A comparative survey of betacoronavirus strain molecular dynamics identifies key ACE2 binding sites

P. X. Rynkiewicz[1], G. A. Babbitt[1*], F. Cui[1], A. O. Hudson[1], M. L. Lynch[2]

Running title – coronavirus binding molecular dynamics

ABSTRACT

Comparative functional analysis of the binding interactions between various betacoronavirus strains and their potential human target proteins, such as ACE1, ACE2 and CD26, is critical to our future understanding and combating of COVID-19. Here, employing large replicate sets of GPU accelerated molecular dynamics simulations, we statistically compare atom fluctuations of the known human target proteins in both the presence and absence of different strains of the viral receptor binding domain (RBD) of the S spike glycoprotein. We identify a common interaction site between the N-terminal helices of ACE2 and the viral RBD in all strains (hCoV-OC43, hCoV-HKU1, MERS-CoV, SARS-CoV1, and SARS-CoV-2) and a second more dynamically complex RBD interaction site involving the ACE2 amino acid sites K353, Q325, and a novel motif, AAQPFLL (386-392) in the more recent cross-species spillovers (i.e. absent in hCoV-OC43). We use computational mutagenesis to further confirm the functional relevance of these sites. We propose a “one touch/two touch” model of viral evolution potentially involved in functionally facilitating binding interactions in zoonotic spillovers. We also observe these two touch sites governing RBD binding activity in simulations on hybrid models of the suspected viral progenitor, batCoV-HKU4, interacting with both the human SARS target, ACE2, and the human MERS target, CD26. Lastly, we confirm that the presence of a common hypertension drug (lisinopril) within the target site of SARS-CoV-2 bound models of ACE1 and ACE2 acts to enhance the RBD interactions at the same key sites in our proposed model. In the near future, we recommend that our comparative computational analysis identifying these key viral RBD-ACE2 binding interactions be supplemented with comparative studies of site-directed mutagenesis in order to screen for current and future coronavirus strains at high risk of zoonotic transmission to humans.

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

STATEMENT OF SIGNIFICANCE

We generated structural models of the spike glycoprotein receptor binding domain from recent and past betacoronavirus outbreak strains aligned to the angiotensin 1 converting enzyme 2 protein, the primary target protein of the SARS-CoV-2 virus causing COVID 19. We then statistically compared computer simulated molecular dynamics of viral bound and unbound versions of each model to identify locations where interactions with each viral strain have dampened the atom fluctuations during viral binding. We demonstrate that all known strains of betacoronavirus are strongly interactive with the N-terminal helix region of ACE2. We also identify a more complex viral interaction with three novel sites that associates with more recent and deadly SARS strains, and also a bat progenitor strain HKU4.

INTRODUCTION

The SARS-CoV-2 virus belongs to the class of positive-strand RNA betacoronaviruses structurally defined by a nucleocapsid interior surrounded by the spike, membrane, and envelope proteins (1). 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 (2). Additionally, the unprecedented infectivity of SARS-CoV-2 brings into question the structural, functional, and evolutionary similarities between it and other coronavirus outbreaks, including those that have evolved into less pathologically severe strains. One prominent candidate for comparative studies of the propensity for viral transmission is the spike protein, 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 Receptor Binding Domains (RBDs) interacting directly with the host, and the S2 subunit which directly fuses the virion to the host cellular membrane (3). 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 (4, 5). The boundary between the S1 and S2 subunits forms an exposed loop in human coronaviruses SARS-CoV-2,OC43, HKU1, and MERS-CoV that is not found in SARS-CoV-1 (6). This suggests a complex evolutionary landscape for SARS-CoV-related coronaviruses, the breadth of which cannot be fully understood through amino acid or nucleotide sequence comparisons alone.

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 (7,8). 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 (9). ACE2 is classified as a zinc metallopeptidase (carboxypeptidase) involved in cleaving a single amino acid residue from angiotensin I (AngI) to generate Ang1-9 (10). In addition, ACE2 also cleaves a single amino acid residue from angiotensin II (Ang II) to generate Ang1-7 (10). 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. The comparative dynamic modeling of present, past, and potential progenitor viral strain interactions to ACE2 and other known targets of past outbreaks, such as CD26, the target of MERS-CoV, can also

potentially illuminate critical functional aspects that facilitate zoonotic spillovers involving the betacoronaviruses.

A powerful tool for *in silico* analysis of protein-protein interactions is all-atom 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 (11). 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 are increasingly broadening these methods of inquiry by allowing further analysis of shorter time scale dynamics of protein-protein interactions (12). In the SARS-CoV-2 pathosystem, MD simulations have already been employed to map the hydrogen bonding topology of the SARS-CoV-2 RBD/ACE2 interaction, explore the merits of drug repurposing, characterize mutations, and determine the stability of binding interactions between the SARS-CoV-2 RBD and several contemporary antiviral drugs (13, 14, 15, 16).

While conventional and accelerated MD 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 challenging. Given recent increases in GPU processor power, statistically comparing 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. So reliable comparisons of dynamics need to further be based upon statistical methods that can fully capture the multi-modal distributional space of protein motion and assign probability scores in the absence of any distributional model that would attempt to describe protein motion.

Here we employ DROIDS 3.0, a combined methodology and software tool for the statistical analysis and visualization of comparative protein dynamics (17,18), to analyze hybrid models of betacoronavirus spike glycoprotein target interactions. DROIDS employs relative entropy (or Kullback-Leibler divergence) calculations on atom fluctuation distributions derived from large ensembles of MD trajectories in order to quantify the local differences in both magnitude and shape atom fluctuations in two different protein functional states. (e.g. viral bound versus unbound, or before versus after mutation). DROIDS also employs a non-parametric two sample Kolmogorov-Smirnov statistical test of significance to identify where molecular dynamics are locally different. P-values from this test are adjusted for the number of sites on the protein using the Benjamini-Hochberg method for adjusting false discovery rates under multiple tests. Here we propose a MD simulation-informed comparative model for interaction between coronavirus receptor binding domains and human ACE2, and discuss MD

simulation as a useful method for protein comparison at the molecular scale. We specifically compare the motions characterizing the interactions of SARS-CoV-2 Spike protein-human ACE2 with similar events in multiple other systems, including comparable respiratory coronavirus outbreaks in humans and in a putative bat progenitor, to gain an evolutionary perspective on the molecular specifics of coronavirus viral entry into human host cells. Through comparative MD simulations, we determine the binding profiles of viral bound and unbound ACE targets. We use this method to compare the binding profiles of common human strains of betacoronavirus (OC43 and HKU1) to recent SARS-like outbreaks (i.e. SARS-CoV-1, SARS-CoV-2 and MERS), as well as the probable bat progenitor strain HKU4. We also investigate whether the presence of a common hypertension drug might act to alter molecular binding recognition by the SARS-CoV-2 spike glycoprotein through its stabilizing effect on ACE receptors.

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) (19) and SARS-CoV-2 (PDB: 6m17) (20). Hybrid models for three more human betacoronavirus variants bound to ACE2 were constructed from homologous single viral structures that were structurally aligned to the SARS-CoV-2 model using UCSF Chimera's MatchMaker superposition tool (21) followed by subsequent deletion of the homologous viral receptor binding domain and any structure outside of the molecular dynamics modeling space defined in Figure 1. These hybrid models of ACE2 interaction included viral receptor binding domains from MERS-CoV (PDB:5x5c) (22), HCoV-OC43 (PDB: 6ohw) (23), and HCoV-HKU1 (PBD: 5gnb) (24) 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 prior to molecular dynamics simulation in Amber18 (25, 26, 27). Loop modeling and refinement were conducted where needed using Modeller in UCSF Chimera (28, 29). A hybrid model of the bat HKU4 beta coronavirus interaction with ACE2 was similarly constructed using PDB: 4qzv (30), HKU4 bound to the MERS target human CD26, as a reference for structural alignment to PDB: 6m17. For site-directed mutagenesis computations, we created mutant models using the swapaa function in UCSF Chimera 1.13 (21) 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. All non-conserved amino acid replacements listed the structural alignment of PDB: 6acg to PDB: 6m17 were used to create the COVID-like SARS model examined in Figure 6. A hybrid model of lisinopril, a common hypertension drug, interacting with viral bound ACE2 (PDB: 6m17) and viral bound ACE1, the intended target of the drug (PDB: 1o86) was prepared similarly to the

hybrid models above (i.e. homologous superposition followed by energy minimization). 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.

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 (25, 27, 11) via the DROIDS v3.0 interface (Detecting Relative Outlier Impacts in Dynamic Simulation) (17,18). 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 (31). Solvation was generated using the Tip3P water model (32) in a 12nm octahedral water box and subsequent charge neutralization with Na+ and Cl- ions. 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. 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 (33). Root mean square atom fluctuations (rmsf) were calculated using the atomicfluct function in CPPTRAJ (34). Force field modifications of the general Amber force field (GAFF2) for the common hypertension drug, lisinopril, bound to ACE1 and ACE2 target proteins were prepared using the QMMM optimization program sqm in the AmberTools antechamber software suite.

