

 Open access • Posted Content • DOI:10.1101/2020.04.23.057042

ACE2 polymorphisms and individual susceptibility to SARS-CoV-2 infection: insights from an in silico study — Source link

Matteo Calcagnile, Patricia Forgez, Antonio Iannelli, Antonio Iannelli ...+3 more authors

Institutions: University of Salento, University of Paris, French Institute of Health and Medical Research, University of Nice Sophia Antipolis

Published on: 24 Apr 2020 - bioRxiv (Cold Spring Harbor Laboratory)

Topics: Genetic predisposition

Related papers:

- [SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor](#)
- [Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation.](#)
- [Comparative genetic analysis of the novel coronavirus \(2019-nCoV/SARS-CoV-2\) receptor ACE2 in different populations](#)
- [Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2.](#)
- [Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor.](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/ace2-polymorphisms-and-individual-susceptibility-to-sars-cov-ql7vlhkjgg>

ACE2 polymorphisms and individual susceptibility to SARS-CoV-2 infection: insights from an *in silico* study

Matteo Calcagnile¹, Patricia Forgez², Antonio Iannelli^{3,4}, Cecilia Bucci¹, Marco Alifano^{5,6*#}, and Pietro Alifano^{1*#}

¹Department of Biological and Environmental Sciences and Technologies, University of Salento, Lecce, Italy

²INSERM UMR-S 1124 T3S, Eq 5 CELLULAR HOMEOSTASIS, CANCER and THERAPY, University of Paris, Campus Saint Germain, Paris, France

³Digestive Disease Department, Archet 2 Hospital, Nice University Hospital, University of Nice Côte d'Azur, Nice, France

⁴INSERM, U1065, Team 8 "Hepatic complications of obesity", University Nice Côte d'Azur, France

⁵Thoracic Surgery Department, Cochin Hospital, APHP Centre, University of Paris, France

⁶INSERM U1138 Team «Cancer, Immune Control, and Escape», Cordeliers Research Center, University of Paris, France

#These authors contributed equally to the work

*Corresponding authors:

Pr. Pietro Alifano

Department of Biological and Environmental Sciences and Technologies, University of Salento, Lecce, Italy

e-mail: pietro.alifano@unisalento.it

Tel: 0039 0832 298856

Pr. Marco Alifano

Thoracic Surgery Department
AP-HP, University of Paris, France

e-mail: marco.alifano@aphp.fr

Tel: 0033 1 58 41 20 64

Abstract

The current SARS covid-19 epidemic spread appears to be influenced by ethnical, geographical and sex-related factors that may involve genetic susceptibility to diseases. Similar to SARS-CoV, SARS-CoV-2 exploits angiotensin-converting enzyme 2 (ACE2) as a receptor to invade cells, notably type II alveolar epithelial cells. Importantly, ACE2 gene is highly polymorphic. Here we have used *in silico* tools to analyze the possible impact of ACE2 single-nucleotide polymorphisms (SNPs) on the interaction with SARS-CoV-2 spike glycoprotein. We found that S19P (common in African people) and K26R (common in European people) were, among the most diffused SNPs worldwide, the only two SNPs that were able to potentially affect the interaction of ACE2 with SARS-CoV-2 spike. FireDock simulations demonstrated that while S19P may decrease, K26R might increase the ACE2 affinity for SARS-CoV-2 Spike. This finding suggests that the S19P may genetically protect, and K26R may predispose to more severe SARS-CoV-2 disease.

Introduction

In principle, any new infectious agent that challenges a totally susceptible population with little or no immunity against it is able to totally infect the population causing pandemics. Pandemics rapidly spread affecting a large part of people causing plenty of deaths with significant social disruption and economic loss. However, if we look at the even worst pandemics in the human history we can realize that ethnic and geographical differences in the susceptibility to disease actually exist, in spite of the infectious sources and transmission routes that are the same for all individuals¹. The current SARS covid-19 (a shortened form of “coronavirus disease of 2019”) epidemic spread appears to be similarly influenced by ethnical and geographical factors. After its initial spread in China, the pandemic is now progressing at an accelerating rate in Western Europe and the United States of America². In these regions, the causative agent, the severe acute respiratory syndrome corona virus -2 (SARS-CoV-2) is spreading incredibly quickly between people, due to its newness – no one on earth has immunity to SARS Covid-19 – and transmission route. Yet, in the other regions of the world, the kinetics of diffusion and mortality seem less impressive, although the world has become highly interconnected as a result of a huge growth in trades and travels².

A multitude of factors may concur to explain the ethnic and geographical differences in pandemic progression and severity, including cultural, social and economic inequality, as well as health care system organization, and climate also. Mostly, considerable individual differences in genetic susceptibility to diseases may be involved³. Genomic predisposition is a major concept in modern medicine, and understanding of molecular bases of genetic predisposition can help to find prevention and treatment strategies for the corresponding diseases³. In the SARS Covid-19, even subtle inter-individual genetic differences may affect both the SARS-CoV-2 viral life cycle and the host innate and acquired immune response.

SARS-CoV-2 is an enveloped positive-stranded RNA virus that replicates in the cytoplasm, and uses envelope spike projections as a key to enter human airway cells⁴. In coronaviruses spike glycoproteins, which forms homotrimers protruding from the viral surface, are a primary determinant of cell tropism, pathogenesis, and host interspecies transmission. Spike glycoproteins comprise two major functional domains: an N-terminal domain (S1) for binding to the host cell receptor, and a C-terminal domain (S2) that is responsible for fusion of the viral and cellular membranes⁵.

Following the interaction with the host receptor, internalization of viral particles into the host cells is accomplished by complex mechanisms that culminate with the activation of fusogenic activity of spike, as a consequence of major conformational changes that, in general, may be triggered by receptor binding, low pH exposure and proteolytic activation⁵.

In some coronaviruses spike glycoproteins are cleaved by furin, a Golgi-resident protease, at the boundary between S1 and S2 domains, and the resulting S1 and S2 subunits remain non-covalently bound in the prefusion conformation with important consequences on fusogenicity⁵. Notably, at variance with SARS-CoV and other SARS-like CoV Spike glycoproteins, SARS-CoV-2 Spike glycoprotein contain a furin cleavage site at the S1/S2 boundary, which is cleaved during viral biogenesis⁶, and may affect the major entry route of viruses into the host cell⁵.

Productive entry of coronaviruses that harbor non-cleaved Spike glycoproteins (such as SARS-CoV) rely on endosomal proteases suggesting that this entry is accomplished by hijacking the host endocytic machinery⁵. Indeed, it has been reported that SARS-CoV infection is inhibited by lysomotropic agents because of the inhibition of the low-pH-activated protease cathepsin L⁷. However, SARS-CoV is also able to fuse directly to the cell membrane in the presence of relevant exogenous proteases, and this entry route is believed to be much more efficient compared to the endocytic route⁸. In fact, proteases from the respiratory tract such as those belonging to the transmembrane protease/serine subfamily (TMPRSS), TMPRSS2 or HAT (TMPRSS11d) are able to induce SARS-CoV spike glycoprotein fusogenic activity^{9,10,11,12}. The first cleavage at the S1-S2 boundary (R667) facilitates the second cleavage at position R797 releasing the fusogenic S2' sub-domain⁵. On the other hand, there is also evidence that cleavage of the ACE2 C-terminal segment by TMPRSS2 can enhance spike glycoprotein-driven viral entry¹³. Notably, it has been very recently demonstrated that also SARS-CoV-2 cell entry depends on TMPRSS2, and is blocked by protease inhibitors¹⁴.

SARS-CoV-2 and respiratory syndrome corona virus (SARS-CoV) Spike proteins share very high phylogenetic similarities (99%), and, indeed, both viruses exploit the same human cell receptor namely angiotensin-converting enzyme 2 (ACE2), a transmembrane enzyme whose expression dominates on lung alveolar epithelial cells^{6,15,16}. This receptor is an 805-amino acid long captopril-insensitive carboxypeptidase with a 17-amino acids N-terminal signal peptide and a C-terminal membrane anchor. It catalyzes the cleavage of angiotensin I into angiotensin 1-9, and of angiotensin II into the vasodilator angiotensin 1-7, thus playing a key role in systemic blood pressure homeostasis, counterbalancing the vasoconstrictive action of angiotensin II, which is generated by cleavage of angiotensin I catalyzed by ACE¹⁷. Although ACE2 mRNA is expressed ubiquitously, ACE2 protein expression dominates on lung alveolar epithelial cells, enterocytes, arterial and venous endothelial cells, and arterial smooth muscle cells¹⁸.

There is evidence that ACE2 may serve as a chaperone for membrane trafficking of an amino acid transporter B0AT1 (also known as SLC6A19), which mediates the uptake of neutral amino

acids into intestinal cells in a sodium dependent manner¹⁹. Recently, 2.9 Å resolution cryo-EM structure of full-length human ACE2 in complex with B0AT1 was presented, and structural modelling suggests that the ACE2-B0AT1 can bind two spike glycoproteins simultaneously^{20,21}. It has been hypothesized that the presence of B0AT1 may block the access of TMPRSS2 to the cutting site on ACE2^{20,21}. B0AT1 (also known as SLC6A19) is expressed with high variability in normal human lung tissues, as shown by analysis of data available in Oncomine from the work by Weiss et al²².

Notably, a wide range of genetic polymorphic variation characterizes the ACE2 gene, which maps on the X chromosome, and some polymorphisms have been significantly associated with the occurrence of arterial hypertension, diabetes mellitus, cerebral stroke, coronary artery disease, heart septal wall thickness and ventricular hypertrophy^{23,24,25}. The association between ACE2 polymorphisms and blood pressure responses to the cold pressor test led to the hypothesis that the different polymorphism distribution worldwide may be the consequence of genetic adaptation to different climatic conditions^{25,26}. In this study we have used a combination of *in silico* tools to analyze the possible impact of ACE2 single-nucleotide polymorphisms (SNPs) on the interaction with SARS-CoV-2 Spike glycoprotein. Results seem to suggest that ACE2 polymorphism can contribute to ethnic and geographical differences in SARS COVID-19 spreading across the world.

Results

3D Protein Data Bank (PDB) models and *in silico* screening of ACE2 SNPs affecting binding with coronavirus Spike proteins. All simulations were carried out using different 3D Protein Data Bank (PDB) X-ray crystallography models from RCSB (<https://www.rcsb.org/>): 2AJF (SARS-CoV Spike Receptor Binding Domain [RBD] /ACE2 complex)²⁷, 6LZG (SARS-CoV-2 Spike RBD /ACE2 complex)²⁸, 6M0J (SARS-CoV-2 Spike RBD /ACE2 complex)²⁹, 6VW1 (chimeric SARS-CoV/SARS-CoV-2 Spike RBD /ACE2 complex)³⁰, and 6M17 (SARS-CoV-2 Spike RBD /ACE2/B0AT1 complex)^{20,21} (Fig. 1a). ClustalO alignments of human ACE2 amino acid sequences, and SARS-CoV, SARS-CoV-2 and chimeric SARS-CoV/SARS-CoV-2 spike RBDs used for X-ray crystallography models are reported in Supplementary Fig. 1. 6VW1 was developed with a chimeric RBD to facilitate crystallization, by using the core from SARS-CoV RBD as the crystallization scaffold and the Receptor Binding Motif (RBM) from SARS-CoV-2 as the functionally relevant unit. Nevertheless, the structures of chimeric SARS-CoV/SARS-CoV-2 Spike RBD /ACE2 complex of 6VW1 and SARS-CoV-2 Spike RBD /ACE2 complex of 6M0J, particularly in RBM region, were highly similar³⁰.

In all models, similar to SARS-CoV RBM, SARS-CoV-2 RBM forms a concave surface that houses a convexity formed by two helices on the exposed surface of ACE2. Strong network of H-bond and salt bridge interactions mediate the receptor-ligand binding. Global energy and several distinctive features of the 3D models with and without glycosylation are reported in Supplementary Table 1. Contact residues are classified as: “permanent” (predicted as binding residues in all 10 models), “stable” (predicted as binding residues in 6 or 7 out of 10 models), “unstable” (predicted as binding residues in 6 or 7 out of 10 models), “hyper-unstable” (1 or 2 models out of 10).

Using the 3D PDB models and EVOEF^{31,32} on SSIPe web-server (<https://zhanglab.cmb.med.umich.edu/SSIPe/>) we screened the entire list of 301 ACE2 SNPs causing missense mutations from the dbSNP and UNIPROT database to identify possible amino acid substitutions that may affect binding interfaces (Supplementary Table 2). SSIPe³³ was used to estimate $\Delta\Delta G$ values associated with each amino acidic substitution, and to generate models of ACE2 polymorphic variants. The list of the amino acid substitutions that may affect the ACE2/Spike, ACE2/B0AT1 and ACE2/ACE2 interfaces is reported in Supplementary Table 3. Twenty-seven substitutions were predicted to influence the ACE2/Spike interface in at least one of the different 3D PDB models. Fifteen and seventeen were predicted to affect the ACE2/B0AT1 and ACE2/ACE2 interfaces, respectively. Some residues, which are described in UNIPROT database (<https://www.uniprot.org/uniprot/Q9BYF1>) as important for the interaction between spike and ACE2, are permanent contact residues (predicted as binding) but all of these are non-polymorphic

(Fig. 1b). In contrast, polymorphic residues are stable, unstable or hyper-unstable. A list of 18 SNPs from dbSNP (S19P, I21T, I21V, E23K, A25T, K26E, K26R, T27A, E35D, E35K, E37K, S43R, E75G, M82I, G326E, E329G, G352V; D355N), which were predicted to affect the ACE2/Spike interface, was used for further analysis.

