Probing the mutational landscape of the SARS-CoV-2 spike protein via quantum mechanical modeling of crystallographic structures

Marco Zaccaria[1¶], Luigi Genovese[2¶*], William Dawson[3], Viviana Cristiglio[4], Takahito Nakajima[3], Welkin Johnson[1], Michael Farzan[4], and Babak Momeni[1*]

[1] Department of Biology, Boston College, Chestnut Hill MA, United States
[2] Université Grenoble Alpes, CEA, INAC-MEM, L_Sim, Grenoble, France
[3] RIKEN Center for Computational Science, Kobe, Japan
[4] Institute Laue Langevin, Grenoble, France
[5] Department of Immunology and Microbiology, The Scripps Research Institute, Jupiter FL, United States

[*] Corresponding authors; luigi.genovese@cea.fr and momeni@bc.edu
[¶] Equal contribution

Abstract

We employ a recently developed quantum mechanical approach (QM-CR), based on complexity reduction of Density Functional Theory calculations, to characterize the interactions of the SARS-CoV-2 spike Receptor Binding Domain (RBD) with ACE2 host receptors and antibodies. QM-CR operates via *ab initio* identification of individual amino acid residues contributions to chemical binding and leads to the identification of the impact of point mutations. Here, we especially focus on the E484K mutation of the viral spike protein. We find that spike residue 484 hinders the spike's binding to the human ACE2 receptor (hACE2). In contrast, the same residue is beneficial in binding to the bat receptor *Rhinolophus macrotis* ACE2 (macACE2). In agreement with empirical evidence, QM-CR shows that the E484K mutation allows the spike to evade categories of neutralizing antibodies like C121 and C144. The simulation also shows how the Delta variant spike binds more strongly to hACE2 compared to the original Wuhan strain, and predicts that a E484K mutation can further improve its binding. Broad agreement between the QM-CR predictions and experimental evidence supports the notion that *ab initio* modeling has now reached the maturity to handle large intermolecular interactions central to biological processes.

Significance Statement

The threat of emerging pathogens, exemplified by the rapid spread of SARS-CoV-2, has motivated investigations into how pathogens may evolve. In surveying possible evolutionary trajectories, wet-bench screens can only sample a small fraction of possibilities because of practical limitations. Mechanistic modeling can partially overcome these limitations by offering: (1) flexibility of *in silico* sampling and (2) insights about underlying interaction mechanisms. Here, we employ a complexity reduction quantum mechanical (QM-CR) approach to describes the intermolecular interactions at the amino acid level. Through this approach, we uncover residues critical to spike-receptor and spike-antibody interactions. We find broad agreement between the QM-CR predictions and experimental evidence, showcasing the ability of *ab initio* modeling to capture biologically-relevant intermolecular interactions.

Introduction

Since SARS-CoV-2 infected the human host, several variants have arisen [1] with distinct changes in the viral spike protein, particularly in the Receptor Binding Domain (RBD). Two trends have been prevalent in the spike evolution: i) selection towards improved binding to host cells [2]; and ii) selection towards evasion of neutralizing antibodies (nAbs) [3–6]. Anticipating the evolutionary trajectory of viruses is a long-established relevant topic in the scientific community [7]. Presently, the main approach in this direction is high-throughput *in vitro* screening of mutants (e.g. [8, 9]); however, such an approach does not directly identify the mechanisms that make a given mutation more, or less, beneficial. In this work, we show how the recent developments in *ab initio* modeling can complement experimental results and offer detailed mechanistic insights.

Traditionally, full QM models of intermolecular interactions are only employed for small molecules of about a hundred atoms [10, 11]; larger molecules have proven computationally challenging for full QM investigations. Nevertheless, *in silico* approaches alternative to full QM have been successful. Molecular docking [12–14], relying on geometrical constraints to assess intermolecular interactions, has been used to survey small-molecule candidates in drug discovery [15]. Force-fields (FFs) have also been successful [16, 17], whenever previous adequate parameterization is available [18]. Hybrid quantum mechanics/molecular mechanics (QM/MM) methods are also common in describing enzyme-substrate systems [19], and have been successfully applied to SARS-CoV-2 [20, 21, 30, 22–29]. QM/MM uses quantum mechanical (QM) simulations for a small portion of the system (tens of atoms) [31], leaving the remaining regions to be modeled with a less computationally demanding MM simulation, driven by FFs.

To mechanistically characterize SARS-CoV-2 spike-receptor and spike-nAb interactions, we apply a recently developed approach for large scale electronic structure calculations: Complexity Reduction in Density Functional Theory calculations [32, 33], hereafter called QM-CR. QM-CR differs from previous approaches in requiring no targeted parameterization or prior knowledge about the nature or sites of interactions, and it is based on full QM calculations on the entire system. QM-CR leverages recent progress in computational chemistry [32, 34] to handle tens of thousands of atoms in a single simulation. This enables us to capture and investigate biological processes involving several hundreds of amino acids, including the SARS-CoV-2 spike interactions. Recent efforts on SARS-CoV-2 have generated structural and biochemical data that can be used to validate QM-CR predictions. In particular, the high level of detail from recent contributions gives us new insight to complement experimental data or analysis based on regression models [35–38].

Importantly, QM-CR can reveal the mechanisms behind intermolecular binding by decomposing interactions into chemical/short-ranged (which imply a shared electron) versus electrostatic/long-ranged (which do not involve shared electrons). We define as "hotspots" amino acids with a significant chemical contribution to the intermolecular interactions. To further investigate the contribution of individual amino acids, single point mutations can be introduced into a protein's (e.g., the spike) primary structure. We employ the BigDFT computer program [39], based on an *ab initio* Density Functional Theory approach on a set of fully atomistic 3D structural models, to simulate intermolecular interactions of interest with a computational cost manageable on modern supercomputers.

In this work, we focus our analysis on the E484K mutation for three main reasons. First, our analysis identifies residue E484 as the main interface weak link in the interaction of the SARS-CoV-2 Wuhan strain with the human receptor ACE2 (hACE2); conversely the same residue is beneficial to binding the bat *Rhinolophus macrotis'* ACE2 (macACE2). Second, we show that an E484K mutation alone can disrupt the neutralizing effect of specific antibodies. In addition, we also highlight the strong modular character of the E484K mutation and show that, if imposed on existing SARS-CoV-2 variants such as Delta, it can enhance binding to hACE2, potentially identifying future viral evolutionary trajectories. Finally, we argue that *ab initio* models are now at the point of providing mechanistic insights on molecular interactions central to biological processes.

Results

We focus our analysis on the impact of the E484K mutation on antibody evasion and cellular

receptor binding. Prior experimental and computational data have shown that spike variants, presenting the E484K mutation in the RBD, can evade antibodies C144 and C121 [38, 40]. E484K is also a typical signature mutation of the RBD of the Gamma and Beta variants. We test our QM model as an agnostic predictor to explain the interaction of the viral spike (the original Wuhan version or the E484K-mutated one) with host receptors and nAbs.

QM-CR underscores hotspots of spike-hACE2 interactions

We examine the interaction between the WT spike RBD and hACE2 as its native substrate. In this analysis, we calculate the overall effect of each amino acid residue on its respective interactor, either on the spike side or on the hACE2 side; the contribution to the binding energy can either be attractive/stabilizing or repulsive/de-stabilizing (Fig 1).

We use Fragment Bond Order (FBO) [32], calculated using the electronic structure of the system in proximity of a given residue, as a measure of the strength of the interaction in the proximity of the interface between the two interacting molecules (Table 1). In Fig 1, we have highlighted residues with large FBO as well as those close to the geometric interface. Residues with both large FBO and interface proximity are determined as major contributors to the intermolecular interactions. This analysis reveals the contribution of each residue to the overall binding performance, highlighting which amino acids facilitate or hinder binding, and how. In the following sections, we use FBO to draw an interaction network of the interface to detail the chemical interactions among residues, and their stabilizing or destabilizing role. Details of the procedure are provided in the supplementary information ("Details of the fragmentation procedure"). As an alternative visualization, the contribution of each amino acid residue to the binding can also be highlighted over the 3D physical arrangement of the two molecules (Fig S1).

