Plausible blockers of Spike RBD in SARS-CoV2 – Molecular design and underlying interaction dynamics from high-level structural descriptors

Sankar Basu[1] *, Pampa Saha[2§], Dhananjay Bhattacharyya[3§], Hirak K Patra[4§]

[1] Department of Microbiology, Asutosh College (affiliated to University of Calcutta), Kolkata 700026, India
[2] Department of Neurological Surgery, University of Pittsburgh, Pittsburgh 15213, USA.
[3] Computational Science Division, Saha Institute of Nuclear Physics, Kolkata 700064, India
[4] Department of Surgical Biotechnology, Division of Surgery and Interventional Science, University College London, London, NW3 2PF, United Kingdom

[§] These authors contributed equally to the work
*Address for all correspondences: SB (nemo8130@gmail.com)

Abstract

COVID-19 is characterized by an unprecedented abrupt increase in transmission rate relative to its endemic evolutionary ancestor, SARS-CoV (2003). The complex molecular cascade of events related to the viral pathogenicity is triggered by the Spike protein upon interacting with the ACE2 receptor on human lung cells through its receptor binding domain (RBD_{Spike}). One potential therapeutic strategy to combat COVID-19 could thus be limiting the infection by blocking this key interaction. In this current study, we adapt a protein design approach to predict and propose non-virulent structural mimics of the RBD_{Spike}, potentially serving as its competitive inhibitors in binding to ACE2. RBD_{Spike} is an independently foldable protein domain, resilient to conformational changes upon mutations and therefore an attractive target for strategic re-design. Interestingly, in spite of displaying an optimal shape fit between their interacting surfaces (attributed to a consequently high mutual affinity), the RBD_{Spike} – ACE2 interaction appears to have a quasi-stable character due to a poor electrostatic match at their interface. Structural analyses of homologous complexes reveal that the RBD_{Spike} has an unusually high degree of solvent-exposed hydrophobic residues, attributed to key evolutionary changes, making it inherently 'reaction-prone'. The designed mimics, aimed to block the viral entry by occupying the available binding sites on ACE2, are tested to have signatures of stable high-affinity binding with ACE2, overriding the native quasi-stable feature. The results show the apt of directly adapting natural examples in rational protein design, wherein, homology-based threading coupled with strategic 'hydrophobic ↔ polar' mutations potentially serves as a breakthrough.

Keywords: *COVID-19; SARS-CoV-2; Protein design; Complementarity; Competitive inhibitor; Homology-based threading in rational protein design*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 September 2020 doi:10.20944/preprints202009.0006.v1

1. Introduction

The world is currently facing an unprecedented global health crisis due to sudden pandemic outbreak of a 'naturally evolving' [1] virus known as Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) [2, 3]. The disease condition associated with SARS-CoV-2 known as COVID-19 was first reported in human subjects in the city of Wuhan, China in Dec'2019 [4]. In a span of 8 months, more than 20 million people got infected with a death toll rising to 750000. The situation has challenged the very foundations of our existing global health management system, threatening with economic crisis that has never been faced before. The SARS-CoV-2 is a positive stranded RNA virus and a β-coronavirus. It shares a significant amount of genomic identity (79.5%) with its related previous strain SARS-CoV which got outspread in 2003 as an endemic, affecting more than 8000 individuals. However, the fatal impact and current nature of SARS-CoV-2 pandemic is indicative of an altogether different functional nature of the virus from that of SARS-CoV [5]. Both SARS-CoV and SARS-CoV-2 were originated from bat but the immediate host from which SARS-CoV-2 got transmitted to human remains to be unclear [5]. The high person-to-person transmission rate of SARS-CoV-2 due to an efficient immune evasion and infectivity are of great concerns from the human intervention perspective [6–14]. So, taking into consideration of the potential pre- and post-symptomatic transmissibility of SARS-CoV-2, it is an urgent biomedical need to contain the spreading of this virus either by designing antiviral drugs or by vaccine development.

The host receptor recognition by SARS-CoV-2 and its entry mechanisms are important determinants of viral infectivity, tissue tropism and pathogenesis. Alongside, these are also the key targets to modulate host immune surveillance and intervene the viral entry into host cells. Mature SARS-CoV-2 expresses envelop anchored trans-membrane Spike (S) glycoproteins that mediate the host cell entry. Distinct pre- and post-fusion conformational states of the S protein have very recently been structurally identified by cryo-electron microscopic (EM) studies [15] with the proposition of a 'surprisingly low kinetic barrier' for the conformational transition. Primed by a conformation dependent proteolytic cleavage, the membrane fusion thus not only acts as the necessary mechanism for the host cell entry of the viral genetic material but also leads to two kinetically-related yet distinct conformations of the S protein. The pre-fusion conformation represents the full-length S protein, while the post-fusion form is a cleaved fragment left embedded on the viral membrane after the cleavage [15]. The post-fusion form is presumed to have subsequent functions, not only limited to the membrane fusion alone, for being strategically decorated with N-linked glycans [15]. Being the initial mediator of the essential host – pathogen interaction cascade, the pre-fusion form appears to be the more vulnerable [15] of the two forms. Both the forms are found to be biologically expressed and assembled as trimers, the post-fusion form is an elongated coiled coil and is thus more stable and rigid. The full-length pre-fusion S protein consist of two domains, the S1 receptor binding domain (henceforth referred to as RBD_{Spike}) and the S2 membrane fusion domain, wherein, the three S1 receptor binding heads are situated on the top of the trimeric membrane fusion S2 stalk [16]. The S1 domain is further consisted of two subdomains - the N-terminal subdomain (NTD) and the C-terminal subdomain (CTD) [17]. The pre-fusion conformation has been resolved structurally at 2.8 to 3.3 Å resolution by several recent cryo-EM studies capturing minor variations between its different (closed / stabilized) states (PDB ID: 6VXX [18], 6CRZ [19], 6XR8 [15]) – in all of which the RBD_{Spike} remains structurally unaltered (see **Supplementary Figure S1**). In addition to the RBD_{Spike}, the pre-fusion form also contains a receptor binding motif (RBM_{Spike}) both of which reside in the CTD in S1 unit [17].

The host cell entry of the SARS-CoV-2 involves a cascade of molecular interactions which has been revealed to be triggered by the binding of the RBD_{Spike} to human angiotensin-converting enzyme-2

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 September 2020 doi:10.20944/preprints202009.0006.v1

(ACE2) embedded on the membranes of human lung cells [20–22]. The experimental structure of SARS-CoV-2 RBD conjugated with the human ACE2 receptor has also been resolved by X-ray crystallography at 2.68 Å (PDB ID: 6VW1) [21]. Upon this RBD_{Spike} – ACE2 interaction, the Spike protein requires a proteolytic cleavage at its S1/S2 junction for S2 to gain an irreversible conformational change which leads to a successful host cell entry. A furin[1] cleavage site has exclusively been found in the S1/S2 boundary of SARS-CoV-2 Spike protein recently [16]. Interestingly, RBD_{Spike} has a stronger affinity for ACE2 than that of the whole Spike protein. This implies a more complex mechanism behind the molecular access of SARS-CoV-2 into the host cell [21]. Moreover, the S1 trimer continuously switches between a 'lying down' and a 'standing up' position onto the S2 subunits [15, 16, 23]. When it is at a 'lying down' position, the S1-RBD remains hidden and unexposed enabling the SARS-CoV-2 to escape the host immune surveillance [16]. It is only the 'standing up' position of S1 that enables it to bind with the ACE2 receptor with a higher affinity compared to that of other related SARS-CoV. Taking into consideration these intricate complex features, SARS-CoV-2 stands out to be one of the most challenging pathogens ever to be contained. In addition, there are significant evolutionary differences in the antigenic properties of SARS-CoV and SARS-CoV-2 [18] in spite of sharing 70% sequence similarity in their RBDs and docking to an identical site in the ACE2 receptor. These strategic critical differences potentially lead to the ineffectiveness of a panel of monoclonal antibodies raised against SARS-CoV towards the neutralization of SARS-CoV-2 [24]. So, developing an effective vaccine targeting the S protein of SARS-CoV-2 remains complex and might take more time than can be afforded in this emergency. So, it is of high value to explore alternative means to design effective antivirals / bio-therapeutics that can successfully target the SARS-CoV-2 host cell entry thereby curbing down its infectivity.

Since onset of the current pandemic, enormous efforts are continuously being made for repurposing already approved drugs [25–27], unfortunately with very limited success. Developing strategically designed small molecules and screening them against the viral infectivity is another approach to find a potential inhibitor to block key interactions of SARS-CoV-2 with host cells. Despite some initial promising outcomes, in most of the cases these drugs are unable to stop the spread of COVID-19. Thus, developing strategic molecules to block the guest-host binding remains a clinically unmet goal. To that end, peptide-based approaches to design antiviral bio-therapeutics might be a fruitful alternative strategy [28]. The availability of experimental atomic structures of the SARS-CoV-2 RBD_{Spike} complexed with the ACE2 receptor [19, 29] serves as a great resource for this purpose, helping in the detailed understanding of the binding mechanism, and, thereby, facilitating the design. The binding affinity of SARS-CoV-2 with the 'ACE2 peptidase domain α-helix' is much stronger than SARS-CoV. Designing a peptide disruptor would therefore be an ideal choice over screening of small molecule inhibitors because of its higher efficacy in covering the extended protein contact interface, potentially acting as a compelling competitive inhibitor [28, 30].

In this present study, we aim to design non-reactive structural mimics of SARS-CoV-2 RBD_{Spike} which can serve as potential competitive inhibitors for its binding to the host ACE2 receptor. These polypeptide-based mimics have been designed to bind stably with high affinity to the interacting surface of ACE2 containing multiple contact 'hotspot's. They would thus potentially interfere with the binding of the native SARS-CoV-2 RBD_{Spike} to ACE2 by already occupying the binding sites. To that end, we adapted a protein design approach with iterative cycles of screening followed by molecular dynamics simulations of the finally selected structural mimics to examine the dynamic stability of the prescribed bound complexes. We followed two alternative sampling strategies for the design, based on (i) alteration of hydrophobic character of the mutable amino acids at the RBD –

1 Proprotein convertase of the host

ACE2 interface and (ii) homology-based threading followed by performing strategic 'gain-of-function' mutations. Scoring of the designed complexes were based on shape and electrostatic complementarities {Sc, EC} [31, 32] which are essential prerequisites of binding affinity and stability and may thus be envisaged as coordinate driven representative measures of the same, as reasoned in the paper. The SARS-CoV-2 RBD_{Spike} is an independently foldable protein domain and remains resilient to conformational changes yet after acquiring a series of mutations along evolution. The prescribed designed complexes are therefore expected to fold naturally as self-sustaining protein units.

Interestingly, the RBD_{Spike} – ACE2 interaction in SARS-CoV-2 appears to have a quasi-stable character in spite of having a high affinity for the interaction. The effect is more pronounced if compared to its evolutionary ancestor, SARS-CoV. This can be further envisaged as having a bouncing nature of the ligand upon receptor binding. This enables the molecule to quickly get released from its receptor-site to be able to bind to a greater number of amenable receptors in nearby cells. It therefore appears that the key molecular player of the most determining interaction in COVID-19 has an inherent structural potential to have a high interaction cross-section with its cognate receptor. This has been vividly surveyed and discussed in the light of molecular evolution of the RBD_{Spike} from SARS-CoV to SARS-CoV-2. Taken together, the current study has both a basic and an applied content and provides a novel approach to design polypeptide-based inhibitors against SARS-CoV-2 RBD_{Spike} – ACE2 binding. Subsequent wet lab experiments and testing of the prescribed designed sequences on biological subjects may potentially offer an alternative powerful therapeutic strategy to combat SARS-CoV-2, due to be carried out in the next phase.

2. Materials and Methods

2.1. Details of experimental structures used in the study

For the all-important pre-fusion form of the viral Spike protein, we used the cryo-EM structures in its 'closed state' (PDB ID: 6VXX; solved at 2.8 Å, 22812 protein atoms), that of a 'stabilized variant' (PDB ID: 6CRZ; solved at 3.30 Å, 25024 protein atoms) and, that of the recent most full-length S protein (PDB ID: 6XR8; solved at 2.9 Å, 25995 protein atoms) from the Protein Data Bank [33]. As a single representative structure, 6VXX was preferred among the three for having the best resolution. The three structures had variation in minor details (missing loops, glycans etc.) which was reflected in their all-atom RMS deviation upon pairwise structural superposition (average: 3.75 Å for an average length of 23255 aligned non-Hydrogen protein atoms). The same average RMS deviation for the RBD in the three structures was even lower (2.2 Å) for a stretch of ~190 aligned residues. Visual structural investigation confirmed that this small deviation was due to the conformational variation of the disordered loop regions while the relative orientation of the secondary structural elements (helices and sheets) were virtually identical in all structures (see **Results and Discussion**). When the three structures were further superposed in turn onto the same (ligand) domain in the ACE2 bound complex (PDB ID: 6VW1), the average all atom RMS deviation reduced even further to 1.36 Å.

The other coordinate files used in the core-study correspond to ligand-receptor protein-protein complexes, pertaining to most if not all representative structures of the RBD_{Spike} (Receptor Binding Domain, UNP Residues: 323-502) - Angiotensin-Converting-Enzyme (ACE2) receptor available at the Protein Data Bank [33] till date (31/08/2020). The ones that were of prime importance among these are RBD_{Spike} of SARS-CoV complexed with human ACE2 (PDB ID: 2AJF; solved at 2.9 Å) and RBD_{Spike} of SARS-CoV-2 complexed with human ACE2 (PDB ID: 6VW1; solved at 2.68 Å). Among the rest of the structures used, there were human strains of the ancestral viral RBD_{Spike} (i.e.,

the 2002-2003 SARS-CoV) complexed with human-civet chimeric receptors (two of them, PDB ID: 3D0G; solved at 2.8 Å & PDB ID: 3D0H; solved at 3.1 Å). There were also civet strain of the viral RBD_{Spike} complexed with human ACE2 (PDB ID: 3SCJ; solved at 3.0 Å) and RBD_{Spike} from SARS-CoV epidemic strain complexed with human-civet chimeric receptor ACE2: (PDB ID: 3SCL, solved at 3.0 Å). Importantly the only complex representative of CoV-2 (PDB ID: 6VW1) had a human ACE2 receptor in it.

Additionally, equivalent / similar complexes from MERS (PDB ID: 4L72; MERS-CoV complexed with human DPP4, solved at 2 Å) and Ebola (PDB ID: 5F18; Viral glycoprotein bound to its endosomal receptor Niemann-pick C1, solved at 3 Å) were also assembled as a mean to compare the receptor-ligand binding in terms of affinity and stability from complementarity estimates. Patches of residues missing due to poor electron densities were modeled using MODELLER [34], wherever applicable. These missing patches essentially mapped to an equivalent stretch in all the RBD_{Spike} which was a disordered loop far from the ACE2 binding site. All sequence alignments, pairwise and multiple, were performed by Muscle [35]. Sequence similarities wherever calculated used the EMBOSS stretcher web-tool implementing global alignment as adapted in (https://www.ebi.ac.uk/Tools/psa/emboss_stretcher/).

2.2. Protein design: side-chain threading and shaking the designed complexes

Our approach was to target and build inhibitors of the SARS-CoV-2 RBD_{Spike} to block its binding sites in the ACE2 receptor. Hence, all mutations were performed on the native ligand molecule alone while retaining the native sequence of the receptor. The strategic attributes of the protein design protocol (sampling and scoring) can be found contextually detailed in the section **3.7** and its subsections in **Results and Discussion**. To fit and thread the mutated side-chains on the native template, Scwrl4.0 [36] was used which samples the side-chain conformations from the Dunbrack's Rotamer library and has its unique fast way of optimally removing steric clash. Subsequent to fitting the mutated side-chains on the native main-chain coordinates, side-chain coordinates of the unaltered amino acid residues were retained from the original native structure (6VW1, chain E). Existing polar hydrogen atoms generated by Scwrl4.0 were subsequently trimmed and all hydrogen atoms were rebuilt afresh by the program REDUCE (v.3.3) [37]. REDUCE geometrically builds hydrogen atoms on the existing heavy atom coordinates by analyzing the local hydrogen bond network, flips -CO and $-NH_2$ groups in amidino groups of Asperagine, Glutamine and takes care of resonating states of histidine as appropriate to the given context. The rebuilt structures were then energy minimized by 500000 steps of steepest dissent and 50000 steps of conjugate gradient method in Gromacs and were subsequently undertaken short (10 ns) all atom molecular dynamic simulations (see section **2.9**) as a mean to consider vibrational perturbation (or shake) due to the performed multi-mutations on the native RBD. The short simulations ensure necessary structural relaxation of the designed co-complexes (upon multi-mutations) by allowing sufficient main- and side-chain flexibility. The designed structures are hence released from being trapped in local energy minima. The post-run time-evolved snapshots (after 10 ns) were taken as the final designed structures.

