article

Betacoronavirus binding dynamics relevant to the functional evolution of the highly transmissible SARS-CoV-2 variant N501Y

Patrick Rynkiewicz[1], Gregory A. Babbitt[1*], Feng Cui[1], André O. Hudson[1], Miranda L. Lynch[2]

[1]Thomas H. Gosnell School of Life Sciences, Rochester Institute of Technology, Rochester NY, USA 14623

[2]Hauptmann-Woodward Medical Research Institute, Buffalo NY, USA 14203

*corresponding email - gabsbi@rit.edu

Running title – comparative betacoronavirus binding dynamics

ABSTRACT

Comparative functional analysis of the binding interactions between various *Betacoronavirus* mutant strains and their potential multiple human target proteins is crucial for a more complete understanding of zoonotic spillovers of viruses that cause diseases like COVID-19. Here, employing hundreds of replicate sets of nanosecond scale GPU accelerated molecular dynamics simulations, we statistically compare atom motions of ACE2 and CD26 target proteins in both the presence and absence of different strains of the viral receptor binding domain (RBD) of the S spike glycoprotein. In all strains, we demonstrate a universally conserved functional binding signature of the viral RBD with the N-terminal helices of ACE2. We also identify a second more dynamically transient interaction of the viral N501 with the previously confirmed ACE2 K353 and two nearby novel sites, Q325 and the AAQPFLL 386-92 motif. We propose a model of the functional evolution of SARS-type zoonotic spillovers involving both (A) a conserved binding interaction with the N-terminal helices of ACE2 that is preadapted from viral interaction of the *Tylonycteris* bat coronavirus progenitor strain HKU4 with the SAMLI 291-5 motif in protein CD26 and (B) a more promiscuous and likely more evolvable interaction between viral N501 and the above-mentioned multiple regions of ACE2 that is preadapted from the bat viral interaction at the CD26 SS 333-4 motif. Our analysis of the highly transmissible N501Y B.1.1.7 and B.1.351 variants in SARS-CoV-2 supports this functional evolutionary two-touch model, identifying progression towards more stable viral binding of the ACE2 N-terminal helices and the ACE2 K353 region.

KEYWORDS Molecular dynamics, molecular evolution, viral binding, COVID 19

INTRODUCTION

In the world-wide COVID-19 pandemic there has been an unprecedented wealth of high-throughput sequencing surveillance of the SARS-CoV-2 virus and near real-time sequence-based analysis of the evolution of local viral strains. Sequence-based analysis can be highly insightful for detailing the phylogenetic history of viral strains. However, it is often difficult to ascertain the functional aspects of this rapid viral evolution without expensive and time-consuming functional binding assays that often need to follow highly specific biosafety level protocols when used with human pathogens (e.g. ELISA-based and solid phase assays for receptor binding specificity as well as microcalorimetry to quantify binding affinity). These functional assays can be aimed at identifying and quantifying the functional repercussions of single non-neutral variants from among the many more neutral variants generated in the normal process of rapid viral evolution. And they can be supplemented by sequence-based phylogenetic tests of selection as well as crystallography and computational structural-functional analyses. However, even when combined, these methods are still often incapable of fully deciphering the precise molecular functions that underlie viral transmission. Ultimately viral transmission is the consequence of soft-matter dynamics responding to weak chemical bonding interactions at many potential sites on both the viral and putative target proteins. Here, we present a systematic approach to the comparative statistical analysis of the dynamic motions of proteins generated from accelerated all-atom molecular dynamics simulations that is subsequently leveraged to characterize the functional binding signatures of viral strains on specific target proteins at single amino acid site resolution. The unprecedented pandemic of SARS-CoV-2 brings with it many questions regarding its functional evolution of both emergent and endemic strains in humans. With the current surveillance of new variants like the highly transmissible N501Y strain that has caused recent lockdowns in the winter of 2020-21 in southeast England, our comparative biophysical analysis of protein binding can offer a new perspective on the functional consequences of viral evolution during a pandemic.

The SARS-CoV-2 virus belongs to the class of positive-strand RNA viruses (genus *Betacoronavirus*) and is structurally defined by a nucleocapsid interior surrounded by the spike, membrane, and envelope proteins (Haan et al. 2000; Fung and Liu 2019). Structural membrane and envelope proteins are known to play a key role in virion assembly and budding from infected cells, making them potential targets for antiviral treatments (Haan et al. 2000). One prominent protein candidate for comparative studies of the propensity for viral transmission is the spike or S glycoprotein, which is responsible for initial viral contact with the target host receptor and subsequent entry into host cells. The spike protein consists of two homotrimeric functional subunits: the S1 subunit consisting of the viral receptor binding domain (RBD) interacting directly with the host, and the S2 subunit which directly fuses the virion to the host cellular membrane (Hoffmann et al. 2018; Shang et al. 2020). The S1 and S2 subunits of

coronaviruses are cleaved and subsequently activated by type-II transmembrane serine proteases and the subtilisin-like serine protease furin, the latter of which evidently primes SARS-CoV-2 for cellular entry but not SARS-CoV-1 (Follis et al. 2006; Shang et al. 2020). The boundary between the S1 and S2 subunits forms an exposed loop in human coronaviruses SARS-CoV-2, hCoV-OC43, hCoV-HKU1, and MERS-CoV that is not found in SARS-CoV-1 (Hoffmann et al. 2020) and suggests a complex evolutionary landscape for emergent SARS-type coronaviruses.

The spike protein of SARS-CoV-2 has been demonstrated to directly and specifically target the human angiotensin converting enzyme 2 (ACE2) receptor, which has a role in the renin–angiotensin system (RAS) critical for regulating cardiovascular and renal functions (Oudit et al. 2003; Ou et al. 2017; Ou et al. 2020). ACE2 is a glycoprotein consisting of 805 amino acid residues that is the primary target for docking and viral entry in many coronavirus infections, including those belonging to SARS-CoV and human alpha-coronavirus NL63 (Hofmann et al. 2005). ACE2 is classified as a zinc metallopeptidase (carboxypeptidase) involved in cleaving a single amino acid residue from angiotensin I (AngI) to generate Ang1-9 (Kuba et al. 2005). In addition, ACE2 also cleaves a single amino acid residue from angiotensin II (Ang II) to generate Ang1-7 (Kuba et al. 2005). ACE2 receptors reside on the surfaces of multiple epithelial cells, with tissue distribution predominantly in the lung and gut linings, but are also found in heart, kidney, liver and blood vessel tissues. Given the dominant role ACE2 plays in SARS-CoV-2 viral entry, understanding how sequence and structural variations affect the molecular dynamics of ACE2/RBD interactions can provide information to guide the rationale design and development of therapeutics targeting this interaction. But it is important to note that SARS-CoV-2 also targets other human protein receptors including several neurolipins (Cantuti-Castelvetri et al. 2020; Daly et al. 2020) and perhaps the known targets of other viral relatives such as the MERS-CoV target protein CD26. Therefore, it is important to demonstrate at a molecular level, how such promiscuity in binding interaction is achieved by the viral RBD and how it may have contributed to zoonotic spillovers from other species to humans. The comparative dynamic modeling of present, past, and potential progenitor viral strain interactions to ACE2 and other known targets of past outbreaks like protein CD26, can also potentially illuminate the dynamics of molecular interactions that may facilitate zoonotic transmission to humans as well as the molecular evolution toward emergent and endemic human viral strains.

A powerful tool for *in silico* analysis of protein-protein interactions is all atom accelerated molecular dynamics (aMD) simulation, which attempts to accurately sample the physical trajectories of molecular structures over time via traditional Newtonian mechanics played out over a modified potential energy landscape that encourages conformational transition states (Pierce et al. 2012). Molecular dynamics (MD) simulations are increasingly broadening these methods of inquiry by allowing further analysis of the rapid time scale dynamics of protein-protein interactions. Protein structures determined via x-ray

crystallography have previously been used to provide insights into viral evolution by comparing angstrom scale distances and conformational similarities. MD simulations have already been employed to map the hydrogen bonding topology of the SARS-CoV-2 RBD/ACE2 interaction, to explore the merits of drug repurposing, to characterize mutations, and to determine the stability of binding interactions between the SARS-CoV-2 RBD and several contemporary antiviral drugs (Ahamad et al. 2020; Al-Khafaji et al. 2020; Mittal et al. 2020; Muralidharan et al. 2020; Wang et al. 2020; Yu et al. 2020). While conventional MD and aMD simulations can provide unprecedented resolution on protein motion over time, interpretations based upon comparisons of single MD trajectories are statistically problematic and difficult to compare. This makes the application of MD to comparative questions regarding functional binding or impacts of mutation quite challenging. Given recent increases in GPU processor power, the statistical comparison of large replicates of aMD simulations is now an available solution to this problem. Because MD trajectories of complex structures like proteins are inherently complex and chaotic, even if they are adequately resampled, they do not often fit traditional model assumptions of many parametric statistical tests. Therefore, reliable comparisons of molecular dynamics need to further be based upon statistical methods that can capture the multimodal distribution space of protein motion and assign statistical significance for divergence in dynamics without over-reliance on the assumptions of normality in the distribution of molecular dynamics over time.

Here we employ DROIDS 3.0, a method and software tool designed to address these challenges in the statistical analysis and visualization of comparative protein dynamics (Babbitt et al. 2018; Babbitt et al. 2020), to subsequently analyze hybrid models of different *Betacoronavirus* strain S spike glycoprotein interactions with their human target proteins. DROIDS software employs a relative entropy (or Kullback-Leibler divergence) calculation on atom fluctuation distributions derived from large numbers of replicates of MD trajectories to quantify the local differences in both the magnitude and shape of the distribution of atom fluctuations in two different functional states of a protein. In this application, we compare viral bound versus unbound dynamics of target proteins to ascertain a computational signature of the local degree of dampened target protein motions during viral binding. DROIDS software also employs a non-parametric two sample Kolmogorov-Smirnov statistical test of significance, subsequently corrected for false discovery rates, to identify where these dampened dynamics upon viral binding are truly and locally significant. First, we characterize the interaction of SARS-CoV-2 Spike protein with human ACE2 both with and without the recently evolved and highly transmissible N501Y variant responsible for recent lockdowns in southeast England. Then, we also analyze models of emergent, endemic and zoonotic coronavirus strains in humans to gain more evolutionary perspective on the functional molecular dynamics of coronavirus viral entry into human host cells. We believe this comparative perspective on viral binding dynamics helps to shed light on evolutionarily conserved protein interactions that have primed certain protein

targets for zoonotic spillover events. Our profiles of binding dynamics include the endemic human strains of *Betacoronavirus* (hCoV-OC43 and hCoV-HKU1), emergent strains responsible for recent SARS-like outbreaks (i.e. SARS-CoV-1, SARS-CoV-2 and MERS-CoV), as well as models of hypothetical interactions with the bat progenitor strain bat CoV-HKU4. We confirm important key sites of binding interaction at the ACE2 N-terminal helices and K353 identified previously by Yan et al. (Yan et al. 2020) and identify two new key ACE2 interaction sites with the viral N501 that exhibit a promiscuous dynamic interplay along with the K353 site with important implications for binding propensity and binding promiscuity in relation to zoonotic coronavirus outbreaks. Importantly, we demonstrate that the new N501Y variant (lineage B.1.1.7) has lost this promiscuous binding in favor of a much more static and stable interaction with K353 alone. Lastly, we present a model of the functional evolution of the binding interactions that facilitate bat to human zoonotic spillover, including how they have changed in both human endemic and emergent viral strains.