SNPs possibly affecting ACE2 glycosylation. Supplementary Table 1 illustrates amino acid glycosylation sites, and structure of the glycosidic chains as inferred from different ACE2/Spike complex PDB models. Putative polymorphic sites (Q60R, N103H, N546D, N546S) from dbSNP database that may affect ACE2 glycosylation are also reported. One of these amino acid variations, N546D, is rather common in South Asia (Supplementary Table 4).

FireDock³⁴ was used to estimate the effects of removal of glycosidic residues or chains on ACE2 interaction with SARS-CoV-2 Spike RBD by calculating ΔG values. The data indicated that removal of glycosidic chains results in either an increased or a decreased ΔG values, depending on the PDB model (Fig. 1c). In particular, removal of glycosidic moieties apparently strengthened the ACE2/Spike interaction in SARS-CoV Spike/ACE2 in the 2AJF model, while it appeared to weaken the interaction between SARS-CoV-2 Spike and ACE2 in the 6VW1 model. In both cases, the effect was mostly due to removal of the terminal beta-mannose (BMA) (Fig. 1c), which was predicted to decorate a glycosidic chain attached to aspartic amino acid residue at position 90 that maps in a helix that is involved in the interaction with Spike, as shown in Fig. 1d. Noteworthy, in the 6VW1 model, BMA is involved in two H-bonds and one pseudo-bond (Fig. 1e), and these bonds are lost in non-glycosylated models. In contrast, in the 2AJF model, the BMA forms only one H-bond, and after removal of terminal BMA, the Thr-41 acquires more grads for binding thereby strengthening the interaction with ACE2. These results seem to suggest that ACE2 glycosylation may play a different role in modulating the interaction with SARS-CoV Spike and SARS-CoV-2 Spike.

Selected ACE2 SNPs affecting binding interfaces

FireDock³⁴ was used to estimate the effects of selected ACE2 SNPs on interaction with SARS-CoV Spike RBD (2AJF model) and SARS-CoV-2 Spike RBD (6M0J and 6WV1 models). Selection was based on both frequency of these SNPs worldwide, and predicted effects on ACE2 binding interfaces. Screening was preceded by correlation analysis of SNPs data from different databases (Supplementary Fig. 2). Network plot (Fig. 2a) and Non-Metric Multidimensional Scaling (NM-MDS, Bray-Curtis index) (Fig. 2b) of the most diffused SNPs demonstrated that S19P and K26R were, among the most diffused SNPs worldwide, the only two SNPs that were able to potentially

affect the interaction of ACE2 with SARS-CoV Spike and SARS-CoV-2 Spike (Supplementary Table 3). In particular, the S19P SNP is rather common in African people with a frequency about 0.3%, while K26R SNP is frequent in European people with a frequency about 0.5% (Supplementary Table 4).

FireDock³⁴ results indicated that the S19P substitution decreased the affinity of ACE2 with Spike in 2AJF and 6VW1 models (Fig. 2c) and similar results were obtained with all other models. Moreover, this amino acid substitution seems also to affect the ACE2 N-terminal cleavage site (Fig. 2d), and when FireDock³⁴ simulations were carried out on ACE2 with the alternative cleavage site, the effects of S19P SNP was much more impressive (Fig. 2 c).

In contrast, the K26R and the less common K26E substitutions appeared to increase the affinity of ACE2 with SARS-CoV-2 Spike (2AJF model), and slightly decrease the affinity of ACE2 with SARS-CoV Spike (6VW1 and 6M17) models (Fig. 3a). As 6VW1 was generated with a chimeric SARS-CoV/SARS-CoV-2 Spike, to support our results we performed an additional simulation by challenging the ACE2 structure from 6VW1 with the Spike structures that were generated by the different models (Fig. 3b), and the results confirmed those shown in Fig. 3a. Noteworthy, the receptor-ligand interactions was much weaker in 6M17 (SARS-CoV-2 Spike RBD /ACE2/B0AT1 complex) with respect to the other models, confirming an inhibitory function of B0AT1. However, in this model, at lower energy values, the effects of K26R/E substitutions were much more evident. FireDock^{34,35} simulations indicate that such an increased affinity between K26R ACE2 and SARS-CoV-2 Spike could be due to an increased number of H-bond and/or pseudo-bonds around Glu-35, Met-82 and Lys-353 (Fig. 3c). Based on this result, the K26R and K26E could genetically predispose to more severe SARS-CoV-2 disease. In addition, several ACE2 SNPs were predicted to affect ACE2/ACE2 homo-dimerization (E668K, N638S, R716H, R710H) or ACE2/B0AT1 interaction (L731F) (Supplementary Fig. 3).

Dynamic models of ACE2: fluctuation, deformation and chaperone requirement for correct topology maintenance

Dynamut³⁶ was used to analyze dynamic features of ACE2 receptor including fluctuation and deformation. Dynamut³⁶ calculated fluctuation and deformation scores for each ACE2 amino acid residue position (Fig. 4a) thus providing dynamic features of the analyzed protein. The model shows that ACE2 (computed by using 3D PDB model 6M17 as input file) is characterized by a high deformation tract that is located immediately upstream of the transmembrane domain (Fig. 4b), whereas the C-terminal tail is characterized by high fluctuation (Fig. 4c).

CHARMM-GUI^{37,38} and VMD/NAMD^{39,40} tools were used to model the interaction of the ACE2 transmembrane domain with the phospholipid membrane. CHARMM-GUI was used to build the model of the phospholipid membrane embedding a single chain of ACE2, while VMD/NAMD tools were used to perform Molecular Dynamics Simulation. Frames shown in Fig. 4d seem to suggest that the hydrophobic domain alone is highly unstable in the membrane confirming that a chaperone is required for correct topology maintenance. This function was assigned to the moonlighting amino acid transporter B0AT1¹⁹.

To investigate dynamic properties of ACE2 globular head, the trans-membrane helix and conserved domains were firstly mapped on a 3D structure. Then, Dynamut³⁶ simulation was carried out on ACE2 by using 6WV1 PDB model (without the transmembrane domain) (Fig. 5ab). Results indicate that some residues of the ACE2 interface, which are involved in the interaction with SARS-CoV-2 Spike glycoprotein can actually fluctuate (Fig. 5cd).

Dynamic properties of SARS-CoV-2 and SARS-CoV spike proteins were also investigated. CABS-flex⁴¹ was used to this purpose, also to compare fluctuation scores of SARS-CoV-2 and SARS-CoV Spike RBD (receptor binding domain) of spike and ACE2. Fig. 5ef depicts the CABS flex model of SARS-CoV and SARS-CoV-2 Spike, respectively while Fig. 5g illustrates the model of ACE2. CABS-flex model (Fig. 5ef) and fluctuation scores indicate that SARS-CoV Spike is characterized by higher flexibility of in a coiled-coil domain, with respect to SARS-CoV-2 Spike (Fig. 5h). This difference may affect the interaction with ACE2.

Variation of the distance between the amino acid residues involved in ACE2 binding interfaces were then analyzed by Molecular Dynamics Simulation, a computer simulation method for analyzing the physical movements of atoms and molecules. Supplementary Fig. 4a and 4c show, respectively, oscillation plot (in ångström) and variance of distance between the amino acid residues in the two helices that are involved in the SARS-CoV-2 Spike interaction, while Supplementary Fig. 4b illustrates 3D images of regions containing the amino acid residues of Supplementary Fig. 4a. Oscillation plots and variance of the distance of amino acid residues between the two helices and beta-sheet, and between the residues of the beta-sheet are illustrated in Supplementary Fig. 4d and 4f, and Supplementary Fig. 4e and 4g, respectively. Overall, the data indicate that, although the two helices in the binding interface with Spike protein form a compact structure, some residue can actually oscillate. Permanent amino acid residues (41-31) exhibit small oscillation by Molecular Dynamics Simulation. In contrast, stable or unstable residues exhibit a different behavior. The residue 21 occupies the first position in the model used (PDB 6M17) for simulation, and, as a consequence, it exhibits the highest oscillation degree. Interestingly, the

distance between the amino acid residues 26-36 and 72-82 shows high variation indicating that Met-82 and Lys-26 may be characterized by considerable degree of freedom.

ACE2 SNPs analyzed by dynamic models: structural effects on binding interfaces

Dynamut36 and ENCoM⁴² were used to compare dynamic features of ACE2 and ACE2 polymorphic variants. Both these tools predict $\Delta\Delta G$ values associated with single amino acid substitution; ENCoM⁴² also predicts $\Delta\Delta E$ values. The Dynamut³⁶ and ENCoM⁴² outputs were used to generate ordination plots (PCA) by PAST to evaluate overall results (Fig. 6a, Supplementary Fig. 5 and Supplementary Table 7).

All 197 amino acid residues that were reported as polymorphic in dbSNP were analyzed. In Supplementary Fig. 5a the ordination plots that were generated by clustering the effects of the single amino acid substitutions according to the Dynamut³⁶ (left panels) and ENCoM⁴² (right panels) outputs are illustrated. The data of the amino acid substitutions in the three ACE2 interfaces (ACE2/SARS-CoV-2 Spike, ACE2/ACE2, ACE2/B0AT1) were then extrapolated (Supplementary Fig. 5b), and were excluded from the analysis that was aimed at predicting structural (indirect) effects of amino acid substitutions on binding interfaces. The resulting subset of data (Supplementary Fig. 5c) was combined by matching Dynamut³⁶ and ENCoM⁴² predictions, and used to generate the final ordination plot shown in Fig. 6a. Results indicate that a considerable number of the amino acid substitutions are able to either stabilize or destabilize the binding interfaces with possible consequences on either ACE2 function and/or ACE2/SARS-CoV-2 Spike interaction. In particular, in Fig. 6b, the effects of amino acid substitutions on structure flexibility were calculated as $\Delta\Delta G$ and/or $\Delta\Delta S$ values comparing ACE2 and each polymorphic variant (red gain in flexibility, blue gain in rigidity). SNPs I468V, V488A and A501T were selected as the most common polymorphisms that gave the highest $\Delta\Delta G$ and $\Delta\Delta S$ values in our dynamic model.

Discussion

Although with limitations and caveats of *in silico* technology, this study addresses the question of whether some ACE2 SNPs may be associated with a different individual susceptibility to COVID-19. To alleviate these limitations, we used a combination of bioinformatics tools, and tested different crystallographic models.

Four months after the spread of the SARS COVID-19, its worldwide distribution remains extremely uneven. Lethality is even more inhomogeneous among and within countries, with figures of 12.6% in Italy⁴³, and 0.6% in South Korea⁴⁴. Although differences in mortality might have various causes, including access and efficiency of health systems, total number of people tested, presence and severity of symptoms in tested populations, they are so impressive that it seems legitimate to search for other factors possibly related to individuals as the elements of a population. Ultimately, infectivity and lethality do not seem linearly related, and probably represent problems to be solved with different, albeit complementary, approaches.

Basic aspects of epidemiology of the disease warrant some considerations: differently from other countries, in South Korea (which adopted a policy of extensive PCR screening), women represent 63% of infected people⁴⁴ as opposed to 50% in Italy⁴³ (where the policy has been to test only severely symptomatic cases for a long time). Lethality figures in women were 0.4% and 8.7% in South Korea and Italy, respectively, as opposed to 1% and 16.4% in men. It could be speculated that women are probably more prone to infection but often present a less severe disease. Although higher incidence of cardiac, respiratory and metabolic co-morbidities are probably responsible for more severe form of infection in men, estrogen-induced upregulation of ACE2 expression would explain increased susceptibility of women to a less severe and often asymptomatic form of disease. Furthermore, the ACE2 gene is located on Xp22, in an area where genes are reported to escape from X-inactivation, further explaining higher expression in females^{45,46}.

On the other hand, it has been hypothesized that, regardless of sex, pharmacological (anti-hypertensive drugs, such as ACE inhibitors and sartans) or environmental factors (NO₂ pollution), capable of inducing an overexpression of ACE2 could be responsible of increased susceptibility to infection and/or greater severity⁴⁷. ACE2 plays an essential role in the renin-angiotensin-aldosterone system, and its loss of function due to the massive binding of viral particles and internalization could constitute an essential element of the pathophysiology of pulmonary and cardiac damage during COVID-19 infection^{47,48}. In this context it should be underlined that ACE2 probably plays a dual role in the dynamic of infection and disease course. While at beginning ACE2 overexpression may increase the entry of the virus into the cell and its replication, its consequent viral-induced loss of function results in an unopposed accumulation of angiotensin II that further

aggravates the acute lung injury response to viral infection. Indeed, in the rodent blockade of the renin-angiotensin-aldosterone system limits the acute lung injury induced by the SARS-CoV-1 spike protein⁴⁹, suggesting that if ACE2 function is preserved (because of increased baseline expression, as especially seen in pre-menopausal women), clinical course of infection might be less severe.