QM-CR identifies the spike E484 residue as the weak link in the binding to the host receptor hACE2

FBO values pinpoint the hotspots of the RBD-hACE2 system (Fig 2). On the hACE2 side (Fig 2, panel a), Q24, T27, D30, K31, H34, E35, E37, D38, Y41, Q42, Y83, and K353 stand out, in agreement with known data [41]. On the spike side (Fig 2, panel b), a more diverse layout emerges, on and off the interface, with several residues displaying repulsion. However, residue E484 shows the unique trait of being simultaneously repulsive *and* at the interface with hACE2, via a short-range interaction with the K31 residue (Figs 2 and S6). Since the chemical interaction is intrinsically attractive, the overall repulsive interaction indicates that another residue in the vicinity cancels the chemical attraction with an electrostatic repulsion. Overall, in the Wuhan

type structure, E484 destabilizes the binding to hACE2. From this analysis, we conclude that the Wuhan spike RBD harbors a sub-optimal residue at the 484 position for hACE2 binding.

To further investigate the impact of E484, we test the model on the available 3D crystal structure of the human homologous ACE2 receptor in *Rhinolophus macrotis*, a host species with a more adapted SARS-CoV-2 interaction [41]. In this simulation (Fig 2 panels a and b, second rows), the E484 residue is instrumental to the binding by being strongly attractive to the *R. macrotis* ACE2 (macACE2). Notably, in both hACE2 and macACE2, the interactor with E484 is the ACE2 residue K31. This means that the macACE2 sequence has residues, proximal to the K31 hotspot, that exert an attractive electrostatic force on E484. A closer inspection of the two sequences reveals that this attractive force comes from the K35 residue, which in hACE2 is replaced by Glutamic Acid. Thus, the model highlights a stark contrast between human and bat receptors.

We further confirmed the role of E484 by introducing the E484K mutation into the viral spike and then assessing the interaction with hACE2 (Fig S2). The E484K mutation improves the spike-hACE2 binding energy by about 32% (Fig S2, bottom right histograms), switching the main hACE2 interacting residue from K31 to E35. Such an interaction, driven by electrostatics, represents a net improvement of the Wuhan-hACE2 network. Conversely, the same mutation does not affect the spike binding energy to macACE2 in the same position, where the bat receptor hosts a lysine. In other terms, for macACE2, K484 clearly does not engage K35, and would actually disappear from the interface (Fig S2). The resulting interaction network is rearranged, and the interface binding energy is not improved by the mutation. Therefore, the model shows a more functional interaction between macACE2 and Wuhan RBD, possibly the result of a longer adaptation by SARS-CoV-2 to *R. macrotis,* compared to the human receptor. In the hACE2 interaction, the E484 spike residue belongs to a sub-optimal sector of the chemical interface, suggesting that other RBD adaptations in this sector are likely to improve the binding.

QM-CR shows how nAb C121 loses binding to the E484K mutated spike

We identify the hotspots between the Wuhan spike RBD and C121 nAbs (Figs 4 and S1) (see results for C144 nAb in Fig S3). Residue E484 is the main spike interactor with C121 nAb. Other relevant sites are residues K444, Y449, F486, Y489, and Q493. On the C121 side, residues Y33, S55, and S75 are pivotal for the Wuhan spike binding. The model estimates that among all the residues contributing to the interaction, the individual contribution of E484 amounts to around 50% of the total. The interaction network (Fig 4) shows E484's binding to residues Y33 and S55 of C121. Once the E484K mutation is imposed, we observe a rearrangement of the

interaction network, and a substantially lower binding energy, between the spike and the antibody. Specifically, E484K reduces the connectivity at the 484 residue in the interaction network, and modifies the interactions on the C121 side towards decreased stability. Only the S52 residue is stabilized by the mutation, but not to the point of compensating for the loss of attraction at other residues. Overall, once the mutation is applied, we observe a substantial decrease of about 25% of the total binding energy, largely because of reduced short-range interactions. The model concludes, with no *a priori* information other than the experimental crystal structure, that the E484 residue is the essential actor in the binding by nAb C121, and that a targeted point mutation will substantially affect said binding. The analysis of C144 nAb shows comparable results. Moreover, C144 undergoes a substantial rearrangement of its interaction network in response to the mutation, arguably a consequence of the original higher connectivity of the residue E484 in the binding, compared to the C121 case: five C144 residues are involved (Y51, S52, G53, G54 and S55) compared to two C121 residues (Y33 and S55) (Fig S3). Interestingly, the importance of E484 also appeared in previous a work in which E484 mutants arise under the selective pressure of nAbs [42].

QM-CR predicts that the E484K mutation strengthens the binding of the Delta spike to hACE2

Starting from the Wuhan strain crystal structure, we generate a virtual crystal structure to represent Delta (B.1.617.2) in conjunction with hACE2 by substituting its characterizing RBD mutations (L452R and T478K) into the Wuhan spike crystal structure. Such residue mutations belong to an off-interface sector of the RBD (see Fig 1). Our simulations identify the same FBO interface residues found for the Wuhan strain. However, differently from the other tested interaction networks, a substantial contribution to the overall binding energy of Delta to hACE2 comes from off-interface residues via their long-range electrostatic effect on their counterparts, highlighting the relevance of including residues beyond the interface region, in the analysis of binding.

Furthermore, when testing the binding of the Delta-hACE2 system after introducing the E484K mutation, the simulation shows that E484K is compatible with the Delta variant and further strengthens the overall binding to hACE2. This *in silico*-generated variant, solely based on theoretical grounds, displays a stronger binding to hACE2 than either E484K or Delta variants individually (Fig 5).

Discussion

Recently developed Complexity Reduction tools in Density Functional Theory calculations have allowed full QM simulations of systems with several thousands of atoms. These advances have bridged the gap that had so far hindered full QM *ab initio* modeling of larger molecules that are often of interest in biology. A computational approach that can capture biologically relevant intermolecular interactions, such as protein-protein interactions, has untapped potential for better mechanistic understanding of biological phenomena at a molecular level.

In this work, we use the BigDFT code to implement an *ab initio* QM simulation of the electronic properties of a given set of atoms as large as a full protein-protein system. Through this model, we decompose the interaction between two biological macro-molecules, spike RBD and receptor/antibody, into the individual energetic contributions of each of the amino acid residues involved. Additionally, the model characterizes the nature of these contributions into two main categories: (1) short-range/chemical and (2) long-range/electrostatic. Ultimately, we infer a network of interactions with amino acid residues of the two interacting molecules as nodes, and the inter-residue binding strength as edges. This interaction network is based on the electronic structure of the protein-protein system.

We focus on the viral spike interaction with ACE2 as its natural receptor, and with nAbs C121 and C144. We demonstrate that a QM model, assessing the interactions among the residues of an intermolecular biological system, enables mechanistic insight into how SARS-CoV-2 interacts with its host. The QM-CR model identifies the E484 residue as the only interface element hindering the binding between the Wuhan strain and hACE2, making it the most evident weak link of the Wuhan spike binding to the human host. The E484K mutation is shown by the model as a direct solution to this hindrance by improving binding to hACE2, and presumably constituting an evolutionary advantage, as supported by its emergence among several successful variants. Interestingly, QM-CR also shows that the E484 residue stabilizes the interaction between the Wuhan viral spike and the bat receptor macACE2 from *Rhinolophus macrotis*. We interpret this as an indication that the Wuhan strain is better adapted to a bat-like ACE2, and the rise of changes at E484 constitutes an adaptation specific to the human host.

In agreement with known data, QM-CR predicts loss of interaction between the SARS-CoV-2 spike and nAbs C121 and C144, once the E484K mutation is imposed on the spike of the Wuhan strain. The RBD residue E484 emerges as the main and fundamental spike fragment enabling the binding event, and therefore neutralization. These data suggest that nAbs challenging the spike at E484—the very residue that most hinders hACE2 interaction—provide an ulterior selective

pressure for the virus to find alternatives to the original phenotype, at this position.

