein binding energy prediction) webserver (https://wenmr.science.uu.nl/prodigy/ accessed on 23 June 2021) was used to calculate the binding affinity either as Gibbs free energy (ΔG, kcal/mol) or as a dissociation constant (Kd, M) at 25 °C between the RBD's and the ACE-2 receptor. Molecule 1 and Molecule 2 were labeled as different chains and merged into a single file using PyMol to upload with a default setting [20].

2.5. In Vitro Test

All chemicals are reagent grade and were obtained from Millipore-Sigma-Aldrich Co., St. Louis, MO, USA and all tissue culture reagents were from Life Technologies (Invitrogen TM) Framingham, MA, USA.

1. SARS-CoV-2 Spike protein (Sp) expression constructs: Spike protein (Sp) constructs were used in the experiments: mRFP-Spike Protein (Sp) Original (mRFP-Sp WT, EX-NV219-M55) and Kappa (*B.1.617.1*, mRFP-Sp Kappa, EX-NV248-M55) were purchased from Genecopoeia, Inc. Rockville, MD, U.S.A. Kappa includes the following eight mutations: K19R, G142D, E154K, L452R, E484Q, D614G, P681R, and Q1031H. Human ACE-2 conjugated mono-GFP (ACE-2-GFP, EX-U1285-M98) was also acquired from Genecopoeia, Inc.
    (1) Based on the kappa plasmid, the delta (*B.1.617.2*, mRFP-Sp Delta) was generated in the lab with the following nine mutations: T19R, 156&157del, R158G, L452R, T478K, D614G, P681R, and Q1031H.
    (2) As predicted, Threonine (Thr) at residue 500 of the delta variant was mutated to Arginine (Arg) to obtain the T500R mutations (mRFP-Sp Delta-T).

2. Antibodies: Primary antibodies used were SARS-COV-2 Spike Protein (E7B3E) (Cell Signaling Technology, Inc. Danvers, MA, U.S.A. #63847, rabbit monoclonal, 1:2000 for WB, 1:200 for IF). SARS-COV-2 Spike Protein (2B3E5) (Cell Signaling Technology, Inc. #52342S, mouse monoclonal, 1:2000 for WB, 1:200 for IF). SARS-CoV-2 Spike RBD (R&D Systems, Inc. Minneapolis, MN, U.S.A. AB#105401, mouse monoclonal, 1:2000 for WB, 1:200 for IF). Beta-actin (13E5) (Cell Signaling Technology, Inc. #4970, rabbit monoclonal, 1:2000 for WB, 1:200 for IF). Beta-actin (C4) (Santa Cruz Bio, Santa Cruz, CA, USA, SC-47778, mouse monoclonal, 1:1000 for WB, 1:100 for IF). Secondary antibodies for immunofluorescence were Alexa Fluora® 488- and Alexa Fluor® 594 conjugated secondary antibodies (Jackson ImmunoResearch, West Grove, PA, USA) and IRDye 800CW Goat anti-Rabbit (Li Cor Biosciences, Lincoln, NE, USA, 926-32211, 1:20,000) and IRDye 680LT Goat anti-Mouse IgG (Li Cor Biosciences, 926-68020, 1:20,000) for Western blotting.
3. Cells lines and culture: HEK 293T cell line (CRL-11268™, ATCC) and COS7 cells (CRL-1651™, ATCC) were cultured in Dulbecco's Modified Eagle's Medium (DMEM) with 10% FBS and incubated in a humidified incubator at 5% CO_2 and 37 °C.
4. Soluble Spike proteins preparation from cells (Freeze & Thaw lysis): 293T cells were set up in p100 dishes and transfected with 8 μg of different Sp plasmids with Lipofectamine 3000 (Thermo Fisher Scientific, Waltham, MA, USA) for 2 days. Cells were then scraped and transferred to 1.5 mL microfuge tube with 1 mL ice-cold PBS. After centrifuge, the pellet was resuspended in 500 μL lysis buffer (600 mM KCl, 20 M Tric-Cl (pH 7.8), and 20% Glycerol) with protease inhibitor (Roche Applied Sciences, Penzberg, Germany). The tube was sealed and frozen using liquid nitrogen, then thawed on ice. After three rounds of the "Freeze–Thaw" process, 250U of Benzonase was added to digest DNA for 10 min at RT (or alternatively via sonication). After a quick vortex, the samples were centrifuged with maximum speed for 30 min in a cold room, and 100 μL per tube of supernatants was aliquoted and saved as ligand samples.
5. Ligand bound assay and Western blot analysis: COS7 cells were set up in p60 dishes. After overnight plating, the total amount of Sp proteins (around 300 ng) was added to the cells. Then, 2 h after incubation, the cell dishes were washed twice with PBS for 5 min to remove unbound ligands. A total of 125 μL SDS sample loading buffer was used to lysis cells and collected into the 1.5 mL micro-centrifuge tubes. The cell lysates in the tube were heated at 100 °C for 7 min and centrifuged for 5 min. The standard purified His-Sp (R & D System) was used as control for quantification. One-third of the total lysates were loaded to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gel and transferred to a PVDF membrane (Immobilon-FL membrane, 0.45 μm, IPFL00010). The membranes were blocked in Intercept Blocking buffer (Li Cor Biosciences, 927-60001) for 30 min at room temperature and then incubated overnight at 4 °C with primary antibodies. The next day, membranes were incubated at room temperature for 1 h with secondary antibodies. The fluorescent signal was assayed using an Odyssey Infrared Imaging system (Li Cor Biosciences).
6. Immunofluorescence: COS 7 Cells cultured on coverslips after spike protein adding were fixed in 4% paraformaldehyde for 12 min at room temperature (RT), then permeabilized TBS with 0.1% Triton X-100 for 5 min. After blocking with 5% BSA/TBS for 30 min at RT, the coverslips were incubated with either primary individual antibodies or combined antibodies from different species in at cold room overnight, followed by incubation of primary antibodies overnight, secondary antibodies for 1 h and DAPI-Fluoromount-G® (0100-20; South-ern Biotech, Birmingham, AL, USA) stain. Images were analyzed using a fluorescence microscope (Nikon Image system).
7. Ligand-receptor endosomal assay: COS7 cells were cultured on coverslips overnight and added with ligand (mRFP-Sp) for 2 h, then cells on a coverslip were quickly treated with 0.025% Trypsin for 1 min after being washed with PBS. Next, cells were

stained following the immunofluorescence procedures. Heatmap was considered for the intensity comparison (LUT).

8. Cell in-fusion assay: 293T cells were set up in p60 dishes and individually transfected with 3 µg of different mRFP-Sp plasmids (Sp Delta and Sp Delta T500R), mRFP, eGFP, and ACE-2-GFP with Lipofectamine 3000 for overnight. All transfected cells were trypsinized and resuspended in a 4 mL culture medium. Different combinations were set (eGFP + mRFP, eGFP + mRFP-Sps, ACE-2-eGFP + mRFP, and ACE-2-eGFP + mRFP-Sps), an example like: 0.5 mL of ACE-2-eGFP transfected 293T cells was mixed with 0.5 mL of 293T cells with mRFP-Sp delta, in 15mL falcon tubes for 15 min at 37 °C cell incubator, then diluted in the 3mL culture medium and settled down on the coverslips for live imaging and were monitored for 4 h, 8 h and overnight to check cell in fusion statues. In-fusion percentages of ACE-2-eGFP were compared among different spike protein variants or mutations.
9. Statistical Analysis: The data were analyzed by Student's *t*-test or two-tailed *t*-tests (two-way ANOVA) with the Prism 8 (GraphPad) or Excel (Microsoft) software. *p*-values < 0.05 were considered statistically significant.