RESULTS

Characterization of the SARS-CoV-2 to ACE2 binding signatures in the original and N501Y mutant strain

The modeled region of binding interaction consisted of the extracellular domain region of human ACE2, and the viral RBD specified in Fig 1A. The main key sites of the ACE2 interface with the viral RBD are shown in Fig 1B, with the N-terminal helices labeled in green and the three main sites that can potentially interact with the viral RBD N501 site are shown in magenta. The DROIDS pipeline (Babbitt et al. 2018; Babbitt et al. 2020) used to characterize the changes in ACE2 molecular dynamics due to viral binding is summarized Fig 1C. More details about this can be found in the methods section here as well as in our two software notes cited above. All the binding models developed and analyzed in this study are listed in Table 1. Comparisons of the molecular dynamics of SARS-CoV-2 RBD bound ACE2 to unbound ACE2 after long range equilibration revealed extensive dampening of atom fluctuations (shown color mapped in blue) at the ACE2 N-terminal helices as well as three downstream sites on ACE2 focused at Q325, K353, and a novel motif AAQPFLL 386-92 (Fig 2A-C). These color mapped structures show the local site-wise mathematical divergence protein backbone atom fluctuation (i.e. root mean square fluctuation averaged over N, C, Cα, and O) when comparing viral bound ACE2 dynamics to unbound ACE2 dynamics. They are presented for both the whole structural model (Fig 2A) and the viral binding interface with ACE2 (Fig 2B) along with sequence-based positional plotting of the signed symmetric Kullback-Leibler or KL divergence (Fig 2C). The site-wise positional profiles of average atom fluctuation in both the viral bound and unbound states (Fig S1), as well as the multiple test-corrected statistical significance of these KL divergences determined by a two-sample Kolmogorov-Smirnov or KS test is also given (Fig S2). NOTE: the

false discovery rate for the p-value adjustment of each test was determined using the Benjamini-Hochberg method to adjust for the fact that a separate p-value is generated for each amino acid in the sequence. Therefore, the p-value given at each site is adjusted for the probability of false discovery that is governed by the length of the polypeptide.

Binding interaction models of past human outbreak strain SARS-CoV-1 in complex with ACE2 (Fig 2D-E) indicates that the binding signature of SARS-CoV-1 and SARS-CoV-2 are nearly identical with both having stable interaction with N-terminal helices of ACE2 and promiscuous interactions with K353 and the same novel sites Q325 and AAQPFLL 386-92. However, the SARS-CoV-1 interactions at these sites appear more moderated than in SARS-CoV-2. Again, the site-wise atom fluctuation profiles and KS tests of significance of the differences in these fluctuations are also given (Fig S1 C-F and S2 C-F).

Characterization of the SARS-CoV-2 to ACE2 binding signatures in the N501Y variant strains B.1.1.7 and B.1.351

An identical characterization of the ACE2 binding signature created by the much more transmissible SARS-CoV-2 B.1.1.7 (UK) variant indicated a similar yet stronger binding profile at the N-terminal helices, with more attraction to the C helix (Fig 3A-C). The binding to the ACE2 K353 by the viral RBD Y501 was also far more stable while the former promiscuous interactions with Q325 and AAQPFLL 386-92 were markedly reduced (Fig 3B-C, with point of the viral RBD mutation N501Y marked by the green star in Fig 3B). In the B.1.351 (South African) variant, we also find a more stable binding pattern overall that involves both the N-terminal helices and K353 site in ACE2 (Fig 3D-E). In this variant we observed a retention of promiscuous binding of Y501 with the k353, Q325 and the AAQPFLL motif as well (Fig 3E). The three mutations prevalent in the RBD of the B.1.351 variant are shown with a green star in Fig 3D. These are N501Y, E484K and K417N/T. Both variants display an extraction of the K353 side chain away from the body of ACE2 and towards the viral RBD Y501 (Fig 3B and 3D) that is more or less maintained throughout the MD simulations.

Characterization of binding signatures of the bat progenitor strain HKU4 with its primary target CD26 and potential zoonotic spillover target ACE2

Binding interaction models of *Tylonycteris* bat progenitor strain batCoV-HKU4 RBD in complex with CD26 (its normal target) and ACE2 were also analyzed in the same manner as SARS-CoV-2 (Fig 4). The bat coronavirus HKU4 was strongly and significantly dampened molecular motion in two precise regions of CD26, the SAMLI 291-295 motif, and double serine motif at SS-333-4 (Fig 4A-C). However, upon interacting with ACE2, the bat coronavirus RBD only dampened the atom fluctuation at a broader region of the ACE2 N-terminal helices (Fig 4D-E) utilizing the same region that interacts with the CD26 SAMLI motif. Thus, the interaction with the ACE2 N-terminal helix appears pre-adapted by the evolution of the interaction with SAMLI

on CD26. The site-wise atom fluctuation profiles and KS tests of significance of the differences in these fluctuations are also given (Fig S3). In both comparative MD simulations, we observe that the bat CoV-HKU4 viral RBD maintains more secondary structure in the form of anti-parallel beta sheet over the both the SAMLI 291-5 stie in CD26 and the N-terminal helices in ACE2.

Survey of other coronavirus binding signatures in human endemic strains and past emergent outbreaks

Binding interaction models of human endemic strains hCoV-OC43 and hCoV-HKU1 in complex with ACE2 were also analyzed identically to previous models. The interaction with the most pathologically benign and longest co-evolved strain hCoV-OC43 (Fig 5A-C) shows highly enhanced dampened motion due to binding only at the N-terminal helices of ACE2, while the interaction with hCoV-HKU1 demonstrates this as well, but also adds moderate promiscuous interactions of N501 with K353 and Q325 (Fig 5D-E). The enhanced binding at the N-terminal helix is clearly associated with the evolution of extended loop structure of the viral RBD at this region of the interface (Fig 5A). The site-wise atom fluctuation profiles and KS tests of significance of the differences in these fluctuations are also given (Fig S1 C-D and S2 C-D). Binding interaction models of past human outbreak strain MERS-CoV in complex with both CD26 and ACE2 (Fig 6) indicate a two-touch interaction with the same CD26 sites as observed with bat CoV-HKU4 binding (i.e. SAMLI 291-5 and SS 333-4) but with an even larger interaction with the several sites just downstream of the double serine motif (Fig6B-C). The MERS-CoV/ACE2 interaction indicates a strong interaction with the N-terminal helices and only a weak interaction with K353 (Fig6D-E). In the MERS-CoV MD simulations we observe a similar maintenance of beta sheet over the SAMLI 291-5 region in CD26, however this is partially lost over the N-terminal helices when the MERS-CoV is presented to ACE2.

Summary of the comparative protein dynamic surveys

In comparisons of molecular dynamics simulations between four strains of human-pathogenic coronaviruses, SARS-CoV-2, SARS-CoV-1, HKU1, and OC43, all four RBDs were associated with statistically significant dampening of molecular motion in the N-terminal helices of human ACE2, indicated by blue color mapping corresponding to a negative Kullback-Leibler (KL) divergence on the ACE2 PDB structure mapped at single amino acid resolution (Figures 2-6). With the exceptions of bat CoV-HKU4 and hCoV-OC43, all RBDs were also associated with dampened molecular motion of ACE2 at residue K353. The absence of a secondary interaction between OC43 and ACE2 is further supported by multiple sequence alignment of the RBD loop region proximal to the K353 site, where OC43 is the only surveyed strain with a polar uncharged threonine residue in place of a small aliphatic residue (Fig. S4). A significant dampening effect is

also present in two novel sites near to K353, the Q325 and 386-AAQPFLL-392 motif, when interacting with SARS-CoV-1 and SARS-CoV-2 (Fig. 2), indicating a transiently dynamic or promiscuous interaction of the viral RBD with these three key ACE2 binding sites. In the recent evolution of the far more transmissible N501Y variants (Fig3), the stabilization of RBD binding interaction with the N-terminal helices and sites Q325, K353, and 386-AAQPFLL-392 in ACE2 while favoring a more stable K353 interaction highlights an important molecular explanation that linking population transmission rate with the evolution of increased binding specificity in this new variant. As of January 2021, ongoing updates on this variant can be found at the following link
European Centre for Disease Prevention and Control. Rapid increase of a SARS-CoV-2 variant with multiple spike protein mutations observed in the United Kingdom – 20 December 2020. ECDC: Stockholm; 2020.

***In silico* Mutagenesis Study of hACE2 and SARS-CoV-2 Bound Dynamics**

To confirm the relative importance of the three promiscuous sites in the binding interaction of coronavirus RBDs to human ACE2, eight total *in silico* targeted mutagenesis studies were performed, on ACE2 and the viral RBD alone (Table 2). K353A, a mutation previously identified to abolish SARS-CoV-1 RBD interaction with ACE2, neutralized dampening of the N-terminal helices during binding with the SARS-CoV-2 RBD (Fig. S7) supporting the observation of decreased infectivity in the mutant (Li et al. 2005). Within the newly identified 386-AAQPFLL-392 motif, a mutation of the nonpolar N-terminal double alanine to the charged polar glutamic acid resulted in an additional strong dampening of molecular motion in the ancestral Q325 site, with a KL divergence of less than -3 between the unbound and bound states (Fig. S8A). Mutating the C-terminal FLL residues to EEE in the 386-AAQPFLL-392 motif resulted in no overall change to the binding characteristics between the SARS-CoV-2 RBD and ACE2 (Fig. S8B). Upon mutating V739 in ACE2 to hydrophilic glutamate, the N-terminal helix region and 386-AAQPFLL-392 motif showed no significant change in molecular motion between bound and unbound states (Fig. S9A). This same pattern of binding characteristics could be seen in the mutation of V185 to glutamate in the loop region of the SARS-CoV-2 RBD, proximal to ACE2 during binding (Fig. S10A). Mutating V185 to leucine in the RBD resulted in the same binding characteristics, resulting in a pattern of dampening in molecular motion that was the same as seen in SARS-CoV-1 (Fig. S10B). All surveyed mutations maintained significant dampening in K353 of ACE2 during binding. The mutation of V739 in ACE2 to leucine resulted in dampening of Q325 in the interaction, as is seen in SARS-CoV-1, and a lack of dampening in the AAQPFLL motif as seen in the SARS-CoV-2 wild type (Fig. S9B). However, mutating V739 in ACE2 to glutamate resulted in pronounced dampening only at the N-terminal helices and K353. Upon mutating polar K408 in the SARS-CoV-2 RBD to nonpolar alanine, the Q325 interaction

disappeared and only interactions with N-terminal helices, K353, and 382-AAQPFLL-392 remained (Fig. S11).

Building a Model for Key ACE2 Binding Regions in Interactions with Coronavirus Receptor Binding Domains

Based on our results comparing regions of dampened molecular motion in ACE2 and CD26 target proteins across zoonotic, emergent, and endemic *Betacoronavirus* strains, we propose a model of protein interaction for coronavirus RBD evolution and human specificity highlighting the role of the two distinct touch points between the viral RBD and the human target proteins (Fig. 7A). The black arrowhead in the model signifies a conserved region of sustained interaction between the coronavirus RBDs and the N-terminal helix region of ACE2. The red arrowhead indicates more promiscuous interactions with Q325 and K353, present individually across SARS-CoV-2 as well as SARS-CoV-1, MERS, and HKU1, as well as the newly identified 386-AAQPFLL-392 motif prominent only in the wild-type SARS-type spike protein RBD. This two-touch model highlights a common interaction point of the viral RBD with the N-terminal helices of ACE2 that is conserved in all known strains of potentially human-interacting *Betacoronavirus*, demonstrated by the universal dampening of ACE2 molecular motions upon viral binding. This interaction corresponds structurally with the viral bat progenitor strain RBD interaction with the SAMLI 291-5 site in CD26. The second and more complex dynamic touch point in the model involves the contact of the viral RBD N501 with the previously described contact site K353, as well as the two nearby novel sites of interaction reported above. It corresponds with the viral bat progenitor strain interaction with SS 333-4 in CD26. This promiscuous interaction site seems considerably apparent in the two pathogenic SARS-type strains, suggesting that the binding interactions during zoonotic spillover may follow a common functional evolutionary path (Fig. 7B), utilizing a preadapted interaction with the SAMLI motif in CD26 to establish binding with the ACE2 N-terminal helices in bat reservoir populations that quickly establishes a 'slippery' two touch binding interaction that allows promiscuous binding in SARS-type emergent strains shortly after zoonotic spillover. In SARS-CoV-2, this has most recently evolved to higher specificity for ACE2 particularly through the strengthening of the ACE2 K353 and the N-terminal interactions. This appears quite different from the evolutionary path of the binding interactions in human endemic strains like hCoV-OC43 (Fig.7B), where an insertion event appears to have enhanced the loop structures near the N-terminal helices of ACE2 and subsequently strengthen binding interactions in this region.