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

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/OpenMM (28, 29). 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

$$rmsf = \frac{1}{4}\sum_{i=1}^{4}\sqrt{\frac{1}{n} * \sum_{j=1}^{n}\left(v_{jx} - w_x\right)^2 + \left(v_{jy} - w_y\right)^2\left(v_{jz} - w_z\right)^2} \#(1)$$

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

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

where rmsf represents the average root mean square deviation of a given atom over time t. 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.

$$KL_{divergence} = \sum_{t=50ps}^{T}\sum_{i=N,C,C_\alpha,O}^{4}\left[\left(rmsf_0 * \log\frac{rmsf_0}{rmsf_1}\right) + \left(rmsf_1 * \log\frac{rmsf_1}{rmsf_0}\right)\right]/8\Big/T \ \#(3)$$

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 3 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 site-wise differences in the *rmsf* distributions in viral bound and unbound ACE2 models. The Benjamini-Hochberg method was used to all the adjust p-values for the false discovery rate generated from a given protein structure.

The data set for this work is reposited at DOI: 10.5281/zenodo.4022863

RESULTS

Human-Pathogenic Coronavirus RBD Binding Comparison

The modeled region consisted of the extracellular domain region of human ACE2, and the viral RBD specified in Figure 1. 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-Liebler (KL) divergence on the ACE2 PDB structure mapped at single amino acid resolution (Fig. 2). With the exception of OC43, all RBDs were also associated with dampened molecular motion of hACE2 residue K353, having a signed symmetric KL divergence in rmsf distribution for residue averaged protein backbone atoms of -1 or less (Fig. 3A-C). 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. S1). A significant dampening effect is also present in the 386-AAQPFLL-392 motif when interacting with SARS-CoV-1 and SARS-CoV-2 (Fig. 3A,B), as well as the Q325 residue when interacting with SARS-CoV-1 and HKU1 (Fig. 3A,C). In an overlaid comparison between bound structures, Q325 was also visibly offset in the SARS-CoV-2 bound structure in comparison to all other human coronavirus bound structures (Fig. S2). The protein site-wise statistical significance of all four human model comparisons are also presented (Fig. S9).

Molecular Dynamics Modeling of Potential Zoonotic RBD/ACE2 Interactions

Interaction models of bat HKU4 and SARS-CoV-2 RBDs in complex with hACE2 and hCD26 resulted in a similar conformation of the RBDs and human receptors (Fig 4A). Bat coronavirus HKU4 significantly dampened molecular motion in two small regions of hCD26, 291-SAMLI-295, and 333-SS-334 (Fig 4B). Interacting with hACE2 however, HKU4 only dampened the N-terminal helices, as seen previously in all human coronavirus RBDs surveyed in this study (Fig 4C). The MERS-CoV strain also results in dampening of motion in the N-terminal helices in the presence of ACE2 (Fig 4D), as well as dampening in the K353 site seen in other human coronavirus strains (Fig. 3) (35). At the single amino acid resolution, it is notable that only K353 and no neighboring residues are significantly dampened, with a KL divergence of less than -1 (Fig. S3C). The protein site-wise statistical significance of the three zoonotic model comparisons are also presented (Fig. S9).

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

Eight total *in silico* targeted mutagenesis studies were performed, on ACE2 and the viral RBD alone (Table 1). K353A, a mutation previously identified to abolish SARS-CoV-1 RBD

interaction with ACE2 (36), neutralized dampening of the N-terminal helices during binding with the SARS-CoV-2 RBD (Fig. S4). Within the newly identified 386-AAQPFLL-392 motif, a mutation of the N-terminal double alanine 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. S5A). 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. S5B). V739 in ACE2, when mutated 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. S6A). 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. S7A). 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. S6B). Mutating V185 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. S7B). All surveyed mutations maintained significant dampening in K353 of ACE2 during binding.

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 in human coronavirus strains, potential zoonotic interactions involving the bat progenitor strain HKU4, and our results in *in silico* mutagenesis, we propose a two-touch model of protein interaction for coronavirus RBD evolution and human specificity (Fig. 5). The black arrowhead in the model signifies a promiscuous region of sustained interaction between the coronavirus RBDs and the N-terminal helix region of ACE2. The red arrowhead indicates 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 touch point of the viral RBD with the N-terminal helices of ACE1 and ACE2 that is likely present in all known strains of potentially human-interacting betacoronaviruses. The second and more complex dynamic touch point in the model involves the two known sites, Q325 and K353, as well as a novel site of interaction reported above. This second interaction site seems considerably more apparent in the zoonotic models of HKU4 and the more recent pathogenic SARS-type strains.

Effect of the Hypertension Drug Lisinopril on SARS-CoV-2 RBD Binding with ACE1/2

Molecular dynamics comparisons of the presence and absence of lisinopril, a common ACE inhibitor prescribed for hypertension, within the viral bound ACE1/2 binding pocket (Fig. 6A) demonstrated significant enhancement of viral binding within the SARS-CoV-2 RBD and

ACE1/2 interface. In ACE2, we found that lisinopril strongly and significantly dampened the atom vibrations in multiple sites in the C-proximal viral region interfacing the surface of ACE2 (Fig. 6B). The viral interaction with the third N-terminal helix in ACE2 is particularly strengthened by the presence of lisinopril in the ACE2 binding pocket (Fig. 6B). In ACE1, the intended target for lisinopril, the enhancement of viral binding in the presence of lisinopril is much more striking. Here, lisinopril was found to dampen atom fluctuation (i.e.stabilize) most of the ACE1 and SARS-CoV-2 structures (Fig. 6C). The dampening of viral motion is particularly pronounced in the single large region of the viral interface that corresponds with the second more complex touch point of our proposed model, described previously (Fig. 6C).

DISCUSSION

Here we present evidence from statistical comparisons of molecular dynamics that identifies several key residues in SARS-CoV-2 and other coronavirus strains' RBD binding with human ACE2. These residues in ACE2 are K353, Q325, and 386-AAQPFLL-392, all of which have complex dynamic interactions that dampened molecular motion in a bound state with the viral RBDs. Additionally, the prevalence of the ACE2 N-terminal helix interactions among all of our human and bat progenitor viral strain models leads us to believe that this is a necessary prerequisite interaction for all zoonotic spillovers involving coronavirus species. From these modeled phenomena we derive a model of viral evolution that describes predisposition viruses of zoonotic origin to species crossover based on an initial interaction with the ACE2 N-terminal helices, referred to as the "one-touch" model. Upon successful infection of the human host, an additional anchor point among K353, Q325, and 386-AAQPFLL-392 in ACE2 appears to be established by the RBD, resulting in the "two-touch" model more specifically seen in highly pathogenic human coronavirus species. Simulated targeted mutagenesis strongly suggests the involvement of the double alanine in the 386-AAQPFLL-392 motif in hydrophobic interactions with the SARS-CoV-2 RBD, and a valine residue present on a loop region of the SARS-CoV-2 RBD proximal to human ACE2 during binding suggests an additional potential hydrophobic interaction. In the presence of the common hypertension drug, lisinopril, bound centrally within the structure of ACE1 or ACE2, the molecular motion within the RBD-ACE2 complex appears to stabilize at many points along the viral-ACE1/2 interface, including the two key touch points in our proposed model.

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. Furthermore, our addition of lisinopril to the simulated binding interaction demonstrates that the stabilization of ACE receptors during drug binding may enhance this molecular recognition between the viral RBD and ACE receptor domain target as a result of further reducing thermal noise in the unbound ACE2 system. It is important to note that this result does not extend to any clinical conclusions, such as suggesting whether or not to take this drug. Rather, it demonstrates the significant effect of external binding molecules on the protein system and provides the impetus for further clinical research on any potential correlation between ACE inhibitor usage and the likelihood and/or severity of SARS-CoV-2 infection.

Some limitations of molecular dynamics as a probative method for functional proteomics are its implicit simplifying assumptions, complex and inherently chaotic computational nature, and high computational expense. Specifically, computational sampling of even the accelerated MD method employed here has strict hardware limitations, and even on NVidia Titan Xp 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, molecular dynamics methods involve some simplification of physics within the system being studied by ignoring atomic charge regulation, charge screening during interaction and macromolecular crowding effects. Insight into long-term dynamics and bond motion in solvent molecules are also limited by current molecular dynamics methods. Glycosylation is another aspect of coronavirus spike protein biology that is not captured by our molecular dynamics simulations, mainly due to lack of glycosylation in most of our key starting structures. In SARS-CoV-1, MERS, and HKU1 viruses, there are 8 to 9 known N-linked glycosylation amino-acid sequins which aid in immune evasion (37).