It has been suggested that polymorphisms in the ACE2 gene could reduce the spike affinity, with subsequent lower susceptibility to infection: in this hypothesis, their geographical / ethnical distribution could explain the strong discrepancies in infection rate and/or lethality observed worldwide⁴⁷. Effectively, we showed by Network plot and Non-Metric Multidimensional Scaling that most of the SNPs diffused worldwide did not affect significantly the interaction of ACE2 with SARS-CoV-2 Spike. S19P was one of the rare polymorphisms able to potentially affect this interaction, by lowering the affinity. This polymorphism is more frequent in African populations, but its diffusion (0.3%) remains too low to explain, except in minimal part, the reduced death toll observed so far in that continent, and, more generally, the enormous differences in geographical spread of infection and lethality. However, it seems clear that the affinity of the virus for ACE2 is a key determinant of its infective potential: in order to choose the experimental model capable of reproducing the essential aspects of human infection, Chan and colleagues⁵⁰ determined *in silico* the spike / ACE2 affinity in primates and in a series of experimental animals, observing that the binding energy is maximal in primates (-62.20 Rosetta energy units (REU)), intermediate in Syrian hamster (-49.96 REU), lower in bat (-39.60 REU). This allowed the authors to predict that hamsters could be infected, which was experimentally confirmed –underlining the reliability of *in silico* modeling- and could be subsequently at the origin of inter-animal transmission. However, hamsters, although developing clinical signs of the infection and relative histopathological changes, did not die⁵⁰: we speculate that lethality may be related to Spike/ACE2 affinity. On the other hand, the lower affinity in bat could explain –besides a better immune control- why these animals are carriers without dying.

In the same study, Chan and colleagues⁵⁰ showed that the binding energy between ACE2 and Spike of SARS-CoV, responsible for the 2002 epidemic, was -39.49 REU as compared to -58.18 of human ACE2. After that epidemic, attempts at developing mouse experimental models, resulted only in mild lung inflammation and rapid viral clearance, until development of transgenic mice expressing human ACE2 under regulation of a global promoter or cytokeratin 18 promoter, which developed rapidly lethal infection after intranasal viral inoculation^{51, 52}. Interestingly, intranasally CoV infected transgenic mice expressing human ACE2 driven by the mouse ACE2 promoter, developed severe disease (on clinical and histopathological grounds, including typical interstitial

pneumonia and widespread extrapulmonary organ damage) without dying; furthermore their viral clearance was markedly slower as compared to wild type mice which had only mild abnormalities⁵³.

So, if modestly SNP-determined lower affinity between spike and ACE2 does not seem to explain the differences in the distribution and lethality of the disease in humans, we hypothesize that the question can be addressed in a specular way: perhaps polymorphisms responsible for higher affinity can be responsible of higher severity of disease, especially when very high affinity receptors are overexpressed because of the above mentioned environmental and pharmacological factors. Obviously underlying diseases would contribute to an even more severe course of the disease, with an intense viral replication capable of infecting in turn a large number of persons, including some individuals with similar ACE2 polymorphisms, and so on. Our *in silico* models allowed us to identify K26R and K26E as SNPs with a possible increase in Spike/ACE2 affinity. K26R SNP is relatively frequent in European people with a frequency about 0.5%, which would correspond to a potential target population of 2,230,000 people at the European Union level.

In addition to FireDock^{34, 35} simulations that led to predict the possible effects of S19P, K26R and K26E ACE2 SNPs, Dynamut³⁶ and ENCoM⁴² tools were used to compare dynamic features of ACE2 and its polymorphic variants in order to analyze the possible indirect effects on binding interfaces of SNPs that are located outside these interfaces. SNPs I468V, V488A and A501T were identified as the most common SNPs that may produce these indirect effects in dynamic models. Although the precise effects of these SNPs on the interaction between ACE2 and SARS-CoV-2 or SARS-CoV Spike proteins have to be determined in more detail, nevertheless, it is desirable to use dynamic modeling to unmask indirect effects of SNPs.

It seems necessary to confirm *in vivo* that, among patients with serious disease and/or fatal outcome, polymorphisms responsible for a very high Spike/ACE2 affinity are more frequent than among patients with less severe/asymptomatic disease or even than in general population. Obviously, the impact of these polymorphisms on severity of outcome should be weighted by appropriate demographic and clinical factors. If these differences were confirmed, this would pave the way for the identification, on a population scale, of healthy individuals whose molecular phenotypes would be responsible for more serious disease. Apart from the usual social distancing measures, which could be reinforced for these cases, targeted drug prevention strategies could be evaluated. It could be logical to assess pharmacological prophylactic interventions, as proposed in categories of healthy people at particular risk of exposure such as care-givers. In particular, chloroquine, interfering with N-terminal glycosylation of ACE2, could lower its affinity for spike, thus representing an interesting candidate. In our *in silico* model, we found that removal of

glycosidic moieties weakened the interaction between SARS-CoV-2 spike and ACE2. The serine protease inhibitor camostat mesylate, approved in Japan to treat unrelated diseases, has been shown to block TMPRSS2 activity^{54,55} and is thus another interesting candidate.

On the other hand, the identification of broader categories of people with lower risk of developing severe disease, could allow a safer exit from the lock-down phase, while facilitating the establishment of a faster herd immunity, and waiting reliable serological tests and, above all, effective vaccines.

On the basis of our *in silico* study we speculate that infection and mortality are determined at individual level by different factors including the amount of expression of ACE2 and its affinity with the spike protein. While the level of ACE2 expression possibly determinates the probability of infection in the presence of an adequate inoculum, the severity of the disease is mainly determined by the phenotype affinity for the spike protein. Clinical studies are urgently required to confirm the present mechanistic hypothesis.

Materials and Methods

Databases. 3D structures of proteins were downloaded from PDB (RCSB Protein Data Bank⁵⁶). We focused our analysis on 2AJF for SARS-CoV²⁷ (DOI:[10.2210/pdb2AJF/pdb](https://doi.org/10.2210/pdb2AJF/pdb)) and 6VW1³⁰ (DOI:[10.2210/pdb6VW1/pdb](https://doi.org/10.2210/pdb6VW1/pdb)), 6M17^{20, 21} (DOI:[10.2210/pdb6M17/pdb](https://doi.org/10.2210/pdb6M17/pdb)) 6LZG²⁸ ([10.2210/pdb6LZG/pdb](https://doi.org/10.2210/pdb6LZG/pdb)), 6M0J²⁹ ([10.2210/pdb6M0J/pdb](https://doi.org/10.2210/pdb6M0J/pdb)) models for SARS-COV-2. dpSNP database^{57,58} was used to identify the ACE2 receptor SNPs, and to select the most diffused ones. Functional information was acquired by UNIPROT database⁵⁹. Chimera⁶⁰ was used as a tool for Image generation, 3D mapping, PDB managing and to analyze the results.

Binding interface characterization. The selected PDB models were analyzed by a structural point of view using Chimera software in order to identify the glycosylation sites and the secondary structures of proteins involved in the binding between ACE2 and Spike protein receptor binding domain (RBD). To estimate the effect of glycosylations we implemented a static model. Chimera was used to remove glycosidic residues, while FireDock^{34,35} was used to compute the global energy scores between the native structures and the de-glycosylated models. On the other side, starting from the entire list of SNPs, SSIPe (EVOEF)³² was used to identify the residues involved in the binding interfaces. A second step, which was carried out with SSIPe (SSIPe)³³, was aimed at estimating the effects of single SNPs, and to generate mutant models. Different SNPs lists were obtained, which were compared, and used to identify the most stable binding amino acid residues. SSIPe analysis performed with the PDB model 6M17 was used to map: the ACE2/Spike protein interaction interface, ACE2 /ACE2 dimerization interface, and B0AT1/ACE2 interaction interface. The model contains ACE2 in the dimeric form (with the hydrophobic domains) and B0AT1, while in all the others models the transmembrane domains, ACE2 /ACE2 interface and B0AT1/ACE2 interface are absent.

Dynamic analysis of ACE2 structure. To obtain a dynamic model of the ACE2 we used different tools. Dynamut³⁶ was used to calculate the general dynamic features of the ACE2 in 6M17 model leading to the identification of two domains characterized by high deformations and high fluctuations scores, respectively. To validate these results and to compare them with Spike proteins fluctuations, we used CABSflex⁴¹ to analyze Spike protein chains in 2AJF and 6VW1 models, and ACE2 chain in 6M17 model. Another Dynamut analysis was performed on 6VW1 model, considering only the structure of the globular head of ACE2. The transmembrane domain was mapped on a 3D file using TMHMM to predict hydrophobic helix. A molecular dynamics approach was used to model the behavior of the transmembrane domain. CHARMM-GUI^{37,38} was used to

implement the model of a DPPC (phosphatide) bilayer that embedded ACE2, while NAMD⁴⁰ was used to perform the analysis via VMD (NAMD GUI)^{39,40}. Chimera was used to select the frames, and to measure the oscillations of interface residues.

Effect of mutations on ACE2 dynamic features. Dynamut³⁶ was used to estimate the effect of SNPs on structural dynamics features. This tool generated two results with the two algorithms ENCoM⁴¹ and Dynamut. A quartile-based clustering was performed with 197 SNPs obtaining two equivalent plots where results were clustered in two ways: ENCoM groups and Dynamut groups. After this step, we extrapolated data related to the three interfaces (ACE2/Spike protein, ACE2/B0AT1 and ACE2/ACE2), because the analyzed SNPs can directly interfere on either binding to Spike proteins, or ACE2 homo-dimerization or ACE2/B0AT1 binding. In order to select the best results from this last subset, the clusters for Dynamut and for ENCoM were combined. PCA plots were generated using PAST software⁶¹.

Docking: dynamic models of ACE2 / Spike protein interactions. To obtain the protein complex models with and without glycosylations we used two software: PathDock⁶² and Gramm-X⁶³, respectively. Resulting complexes were submitted to FireDock^{34,35} to evaluate the energy score. As receptors and ligands we used all chains of ACE2 in PDB models (2AJF, 6VW1, 6LZG, 6M0J, 6M17), and the models of isoform X1 and X3, which were generated by using SwissModel⁶⁴ tool, while all chains of the Spike protein in the PDB files were used as ligands. For each software, the total simulations number was 35. All the obtained results were screened excluding: i) the high-energy complexes, ii) bad-orientated solutions (overlapping of receptor and ligand chain) and iii) off-target solutions (binding in membrane helix, in B0AT1 interface or ACE2 / ACE2 homo-dimerization interface). The selected results were used as inputs to superposition using: the docking solutions as a model, the chains of B0AT1(2X) / ACE2 (dimeric) from 6M17 model, and Spike protein trimeric structure, which were obtained by using I-Tasser⁶⁵ (<https://zhanglab.ccmb.med.umich.edu/COVID-19/>). To obtain multiple conformations we performed more than one superposition (Chimera) on the solutions changing the aligned chains. The final results were analyzed using Chimera.

African and European static and dynamic models. To gain information about the geographical distribution and abundance of ACE2 SNPs, we analyzed dbSNP^{57,58} and available databases: GnomAD-Exomes, TopMed, ExAC, GnomAD-Genomes, GO Exome Sequencing Project, 1000Genomes. The frequency values of the most abundant SNPs (1KGB database) were reported in

Supplementary Table S4. Data were analyzed by PAST⁶¹ that generated NM-MDS ordination plots. Additionally, network and between-database correlation analyses (PAST⁶¹) were carried out in order to clarify the relationship between SNPs, and the correlation between data of all databases.

To model the binding of S19P ACE2 we implemented two types of analyses: static analysis and dynamic analysis. The models with substitutions of one single amino acid were selected from SIPPe results (static model), and were submitted to FireDock^{34, 35} server. In this step we focused on mutant and wild type models of 6VW1 and 2AJF. Serine in position 19 is the first residue of the mature chain of ACE2, while residues from 1 to 18 form a signal peptide. To prove if this SNP changes the cut site of the ACE2 precursor we used Signal IP 5.0⁶⁶. The sequence of this predicted mutant was submitted to I-Tasser server⁶⁵ in order to obtain the mutant models. Using this predicted model as a receptor and SIPPe complexes as a reference we superposed the structures obtaining two static models (one for ligand: 2AJF and 6VW1). The dynamic models were implemented using Gram-X (ligands Spike chains of 6VW1 and 2AJF; receptor mutant models). All models obtained were submitted to FireDock^{34, 35} in order to obtain a value of global energy.

To model the binding K26R or K26E ACE2 we used a set of static models. Firstly, starting from the dataset of the mutant models selected by SIPPe³³ results, we calculated changes in global energy. This step was repeated for three models: 6VW1, 2AJF and 6M17. For this last model, we calculated the variation in terms of global energy in two conformations, using as reference models both: i) ACE2-spike chains B-E and ii) ACE2-Spike chains D-F. In order to estimate the effect of different ligands, we used the models 6VW1 from SIPPe as reference structure and receptor, and all Spike models reported in this study as ligands. In a similar manner, other static models from SSIPe were used as models to estimate the variation in terms of free energy related to polymorphisms that map on the ACE2 dimerization interface and ACE2-B0AT binding interface.