DISCUSSION

Here we conducted statistical comparisons of hundreds of replicate sets of nanosecond scale GPU accelerated molecular dynamics simulations on viral bound and unbound target

proteins in an effort to survey various emergent, endemic, and potentially zoonotic *Betacoronavirus* strains in humans. We present evidence that strongly suggests a common route of functional molecular evolution occurring at the binding interface of the viral receptor binding domain and their primary protein targets ACE2 and CD26. We propose a two-touch model of viral evolution that may be evolving similarly during SARS-type *Betacoronavirus* zoonotic spillover events from *Tylonycteris* bat to human populations. We present computationally derived signatures of receptor binding domain interactions with target human proteins, in the *Tylonycteris* bat progenitor strain batCoV-HKU4 and five human strains, the emergent MERS-CoV, SARS-CoV-1 and SARS-CoV-2 strains and the endemic hCoV-OC43 and hCoV-HKU1 strains. Our studies of the physics of viral binding at single amino acid site resolution demonstrate that the coronavirus spike glycoprotein receptor binding domain (RBD) has a strong, static, and well-defined binding interaction occurring in two distinct regions, or "touch points" of human CD26 (also likely the CD26 ortholog in bats as well). In its interactions with the human ACE2 target protein, the spike glycoprotein RBD appears pre-adapted through its binding interaction involving one of these sites, the SAMLI 291-5 motif on CD26, to further allow broad binding interactions with the whole N-terminal helical region of ACE2. This preadaptation to binding ACE2 N-terminal helices appears common to all betacoronavirus strains we have examined and is very well demonstrated by dampened molecular motion in all our simulations. The prevalence of the ACE2 N-terminal helix interactions among all our human and bat progenitor viral strain models leads us to believe that this is pre-adapted binding interaction may be a key prerequisite interaction for zoonotic spillovers serving as a physical anchor, or primary "touch point" that allows further evolution of various coronavirus RBD interaction with ACE2 in humans. In both SARS-CoV-1 and SARS-CoV-2, and to a lesser extent in the human coronavirus strain, hCoV-HKU1, we identify three key binding sites in ACE2 that have complex and perhaps transient dynamic interactions with the N501 RBD site in the early stages of this evolution. Interactions at all three of these sites are clearly evidenced by our comparative statistical method applied to molecular dynamics. These three sites in ACE2 are the previously implicated K353 site (Yan et.al 2020) as well as two new novel flanking sites at Q325 and 386-AAQPFLL-392. When the viral RBD appears to transiently interact with these three residues, we refer to this interaction as forming a "slippery touch point" during RBD binding (Fig. 7B). This interaction is not apparent in simulations on the bat progenitor strain batCoV-HKU4, on the endemic hCoV-OC43, nor the MERS-CoV interactions with ACE2 (NOTE: the MERS primary human target protein was CD26). This transient viral RBD interaction with multiple sites on ACE2 in emergent SARS-type strains suggests that the interaction with K353 reported by (Yan et al. 2020), is not as stable as their initial structural analysis might imply. We hypothesize that the binding interactions at these three sites are perhaps more evolvable in SARS-type strains than the interactions at the N-terminal helices. The recent appearance of the highly transmissible N501Y (lineage B.1.1.7) variant in southeast England during the winter of

2020-21 and the later emergence of the N501Y/E484K/K417N|T mutations in the South African B.1.351 and Brazilian P.1 variants appears to further support the functional evolution of more stable binding at this same location on the ACE2 target protein. When we characterized the effect of the mutation on the binding dynamics in this region, we found that this one mutation has greatly stabilized the Y501 interaction with the ACE2 K353 site and reduced the transient and promiscuous interactions with the other two flanking binding sites. Thus, the stabilization of two touch points in the RBD binding has enhanced the binding stability, and by extension perhaps also the transmissibility of the SARS-CoV-2 virus, bringing it much closer to the molecular binding profile of the bat progenitor strain batCoV-HKU4 with its primary target protein, CD26. Interestingly, the human endemic strain hCoV-OC43 appears to have evolved towards enhanced stability on ACE2 via a quite different mechanism, losing this second touch point altogether in favor of a strong broad interaction with the ACE2 N-terminal helices that is facilitated by the evolution of enhanced loop structures on the viral RBD near its interface at this first touch point on ACE2. The slightly more pathogenic human strain hCoV-HKU1 also shows shares this feature with hCoV-OC43, but still also shows weak interactions with K353.

It has been recently hypothesized and demonstrated that promiscuity in protein-protein interactions is related to a lack of protein stability (Cohen-Khait et al. 2017) and it would also appear, through our work on this system, that a lack of protein stability may also be related to the molecular facilitation of zoonotic spillovers of certain viral pathogens as well. We believe the complex dynamic interaction at the second 'slippery touch point' in our two-touch model may represent the molecular manifestation of a lack of binding specificity that may characterize many viral binding interactions during emergent SARS-type outbreaks when the co-evolution between a viral binding domain and a potential host target receptor is still historically very recent. Our simulations suggest several alternative paths of viral evolution in emergent versus endemic strains that have both favored a more precise and specific targeting of human ACE2, which may also be associated with enhanced transmissibility as well.

While our approach offers the considerable advantage of combining a comparative statistical method and a physics-based modeling approach towards addressing functional molecular evolution, it is not without some pitfalls. Some potential limitations of MD simulations as a probative method for functional molecular evolution are its many implicit simplifying computational assumptions, its complex and inherently stochastic nature, and vary high computational expense (i.e. due to femtosecond time steps). Specifically, computational sampling of even the accelerated MD method employed here has strict hardware limitations, and even on modern graphics cards our simulations can typically have a cumulative runtime of several weeks to generate the proper statistical replication to compare physical time frames of only several hundreds of nanoseconds. In addition, MD simulations always involve some simplification of physics within the system being studied as it invariably ignores atomic charge

regulation, bond motions in the solvent, charge screening during interaction, and other macromolecular crowding effects. Insight into long-term micro-millisecond dynamics in large explicit solvent systems are still limited by currently available hardware, even when creative algorithms for accelerating MD simulations are used.

Glycosylation is another aspect of coronavirus spike protein biology that is not fully captured by our MD simulations, mainly due to lack of glycosylation in the functional binding interface of most of our key starting structures. We would argue that while these post-translational modifications can have in immediate impacts on dynamics, their probable lack of heritability may also minimize their role during viral evolution during outbreaks. In SARS-CoV-1, MERS-CoV, and hCoV-HKU1, there are 8 to 9 known N-linked glycosylation amino-acid sequons which aid in immune evasion (Turoňová et al. 2020; Watanabe, Allen, et al. 2020; Watanabe, Berndsen, et al. 2020; Yao et al. 2020) however their consistent heritability within viral populations during epidemics remains yet to be documented. Viral spike proteins are known to be extensively glycosylated to facilitate immune evasion in a process known as glycan shielding (Bagdonaite and Wandall 2018; Casalino et al. 2020). The full SARS-CoV-2 spike protein has been determined to contain 22 potential N-linked glycan sequons across each full length protomer (Watanabe, Allen, et al. 2020). Various expression systems and recombinant techniques have been employed for determining the potential occupancy and antigenicity of coronavirus glycan sequons, providing an overview of potentially functionally meaningful glycosylation sites but not comprehensively reviewing their presence in clinical samples (Watanabe, Allen, et al. 2020). Our study focuses on the conformational shielded receptor binding domain, which in both MERS-CoV and SARS-CoV-2 has been predicted to have only one to two N-glycosylation sites, significantly fewer than the rest of the spike protomer (Watanabe, Allen, et al. 2020; Watanabe, Berndsen, et al. 2020). Recent mapping of glycosylation sites in SARS-CoV-2 (Yao et al. 2020) also confirms only one site of RBD glycosylation (at N343) falling within our modeling region (i.e. sites 335-518) and which is far removed from the ACE2 interface and the site of the N501Y mutation. Nevertheless, glycosylation sites proximal to the SARS-CoV-2 RBD have been implicated in stabilizing the RBD hinge, but the potential functional importance of glycans at the interface of the viral RBD and human ACE2, and particularly their consistent heritability at the population-level, remain largely unknown. And while glycans are perhaps very important to understanding the extreme variation in clinical presentations of COVID-19, given that they are most likely non-heritable secondary modifications of structure during viral evolution, and are largely absence within the ACE2-RBD binding interface (Ke et al. 2020), we can justify having excluded them in our evolutionary studies of comparative molecular dynamics in emergent and endemic human *Betacoronavirus* strains.

Despite these limitations, we conclude that our identification of additional key residues in the binding interaction between the SARS-CoV-2 RBD and human ACE2 receptor, as well as

our evolutionary exploration of the two-touch model of RBD evolution provide a conceptual framework for future functional mutagenesis studies of this system. This will be especially important for understanding the functional evolution of transmissibility in new variants like N501Y, recently responsible for the large community lockdowns during the COVID-19 pandemic in southeast England in late 2020. Future surveys of *Betacoronavirus* circulating in past, present, and future human populations as well as molecular and clinical investigations of SARS-CoV-2 infection will likely continue to be further informed by the model interpretation of binding interaction that we present in this study. In future studies of computational models derived from potential zoonotic coronavirus strains, our binding model can lend greater interpretability to observations regarding evolutionary diversity in coronaviruses infecting reservoir species like bats, birds, and small mammals. Finally, as recent work has called out the importance of specific molecular motions in forming novel therapeutic targets for intervention in emerging zoonotic spillover events (Pierri 2020), tools for highlighting functionally conserved dynamics of conformational alterations of the interaction host and pathogen proteins will prove a valuable addition to the arsenal of modeling approaches available for drug development.

MATERIALS AND METHODS

PDB structure and hybrid model preparation

Structures of the two main human SARS variants of beta-coronavirus spike glycoprotein receptor binding domain (RBD) bound to human ACE2 receptor protein were obtained from the Protein Data Bank (PDB). These were SARS-CoV-1 (PDB: 6acg) (Song et al. 2018) and SARS-CoV-2 (PDB: 6m17) (Yan et al. 2020). Three additional hybrid models of viral RBD interaction with ACE2 consisting of human *Betacoronavirus* variants MERS-CoV (PDB: 5x5c, and PDB 4kr0) (Yuan et al. 2017), hCoV-OC43 (PDB: 6ohw) (Tortorici et al. 2019), and hCoV-HKU1 (PDB: 5gnb) (Ou et al. 2017) bound to ACE2 were generated by creating an initial structural alignment to the SARS-CoV-2 RBD/ACE2 model (PDB: 6m17) using UCSF Chimera's MatchMaker superposition tool (Pettersen et al. 2004) followed by deletion of the SARS-CoV-2 RBD in the model and any structures outside of the molecular dynamics modeling space defined in Figure 1A, leaving only the viral RBD and ACE2 receptor domain. These hybrid models of ACE2 interaction included viral receptor binding domains from MERS-CoV (PDB:5x5c and 4kr0) (Yuan et al. 2017), HCoV-OC43 (PDB: 6ohw) (Tortorici et al. 2019), and HCoV-HKU1 (PBD: 5gnb) (Ou et al. 2017) bound to the ACE2 protein from the SARS-CoV-1 model. Unbound forms of the ACE2 structure were obtained by deleting the viral structure in PDB: 6m17 and performing energy minimization for 2000 steps and 10ns of equilibration of molecular dynamics simulation in Amber18 (Case et al. 2005; Götz et al. 2012; Pierce et al. 2012; Salomon-Ferrer et al. 2013) prior to setting up production runs for the sampling regime (described below). Loop modeling and refinement were conducted where needed using Modeller in UCSF Chimera (Sali and Blundell 1993; Fiser et al. 2000). A hybrid model of the bat HKU4 beta coronavirus interaction with ACE2 was similarly

constructed using PDB: 4qzv (Wang et al. 2014), batCoV-HKU4 bound to the MERS target human CD26, as a reference for structural alignment to PDB: 6m17. Our hybrid models of the viral RBD consisted of the largely un-glycosylated region (represented in PDB: 6m17 from site 335-518). Only one glycan was removed using swapaa in UCSF Chimera at ASN 343 on the viral RBD located on the opposite side of the RBD-ACE2 interface (PDB: 6m17). Five glycans on the ACE2 receptor domain were also removed, none occurring near the binding interface (ASN 53, ASN 90, ASN 103, ASN 322, ASN 546). Single mutation models were also similarly constructed to examine the effects of single mutations of potentially large effect. These models are summarized in Table 1 and are available in a separate supplemental download file titled PDBmodels.zip.