We conclude that our identification of key residues in the binding interaction between the SARS-CoV-2 RBD and human ACE2 receptor, as well as our two-touch model of RBD evolution provide a conceptual framework for future functional mutagenesis studies of this system. Future surveys of betacoronavirus strains circulating in past, present, and future human populations as well as molecular and clinical investigations of SARS-CoV-2 infection in the presence of other ACE inhibitors like lisinopril will be informed by the model interpretation of binding interaction that we present in this study. In further 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. Furthermore, simulation on supercomputing infrastructure including glycosylation chemistry in the structures will increase the accuracy and interpretability of molecular dynamics findings on this protein interaction system.

Table 1. ***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	K325	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

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 .tar.gz or .zip file

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

PXR, FC and GAB designed and executed the biological study. GAB designed the statistical methods and wrote computer code to implement the methods. PXR, 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). We also acknowledge the Nvidia corporation for a hardware support grant.

REFERENCES

1. Tan, Y.-J., Lim, S. G., and Hong, W. (2005) Characterization of viral proteins encoded by the SARS-coronavirus genome. *Antiviral Res*. 65:69–78
2. De Haan, C. A. M., Vennema, H., & Rottier, P. J. M. (2000). Assembly of the Coronavirus Envelope: Homotypic Interactions between the M Proteins. *Journal of Virology*, 74(11):4967–4978.
3. Hoffmann, M., Hofmann-Winkler, H., and Pöhlmann, S. (2018) Priming Time: How Cellular Proteases Arm Coronavirus Spike Proteins. in *Activation of Viruses by Host Proteases* (Böttcher-Friebertshäuser, E., Garten, W., and Klenk, H. D. eds), pp. 71–98, Springer International Publishing, Cham, 10.1007/978-3-319-75474-1_4
4. Shang, J., Wan, Y., Luo, C., Ye, G., Geng, Q., Auerbach, A., and Li, F. (2020) Cell entry mechanisms of SARS-CoV-2. *PNAS*. 117:11727–11734
5. Follis, K. E., York, J., and Nunberg, J. H. (2006) Furin cleavage of the SARS coronavirus spike glycoprotein enhances cell–cell fusion but does not affect virion entry. *Virology*. 350:358–369
6. Hoffmann, M., Kleine-Weber, H., and Pöhlmann, S. (2020) A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. *Molecular Cell*. 78:779-784.e5
7. Ou, X., Liu, Y., Lei, X., Li, P., Mi, D., Ren, L., Guo, L., Guo, R., Chen, T., Hu, J., Xiang, Z., Mu, Z., Chen, X., Chen, J., Hu, K., Jin, Q., Wang, J., and Qian, Z. (2020) Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat Commun*. 11:1620
8. Oudit, G. Y., Crackower, M. A., Backx, P. H., and Penninger, J. M. (2003) The role of ACE2 in cardiovascular physiology. *Trends Cardiovasc. Med.* 13:93–101
9. Hofmann, H., Pyrc, K., van der Hoek, L., Geier, M., Berkhout, B., and 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

10. Kuba, K., Imai, Y., Rao, S., Gao, H., Guo, F., Guan, B., Huan, Y., Yang, P., Zhang, Y., Deng, W., Bao, L., Zhang, B., Liu, G., Wang, Z., Chappell, M., Liu, Y., Zheng, D., Leibbrandt, A., Wada, T., Slutsky, A. S., Liu, D., Qin, C., Jiang, C., and Penninger, J. M. (2005) A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus–induced lung injury. *Nature Medicine*. 11:875–879

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

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

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

14. Muralidharan, N., R. Sakthivel, D. Velmurugan, and M.M. Gromiha. 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.* 1–6.

15. Ahamad, S., H. Kanipakam, and D. Gupta. 2020. Insights into the structural and dynamical changes of spike glycoprotein mutations associated with SARS-CoV-2 host receptor binding. *Journal of Biomolecular Structure and Dynamics*. 0:1–13.

16. Al-Khafaji, K., AL-Duhaidahawi, D., and Tok, T. T. (2020) Using integrated computational approaches to identify safe and rapid treatment for SARS-CoV-2. *Journal of Biomolecular Structure and Dynamics*. 0:1–9

17. Babbitt, G.A., Mortensen, J.S., Coppola, E.E., Adams, L.E., and Liao J.K. (2018) DROIDS 1.20: A GUI-Based Pipeline for GPU-Accelerated Comparative Protein Dynamics. *Biophys J*. 114:1009–1017

18. Babbitt, G.A., Fokoue, E.P., Evans, J.R., Diller, K.I., Adams, L.E. (2020) DROIDS 3.0-Detecting Genetic and Drug Class Variant Impact on Conserved Protein Binding Dynamics. Biophys. J. 118:541–551.

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

20. 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

21. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*. 25:1605–1612

22. Yuan, Y., Cao, D., Zhang, Y., Ma, J., Qi, J., Wang, Q., Lu, G., Wu, Y., Yan, J., Shi, Y., Zhang, X., and Gao, G. F. (2017) Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nature Communications*. 8:15092

23. Tortorici, M. A., Walls, A. C., Lang, Y., Wang, C., Li, Z., Koerhuis, D., Boons, G.-J., Bosch, B.-J., Rey, F. A., de Groot, R. J., and Veesler, D. (2019) Structural basis for human coronavirus attachment to sialic acid receptors. *Nat. Struct. Mol. Biol.* 26:481–489

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

25. Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs. *Journal of Computational Chemistry*. 26:1668–1688

26. Götz, A.W., Williamson, M.J., Xu, D., Poole, D., Le Grand, S., Walker, R.C. (2012) Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* 8:1542–1555

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

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

29. Šali, A., and Blundell, T. L. (1993) Comparative Protein Modelling by Satisfaction of Spatial Restraints. *Journal of Molecular Biology*. 234:779–815

30. Wang, Q., Qi, J., Yuan, Y., Xuan, Y., Han, P., Wan, Y., Ji, W., Li, Y., Wu, Y., Wang, J., Iwamoto, A., Woo, P.C.Y., Yuen, K.-Y., Yan, J., Lu, G., Gao, G.F. (2014) Bat origins of MERS-CoV supported by bat coronavirus HKU4 usage of human receptor CD26. *Cell Host Microbe.* 16:328–337

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

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

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

34. Roe, D.R., Cheatham, T.E., 2013. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* 9:3084–3095

35. Mohd, H.A., Al-Tawfiq, J.A., Memish, Z.A. (2016) Middle East Respiratory Syndrome Coronavirus (MERS-CoV) origin and animal reservoir. *Virology Journal.* 13:87.

36. Li, W., Zhang, C., Sui, J., Kuhn, J. H., Moore, M. J., Luo, S., Wong, S.-K., Huang, I.-C., Xu, K., Vasilieva, N., Murakami, A., He, Y., Marasco, W. A., Guan, Y., Choe, H., and Farzan, M. (2005)

Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J.* 24:1634–1643

37. Watanabe, Y., Berndsen, Z. T., Raghwani, J., Seabright, G. E., Allen, J. D., Pybus, O. G., McLellan, J. S., Wilson, I. A., Bowden, T. A., Ward, A. B., and Crispin, M. (2020) Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. *Nature Communications*. 11:2688

FIGURES

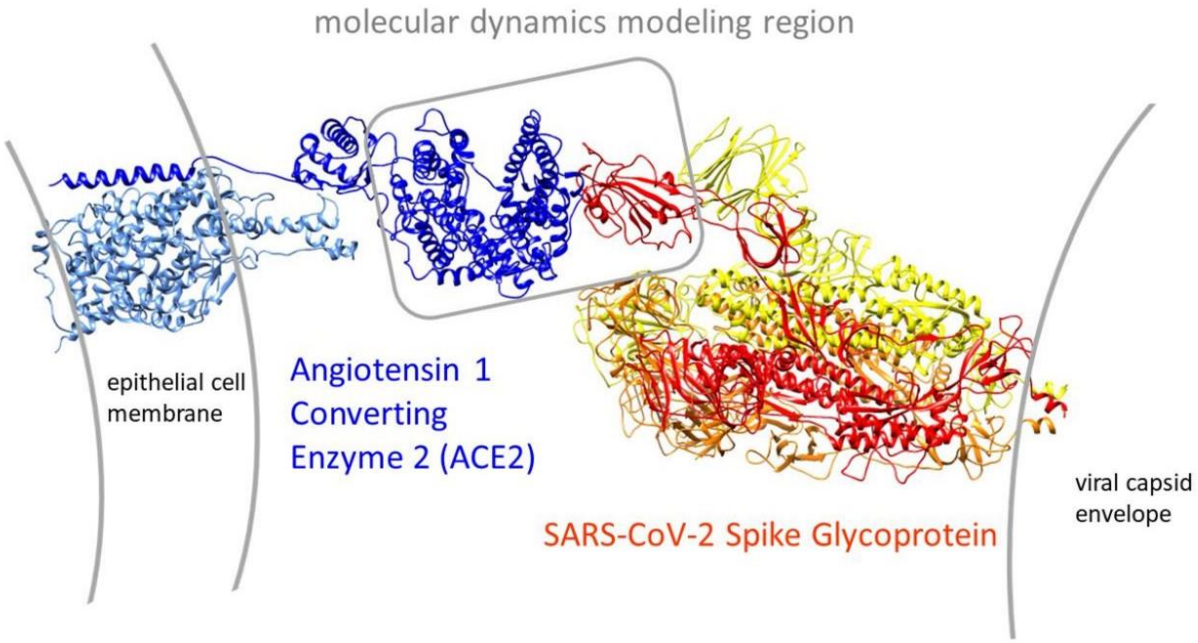


Figure 1. Overview of the open conformation SARS-CoV-2 spike glycoprotein (PDB: 6vyb) interaction with its human target cell receptor, angiotensin 1 converting enzyme 2 (ACE2 PDB: 6m17). The gray box in the center shows our modeling region that bounds most of the subsequent molecular dynamics simulations in the study.