References

1. Cohn, S. K. Pandemics: waves of disease, waves of hate from the Plague of Athens to A.I.D.S. *Hist J* **85**, 535-555 (2012). DOI: [10.1111/j.1468-2281.2012.00603.x](https://doi.org/10.1111/j.1468-2281.2012.00603.x)
2. Fang, L., Karakiulakis, G. & Roth, M. Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? *Lancet Respir Med* **8**, (2020). DOI: [10.1016/S2213-2600\(20\)30116-8](https://doi.org/10.1016/S2213-2600(20)30116-8).
3. Knight, J.C. Genomic modulators of the immune response. *Trends Genet* **29**, 74-83 (2013). DOI: [10.1016/j.tig.2012.10.006](https://doi.org/10.1016/j.tig.2012.10.006).
4. Chan, J.F.W. et al. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg Microbes Infect* **28**, 9, 221-236 (2020a). DOI : [10.1080/22221751.2020.1719902](https://doi.org/10.1080/22221751.2020.1719902).
5. Belouzard, S., Millet, J.K., Licitra, B.N. & Whittaker, G. R. Mechanisms of coronavirus cell entry mediated by the viral spike protein. *Viruses* **4**, 1011-1033 (2012). DOI: [10.3390/v4061011](https://doi.org/10.3390/v4061011).
6. Walls, A.C. et al. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **181**, 2, 281-292.e6 (2020). DOI : [10.1016/j.cell.2020.02.058](https://doi.org/10.1016/j.cell.2020.02.058).
7. Simmons, G. et al. Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry. *Proc Natl Acad Sci USA* **102**, 11876–11881 (2005). DOI: [10.1073/pnas.0505577102](https://doi.org/10.1073/pnas.0505577102).
8. Matsuyama, S., Ujike, M., Morikawa, S., Tashiro, M. & Taguchi, F. Protease-mediated enhancement of severe acute respiratory syndrome coronavirus infection. *Proc Natl Acad Sci USA* **102**, 12543–12547 (2005). DOI: [10.1073/pnas.0503203102](https://doi.org/10.1073/pnas.0503203102).
9. Bertram, S. et al. Cleavage and activation of the severe acute respiratory syndrome coronavirus spike protein by human airway trypsin-like protease. *J Virol* **85**, 13363–13372 (2011). DOI: [10.1128/JVI.05300-11](https://doi.org/10.1128/JVI.05300-11).
10. Glowacka, I. et al. Evidence that TMPRSS2 activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces viral control by the humoral immune response. *J Virol* **85**, 4122–4134 (2011). DOI: [10.1128/JVI.02232-10](https://doi.org/10.1128/JVI.02232-10).
11. Kam, Y.W., et al. Cleavage of the SARS coronavirus spike glycoprotein by airway proteases enhances virus entry into human bronchial epithelial cells in vitro. *PLoS One* **4**, 11, e7870 (2009): DOI: [10.1371/journal.pone.0007870](https://doi.org/10.1371/journal.pone.0007870).
12. Shulla, A. et al. A transmembrane serine protease is linked to the severe acute respiratory syndrome coronavirus receptor and activates virus entry. *J Virol* **85**, 873–882 (2011). DOI: [10.1128/JVI.02062-10](https://doi.org/10.1128/JVI.02062-10).

13. Heurich, A. et al. TMPRSS2 and ADAM17 cleave ACE2 differentially and only proteolysis by TMPRSS2 augments entry driven by the severe acute respiratory syndrome coronavirus spike protein. *J Virol* **88**, 2, 1293–307 (2014). DOI:[10.1128/JVI.02202-13](https://doi.org/10.1128/JVI.02202-13).
14. Hoffmann, M. et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* 2020 **181**, 271-280.e8 (2020). DOI: [10.1016/j.cell.2020.02.052](https://doi.org/10.1016/j.cell.2020.02.052).
15. Letko, M., Marzi, A. & Munster, V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat Microbiol* **5**, 562-569 (2020). DOI:[10.1038/s41564-020-0688-y](https://doi.org/10.1038/s41564-020-0688-y).
16. Wrapp, D. et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*, **367**, 1260-1263 (2020). DOI:[10.1126/science.abb2507](https://doi.org/10.1126/science.abb2507).
17. Zheng, Y.Y., Ma, Y.T., Zhang, J.Y., Xie, X. COVID-19 and the cardiovascular system. *Nat Rev Cardiol* **17**, 259-260 (2020). DOI:[10.1038/s41569-020-0360-5](https://doi.org/10.1038/s41569-020-0360-5).
18. Hamming, I. et al. Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *J Pathol* **203**, 631–637 (2004). DOI:[10.1002/path.1570](https://doi.org/10.1002/path.1570).
19. Kowalczyk, S. et al. A protein complex in the brush-border membrane explains a Hartnup disorder allele. *FASEB J* **22**, 2880-2887 (2008). DOI:[10.1096/fj.08-107300](https://doi.org/10.1096/fj.08-107300).
20. Yan, R., Zhang, Y., Li, Y., Xia, L., Zhou, Q. Structure of dimeric full-length human ACE2 in complex with B0AT1.
Preprint at <https://www.biorxiv.org/content/10.1101/2020.02.17.951848v1> (2020). DOI : 10.1101/2020.02.17.951848
21. Yan, R. et al. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **367**, 6485, 1444-1448, (2020). DOI:[10.1126/science.abb2762](https://doi.org/10.1126/science.abb2762).
22. Weiss J. et al. Frequent and focal FGFR1 amplification associates with therapeutically tractable FGFR1 dependency in squamous cell lung cancer. *Sci Transl Med* **2**:62ra93, (2010). DOI: 10.1126/scitranslmed.3001451.
23. Lu, N. et al. ACE2 gene polymorphism and essential hypertension: an updated meta-analysis involving 11,051 subjects. *Mol Biol Rep* **39**, 6, 6581–6589 (2012). DOI:[10.1007/s11033-012-1487-1](https://doi.org/10.1007/s11033-012-1487-1).
24. Pinheiro, D.S. et al. The combination of ACE I/D and ACE2 G8790A polymorphisms reveals susceptibility to hypertension: A genetic association study in Brazilian patients. *PLoS ONE* **14**, e0221248 (2019). DOI:[10.1371/journal.pone.0221248](https://doi.org/10.1371/journal.pone.0221248).

25. Zhang, Q. et al. Association of Angiotensin-Converting Enzyme 2 gene polymorphism and enzymatic activity with essential hypertension in different gender: A case-control study. *Medicine (Baltimore)* **97**, 42, e12917 (2018). DOI: [10.1097/MD.00000000000012917](https://doi.org/10.1097/MD.00000000000012917).
26. Huang, J. et al. Polymorphisms of ACE2 are associated with blood pressure response to cold pressor test: the GenSalt study. *Am J Hypertens* **25**, 8, 937–942 (2012). DOI: [10.1038/ajh.2012.61](https://doi.org/10.1038/ajh.2012.61).
27. Li, F., Li, W., Farzan, M. & Harrison, S. C. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **309**, 5742, 1864-1868 (2005). DOI: [10.1126/science.1116480](https://doi.org/10.1126/science.1116480).
28. Wang et al., 2020 <http://www.rcsb.org/structure/6LZG>
29. Lan, J. et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 1-9. (2020). DOI: [10.1038/s41586-020-2180-5](https://doi.org/10.1038/s41586-020-2180-5).
30. Shang J, F et al. Structural basis of receptor recognition by SARS-CoV-2. *Nature* 1-8, (2020) DOI: [10.1038/s41586-020-2179-y](https://doi.org/10.1038/s41586-020-2179-y).
31. Pearce, R., Huang, X., Setiawan, D. & Zhang, Y. EvoDesign: Designing protein-protein binding interactions using evolutionary interface profiles in conjunction with an optimized physical energy function. *J Mol Biol* **431** 2467-2476 (2019). DOI: [10.1016/j.jmb.2019.02.028](https://doi.org/10.1016/j.jmb.2019.02.028).
32. Huang X., Pearce R. & Zhang Y. EvoEF2: accurate and fast energy function for computational protein design. *Bioinformatics* **36**,1135-1142 (2020). DOI: [10.1093/bioinformatics/btz740](https://doi.org/10.1093/bioinformatics/btz740).
33. Huang, X., Zheng, W., Pearce, R., & Zhang, Y. SSIPe: accurately estimating protein–protein binding affinity change upon mutations using evolutionary profiles in combination with an optimized physical energy function. *Bioinformatics* **btz926**, (2019). DOI: [10.1093/bioinformatics/btz926](https://doi.org/10.1093/bioinformatics/btz926).
34. Andrusier, N., Nussinov, R., & Wolfson, H. J. (2007). FireDock: fast interaction refinement in molecular docking. *Proteins* **69**, 1, 139-159, (2007). DOI: [10.1002/prot.21495](https://doi.org/10.1002/prot.21495).
35. Mashiach, E., Schneidman-Duhovny, D., Andrusier, N., Nussinov, R., & Wolfson, H. J. FireDock: a web server for fast interaction refinement in molecular docking. *Nucleic Acids Res* **36**, W229-W232, (2008). DOI: [10.1093/nar/gkn186](https://doi.org/10.1093/nar/gkn186).
36. Rodrigues, C. H., Pires, D. E., & Ascher, D. B. DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Res* **46**, W1, W350-W355 (2018). DOI: [10.1093/nar/gky300](https://doi.org/10.1093/nar/gky300).

37. Jo, S., Kim, T., Iyer, V. G., & Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem* **29**, 1859-1865 (2008). DOI:[10.1002/jcc.20945](https://doi.org/10.1002/jcc.20945).
38. Jo, S., Lim, J. B., Klauda, J. B., & Im, W. CHARMM-GUI Membrane Builder for mixed bilayers and its application to yeast membranes. *Biophys J* **97**, 50-58, (2009). DOI: [10.1016/j.bpj.2009.04.013](https://doi.org/10.1016/j.bpj.2009.04.013).
39. Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: visual molecular dynamics. *J mol graph* **14**, 33-38 (1996). DOI:[10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
40. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kalé, L. & Schulten, K. Scalable molecular dynamics with NAMD. *J Comput chem* **26**, 1781-1802, (2005). DOI:[10.1002/jcc.20289](https://doi.org/10.1002/jcc.20289).
41. Kuriata, A., Gierut, A. M., Oleniecki, T., Ciemny, M. P., Kolinski, A., Kurcinski, M., & Kmiecik, S. CABS-flex 2.0: a web server for fast simulations of flexibility of protein structures. *Nucleic Acids Res* **46**, W1, W338-W343 (2018). DOI:[10.1093/nar/gky356](https://doi.org/10.1093/nar/gky356).
42. Frappier, V., Chartier, M., & Najmanovich, R. J. ENCoM server: exploring protein conformational space and the effect of mutations on protein function and stability. *Nucleic Acids Res* **43**, W1, W395-W400, (2015). DOI: [10.1093/nar/gkv343](https://doi.org/10.1093/nar/gkv343).
43. Task force COVID-19 del Dipartimento Malattie Infettive e Servizio di Informatica, Istituto Superiore di Sanità - Epidemia COVID-19. Aggiornamento nazionale. 16 aprile 2020. https://www.epicentro.iss.it/coronavirus/bollettino/Bollettino-sorveglianza-integrata-COVID-19_16-aprile-2020.pdf
44. Shim, E., Tariq, A., Choi, W., Lee, Y. & Chowell, G. Transmission potential and severity of COVID-19 in South Korea. *Int J Infect Dis* **93**, P339-344, (2020). DOI: [10.1016/j.ijid.2020.03.031](https://doi.org/10.1016/j.ijid.2020.03.031).
45. Carrel, L. & Willard, H.F. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* **434**, 400-404, (2005). DOI:[10.1038/nature03479](https://doi.org/10.1038/nature03479).
46. Talebizadeh, Z., Simon, S.D. & Butler, M.G. X chromosome gene expression in human tissues: male and female comparisons. *Genomics* **88**, 675-681, (2006). DOI: [10.1016/j.ygeno.2006.07.016](https://doi.org/10.1016/j.ygeno.2006.07.016).
47. Alifano, M., Alifano P. & Forgez P, Iannelli A. Renin-angiotensin system at the heart of COVID-19 pandemic. *Biochimie* **174**, 30-33 (in press). DOI: [10.1016/j.biochi.2020.04.008](https://doi.org/10.1016/j.biochi.2020.04.008).
48. Gheblawi, M. et al. Angiotensin converting enzyme 2: SARS-CoV-2 receptor and regulator of the renin-angiotensin system. *Circ Res*, (2020). DOI:[10.1161/CIRCRESAHA.120.317015](https://doi.org/10.1161/CIRCRESAHA.120.317015).