2.2.1. Scrambled Sequences as Negative Control

To serve as negative controls, a pool of scrambled sequences was constructed having an identical composition to that of the presumably potential solutions obtained from the protein design results (i.e., its different variant protocols adapted). For these sequences (hits), amino acid compositions were computed and grouped into six classes: C1. hydrophobic & branched-chain (Ala, Val, Leu, Ile, Met), C2. hydrophobic & aromatic (Phe, Tyr, Trp), C3. polar (Ser, Thr, Asn, Gln), C4. positively

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 September 2020 doi:10.20944/preprints202009.0006.v1

charged (Lys, Arg, His), C5. negatively charged (Asp, Glu), C6. helix breaker and disulfide forming (Gly, Pro, Cys). Compositions (in terms of percentage of each class) were averaged over the 'hits' which served as a compositional consensus. Randomly reshuffled sequences were then generated (hundreds of them) with identical compositions implementing the Fisher-Yates Shuffle algorithm (http://www.programming-algorithms.net/article/43676/Fisher-Yates-shuffle). These 'scrambled' sequences together served as potential negative controls to the computational prediction (see **Results and Discussion**), as a mean to physically verify and cross-validate the importance of crucial and/or conserved amino acid positions in the native sequence (as in 6VW1, chain E) over and above merely meeting the compositional criteria.

2.3. Contact map at the Interface

Amino acid residues buried upon association were identified by a net (non-zero) change in their atomic solvent Accessible Surface Areas (ΔASA≠0) between their bound and free forms. The interfacial atomic contacts were identified when any two heavy atoms coming from two amino acid residues residing at each molecular interfacial surface were found within 4 Å of each-other. A slight relaxation (4.5 Å) of this very stringent cutoff were also attempted. This collection of residue-wise atomic contacts served as the contact map at the receptor-ligand interface – which were vividly and explicitly used as one of the indicators to choose the mutations for the protein design experiment. The same standard cutoff was also used to identify salt-bridges [38, 39] at the receptor - ligand interface.

2.5. Shape and Electrostatic Complementarity

The semi-empirical function of shape correlation statistic (Sc) as formulated by Lawrence and Colman [31] was adopted as a mean to evaluate the Shape Complementarity of the bound protein complexes at their interface. The program Sc (version 2.0, © Michael Lawrence) attributed to the original paper was used to serve the purpose. implicit to this program, first, the molecular (Connoly) surfaces [40] were constructed, sampled at 15 dots / Å^2 for both interacting molecular partners separately. The nearest neighboring dot surface points were identified within a maximum distance of 3.5 Å and the following measure (S_{dp}) computed for each pair of nearest neighboring dot points.

$$S_{dp} = n_A.n_B exp(-wd^2); Sc = median\{S_{dp}\} \quad \text{- (1)}$$

where, n_A and n_B refers to the unit normal vectors, one outwardly and the other inwardly oriented, corresponding to the two dot points A and B coming from the two interfacial molecular surfaces; d is their distance and w is a scaling constant set to 0.5. Median of this distribution is taken as Sc.

Electrostatic Complementarity (EC) at the protein-protein interfaces was adopted as originally prescribed by McCoy et al., [32] wherein, the surface electrostatic potential was computed for each interfacial protein surface twice, one time each for the contribution of each partner molecule (taken as 'target' and 'neighbor'). The surface electrostatic potentials were computed by numerically solving the Poisson-Boltzmann equation (using Delphi v8.4 [41]) implementing its finite difference method, wherein, the protein dielectric was modeled as a smooth Gaussian function of distance from its center of mass [42]). This returns two troughs of potential values for each intrefacial surface and the negative of the Pearson's correlation is defined as the EC at each interfacial surface (see eq.2). The average of the two EC's obtained for the two interfacial surfaces (EC_1, EC_2) is taken as EC at the interface.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 September 2020 doi:10.20944/preprints202009.0006.v1

$$EC_{1,2} = -\left(\frac{\sum_{i=1}^{N}(\Phi(i)-\bar{\Phi}).(\Phi\prime(i)-\bar{\Phi}\prime)}{\sum_{i=1}^{N}(\Phi(i)-\bar{\Phi})^2.\sum_{i=1}^{N}(\Phi\prime(i)-\bar{\Phi}\prime)^2}\right); EC = (EC_1 + EC_2)/2 \quad \text{- (2)}$$

In the above equation (eq. 2), if an interacting molecular surface consisting of N dot surface points is taken as the 'target' molecule, $\Phi(i)$ represents the electrostatic potential on its i^{th} point realized due to its own atoms and $\Phi'(i)$, due to the charged atoms of its molecular partner, taken as 'neighbor'. $\bar{\Phi}$ and $\bar{\Phi}'$ are the mean potentials of $\Phi(i)$ and $\Phi'(i)$, $i=1..N$ respectively. $EC_{1,2}$ may interchangeably represent both EC_1 and EC_2 with the necessary swapping of 'target' and 'neighbor' and the corresponding potential terms ($\Phi \leftrightarrow \Phi'$).

Both Sc and EC are essentially correlation measures ranging from -1 (perfect anti-complementarity) to 1 (perfect complementarity) having the same sense of directions (higher the better). The non-rigid optimal ranges for Sc and EC can be reasonably approximated as {0.55 to 0.75} and {0.45 to 0.65} respectively as has been found in protein complexes coming from a wide range of biological origins [31, 32].

2.6. Complementarity Plot (CP_{int} and CP_{dock})

Complementarity Plot [43–47] refers to a two dimensional plot of the ordered-pair values of shape and electrostatic complementarities along its X and Y axes. The plot estimates the probabilistic correctness of an experimentally solved or a computationally built atomic model of a globular protein or a protein-protein complex, based on harmony of the embedded side-chains at their respective protein environments with their local and non-local neighborhood. The harmony with respect to the local and non-local neighborhood is estimated in terms of shape and electrostatic complementarity of buried and partially buried amino acid residues. The complementarity plot has three variants. The first two of them, namely, CP and CP_{int} are residue-wise plots plotting the ordered-pair complementarity values computed for buried or partially buried amino acid residues at the protein interior and interface respectively. The third variant, CP_{dock} was originally proposed as a protein-protein docking scoring function [47] and is based on the aforementioned single {Sc, EC} values obtained for the whole protein-protein interface (**Figure 1**). In all the three variant plots, the the resultant points may be found located to either of the three regions in the plots: 'probable', 'less probable' or 'improbable' based on their probabilistic feasibility to fit into a folded protein or a protein co-complex model. As can be seen (**Figure 1**), the 'probable' and 'less probable' regions in CP_{dock} (and those in the other variant CP's) are primarily covered (>85% area) by the first (+,+) quadrant of the plot with Sc, EC both attaining positive values. Such 'both positive' points would thus render a higher probabilistic feasibility of two proteins to interact and this probability would increase with the closeness of the point from the 'probable' and/or 'less probable' regions. Depending on the requirement, both CP_{int} and CP_{dock} were used in the study. CP_{dock} was used for screening and scoring the complexes, while CP_{int} was utilized for shortlisting and identifying the amino acid residues at the interface to be attempted for mutations in the protein-design pipeline.

2.7. Accessibility Score

The accessibility score (***rGb***) compares the hydrophobic burial profile (i.e., the distribution of amino acids as a function of solvent exposure) of a globular protein or a protein-protein complex with respect to corresponding native distributions, enumerated from standard databases. The score is also applicable to peptide fragments or protein domains. The accessibility score is an integral part of the structure validation protocol prescribed in the Complementarity Plot [44, 45]. Mathematically, the score is based on normalized conditional probability (or propensity) estimates of residue types given their burial (and hence the name: ***rGb***) and can be formulated as follows.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 September 2020 doi:10.20944/preprints202009.0006.v1

$$rGb = \frac{1}{N_{res}} \sum_{i=1}^{N_{res}} log_{10}\,(Pr_i) \qquad \text{- (3)}$$

where N_{res} is the sequence length of the input polypeptide chain and Pr_i is the propensity of a particular amino acid (Val, Asn, His etc.) to acquire a particular degree of solvent exposure.

A value of ***rGb*** > 0.011 [44] (and higher the better) renders the input atomic model affirmative with regard to the 'native-like' distribution of amino acids as a function of solvent exposure while a value less than that means hydrophobic residues are exposed to the solvent causing the molecule stay in an unfavorable/frustrated disordered (high entropy) state! A negative value emphasizes this instability which may be extended to explain reaction-prone nature of the said fragment.

2.8. Fold recognition

Complementarity-based fold recognition measures (CS_{gl}, CS_{cp}) [43] were implemented to test the compatibility of the designed sequences to the fold (i.e., main-chain trajectory) of the RBD_{Spike}. A (μ - 3σ)2 baseline on the complementarity scores (CS_{gl}:2.4, CS_{cp}:0.01) was set as a threshold value to determine the compatibility of a designed sequence to the given fold (μ, σ taken from the original reference).

2.9. Molecular Dynamic Simulation (short and long)

Molecular dynamic simulations were used in the study to serve a two-fold purpose. As described previously (see section **2.2**), the short simulations were run implicit to the design protocol as a mean to incorporate vibrational perturbation to the *in-silico* designed complex. In contrast, long simulations were conducted to study the dynamics and stability of the binding of the finally chosen complexes. In either case, the same protocol was followed but for changing the duration of the production phase. Explicit-water Molecular Dynamics (MD) simulation were performed in GROMACS v.2018.1 [48, 49] using the AMBER99SB-ILDN protein force-field [50], TIP3P water model and 'solvent' as the ion replacing system associated with the MD package. Periodic boundary conditions were used, solvation and charge neutralization of the proteins were subsequently followed by two rounds of energy minimization (500000 steps of steepest descent followed by 50000 steps of conjugate gradient) using the in-built PROMD module [49] within GROMOS96. The energy minimized protein – solvent system was then equilibrated in an NVT ensemble followed by an NPT ensemble for 100 ps and 5 ns respectively. The initial temperature set for the NVT ensemble was 100 K which was gradually raised to 300 K at constant volume and was kept the same for the entire NPT equilibration while the pressure being maintained at 1 bar. The simulation systems were large, consisting of a total number of 246148 (± 10) atoms at an average. The production runs were done in an NPT mode for 200 ns for the long MD simulation runs (10 ns for the short ones) with a time-step of 2 fs for each equilibrated protein–solvent system. The 'cubic' simulation boxes were built by considering an initial length of at least 6 Å from the surface of the protein co-complex (placed at the center of the box) to each cubic face. This lead to an average box-dimension of ~ 135 Å × 135 Å × 135 Å. To maintain constant temperature, Berendsen's temperature bath was used with a coupling constant of 2 ps, while barostat with a coupling constant of 1 ps was used to regulate the constant pressure. The 'lincs' algorithm was used to restrain bond-lengths for all bonds. For the short simulation runs, the final snapshots (at 10ns) were stored and used as the final designed structures for scoring. For the long simulation runs, trajectories were written at an interval of 2 ps, resulting in 100,000 frames (or time-stamps). Binding stability and other related

2 μ: mean; σ: standard deviation

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 September 2020 doi:10.20944/preprints202009.0006.v1

dynamical analyses were all performed on the post-equilibrium 200 ns long trajectories (for the finally selected designed proteins). For each simulated protein-complex trajectory, all post-simulation analyses were done 4000 snapshots collected at 50 ps interval. This sampling may be considered of sufficient resolution to capture the molecular events under investigation.

2.10. Measuring the dynamic stability of the proposed 'optimal' solution

To quantitatively assess the dynamic stability of the proposed 'optimal' solutions, CP_{dock} was run on their whole dynamic trajectory (i.e., on the selected collection of snapshots representing the trajectory). Alongside analyzing the dynamic persistence of the attained Sc, EC values individually, the ordered pair treatment of {Sc, EC} was also invoked by estimating the distance of the corresponding points in the plot from the 'probable' region. To that end, a 2D Euclidean distance measure (E2d) was formulated based on a binary logic. If the {Sc, EC} point in the plot was found to be located on a 'probable' grid, then E2d was set to zero. Otherwise E2d was computed as the 2D Euclidean grid distance from the mid-point of the nearest edge belonging to the 'probable' grid nearest to that point in the plot. It can be formally proven that E2d is a metric in an R2 vector-space (proof not given).

3. Results and Discussion

3.1. Molecular evolution of the SARS-CoV-2 RBD_{Spike}: reviewing key residues

SARS-CoV-2 has a high rate of transmission in human [51–54] (though the fatality rate is low) while sprading only nominally within other close species (civet, rodents, ferrets, other primates etc.). Evolutionary genomic studies have revealed that the RBD_{Spike} is the most variable part of the corona virus genome [20, 55]. Furthermore, recent literature on the proximal origin of SARS-CoV-2 [1] has highlighted the essential effective difference between RBD_{Spike} of CoV and CoV-2 to be localized within a 51 amino acid stretch (residues: 442-491 in CoV; 455-505 in CoV-2) at the ACE2 binding site. Let this stretch be henceforth referred to as the 'Spike RBD hotspot'. A visual structural examination revealed that the stretch primarily mapped to a long partially folded disordered loop with a small anti-parallel β-strand embedded in it (**Supplementary Figure S2**). The hotspot region includes six 'critical' amino acid positions that physically bind to the receptor out of which five are mutated in CoV-2 with respect to CoV (Y442→L455, L472→F486, N479→Q493, D480→S494, T487→N501) [1]. The overall composition or physico-chemical consensus (in terms of hydrophobicity, charge, polarity, aromaticity, amino acid volume etc.) upon these evolutionary changes remains almost unaltered in the two viral species. The only noticeable effective difference is in the mutation of one negatively charged amino acid to a polar residue (D480: CoV→S494: CoV-2). In a sense, the mutations collectively appear to be a reshuffling of the overall discrete sequence space (consisting of the aforementioned crucial positions). So, based on the above hypothesis [1], it is quite surprising that how this small localized change could alone lead to such an incredibly high increase in transmission rate in CoV-2 with respect to that in CoV! To portray a more comprehensive picture of the evolutionary event, the observation window was broaden to the aligned full-length sequences of the two homologous protein domains (RBD_{Spike}). As a matter of fact, the total number of point mutations between RBD_{Spike} of CoV and CoV-2 are found to be 17, twelve out of which have an alternating hydrophobic character (i.e., polar / charged ↔ hydrophobic). Interestingly, all these mutations are situated within the 'Spike RBD hotspot' defined above.

3.2. Affinity and stability of binding from local and non-local measures of Complementarity

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 September 2020 doi:10.20944/preprints202009.0006.v1

The coupling between the dual attributes of complementarity is well known in biomolecular recognition, concerning shape and electrostatic matching of the interacting molecular surfaces [43, 56–59]. It was also realized subsequently that shape complementarity (Sc) is a necessary criterion for macro-molecular binding while electrostatic complementarity (EC) is sufficient [56, 60, 61]. For oligomer formation in proteins, where large surface area (~1600 $Å^2$) [62] are required to get buried upon complexation, surfaces have to be carefully tailored for the complementary interlocking of side-chains at the interface. This close association between the interacting molecular partners enhances the effective match between their protrusions and crevices so that extended areas can move into close contact [31, 43, 63, 64]. A poor complementarity in shape between two macro-molecular surfaces, therefore, stands out to be a strong forbidding factor for their close association. For example, two purely convex surfaces (say, two spheroids or ellipsoids) lack the steric fit to bind.

On the other hand, complementarity in surface electrostatic potential serves as a secondary criterion in macro-molecular interactions, especially for proteins. The inter-relation of electrostatic forces and protein stability is well known [57]. For example, optimizing Coulomb interactions through charge substitution on the protein surface leads to increased stability [65–68]. However, the same may not be achieved by a mere non-strategic increase in the net charge (positive or negative) as electrostatic repulsion may interfere within the folded state [65, 69, 70]. Along the same line, complementarity in surface charge and/or net charge were ruled out as the representative complementarity term in protein co-complexes [32] and was corrected by redefining EC as the correlation in surface electrostatic potentials. Sub-optimal EC values (even negative values) have been found to result occasionally from unfavorable or repulsive interactions in protein co-complexes, also in protein-ligand interactions [71], often compensated by strong counterbalancing geometric fit [58]. Such instances have been found in statistically considerable proportion (in ~20% of the cases) in native protein-protein complexes [60], wherein, compensatory elevated Sc values have frequently been recorded [60]. Such obligate interactions[3] are generally found to be transient in nature, often linked with signaling pathways [72–75].

The long- and short-range nature of the forces giving rise to EC and Sc respectively leads to their corresponding stringent and relaxed criteria. Accordingly, the height and width of the 'probable' regions vary in the complementarity plots (**Figure 1**). From this conceptual platform, it is quite logical to envisage shape complementarity (Sc) as an attractant factor in macro-molecular binding representing the mutual affinity of the two molecular partners to bind. On the other hand, since, satisfaction of electrostatic matching at the interaction-surface to an adequate extent works favorably to stabilize the bound complex, EC may plausibly be treated as the analogous structural parameter representing binding stability.