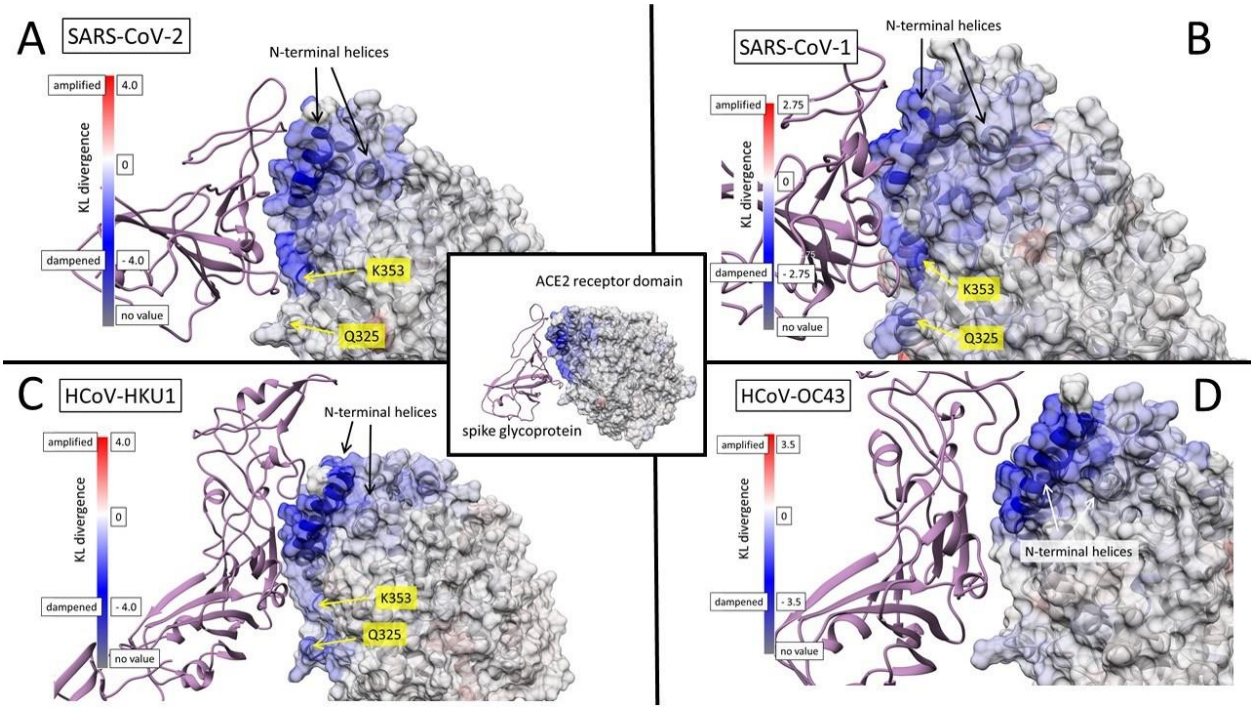


Figure 2. Color maps of dampened atom fluctuations in human ACE2 receptor protein due to the binding of betacoronavirus spike glycoprotein. Shift in atom fluctuations were calculated as the signed symmetric Kullback-Leibler divergence (i.e. relative entropy) of distributions the root mean square time deviations (rmsf or FLUX) for residue averaged protein backbone atoms (i.e. N, Cα, C and O) for ACE2 in its spike bound versus unbound dynamic state. The (A) SARS-CoV-2 (COVID19) interaction model with ACE2 model (PDB: 6m17) was used as a structural alignment template to model the potential interactions of other human betacoronavirus spike glycoproteins (B) SARS CoV-1 from PDB: 6acg (C) HCoV-HKU1 from PDB: 5gnb, and (D) HCoV-OC43 from PDB:6ohw. After 1ns of heating and equilibration, the molecular dynamics for the viral bound and unbound states were each sampled over 200 production runs at 0.2ns in charge neutralized explicit solvent periodic boundary conditions at temperature of 300K and pressure of 1 atmosphere. Note: blue mapping indicates dampening of atom motion while red indicates amplification of motion.

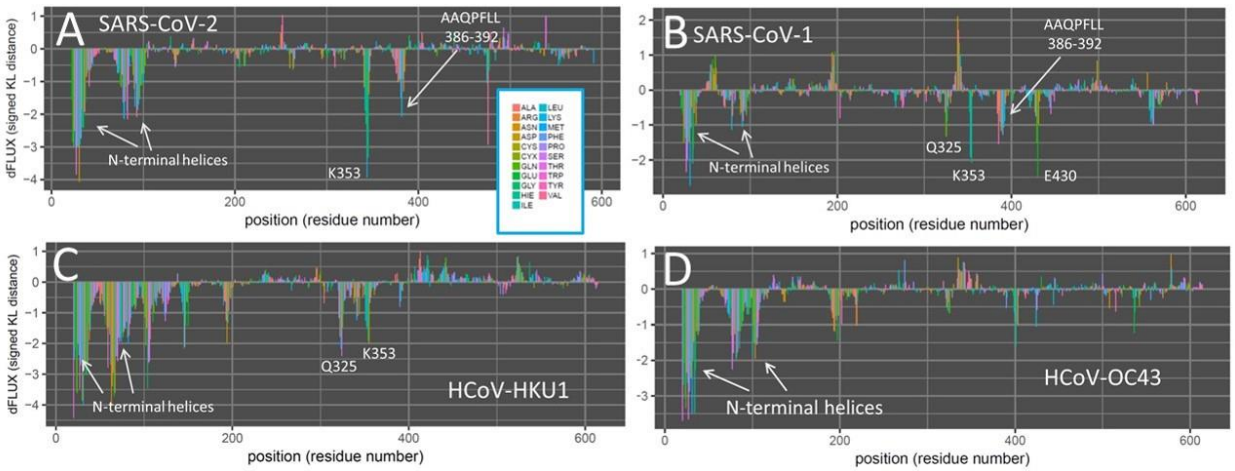


Figure 3. Positional plots of changes to atom fluctuations in human ACE2 protein due to the binding of betacoronavirus spike glycoprotein. Shift in atom fluctuations (Y axis) due to binding were calculated as the signed symmetric Kullback-Leibler divergence of distributions the root mean square time deviations for the residue averaged protein backbone atoms (i.e. N, Cα, C and O) on each amino acid (X axis) for ACE2 comparing spike glycoprotein bound versus unbound dynamic state. The KL divergences in atom fluctuation are shown for the four models in Figure 2 (A) SARS-CoV-2 (COVID19) (PDB: 6m17), (B) SARS CoV-1 from PDB: 6acg, (C) HCoV-HKU1 from PDB: 5gnb, and (D) HCoV-OC43 from PDB:6ohw.