Mutant model construction and rationale

To generate *in silico* site-directed mutagenesis computations, we created mutant models using the swapaa function in UCSF Chimera 1.13 by first swapping amino acids using optimal configurations in the Dunbrack rotamer library, then using 2000 steepest gradient descent steps of energy minimization to relax the regions around the amino acid replacement. Mutant models were chosen to test the predicted sites functional role mitigating local binding between the viral RBD and ACE2 by comparing amino acid replacements with either very similar or very different sidechain properties. This is summarized in Table 2. The SARS-CoV-2 N501Y mutant model was constructed using PDB:6m17. This model of the viral RBD only extended from site 335-518 and therefore only included this single mutation (N501Y) from among the seven other amino acid replacements and two deletions found to occur on the B.1.1.7 lineage (The total set of mutations defining this variant are del 69-70, del 144, N501Y, A570D, D614G, P681H, T716I and S982A). Because of its proximity to the binding interface, the N501Y mutation is believed to be one of the most functional. In the B.1.351 lineage, we were able to model three mutations in the viral RBD. These were N501Y, E484K and B417T.

Molecular dynamics simulation protocols

To facilitate the proper statistical confidence of the differences in rapid vibrational states in our model comparisons, many individual graphic processing unit (GPU) accelerated molecular dynamics (MD) simulations were produced in order to build large statistical replicate sets of MD trajectories. All accelerated MD simulations were prepared with explicit solvation and conducted using the particle mesh Ewald method employed by pmemd.cuda in Amber18 (Ewald 1921; Darden et al. 1993; Case et al. 2005; Pierce et al. 2012; Salomon-Ferrer et al. 2013) via the DROIDS v3.0 interface (Detecting Relative Outlier Impacts in Dynamic Simulation) (Babbitt et al. 2018; Babbitt et al. 2020). Simulations were run on a Linux Mint 19 operating system mounting two Nvidia Titan Xp graphics processors. Explicitly solvated protein systems were prepared using tLeap (Ambertools18) using the ff14SB protein force field (Maier et al.

2015) and protonated at all available sites using pdb4amber in Ambertools18. Solvation was generated using the Tip3P water model (Mao and Zhang 2012) in a 12nm octahedral water box and subsequent charge neutralization with Na+ and Cl- ions. All comparative molecular dynamics binding analyses utilized the following protocol. After energy minimization, heating to 300K, and 10ns equilibration, an ensemble of 200 MD production runs each lasting 0.2 nanoseconds was created for both viral bound and unbound ACE2 receptor (Figure 1B). Two examples of the 10ns RMSF equilibration are also shown in Figure 1C. Each MD production run was preceded by a single random length short spacing run selected from a range of 0 to 0.1ns to mitigate the effect of chaos (i.e. sensitivity to initial conditions) in the driving ensemble differences in the MD production runs. All MD was conducted using an Andersen thermostat under a constant pressure of 1 atmosphere (Andersen 1980). Root mean square atom fluctuations (rmsf) were calculated using the atomicfluct function in CPPTRAJ (Roe and Cheatham 2013). To assess the adequacy of the 10ns equilibration step prior to sampling in our many comparisons of strains, a longer range equilibration and production sampling was conducted upon only the SARS-CoV-2 model (PDB: 6m17). This simulation involved 100ns of equilibration followed by a 100 x 0.5ns sampling regime on the viral bound and unbound ACE2. This result is presented in Figure 2.

Comparative dynamics of bound/unbound and mutant/wild type functional states

Computational modeling of bound and unbound protein structure dynamics can identify both proximal and distal differences in molecular motion as a result of the binding interaction. In the interest of studying functional dynamics, we focus on regions with highly dampened molecular motion at the interface between the viral RBDs and ACE2. Given the chaotic nature of individual molecular dynamics simulations, ensembles on the order of hundreds of samples are required for comparing functional states and deriving statistically significant differences between them. Furthermore, binding interactions can potentially be disrupted or prevented entirely by dynamic thermal noise that exists inherently in the system. In our study, the dampened atom fluctuations of receptors ACE2, CD26, and ACE1 upon binding are indicative of molecular recognition, in the sense that weak bonding interactions overcome baseline thermal motion and lead to a persistent functional state.

The molecular dynamics of the viral bound and unbound models of CD26/ACE2 and lisinopril bound and unbound models of viral bound ACE1/ACE2 were generated and statistically compared using the DROIDS v3.0 comparative protein dynamics software interface for Amber18 (Babbitt et al. 2018; Babbitt et al. 2020). The symmetric Kullback-Leibler (KL) divergence or relative entropy between the distributions of atom fluctuation (i.e. root mean square fluctuation or *rmsf* taken from 0.01 ns time slices of total MD simulation time) on viral bound and unbound ACE2 were computed using DROIDS and color mapped to the protein backbone with individual amino acid resolution to the bound structures using a temperature

scale (i.e. where red is hot or amplified fluctuation and blue is cold or dampened fluctuation). The *rmsf* value is thus

$$rmsf = \frac{1}{4} \sum_{i=N,C,C_\alpha,O}^{4} \sqrt{\frac{1}{n} * \sum_{j=1}^{n} \left(v_{jx} - w_x\right)^2 + \left(v_{jy} - w_y\right)^2 \left(v_{jz} - w_z\right)^2} \quad (1)$$

where v represents the set of XYZ atom coordinates for i backbone atoms (C, N, O, or Cα) for a given amino acid residue over j time points and w represents the average coordinate structure for each MD production run in a given ensemble. The Kullback Leibler divergence (i.e. relative entropy) or similarity between the root mean square fluctuation (rmsf) of two homologous atoms moving in two molecular dynamics simulations representing a functional binding interaction (i.e. where 0 = unbound state and 1 = viral bound state) can be described by

$$KL_{divergence} = \sum_{t=50ps}^{T} \left[\left(rmsf_0 * \log\frac{rmsf_0}{rmsf_1}\right) + \left(rmsf_1 * \log\frac{rmsf_1}{rmsf_0}\right)\right] \Big/ T \quad (2)$$

where rmsf represents the average root mean square deviation of a given atom over time. More specifically, the rmsf is a directionless root mean square fluctuation sampled over an ensemble of MD runs with similar time slice intervals. Because mutational events at the protein level are typically amino acid replacements, this calculation is more useful if applied to resolution of single amino acids rather than single atoms. Because only the 4 protein backbone atoms (N, Cα, C and O) are homologous between residues, the R groups or side chains are ignored in the calculation and equation 3 can be applied. Since the sidechain atoms always attach to this backbone, rmsf still indirectly samples the dynamic effect of amino acid sidechain replacement as they are still present in the simulation. The reference state of the protein is unbound while the query state is viral bound. Therefore, this pairwise comparison represents the functional impact of viral binding on the ACE2 protein's normal unbound motion, where it is expected that viral contact would typically dampen the fluctuation of atoms at the sites of binding to some degree. This calculation in equation 2 is used to derive all the color maps and KL divergence (i.e. dFLUX) values presented in the Figures. Multiple test-corrected two sample Kolmogorov-Smirnov tests are used to determine the statistical significance of local site-wise differences in the *rmsf* distributions in viral bound and unbound ACE2 models. The test was applied independently to each amino acid site. The Benjamini-Hochberg method was used to all the adjust p-values for the false discovery rate generated from the multiple sites of a given protein structure.

DATA AVAILABILITY STATEMENT

Structural models, images, movies and data sets for the post-processed molecular dynamics presented in this work are deposited at Zenodo CERN under the DOI 10.5281/zenodo.4421581 https://zenodo.org/record/4421581

A movie file comparing SARS-CoV-2 RBD interaction with ACE2, in both wildtype and N501Y mutant strains is also available at https://youtu.be/ptn1_BBJi70

SUPPORTING MATERIAL

The main repository for DROIDS 3.0 and maxDemon 1.0 can be found at the GitHub repository link below. Please follow the link to "Releases" and download the latest release as a .tar.gz or .zip

https://github.com/gbabbitt/DROIDS-3.0-comparative-protein-dynamics

A website for DROIDS is also available here.

https://people.rit.edu/gabsbi/

AUTHOR CONTRIBUTIONS

PR, FC and GAB designed and executed the biological study. GAB designed the statistical methods and wrote computer code to implement the methods. PR, GAB, FC, AOH, and MLL interpreted the results and wrote the manuscript.

ACKNOWLEDGEMENTS

We acknowledge funding support from the National Science Foundation (NSF Award Number 2029885) and the National Institutes of Health (NIH grant GM116102). We also acknowledge the Nvidia Corporation for a hardware support grant.

FIGURE LEGENDS

Figure 1. (A) An overview of the open conformation SARS-CoV-2 spike glycoprotein (PDB: 6vyb) interaction with its human target cell receptor, angiotensin 1 converting enzyme 2 (ACE2 PDB: 6m17). The green box in the center shows our modeling region that bounds the subsequent molecular dynamics simulations in the study. (B) The modeling region of the viral receptor binding domain (RBD) in red and ACE2 target protein in blue is shown along with the most relevant features of the binding interface; the N-terminal helices of ACE2 shown in green and the three main sites of the RBD N501 site interactions shown in magenta. (C) An overview of the DROIDS comparative molecular dynamic pipeline for binding signature analysis used throughout this study. The signed/symmetric KL divergence (or relative entropy) in the distribution of viral bound vs. unbound ACE2 atom fluctuations was collected over 200

production runs and is used to measure the difference in motion for each single amino acid's backbone (i.e. N, C, Cα, and O) upon binding by the virus.

Figure 2. DROIDS binding signature of dampened atom fluctuations in human ACE2 receptor protein upon interaction with both SARS-CoV-2 spike glycoprotein (modeled from PDB: 6m17) and the upon interaction with the past human outbreak strain SARS-CoV-1 spike glycoprotein (modeled from PDB: 6acg). Shift in atom fluctuations were calculated as the signed symmetric Kullback-Leibler (KL) divergence (i.e. relative entropy or dFLUX) of distributions of the root mean square time deviations (rmsf) for residue averaged protein backbone atoms (i.e. N, Cα, C and O) for ACE2 in its spike bound (PDB: 6m17/6acg) versus unbound dynamic state (PDB: 6m17/6acg without the viral receptor binding domain or RBD). Here we show color mapping (A, B, D) and sequence positional plotting (C, E) of dampening of atom motion on viral RBD-protein target interface in blue for (A-C) the original strain of SARS-CoV-2 and (D-E) its predecessor, SARS-CoV-1 . The sequence profile of the KL divergence between SARS-CoV-2 viral bound and unbound ACE2 produces strong negative peaks indicating key residue binding interactions with the N-terminal helices (white) and the N501 interactions at Q325, K353 and the AAQPLL 386-392 motif (yellow). The interactions at these sites are more moderated in the SARS-CoV-1 interaction with the ACE2 target protein.