49. Kuba, K. et al. A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus-induced lung injury. *Nat Med* **11**, 875-879, (2005). DOI:[10.1038/nm1267](https://doi.org/10.1038/nm1267).
50. Chan, J.F.W. et al. Simulation of the Clinical and Pathological Manifestations of Coronavirus Disease 2019 (COVID-19) in Golden Syrian Hamster Model: Implications for Disease Pathogenesis and Transmissibility. *Clin Infect Dis pii: ciaa325*, (2020b). DOI:[10.1093/cid/ciaa325](https://doi.org/10.1093/cid/ciaa325).
51. McCray, P.B. Jr et al. Lethal infection in K18-hACE2 mice infected with SARS-CoV. *J Virol* **81**, 813–821 (2007). DOI: [10.1128/JVI.02012-06](https://doi.org/10.1128/JVI.02012-06).
52. Tseng, C.T et al. SARS coronavirus infection of mice transgenic for the human angiotensin-converting enzyme 2 (hACE2) virus receptor. *J Virol* **81**, 1162–1173, (2007). DOI: [10.1128/JVI.01702-06](https://doi.org/10.1128/JVI.01702-06)
53. Yang X.H. et al. Mice transgenic for human angiotensin-converting enzyme 2 provide a model for SARS coronavirus infection. *Comp Med* **57**, 450-459, (2007).
54. Kawase, M., Shirato, K., van der Hoek, L., Taguchi, F. & Matsuyama, S. Simultaneous treatment of human bronchial epithelial cells with serine and cysteine protease inhibitors prevents severe acute respiratory syndrome coronavirus entry. *J Virol* **86**, 6537-6654, (2012) DOI:[10.1128/JVI.00094-12](https://doi.org/10.1128/JVI.00094-12)
55. Zhou, Y. et al. Protease inhibitors targeting coronavirus and filovirus entry. *Antiviral Res* **116**, 76-84 (2015). DOI:[10.1016/j.antiviral.2015.01.011](https://doi.org/10.1016/j.antiviral.2015.01.011)
56. Rose, P. W. et al. The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Res* **45**, D1, D271-D281, (2017). DOI:[10.1093/nar/gkw1000](https://doi.org/10.1093/nar/gkw1000)
57. Sherry, S. T. et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* **29**, 308-311, (2001). DOI:[10.1093/nar/29.1.308](https://doi.org/10.1093/nar/29.1.308)
58. Bhagwat, M. Searching NCBI's dbSNP database. *Curr Protoc Bioinformatics* **32**, 1-19. DOI: [10.1002/0471250953.bi0119s32](https://doi.org/10.1002/0471250953.bi0119s32)
59. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic. Acids Res* **45**, D158–D169, (2017). DOI:[10.1093/nar/gkw1099](https://doi.org/10.1093/nar/gkw1099).
60. Pettersen, E. F., et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612 (2004).DOI:[10.1002/jcc.20084](https://doi.org/10.1002/jcc.20084)
61. Hammer, Ø., Harper, D. A., & Ryan, P. D. PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia electronica*, **4**, 9, (2001).

62. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., & Wolfson, H. J. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res* **33**, W363-W367 (2005). DOI:[10.1093/nar/gki481](https://doi.org/10.1093/nar/gki481)
63. Tovchigrechko, A. & Vakser, I. A. GRAMM-X public web server for protein–protein docking. *Nucleic Acids Res* **34**, W310-W314 (2006). DOI:[10.1093/nar/gkl206](https://doi.org/10.1093/nar/gkl206)
64. Kiefer, F. et al. The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res* **37**, D387-D392 (2009). DOI:[10.1093/nar/gkn750](https://doi.org/10.1093/nar/gkn750)
65. Yang, J. et al. The I-TASSER Suite: protein structure and function prediction. *Nat Methods* **12**, 1, 7 (2015). DOI: [10.1038/nmeth.3213](https://doi.org/10.1038/nmeth.3213)
66. Almagro Armenteros, J.J. et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* **37**, 420-423 (2019). DOI:[10.1038/s41587-019-0036-z](https://doi.org/10.1038/s41587-019-0036-z)

Acknowledgements

We wish to thank prof. Diane Damotte (University of Paris) for advice and critical reading of the manuscript.

Author contributions

M.A., P.A.: conception, coordination, designing, writing.

M.C.: experimental set-up, pipeline development, *in silico* analysis;

P.F., A.I., designing, data providing;

Competing interests

The authors declare no competing interests.

Materials & Correspondence

The authors declare no competing interests.