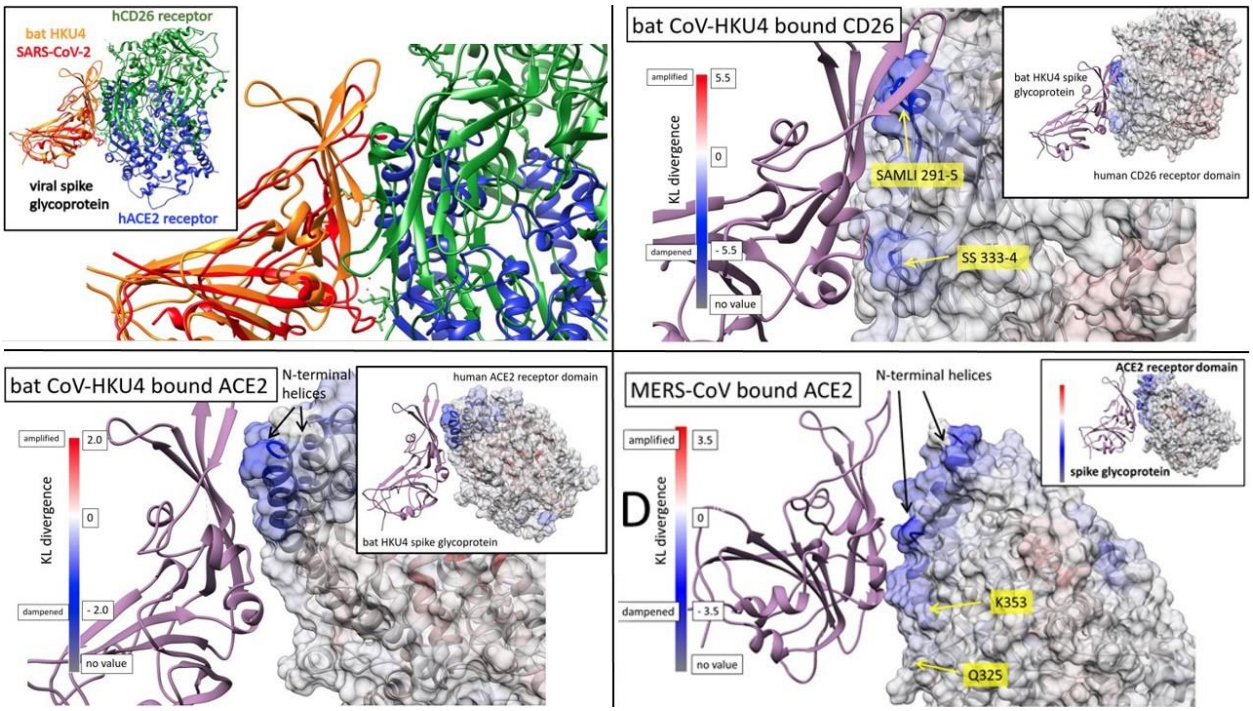


Figure 4. Color maps of dampened atom fluctuations due to binding in potential zoonotic encounters with novel betacoronavirus spike glycoproteins. (A) The structural alignment of the viral proteins from the SARS-CoV-2 interaction model with ACE2 model (PDB: 6m17) and a likely outbreak progenitor model of bat HKU4 interaction with human CD26 (PDB: 4qzv), the receptor of the Middle East Respiratory Syndrome or MERS outbreak of 2012 was used to model the potential interactions of bat HKU4 betacoronavirus interacting with human ACE2. The models of zoonotic encounters include (B) bat HKU4 interacting with human CD26 (PDB: 4qzv), (C) bat HKU4 interacting with human ACE2 (from PDB: 6m17), and (D) MERS - CoV virus from PDB: 5x5c interacting with human ACE2 from PDB: 6m17. Shift in atom fluctuations were calculated as the signed symmetric Kullback-Leibler divergence (i.e. relative entropy) of distributions the root mean square time deviations (rmsf or FLUX) for residue averaged protein backbone atoms (i.e. N, Cα, C and O) for ACE2 in its spike bound versus unbound dynamic state. After 1ns of heating and equilibration, the molecular dynamics for the viral bound and unbound states were each sampled over 200 production runs at 0.2ns in charge neutralized explicit solvent periodic boundary conditions at temperature of 300K and pressure of 1 atmosphere. Note: blue mapping indicates dampening of atom motion while red indicates amplification of motion.

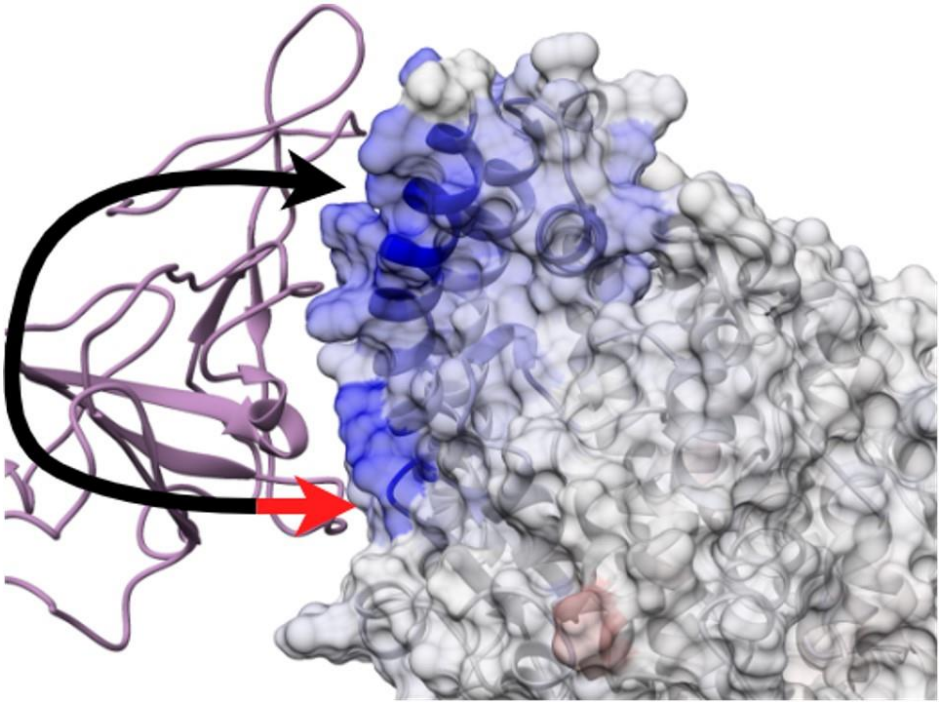


Figure 5. Two touch model of protein interaction. Black arrowhead - a promiscuous viral site of rapid dynamic interaction common across all outbreaks and also present in potential zoonotic encounters. Red arrowhead - a site active in more pathogenic outbreaks characterized by more complex dynamics enabling articulation between novel sites on human ACE2.