In order to have some numeric check of the proposed hypothesis, a pilot calculation on an available dataset was also attempted. The dataset comprises of binding thermodynamic parameters (K_d, $\Delta G_{binding}$) for protein co-complexes having experimental coordinates (https://bmm.crick.ac.uk/~bmmadmin/Affinity/) [76]. Hence, it was possible to compute the complementarity estimates {Sc, EC} from the corresponding atomic coordinates. The database contains 178 protein-protein complexes from various biological origins – all of which were assembled and used in the calculation. In case of more than one annotated chains for either the receptor and/or the ligand, they were merged to make two chains in total (e.g., for 4FQI, chains {H, L}→H : {A, B, E, F, C}→A). Binding affinity and energy values were undertaken min-max scaling ($Xs=\{X-X_{min}\}/\{X_{max}-X_{min}\}$) as a mean to normalize their arbitrarily wide range of raw values into

3 interactions required instantaneous / short-termed, e.g., those involved in signal transduction

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 September 2020 doi:10.20944/preprints202009.0006.v1

[0, 1]. EC having the wider spread could best be fitted linearly to $\Delta G_{binding}$ (Pearson's Correlation: 0.63, P-value<0.00001) while Sc having a far more stringent interaction-threshold (~0.45) had a hyperbolic fit as an empirical function of K_d, reminiscent of a saturation profile (**Supplementary Figure S3**). Although this semi-empirical correlation may not be interpreted as all-conclusive or comprehensive and may prove to be more complex from numbers obtained from a larger dataset (when available), it certainly appears both logically and numerically that affinity and stability of interacting proteins may reasonably be well approximated by the high level structural descriptors, Sc and EC.

3.3. Evolution of the CoV-2 RBD_{Spike} – ACE2 interaction dynamics

Based on the conceptual foundations discussed in section **3.1**, the relative Sc, EC values computed for SARS-CoV (2AJF) and SARS-CoV-2 (6VW1) were insightful. 2AJF has an Sc of 0.42 with an EC of 0.21. Together, these values rationalize the binding, since, both numbers are appreciably positive, falling in the 'both-positive' (+,+) first quadrant of CP_{dock} (section **2.8**). However, the ordered pair {Sc, EC} values also indicate that the binding is sub-optimal with respect to their corresponding reference ranges – which is clearly reflected from the location of the corresponding point in CP_{dock} (**Figure 2**). In more elaborate terms, the point falls outside the optimal or near-optimal zones, i.e., outside the 'probable' and 'less probable' regions in the plot (see section **2.6** in **Materials and Methods**). In contrast, in 6VW1, Sc is found to be 0.56 (14% increase w.r.t. CoV) while EC is as low as 0.13 (~8% drop w.r.t. CoV). Again, since, both values are positive, the resultant {Sc, EC} point in CP_{dock} hits the first (+,+) quadrant of the plot (**Figure 2**), thereby, rationalizing the binding (please refer to section **2.6**). Visual investigation of the two {Sc, EC} points from 2AJF (CoV), 6VW1 (CoV-2) side-by-side on CP_{dock} further revealed their comparative interaction dynamics which is evolutionarily insightful. Biochemical solution studies elsewhere [21] had already confirmed that the RBD_{Spike} has a significantly greater affinity towards ACE2 relative to that in CoV. The same is also reflected in their corresponding Sc values. The 14% increase in Sc in CoV-2 relative to that in CoV actually makes the Sc value hit its non-rigid optimal range (refer to section **2.5**). As a result of this appreciably increased shape matching, the RBD_{Spike} in CoV-2 would have a much higher affinity for ACE2 than that of CoV and would therefore be attracted much faster to its cognate receptor. However, at the same time, it renders a sub-optimal EC value (0.13) upon interacting with ACE2. In elaborate terms, the receptor and the ligand contact-surfaces share just 13% match between their surface electrostatic potentials coming from the electric fields of their own and that of their partner's (see **Figure 3**). By definition (see section **2.5**), this means weak anti-correlation in surface potentials at the interface, as the close association of two perfectly anti-correlated electrostatic surfaces would ideally return a value of EC=1 [32]. Hence, yet being attracted to ACE2 faster than that in CoV, the RBD_{Spike} in CoV-2 would also get released from the receptor faster as the unfavorable electrostatic interactions would act against a stable binding. The lower stability in the co-complex in CoV-2 relative to that in CoV can also be cross-validated by comparing the 'dG_separated' values for both, computed by structure driven thermodynamic calculations using Rosetta [23]. Interestingly, in spite of the sub-optimal EC, the increase in Sc in CoV-2 relative to CoV results in a right-shift along the horizontal axis of the corresponding resultant point (CoV-2) in CP_{dock} making the point map to the near optimal zone (~ 'less probable' region). Overall, the RBD_{Spike} – ACE2 interaction in CoV-2 does appear to have a quasi-stable character in spite of having a high affinity. At the same time, it is also interesting to reveal that a disease with such a high rate of transmission is actually triggered by a quasi-stable interaction – which may potentially instigate parallel research endeavors to further explore the phenomenon at more complex molecular hierarchies.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 September 2020 doi:10.20944/preprints202009.0006.v1

In order to carry out a comparison among the available homologs, Sc, EC were computed for all six RBD_{Spike} - ACE2 complexes (see section **2.1** in **Materials and Methods**) and were plotted together in CP_{dock}. Both, Sc and EC ranged across sub-optimal to near-optimal values making the corresponding points scattered over 'improbable', 'less probable' to near-'probable' regions of CP_{dock}. Noticeably, the civet strain of 3SCJ has the closest approach (**Figure 2**) towards optimality in terms of Sc, EC values, or, in other words, in terms of the closeness from the 'probable' region, when plotted in CP_{dock}. Interestingly, the {Sc, EC} points corresponding to all the homologs was found to cluster around the left-bottom (south west) of the 'probable' (optimal) region in CP_{dock} (**Figure 2 B**). Such a distribution of points in CP_{dock} is indicative of sub-optimal quasi-stable binding of the two molecular partners along evolution. This was also prominent from a structural display of the molecular interface (**Figure 2 A**). For instance, there was no deep grooves nor any binding pockets on the receptor where the ligand may stably fit with high affinity. Neither there were signs of any conformation-induced knotting upon binding nor other known / intuitive structural models that might map to 'high affinity stable binding'. Rather, the binding appears to be reminiscent of a 'molecular handshake' [77] rather than a molecular hug or cling, both from CP_{dock} and from the corresponding structural displays. It is also noteworthy that the part of the 'ACE2 peptidase domain' involved physically in the binding to RBD_{Spike} is actually a single α-helix. The same relative trends among the homologous structures (**Table 1**) are also naturally reflected from CP-based global (Complementarity score, CS_l) and local measures [44, 45].

3.4. Comparison with equivalent complexes from MERS, Ebola

As a point of reference, equivalent complexes from other deadly viral diseases in human were surveyed in a likewise manner. MERS (PDB ID: 4L72) CoV RBD, when bound to its human-receptor Dipeptyl transferase (DPP4) had substantially better shape fit and electrostatic matching along extended mutually compatible surfaces (**Figure 4**, top panel). On the other hand, the Ebola Viral Glycoprotein bound to its endosomal receptor Niemann-Pick C1, displayed signatures of knotting upon binding induced conformational changes naturally having far greater surface fit coupled with optimal electrostatic matching (**Figure 4**, bottom panel).

3.5. Reaction-prone nature of the CoV-2 RBD_{Spike} binding site

As elaborated in the above sections, when compared with analogous ligand-receptor complexes from related viral strains in the human host, the RBD_{Spike} – ACE2 interface in SARS-CoV-2 does appear to be different and rare. All analyses unequivocally indicate that the interface maps to protein co-complexes involving transient interactions [74] which is likely to be causally linked to its presumably unique *modus operandii*. To cross-validate this observation, other independent approaches were also adopted concerning the study of the interface. This included (i) calculation of the accessibility score (*rGb*) of the complex and different relevant molecular fragments, and, (ii) a detailed analysis of the contact map at the interface. As a matter of interest, 6VW1 (i.e., the only representative interface structure from CoV-2) alone was chosen for the analyses.

As detailed in **Materials and Methods** (section **2.7**), a value of *rGb* greater than 0.011 (and higher the better) qualifies a globular protein / protein complex / peptide fragment / protein domain to be considered native-like in terms of hydrophobic burial or the distribution of amino acid residues with respect to solvent exposure. Any value less than this empirical threshold renders the input protein molecule non-native like which physically means that hydrophobic residues are exposed to the solvent. This would cause the molecule to stay in an unfavourable / frustrated disordered (high entropy) state! A negative value virtually guarantees this instability which may be extended to depict a reaction-prone nature of the said protein fragment.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 September 2020 doi:10.20944/preprints202009.0006.v1

With this understanding, *rGb* was computed for the (i) whole native protein complex (referred to as 6VW1_AE in **Table 2**) and different relevant molecular fragments, namely, (ii) the ligand (E) chain E or the RBD_{Spike} alone (6VW1_E in **Table 2**), (iii) the 'Spike RBD hotspot' (residues: 455-505, see section **3.1**) where all key mutations are localized (6VW1_E_hotspot in **Table 2**) and (iv) the actual binding site or the collection of mapped interfacial residues on chain E as found in the contact map (6VW1_E_bs in **Table 2**). The trend of *rGb* was found to be descending in large fractions from (i) the whole complex to (iv) the binding site (6VW1_E_bs). The relative numbers clearly indicate that the complex has the most optimum (or native-like) distribution of hydrophobic burial (***rGb***: 0.052, **Table 2**) in the whole set which is substantially better than the ligand chain alone (***rGb***: 0.028). The high negative value (***rGb***: -0.055) obtained for 6VW1_E_bs speaks for its high reaction-proneness [78]. In other words, the high degree of unfavorable hydrophobic exposure makes the binding site (6VW1_E_bs) critically scurried or strained in its free state. Thus, it is always in a crisis need to embed itself within a befitting complementary surface of an appropriate binding partner.

For another level of cross-checking, the contact map at the interface (see **Materials and Methods**, section **2.3**) was also rigorously scrutinized. The interface was large with an accessible surface area buried upon complexation (ΔASA) of 1644.4 $Å^2$ considering both molecular partners. It involved 23 inter-residue contacts between the residues coming from the two molecular partners totaling 96 pairwise atomic contacts between their side-chain atoms. The interface appears to have many rare interesting features. From the ***rGb*** calculations stated above, it was already clear that the RBD_{Spike} interfacial surface had several exposed hydrophobic residues, hence, it is perhaps of no surprise that the contact map consisted of several hydrophobic residues coming from the ligand (RBD_{Spike}). Interestingly enough, most of these hydrophobic residues were found to be in contact with hydrophilic residues coming from the receptor. Furthermore, a large majority of these hydrophobic residues were in fact bulky aromatic amino acids (**Suppleentary Table 1**). They were mostly found to be in contact with either 'elongated positively charged' (Lys) or 'aromatic yet polar amino acids' (His) coming from the receptor. The corresponding interactions mapped to close hydrophobic packing between extended chains of successive mythelene groups (-(-CH_2)$_4$) of the Lysine(s) and the aromatic ring (31-Lys-A – 489-Tyr-E, 353-Lys-A – 505-Tyr-E) (**Figure 5 A, B**). There were also instances of polar interactions involving aromatic components (34-His-A – 453-Tyr-E) (**Figure 5 D**), although, there were no clear signatures of any cation–Π or Π-Π stacking between the charged residues and the aromatic rings. However, there were instances of regular aromatic stacking with a slide and an open angle separating the otherwise-parallel aromatic rings (83-Tyr-A – 486-Phe-E) (**Figure 4 E**). Also, there were hydrophobic packing (79-Leu-A – 486-Phe-E, 34-His-A – 455-Leu-E) and electrostatic interactions involving polar atoms (24-Gln-A – 487-Asn-E, 42-Gln-A – 498-Gln-E, 34-His-A – 493-Gln-E) (**Figure 5 F**). Interestingly, there was a salt-bridge (31-Lys-A – 484-Glu-E) as well at the interface (**Figure 5 C**) whose presence may be further destabilizing due to desolvation effects – as has been found for salt-bridges in general at protein-interfaces [32, 38, 79]. Overall, it genuinely appears that the interface high potential to harbor and withstand unfavorable electrostatic interactions – which may be causal to the resultant sub-optimal electrostatic complementarity (EC = 0.13).

3.6. Inherent evolutionary features of RBD_{Spike} naturally aiding the design of its structural mimics:

The primary objective of the current study was to develop non-reactive structural mimics of the RBD_{Spike} that could bind to the ACE2 receptor stably with high affinity. These mimics would thus serve as potential competitive inhibitors of the viral RBD_{Spike} by occupying the binding sites in the ACE2 receptors. To that end, a protein design approach was adopted aiming to raise the EC to the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 September 2020 doi:10.20944/preprints202009.0006.v1

extent possible from its sub-optimal native reference (EC_{6VW1}=0.13) of the designed complexes while retaining or raising Sc at or from its already near-optimal range (Sc_{6VW1}=0.56). The conceptual foundations of the 'plausibility of the design strategy' relied on a two-fold fact. Firstly, the RBD_{Spike} is an independently foldable domain which is self-sustained as a protein unit and can undergo folding independent to that of the rest of the Spike protein [21]. Secondly, the RBD_{Spike} is resilient to conformational changes upon multi mutations, as has been evident from structural analyses (see section **3.3**) of the homologs. This means that the basic fold in RBD_{Spike} remains unaltered in spite of the evolutionary sequence variations. The pairwise sequence similarity of the CoV RBD_{Spike} sequences with respect to 6VW1 (CoV-2) was found to be ~69%. RMS deviations (C^{α}) upon superposing the CoV RBD_{Spike} – ACE2 structures (see **Materials and Methods**, section **2.1**) onto 6VW1 were found varying from 1.29 Å (for 3SCL) to 3.18 Å (for 3D0G) (**Supplementary Figure S4**). Furthermore, there were virtually no confomational changes of the RBD_{Spike} upon binding to the ACE2 receptor with respect to its structure in free form (6VXX). RMS deviation (C^{α}) upon superposing the RBD_{Spike} from 6VW1 onto the free and full structure of the Spike protein (6VXX) was 0.893 Å. Together this means that one may simply administer the finally selected designed mimics without having to bother about their folding (*ab-initio*) as long as their sequences fit the fold. Test of this fitness with the given fold (i.e., fold compatibility) of the designed sequences was made by *state-of-the-art* scoring functions for fold recognition (see **Materials and Methods**, section **2.8**).

3.7. The Protein Design Strategy: Sampling and scoring

As mentioned in several earlier sections, a protein-design approach was adopted aiming to develop non-reactive structural mimics of the RBD_{Spike} which may serve as potential competitive inhibitors of the native viral Spike protein to act against the viral pathogenicity. As was found out, the interacting surfaces of CoV-2 RBD_{Spike} and ACE2 has a high shape fit (Sc: 0.56) mapping to its optimal range (section **2.5**) coupled with a sub-optimal electrostatic matching at the interface (EC: 0.13). Together, these may be interpreted in terms of having a high affinity yet with a low stability upon binding. Aligned observations have also been proclaimed by biochemical solution assays [21] and calculation of structure based thermodyanmic parameters [23] carried out in other studies. This quasi-stable nature of the binding potentially triggers a fast-release of the ligand from the receptor, making them amenable to interact with a greater number of cells having surface-exposed ACE2 receptors. So, the primary objective in the designed RBD mimics was to increase the EC at the interface which would make the interaction more stable. Combining the shape affinity factor, the design problem aimed to improve EC while retaining Sc at least native-like in that 'near-optimal to optimal' range. Experimental structural studies in an aligned direction have already demonstrated the favorable effect of key residue substitutions performed across the whole C-terminal domain of the CoV-2 Spike protein harboring the RBD (see **Supplementary Figure 1**). Such key-substitutions have been found to strengthen the Spike – ACE2 interaction leading to a 4-fold increased affinity for receptor binding than that of the native co-complex [24]. For our purpose, we had chosen to operate on the RBD_{Spike} itself. When the native complexes from the homologs (see **Materials and Methods**, section **2.1**) were superposed onto 6VW1, the average pairwise C^{α}-RMS deviation was found to be 2.05 Å. This evolutionary structural conservation meant that mutations at the ligand (RBD_{Spike}) interface can directly be performed on the native bound complex (6VW1) itself. In a sense, the bound complexes were treated like unified globular proteins, wherein, the design protocol may be considered analogous to performing a 'hydrophobic core design' or a 'full sequence design' in globular proteins. Any protein design protocol has two essential steps: (i) sampling and (ii) scoring. For the current study, sampling (i.e., incorporating strategic mutations) was attempted by essentially two approaches, consistent with the main objective of raising the EC while retaining an at least native-like Sc. In the first of the two approaches attempts were made to alter the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 September 2020 doi:10.20944/preprints202009.0006.v1

hydrophobic character of the amino acid residues at the interface while keeping their shape and size as similar as possible. Intuitively, this could alter and possibly raise the EC while keeping Sc similar. An equivalent strategy, earlier, was found fruitful in incorporating unbalanced partial charges into native globular protein interiors and detecting the local 'electrostatic' errors in-turn [44]. In the second approach, homologous sequences (i.e., direct examples from nature) that were already found to hit appreciably higher EC values were threaded on the native RBD_{Spike} template in 6VW1. Strategic mutations were performed on this threaded homologous sequence based on the contact map at the interface. All mutations in the aforementioned two approaches were performed on the ligand molecule alone retaining the receptor as it is. Scoring and raking of the complexes were primarily based on the complementarity measures (see **Materials and Methods**, section **2.5, 2.6**). Fitness or compatibility of each designed sequence with respect to the native fold was tested by fold recognition measures also based on complementarity (section **2.8**).