Figure1

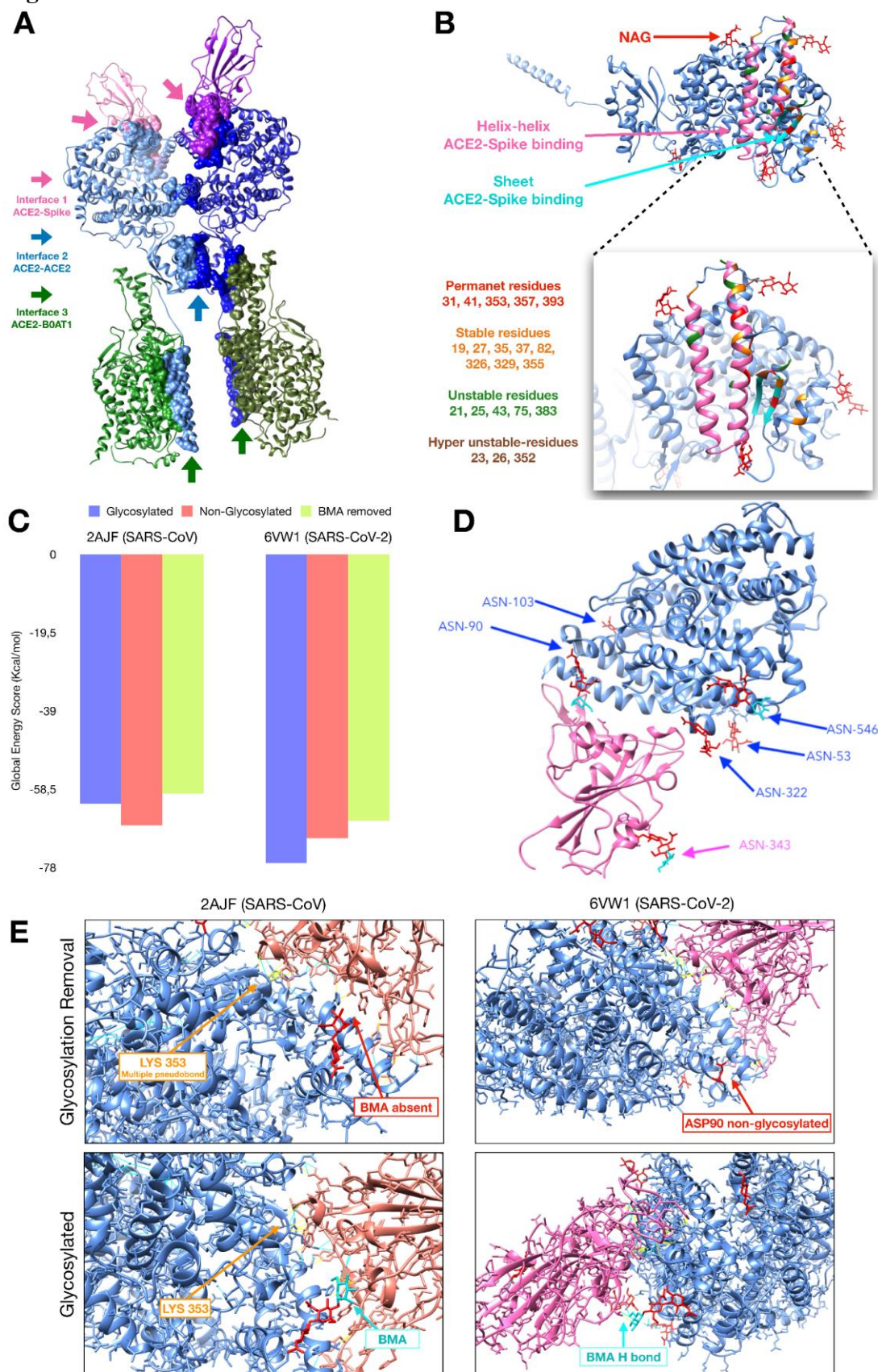


Figure 2

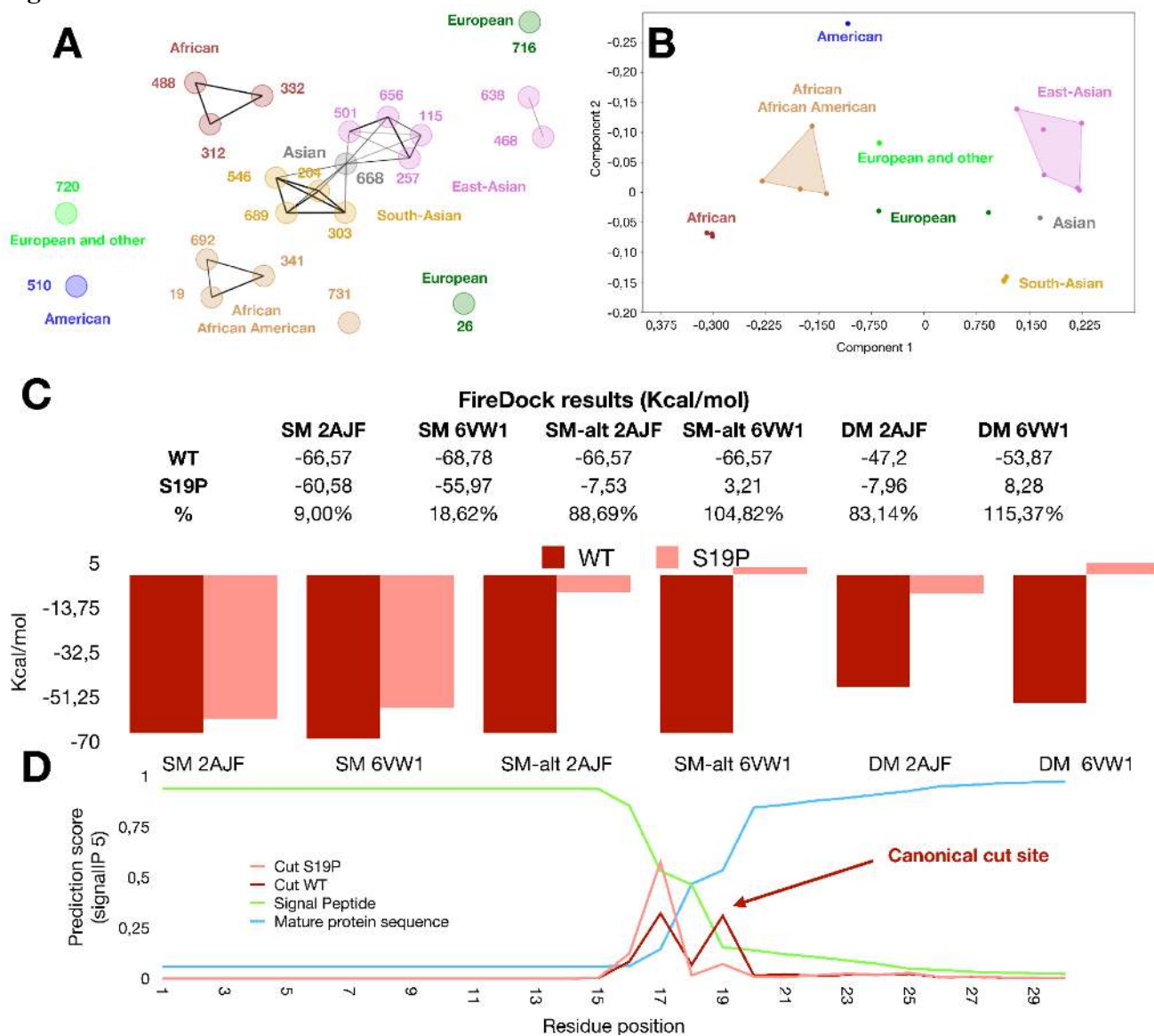


Figure 3

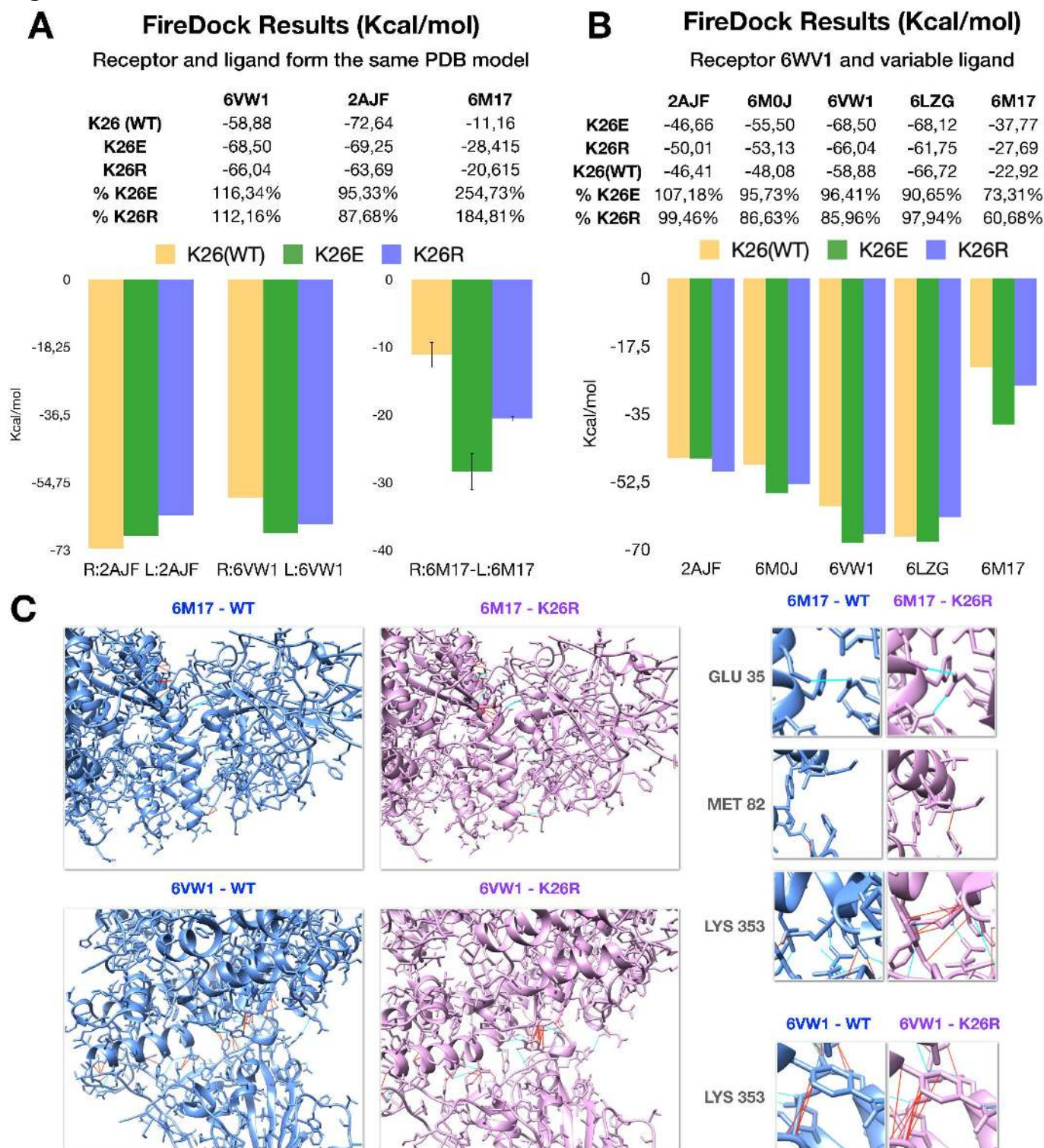


Figure 4

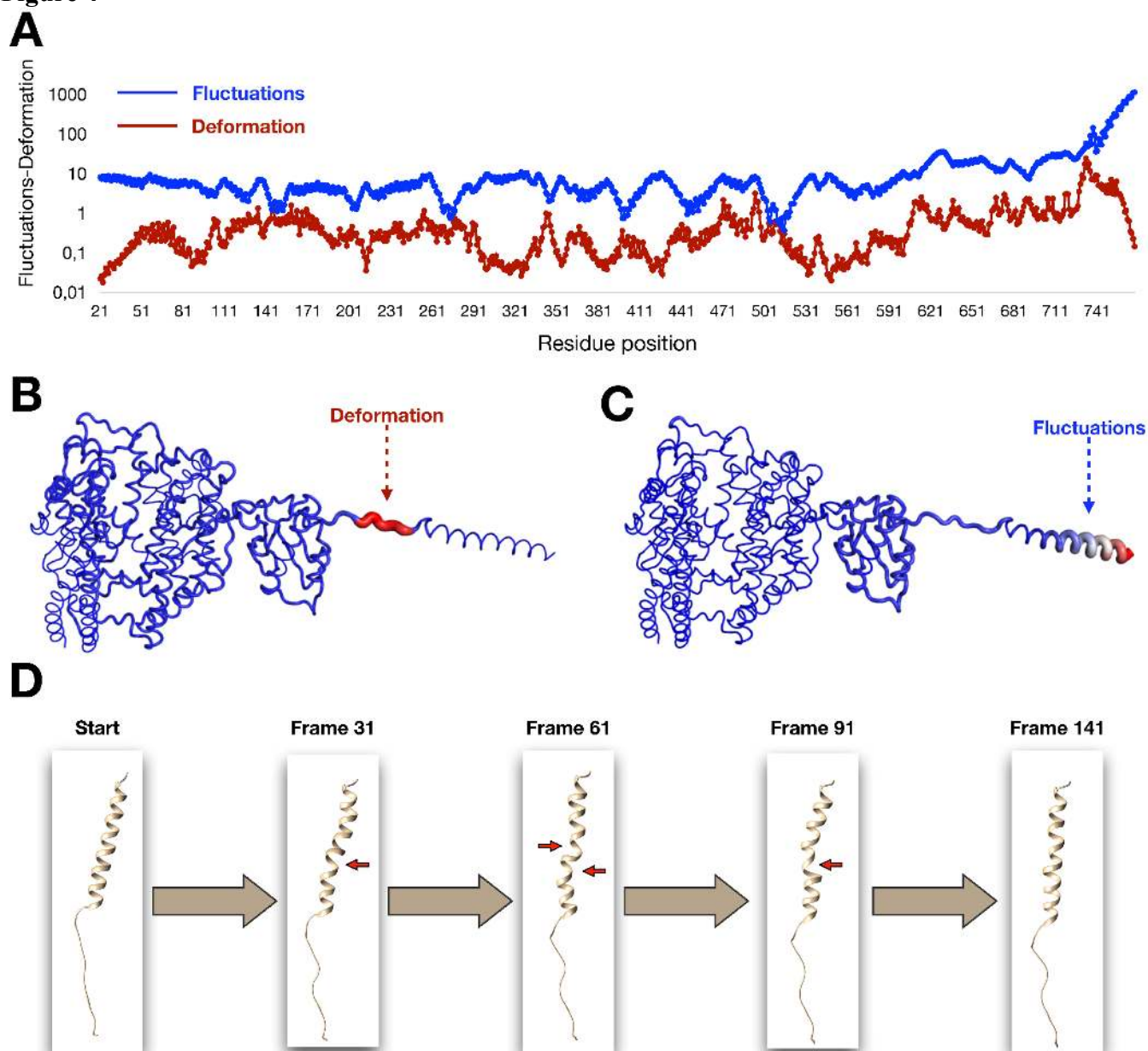


Figure 5

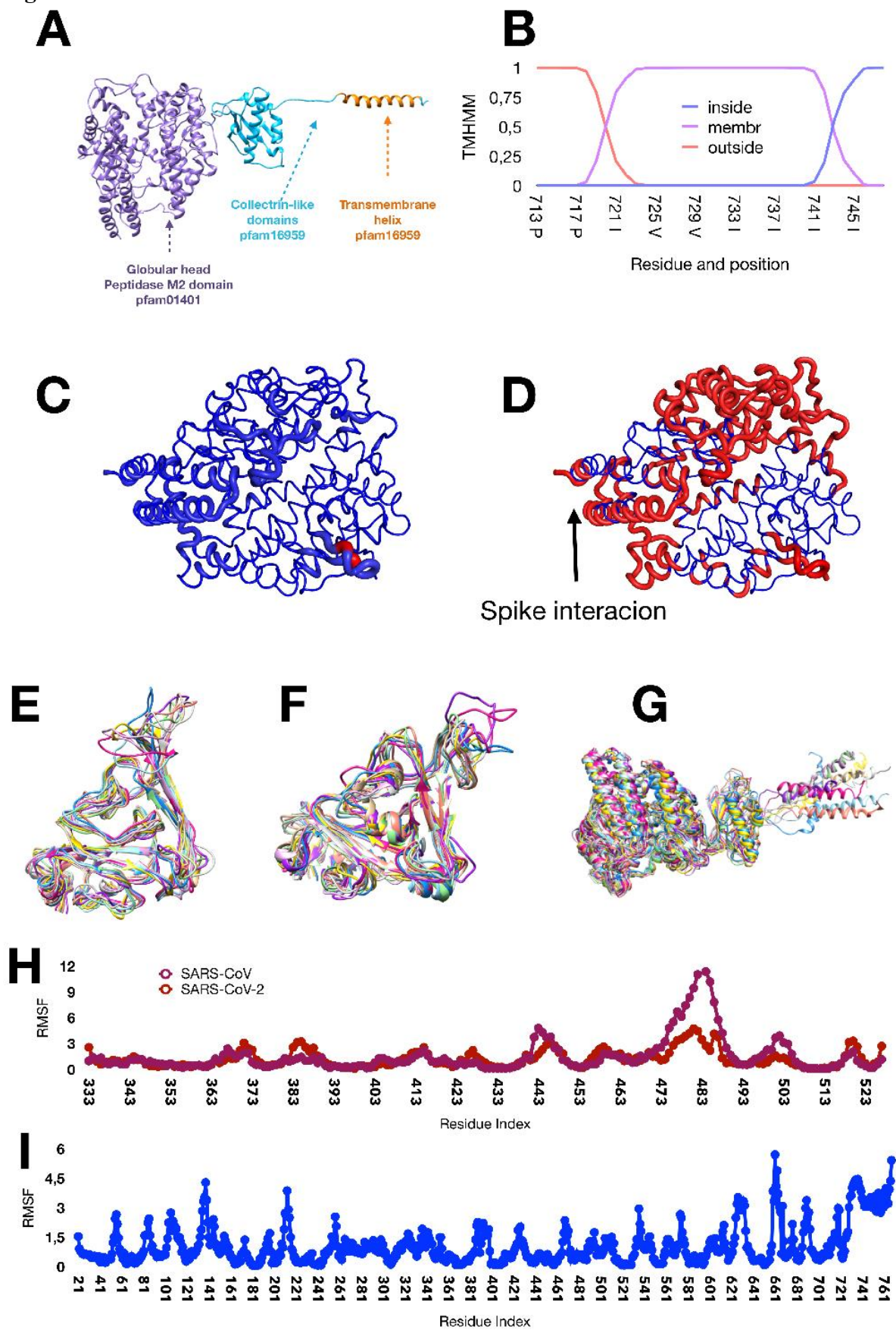
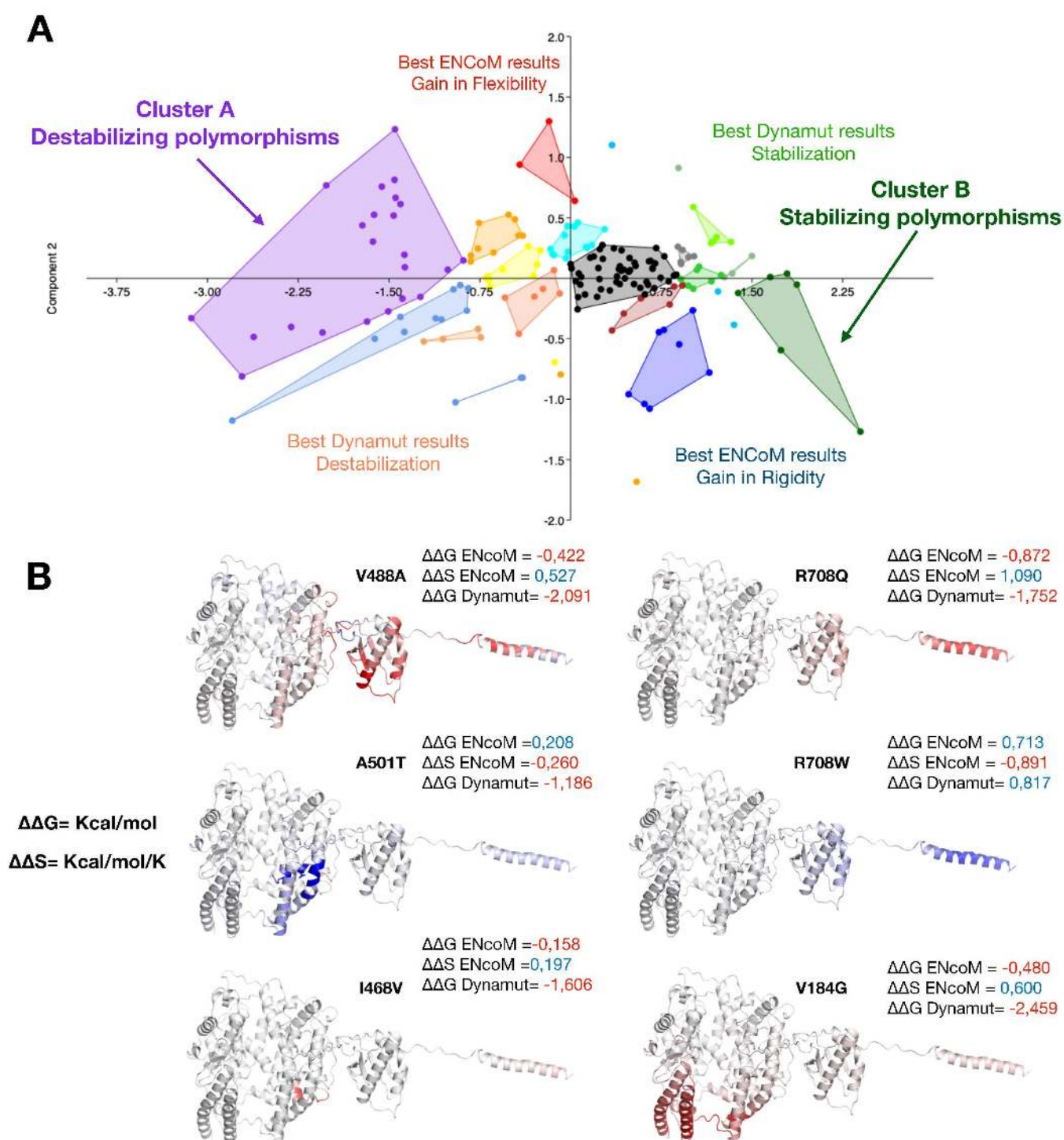


Figure 6



Legends to Figures

Fig. 1. 3D models of SARS-CoV-2 Spike/ACE2 complex, and effects of ACE2 glycosylation on binding interfaces. **a** 3D model of SARS-CoV-2 Spike/ACE2/B0AT1 complex. The arrows indicate the principal binding interfaces between all proteins involved in this complex. **b** Domains and ACE2 amino acid residues involved the interaction with Spike. “Permanent” (red), “stable” (orange), “unstable” (green), and “hyper-unstable” residues are indicated. **c** FireDock was used to estimate ΔG values by using 2AJF (SARS-CoV Spike/ACE2 complex) and 6VW1 (SARS-CoV-2 Spike/ACE2 complex) after removal of the glycosydic chains (red bars) or the terminal beta-mannose (BMA) (green bars). **d** 3D 6VW1 model with highlighting glycosydic chains (red, N-acetyl-glucosamine [NAG]; cyan, BMA). **e** Predicted effects of removal of the glycosydic chains or the BMA in 2AJF and 6VW1 models. In 6VW1, the BMA forms two H-bonds and one pseudo-bond, while these bonds are lost in non-glycosylated model. In contrast, in the 2AJF, the glycosydic chain forms only one H-bond and, after removing BMA the Thr-41 has more grads of binding.