By analyzing the competition between short- and long-range interaction contributions, we have shown that, compared to the Wuhan strain, the charge-shift E484K mutation substantially increased (by about 30%) the binding energy to hACE2. On the RBD side, the model also highlights how the effect of E484K is focused on the 484 position, with limited off-target repercussions for the spike's binding (Fig 3). We argue that this trait qualifies the E484K mutation as highly "RBD-modular" and readily achievable in an already well-adapted spike structure. The contribution of E484K to the binding is largely long-range/electrostatic, therefore less dependent on a specific steric conformation. Our simulations are motivated by the available empirical data in identifying the E484K mutation as a particularly likely evolutionary outcome, based on increased SARS-CoV-2 infectivity and antibody evasion. We thus examined the potential impact of the E484K mutation on spike-hACE2 binding in the background of the Delta variant. Our model suggests that E484K affects spike-hACE2 and spike-nAb binding in a modular fashion. Thus, if acquired by the Delta strain, E484K further increases binding, possibly contributing to increased infectivity. We acknowledge that infectivity is a multi-factor process of which receptor binding is only one among multiple actors.

Our investigation is focused on characterizing individual amino acid contributions to the different performance of alternative spike structures in binding hACE2, especially to assess the hypothetical relevance of present and future single point mutations imposed on available crystal structures. Binding to ACE2 is the first step for SARS-CoV-2 infection, and is therefore central to the overall fitness of a given viral variant. In the context of viral evolution towards improved human ACE2 binding, we intend to identify the structural traits that represent the objects of selection; when compared to the closest experimental dataset available [9], the quantities we compute provide QM simulations which largely align with empirical results (Fig S4).

The QM-CR approach is performed on all-atom *in silico* structures as inputs. In this context, we have applied the QM-CR method to crystal structures available in the PDB database, as well as variations of them, whenever crystalized structures are unavailable. Our analysis does not take into account conformational changes (which recent work has shown take place on the order of microseconds for spike-hACE2 interactions [43, 44]; conformational changes would require applying QM-CR to a population of structures coming from, for instance, subsampled MD trajectories [45]. Furthermore, due to the nature of the QM-CR analysis and the use of a single frame, interaction energies do not account for entropic effects or rearrangement (electronic or nuclear) after disassociation. Interactions have also been partitioned into per-amino-acid contributions, which introduces some error terms, however this can be controlled using measures

provided by the QM-CR methodology (see Supporting Information, "details of fragmentation procedure"). For this study, *i.e*. the case of E484K, the model's predictions align with available empirical data even when using the initial virtual crystal structures. In this specific case, this may be due to the long-range impact of E484K as a charge-shift mutation. Moreover, in the vicinity of the interface, the QM-CR approach produces an interaction network, which at the very least encodes the first-order effects that a mutation can induce in the chemical bonds of the interface.

The crystal structures employed for the Delta spike variant are not associated to an experimental result. They are virtual approximations, obtained via local energy minimization. The approximation assumes that no major structural changes from the reference Wuhan spike occur when single point mutations are introduced. In the supplementary information, we show evidence that such an approximation is reasonable, at least for the combination of mutations characterizing the Beta variant RBD: E484K, N501Y, and K417N. We employ a well-established DFT approximation, PBE+D3, which provides reliable information on coarse-grained quantities and trends [32, 46], and simulates structures in their relaxed positions [47]. Overall, we deem our method to be a balanced compromise between accuracy and modeling complexity.

The maturity of large-scale quantum mechanical calculations represents a unique opportunity to employ full QM approaches to uncover the interaction mechanisms. Such mechanisms are presently inaccessible to other, more conventional computational approaches. We also show that an *ab initio* modeling in QM-CR provides insights useful for comparison with experimental data, supporting its capability to offer predictive power for intermolecular interactions of biological relevance. Finally, we argue that QM-CR can be correlated to high throughput calculations of libraries of mutated structures aimed at identifying potential antibody escape routes for SARS-CoV-2 and, being unbiased and agnostic, can be readily applied to other biological systems.

Methods

Computational approach

We perform a full Quantum Mechanical (QM) model, as implemented in the BigDFT computer program suite [48]. The approach employs the formalism of Daubechies wavelets to express the electronic structure of the assemblies in the framework of the Kohn-Sham (KS) formalism of Density Functional Theory (DFT) [39]. The electronic structure is expressed, by both the density matrix and the Hamiltonian operator, in an underlying basis set of support functions—a set of localized functions adapted to the chemical environment of the system. Such functions are expressed in Daubechies wavelets, typically using one to four support functions per atom as the

basis set. The electronic density matrices, as well as the Hamiltonian expressed in the BigDFT basis set, are analyzed to provide quantum observables of the systems. The code provides efficient and accurate QM results for full systems of large sizes, delivering excellent performance on massively parallel supercomputers. In the present study, we employ the PBE approximation corrected by dispersion D3 correction terms [49] and Hartwigsen-Goedecker-Hutter (HGH) pseudopotentials [50]. The CheSS library [51] has been employed to calculate the system's density matrix. A comparison of the inclusion of an implicit solvent, with respect to gas phase calculations shows that interaction energies at interface residues are only marginally affected by the presence of the solvent (Fig S5).

Each calculation includes approximately 12,000 atoms and requires about 2 h of wall-time on 32 compute nodes of the IRENE-Rome supercomputer, at the TGCC Supercomputing center in Saclay (Paris, France). A similar approach has been previously used, in conjunction with the other atomistic techniques described in the introduction, to investigate the interaction patterns of the SARS-CoV-2 main protease with natural peptidic substrates, and to design peptide inhibitors tested *in vitro* [45].

Procedure

Starting from a representative 3D model of the molecules as our input, we calculate the system's electronic structure, from which we extract various quantities. We draw a contact network to identify relevant chemical interactions among the spike RBD and the various interactors considered in this study. The strength of the inter-residue interaction is quantified by the Fragment Bond Order (FBO) [52], calculated using the electronic structure of the system in proximity of a given residue. Such an approach has been previously described in detail [39, 53] and is summarized in Table 1.

We use the FBO to identify the interface residues, defined as the amino acids of the counter-ligand that have a non-negligible value, above a set threshold of the FBO, with the ligand. In contrast to a simple geometrical indicator like the RBD-ligand distance, the FBO provides a metric that enables a non-empirical identification of steric hot-spot interactions. We here identify as chemical hot-spot interface residues the amino acids which exhibit a FBO value with the ligand larger than 7×10^{-3}. Such threshold is obtained comparing the hydrogen bonding interaction network of the SARS-CoV-2 main protease to its natural peptidic substrates, derived from traditional FF analysis and the equivalent FBO network [32].

Once the chemical connection among amino acids is identified, we assign to each residue its contribution to the binding interaction between the two subsystems. We calculate these

interaction terms from the output of the DFT code and interpret them as two parts. First, a long-range electrostatic attraction/repulsion term, defined from the electron distributions of each of the fragments (even when far apart, two fragments may still interact). The remaining term, which can only be attractive, is provided by the chemical binding between the fragments, and is non-zero only if the electronic clouds of the fragments superimpose (short-range). This term is correlated with the FBO strength, and we identify it as the chemical interaction.

By including long-range electrostatic terms, the decomposition enables us to single out relevant residues not necessarily residing at the interface. In this way, the model provides an *ab initio* representation of the RBD-ligand interactions as the final output.

Crystal structures and generation of mutant virtual structures

Crystallographic structures are obtained from the RCSB database [54] using PDB entries 6M0J (hACE2), 7K8X (nAb C121), 7K90 (nAb C144), and 7C8J (macACE2). Protonation of histidines and other titratable residues is assigned a pH of 7, based on the PDBFixer tool in OpenMM [55, 56].