3. Results

Previous bioinformatic analyses have shown that the spike protein RBD generates proteins that are essential for host binding and prone to new mutations [40,41]. To investigate and compare the RBD, we initially used the amino acid sequences of the Original variant obtained from NCBI and then manually edited the residues for the other five naturally occurring variants, namely the Alpha variant, Beta variant, Gamma variant, Delta variant, and Omicron variant (Figure S1). These sequences were then processed using trRosetta suite and AlphaFold v2.1.0 individually. Although the experimentally solved structures are available on public databases, our aim was to evaluate the algorithm so that it can regularly predict protein structures even in cases when no similar structure is known/available. The predicted models generated by trRosetta and AlphaFold v2 were then superimposed against the laboratory derived PDB: 6M0J through the "super" command of PyMol (Figure 1). We observed that for most pairs of the trRosetta-generated models and PDB: 6M0J residues, the RMSD ranges from 0.69 to 0.77 Å, and AlphaFold-generated models give RMSD ranges from 0.83 to 1.6 Å. Lower RMSD indicates a good alignment of the RBDs. Thus, we considered trRosetta-generated models for further analysis (Table 1). Qualitative structural changes observed through superimposing the trRosetta-generated naturally occurring variants in Figure 1A-a suggest that the mutations were focused on the interaction interface located on the bottom section of the RBD (Figure 1A-b,A-c). The Receptor-Binding Motif (RBM) is the main functional motif in the RBD, and upon highlighting the residues (Figure 1B-a) of 6M0J, we observe structural changes in comparison to the Delta and Omicron sequences (Figure 1B-b,B-c).

Table 1. RMSD (root mean square deviation of atomic positions) score for both trRosetta-generated and AlphaFold-generated RBD models superimposed on PDB: 6M0J. The root-mean-square deviation of atomic positions is the measure of the average distance (Å) between the atoms of superimposed proteins.

Models	Root-Mean-Square Deviation (RMSD) for trRosetta-Generated Models (Å)	Root-Mean-Square Deviation (RMSD) for AlphaFold v2.1.0-Generated Models (Å)
Original Variant	0.705 for 1434 atoms	1.257 for 1442 atoms
Alpha Variant	0.694 for 1446 atoms	1.596 for 1444 atoms
Beta Variant	0.777 for 1434 atoms	0.969 for 1505 atoms
Gamma Variant	0.762 for 1474 atoms	1.199 for 1488 atoms
Delta Variant	0.785 for 1466 atoms	1.407 for 1370 atoms
Omicron Variant	0.732 for 1421 atoms	0.828 for 1482 atoms

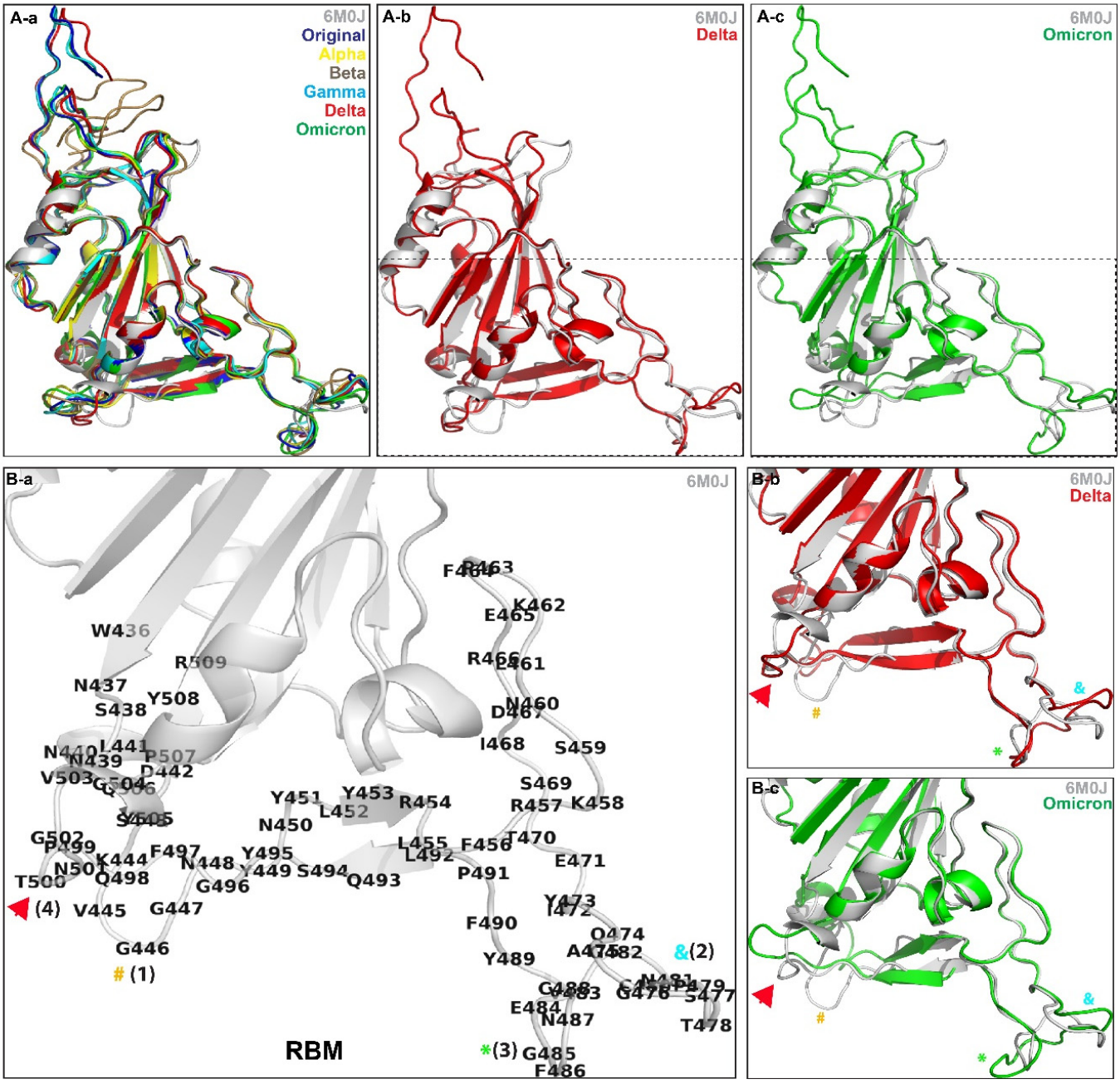


Figure 1. (**A-a**) All the trRosetta-generated Receptor-Binding Domain (RBD) structures were superimposed on 6M0J. (**A-b**) Structural changes between the RBM of 6M0J and the Delta sequence. (**A-c**) Structural changes of the RBM between 6M0J and the Omicron sequence. (**B-a**) RBM of 6M0J; (1), (2), (3), (4) are the changes observed. (**B-b**) Structural changes of the RBM between 6M0J and the Delta sequence. (**B-c**) Structural changes of the RBM between 6M0J and the Omicron sequence.

The trRosetta-generated structures of RBD were then docked with the ACE-2 receptor from 6M0J using the HADDOCK server. The HADDOCK program generated several complexes and clustered them. The cluster with the highest HADDOCK score and lowest electrostatic energy was selected. The best complex was chosen based on cluster size, HADDOCK score, and electrostatic energy. The HADDOCK generates water-refined interaction conformations of both interacting molecules and groups similar conformations in clusters. The superimposed structures of all the complex structures in one cluster were generated along with the plots of HADDOCK scores, Cluster size, RMSD, Van der Waals

energy, electrostatic energy, desolvation energy, restraint violation energy, buried surface area, and Z-score. The best possible interaction complex for all receptors with the best HADDOCK score was then selected for refinement (Table S2). Considering the fact that both the Original variant sequence and the crystal structure of SARS-CoV-2 spike RBD in 6M0J was from the Wuhan strain of COVID-19, the HADDOCK scores from the RBD-ACE2 complex of the Original variant and the 6M0J were taken as thresholds. The docked complexes were further analyzed in PRODIGY for the number and types of interactions between the RBD and the ACE-2 receptor. PRODIGY generates a binding affinity value (ΔG) and a total number of interactions between both molecules. Additionally, the information on the total number of charged–charged interactions, charged–polar, charged–apolar, polar–polar, polar–apolar, and apolar–apolar interactions was generated in the analysis provided in Table S3.