Figure 3. DROIDS binding signature of dampened atom fluctuations in human ACE2 receptor protein upon interaction with both the highly transmissible B.1.1.7 (UK) variant and the B.1.351 (South African and Brazilian) variant of SARS-CoV-2 spike glycoprotein (modeled from PDB: 6m17). Shift in atom fluctuations were calculated as the signed symmetric Kullback-Leibler (KL) divergence (i.e. relative entropy or dFLUX) of distributions of the root mean square time deviations (rmsf) for residue averaged protein backbone atoms (i.e. N, Cα, C and O) for ACE2 in its spike bound (PDB: 6m17) versus unbound dynamic state (PDB: 6m17 without the viral receptor binding domain or RBD). Here we show color mapping (A, B, D) and sequence positional plotting (C, E) of dampening of atom motion on viral RBD-protein target interface in blue for (A-C) the B.1.1.7 variant and (D-E) the B.1.351 variant of SARS-CoV-2. The sequence profile of the KL divergence between viral bound and unbound ACE2 produces negative peaks indicating key residue binding interactions with the N-terminal helices (white) and the N501 interactions at Q325, K353 and the AAQPLL 386-392 motif (yellow). The N501Y, E484K and K417N/T mutations are marked by the green stars (2B and 2D).

Figure 4. DROIDS binding signature of dampened atom fluctuations in human CD26 and ACE2 receptor proteins upon interaction with the bat progenitor strain batCoV-HKU4 glycoprotein (modeled from PDB: 4qzv and 6m17). Here we show color mapping (A, B, D) and sequence positional plotting (C, E) of dampening of atom motion on the viral RBD-protein target interface

in blue for (A-C) its known targeting of CD26 and (A,D-E) its hypothetical targeting of ACE2. The sequence profile of the KL divergence between viral bound and unbound target proteins produces two strong negative peaks indicating key residue binding interactions with the (C) the SAMLI and double serine motifs (shown in yellow) on CD26 and (E) the N-terminal helices on ACE2.

Figure 5. DROIDS binding signature of dampened atom fluctuations in human ACE2 receptor proteins upon interaction with two human commensal strains hCoV-OC43 and hCoV-HKU1 glycoprotein (modeled from PDB: 6ohw, 5gnb and 6m17). Here we show color mapping (A, B, D) and sequence positional plotting (C, E) of dampening of atom motion on the viral RBD-protein target interface in blue for (A-C) the targeting of ACE2 by the most benign strain hCoV-OC43 and (D-E) its less benign counterpart hCoV-HKU1. The sequence profile of the KL divergence between viral bound and unbound target proteins produces strong negative peak indicating key residue binding interactions (C, E) with the N-terminal helices on ACE2 and (E) moderate interactions with K353 and Q325 observed only in hCoV-HKU1.

Figure 6. DROIDS binding signature of dampened atom fluctuations in human ACE2 receptor proteins upon interaction with the past human outbreak strain MERS-CoV spike glycoprotein (modeled from PDB: 4kr0, 5x5c, and 6m17). Here we show color mapping (A, B, D) and sequence positional plotting (C, E) of dampening of atom motion on the viral RBD-protein target interface in blue for (A-C) the targeting of CD26 by the MERS-CoV and (D-E) the hypothetical targeting of ACE2 by MERS-CoV. The sequence profile of the KL divergence between viral bound and unbound target proteins produces strong negative peaks indicating key residue binding interactions with (C) the SAMLI 291-5 and SS333-4 motifs on CD26 and € with the N-terminal helices on ACE2. This is only a very weak interaction with K353 (yellow) on the ACE2.

Figure 7. A functional evolutionary model of SARS-type viral protein binding interactions during zoonotic spillover. (A) The two touch model consists of a unique combination of universally evolutionary conserved ACE2 interaction at the N-terminal helices likely preadapted from its interaction with the SAMLI motif on CD26 (shown with black arrow) and a much more promiscuous, dynamically complex , and evolvable interaction with K353 and surrounding sites (shown with red arrow) . (B) Given the conserved nature of binding of CD26 at SAMLI and ACE2 at the N-terminal helices of ACE2 by the same region of the viral receptor binding domain (RBD) across the large diversity of strains we surveyed, we hypothesize this region of the virus is likely responsible for instances of zoonotic spillover to humans. The more pathogenic SARS-like outbreaks associated with ACE2 seem to develop transient dynamics of N501 that involve

promiscuous interactions at three closely located sites (K353 and the two flanking sites Q325 and AAQPFLL 386-92). This creates a 'slippery two touch interaction with the ACE2 target. In the much more transmissible new strain, N501Y (lineage B.1.1.7), the promiscuous interactions are mostly lost, favoring a much more stable binding interaction with K353 alone.

Table 1. **Hybrid models comparatively analyzed in the course of this study**

Effect of…on MD	**bound structure**	**unbound structure**	**PDB ID's utilized to create model**	**Figure panel(s)**
viral binding	SARS-CoV-2 RBD (wildtype and N501Y) and ACE2	ACE2	6m17	Figure 2 A-E
viral binding	SARS-CoV-1 RBD and ACE2	ACE2	6acg	Figure 5 A-C
viral binding	hCoV-HKU1 RBD and ACE2	ACE2	5gnb and 6m17	Figure 4D-E
viral binding	hCoV-OC43 RBD and ACE2	ACE2	6ohw and 6m17	Figure 4 A-C
viral binding	bat CoV-HKU4 and CD26	CD26	4qzv	Figure 3 A-C
viral binding	bat CoV-HKU4 and ACE2	ACE2	4qzv and 6m17	Figure 3 D-E
viral binding	MERS and CD26	CD26	5x5c and 4qzv	5D and S3C

Table 2. ***In silico*** **targeted mutagenesis studies of the SARS-CoV-2 receptor-spike interaction with various amino acid mutations**. Mutations were conducted with the *swapaa* command in UCSF Chimera, and bound structures were minimized with 2000 steps of steepest descent. D (blue) signifies dampened molecular motion, NC (yellow) indicates no significant change in molecular motion as a result of binding, and A (orange) indicates an amplification in molecular motion after RBD binding. Significance criteria was determined as having <= -1 KL divergence or less (dampened motion during binding), or >=1 KL divergence or more (amplified motion during binding). N-terminal helix and AAQPFLL motif categorization was determined as dampened or amplified when most of their constituent amino acids had a KL divergence of <= -1 or >= 1, respectively.

Protein	*In silico* Mutation	N-terminal Helix	Q325	K353	386-AAQPFLL-392
	Wild Type	D	NC	D	D
ACE2	**K353A**	NC	NC	D	D
ACE2	**386-AA/EE-387**	D	D	D	D
ACE2	**390-FLL/EEE-392**	D	NC	D	D
ACE2	**V739E**	NC	NC	D	NC
ACE2	**V739L**	D	D	D	NC
ACE2	**K408A**	D	NC	D	D
RBD	**V185E**	NC	NC	D	NC
RBD	**V185L**	D	D	D	NC

REFERENCES

Ahamad S, Kanipakam H, Gupta D. 2020. Insights into the structural and dynamical changes of spike glycoprotein mutations associated with SARS-CoV-2 host receptor binding. *J. Biomol. Struct. Dyn.* 0:1–13.

Al-Khafaji K, AL-Duhaidahawi D, Tok TT. 2020. Using integrated computational approaches to identify safe and rapid treatment for SARS-CoV-2. *J. Biomol. Struct. Dyn.* 0:1–9.

Andersen HC. 1980. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* 72:2384–2393.

Babbitt GA, Fokoue EP, Evans JR, Diller KI, Adams LE. 2020. DROIDS 3.0-Detecting Genetic and Drug Class Variant Impact on Conserved Protein Binding Dynamics. *Biophys. J.* 118:541–551.

Babbitt GA, Mortensen JS, Coppola EE, Adams LE, Liao JK. 2018. DROIDS 1.20: A GUI-Based Pipeline for GPU-Accelerated Comparative Protein Dynamics. *Biophys. J.* 114:1009–1017.

Bagdonaite I, Wandall HH. 2018. Global aspects of viral glycosylation. *Glycobiology* 28:443–467.

Cantuti-Castelvetri L, Ojha R, Pedro LD, Djannatian M, Franz J, Kuivanen S, Meer F van der, Kallio K, Kaya T, Anastasina M, et al. 2020. Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity. *Science* [Internet]. Available from: https://science.sciencemag.org/content/early/2020/10/19/science.abd2985

Casalino L, Gaieb Z, Goldsmith JA, Hjorth CK, Dommer AC, Harbison AM, Fogarty CA, Barros EP, Taylor BC, McLellan JS, et al. 2020. Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent. Sci.* 6:1722–1734.

Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ. 2005. The Amber biomolecular simulation programs. *J. Comput. Chem.* 26:1668–1688.

Daly JL, Simonetti B, Klein K, Chen K-E, Williamson MK, Antón-Plágaro C, Shoemark DK, Simón-Gracia L, Bauer M, Hollandi R, et al. 2020. Neuropilin-1 is a host factor for SARS-CoV-2 infection. *Science* [Internet]. Available from: https://science.sciencemag.org/content/early/2020/10/19/science.abd3072

Darden T, York D, Pedersen L. 1993. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.

Ewald PP. 1921. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.* 369:253–287.

Fiser A, Do RK, Sali A. 2000. Modeling of loops in protein structures. *Protein Sci. Publ. Protein Soc.* 9:1753–1773.

Follis KE, York J, Nunberg JH. 2006. Furin cleavage of the SARS coronavirus spike glycoprotein enhances cell-cell fusion but does not affect virion entry. *Virology* 350:358–369.

Fung TS, Liu DX. 2019. Human Coronavirus: Host-Pathogen Interaction. *Annu. Rev. Microbiol.* 73:529–557.

Götz AW, Williamson MJ, Xu D, Poole D, Le Grand S, Walker RC. 2012. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* 8:1542–1555.

Haan CAM de, Vennema H, Rottier PJM. 2000. Assembly of the Coronavirus Envelope: Homotypic Interactions between the M Proteins. *J. Virol.* 74:4967–4978.

Hoffmann M, Hofmann-Winkler H, Pöhlmann S. 2018. Priming Time: How Cellular Proteases Arm Coronavirus Spike Proteins. *Act. Viruses Host Proteases*:71–98.

Hoffmann M, Kleine-Weber H, Pöhlmann S. 2020. A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. *Mol. Cell* 78:779-784.e5.

Hofmann H, Pyrc K, van der Hoek L, Geier M, Berkhout B, Pöhlmann S. 2005. Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry. *Proc. Natl. Acad. Sci. U. S. A.* 102:7988–7993.

Ke Z, Oton J, Qu K, Cortese M, Zila V, McKeane L, Nakane T, Zivanov J, Neufeldt CJ, Cerikan B, et al. 2020. Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature*:1–5.

Kuba K, Imai Y, Rao S, Gao H, Guo F, Guan B, Huan Y, Yang P, Zhang Y, Deng W, et al. 2005. A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus–induced lung injury. *Nat. Med.* 11:875–879.

Li W, Zhang C, Sui J, Kuhn JH, Moore MJ, Luo S, Wong S, Huang I, Xu K, Vasilieva N, et al. 2005. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J.* 24:1634–1643.

Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. 2015. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* 11:3696–3713.

Mao Y, Zhang Y. 2012. Thermal conductivity, shear viscosity and specific heat of rigid water models. *Chem. Phys. Lett.* 542:37–41.

Mittal A, Manjunath K, Ranjan RK, Kaushik S, Kumar S, Verma V. 2020. COVID-19 pandemic: Insights into structure, function, and hACE2 receptor recognition by SARS-CoV-2. *PLoS Pathog.* [Internet] 16. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7444525/

Muralidharan N, Sakthivel R, Velmurugan D, Gromiha MM. 2020. Computational studies of drug repurposing and synergism of lopinavir, oseltamivir and ritonavir binding with SARS-CoV-2 protease against COVID-19. *J. Biomol. Struct. Dyn.* 0:1–6.

Ou X, Guan H, Qin B, Mu Z, Wojdyla JA, Wang M, Dominguez SR, Qian Z, Cui S. 2017. Crystal structure of the receptor binding domain of the spike glycoprotein of human betacoronavirus HKU1. *Nat. Commun.* 8:15216.