Virtual structures are generated by imposing point mutations on the original structure. Structure relaxations are performed by optimizing the crystal geometry with the OpenMM package using the AMBER FF14SB force field [57]. While such optimized structures do not represent the full panorama of conformations that might exist at a finite temperature, the resulting structures are interpreted as one plausible representative among the possible conformations of the system. To further verify this statement, we compared the difference in the interaction pattern obtained from the experimental crystal structure of the Beta variant in conjunction with hACE2 (PDB 7VX4), to the same quantity from the combined action of each point-mutation characterizing the Beta RBD (E484K, N501Y and K417N), applied on virtual crystals derived from WT-RBD (6M0J). We verify (see Supplementary Information) that the interaction difference on the RBD of the two real crystals corresponds to the overall sum of the contributions of each of the point mutations. This fact, on the one hand, confirms the modular impact of each point mutation to the overall binding, on the other hand, suggests that the impact due to conformational rearrangements is of higher order, for this variant.

Acknowledgments

We acknowledge useful discussions with Michel Masella, Lorenzo Fontolan, and Brigitte Lawhorn. LG also acknowledges support from the MaX EU Center of Excellence, and from French National Computing Resources (projects spe0011 and gen12049). BM and MZ were

supported by an Ignite grant from Boston College and by an Award for Excellence in Biomedical Research from the Smith Family Foundation. This work used computational resources of the supercomputer Fugaku provided by RIKEN through the HPCI System Research Project (Project ID: hp200179).

Data Availability

All data is included in the manuscript and/or supporting information.

Author Contributions

Conceptualization: MZ, LG, MF, WJ, BM.

Formal analysis: LG, WD, VC.

Funding acquisition: LG, TN, BM, WD.

Investigation: MZ, LG, MF, WJ, BM.

Methodology: MZ, LG, WD.

Software: LG, WD, VC.

Supervision: MF, WJ, BM.

Writing – original draft: MZ, LG, BM.

Writing – review and editing: MZ, LG, MF, WD, BM.

References

1. Yurkovetskiy L, Wang X, Pascal KE, Tomkins-Tinch C, Nyalile TP, Wang Y, et al. Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. *Cell* 2020; **183**: 739-751.e8.

2. Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science (80-)* 2021; **372**: eabg3055.

3. Dejnirattisai W, Zhou D, Ginn HM, Duyvesteyn HME, Supasa P, Case JB, et al. The antigenic anatomy of SARS-CoV-2 receptor binding domain. *Cell* 2021; **184**: 2183-2200.e22.

4. Thomson EC, Rosen LE, Shepherd JG, Spreafico R, da Silva Filipe A, Wojcechowskyj JA, et al. Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. *Cell* 2021; **184**: 1171-1187.e20.

5. Jangra S, Ye C, Rathnasinghe R, Stadlbauer D, Alshammary H, Amoako AA, et al. SARS-CoV-2 spike E484K mutation reduces antibody neutralisation. *The Lancet Microbe* 2021; **2**: e283–e284.

6. Starr TN, Greaney AJ, Addetia A, Hannon WW, Choudhary MC, Dingens AS, et al. Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. *Science (80-)* 2021; **371**: 850–854.

7. Luo R, Delaunay-Moisan A, Timmis K, Danchin A. SARS-CoV-2 biology and variants: anticipation of viral evolution and what needs to be done. *Environ Microbiol* 2021; **23**: 2339–

2363.

8. Chan KK, Dorosky D, Sharma P, Abbasi SA, Dye JM, Kranz DM, et al. Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2. *Science (80-)* 2020; **369**: 1261–1265.

9. Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Dingens AS, et al. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* 2020; **182**: 1295-1310.e20.

10. Schneider G. Virtual screening: an endless staircase? *Nat Rev Drug Discov 2010 94* 2010; **9**: 273–276.

11. Gorgulla C, Boeszoermenyi A, Wang Z-F, Fischer PD, Coote PW, Padmanabha Das KM, et al. An open-source drug discovery platform enables ultra-large virtual screens. *Nature* 2020; **580**: 663–668.

12. Brooijmans N, Kuntz ID. Molecular recognition and docking algorithms. *Annu Rev Biophys Biomol Struct* 2003; **32**: 335–373.

13. Guedes IA, de Magalhães CS, Dardenne LE. Receptor-ligand molecular docking. *Biophys Rev* . 2014. Springer. , **6**: 75–87

14. Pagadala NS, Syed K, Tuszynski J. Software for molecular docking: a review. *Biophys Rev* . 2017. Springer Verlag. , **9**: 91–102

15. Meng X-Y, Zhang H-X, Mezei M, Cui M. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Curr Comput Aided-Drug Des* 2012; **7**: 146–157.

16. Dauber-Osguthorpe P, Hagler AT. Biomolecular force fields: where have we been, where are we now, where do we need to go and how do we get there? *J Comput Aided Mol Des* 2019; **33**: 133–203.

17. Schlick T, Portillo-Ledesma S. Biomolecular modeling thrives in the age of technology. *Nat Comput Sci 2021 15* 2021; **1**: 321–331.

18. van der Spoel D. Systematic design of biomolecular force fields. *Curr Opin Struct Biol* 2021; **67**: 18–24.

19. Senn HM, Thiel W. QM/MM methods for biomolecular systems. *Angew Chemie - Int Ed* 2009; **48**: 1198–1229.

20. Bai C, Warshel A. Critical Differences between the Binding Features of the Spike Proteins of SARS-CoV-2 and SARS-CoV. *J Phys Chem B* 2020; **124**: 5907–5912.

21. Casalino L, Dommer AC, Gaieb Z, Barros EP, Sztain T, Ahn SH, et al. AI-driven multiscale simulations illuminate mechanisms of SARS-CoV-2 spike dynamics. *Int J High Perform Comput Appl* 2021; **35**: 432–451.

22. Toor HG, Banerjee DI, Lipsa Rath S, Darji SA. Computational drug re-purposing targeting the spike glycoprotein of SARS-CoV-2 as an effective strategy to neutralize COVID-19. *Eur J Pharmacol* 2021; **890**: 173720.

23. Ling R, Dai Y, Huang B, Huang W, Yu J, Lu X, et al. In silico design of antiviral peptides targeting the spike protein of SARS-CoV-2. *Peptides* 2020; **130**: 170328.

24. Teruel N, Mailhot O, Najmanovich RJ. Modelling conformational state dynamics and its role on infection for SARS-CoV-2 Spike protein variants. *PLOS Comput Biol* 2021; **17**: e1009286.

25. Ching WY, Adhikari P, Jawad B, Podgornik R. Ultra-large-scale ab initio quantum chemical computation of bio-molecular systems: The case of spike protein of SARS-CoV-2 virus. *Comput Struct Biotechnol J* 2021; **19**: 1288–1301.

26. Adhikari P, Li N, Shin M, Steinmetz NF, Twarock R, Podgornik R, et al. Intra- and intermolecular atomic-scale interactions in the receptor binding domain of SARS-CoV-2 spike protein: implication for ACE2 receptor binding. *Phys Chem Chem Phys* 2020; **22**: 18272–18283.

27. Basu A, Sarkar A, Maulik U. Molecular docking study of potential phytochemicals and their effects on the complex of SARS-CoV2 spike protein and human ACE2. *Sci Reports 2020 101* 2020; **10**: 1–15.

28. Isaac-Lam MF. Molecular modeling of the interaction of ligands with ACE2–SARS-CoV-2 spike protein complex. *Silico Pharmacol 2021 91* 2021; **9**: 1–16.

29. Francés-Monerris A, Hognon C, Miclot T, García-Iriepa C, Iriepa I, Terenzi A, et al. Molecular Basis of SARS-CoV-2 Infection and Rational Design of Potential Antiviral Agents: Modeling and Simulation Approaches. *J Proteome Res* 2020; **19**: 4291–4315.

30. Qiao Z, Zhang H, Ji HF, Chen Q. Computational View toward the Inhibition of SARS-CoV-2 Spike Glycoprotein and the 3CL Protease. *Computation* 2020; **8**: 53.