Previous literature has established that the spike protein of SARS-CoV-2 binds to the ACE2 receptor via its receptor-binding domain (RBD). The cartoon representation of the refined structure and interaction complexes generated by PRODIGY was then visualized on PyMol (Figure 2). We observe that the binding affinity (ΔG) of the Alpha and Omicron variants are the lowest (−14.5 kcal/mol and −14.2 kcal/mol, respectively). The higher binding interaction with the ACE2 receptor might reflect a greater ability for potential immune evasion risk and high transmissibility [42] as we have observed with the Omicron variant. We also observe the presence of hydrogen bonding between G496 of the RBM and K671 of the ACE-2 receptor of the Original variant. In addition, T500 changes for the Alpha, Beta, and Gamma variants, where we can also see it bind to P499 of the RBM (Figure 2D–F). However, the binding of G496 is consistent between the Delta and Omicron variants, even with the natural mutation of G496S in Omicron (Figure 2G,H).

Considering the variations of interactions observed between the residues of the RBM of the naturally occurring variants binding to K671 of the ACE-2 receptor, we generated new variants with mutations on the corresponding amino acids on the RBM of the naturally occurring variants and grouped them as Variant 1 to Variant 11 (Table 2). To understand if the change in the interaction between K671 and P499/T500 affects the binding affinity, Variant 1 was generated with all the mutations found in the Alpha and Delta variants. Then, we generated various known point mutations in the Original variant, such as N501Y (Variant 2) and S494P (Variant 3). Since S494P is absent in the Delta and Omicron variants (Figure S1), we generated those mutations to see the alterations in their interactions (Variants 4 and 5). We then generated mutations at P499 and T500, as these were part of conserved residues in the naturally occurring variants. Considering the Mutagenetix database, http://mutagenetix.utsouthwestern.edu accessed on 23 June 2021, we selected the most possible incidental mutation of Proline to Serine (Variant 6, 7, and 8) for P499 and the least possible mutation of Threonine to Arginine (Variant 9, 10, and 11) for T500.

All the "artificial" variants underwent the same aforementioned methodology to generate a PDB structure using trRosetta with RMSD ranging from 0.59 to 0.81 Å, which indicated a good alignment of the RBDs (Table S4) and was further analyzed using HADDOCK and PRODIGY.

The trRosetta-generated structures of RBDs of the artificially generated variants were then docked with the ACE-2 receptor from PDB: 6M0J using the HADDOCK server. Variants 2, 3, 6, 7, and 8 showed lower HADDOCK scores compared to the naturally occurring variants, suggesting a stronger interaction between their RBDs and the ACE-2 receptor (Table S5). Taking into account that Variant 2 [43] and Variant 3 [43] are already recognized and have been reported to enhance the affinity of RBDs to the ACE-2 receptor, we considered looking through the other mutations that currently do not exist, such as Variant 6 (−123.5 ± 1.2), Variant 7 (−111.8 ± 2.9), and Variant 8 (−127.1 ± 3.1). These were further analyzed in PRODIGY for the number and types of interactions (Table S6).

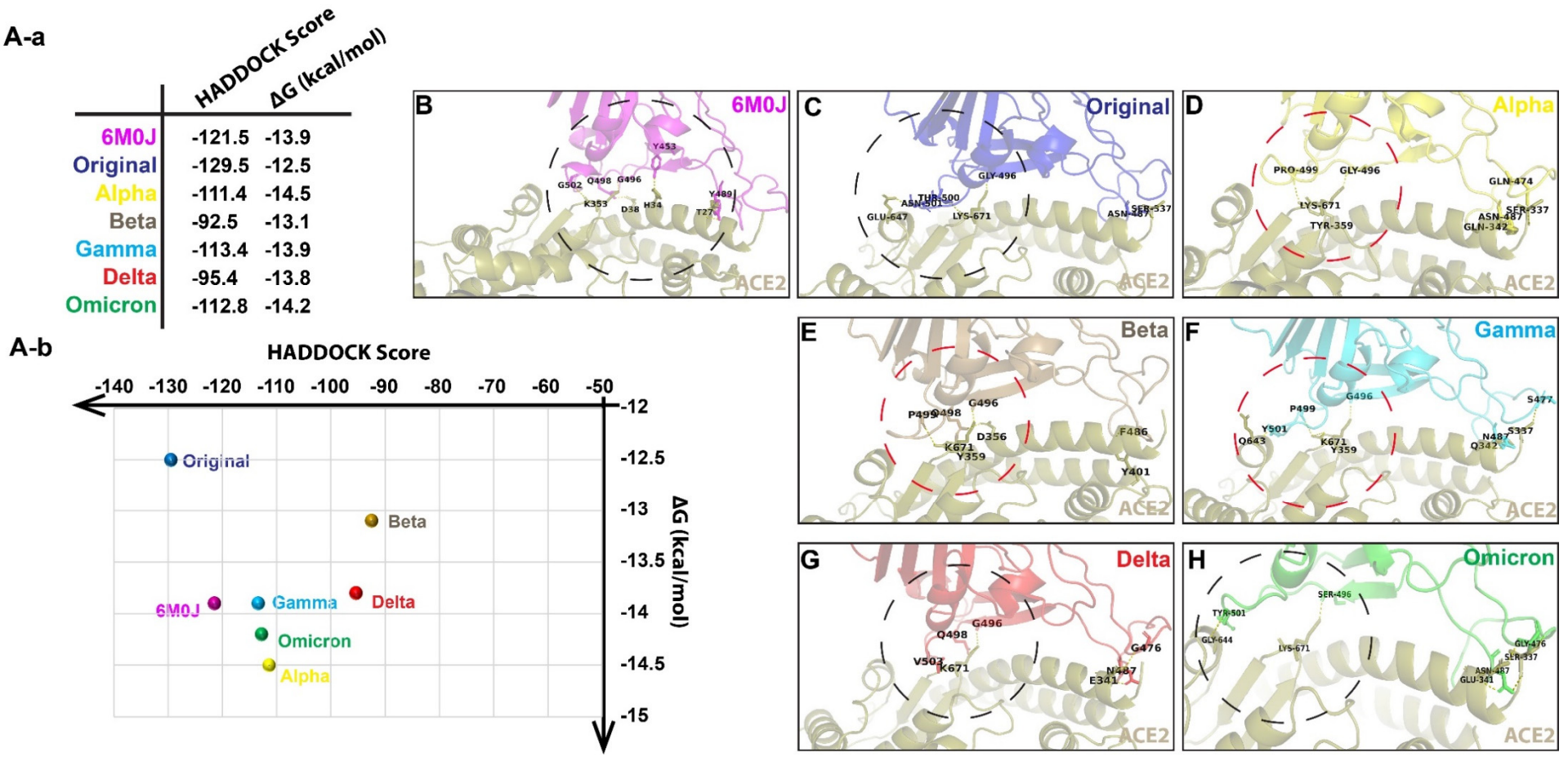


Figure 2. (**A-a**)The HADDOCK and PRODIGY analysis of interactions between the naturally occurring variants' RBDs and the ACE-2 receptor from 6M0J. (**A-b**) A distribution map of the scores obtained from HADDOCK and PRODIGY. (**B**) A cartoon representation of the interaction between 6M0J RBD and the ACE-2 receptor. (**C**) A cartoon representation of the interaction between the Original RBD and the ACE-2 receptor. (**D**) A cartoon representation of the interaction between the Alpha RBD and the ACE-2 receptor. (**E**) A cartoon representation of the interaction between the Beta RBD and the ACE-2 receptor. (**F**) A cartoon representation of the interaction between the Gamma RBD and the ACE-2 receptor. (**G**) A cartoon representation of the interaction between the Delta RBD and the ACE-2 receptor. (**H**) A cartoon representation of the interaction between the Omicron RBD and the ACE-2 receptor.