Ou X, Liu Y, Lei X, Li P, Mi D, Ren L, Guo L, Guo R, Chen T, Hu J, et al. 2020. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat. Commun.* 11:1620.

Oudit GY, Crackower MA, Backx PH, Penninger JM. 2003. The role of ACE2 in cardiovascular physiology. *Trends Cardiovasc. Med.* 13:93–101.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25:1605–1612.

Pierce LCT, Salomon-Ferrer R, Augusto F. de Oliveira C, McCammon JA, Walker RC. 2012. Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *J. Chem. Theory Comput.* 8:2997–3002.

Pierri CL. 2020. SARS-CoV-2 spike protein: flexibility as a new target for fighting infection. *Signal Transduct. Target. Ther.* 5:1–3.

Roe DR, Cheatham TE. 2013. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* 9:3084–3095.

Sali A, Blundell TL. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234:779–815.

Salomon-Ferrer R, Götz AW, Poole D, Le Grand S, Walker RC. 2013. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* 9:3878–3888.

Shang J, Wan Y, Luo C, Ye G, Geng Q, Auerbach A, Li F. 2020. Cell entry mechanisms of SARS-CoV-2. *Proc. Natl. Acad. Sci.* 117:11727–11734.

Song W, Gui M, Wang X, Xiang Y. 2018. Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoS Pathog.* 14:e1007236.

Tortorici MA, Walls AC, Lang Y, Wang C, Li Z, Koerhuis D, Boons G-J, Bosch B-J, Rey FA, de Groot RJ, et al. 2019. Structural basis for human coronavirus attachment to sialic acid receptors. *Nat. Struct. Mol. Biol.* 26:481–489.

Turoňová B, Sikora M, Schürmann C, Hagen WJH, Welsch S, Blanc FEC, Bülow S von, Gecht M, Bagola K, Hörner C, et al. 2020. In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science* 370:203–208.

Wang Q, Qi J, Yuan Y, Xuan Y, Han P, Wan Y, Ji W, Li Y, Wu Y, Wang J, et al. 2014. Bat origins of MERS-CoV supported by bat coronavirus HKU4 usage of human receptor CD26. *Cell Host Microbe* 16:328–337.

Wang Y, Liu M, Gao J. 2020. Enhanced receptor binding of SARS-CoV-2 through networks of hydrogen-bonding and hydrophobic interactions. *Proc. Natl. Acad. Sci.* 117:13967–13974.

Watanabe Y, Allen JD, Wrapp D, McLellan JS, Crispin M. 2020. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* 369:330–333.

Watanabe Y, Berndsen ZT, Raghwani J, Seabright GE, Allen JD, Pybus OG, McLellan JS, Wilson IA, Bowden TA, Ward AB, et al. 2020. Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. *Nat. Commun.* 11:2688.

Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. 2020. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* 367:1444–1448.

Yao H, Song Y, Chen Y, Wu N, Xu J, Sun C, Zhang J, Weng T, Zhang Z, Wu Z, et al. 2020. Molecular Architecture of the SARS-CoV-2 Virus. *Cell* 183:730-738.e13.

Yu J, Qiao S, Guo R, Wang X. 2020. Cryo-EM structures of HKU2 and SADS-CoV spike glycoproteins provide insights into coronavirus evolution. *Nat. Commun.* 11:3070.

Yuan Y, Cao D, Zhang Y, Ma J, Qi J, Wang Q, Lu G, Wu Y, Yan J, Shi Y, et al. 2017. Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nat. Commun.* 8:15092.

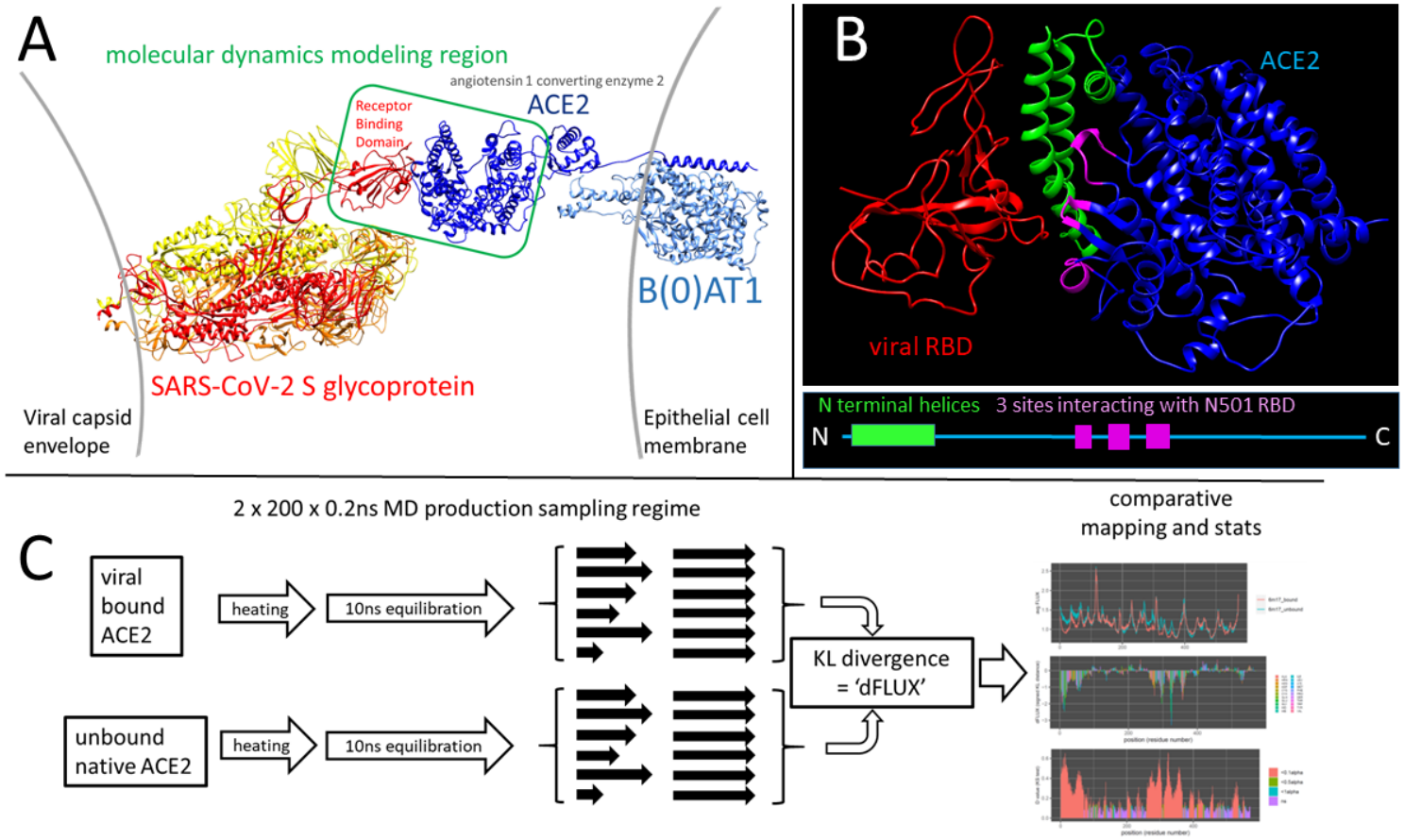

Figure 1. (A) An overview of the open conformation SARS-CoV-2 spike glycoprotein (PDB: 6vyb) interaction with its human target cell receptor, angiotensin 1 converting enzyme 2 (ACE2 PDB: 6m17). The green box in the center shows our modeling region that bounds the subsequent molecular dynamics simulations in the study. (B) The modeling region of the viral receptor binding domain (RBD) in red and ACE2 target protein in blue is shown along with the most relevant features of the binding interface; the N-terminal helices of ACE2 shown in green and the three main sites of the RBD N501 site interactions shown in magenta. (C) An overview of the DROIDS comparative molecular dynamic pipeline for binding signature analysis used throughout this study. The signed/symmetric KL divergence (or relative entropy) in the distribution of viral bound vs. unbound ACE2 atom fluctuations was collected over 200 production runs and is used to measure the difference in motion for each single amino acid's backbone (i.e. N, C, Cα, and O) upon binding by the virus.

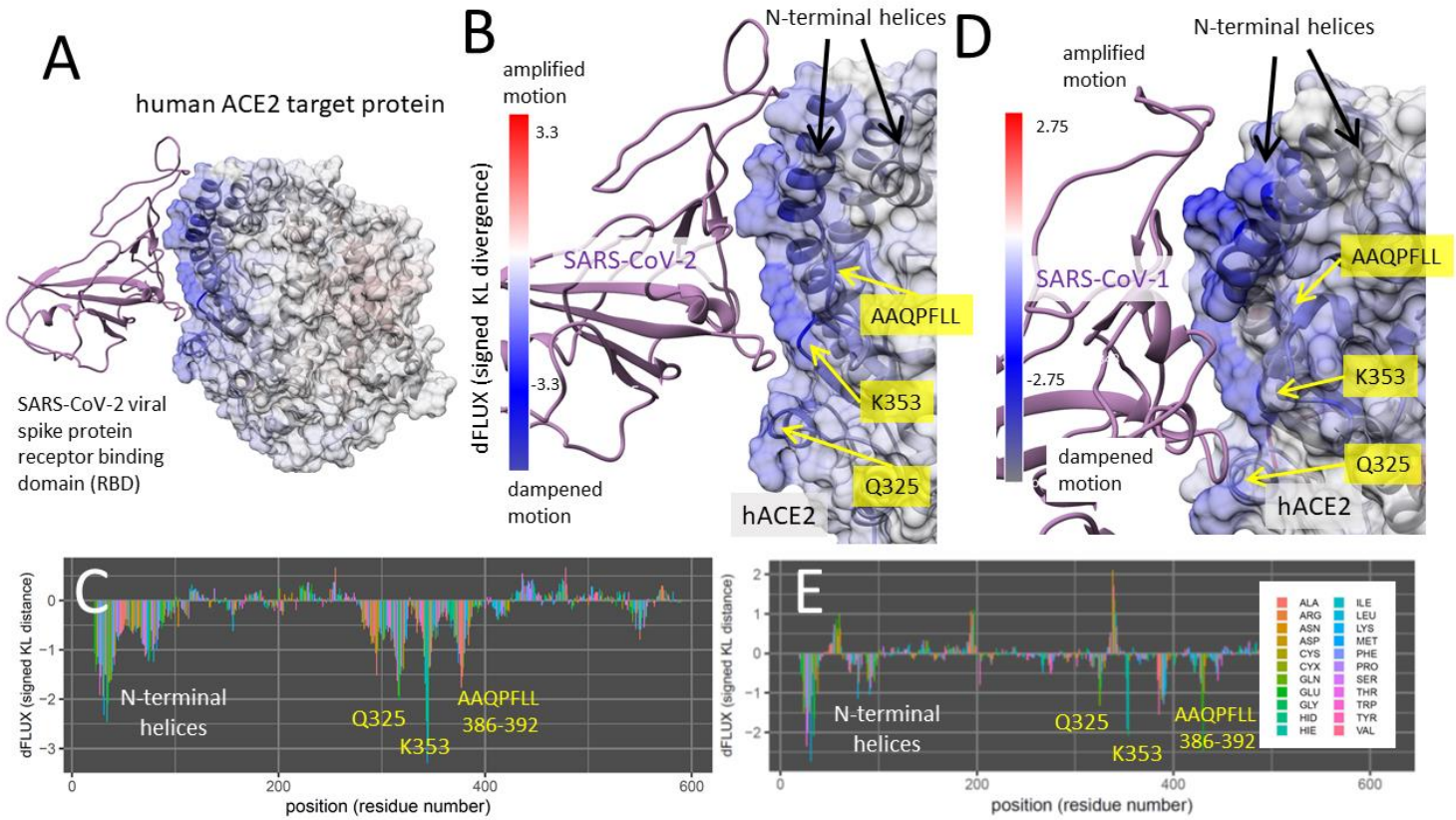

Figure 2. DROIDS binding signature of dampened atom fluctuations in human ACE2 receptor protein upon interaction with both SARS-CoV-2 spike glycoprotein (modeled from PDB: 6m17) and the upon interaction with the past human outbreak strain SARS-CoV-1 spike glycoprotein (modeled from PDB: 6acg). Shift in atom fluctuations were calculated as the signed symmetric Kullback-Leibler (KL) divergence (i.e. relative entropy or dFLUX) of distributions of the root mean square time deviations (rmsf) for residue averaged protein backbone atoms (i.e. N, Cα, C and O) for ACE2 in its spike bound (PDB: 6m17/6acg) versus unbound dynamic state (PDB: 6m17/6acg without the viral receptor binding domain or RBD). Here we show color mapping (A, B, D) and sequence positional plotting (C, E) of dampening of atom motion on viral RBD-protein target interface in blue for (A-C) the original strain of SARS-CoV-2 and (D-E) its predecessor, SARS-CoV-1 . The sequence profile of the KL divergence between SARS-CoV-2 viral bound and unbound ACE2 produces strong negative peaks indicating key residue binding interactions with the N-terminal helices (white) and the N501 interactions at Q325, K353 and the AAQPLL 386-392 motif (yellow). The interactions at these sites are more moderated in the SARS-CoV-1 interaction with the ACE2 target protein.