3.7.1. Design strategy-1: Altering the hydrophobic character of the amino acids

First, from the distribution of interfacial amino acid residues of the ligand chain (6VW1_E) in the residue-wise Complementarity Plots (CP_{int}), residues falling in the 'less-probable' and 'improbable' regions (please refer to **Figure 1**) were accumulated. They were then united with critical residue positions at the binding site (the 'Spike RBD hotspot', residues: 455-505 : see section **3.1**) said to be harboring determining evolutionary mutations [1]. The full set (S1) consisted of 11 amino acids in total (**Supplementary Table S2**) and but for the case of 417-Val the rest of the residues were covered within the aforementioned 'hotspot' region. Out of the eleven amino acids selected, four were bulky aromatics, three branched chain hydrophobic, and the rest polar. As a first trial (**strategy-1a**), mutations were made in this set of 11 residues alone. The raw combinatorial space considering all possible amino-acid mutations is of the astronomical order. To curtail it down to the limits of finite sampling, *ad-hoc* filters involving semi-empirical rules of thumb (detailed as follows) were judiciously incorporated. Each designed sequence was unique as the sampling involved random seeds. Coupled with the random seeds a weighting scheme was further adopted. For 50% of cases, the amino acids were mutated to (i) residues with alternating hydrophobic character and/or structural properties (S↔S, A↔S, V↔T, L↔N, F↔Y, L↔D, I↔M, M↔R, E↔R, E↔Q, D↔N, R↔M, R↔E, etc. : antonymous changes) and for the other 50%, to (ii) amino acids with similar properties (G↔P, V↔L, F↔W, K↔R, E↔D, Q↔N, H↔Y, S↔T etc. : synonymous changes). Care was taken to retain their size and/or shape as much as possible. This 1:1 ratio of weights was further varied from 2:1 to 1:1. The intent was to raise the residue-wise electrostatic complementarity (E_m) of amino acids falling into the 'improbable' / 'less probable' regions of CP_{int} in such extents that they can make it to the 'probable' regions. It was subsequently realized that electrostatic matching is essentially a global effect and need not necessarily affect the mutated residue itself. Hence, in an alternative approach (**strategy-1b**), the contact map of the interface was surveyed (see **Materials and Methods**, section **2.3**) and the ligand residues involved in this set (S2: 13 of them) were chosen as the target positions (**Supplementary Table S2**) to perform the mutations keeping the same sampling strategy. There was appreciable overlap (~46%) between the two sets, S1 and S2.

For each of the two aforementioned strategies (**1a & 1b**) 50 redesigned sequences were constructed and tested in CP_{dock}. Each individual case was carefully scrutinized with visual intervention at all stages of the design protocol. When plotted in CP_{dock}, they were fairly closely spaced creating a south-west island (**Figure 6 A**) relative to the center of the optimal-zone in CP_{dock} (i.e., the 'probable' region). The points were more closely clustered for the first set (**strategy-1a**) relative to the second (**strategy-1b**) in terms of both complementarity measures: Sc, EC as reflected in their

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 September 2020 doi:10.20944/preprints202009.0006.v1

corresponding range of values (Set-1a: [0.354 – 0.544] in Sc, [0.103 – 0.271] in EC; Set-1b: [0.454 – 0.733] in Sc, [0.018 – 0.318] in EC for strategies 1a & 1b respectively).

In spite of being more closely clustered, set-1a mapped to values further away from the optimal zone relative to set-1b. On the other hand, set-1b appeared to have a greater chance of returning false positive points falling in the 'improbable' regions (sub-optimal zones) of the plot (**Figure 5 B**). The top 25 sequences from each set were then filtered based on their residence in CP_{dock} (relative to the optimal zone). All filtered sequences successfully passed the test for fold-compatibility (averages: 2.76 ±0.17 in CS_{gl}; 0.016 ±0.0001 in CS_{cp}). These sequences were more closely spaced in CP_{dock} relative to the corresponding original sets. Set-1b mapped more into the 'probable' / 'less probable' regions (i.e., optimal / near-optimal zone) relative to Set-1a, though, with a greater number of false-positives (**Figure 6 A, B**). To serve as negative controls, 'scrambled' sequences (see section **2.2.1**) were generated for each set by random reshuffling of the designed sequences and plotted alongside the 'hits' in the two sets (1a & 1b). Clear discriminatory clusters were obtained for the 'hits' and the 'scrambled' sequences (see **Materials and Methods**, section **2.2.1**) with virtually no overlap (**Figure 6**). All points in the corresponding 'random' clusters (**Figure 6,** red dots) representing the scrambled sequences were found at the 'improbable' regions of the plot, indicating that they were unambiguously sub-optimal.

3.7.2. Design strategy-2: Homology-based protein design: taking templates from Nature itself

In several well-posed hard-to-solve bioinformatics problems direct adoption of empirical natural strategies [80–83] coupled with trial-and-error modulations has found much scope and penetration. This includes the very problem of protein structure prediction (considered to be the 'holly grail of structural biology') or other related sub-problems emerging from the core of the protein folding problem (e.g., fold recognition [43], protein design [84] etc.). The 'fragment assembly simulated annealing' strategy [82, 85] as in Rosetta is based on natural examples – which is arguably the best structure prediction methodology till date. With the same intuition, we also attempted the direct use of empirical natural examples in our design pipeline, as an alternative to changing the hydrophobic character of amino acids at the interface (**strategy-1, a & b**). In that line, we picked up the RBD_{Spike} sequence from 3SCJ (i.e., the civet strain from predicted SARS-CoV: **Table 1**) motivated by its complementarity estimates (Sc: 0.52, EC: 0.32) – together which stood out to be the best among the homologous. Consequently, 3SCJ also had the closest approach to the 'probable' region of CP_{dock} (please refer to **Figure 2**) relative to the other homologous, which is to say the closest to being an optimal solution. The sequence of 3SCJ and 6VW1 were aligned, and, the aligned 3SCJ sequence (target) was directly threaded onto the main-chain trajectory of the ligand in 6VW1 (template). The threading protocol followed three simple rules of thumb. R1) For a deletion in the target sequence with respect to the template, the template amino acid was incorporated to fill the gap. R2) In case of substitution(s), the obvious choice was the target amino acid. R3) For identical amino acids in the corresponding positions in the template and the target, choosing either of the two meant the same. As a matter of fact, there were no insertions in the target with respect to the template (i.e., no gaps in the template).

Subsequent to threading, dynamic perturbations were introduced to the designed complexes (see **Materials and Methods**, section **2.9**) and the final atomic models were surveyed for their contact maps at the receptor – ligand interface. Absurdities in atomic contacts (design artifacts) such as those between two positively or two negatively charged amino acids (Lys-Lys, Glu-Asp etc.) were obviated, wherever found, by mutating the corresponding amino acid in the originally threaded sequence (e.g., Lys→Glu, Glu→Arg etc.). Such 'artifact cleaning mutations' were chosen based on overall knowledge of atomic interactions in proteins. Such mutations often involved alteration in the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 September 2020 doi:10.20944/preprints202009.0006.v1

hydrophobic character of the amino acids as well. This process gave rise to an iterative (threading → mutation → contact-map)$_n$ cycle in the protein design pipeline. Each resultant contact map was rigorously and manually scrutinized wherein other mutable positions were jotted down that could intuitively raise the EC while retaining the Sc. At instances, drastic changes like deleting a bulky side-chain (e.g., Phe→Ala) were also attempted. Charged amino acids were introduced as well as eliminated to favor and forbid the formation of salt-bridges. To eliminate the negative charge in Glu, Asp, they were mutated to corresponding polar variants (Gln, Asn). Attempts were also made to deliberately incorporate extended hydrophobic packing (i.e., introducing Ile, Met at strategic places etc.) as well as aromatic stacking (introducing Tyr, His etc.) at the interface. The final evaluation of the complexes were made by the complementarity measures and their mapping in CP_{dock}. Again, a total of 50 redesigned alternatives were constructed and tested in CP_{dock}. Among the given alternatives, this set could fairly cover all non-redundant 'presumably sensitive' point mutations and their combinations. Each individual case was carefully scrutinized with visual structural intervention of their redesigned interfaces to remove design artifacts. When plotted in CP_{dock}, their population distribution in a close cluster ensured empirical thresholds in both measures to be naturally satisfied (Sc_{min}: 0.402, EC_{min}: 0.168). In other words, the range of values obtained in the whole set were tight in both complementarity measures (Sc: [0.4 – 0.72], EC: [0.17 – 0.34]). Obtaining such tightly spaced numbers does not seem to be possible by random design or a mere reshuffling of sequence. To test this, scrambled sequences (see **Materials and Methods**, section **2.2.1**) were generated and undertaken the same analysis. Just as the cases for strategies 1a & 1b, clear discriminatory clusters obtained for the hits and the scrambled sequences (**Figure 6**) with practically no overlap. The disjointedness of the two clusters was clearer and more convincing than the earlier two sets (strategies 1a & 1b).

An apparent saturation was ensured in terms of covering arguably the whole spectrum of 'sensitive' mutations attempted on the plausible mutational hot-spots. The analyses were greatly helped by the rigorous repeated use of visual structural examination. Interestingly, shape complementarity of the 'hits' in this third set (Strategy-2) appeared to have a much wider range (almost twice) than corresponding ranges of electrostatic complementarity. More interestingly, there was not a single case with the EC raised to 40%. The difference in geometric fit among the designed sequences may cause from mutations either resulting in undue holes being created at the interface or leading to short contacts. The two events involve truncation and forced incorporation of bulky groups (e.g., Gly→Trp & Tyr→Val respectively) at the designed interface. At the same time, there appears to be natural evolutionary constraints on the upper limit of EC at this interface, which does not seem possible to be oversteped by different levels of protein engineering using the pull of 20 naturally occurring amino acids. The resultant EC values (natural as well as designed) physically mean quasi-stable to stable binding. The ones that are stable (i.e., optimal in terms of CP_{dock}) were the ones of interest to be considered further. Overall, there appears to be strong natural and evolutionary control over the dynamics of RBD_{Spike} – ACE2 binding. The top 25 sequences were filtered based on their residence relative to the optimal zone in CP_{dock}, and, considered further. The filtering also accompanied careful individual visual re-scrutiny of their interface. Its but trivial that these sequences were more closely spaced in CP_{dock} and mapped to the 'probable' / 'less probable' regions (i.e., optimal / near-optimal zones). Again, all filtered sequences were successfully validated for fold-compatibility (averages: 2.84 ±0.16 in CS_{gl}; 0.017 ±0.0002 in CS_{cp}).

It was unambiguous from the comparison of the three plots constructed for the three different sets of design (**Figure 6**) that the predicted solutions gradually improved from set-1 to set-3 reflected in the gradual north-eastern shift of the clusters (black dots in the plots). In other words, the homology-based design performed the best among the three. It was also evident from these results that the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 September 2020 doi:10.20944/preprints202009.0006.v1

'scrambled' sequences may indeed serve as negative controls in the future experimental validation of the current hypothesis.

A demonstrative example is cited in **Figure 7**, wherein, a case consisting of three designed sequences (HM0, HM3, HM5) selected from the pool (set-2) collectively portrays the impact of strategic point mutations. For HM5, the designed sequence contains a single point mutation (493-Q→N) with respect to the initially threaded sequence (3SCJ_E on 6VW1_E, referred to as HM0 in **Figure 7**). In the third case (HM5), the designed sequence further contains a second strategic point mutation (505-Y→H) over and above the earlier mutation. Here in this particular triad, the one with the single point mutation (HM3) gives somewhat better numbers (Sc: 0.710, EC: 0.251) than the one (HM5) with the additional aromatic mutation (Sc: 0.605, EC: 0.292), both better than the threaded sequence alone (HM0; Sc: 0.546, EC: 0.297). This demonstrates the scope and benefit of strategic point mutations to be invoked on the threaded homologous sequence to further improve the solution. Taken together with the natives (6VW1_E, 3SCJ_E), the results show a gradual shift towards a more balanced optimal solution upon threading (HM0) followed by subsequent strategic point mutations (HM3, HM5). The full-length sequences of these designed RBD mimics are provided in **Supplementary Dataset S1**.

3.8. Dynamic persistence of the binding of the selected designed structural mimics

Two best predicted solutions (HM19, HM21) designed from strategy-2 (sequences provided in **Supplementary Dataset S1**) were undertaken for long MD simulations (see **Materials and Methods**, section **2.9**) to study the dynamic persistence of the binding parameters. They had attained {Sc, EC} values of {0.614, 0.267} and {0.687, 0.310} respectively. To that end, all atom explicit water MD simulation production runs were performed for 200 ns each, wherein, simulated coordinates were accumulated at an interval of 50 ps resulting in 4000 snapshots (or time-stamps) for each simulated protein-complex. The C^{α}-RMS deviations for the whole protein complexes were found to be 3.15 (±0.36) Å and 3.31 (±0.64) Å for HM19 and HM21 respectively with respect to their initial snapshots, averaged over entire trajectories (**Supplementary Figure S5**). The dynamic persistence of the complementarity measures were analyzed by running CP_{dock} on each of the snapshots and then plotting time-series plots individually for Sc, EC and a 2D Euclidean distance measure (E2d: see **Materials and Methods**, section **2.10**) estimating the distance of the corresponding point from the 'probable' region of the plot. It is to be noted that E2d renders a value of 'zero' if the point falls into the 'probable' region of the plot.

For HM19, the average (and standard deviations) were found to be 0.670 (±0.047) for Sc, 0.283 (±0.078) for EC and 0.00421 (±0.012) for E2d (**Supplementary Figure S6**) while the fraction of snapshots falling in the 'probable', 'less probable' and 'improbable' regions of the plot were found to be 83.04%, 15.9% and 1.06% respectively (**Figure 8**). The same numbers for HM21 were found to be 0.676 (±0.048) for Sc, 0.247 (±0.072) for EC and 0.0048 (±0.0123) for E2d while the fraction of snapshots falling in the 'probable', 'less probable' and 'improbable' regions of the plot were found to be 81.15%, 18.20% and 0.75% respectively. Within the entire 200 ns trajectories, Sc could maximally be raised to 0.804, 0.797 for HM19, HM21 while their corresponding highest EC values attained were 0.583, 0.518 respectively. All numbers unequivocally indicate that the binding is dynamically stable and of high affinity. A representative case of the directed improve in the matching of electrostatic contact surfaces is portrayed in **Figure 9** citing the example of HM21, wherein, a comparison with **Figure 3** reveals the improvement in EC from the sub-optimal to the optimal range.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 September 2020 doi:10.20944/preprints202009.0006.v1

3.9. Nullifying the feasibility of the proposed designed therapeutics to compete with the ACE2 – angiotensin II binding

Angiotensin Converting Enzyme 2 (ACE2), a vital counter-regulatory component of the renin-angiotensin system (RAS), has recently got great attention in COVID-19 research for acting as a doorway to SARS-CoV-2 into the host cells [86–90]. Upon low blood flow, kidney cells convert the circulating pro-renins into renins which further take part in catalyzing angiotensinogen secreted by liver cells into angiotensin I [89]. Membrane bound angiotensin converting enzyme (ACE) present on vascular endothelial cell surface in lungs, thereafter, catalyzes the angiotensin I into angiotensin II which is a vasoconstrictor [89]. As a result, angiotensin II causes blood vessels to be constricted to increase blood pressure through engaging type 1 angiotensin receptor (AT1R) [90, 91]. Angiotensin II also increases blood pressure by stimulating adrenal cortex cells to secrete the aldosterone hormone. ACE2 is also expressed on heart, kidney cells along with the type 2 alveolar epithelial cells in the lungs [88]. To counterbalance the effect of RAS mediated increased blood pressure, ACE2 acts on angiotensin II and cleaves this peptide vasoconstrictor into Ang-(1-7) that further acts on Mas receptor releasing increased blood pressure [92]. Importantly, Ang-(1-7) and Mas receptor signaling can attenuate inflammatory reactions to take place through the production of nitric oxide. Apart from promoting vasoconstriction, angiotensin II –AT1R signaling has deleterious effects on leucocytes, endothelial cells and vascular smooth muscles through eliciting pro-inflammatory cascades by the production of cytokines, chemokines and adhesion molecules [93]. So under normal physiological condition, a fine balance between ACE-angiotensin II and ACE2-Ang-(1-7) has to be maintained in order to control the blood pressure and inflammation. As because SARS-CoV-2 utilizes the membrane bound ACE2 receptor to gain entry into host cells, so this is a condition where the viral Spike protein bound ACE2 receptors will be less available to angiotensin II. As a result, an equilibrium shift towards the increased activity of ACE-angiotensin II might drive acute lung injury. In addition to that, it has been hypothesized that SARS-CoV-2 – ACE2 binding causes increased internalization and shedding off of ACE2 receptor making it further unavailable for angiotensin II and the less production of Ang-(1-7). This can induce blood pressure along with direct parenchymal injury [94].