31. Kulik HJ, Zhang J, Klinman JP, Martínez TJ. How large should the QM region be in QM/MM calculations? the case of catechol O-methyltransferase. *J Phys Chem B* 2016; **120**: 11381–11394.

32. Dawson W, Mohr S, Ratcliff LE, Nakajima T, Genovese L. Complexity Reduction in Density Functional Theory Calculations of Large Systems: System Partitioning and Fragment Embedding. *J Chem Theory Comput* 2020; **16**: 2952–2964.

33. Mohr S, Masella M, Ratcliff LE, Genovese L. Complexity Reduction in Large Quantum Systems: Fragment Identification and Population Analysis via a Local Optimized Minimal Basis. *J Chem Theory Comput* 2017; **13**: 4079–4088.

34. Ratcliff LE, Mohr S, Huhs G, Deutsch T, Masella M, Genovese L. Challenges in large scale quantum mechanical calculations. *Wiley Interdiscip Rev Comput Mol Sci* 2017; **7**: e1290.

35. Laffeber C, de Koning K, Kanaar R, Lebbink JHG. Experimental Evidence for Enhanced Receptor Binding by Rapidly Spreading SARS-CoV-2 Variants. *J Mol Biol* 2021; **433**.

36. Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, et al. SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol 2021 197* 2021; **19**: 409–424.

37. Rawat P, Jemimah S, Ponnuswamy PK, Gromiha MM. Why are ACE2 binding coronavirus strains SARS-CoV/SARS-CoV-2 wild and NL63 mild? *Proteins* 2021; **89**: 389–398.

38. Sharma D, Rawat P, Janakiraman V, Gromiha MM. Elucidating important structural features for the binding affinity of spike - SARS-CoV-2 neutralizing antibody complexes. *Proteins Struct Funct Bioinforma* 2022; **90**: 824–834.

39. Ratcliff LE, Dawson W, Fisicaro G, Caliste D, Mohr S, Degomme A, et al. Flexibilities of wavelets as a computational basis set for large-scale electronic structure calculations. *J Chem Phys* 2020; **152**: 194110.

40. Weisblum Y, Schmidt F, Zhang F, DaSilva J, Poston D, Lorenzi JCC, et al. Escape from neutralizing antibodies 1 by SARS-CoV-2 spike protein variants. *Elife* 2020; **9**: 1.

41. Mou H, Quinlan BD, Peng H, Liu G, Guo Y, Peng S, et al. Mutations derived from horseshoe bat ACE2 orthologs enhance ACE2-Fc neutralization of SARS-CoV-2. *PLoS Pathog* 2021; **17**.

42. Andreano E, Piccini G, Licastro D, Casalino L, Johnson N V., Paciello I, et al. SARS-CoV-2 escape from a highly neutralizing COVID-19 convalescent plasma. *Proc Natl Acad Sci U S A* 2021; **118**.

43. Spinello A, Saltalamacchia A, Magistrato A. Is the Rigidity of SARS-CoV-2 Spike Receptor-Binding Motif the Hallmark for Its Enhanced Infectivity? Insights from All-Atom Simulations. *J Phys Chem Lett* 2020; **11**: 4785–4790.

44. Spinello A, Saltalamacchia A, Borišek J, Magistrato A. Allosteric Cross-Talk among Spike's Receptor-Binding Domain Mutations of the SARS-CoV-2 South African Variant Triggers an Effective Hijacking of Human Cell Receptor. *J Phys Chem Lett* 2021; **12**: 5987–5993.

45. Chan HTH, Moesser MA, Walters RK, Malla TR, Twidale RM, John T, et al. Discovery of SARS-CoV-2 Mpro peptide inhibitors from modelling substrate and ligand binding. *Chem Sci* 2021; **12**: 13686–13703.

46. Perdew JP, Burke K, Ernzerhof M. Generalized Gradient Approximation Made Simple. *Phys Rev Lett* 1996; **77**: 3865–3868.

47. Gundelach L, Fox T, Tautermann CS, Skylaris CK. Protein–ligand free energies of binding from full-protein DFT calculations: convergence and choice of exchange–correlation functional. *Phys Chem Chem Phys* 2021; **23**: 9381–9393.

48. Mohr S, Ratcliff LE, Genovese L, Caliste D, Boulanger P, Goedecker S, et al. Accurate and efficient linear scaling DFT calculations with universal applicability. *Phys Chem Chem Phys* 2015; **17**: 31360–31370.

49. Grimme S, Antony J, Ehrlich S, Krieg H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J Chem Phys* 2010; **132**: 154104.

50. Willand A, Kvashnin YO, Genovese L, Vázquez-Mayagoitia Á, Deb AK, Sadeghi A, et al. Norm-conserving pseudopotentials with chemical accuracy compared to all-electron calculations. *J Chem Phys* 2013; **138**.

51. Mohr S, Dawson W, Wagner M, Caliste D, Nakajima T, Genovese L. Efficient Computation of Sparse Matrix Functions for Large-Scale Electronic Structure Calculations: The CheSS Library. *J Chem Theory Comput* 2017; **13**: 4684–4698.

52. Dawson W, Degomme A, Stella M, Nakajima T, Ratcliff LE, Genovese L. Density functional theory calculations of large systems: Interplay between fragments, observables, and computational complexity. *Wiley Interdiscip Rev Comput Mol Sci* 2021; e1574.

53. Mohr S, Ratcliff LE, Boulanger P, Genovese L, Caliste D, Deutsch T, et al. Daubechies wavelets for linear scaling density functional theory. *J Chem Phys* 2014; **140**.

54. RCSB PDB. https://www.rcsb.org/. Accessed 23 Oct 2021.

55. OpenMM. https://openmm.org/. Accessed 23 Oct 2021.

56. GitHub - openmm/pdbfixer: PDBFixer fixes problems in PDB files. https://github.com/openmm/pdbfixer. Accessed 27 Nov 2021.

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

58. Siebenmorgen T, Zacharias M. Computational prediction of protein–protein binding affinities.

Wiley Interdiscip Rev Comput Mol Sci 2020; **10**: e1448.

59. GISAID - Initiative. https://www.gisaid.org/. Accessed 27 Nov 2021.

60. Dawson W, Mohr S, Ratcliff LE, Nakajima T, Genovese L. Complexity Reduction in Density Functional Theory Calculations of Large Systems: System Partitioning and Fragment Embedding. 2019; 1–13.

Figure Captions

Fig 1. Mechanistic characterization of the binding between Wuhan strain's spike and hACE2. Data are plotted on the sequence of hACE2 (panel a) and the spike RBD (panel b). Letters represent single amino acid residues; yellow bars indicate interface residues, identified with the FBO threshold. "FBO" is the Fragment Bond Order values, and "Distance" is the distance of a residue to the nearest atom of its ligand. "Interaction" is the chemical/electrostatic force shown as attractive (blue) or repulsive (red), with darker colors indicating stronger effects.

Fig 2. Mechanistic characterization of Wuhan and mutated (E484K) spike binding to hACE2. Data are plotted on hACE2 (panel a) and on the Wuhan spike (panel b) primary structure bound to the Wuhan spike (WT) and the mutated one (E484K). Amino acids are represented by the corresponding letters and numbered on the histogram's horizontal axis. Interface residues are highlighted by yellow bars and their overall effect on the other molecule is indicated by red (repulsive) or blue (attractive) tiles. Histograms underneath the sequences show the relative change in binding energy of the E484K mutated variant relative to the Wuhan strain, with positive and negative values indicating weaker and stronger binding, respectively. Bottom right histograms represent the overall binding energy of hACE2 with the Wuhan spike versus the mutated one, partitioned into chemical and electrostatic contributions. Interaction networks (Wuhan spike-hACE2 to the right, and mutated spike-hACE2 to the left), including FBO-interface residues and their coordinated interactors are shown (panel c). Squares depict spike residues and circles depict hACE2 residues, with red color for repulsive and blue color for attractive energy. Yellow outlines highlight interface residues. Bonds are purple when intermolecular or black when intramolecular.