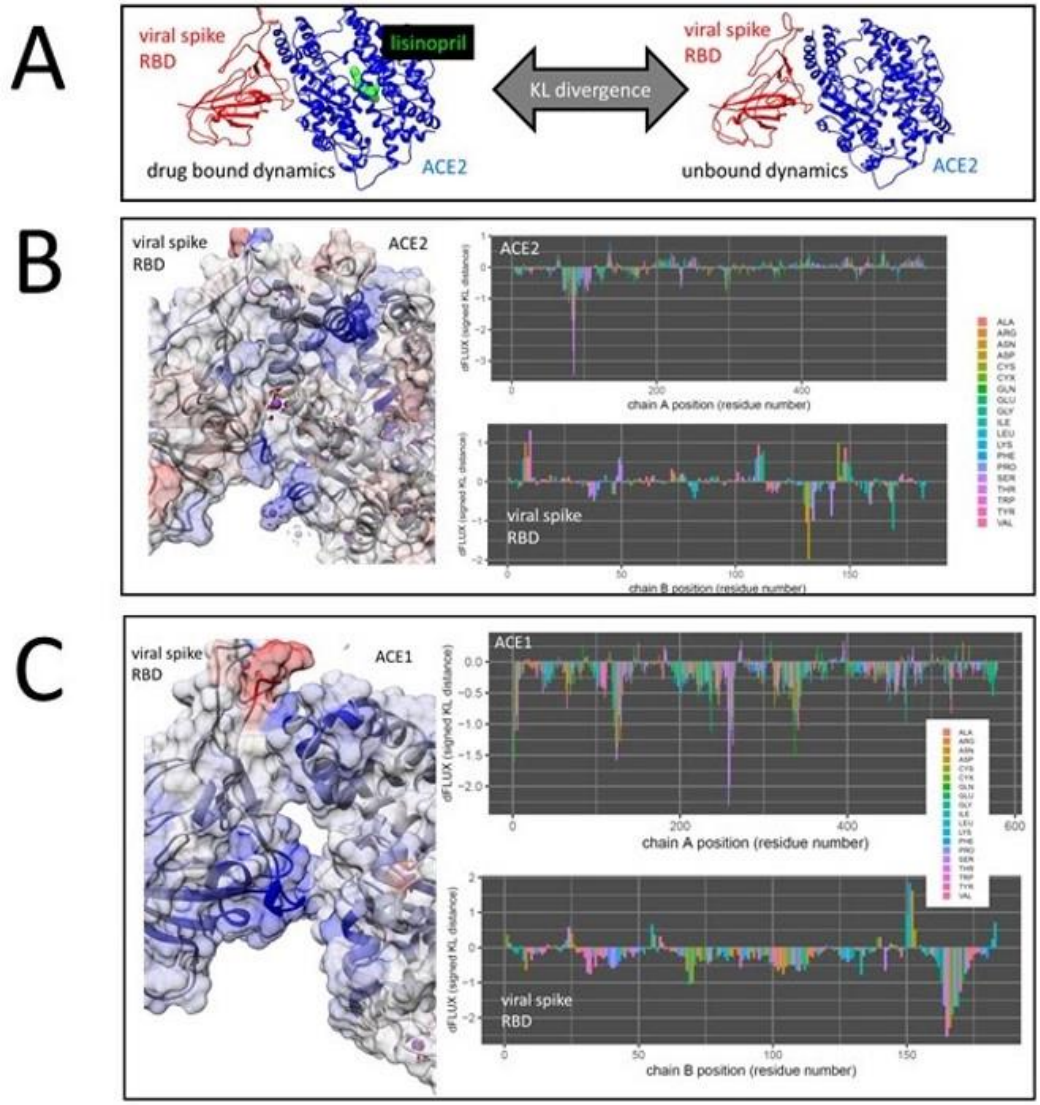
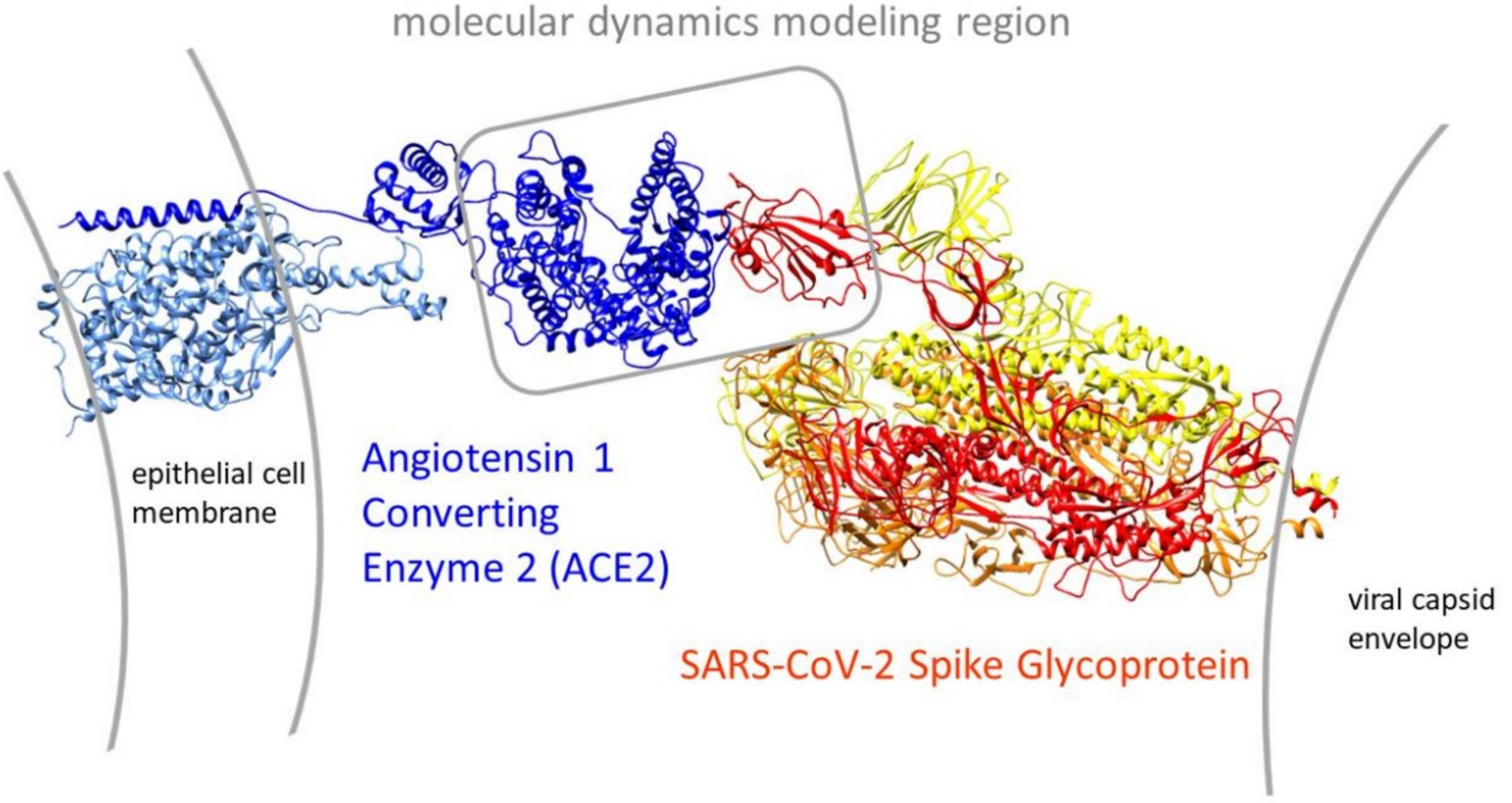
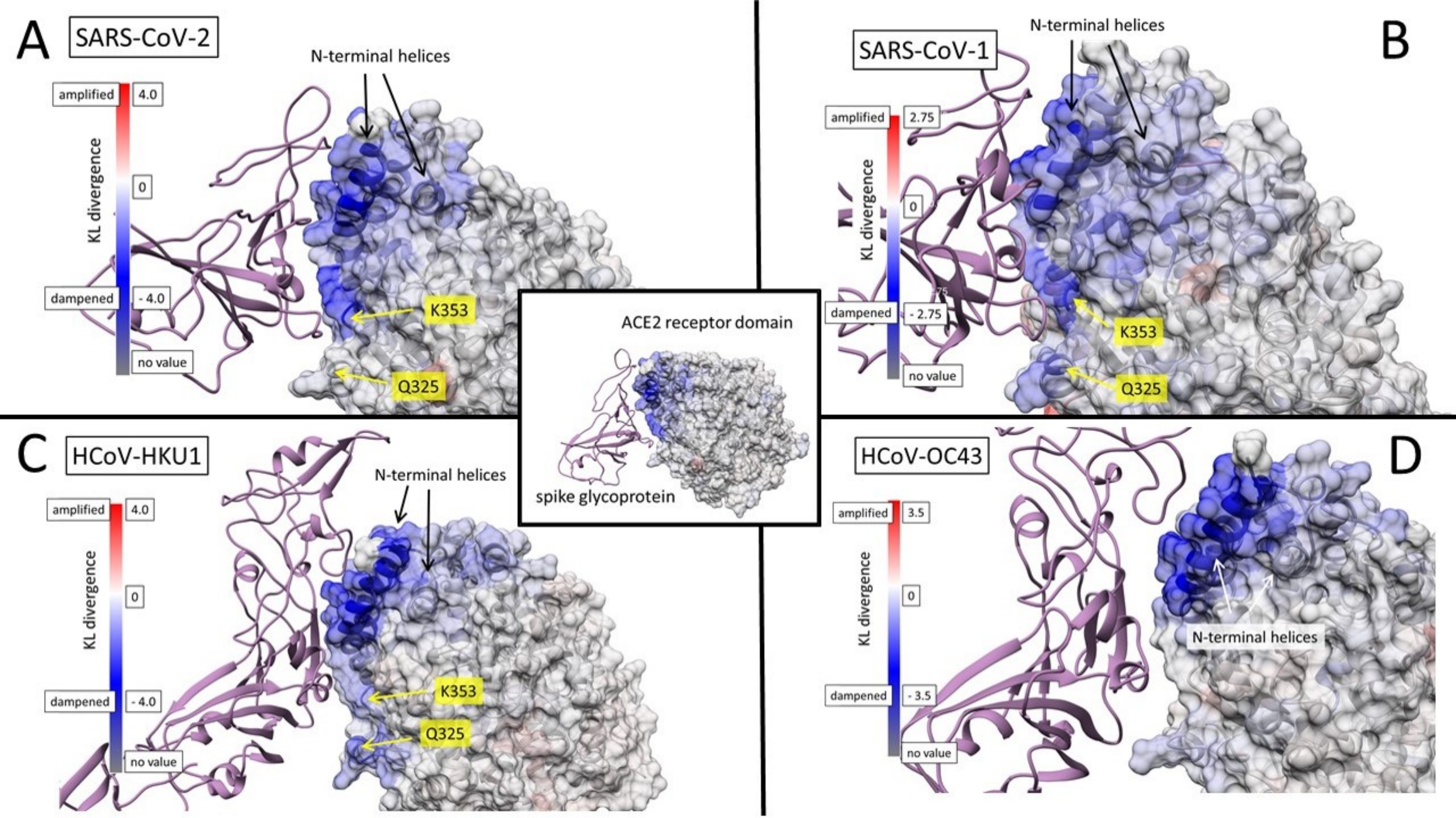
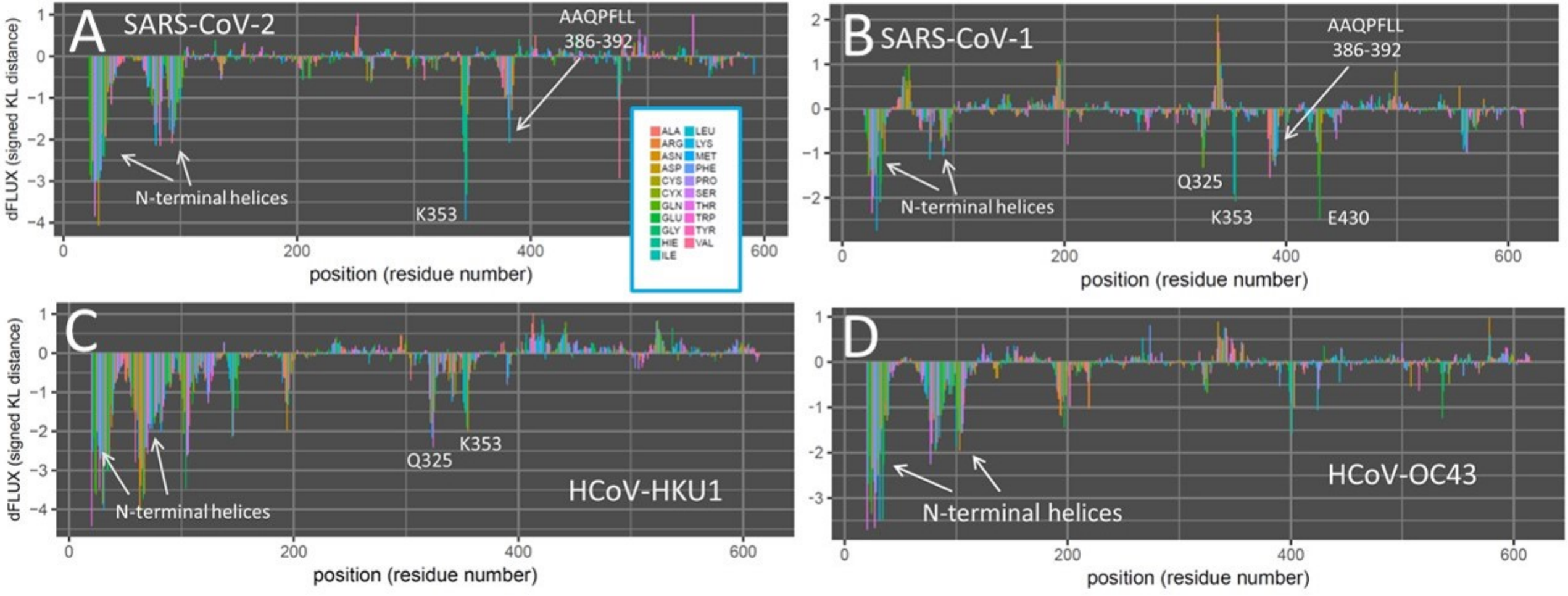
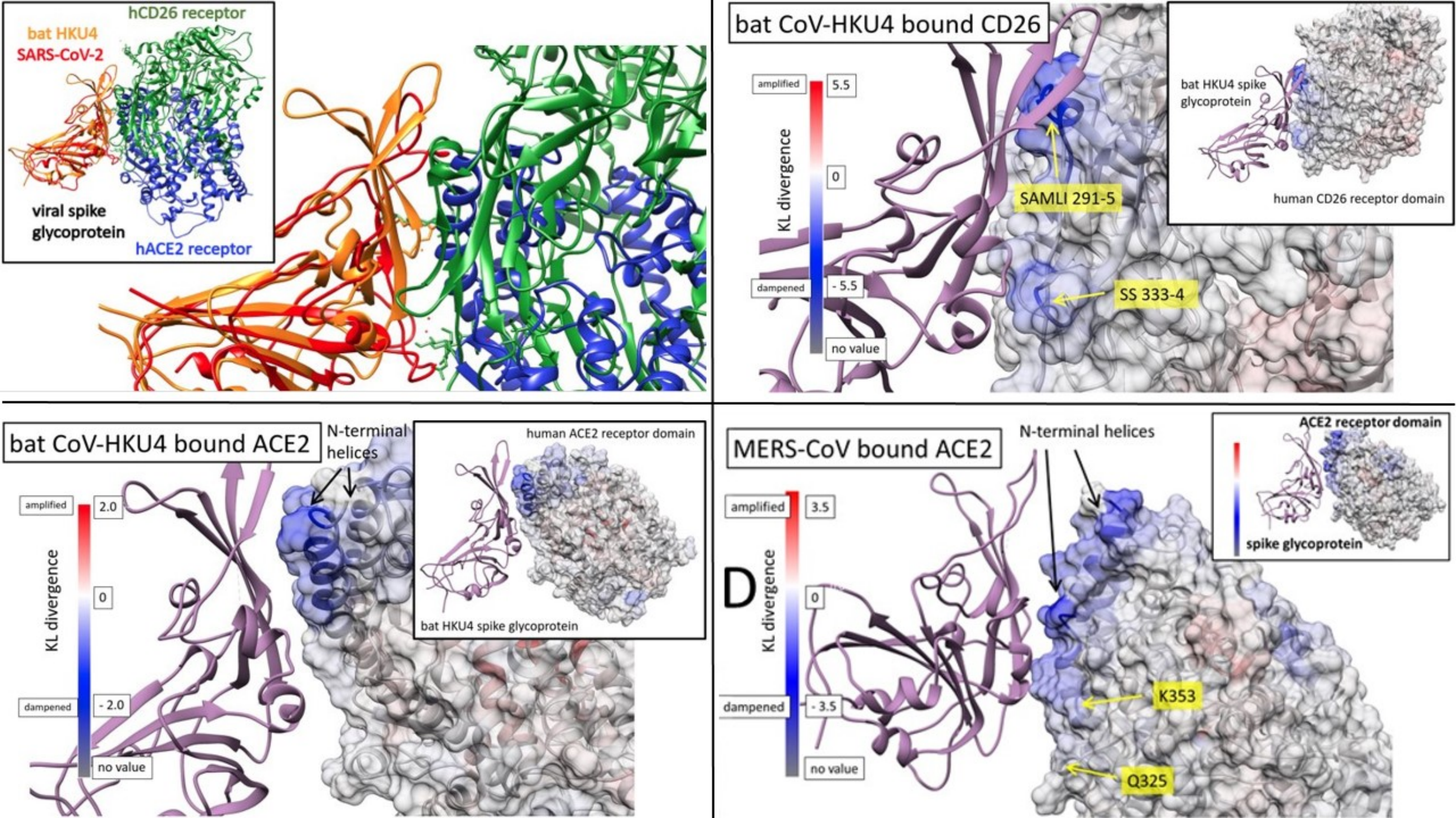