Table 2. Artificially generated variants with their residues mutated that do not naturally exist as a pathogenic variant.

Type of Variants	Residues Mutated
Variant 1	Mutations from both the Alpha and Delta variants
Variant 2	Mutation only at N501Y residue in the Original variant
Variant 3	Mutation only at S494P residue in the Original variant
Variant 4	Mutation only at S494P residue in the Delta variant
Variant 5	Mutation only at S494P residue in the Omicron variant
Variant 6	Mutation only at P499S residue in the Original variant
Variant 7	Mutation only at P499S residue in the Delta variant
Variant 8	Mutation only at P499S residue in the Omicron variant
Variant 9	Mutation only at T500R residue in the Original variant
Variant 10	Mutation only at T500R residue in the Delta variant
Variant 11	Mutation only at T500R residue in the Omicron variant

The cartoon representation of the refined structure and interaction complexes generated by PRODIGY is provided in Figure 3B,D. The P499S mutation in Delta showed new interactions of S499 with T41 of the ACE-2 receptor (H-bond length of 3.2 Å), G498 with A38 (H-bond length of 2.1 Å), and that G496 bonds with L353 (H-bond length of 2.5 Å) instead of K671 (H-bond length of 3.2 Å) (Figure 3B). The P499S mutation in Omicron shows interactions of T500 and S496 with K671 of the ACE-2 receptor (H-bond length of 3.5 Å and 3.2 Å, respectively), which was initially absent in the naturally occurring Omicron variant. The mutation also changes the interaction of Y501 with G644 to Y501 with T642 (Figure 3D). The interaction was further studied through PRODIGY, where the presence of P499S in Delta and Omicron showed binding affinities of −12.2 kcal/mol and −14.5 kcal/mol, respectively, in comparison with Delta to ACE-2 (−13.8 kcal/mol) and Omicron to ACE-2 (−14.2 kcal/mol) (Figure 3). Additionally, both P499S mutations in Delta and Omicron show increased electrostatic energy in comparison to the mutation in the Original variant, which might be responsible for the higher binding affinity. To further understand the interaction of T500 with K671, we created T500R mutations in the Original, Delta, and Omicron variants. We observed a major difference in the binding affinity between the naturally occurring variants and the mutated variants. This was reflected in the cartoon representation, which suggests a complete loss of the H-bond of K671 in Delta and of G644 in Omicron. However, both mutations still show a decent binding affinity compared to other variants (Figure 3). Additionally, variations in electrostatic energy and Van der Waals energy were clearly observed, which might be responsible for the alterations in binding affinity.

To determine the efficacy of whether the predicted T500R mutation changes the cell binding affinity in vitro, we carried out some preliminary tests for its binding to the spike protein and infiltration using a cell-cell fusion assay. Initially, plasmids of the Delta variant and the T500R Delta variant were made in conjugation with mRFP as described in Materials and Methods (Figure S1). To prepare the ligand in natural statue, soluble lysates were prepared via the "Freeze–Thaw" method from transfected 293T cells (Figure S2A). After filtering through the low retention 0.25 µm filter, the cell lysates containing spike proteins were added into pre-cultured COS7 cells either on dish for Western blot or coverslip for staining. Compared with mRFP, the mRFP-Sp was detected on cells, with a majority surrounding the membrane, and the 100 ng/mL saturation scale was reached (Figure S2B). To load the same amount of spike proteins, the lysates from the Delta variant and its mutation were quantified via Western blot with standard purified spike protein (purchased from company) (Figure 4A-a). Then, the cell-bound assay was performed to dissect the difference between the Delta variant and its mutation (T500R). Figure 4A-b surprisingly shows that the T500R mutation enhanced the cell bound to 30% (Figure 4C) after statistical analysis and was further confirmed with immunofluorescence (Figure 4B).

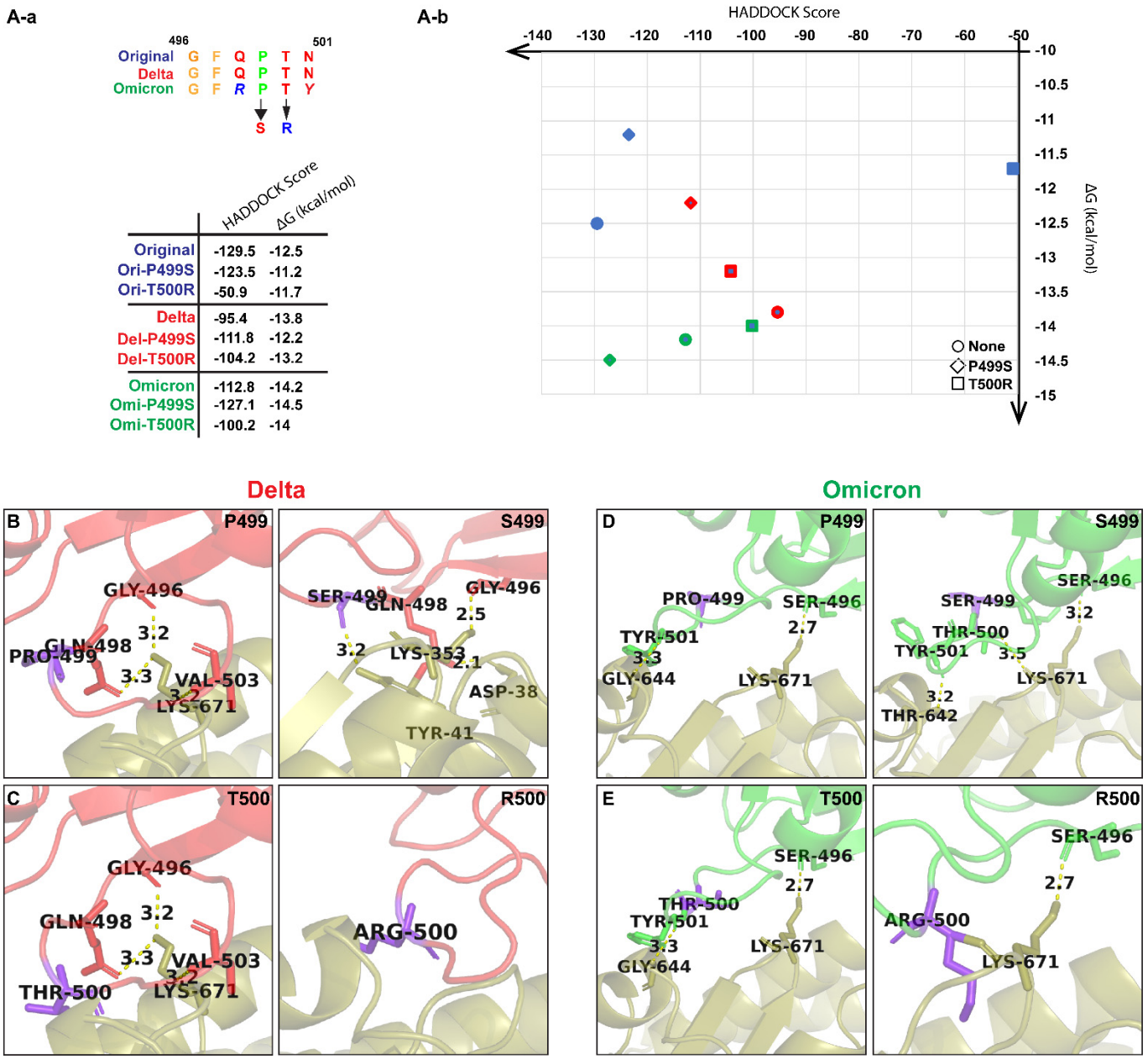


Figure 3. (**A-a**) Mutations in the RBM and the difference between their HADDOCK scores and binding affinity. (**A-b**) A distribution map of the scores obtained from HADDOCK and PRODIGY. (**B**) A cartoon representation of the interaction between P499 and S499 in the Delta RBD and the ACE-2 receptor. (**C**) A cartoon representation of the interaction between T500 and R500 in the Delta RBD and the ACE-2 receptor. (**D**) A cartoon representation of the interaction between P499 and S499 in the Omicron RBD and the ACE-2 receptor. (**E**) A cartoon representation of the interaction between T500 and R500 in the Omicron RBD and the ACE-2 receptor.