Fig 3. Mechanistic characterization of the Wuhan spike binding to the human ACE2 (hACE2) and *R. macrotis* ACE2 (macACE2). Data are plotted on the ACE2 primary structure (a), and on the Wuhan spike RBD (b), when binding to the human (hACE2) and the bat (macACE2) receptor. Amino acid residues are labeled with letters and numbered. Interface residues are highlighted with a yellow bar, red tiles are repulsive residues, and blue tiles are attractive residues; see the rest of the figure for energy scales. The interaction networks (bottom) represent the hACE2-spike system on the left, and macACE2-spike on the right; circles are ACE2 residues, squares are spike residues. Interface residues are highlighted with a yellow bar, red tiles are repulsive residues, and blue tiles are attractive residues. Bonds are purple when intermolecular or black when intramolecular, and their thickness represents the strength of the FBO between residues.

Fig 4. Mechanistic characterization of C121 binding to the Wuhan strain spike protein, and energetic changes as a result of the E484K spike mutation. Data are plotted on the spike primary structure (panel a) and on C121's Heavy-Chain (panel b) considering the different bindings via the Wuhan spike (WT) and the mutated one (E484K). Amino acids are represented by letters and numbered on the histogram's horizontal axis. Histograms underneath the sequences represent the relative change in binding energy of the second row relative to the first one (Wuhan strain). The bottom right histograms represent the overall binding energy of C121 with the Wuhan spike (left) and the mutated one (right) and its characterization as chemical or electrostatic. The row above each sequence shows the chemical or

electrostatic forces as attractive (blue) or repulsive (red), with darker colors indicating stronger effects.

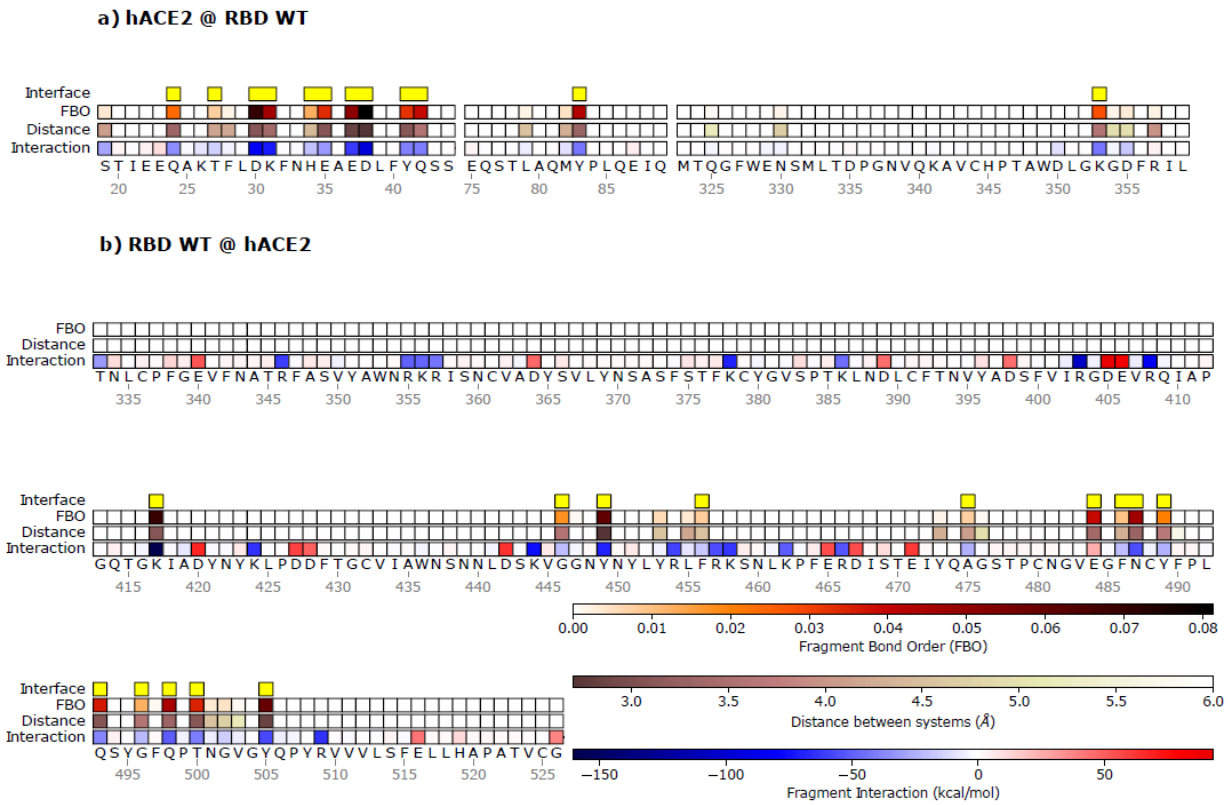
Fig 4 (cont.). Interaction networks with C121 nAbs are shown (panel c). Network nodes are represented in red (repulsive) or blue (attractive) based on their effect on their counterparts. Residues at the binding interface are highlighted by a yellow outline. Bonds are plotted as purple when intermolecular or black when intramolecular and their thickness is related to the strength of the FBO between residues.

Fig 5. Mechanistic characterization of spike-hACE2 binding suggests that Delta+E484K spike has stronger hACE2 binding than the Delta variant. Data are plotted on hACE2 (panel a) primary structure bound to the Wuhan spike (WT), Delta spike (δ), and Delta + 484K spike (δ + 484K). Amino acids are represented by the corresponding letters and numbered on the histogram's horizontal axis. Interface residues are highlighted by yellow bars and their overall effect on the other molecule is indicated by red (repulsive) or blue (attractive) squares (energy scale is identical to the one employed in the other figures). Histograms underneath the sequences show the relative change in binding energy (green: Delta compared to Wuhan; red: Delta+E484K compared to Delta). Data are plotted on the viral spike (panel b) primary structure bound to the Wuhan spike (WT), Delta spike (δ), and Delta + 484K spike (δ + 484K). Amino acids are represented by the corresponding letters and numbered on the histogram's horizontal axis. Interface residues are highlighted by yellow bars and their overall effect on the other molecule is indicated by red (repulsive) or blue (attractive) squares (energy scale is identical to the one employed in the other figures). Histograms underneath the sequences show the relative change in binding energy (green: Delta compared to Wuhan; red: Delta+E484K compared to Delta). Bar plots on the bottom right represent the overall binding energy of hACE2 with the Wuhan, Delta, and Delta+E484K strains, partitioned into chemical or electrostatic contributions.

Table 1. Prospectus of the main concepts and quantities constituting the model. All the elements here discussed are general and therefore applicable, without previous parameterization, to any given set of atoms for which atomistic structural representations are available.