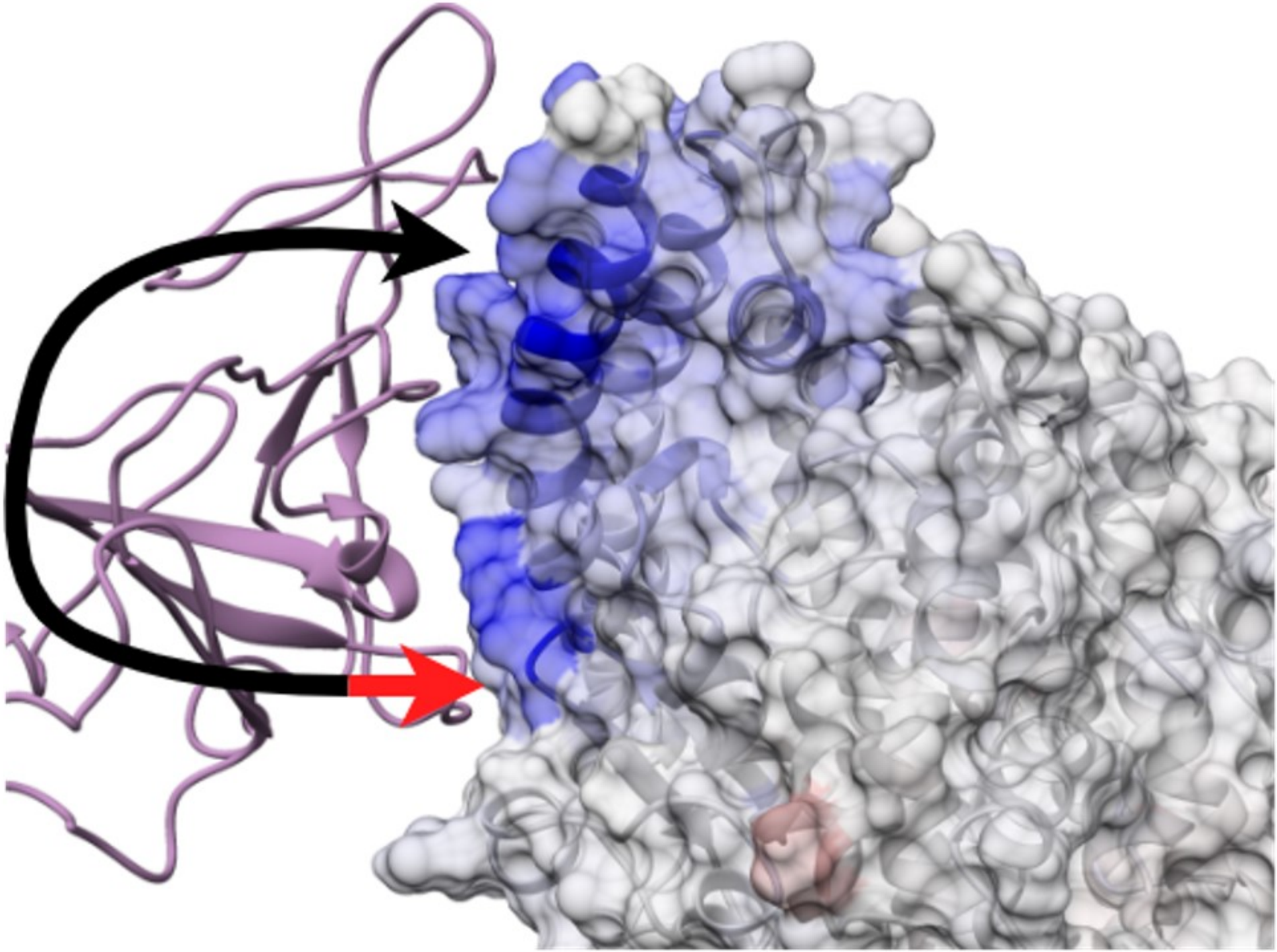
Figure 6. Enhanced binding effect of SARS-CoV-2 spike glycoprotein on human ACE protein paralogs in the presence of the hypertension drug lisinopril. A) KL divergence comparing two viral bound ACE2 models was calculated and color mapped. The comparison only differs with regard to the presence and absence of the hypertension drug lisinopril. B) Upon lisinopril binding to ACE2, stabilization of the virus-human structural interface occurs at both the one touch (i.e. N-terminal helices) and two touch regions of our model. C) Upon lisinopril binding to ACE1, the intended target of the drug, stabilization of the virus-human structural interface also occurs at both the one touch (i.e. N-terminal helices) and two touch regions of our model.