Fig. 2. Selected ACE2 SNPs affecting binding interfaces: S19P. **a** Network plot of the most diffused ACE2 SNPs causing missenses worldwide. **b** Non-Metric Multidimensional Scaling (NM-MDS, Bray-Curtis index) the most diffused ACE2 SNPs. **c** FireDock results predicting the effects of ACE2 S19P amino acid replacement on ACE2/Spike interaction in 2AJF (SARS-CoV Spike/ACE2), and 6VW1 (chimeric SARS-CoV/SARS-CoV-2 Spike RBD /ACE2) models. Results predicting the effects of ACE2 S19P with the alternative (ALT) N-terminal cleavage site (**d**) are also shown in either 2AJF and 6VW1 static models (SM-alt), or 2AJF and 6VW1 dynamic (docking) models (DM). **d** Predicted effect of ACE2 S19P on ACE2 N-terminal cleavage site.

Fig. 3. Selected ACE2 SNPs affecting binding interfaces: K26R/E. **a** FireDock results predicting the effects of ACE2 K26R and K26E amino acid replacements on ACE2/Spike interaction in 2AJF (SARS-CoV Spike/ACE2), and 6VW1 (chimeric SARS-CoV/SARS-CoV-2 Spike RBD /ACE2) and 6M17 (SARS-CoV-2 Spike RBD /ACE2/B0AT1). **b** FireDock results that were obtained by challenging the ACE2 structure from 6VW1 with the Spike structures that were generated by the different models as shown in the bottom of the histogram. **c** Possible effects of K26R substitutions on H-bonds (cyan) and pseudo-bonds (red) from 6M17 (upper images) and 6VW1 (upper panels) models as illustrated (larger panels: left, wild type ACE2; right, K26R ACE2). Smaller panels on the right are magnifications of regions the respective panels on the left as indicated.

Fig. 4. Dynamut models of ACE2. **a** Score for fluctuation and deformation of ACE2 calculated by Dynamut. **b** ACE2 region that is mostly subject to deformation. **c** ACE2 region that is mostly subject to fluctuation. **D**) Distortion of the hydrophobic transmembrane domain during Molecular Dynamic Simulation.

Fig. 5. Structural and dynamic features of ACE2, SARS-CoV-2 and SARS-CoV Spike proteins as inferred by Dynamut and CABSflex. **a** Conserved domains of ACE2. **b** Transmembrane region of ACE2 identified by TMHMM. **c-d** Dynamut simulations of ACE2 structure by using the 6WV1 PDB model (without the transmembrane domain) showing that some residues of the ACE2 interface that bind the Spike protein can fluctuate. **e-g** CABSflex simulations of SARS-CoV (**e**), SARS-CoV-2 Spike (**f**), and ACE2 (**g**). **h** CABSflex fluctuation scores of SARS-CoV Spike and SARS-CoV-2 Spike. **i** CABSflex fluctuation scores of ACE2.

Fig. 6. ACE2 SNPs analyzed by Dynamut and ENCoM. **a** Ordination plot (PCA) obtained with PAST to evaluate the results of Dynamut and ENCoM. This set of data includes all SNPs by dbSNP database, excluding SNPs in contact interfaces. **b** Effects on structure flexibility calculated as $\Delta\Delta G$ and $\Delta\Delta S$ variations between ACE2 and some polymorphic variants (red, gain in flexibility, blue, gain in rigidity). Illustrated SNPs were selected as the most common polymorphisms that gave the highest $\Delta\Delta G$ and $\Delta\Delta S$ variations.

Supplementary Figure 1

ACE2 ALIGNMENT of the sequences from PDB models

```

6M17          --IEEQAKTFLDKFNHEAEDLFYQSSLASWNYNTNITEENVQNMNAGDKWSAFLKEQST      58
6WV1/1AJF/6LZG/6M0J          STIEEQAKTFLDKFNHEAEDLFYQSSLASWNYNTNITEENVQNMNAGDKWSAFLKEQST      60
          *****

6WV1/1AJF/6LZG/6M0J/6M17     LAQMYPLQEIQNLTVKLQLQALQQNGSSVLSSEDKSKRLNTILNTMSTIYSTGKVCNPDNP      118/120
          *****

6WV1/1AJF/6LZG/6M0J/6M17     QECLLLEPGLNEIMANSLDYNERLWAWESWRSEVGKQLRPLYEEYVVLKNEMARANHYED      178/180
          *****

6WV1/1AJF/6LZG/6M0J/6M17     YGDYWRGDYEVNGVDGYDSRGLIEDVEHTFEEIKPLYEHLHAYVRAKLMNAYPSYISP      238/240
          *****

6WV1/1AJF/6LZG/6M0J/6M17     IGCLPAHLLGDMWRFWNTNLYSLTVFPFQKPNIDVTDAMVDQAWDAQRIKFAEKFFVSV      298/300
          *****

6WV1/1AJF/6LZG/6M0J/6M17     GLPNMTQGFWENSMLTDPGNVQKAVCHPTAWDLGKDFRILMCTKVTMDDFLTAHHEMGGH      358/360
          *****

6WV1/1AJF/6LZG/6M0J/6M17     IQYDMAAAQPFLLRNGANEGFHEAVGEIMSLSAATPKHLKSIGLLSPDFQEDNETEIFN      418/420
          *****

6WV1/1AJF/6LZG/6M0J/6M17     LLKQALTIVGTLPFTYMLEKWRWVFKGEIPKDWKMKWEMKREIVGVVEPVPHDETYC      478/480
          *****

6WV1/1AJF/6LZG/6M0J/6M17     DPASLFHVSNDSYFIRYYTRTLYQFQFQEAALCQAAKHEGPLHKCDISNSTEAGQKLFNML      538/540
          *****

6M17          RLKSEPWTLALENVVGAKNMNRPLLNYFEPLFTWLKDNKNSFVGWSTDWSPYADQSI      598
6WV1          RLKSEPWTLALENVVGAKNMNRPLLNYFEPLFTWLKDNKNSFVGWSTDWSPYA----      596
2AJF         RLKSEPWTLALENVVGAKNMNRPLLNYFEPLFTWLKDNKNSFVGWSTDWSPYAD---      597
6LZG         RLKSEPWTLALENVVGAKNMNRPLLNYFEPLFTWLKDNKNSFVGWSTDWSPYA----      596
6M0J         RLKSEPWTLALENVVGAKNMNRPLLNYFEPLFTWLKDNKNSFVGWSTDWSPYAD---      597
          *****

6M17          KVRISLKSALGDKAYEWNDEMFLFRSSVAYAMRQYFLKVKQMILFGEEDVRVANLKPR      658
6M17          ISFNFFVTAPKNVSDIIPRTEVEKAIRMSRSRINDAFRLNDSLEPLGIQPTLGGPPNQPP      718
6M17          VSIWLVFVGVVMGIVVGVIVLIFGTGIRDR 748
  
```

SPIKE alignment of the sequences from PDB models

```

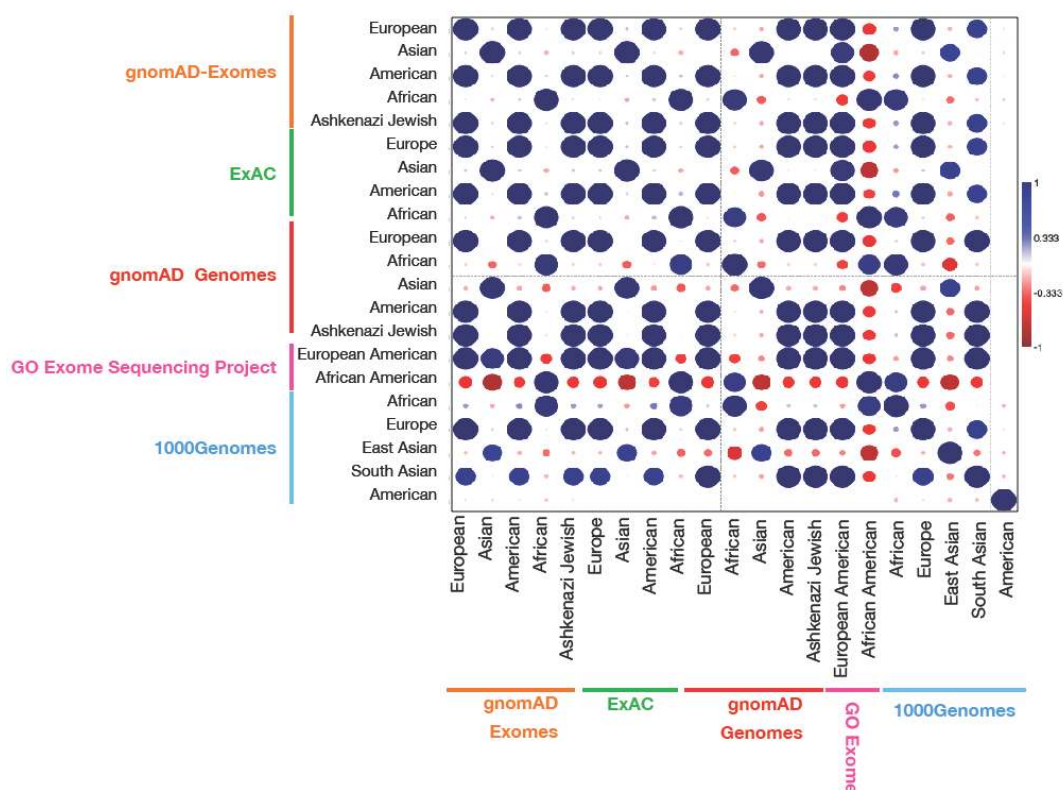
2AJF          ---CPFGVEFNATKFPVAVYAWERKKISNCVADYSVLYNSTFFSTFKCYGVSATKL----- 52
6WV1          -NLCPFGEVFNATKFPVAVYAWERKKISNCVADYSVLYNSTFFSTFKCYGVSATKLNDLCF 59
6LZG         TNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNLDLCF 60
6M0J         TNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNLDLCF 60
6M17         ---CPFGVEFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNLDLCF 57
          *****:* *****:**:*****:*****:***** **

2AJF          -NVYADSFVVKGDDVRQIAPGQTGVIADYNYKLPDDFMGCVLAWNTRNIDATSTGNYNYK 111
6WV1         SNVYADSFVVKGDDVRQIAPGQTGVIADYNYKLPDDFMGCVLAWNTRNIDATSTGNYNYK 119
6LZG         TNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNLDKSVGGNYNYL 120
6M0J         TNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNLDKSVGGNYNYL 120
6M17         TNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNLDKSVGGNYNYL 117
          *****:***:***** ***** ***** **:***:*:*: *****

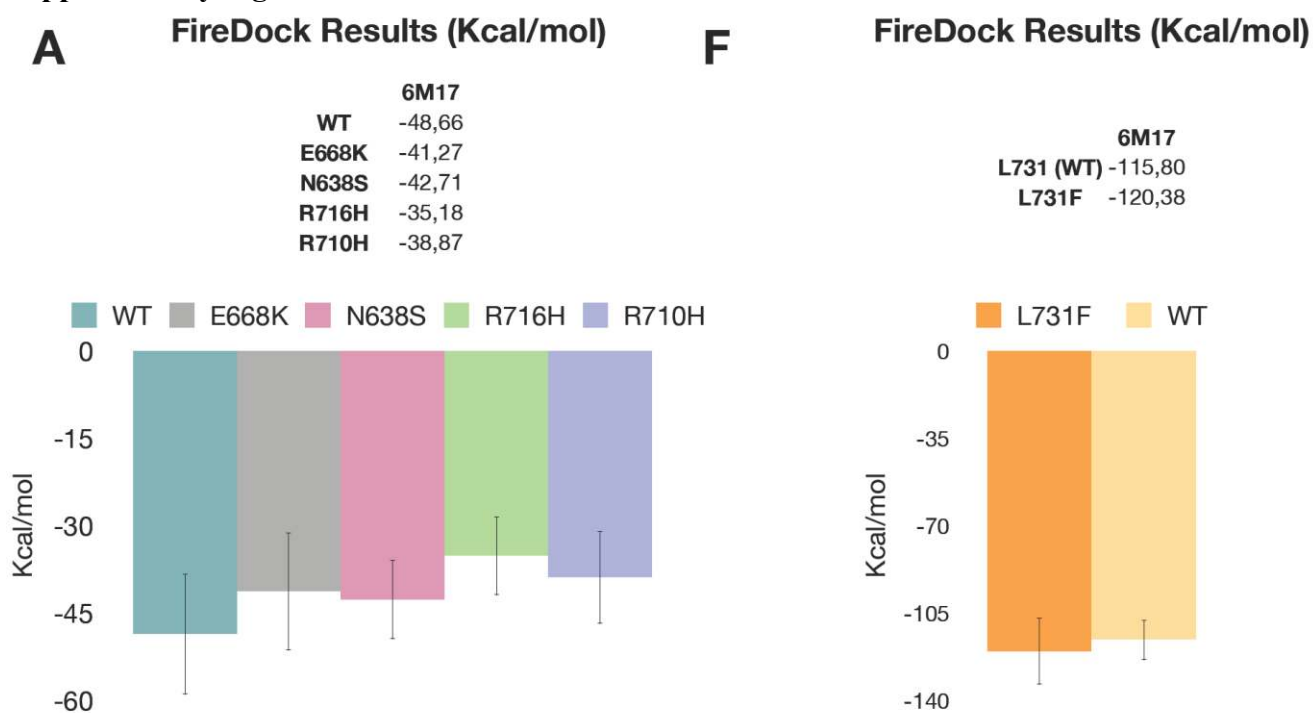
2AJF          YRYLRHGKLRPFERDISNVFSPDGKPCPTP-PALNCYWPLNDYGFYTTTGGYQPYRVVV 170
6WV1         YRLFRRKSNLKPFFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV 179
6LZG         YRLFRRKSNLKPFFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRV-V 179
6M0J         YRLFRRKSNLKPFFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV 180
6M17         YRLFRRKSNLKPFFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV 177
          ** :*:*:*:*****. :. ....* .:*****:***** *:*:***** *

2AJF          LSFEL----- 174
6WV1         LSFELLNAP-TVCGP 193
6LZG         LSFELHAPATVCGP 194
6M0J         LSFELHAPATVCG- 194
6M17         LSFELL----- 183
          *****
  