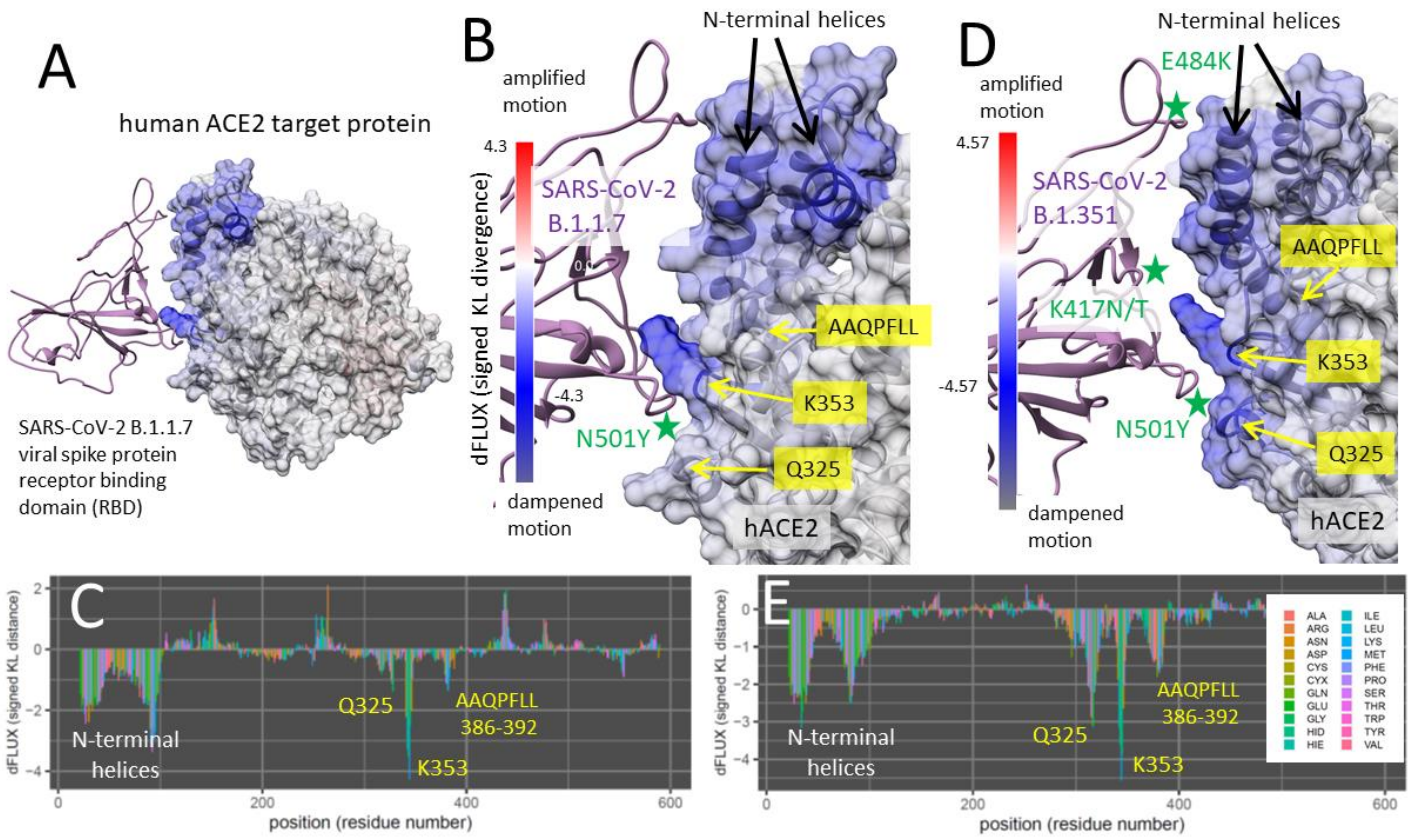

Figure 3. DROIDS binding signature of dampened atom fluctuations in human ACE2 receptor protein upon interaction with both the highly transmissible B.1.1.7 (UK) variant and the B.1.351 (South African and Brazilian) variant of SARS-CoV-2 spike glycoprotein (modeled from PDB: 6m17). Shift in atom fluctuations were calculated as the signed symmetric Kullback-Leibler (KL) divergence (i.e. relative entropy or dFLUX) of distributions of the root mean square time deviations (rmsf) for residue averaged protein backbone atoms (i.e. N, Cα, C and O) for ACE2 in its spike bound (PDB: 6m17) versus unbound dynamic state (PDB: 6m17 without the viral receptor binding domain or RBD). Here we show color mapping (A, B, D) and sequence positional plotting (C, E) of dampening of atom motion on viral RBD-protein target interface in blue for (A-C) the B.1.1.7 variant and (D-E) the B.1.351 variant of SARS-CoV-2. The sequence profile of the KL divergence between viral bound and unbound ACE2 produces negative peaks indicating key residue binding interactions with the N-terminal helices (white) and the N501 interactions at Q325, K353 and the AAQPLL 386-392 motif (yellow). The N501Y, E484K and K417N/T mutations are marked by the green stars (2B and 2D).

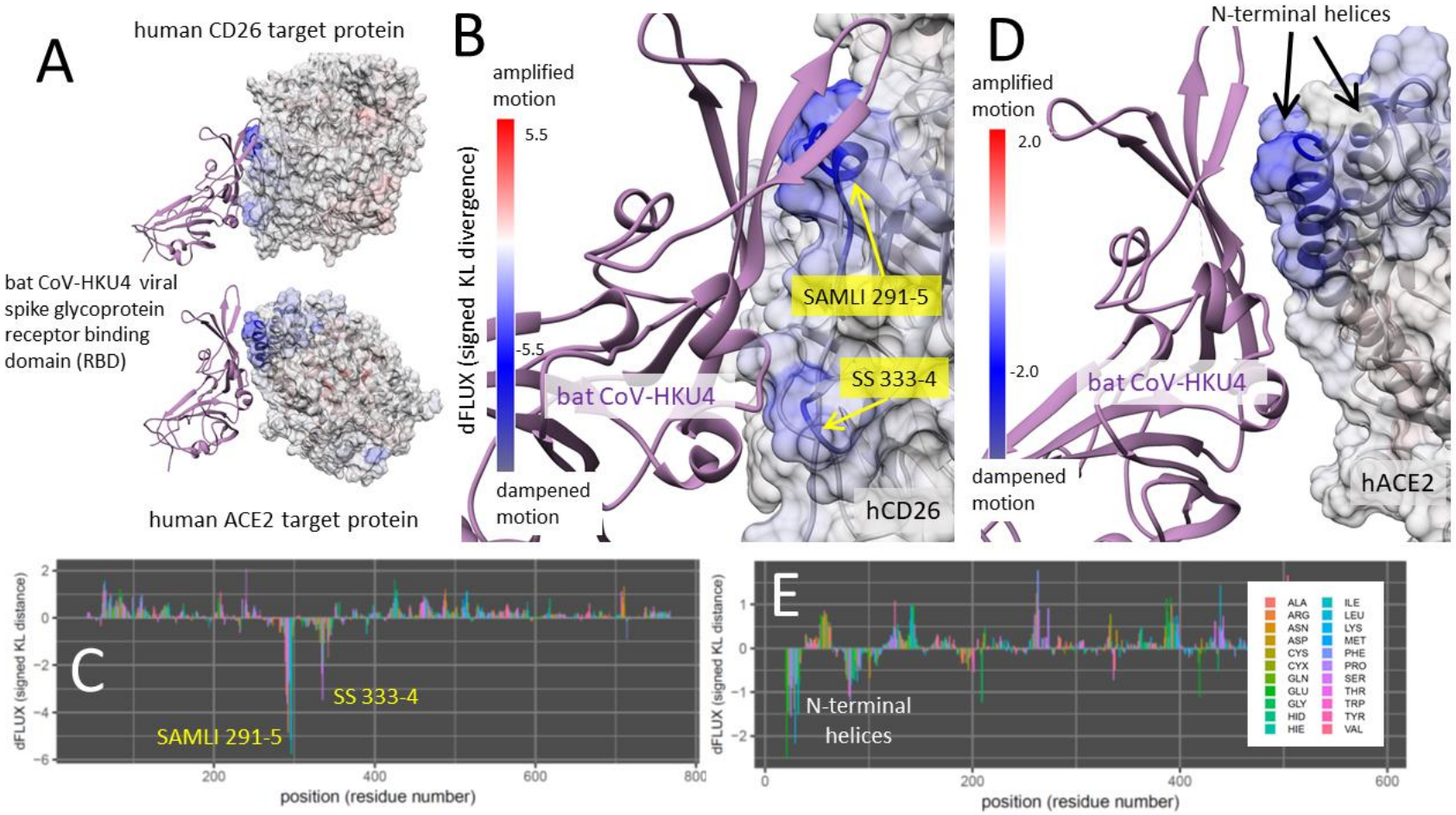

Figure 4. DROIDS binding signature of dampened atom fluctuations in human CD26 and ACE2 receptor proteins upon interaction with the bat progenitor strain batCoV-HKU4 glycoprotein (modeled from PDB: 4qzv and 6m17). Here we show color mapping (A, B, D) and sequence positional plotting (C, E) of dampening of atom motion on the viral RBD-protein target interface in blue for (A-C) its known targeting of CD26 and (A,D-E) its hypothetical targeting of ACE2. The sequence profile of the KL divergence between viral bound and unbound target proteins produces two strong negative peaks indicating key residue binding interactions with the (C) the SAMLI and double serine motifs (shown in yellow) on CD26 and (E) the N-terminal helices on ACE2.