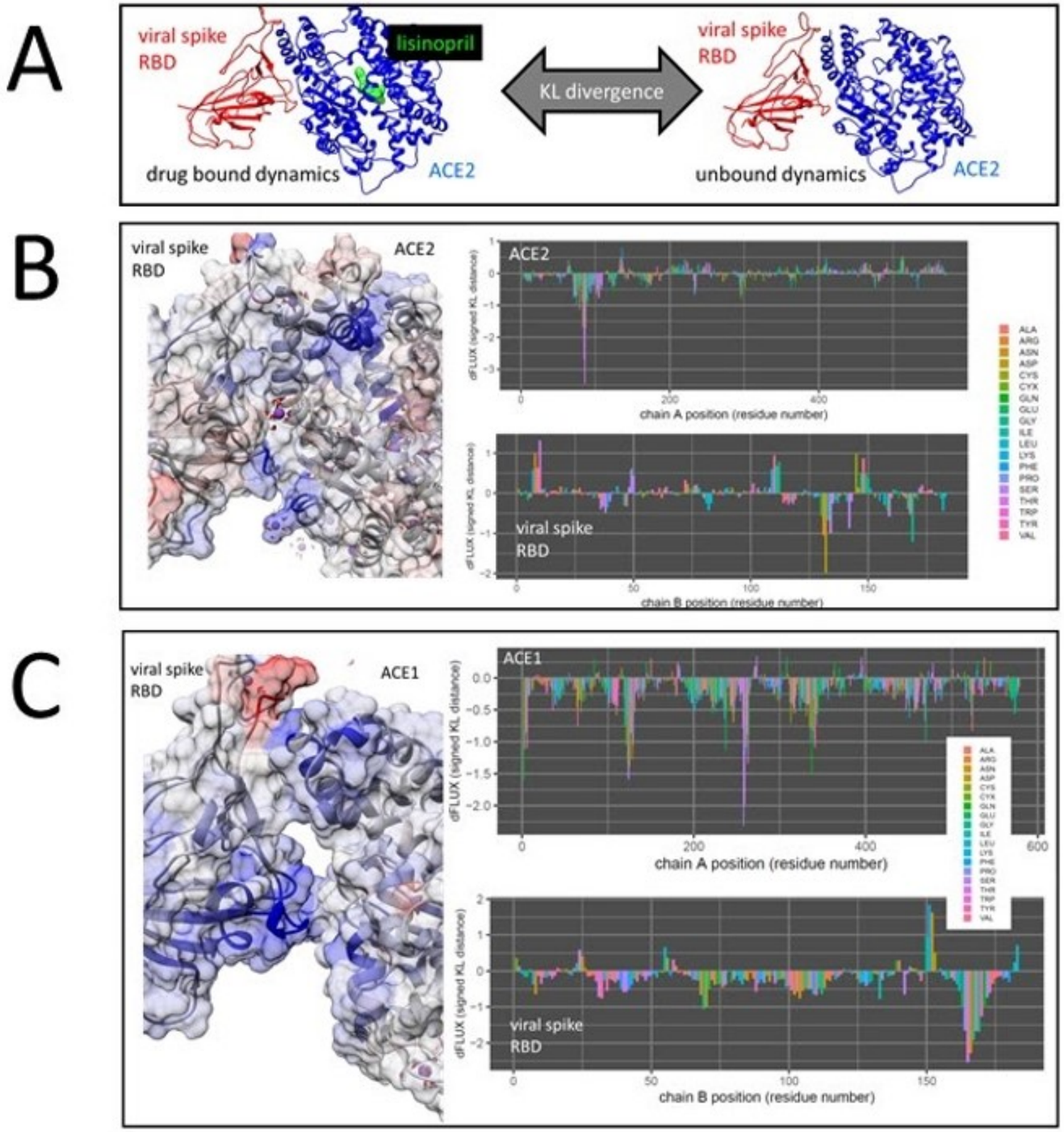

molecular dynamics modeling region
epithelial cell membrane
Angiotensin 1 Converting Enzyme 2 (ACE2)
SARS-CoV-2 Spike Glycoprotein
viral capsid envelope


A
SARS-CoV-2
N-terminal helices
amplified
4.0
KL divergence
0
dampened
- 4.0
no value
K353
Q325
B
SARS-CoV-1
N-terminal helices
amplified
2.75
KL divergence
0
dampened
- 2.75
no value
K353
Q325
ACE2 receptor domain
spike glycoprotein
C
HCoV-HKU1
N-terminal helices
amplified
4.0
KL divergence
0
dampened
- 4.0
no value
K353
Q325
D
HCoV-OC43
amplified
3.5
KL divergence
0
dampened
- 3.5
no value
N-terminal helices


A
SARS-CoV-2
AAQPFLL
386-392
N-terminal helices
K353
dFLUX (signed KL distance)
position (residue number)
ALA
ARG
ASN
ASP
CYS
CYX
GLN
GLU
GLY
HIE
ILE
LEU
LYS
MET
PHE
PRO
SER
THR
TRP
TYR
VAL
B
SARS-CoV-1
AAQPFLL
386-392
N-terminal helices
Q325
K353
E430
position (residue number)
C
HCoV-HKU1
N-terminal helices
Q325
K353
dFLUX (signed KL distance)
position (residue number)
D
HCoV-OC43
N-terminal helices
position (residue number)


hCD26 receptor
bat HKU4
SARS-CoV-2
viral spike glycoprotein
hACE2 receptor
bat CoV-HKU4 bound CD26
amplified
5.5
KL divergence
0
dampened
- 5.5
no value
SAMLI 291-5
SS 333-4
bat HKU4 spike glycoprotein
human CD26 receptor domain
bat CoV-HKU4 bound ACE2
N-terminal helices
amplified
2.0
KL divergence
0
dampened
- 2.0
no value
human ACE2 receptor domain
bat HKU4 spike glycoprotein
MERS-CoV bound ACE2
N-terminal helices
D
amplified
3.5
KL divergence
0
dampened
- 3.5
no value
K353
Q325
ACE2 receptor domain
spike glycoprotein




A
viral spike
RBD
lisinopril
drug bound dynamics
ACE2
KL divergence
viral spike
RBD
unbound dynamics
ACE2
B
viral spike
RBD
ACE2
ACE2
dFLUX (signed KL distance)
chain A position (residue number)
viral spike
RBD
chain B position (residue number)
ALA
ARG
ASN
ASP
CYS
CYX
GLN
GLU
GLY
ILE
LEU
LYS
PHE
PRO
SER
THR
TRP
TYR
VAL
C
viral spike
RBD
ACE1
ACE1
dFLUX (signed KL distance)
chain A position (residue number)
viral spike
RBD
chain B position (residue number)