```

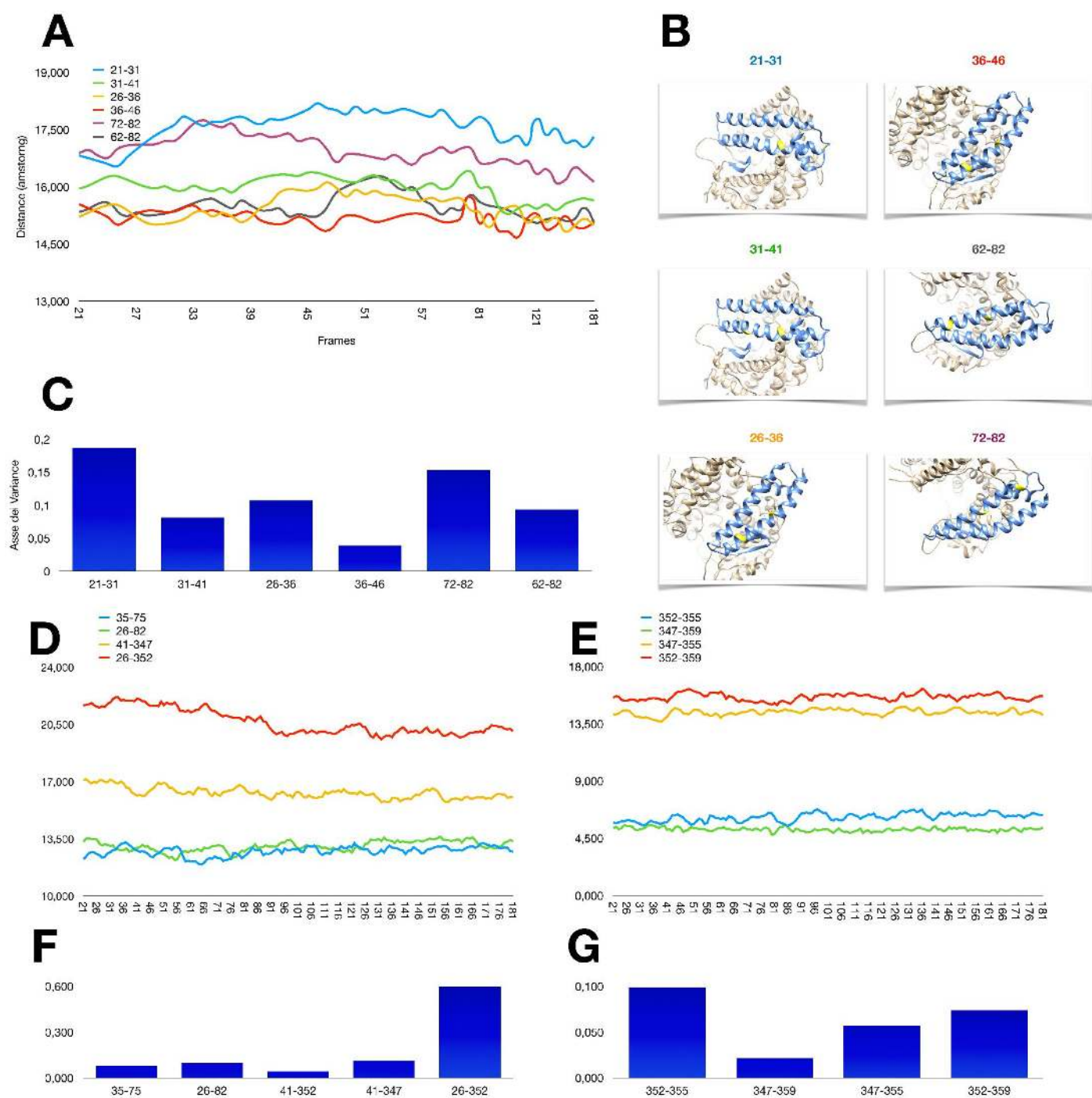
Supplementary Figure 2



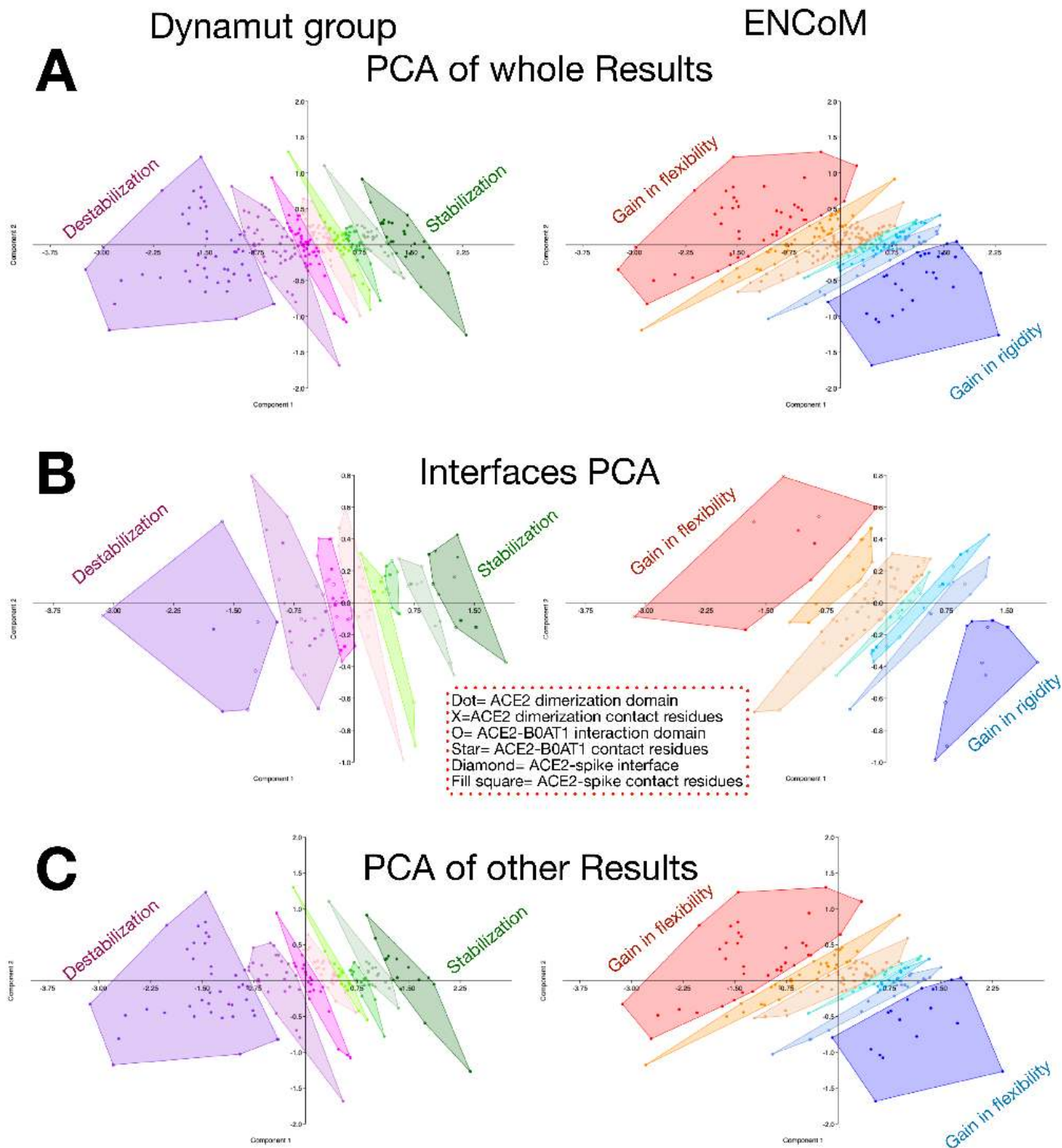
Supplementary Figure 3



Supplementary Figure 4



Supplementary Figure 5



Legends to Supplementary Figures

Supplementary Fig. 1. ClustalO alignments. ClustalO alignments of human ACE2 amino acid sequences, and SARS-CoV, SARS-CoV-2 and chimeric SARS-CoV/SARS-CoV-2 Spike RBDs used for X-ray cristallography models

Supplementary Fig. 2. ACE2 SNPs data from different databases. Correlation analysis of ACE2 SNPs data from different databases is illustrated.

Supplementary Fig. 3. Selected ACE2 SNPs affecting binding interfaces. FireDock results predicting the effects of several ACE2 amino acid replacements on ACE2/ACE2 (left) and ACE2/B0AT1 (right) interaction.

Supplementary Fig. 4. Oscillation of amino acid residues in ACE2 binding interfaces. **a** Oscillation plot of the distance between ACE2 amino acid residues mapping in the two helices that are involved in binding with Spike proteins. **b** 3D images of regions containing the amino acid residues shown in panel **a**. **c** Variance of the distance between ACE2 amino acid residues mapping in the two helices that are involved in binding with Spike proteins. **d-e** Oscillation plot of the distance between the two helices and beta-sheet (**d**), and between the residues of the beta-sheet (**e**). **f-g** Variance of the distance of amino acid residues between the two helices and beta-sheet (**f**), and between the residues of the beta-sheet (**g**).

Supplementary Fig. 5. Ordination plots of Dynamut and ENCoM results. **a** Ordination plot (PCA) obtained using PAST to evaluate the results of Dynamut (left panels) and ENCoM (right panles) outputs. Clustering involved 197 SNPs and was based on $\Delta\Delta G$ and/or $\Delta\Delta S$ values comparing ACE2 and each polymorphic variant. **b** Ordination plot (PCA) showing the results of SNPs mapping within the ACE2 interfaces (ACE2/SARS-CoV-2 Spike, ACE2/ACE2, ACE2/B0AT1). **c** Ordination plot (PCA) showing the results of SNPs Ordination plot (PCA) showing the results of SNPs that mapping outside the ACE2 interfaces.

Supplementary Tables (Excel File)

Supplementary Table 1. ACE2 PDB models and glycosylation.

Supplementary Table 2. EVOEF results.

Supplementary Table 3. SSIPe results.

Supplementary Table 4. ACE2 SNPs metadata.

Supplementary Table 5. GrammX-FireDock results.

Supplementary Table 6. PatchDock-FireDock results.

Supplementary Table 7. ENCoM and Dynamut results.