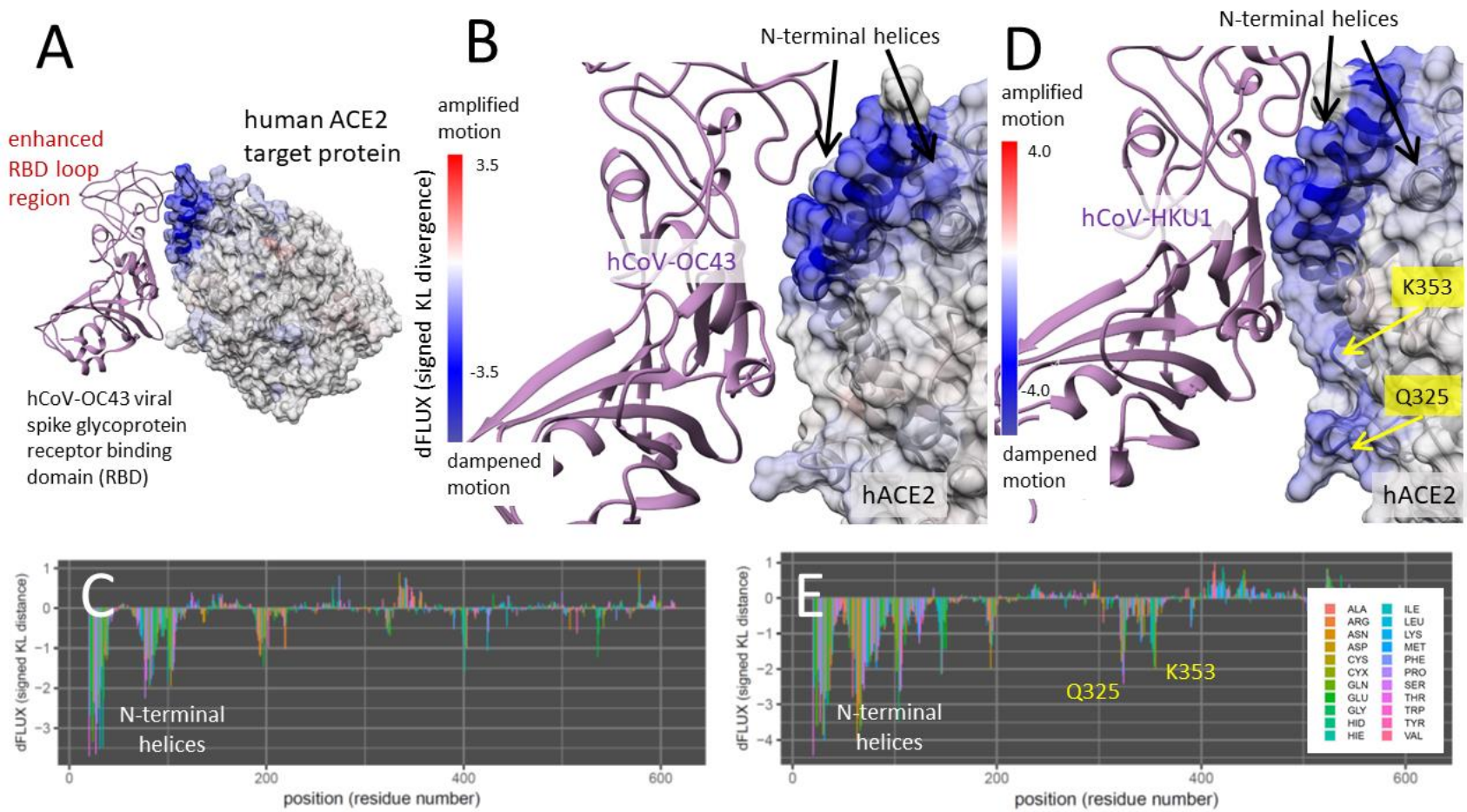

Figure 5. DROIDS binding signature of dampened atom fluctuations in human ACE2 receptor proteins upon interaction with two human commensal strains hCoV-OC43 and hCoV-HKU1 glycoprotein (modeled from PDB: 6ohw, 5gnb and 6m17). Here we show color mapping (A, B, D) and sequence positional plotting (C, E) of dampening of atom motion on the viral RBD-protein target interface in blue for (A-C) the targeting of ACE2 by the most benign strain hCoV-OC43 and (D-E) its less benign counterpart hCoV-HKU1. The sequence profile of the KL divergence between viral bound and unbound target proteins produces strong negative peak indicating key residue binding interactions (C, E) with the N-terminal helices on ACE2 and (E) moderate interactions with K353 and Q325 observed only in hCoV-HKU1.

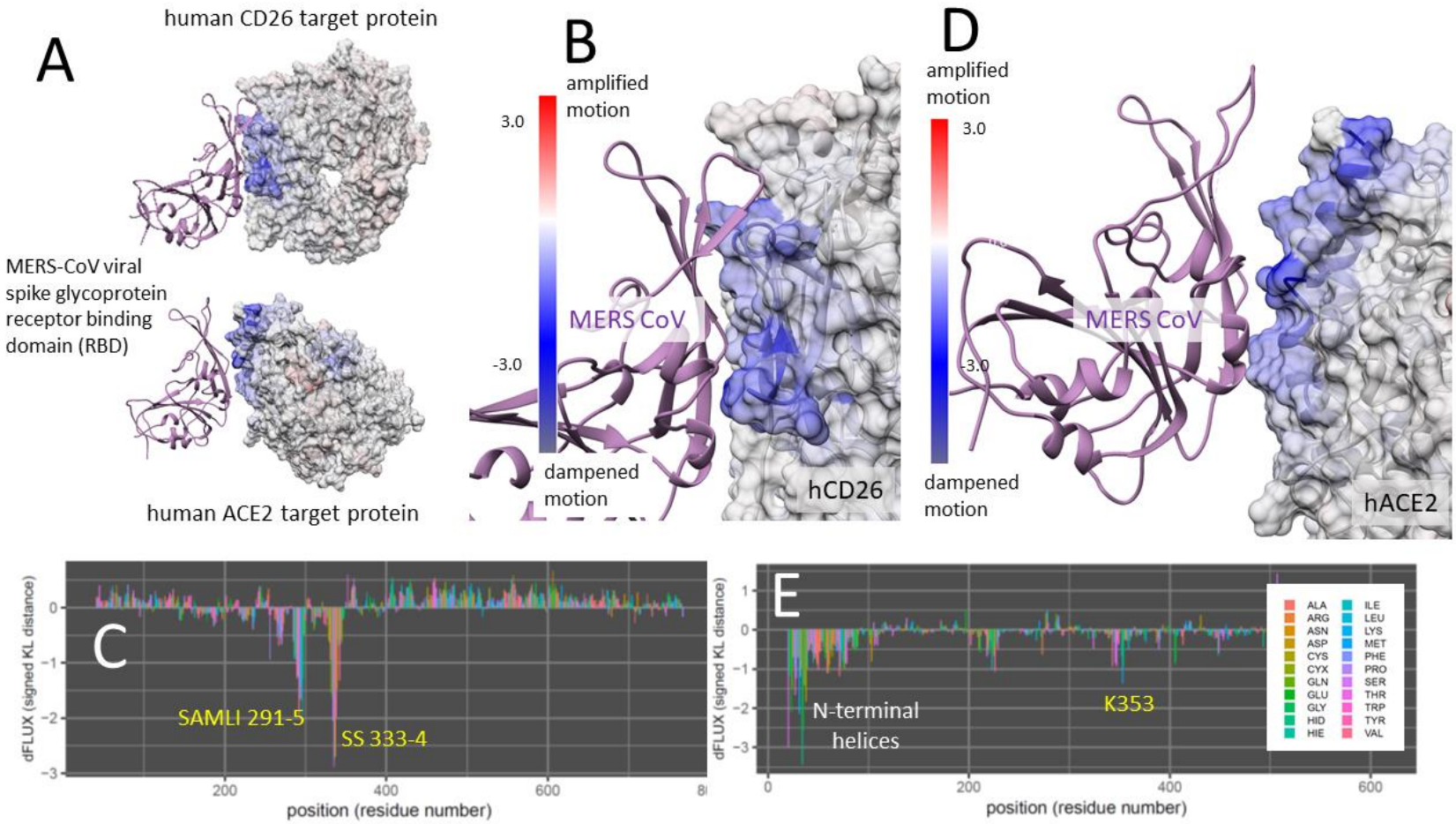

Figure 6. DROIDS binding signature of dampened atom fluctuations in human ACE2 receptor proteins upon interaction with the past human outbreak strain MERS-CoV spike glycoprotein (modeled from PDB: 4kr0, 5x5c, and 6m17). Here we show color mapping (A, B, D) and sequence positional plotting (C, E) of dampening of atom motion on the viral RBD-protein target interface in blue for (A-C) the targeting of CD26 by the MERS-CoV and (D-E) the hypothetical targeting of ACE2 by MERS-CoV. The sequence profile of the KL divergence between viral bound and unbound target proteins produces strong negative peaks indicating key residue binding interactions with (C) the SAMLI 291-5 and SS333-4 motifs on CD26 and € with the N-terminal helices on ACE2. This is only a very weak interaction with K353 (yellow) on the ACE2.

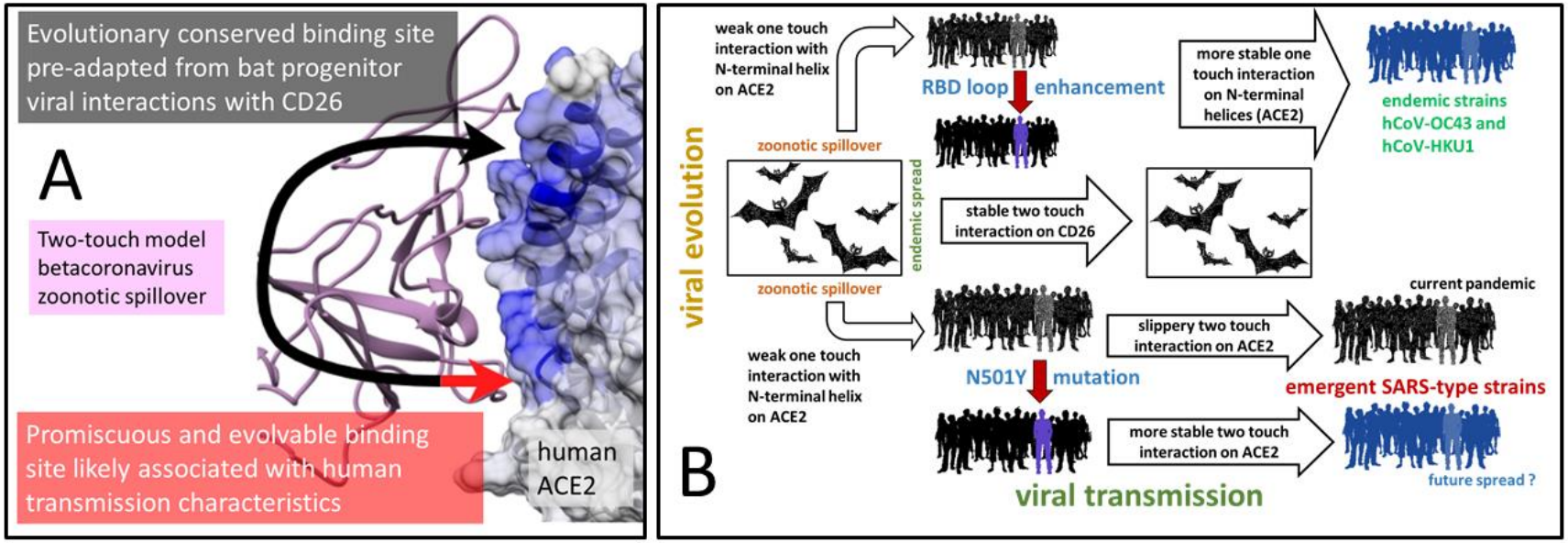

Figure 7. A functional evolutionary model of SARS-type viral protein binding interactions during zoonotic spillover. (A) The two touch model consists of a unique combination of universally evolutionary conserved ACE2 interaction at the N-terminal helices likely preadapted from its interaction with the SAMLI motif on CD26 (shown with black arrow) and a much more promiscuous, dynamically complex, and evolvable interaction with K353 and surrounding sites (shown with red arrow) . (B) Given the conserved nature of binding of CD26 at SAMLI and ACE2 at the N-terminal helices of ACE2 by the same region of the viral receptor binding domain (RBD) across the large diversity of strains we surveyed, we hypothesize this region of the virus is likely responsible for instances of zoonotic spillover to humans. The more pathogenic SARS-like outbreaks associated with ACE2 seem to develop transient dynamics of N501 that involve promiscuous interactions at three closely located sites (K353 and the two flanking sites Q325 and AAQPFLL 386-92). This creates a 'slippery two touch interaction with the ACE2 target. In the much more transmissible new strain, N501Y (lineage B.1.1.7), the promiscuous interactions are mostly lost, favoring a much more stable binding interaction with K353 alone.