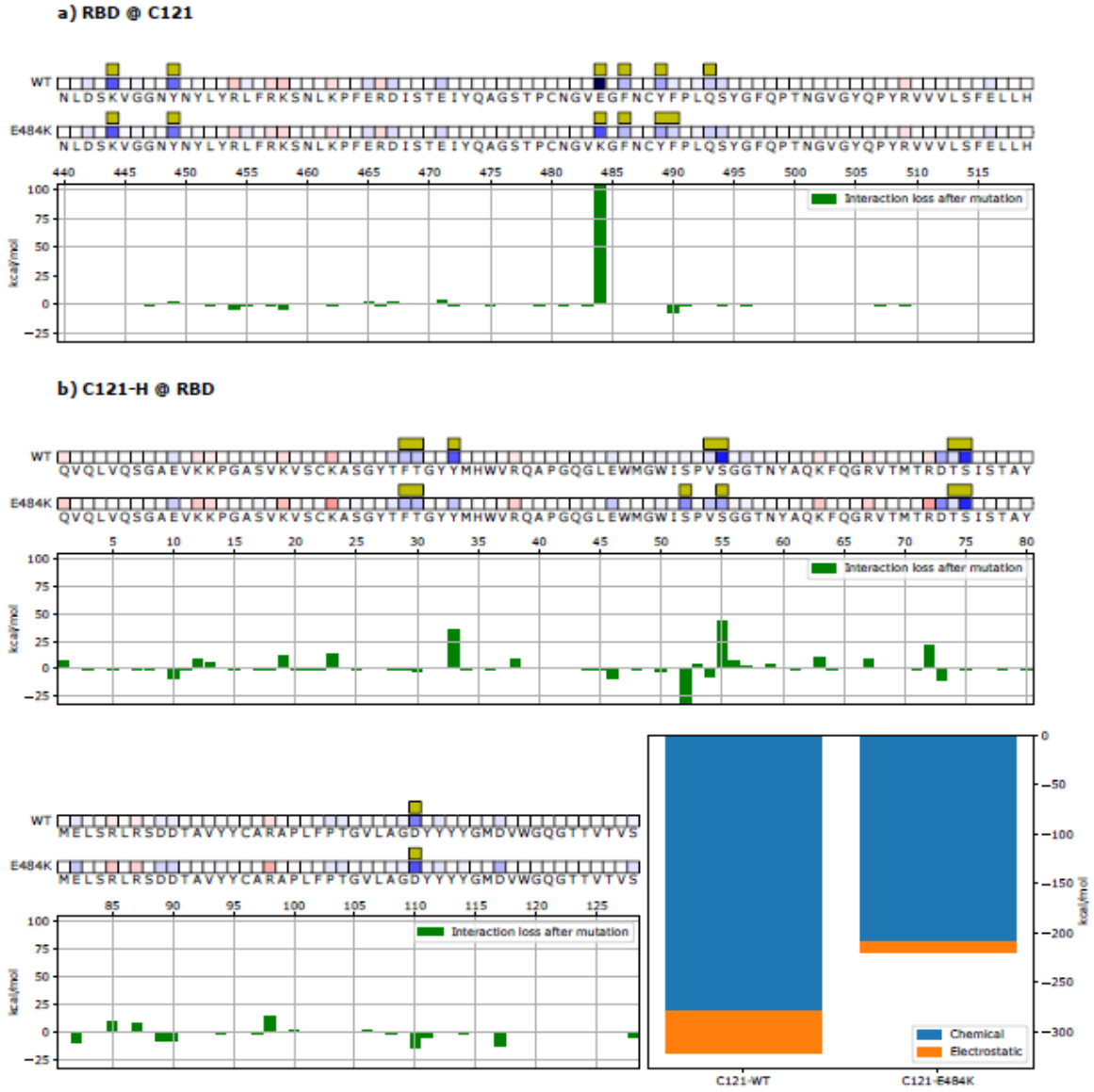
Electron Density	The distribution of electrons in a given molecular system. The electron density determines the nature and strength of the chemical bonds between interacting molecules. Such an “electron cloud" is the main emerging property of the underlying atomic structure in defining the chemical characteristics of a molecule.
Fragment	The modular elements into which the electron cloud can be partitioned, for example, an amino acid. The model partitions the electron cloud into physically consistent regions and/or verifies the consistency of a pre-defined partitioning; every such region is defined as a fragment.
Fragment Bond Order (FBO)	The descriptor of the inter-fragment interactions. FBO is the main quantity used in the model to represent the connection pattern of the fragments of interacting molecules.
Fragment Interactions	From the results of the model and the features of the fragments it is then possible to calculate the interaction strength between any two fragments. Such interaction has both a chemical/short-range term that is always attractive, and an electrostatic/long-range term that can be attractive or repulsive.
Final Output	At the end of the simulation, BigDFT provides a simple representation of the strength of interaction between fragments of the two molecules. The model can describe the energy and nature of the acting chemical bonds. This enables a mechanistic explanation and/or prediction of how specific amino acid substitutions or deletions, in spikes or nAbs, impact the interactions with their hACE2 substrate or the viral spike, respectively.
Hardware Requirements	The model requires massively parallel calculations via high performance computing. On a modern supercomputer, hundreds of simulations can be performed in a time frame of one hour.