Our current work has considered the possibility of whether or not our designed plausible therapeutics can compete with the binding site of angiotensin II on ACE2 and may thereby disrupt the balance in RAS. In this regard, the NMR structure of angiotensin-II (PDB ID: 1N9V) was surveyed which shows zero deviations among its models. When, 1N9V was superposed onto the ligand (E) chain of 6VW1, the peptide is found away from the binding site (**Supplementary Figure S7**) having an RMS deviation of 4.28 Å. Based on a pairwise sequence alignment (in CLUSTAL-OMEGA [95]), the angiotensin-II sequence was then threaded onto the receptor binding site of the RBD_{Spike}, previously referred to as '6VW1_E_bs' (section **3.5**). The corresponding atomic model was subsequently built which resulted in an RMS deviation of 3.46 Å considering a stretch of just 8 mapped amino acids. Thus, the two molecular object does not seem to have any appreciable structural resemblance. Furthermore, when this built atomic model is placed onto the complex structure (6VW1), it has no proximity with the ACE-II receptor (displayed as solid surface in **Supplementary Figure S7**, bottom panel). No atoms were found at the native RBD_{Spike} – ACE2 interface. Angiotensin-II is a small bent linear peptide (**Supplementary Figure S7**, top panel) while the Spike protein binding site in ACE2 is a single α-helix (section **3.3**). From a structural perspective, they therefore have little chance to fit together into a valid binding model. Rather, a deep groove or a pocket is generally required to engulf small molecules such as angiotensin-II without having the necessity to have a proper shape and/or electrostatic match [96–98]. This further nullifies the possibility of a binding conflict with angeotensin II at the Spike binding site of ACE2. Taken together, there does not seem to be any convincing structural rationale to favor a plausible

conflict between the proposed therapeutic intervention and the ACE2 – angiotensin-II binding which is central to the RAS.

3.10. Comparing the proposed therapeutic intervention with the current *state-of-the-art*

One of the prime focuses of the recent research advances on anti-viral therapeutics for SARS-CoV-2 has been on utilizing the already available knowledge on the host cell entry mechanisms of SARS-CoV, MERS and other coronaviruses. Three general pathways that could lead to the development of potential antiviral therapeutics are (i) repurposing through the testing of pre-existing antiviral drugs, (ii) by high throughput screening of small molecules and (iii) through the redevelopment of new drugs or neutralizing antibodies or vaccines. Our current study proposes a non-trivial protein design approach to develop antiviral therapeutics that might act as potential competitive inhibitors of the SARS-CoV-2 RBD_{Spike}. After gaining insight into host cell entry mechanisms, importantly through the revelation of X-ray crystallographic structure of SARS-CoV-2 Spike protein binding to its cognate receptor, ACE2, on human cells [21, 24, 99], the drug designing methods are primarily revolving around the S protein subdomain blockers for obvious reasons.

There are also peptide-based approaches involving strategic contextual design of hybrid and fusion peptides. Such a hybrid peptide has been computationally constructed by linking two discontinuous fragments of ACE2 (residues: 22-44 and 351-357) by a linker glycine [100]. In addition to designing of small peptides from ACE2 sequence, clinical grade soluble hACE2 has proven to be a promising therapeutic candidate molecule which has shown to block the entry and growth of SARS-CoV-2 in blood vessel and kidney organoids system [101]. In order to develop potential therapeutics against SARS-CoV-2, the RBD in S1 subunit has not only been targeted but also the heptad repeat 1 (HR1) and HR2 domain in S2 subunit has been targeted [102]. Lipo-peptide such as EK1C4 has been demonstrated to be the most potent fusion inhibitor [102, 103]. Further, evidences have been put forward in support of significant efficacies of peptide inhibitors derived from the HR2 domain which can block the viral and host cell membrane fusion [104].

Alternatively, it has been shown by wet-lab experiments in hACE2 expressing cells that the recombinant RBD could block the entry of both the SARS-CoV and SARS-CoV-2 into the host cells [105]. A recent MD simulation study coupled with bio-layer interferometry [106] has targeted the ACE2 peptidase domain (PD) α1 helix where the SARS-CoV-2 RBD binding actually occurs. This 23-mer peptide fragment (residues 21-43) can effectively bind to SARS-CoV-2 RBD at a very low nano-molar affinity (K_d=47 nM) thereby posing a high possibility to interfere with the viral entry into host [106]. Importantly, although their peptide-based drug designing approach means to bypass the alteration in ACE2 physiological functions, but, it still remains to be checked the actual effect of their RBD_{Spike} blocker in terms of titters in human system. Such approaches are essentially aiming for an ‘antigen arrest’ before the pathogen reaches the host pulmonary system. A similar approach has also been adapted using nanobodies for directed delivery of neutralizing antibodies of the RBD_{Spike} [107]. In complete contrast, our approach takes the alternative route to develop therapeutics which may potentially block the RBD_{Spike} binding site on the cognate receptor, ACE2. We take advantage of the quasi-stable native binding of RBD_{Spike} to ACE2 in SARS-CoV-2 and aim to appreciably increase the binding stability while retaining high affinity. The mutations were directly performed on the native experimental co-complex and the proposed designed variants are end-products of cycles of rigorous computational screening through high-level structural descriptors. The proposed ‘high affinity stable binding’ in the designed structural mimics should therefore serve as potential blockers of the native Spike protein for its cognate receptor. Aligned approaches have shown the effect of key residue substitutions in SARS-CoV-2-CTD (see **Supplementary Figure 1**) leading to a 4-fold increased affinity for receptor binding than that of the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 September 2020 doi:10.20944/preprints202009.0006.v1

native co-complex [24]. We further structurally cross-checked that the designed structural mimics does not seem to have a realistic chance to interfere with the physiological function of ACE2 (section **3.10**). Furthermore, the prescribed designed structural mimics being substantially smaller in size (of the order of $1/100^{th}$) than those of the full virus particles should be able to reach the binding sites at a much faster time-scale.

4. Conclusion

Quasi-stable binding appears to be one of the essential features of SARS-CoV-2 RBD_{Spike} – ACE2 interaction. Having said that, the ligand possesses a high affinity towards its cognate receptor in the human host. This 'affinity-stability trade-off' seems to be evolutionarily fine-tuned in the corresponding protein family and fold – as revealed from the study of homologous complexes (section **3.3**). The current study unravels this fine-tuning from coordinate driven local and non-local complementarity measures, {Sc, EC}. While being counterbalanced by compensatory shape constraints (attributed to high affinity), attainment of sub-optimal electrostatic matching at the interface certainly appears to be a characteristic feature of this oligomeric association, conserved through evolution. Even in the designed co-complexes, other than a low fraction of snapshots (varying from 1.5 to 5.5%) in the long MD simulations, the maximally elevated EC value was found not more than 40% (section **3.7.1, 3.7.2**). That too, given the design-protocol being directed to raise the EC (section **3.7**). The amino acid composition of the extracellular RBD_{Spike} does seem to be non-trivial (compared to those in native globular proteins) involving solvent-exposed hydrophobic residues. This appears to be causally related to the conformational switch between a 'standing up' and a 'lying down' position of the RBD and the corresponding change in the membrane environment it encounters. As a consequence, the interface has a bunch of hydrophobic – polar interactions coupled with weak aromatic stabilization. Furthermore the co-complex structure is reminiscent of a molecular handshake rather than a hug or a cling (section **3.3**) like those found in analogous co-complexes in other related respiratory viral disorders (section **3.4**). Together, this leads to the high reaction-prone nature of the RBD_{Spike}. Also, the RBD_{Spike} – ACE2 interaction is intricately coupled with the host-protease mediated peptide-cleavage (as detailed in the **Introduction**), which, being an enzymatic reaction, involves covalent bond-breaking and -making. The interaction is thus concomitantly linked to the ephemeral formation of a transition state (TS) involving a saddle point, as is natural to enzyme kinetics. This also strongly speaks in favor of the quasi-stable nature of the RBD_{Spike} – ACE2 interaction. Quantum chemical calculations (in combination with molecular mechanics) may be invoked to reveal the plausible mechanism of the associated enzymetic 'cleavage' reaction (outside the scope of the current study). This quasi-stable character of the interaction has been utilized beneficially towards the directed design of the structural mimics aimed to serve as plausible blockers of the RBD_{Spike} – ACE2 interaction. The objective of the exercise was to improve the interaction-stability while keeping intact the already attained high affinity so that the designed mimics can actually block the host-pathogen interaction by stably occupying the binding sites on the receptor. Accordingly, EC was designed to be raised in an iterative design cycle while retaining the Sc at least native-like. As a matter of fact, both complementarity measures could be raised substantially in the best predicted designed complexes with respect to their native estimates. Starting from a native {Sc, EC} value of {0.56, 0.13}, we could statistically hit the $\{0.6 \rightarrow 0.7\}$ and $\{0.3 \rightarrow 0.4\}$ ranges in Sc and EC respectively (section **3.7.2**) for the designed interfaces, in spite of the evolutionary constraints in EC. Together these ranges in {Sc, EC} falls in the 'optimal' zone for high affinity stable binding in protein co-complexes – as revealed from the mapping of the corresponding points in CP_{dock}. This was possible by means of realizing the benefits of directly adapting natural examples in the design process. Thus, the use of homology-based design coupled with strategic mutations altering the hydrophobic characters of key amino acids appeared to be judicious in achieving the desired goal. Hence, the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 September 2020 doi:10.20944/preprints202009.0006.v1

paper may also be viewed to have presented a design methodology *per se*, the applicability and robustness of which are to be tested across related host-pathogen systems.

~

Tables

Table 1. Comparison the complementary estimates of the homologous RBD_{Spike} – ACE2 bound complexes. In the Table, P_{count} refers to the percentage of interfacial residues falling in the 'improbable' regions of the plot; N_{lsp}, N_{imp}: numbers of interfacial residues falling respectively in the 'less probable' and 'improbable' regions of the residue-wise Complementarity Plot (CP_{int}; see section **2.6**).

PDB ID	SPIKE RBD SOURCE	STRAIN	ACE2	Sc	EC	CS_I	P_{count}	N_{lsp}	N_{imp}
3D0G	2002-2003 SARS-CoV	HUMAN	HUMAN-CIVET CHIMERIC	0.17	-0.65	-6.56	58.3	0	4
3D0H	2002-2003 SARS-CoV	CIVET[4]	HUMAN-CIVET CHIMERIC	0.21	-0.45	-1.92	47.1	2	5
3SCL	SARS-CoV	EPIDEMIC	HUMAN-CIVET CHIMERIC	0.38	0.06	1.18	22.9	6	5
3SCJ	PREDICTED SARS-CoV	CIVET	HUMAN	0.52	0.32	1.45	20.0	7	0
2AJF	SARS-CoV	HUMAN	HUMAN	0.42	0.21	1.09	23.5	5	3
6VW1	SARS-COV-2 COVID-19	CHIMERIC	HUMAN	0.56	0.13	1.15	14.7	5	3

Table 2. Reaction proneness of the RBD_{Spike} binding site surveyed by the accessibility (*rGb*) score

Input protein fragment	*rGb*
6VW1_AE	0.052
6VW1_E	0.028
6VW1_E_hotspot	0.019
6VW1_E_bs	-0.055
threshold for native-like features	0.011

Figures

4 *Paguma larvata*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 September 2020 doi:10.20944/preprints202009.0006.v1

Figure 1. The Complementarity Plots (CP_{int} and CP_{dock}). The composite figure represents the two variants of the Complementarity Plot CP_{int} and CP_{dock}. CP_{int} (upper panel) is the residue-wise plot, plotting the residue-wise complementarity estimates, S_m vs. E_m [43] for interfacial residues – which is further distributed into three sub-plots (CP1, CP2, CP3) based on their burial of solvent exposure of the plotted residues. CP_{dock} (lower panel) is for the whole interface {Sc, EC}. The inner island colored in 'purple', the outer rim in 'mauve' and the rest in 'sky blue' corresponds to the 'probable', 'less probable' and 'improbable' regions of the plots. The pictorial demonstration is made on the very structure of 6VW1 (i.e., the RBD_{Spike} – ACE2 complex in CoV-2) displayed at the right-bottom of the composite diagram. The interfacial residues of the ligand (RBD_{Spike}: cyan cartoon) which are in physical contact with the receptor (ACE2: orange-yellow) are presented as their van der wall's dot surfaces colored according to their corresponding residences in CP_{int} ('probable': violet, 'less probable': magenta, 'improbable': violet-purple).

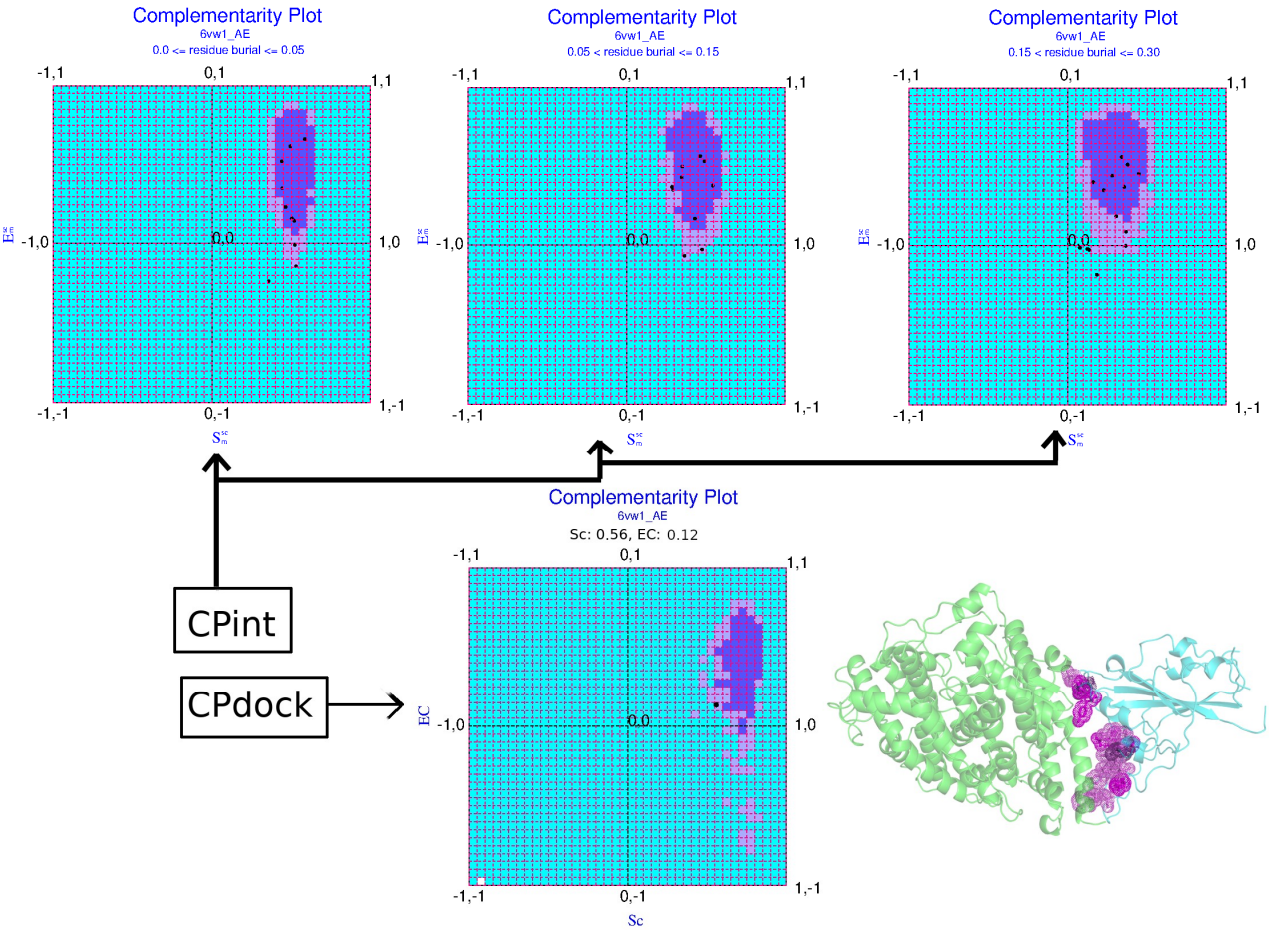


Figure 2. The dynamics of RBD_{Spike} – ACE2 binding from complementarity estimates. The left panel of the figure shows the superposed RBD_{Spike} – ACE2 co-complexes from the homologs (see

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 September 2020 doi:10.20944/preprints202009.0006.v1

section **3.3**) in cartoon covered with mesh representation. The receptors and the ligands are colored in green and magenta respectively. The right panel shows the mapping of their corresponding {Sc, EC} points in CP_{dock} as per mentioned in the embedded legend.