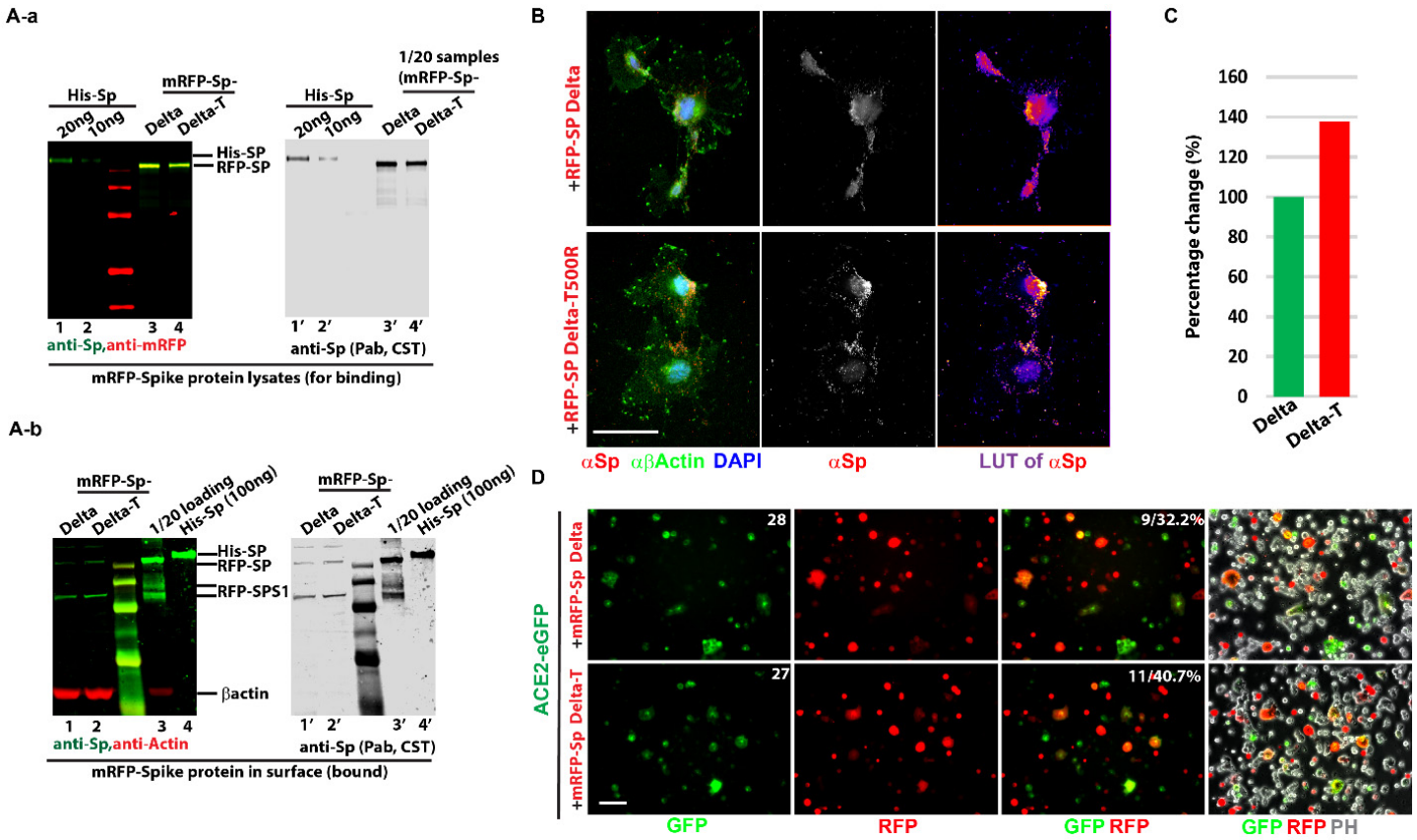


Figure 4. T500R increases the spike protein's cell-bound affinity. (**A-a**) Quantifications of spike protein Delta variant (Delta) and Delta T500R (Delta-T) in transfected cell lysates (before binding). The purchased His-Sp was loaded for quantification. (**A-b**) More bound spike protein on cells was detected in T500R mutation. Spike proteins (Delta and Delta-T) were added to COS7 cells for bound, and the cell uptake level of spike proteins was checked via WB. Three experiments were repeated. (**B**) T500R mutation resulted in more endosomal entering. Delta and Delta-T were added to COS7 cells on coverslip and stained with actin (green) and spike protein (red) after a one-minute treatment with 0.0025% Trypsin to remove surface unbound ligands. Heatmap (LUT) was presented to indicate the intensity of spike protein's stain. Warm color means high intensity. (**C**). Graph representation of the ratio of T500R/Delta (Delta-T/Delta) band's intensity. (**D**) Cell in-fusion assay was performed to show T500R mutation enhances the ACE-2 binding affinity. Cell in-fusion was observed in the combination of ACE2-eGFP with RFP Sp-Delta or Delta-T. The efficacy was calculated as the percentage of Cells No. of "overlap of RFP with GFP"/Cells No. of "GFP". The number was labeled inside the photos. Scale bar: 20 μm.

SARS-CoV-2 uses clathrin-mediated endocytosis to gain access to cells, and thus, endosomes tend to form inside cells once ligands associate with receptors [44]. More endosome formation was observed for the Delta T500R mutation under heat map (LUT) presentation. To investigate whether the increased association of T500R on cells was through ACE-2, ACE-2-eGFP plasmid was transfected and confirmed correct expression pattern via antibody (Figure S3A). Next, ACE-2-eGFP and mRFP-Sps (Delta and its T500R mutation) were transfected for cell in-fusion assay (Figure S3B). This was to monitor the binding activity of the spike proteins by using SARS-CoV-2-S-transfected 293T as effector cells and ACE-2-expressing 293T as target cells [45]. The cells in the fusion phenome were only observed in the combination of ACE-2-eGFP and mRFP-Sps instead of others (Figure S3C). Compared with Delta, the T500R mutation with ACE-2-eGFP exhibited a high capacity for membrane fusion and was efficient in mediating virus fusion and entry into target cells.

The in vitro data suggest that T500R enhances the spike protein's efficacy for fusion via the ACE-2 receptor's binding affinity and correlates with the in-silico results. These results indicate that our algorithm and simulations agree with the in vitro results, and thus mutations at P499 or T500 that equally increase the RBM's binding affinity to ACE-2 could still be a potential target.