601 Figure 1



602

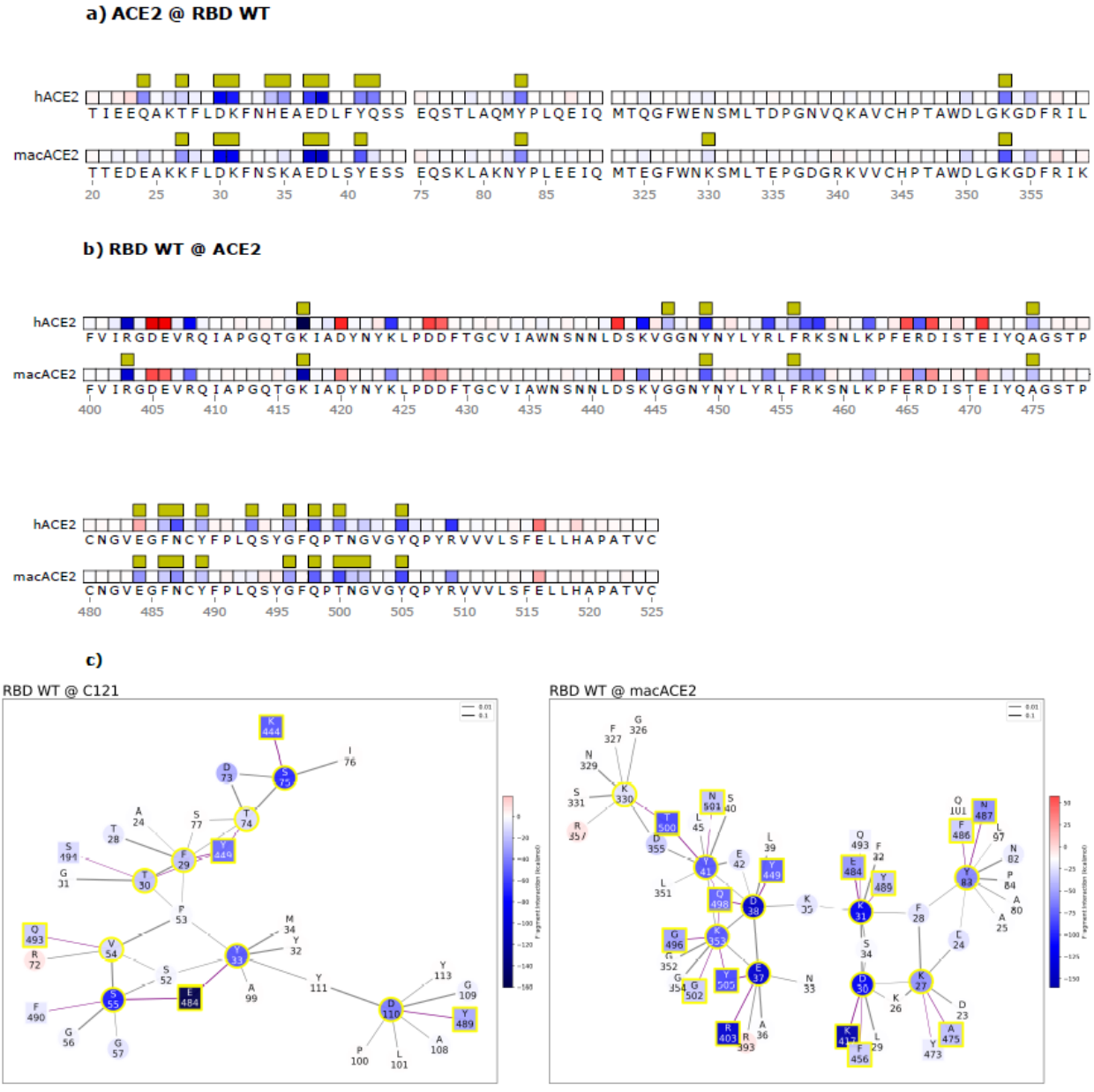
603 Figure 2



604

605

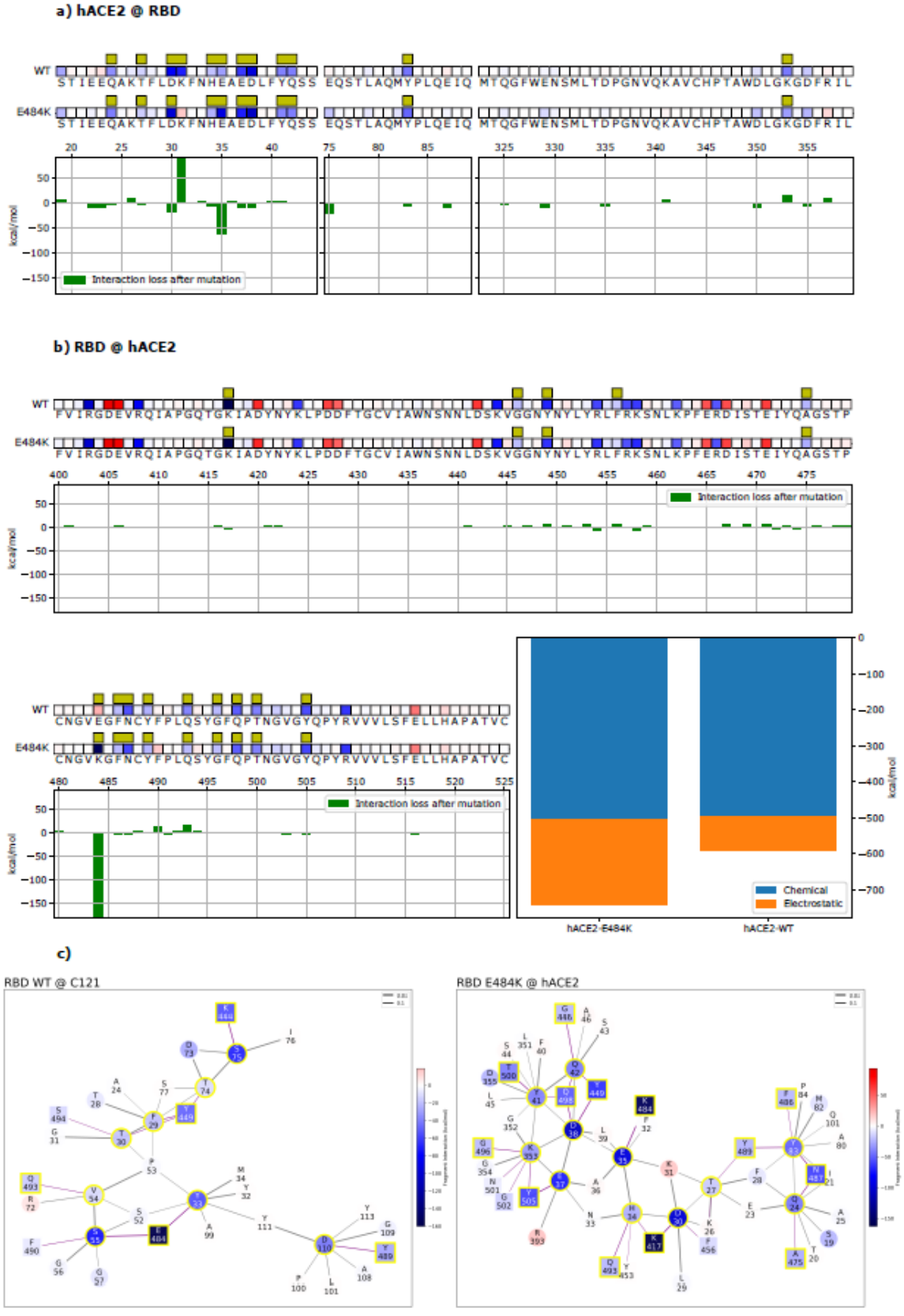
606 Figure 3



607

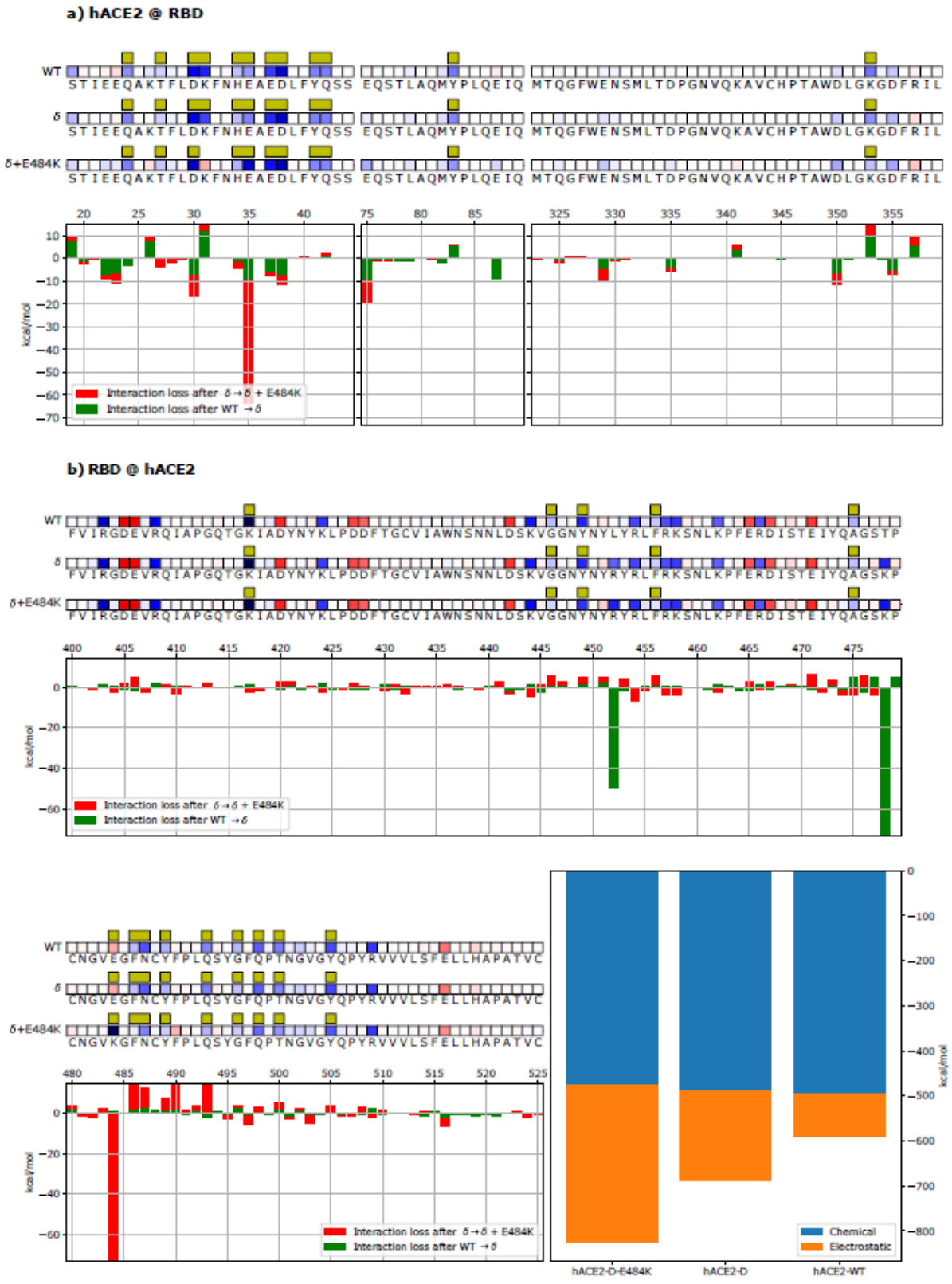
608

609 Figure 4



610

611 Figure 5



612