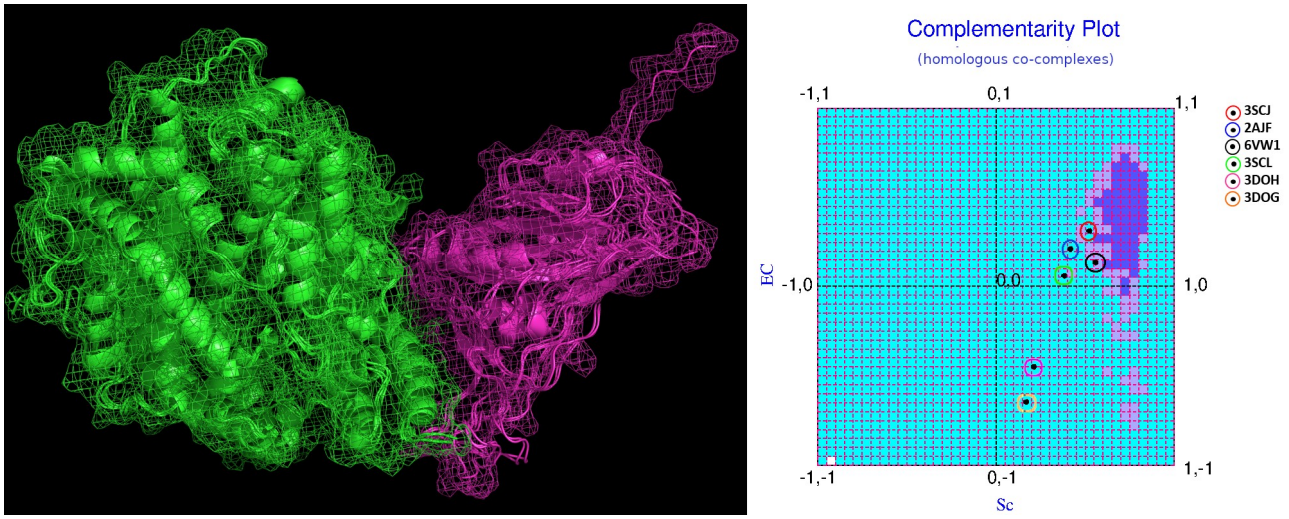


Figure 3. Electrostatic surface of the native co-complex in 6VW1. Panel A) and B) map the electrostatic potential surface of the ligand due to the electric fields coming from the ligand itself

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 September 2020 doi:10.20944/preprints202009.0006.v1

(self) and the receptor (partner) respectively. Likewise, panel C) and D) map the electrostatic surface of the receptor due to the electric fields coming from the receptor (self) and the ligand (partner) respectively. In each panel, the thick white arrows indicate whether the surface potentials are due to 'self' (panels A, C) or 'partner' (panels B, D). Further, in each panel, the molecular partner represented as 'cartoon' is colored 'yellow', if it is contributing to the potential (i.e., in case of partner-potentials), and, 'dim gray' otherwise (self-potentials). The electrostatic surface coloring was done in Chimera [108] using Delphi [41] electrostatic focusing files (.cube) with a color scale set to -10 kT/e for 'pure blue' to +10 kT/e for 'pure red'. As can be seen, there is very little match of counter-colors (red and blue's) between corresponding patches on both 'contact surfaces' (ligand and receptor) due their respective self- and partner-potentials – which means weak anti-correlation due to unfavorable electrostatic interactions.

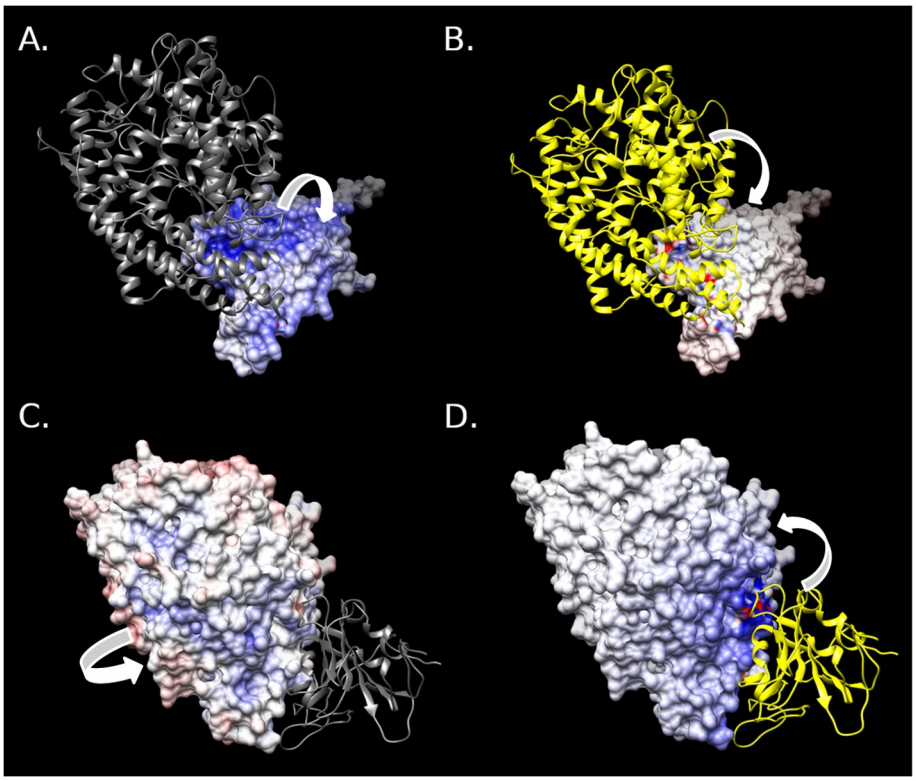
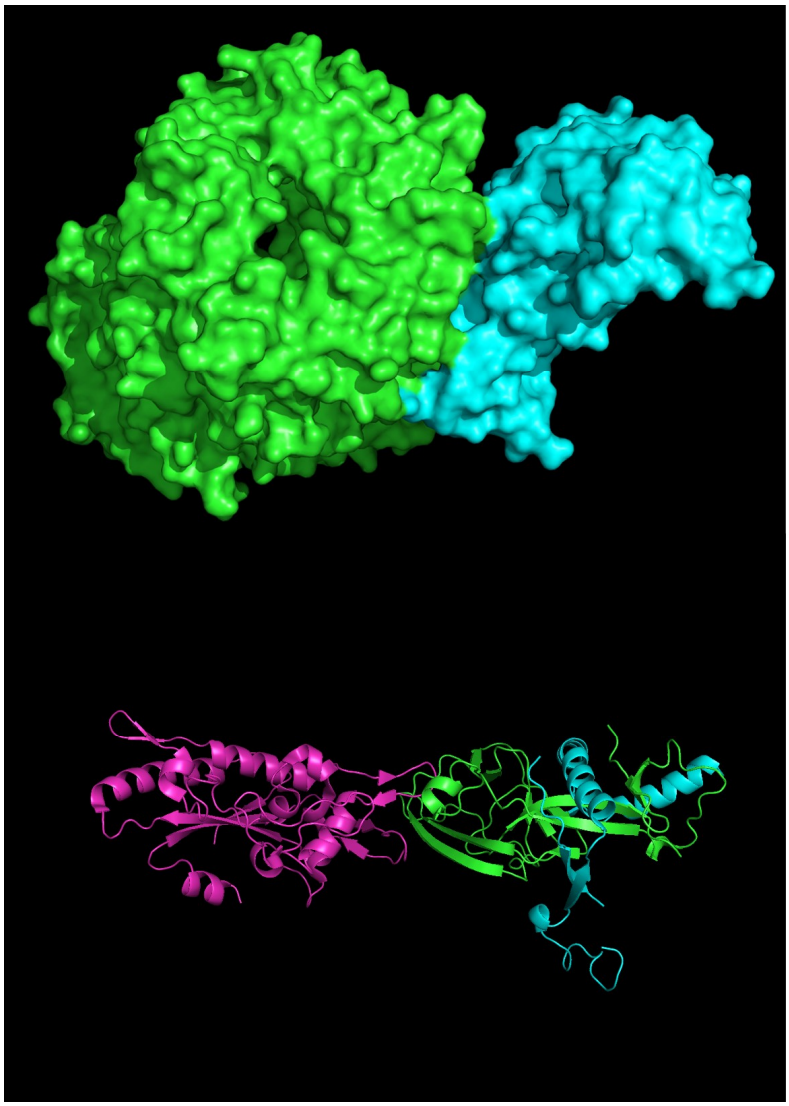
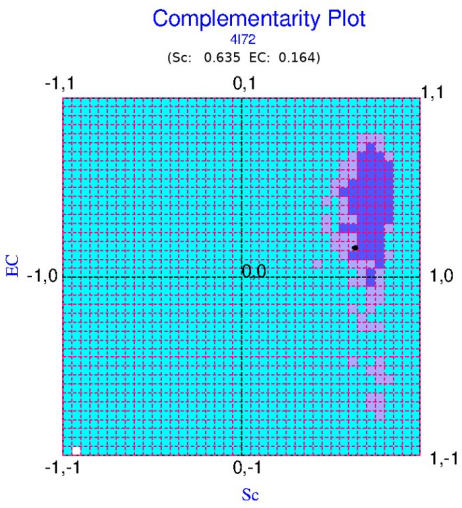


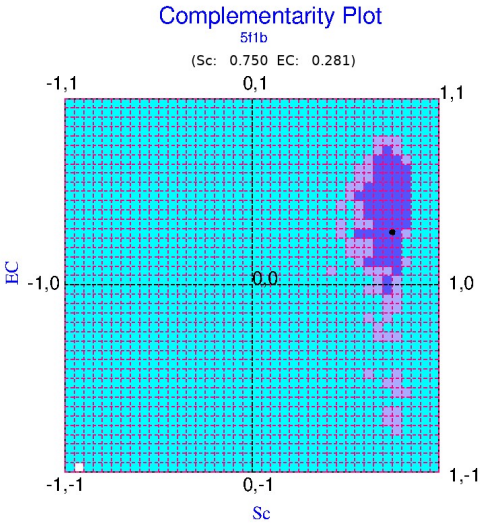
Figure 4. Analogous complexes of SARS-CoV-2 - RBD_{Spike} – ACE2 in MERS, Ebola: Dynamics of binding from complementarity. The top and bottom panels of the figure represent

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 September 2020 doi:10.20944/preprints202009.0006.v1

the co-complexes in MERS and Ebola respectively, their structures on the left and the corresponding mapping of their {Sc, EC} points in CP_{dock} on the right.

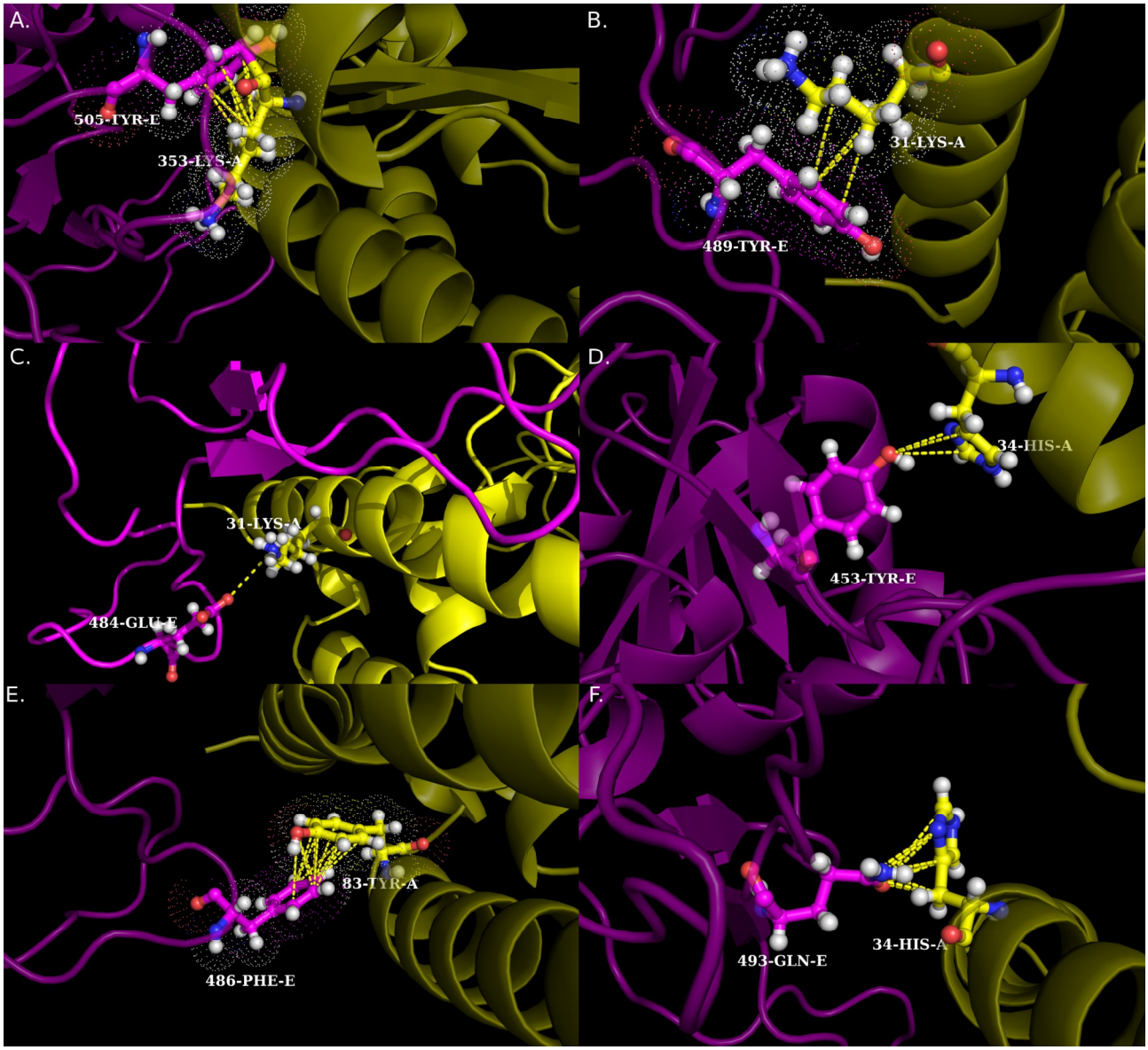






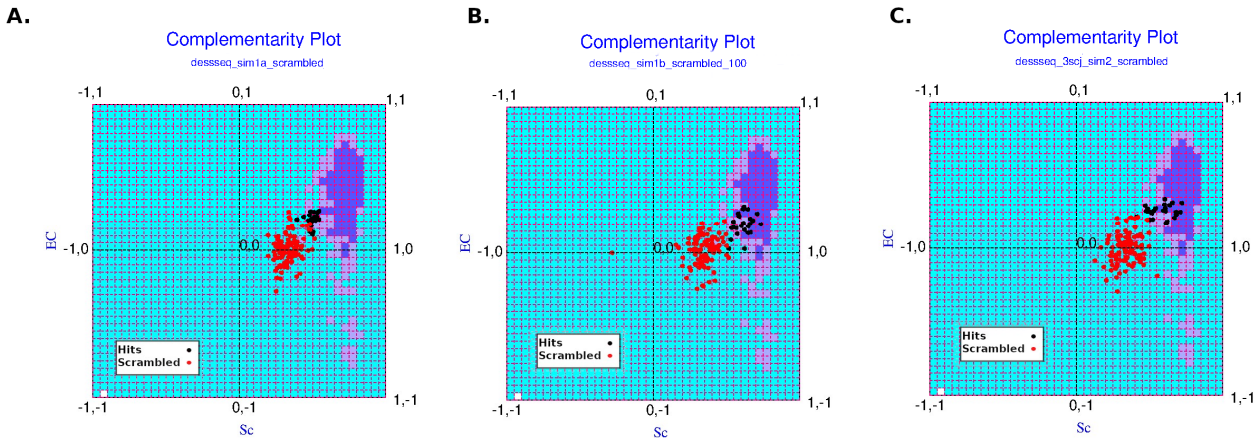
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 September 2020 doi:10.20944/preprints202009.0006.v1

Figure 5. The RBD_{Spike} – ACE2 interface in SARS-CoV-2: non-trivial interactions. A) and B) represents extended packing between aromatic rings and consecutively connected mythelene groups of elongated charged amino acids; C) portrays the only salt-bridge found at the interface; D) and F) are instances of polar atom mediated interactions involving an aromatic ring while E) presents aromatic stacking with a slide and an open angle.