4. Discussion

The widely applied and most popular web-based platform trRosetta suite has resolved some of the greatest challenges in protein folding, presenting a new era of opportunities. It highlights the combination of Rosetta-based optimization with predicted information and additional components of the Rosetta energy function, thus generating protein models. This allows researchers to access de novo protein structures in future research, as protein folding remains a challenging field among researchers. Probing this technology and its myriad uses will similarly yield numerous applications that can further our fundamental understanding of life. We here evaluate and demonstrate an accurate and fully automated pipeline to predict protein folding and its key interactions with receptors. This specific approach of using trRosetta algorithm was intriguing to us because it was an available web-based platform and presented a scalable, efficient method for future researchers to conduct preliminary analysis into new mutations and variants that may not yet exist under natural conditions. Additionally, several studies have been reported on protein/peptide prediction of/against novel coronavirus SARS-COV-2 using trRosetta modeling suite [23,46–48]. Instead of time-consuming laboratory work to alter the genetic sequences of pathogens, large-scale computer generation of proteins could provide an avenue to narrow down mutations to study, discern patterns that can only be seen with large quantities of data and identify new therapeutics.

Previous studies conducted on COVID-19 pathogenesis suggest that its spike protein is critical in both ACE-2 receptor binding and viral fusion into the host cell membrane [49]. Indeed, the spike protein has also been the primary immunogen target for the currently available COVID-19 vaccines [50,51]. Until now, a countless number of mutations (e.g., T19R, A67V, D80A, D80G, T95I, G142D, Y145D, R158G, L212I, D215G, A222V, N234Q, W258L, N331Q, G339D, N343Q, R346K S371L, S373P, and S375F) have been reported in the S1 protein. Our approach was first to evaluate both trRosetta and AlphaFold algorithms with the RBD sequence of naturally occurring variants whose mutations were already previously characterized, namely the Original variant from the Wuhan, China outbreak, the Alpha variant, the Beta variant, the Gamma variant, the Delta variant, and the Omicron variant. The generated models were superimposed on the RCSB PDB:6M0J structure. The RMSD values from the AlphaFold varied greatly, so we used the trRosetta-generated models and applied the same methodology to the mutant variants [46]. The aim was to search for and identify a potential mutation with high binding affinity to the ACE-2 receptor in comparison to the known mutations.

This study was mainly focused on investigating the Receptor-Binding Motif (RBM) of the spike protein, as that portion of the protein is what specifically interacts with the ACE-2 receptor during viral infiltration and thus increases the transmissibility. The "super" command on PyMol allowed for a simple yet direct way to establish and measure the differences and accuracy of the RBD structures between different variants. After superimposing the variants, we could see some obvious differences (beta sheets disappearing, alpha helices having different amino acids on them, etc.) (Figure 1). The qualitative structural observations were performed as a preliminary way to compare the variants. The generated models were further validated using the modeling approach considering the interaction between the RBD and the ACE-2 receptor. The interaction was stabilized by several non-covalent interactions and distinctive binding stimulations. We observe a higher binding interaction of the Alpha and Omicron variants in comparison to the Original variant, suggesting that these variants could be highly transmissible. The high transmissibility of the Alpha variant resulted in the origin of the third wave of COVID-19 encountered in France [52]. However, the introduction of vaccinations coupled with acquired collective immunity to this allowed the expansion and circulation of a better-adapted variant, Delta, which eventually was monitored with vaccine introduction. Later, the reduced efficacy of vaccinations contributed to the selection of the Omicron variant, which showed a greater temporal shift [52]

The identification of P499S as a potential deleterious mutation highlights the fact that there are numerous potentially dangerous mutations that could give rise to even stronger variants than the Delta and Omicron variants causing a significant number of cases around the globe. Such mutations are not an improbability either, as the various case surges around the world present ample opportunity for various point mutations to be selected and for these variants to cause significant damage. This has already been proven by the various mutations increasing their presence in the gene pool significantly, as their docking ability may have the potential to supersede currently fixed mutations. On correlating the binding affinity, we also investigated the residue T500 and created the least likely mutagenesis to observe its effect on binding affinity. To evaluate the efficacy of the predicted mutation, we undertook a preliminary investigation of the mutation in vitro by monitoring the spike protein binding and cell-infusion assay, which correlated with the in-silico data. Although further experimental testing on P499S and T500R must be undertaken to confirm these conjectures, it would be irresponsible to ignore the potential threat of a more serious variant mutating at a time when transmission remains high across the globe.

5. Conclusions

In conclusion, we suggest that the higher binding affinity of P499S and T500R indicates that they could be one of the potential mutations yet to be present in nature. Omicron is a known fast-paced variant regarding transmissibility, and as per our data, these mutations show higher binding affinity in comparison to the Omicron variant. This presents additional corroboration that our computational method may be useful in determining potential future mutations. This computational model for variant investigation and de novo prediction requires further refinement. However, it still poses a significant opportunity for researchers to identify and prophylactically predict dangerous variants in relation to emerging diseases. Although the model cannot substitute lab work, it can complement research by successfully presenting a wide variety of potentially deleterious mutations and variants for broad scientific inquiry in the field.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/biom12111665/s1, Figure S1: Scheme of COVID-19 Spike protein. (A) Simple diagram of Spike protein. The three antibodies' recognized regions are shown. RBD: Receptor-Binding Domain; RBM: ACE-2 Receptor-Binding Motif; FP: furin process domain; (B) Amino acid alignment of RBM in all different known variants. (C) Spike mutation of Delta Variant was tagged with mREF in N terminal. T500% mutation was mutated in the RBM mutation. Figure S2: mREF-Spike protein binds to cell membrane. (A) Strategy of cell lysate preparation and ligand-cell-bound assay. 293T cells were transfected with plasmids for two days, and then processed cell lysis was via freeze–thaw

method. After quantification by Western blot, the exact number of protein lysates were filtered to ass in pre-cultured COS 7cells. (B) mRFP-Sp protein was detected majority around cell surface either via RFP (data not shown) or Spike protein antibodies' staining. (C) Binding affinity by dosage. Different concentrations of spike protein lysates were added to measure the binding curve. Around 100 ng spike protein is reaching the station platform. Scale bar: 20 µm. Figure S3: Cell in fusion assay. (A) ACE2-eGFP was transfected in COS7 cells and correctly confirmed expression pattern by ACE2 antibody. (B) 293T cells were separately transfected with eGFP, ACE2-eGFP, mRFP, and mRFP-Sp. Overnight transfected cells were trypsinized, and ACE2-eGFP was mixed with mRFP-Sp for cells in fusion assay. Other combinations were taken as control. (C) Cell in fusion was only observed in the combination of ACE2-eGFP and mRFP-Sp. No infusion phenome was observed in control combinations. Scale bar: 20 µm. Table S1: The list of mutations that each variant possessed compared to the original spike protein sequence from the COVID-19 virus pathogen identified in Wuhan, China, provided by the CDC (https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html, accessed on 23 June 2021). Table S2: The table represents the HADDOCK parameters of naturally occurring variants after refinement. Table S3: The table represents the binding affinity and number of interactions between the trRosetta-generated RBDs and the ACE receptor of 6M0J generated by the PRODIGY server. Table S4: RMSD (root mean square deviation of atomic positions) score for trRosetta-generated mutated RBD models. The root-mean-square deviation of atomic positions is the measure of the average distance (Å) between the atoms of superimposed proteins. Table S5: The table represents the HADDOCK parameters of all the manually mutated variants after refinement. Table S6: The table represents the binding affinity and number of interactions between the trRosetta-generated RBDs and the ACE receptor of 6M0J generated by the PRODIGY server.