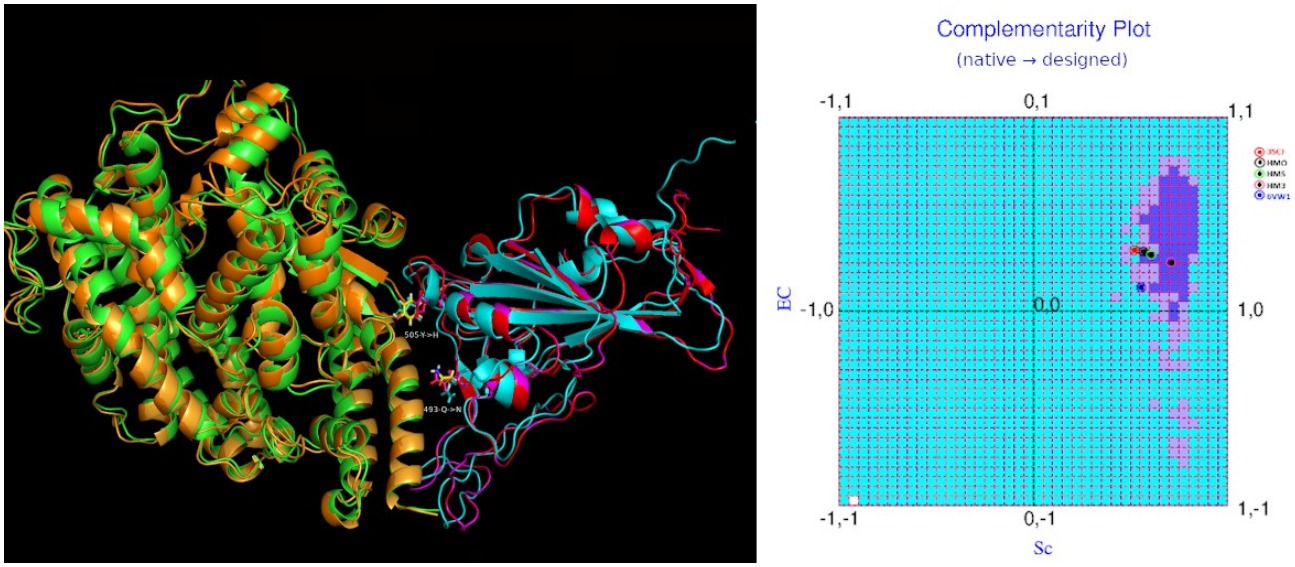
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 September 2020 doi:10.20944/preprints202009.0006.v1

Figure 6. The solution space: From alteration of hydrophobic character to homology-based design. Panels (A) (B) and (C) represents the solution space for strategies 1a, 1b and 2 respectively (as referred in section **3.7.1**, **3.7.2**). The red dots represent the {Sc, EC} points obtained for the corresponding scrambled sequences.



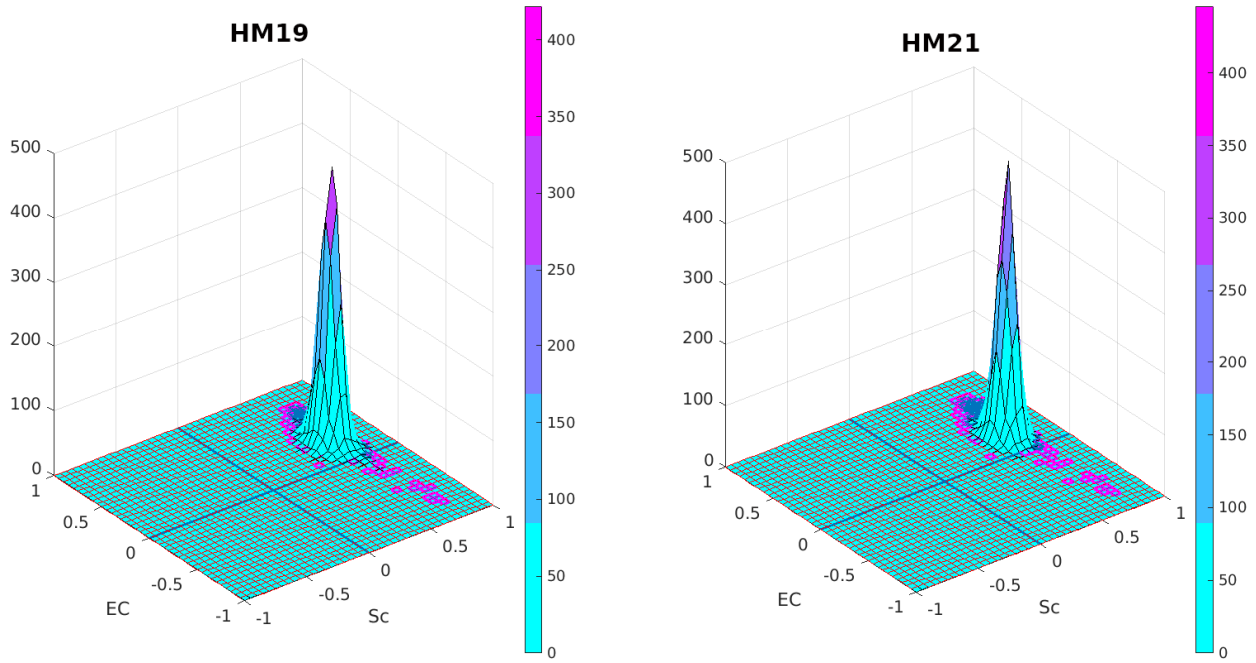
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 September 2020 doi:10.20944/preprints202009.0006.v1

Figure 7. Homology-based design of the CoV-2 RBD_{Spike}: Signatures of stable high affinity binding. The top panel displays the superposed co-complexes of HM0, HM3 and HM5 with their designed RBD_{Spike} chains colored in cyan, magenta and red respectively. The mutations are highlighted in form of sticks. The bottom panel shows the mapping of their corresponding {Sc, EC} points in CP_{dock} as per mentioned in the embedded legend.



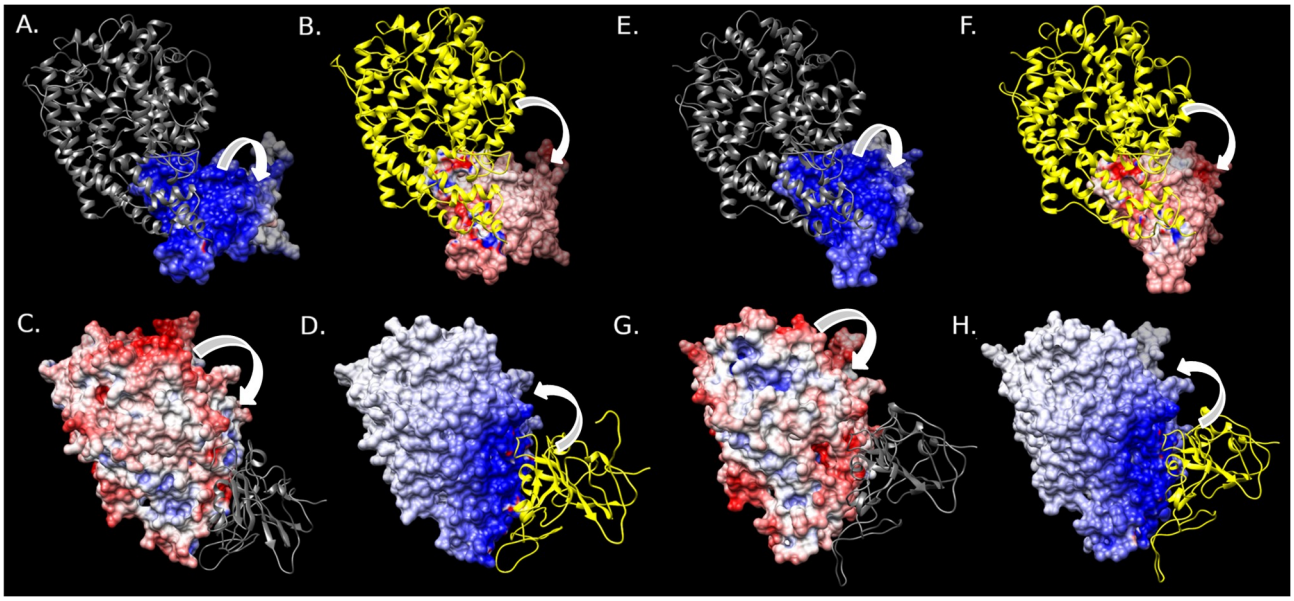
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 September 2020 doi:10.20944/preprints202009.0006.v1

Figure 8. Population density plots for the simulated trajectories corresponding to the selected designed complexes. These plots may also be described as the three-dimensional versions of CP_{dock}, wherein, the third dimension (i.e., the height) represents the frequency distribution of the points spanning the 2D {Sc, EC} map. Colorbars given beside each plot represent the raw frequencies corresponding to each X-Y square-grid of the 2D plot. The grid-contours of the 'probable' and 'less probable' regions (partly visible as shadows on the X-Y plane) are demarcated by colors: magenta and blue respectively.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 September 2020 doi:10.20944/preprints202009.0006.v1

Figure 9. Electrostatic surface representation of one of the best predicted designed co-complexes. Panel A-D represent the electrostatic surface map of HM21 subsequent to its design while panel E-H represent a snapshot picked up from its long MD trajectory mapping to its highest attained EC (section **3.8**). Rest of the figure may be described likewise to that of **Figure 3**. Briefly, panels A, C represent 'self-potentials' while B, D represent 'partner-potentials' realized on the ligand and receptor surface respectively for HM21. Self- and partner-potentials are as defined in the legend of **Figure 3**. Panels E, G and B, D represent equivalent cases to those of A, C and B, D respectively for the selected MD snapshot. Arrows indicate whether the surface potentials are due to 'self' (panels A, C, E, G) or 'partner' (panels B, D, F, H). Coloring of 'cartoon's are as in **Figure 3**. A direct comparison with **Figure 3** clearly shows that the match in counter-colors (red and blue's) improves appreciably between corresponding patches on the contact surfaces (due to their respective self- and partner-potentials) from that of the native co-complex. This reflects that the native weak anti-correlation in electrostatic surface potential could be significantly strengthened by the strategic design.



Acknowledgment & Funding

HKP would like to acknowledge MIIC at Linkoping University and UCL for activity support. We thank Souparno Adhikary for his initial initiative for simulation study. We sincerely thank Dr. Devlina Charkravarty for her kind favor in citing us the Affinity2 database used in this study.

Authors Contribution

The initial idea about a competitive inhibitor approach was conceived by HKP in a discussion with SB. Based on this idea, SB conceptualized the problem, designed the complete study and performed all the calculations. PS contributed with background survey, developed and drafted most of the basic structure of the introduction and actively helped in revising the MS. HKP intervened at different phases with crucial updates, inputs and revising the MS. SB wrote the main draft of the paper with active help from PS, HKP and rigorously revised the paper until satisfactory. DB helped with the molecular dynamic simulation part, contributed with results and participated with SB in fruitful discussions. All authors read and approved the final manuscript.

References

1. Andersen KG, Rambaut A, Lipkin WI, et al (2020) The proximal origin of SARS-CoV-2. Nature Medicine 26:450–452. https://doi.org/10.1038/s41591-020-0820-9

2. Li Q, Guan X, Wu P, et al (2020) Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. New England Journal of Medicine 382:1199–1207. https://doi.org/10.1056/NEJMoa2001316

3. Huang C, Wang Y, Li X, et al (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395:497–506. https://doi.org/10.1016/S0140-6736(20)30183-5

4. Zhu N, Zhang D, Wang W, et al (2020) A Novel Coronavirus from Patients with Pneumonia in China, 2019. New England Journal of Medicine 382:727–733. https://doi.org/10.1056/NEJMoa2001017

5. Lu R, Zhao X, Li J, et al (2020) Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. The Lancet 395:565–574. https://doi.org/10.1016/S0140-6736(20)30251-8

6. Chu H, Chan JF-W, Wang Y, et al (2020) Comparative replication and immune activation profiles of SARS-CoV-2 and SARS-CoV in human lungs: an ex vivo study with implications for the pathogenesis of COVID-19. Clin Infect Dis. https://doi.org/10.1093/cid/ciaa410

7. Dong L, Tian J, He S, et al (2020) Possible Vertical Transmission of SARS-CoV-2 From an Infected Mother to Her Newborn. JAMA 323:1846–1848. https://doi.org/10.1001/jama.2020.4621

8. Ferretti L, Wymant C, Kendall M, et al (2020) Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science 368:. https://doi.org/10.1126/science.abb6936

9. Kim Y-I, Kim S-G, Kim S-M, et al (2020) Infection and Rapid Transmission of SARS-CoV-2 in Ferrets. Cell Host Microbe. https://doi.org/10.1016/j.chom.2020.03.023

10. Li R, Pei S, Chen B, et al (2020) Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science 368:489–493. https://doi.org/10.1126/science.abb3221

11. Morawska L, Cao J (2020) Airborne transmission of SARS-CoV-2: The world should face the reality. Environ Int 139:105730. https://doi.org/10.1016/j.envint.2020.105730

12. Sanche S, Lin YT, Xu C, et al (2020) High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2. Emerging Infect Dis 26:. https://doi.org/10.3201/eid2607.200282

13. Tong Z-D, Tang A, Li K-F, et al Potential Presymptomatic Transmission of SARS-CoV-2, Zhejiang Province, China, 2020 - Volume 26, Number 5—May 2020 - Emerging Infectious Diseases journal - CDC. https://doi.org/10.3201/eid2605.200198

14. Wei WE (2020) Presymptomatic Transmission of SARS-CoV-2 — Singapore, January 23–March 16, 2020. MMWR Morb Mortal Wkly Rep 69:. https://doi.org/10.15585/mmwr.mm6914e1

15. Cai Y, Zhang J, Xiao T, et al (2020) Distinct conformational states of SARS-CoV-2 spike protein. Science. https://doi.org/10.1126/science.abd4251

16. Shang J, Wan Y, Luo C, et al (2020) Cell entry mechanisms of SARS-CoV-2. PNAS 117:11727–11734. https://doi.org/10.1073/pnas.2003138117

17. Jaimes JA, André NM, Chappie JS, et al (2020) Phylogenetic Analysis and Structural Modeling of SARS-CoV-2 Spike Protein Reveals an Evolutionary Distinct and Proteolytically Sensitive Activation Loop. Journal of Molecular Biology 432:3309–3325. https://doi.org/10.1016/j.jmb.2020.04.009

18. Walls AC, Park Y-J, Tortorici MA, et al (2020) Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell 181:281-292.e6. https://doi.org/10.1016/j.cell.2020.02.058

19. Wrapp D, Wang N, Corbett KS, et al (2020) Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science 367:1260–1263. https://doi.org/10.1126/science.abb2507

20. Zhou P, Yang X-L, Wang X-G, et al (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579:270–273. https://doi.org/10.1038/s41586-020-2012-7

21. Shang J, Ye G, Shi K, et al (2020) Structural basis of receptor recognition by SARS-CoV-2. Nature 581:221–224. https://doi.org/10.1038/s41586-020-2179-y

22. Letko M, Marzi A, Munster V (2020) Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. Nature Microbiology 5:562–569. https://doi.org/10.1038/s41564-020-0688-y

23. Mercurio I, Tragni V, Busto F, et al (2020) Protein structure analysis of the interactions between SARS-CoV-2 spike protein and the human ACE2 receptor: from conformational changes to novel neutralizing antibodies. Cell Mol Life Sci. https://doi.org/10.1007/s00018-020-03580-1

24. Wang Q, Zhang Y, Wu L, et al (2020) Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. Cell 181:894-904.e9. https://doi.org/10.1016/j.cell.2020.03.045

25. McKee DL, Sternberg A, Stange U, et al (2020) Candidate drugs against SARS-CoV-2 and COVID-19. Pharmacol Res 157:104859. https://doi.org/10.1016/j.phrs.2020.104859

26. Wu C, Liu Y, Yang Y, et al (2020) Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. Acta Pharmaceutica Sinica B. https://doi.org/10.1016/j.apsb.2020.02.008

27. Zhou Y, Hou Y, Shen J, et al (2020) Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. Cell Discovery 6:1–18. https://doi.org/10.1038/s41421-020-0153-3

28. Josephson K, Ricardo A, Szostak JW (2014) mRNA display: from basic principles to macrocycle drug discovery. Drug Discov Today 19:388–399. https://doi.org/10.1016/j.drudis.2013.10.011