Author Contributions: Conceptualization, T.J. and Y.Y.; methodology, T.J., W.W., F.Z. and E.Y.Z.; software, S.B., T.J. and F.Z.; validation, T.J., W.W., S.B. and Y.Y.; formal analysis, T.J., S.B. and W.W.; investigation, W.W. and E.Y.Z.; resources, Y.Y.; data curation, S.B., T.J. and W.W.; writing—original draft preparation, T.J., S.B. and W.W.; writing—review and editing, S.B., T.J., W.W. and Y.Y.; visualization, T.J., W.W. and S.B.; supervision, Y.Y.; project administration, Y.Y.; funding acquisition, Y.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by gift funds from multiple generous donors.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wilcox, G.L.; Poliac, M.; Liebman, M.N. Neural Network Analysis of Protein Tertiary Structure. *Tetrahedron Comput. Methodol.* **1990**, *3*, 191–204. [CrossRef]
2. Rost, B.; Sander, C. Combining Evolutionary Information and Neural Networks to Predict Protein Secondary Structure. *Proteins Struct. Funct. Bioinf.* **1994**, *19*, 55–72. [CrossRef] [PubMed]
3. Simons, K.T.; Kooperberg, C.; Huang, E.; Baker, D. Assembly of Protein Tertiary Structures from Fragments with Similar Local Sequences Using Simulated Annealing and Bayesian Scoring Functions. *J. Mol. Biol.* **1997**, *268*, 209–225. [CrossRef]
4. Bohr, H.; Bohr, J.; Brunak, S.; Cotterill, R.M.J.; Fredholm, H.; Lautrupt, B.; Petersen, S.B. A Novel Approach to Prediction of the 3-Dimensional Structures of Protein Backbones by Neural Networks. *FEBS Lett.* **1990**, *261*, 43–46. [CrossRef]
5. Fariselli, P.; Casadio, R. A Neural Network Based Predictor of Residue Contacts in Proteins. *Protein Eng.* **1999**, *12*, 15–21. [CrossRef]
6. Xue, B.; Faraggi, E.; Zhou, Y. Predicting Residue-Residue Contact Maps by a Two-Layer, Integrated Neural-Network Method. *Proteins Struct. Funct. Bioinf.* **2009**, *76*, 176–183. [CrossRef]
7. Wu, S.; Szilagyi, A.; Zhang, Y. Improving Protein Structure Prediction Using Multiple Sequence-Based Contact Predictions. *Structure* **2011**, *19*, 1182–1191. [CrossRef]
8. Zhang, G.Z.; Huang, D.S.; Quan, Z.H. Combining a Binary Input Encoding Scheme with RBFNN for Globulin Protein Inter-Residue Contact Map Prediction. *Pattern Recognit. Lett.* **2005**, *26*, 1543–1553. [CrossRef]
9. Vullo, A.; Walsh, I.; Pollastri, G. A Two-Stage Approach for Improved Prediction of Residue Contact Maps. *BMC Bioinf.* **2006**, *7*, 180. [CrossRef]
10. Ovchinnikov, S.; Kim, D.E.; Wang, R.Y.R.; Liu, Y.; Dimaio, F.; Baker, D. Improved de Novo Structure Prediction in CASP11 by Incorporating Coevolution Information into Rosetta. *Proteins Struct. Funct. Bioinf.* **2016**, *84*, 67–75. [CrossRef]

11. Sun, H.P.; Huang, Y.; Wang, X.F.; Zhang, Y.; Shen, H. Bin Improving Accuracy of Protein Contact Prediction Using Balanced Network Deconvolution. *Proteins Struct. Funct. Bioinf.* **2015**, *83*, 485–496. [CrossRef]
12. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [CrossRef]
13. Rohl, C.A.; Strauss, C.E.M.; Misura, K.M.S.; Baker, D. Protein Structure Prediction Using Rosetta. *Methods Enzymol.* **2004**, *383*, 66–93. [CrossRef]
14. Moult, J.; Fidelis, K.; Kryshtafovych, A.; Schwede, T.; Tramontano, A. Critical Assessment of Methods of Protein Structure Prediction (CASP)—Round XII. *Proteins Struct. Funct. Bioinf.* **2018**, *86*, 7–15. [CrossRef] [PubMed]
15. Kryshtafovych, A.; Schwede, T.; Topf, M.; Fidelis, K.; Moult, J. Critical Assessment of Methods of Protein Structure Prediction (CASP)-Round XIII. *Proteins Struct. Funct. Bioinf.* **2019**, *87*, 1011–1020. [CrossRef] [PubMed]
16. Kryshtafovych, A.; Schwede, T.; Topf, M.; Fidelis, K.; Moult, J. Critical Assessment of Methods of Protein Structure Prediction (CASP)-Round XIV. *Proteins Struct. Funct. Bioinf.* **2021**, *89*, 1607–1617. [CrossRef] [PubMed]
17. Liu, S.; Gao, Y.; Vakser, I.A. DOCKGROUND Protein-Protein Docking Decoy Set. *Bioinf. Appl. NOTE* **2008**, *24*, 2634–2635. [CrossRef]
18. Koehler Leman, J.; Weitzner, B.D.; Lewis, S.M.; Adolf-Bryfogle, J.; Alam, N.; Alford, R.F.; Aprahamian, M.; Baker, D.; Barlow, K.A.; Barth, P.; et al. General Overview and Challenges Macromolecular Modeling and Design in Rosetta: Recent Methods and Frameworks. *Nat. Methods* **2020**, *17*, 665–680. [CrossRef]
19. Orchard, S.; Ammari, M.; Aranda, B.; Breuza, L.; Briganti, L.; Broackes-Carter, F.; Campbell, N.H.; Chavali, G.; Chen, C.; Del-Toro, N.; et al. The MIntAct Project-IntAct as a Common Curation Platform for 11 Molecular Interaction Databases. *Nucleic Acids Res.* **2013**, *42*, D358–D363. [CrossRef]
20. Szklarczyk, D.; Gable, A.L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N.T.; Morris, J.H.; Jensen, L.J.; et al. STRING V11: Protein-Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets. *Nucleic Acids Res.* **2018**, *47*, 607–613. [CrossRef]
21. Cong, Q.; Anishchenko, I.; Ovchinnikov, S.; Baker, D. Protein Interaction Networks Revealed by Proteome Coevolution. *Science* **2019**, *365*, 185–189. [CrossRef] [PubMed]
22. Senior, A.W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Žídek, A.; Nelson, A.W.R.; Bridgland, A.; et al. Improved Protein Structure Prediction Using Potentials from Deep Learning. *Nature* **2020**, *577*, 706–710. [CrossRef] [PubMed]
23. Du, Z.; Su, H.; Wang, W.; Ye, L.; Wei, H.; Peng, Z.; Anishchenko, I.; Baker, D.; Yang, J. The TrRosetta Server for Fast and Accurate Protein Structure Prediction. *Nat. Protoc.* **2021**, *16*, 5634–5651. [CrossRef] [PubMed]
24. Singh, A. Deep Learning 3D Structures. *Nat. Methods* **2020**, *17*, 249. [CrossRef] [PubMed]
25. Jianyi, Y.; Ivan, A.; Hahnbeom, P.; Zhenling, P.; Sergey, O.; David, B. Improved Protein Structure Prediction Using Predicted Interresidue Orientations. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 1496–1503. [CrossRef]
26. Norn, C.; Wicky, B.I.M.; Juergens, D.; Liu, S.; Kim, D.; Tischer, D.; Koepnick, B.; Anishchenko, I.; Players, F.; Baker, D.; et al. Protein Sequence Design by Conformational Landscape Optimization. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2017228118. [CrossRef]
27. Pleiner, T.; Tomaleri, G.P.; Januszyk, K.; Inglis, A.J.; Hazu, M.; Voorhees, R.M. Structural Basis for Membrane Insertion by the Human ER Membrane Protein Complex. *Science* **2020**, *369*, 433–436. [CrossRef]
28. O'donnell, J.P.; Phillips, B.P.; Yagita, Y.; Juszkiewicz, S.; Wagner, A.; Malinverni, D.; Keenan, R.J.; Miller, E.A.; Hegde, R.S. The Architecture of EMC Reveals a Path for Membrane Protein Insertion. *eLife* **2020**, *9*, e57887. [CrossRef]
29. Melia, T.J.; Reinisch, K.M. A Possible Role for VPS13-Family Proteins in Bulk Lipid Transfer, Membrane Expansion and Organelle Biogenesis. *J. Cell Sci.* **2022**, *135*, jcs259357. [CrossRef]
30. Yu, T.; Hu, Y.; Zhang, Y.; Zhao, R.; Yan, X.; Dayananda, B.; Wang, J.; Jiao, Y.; Li, J.; Yi, X. Whole-Genome Sequencing of Acer Catalpifolium Reveals Evolutionary History of Endangered Species GBE. *Genome Biol. Evol.* **2021**, *13*, evab271. [CrossRef]
31. Seong, K.; Krasileva, K. V Computational Structural Genomics Unravels Common Folds and Novel Families in the Secretome of Fungal Phytopathogen Magnaporthe Oryzae. *Mol. Plant-Microbe Interact.* **2021**, *34*, 1267–1280. [CrossRef] [PubMed]
32. Alici, H.; Uversky, V.N.; Kang, D.E.; Woo, A.; Coskuner-Weber, O. Structures of the Wild-Type and S59L Mutant CHCHD10 Proteins Important in Amyotrophic Lateral Sclerosis−Frontotemporal Dementia. *ACS Chem. Neurosci.* **2022**, *13*, 1273–1280. [CrossRef] [PubMed]
33. Vishwakarma, P.; Melarkode Vattekatte, A.; Shinada, N.; Diharce, J.; Martins, C.; Cadet, F.; Gardebien, F.; Etchebest, C.; Nadaradjane, A.A.; De Brevern, A.G. V_HH Structural Modelling Approaches: A Critical Review. *Int. J. Mol. Sci.* **2022**, *23*, 3721. [CrossRef]
34. Jaan, S.; Zaman, A.; Ahmed, S.; Shah, M.; Chandra Ojha, S. MRNA Vaccine Designing Using Chikungunya Virus E Glycoprotein through Immunoinformatics-Guided Approaches. *Vaccines* **2022**, *10*, 1476. [CrossRef]
35. Callaway, E. "It Will Change Everything": DeepMind's AI Makes Gigantic Leap in Solving Protein Structures. *Nature* **2020**, 588. [CrossRef] [PubMed]
36. Belouzard, S.; Millet, J.K.; Licitra, B.N.; Whittaker, G.R. Mechanisms of Coronavirus Cell Entry Mediated by the Viral Spike Protein. *Viruses* **2012**, *4*, 1011–1033. [CrossRef] [PubMed]
37. Nagesha, S.N.; Ramesh, B.N.; Pradeep, C.; Shashidhara, K.S.; Ramakrishnappa, T.; Krishnaprasad, B.T.; Jnanashree, S.M.; Manohar, M.; Arunkumar, N.; Yallappa; et al. SARS-CoV 2 Spike Protein S1 Subunit as an Ideal Target for Stable Vaccines: A Bioinformatic Study. *Mater. Today Proc.* **2021**, *49*, 904–912. [CrossRef]

38. Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; et al. Structure of the SARS-CoV-2 Spike Receptor-Binding Domain Bound to the ACE2 Receptor. *Nature* **2020**, *581*, 215–220. [CrossRef]
39. Saha, P.; Kumar Banerjee, A.; Tripathi, P.P.; Srivastava, A.K.; Ray, U.; Banerjee, A.K. A Virus That Has Gone Viral: Amino Acid Mutation in S Protein of Indian Isolate of Coronavirus COVID-19 Might Impact Receptor Binding, and Thus, Infectivity. *Biosci. Rep.* **2020**, *40*, 20201312. [CrossRef]
40. Chen, J.; Wang, R.; Wang, M.; Wei, G.W. Mutations Strengthened SARS-CoV-2 Infectivity. *J. Mol. Biol.* **2020**, *432*, 5212–5226. [CrossRef]
41. Andersen, K.G.; Rambaut, A.; Lipkin, W.I.; Holmes, E.C.; Garry, R.F. The Proximal Origin of SARS-CoV-2. *Nat. Med.* **2020**, *26*, 450–452. [CrossRef] [PubMed]
42. Alexander, M.R.; Schoeder, C.T.; Brown, J.A.; Smart, C.D.; Moth, C.; Wikswo, J.P.; Capra, J.A.; Meiler, J.; Chen, W.; Madhur, M.S. Which Animals Are at Risk? Predicting Species Susceptibility to COVID-19. *bioRxiv* **2020**, bioRxiv:2020.07.09.194563.
43. Alenquer, M.; Ferreira, F.; Lousa, D.; Valério, M.; Medina-Lopes, M.; Bergman, M.L.; Gonçalves, J.; Demengeot, J.; Leite, R.B.; Lilue, J.; et al. Signatures in SARS-CoV-2 Spike Protein Conferring Escape to Neutralizing Antibodies. *PLoS Pathog.* **2021**, *17*, e1009772. [CrossRef] [PubMed]
44. Bayati, A.; Kumar, R.; Francis, V.; McPherson, P.S. SARS-CoV-2 Infects Cells after Viral Entry via Clathrin-Mediated Endocytosis. *J. Biol. Chem.* **2021**, 296. [CrossRef] [PubMed]
45. Wang, L.; Zhao, J.; Nguyen, L.N.T.; Adkins, J.L.; Schank, M.; Khanal, S.; Nguyen, L.N.; Dang, X.; Cao, D.; Krishna, B.; et al. Blockade of SARS-CoV-2 Spike Protein-Mediated Cell-Cell Fusion Using COVID-19 Convalescent Plasma. *Sci. Rep.* **2021**, *11*, 5558. [CrossRef]
46. Heo, L.; Feig, M. Modeling of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Proteins by Machine Learning and Physics-Based Refinement. *bioRxiv* **2020**, bioRxiv:2020.03.25.008904. [CrossRef]
47. Kalita, P.; Padhi, A.K.; Zhang, K.Y.J.; Tripathi, T. Design of a Peptide-Based Subunit Vaccine against Novel Coronavirus SARS-CoV-2. *Microb. Pathog.* **2020**, *145*, 104236. [CrossRef]
48. Tripathy, S.P.; Ponnapati, M.; Jacobson, J.M.; Chatterjee, P. Sub-Picomolar Detection of SARS-CoV-2 RBD via Computationally-Optimized Peptide Beacons. *bioRxiv* **2021**, bioRxiv:2021.06.04.447114.
49. Huang, Y.; Yang, C.; Xu, X.-F.; Xu, W.; Liu, S.-W. Structural and Functional Properties of SARS-CoV-2 Spike Protein: Potential Antivirus Drug Development for COVID-19. *Acta Pharmacol. Sin.* **2020**, *41*, 1141–1149. [CrossRef]
50. Wrobel, A.G.; Benton, D.J.; Benton, D.J.; Xu, P.; Roustan, C.; Martin, S.R.; Rosenthal, P.B.; Skehel, J.J.; Gamblin, S.J. SARS-CoV-2 and Bat RaTG13 Spike Glycoprotein Structures Inform on Virus Evolution and Furin-Cleavage Effects. *Nat. Struct. Mol. Biol.* **2020**, *27*, 763–767. [CrossRef]
51. Heinz, F.X.; Stiasny, K. Distinguishing Features of Current COVID-19 Vaccines: Knowns and Unknowns of Antigen Presentation and Modes of Action. *NPJ Vaccines* **2021**, *6*, 104. [CrossRef] [PubMed]
52. Wurtzer, S.; Levert, M.; Dhenain, E.; Accrombessi, H.; Manco, S.; Fagour, N.; Goulet, M.; Boudaud, N.; Gaillard, L.; Bertrand, I.; et al. From Alpha to Omicron BA.2: New Digital RT-PCR Approach and Challenges for SARS-CoV-2 VOC Monitoring and Normalization of Variant Dynamics in Wastewater. *Sci. Total Environ.* **2022**, *848*, 157740. [CrossRef] [PubMed]