29. Wan Y, Shang J, Graham R, et al (2020) Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. Journal of Virology 94:. https://doi.org/10.1128/JVI.00127-20

30. Smith MC, Gestwicki JE (2012) Features of protein-protein interactions that translate into potent inhibitors: topology, surface area and affinity. Expert Rev Mol Med 14:e16. https://doi.org/10.1017/erm.2012.10

31. Lawrence MC, Colman PM (1993) Shape complementarity at protein/protein interfaces. J Mol Biol 234:946–950. https://doi.org/10.1006/jmbi.1993.1648

32. McCoy AJ, Chandana Epa V, Colman PM (1997) Electrostatic complementarity at protein/protein interfaces. J Mol Biol 268:570–584. https://doi.org/10.1006/jmbi.1997.0987

33. Berman HM, Westbrook J, Feng Z, et al (2000) The Protein Data Bank. Nucl Acids Res 28:235–242. https://doi.org/10.1093/nar/28.1.235

34. Eswar N, Webb B, Marti-Renom MA, et al (2006) Comparative Protein Structure Modeling Using Modeller. Curr Protoc Bioinformatics 0 5:Unit-5.6. https://doi.org/10.1002/0471250953.bi0506s15

35. MUSCLE: multiple sequence alignment with high accuracy and high throughput. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC390337/. Accessed 26 Feb 2017

36. Krivov GG, Shapovalov MV, Dunbrack RL (2009) Improved prediction of protein side-chain conformations with SCWRL4. Proteins 77:778–795. https://doi.org/10.1002/prot.22488

37. Word JM, Lovell SC, Richardson JS, Richardson DC (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. J Mol Biol 285:1735–1747. https://doi.org/10.1006/jmbi.1998.2401

38. Basu S, Mukharjee D (2017) Salt-bridge networks within globular and disordered proteins: characterizing trends for designable interactions. J Mol Model 23:206. https://doi.org/10.1007/s00894-017-3376-y

39. Basu S, Biswas P (2018) Salt-bridge dynamics in intrinsically disordered proteins: A trade-off between electrostatic interactions and structural flexibility. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 1866:624–641. https://doi.org/10.1016/j.bbapap.2018.03.002

40. Connolly ML (1983) Analytical molecular surface calculation. Journal of Applied Crystallography 16:548–558. https://doi.org/10.1107/S0021889883010985

41. DelPhi Suite: New Developments and Review of Functionalities - Li - 2019 - Journal of Computational Chemistry - Wiley Online Library. https://onlinelibrary.wiley.com/doi/full/10.1002/jcc.26006. Accessed 25 May 2020

42. Li L, Li C, Zhang Z, Alexov E (2013) On the Dielectric “Constant” of Proteins: Smooth Dielectric Function for Macromolecular Modeling and Its Implementation in DelPhi. J Chem Theory Comput 9:2126–2136. https://doi.org/10.1021/ct400065j

43. Basu S, Bhattacharyya D, Banerjee R (2012) Self-Complementarity within Proteins: Bridging the Gap between Binding and Folding. Biophys J 102:2605–2614. https://doi.org/10.1016/j.bpj.2012.04.029

44. Basu S, Bhattacharyya D, Banerjee R (2014) Applications of complementarity plot in error detection and structure validation of proteins. Indian J Biochem Biophys 51:188–200

45. Basu S, Bhattacharyya D, Wallner B (2014) SARAMAint: The Complementarity Plot for Protein–Protein Interface. Journal of Bioinformatics and Intelligent Control 3:309–314. https://doi.org/10.1166/jbic.2014.1103

46. Basu S, Bhattacharyya D, Banerjee R (2013) SARAMA: A Standalone Suite of Programs for the Complementarity Plot—A Graphical Structure Validation Tool for Proteins. Journal of Bioinformatics and Intelligent Control 2:321–323. https://doi.org/10.1166/jbic.2013.1059

47. Basu S (2017) CPdock: the complementarity plot for docking of proteins: implementing multi-dielectric continuum electrostatics. J Mol Model 24:8. https://doi.org/10.1007/s00894-017-3546-y

48. Berendsen HJC, van der Spoel D, van Drunen R (1995) GROMACS: A message-passing parallel molecular dynamics implementation. Computer Physics Communications 91:43–56. https://doi.org/10.1016/0010-4655(95)00042-E

49. Scott WRP, Hünenberger PH, Tironi IG, et al (1999) The GROMOS Biomolecular Simulation Program Package. J Phys Chem A 103:3596–3607. https://doi.org/10.1021/jp984217f

50. K L-L, S P, K P, et al (2010) Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field. In: Proteins. https://pubmed.ncbi.nlm.nih.gov/20408171/. Accessed 22 May 2020

51. Shereen MA, Khan S, Kazmi A, et al (2020) COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. Journal of Advanced Research 24:91–98. https://doi.org/10.1016/j.jare.2020.03.005

52. Shim E, Tariq A, Choi W, et al (2020) Transmission potential and severity of COVID-19 in South Korea. International Journal of Infectious Diseases 93:339–344. https://doi.org/10.1016/j.ijid.2020.03.031

53. Bi Q, Wu Y, Mei S, et al (2020) Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. The Lancet Infectious Diseases 0: https://doi.org/10.1016/S1473-3099(20)30287-5

54. Yuen K-S, Ye Z-W, Fung S-Y, et al (2020) SARS-CoV-2 and COVID-19: The most important research questions. Cell Biosci 10:. https://doi.org/10.1186/s13578-020-00404-4

55. Wu F, Zhao S, Yu B, et al (2020) A new coronavirus associated with human respiratory disease in China. Nature 579:265–269. https://doi.org/10.1038/s41586-020-2008-3

56. Wang L, Friesner RA, Berne BJ (2010) Competition of electrostatic and hydrophobic interactions between small hydrophobes and model enclosures. J Phys Chem B 114:7294–7301. https://doi.org/10.1021/jp100772w

57. Zhou H-X, Pang X (2018) Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. Chem Rev 118:1691–1741. https://doi.org/10.1021/acs.chemrev.7b00305

58. Gabb HA, Jackson RM, Sternberg MJ (1997) Modelling protein docking using shape complementarity, electrostatics and biochemical information. J Mol Biol 272:106–120. https://doi.org/10.1006/jmbi.1997.1203

59. Bk S, Id K (1991) Protein docking and complementarity. J Mol Biol 221:327–346. https://doi.org/10.1016/0022-2836(91)80222-g

60. Basu S, Wallner B (2016) Finding correct protein-protein docking models using ProQDock. Bioinformatics 32:i262–i270. https://doi.org/10.1093/bioinformatics/btw257

61. Tsuchiya Y, Kinoshita K, Nakamura H (2006) Analyses of homo-oligomer interfaces of proteins from the complementarity of molecular surface, electrostatic potential and hydrophobicity. Protein Eng Des Sel 19:421–429. https://doi.org/10.1093/protein/gzl026

62. Lo Conte L, Chothia C, Janin J (1999) The atomic structure of protein-protein recognition sites. J Mol Biol 285:2177–2198

63. Banerjee R, Sen M, Bhattacharya D, Saha P (2003) The jigsaw puzzle model: search for conformational specificity in protein interiors. J Mol Biol 333:211–226

64. Yan Y, Huang S-Y (2019) Pushing the accuracy limit of shape complementarity for protein-protein docking. BMC Bioinformatics 20:1–10. https://doi.org/10.1186/s12859-019-3270-y

65. Dell'Orco D, Xue W-F, Thulin E, Linse S (2005) Electrostatic Contributions to the Kinetics and Thermodynamics of Protein Assembly. Biophysical Journal 88:1991–2002. https://doi.org/10.1529/biophysj.104.049189

66. Makhatadze GI, Loladze VV, Gribenko AV, Lopez MM (2004) Mechanism of Thermostabilization in a Designed Cold Shock Protein with Optimized Surface Electrostatic Interactions. Journal of Molecular Biology 336:929–942. https://doi.org/10.1016/j.jmb.2003.12.058

67. Marti DN, Rudolf Bosshard H (2003) Electrostatic Interactions in Leucine Zippers: Thermodynamic Analysis of the Contributions of Glu and His Residues and the Effect of Mutating Salt Bridges. Journal of Molecular Biology 330:621–637. https://doi.org/10.1016/S0022-2836(03)00623-5

68. Ojennus DD, Lehto SE, Wuttke DS (2003) Electrostatic interactions in the reconstitution of an SH2 domain from constituent peptide fragments. Protein Science 12:44–55. https://doi.org/10.1110/ps.0227903

69. Loladze VV, Makhatadze GI (2002) Removal of surface charge-charge interactions from ubiquitin leaves the protein folded and very stable. Protein Science 11:174–177. https://doi.org/10.1110/ps.29902

70. Schwehm JM, Fitch CA, Dang BN, et al (2003) Changes in Stability upon Charge Reversal and Neutralization Substitution in Staphylococcal Nuclease Are Dominated by Favorable Electrostatic Effects. Biochemistry 42:1118–1128. https://doi.org/10.1021/bi0266434

71. Kahraman A The Geometry and Physicochemistry of Protein Binding Sites and Ligands and their Detection in Electron Density Maps. 221

72. Maleki M, Vasudev G, Rueda L (2013) The role of electrostatic energy in prediction of obligate protein-protein interactions. Proteome Science 11:S11. https://doi.org/10.1186/1477-5956-11-S1-S11

73. Zhang Z, Witham S, Alexov E (2011) On the role of electrostatics on protein-protein interactions. Phys Biol 8:035001. https://doi.org/10.1088/1478-3975/8/3/035001

74. Kundrotas PJ, Alexov E (2006) Predicting 3D structures of transient protein-protein complexes by homology. Biochim Biophys Acta 1764:1498–1511. https://doi.org/10.1016/j.bbapap.2006.08.002

75. Kundrotas PJ, Alexov E (2006) Electrostatic Properties of Protein-Protein Complexes. Biophysical Journal 91:1724–1736. https://doi.org/10.1529/biophysj.106.086025

76. Vreven T, Moal IH, Vangone A, et al (2015) Updates to the Integrated Protein–Protein Interaction Benchmarks: Docking Benchmark Version 5 and Affinity Benchmark Version 2. Journal of Molecular Biology 427:3031–3041. https://doi.org/10.1016/j.jmb.2015.07.016

77. Patel DJ (1994) A molecular handshake. Nature 367:688–689. https://doi.org/10.1038/367688a0

78. Levine L, Edsall G (1981) Tetanus Toxoid: What Determines Reaction Proneness? J Infect Dis 144:376–376. https://doi.org/10.1093/infdis/144.4.376

79. Honig B, Yang AS (1995) Free energy balance in protein folding. Adv Protein Chem 46:27–58

80. Chaudhury S, Berrondo M, Weitzner BD, et al (2011) Benchmarking and analysis of protein docking performance in Rosetta v3.2. PLoS ONE 6:e22477. https://doi.org/10.1371/journal.pone.0022477

81. Simoncini D, Zhang KYJ (2013) Efficient Sampling in Fragment-Based Protein Structure Prediction Using an Estimation of Distribution Algorithm. PLOS ONE 8:e68954. https://doi.org/10.1371/journal.pone.0068954

82. Rohl CA, Strauss CEM, Misura KMS, Baker D (2004) Protein Structure Prediction Using Rosetta. In: Methods in Enzymology. Academic Press, pp 66–93

83. Lyskov S, Gray JJ (2008) The RosettaDock server for local protein-protein docking. Nucleic Acids Res 36:W233-238. https://doi.org/10.1093/nar/gkn216

84. Leaver-Fay A, Tyka M, Lewis SM, et al (2011) ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. Meth Enzymol 487:545–574. https://doi.org/10.1016/B978-0-12-381270-4.00019-6

85. Simons KT, Kooperberg C, Huang E, Baker D (1997) Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. J Mol Biol 268:209–225. https://doi.org/10.1006/jmbi.1997.0959

86. Aleksova A, Ferro F, Gagno G, et al COVID-19 and renin-angiotensin system inhibition: role of angiotensin converting enzyme 2 (ACE2) - Is there any scientific evidence for controversy? Journal of Internal Medicine n/a: https://doi.org/10.1111/joim.13101

87. Ingraham NE, Barakat AG, Reilkoff R, et al (2020) Understanding the Renin-Angiotensin-Aldosterone-SARS-CoV-Axis: A Comprehensive Review. European Respiratory Journal. https://doi.org/10.1183/13993003.00912-2020

88. South AM, Tomlinson L, Edmonston D, et al (2020) Controversies of renin–angiotensin system inhibition during the COVID-19 pandemic. Nature Reviews Nephrology 16:305–307. https://doi.org/10.1038/s41581-020-0279-4

89. Peach MJ (1977) Renin-angiotensin system: biochemistry and mechanisms of action. Physiol Rev 57:313–370. https://doi.org/10.1152/physrev.1977.57.2.313

90. Paz Ocaranza M, Riquelme JA, García L, et al (2020) Counter-regulatory renin-angiotensin system in cardiovascular disease. Nat Rev Cardiol 17:116–129. https://doi.org/10.1038/s41569-019-0244-8

91. Sparks MA, Crowley SD, Gurley SB, et al (2014) Classical Renin-Angiotensin System in Kidney Physiology. Compr Physiol 4:1201–1228. https://doi.org/10.1002/cphy.c130040

92. South AM, Shaltout HA, Washburn LK, et al (2019) Fetal programming and the angiotensin-(1-7) axis: a review of the experimental and clinical data. Clin Sci 133:55–74. https://doi.org/10.1042/CS20171550

93. Dandona P, Dhindsa S, Ghanim H, Chaudhuri A (2007) Angiotensin II and inflammation: the effect of angiotensin-converting enzyme inhibition and angiotensin II receptor blockade. J Hum Hypertens 21:20–27. https://doi.org/10.1038/sj.jhh.1002101

94. Kuba K, Imai Y, Rao S, et al (2005) A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus–induced lung injury. Nature Medicine 11:875–879. https://doi.org/10.1038/nm1267

95. Sievers F, Higgins DG (2018) Clustal Omega for making accurate alignments of many protein sequences: Clustal Omega for Many Protein Sequences. Protein Science 27:135–145. https://doi.org/10.1002/pro.3290

96. Stockwell GR, Thornton JM (2006) Conformational diversity of ligands bound to proteins. J Mol Biol 356:928–944. https://doi.org/10.1016/j.jmb.2005.12.012

97. Kahraman A, Morris RJ, Laskowski RA, Thornton JM (2007) Shape Variation in Protein Binding Pockets and their Ligands. Journal of Molecular Biology 368:283–301. https://doi.org/10.1016/j.jmb.2007.01.086

98. Kahraman A, Morris RJ, Laskowski RA, et al (2010) On the diversity of physicochemical environments experienced by identical ligands in binding pockets of unrelated proteins. Proteins 78:1120–1136. https://doi.org/10.1002/prot.22633

99. Lan J, Ge J, Yu J, et al (2020) Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 581:215–220. https://doi.org/10.1038/s41586-020-2180-5

100. Huang X, Pearce R, Zhang Y (2020) De novo design of protein peptides to block association of the SARS-CoV-2 spike protein with human ACE2. Aging (Albany NY) 12:11263–11276. https://doi.org/10.18632/aging.103416

101. Monteil V, Kwon H, Prado P, et al (2020) Inhibition of SARS-CoV-2 Infections in Engineered Human Tissues Using Clinical-Grade Soluble Human ACE2. Cell 181:905-913.e7. https://doi.org/10.1016/j.cell.2020.04.004

102. Xia S, Liu M, Wang C, et al (2020) Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. Cell Research 30:343–355. https://doi.org/10.1038/s41422-020-0305-x

103. Xia S, Yan L, Xu W, et al (2019) A pan-coronavirus fusion inhibitor targeting the HR1 domain of human coronavirus spike. Science Advances 5:eaav4580. https://doi.org/10.1126/sciadv.aav4580

104. Xia S, Zhu Y, Liu M, et al (2020) Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein. Cellular & Molecular Immunology 17:765–767. https://doi.org/10.1038/s41423-020-0374-2

105. Tai W, He L, Zhang X, et al (2020) Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. Cellular & Molecular Immunology 17:613–620. https://doi.org/10.1038/s41423-020-0400-4

106. Zhang G, Pomplun S, Loftis AR, et al (2020) Investigation of ACE2 N-terminal fragments binding to SARS-CoV-2 Spike RBD. bioRxiv 2020.03.19.999318. https://doi.org/10.1101/2020.03.19.999318

107. J H, A LB, Rr R, et al (00:00:00.0) Neutralizing nanobodies bind SARS-CoV-2 spike RBD and block interaction with ACE2. Nature Structural & Molecular Biology. https://doi.org/10.1038/s41594-020-0469-6

108. Pettersen EF, Goddard TD, Huang CC, et al (2004) UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612. https://doi.org/10.1002/jcc.20084