bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470748; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Investigating the mutational landscape of the SARS-CoV-2 Omicron variant via *ab initio* quantum mechanical modeling

Luigi Genovese[1◑*], Marco Zaccaria[2◑], Michael Farzan[3], Welkin Johnson[2], and Babak Momeni[2].

1 Université Grenoble Alpes, CEA, INAC-MEM, L_Sim, Grenoble, France
2 Department of Biology, Boston College, Chestnut Hill MA, United States
3 Department of Immunology and Microbiology, The Scripps Research Institute, Jupiter FL, United States

◑Equal contribution
* luigi.genovese@cea.fr

Abstract

SARS-CoV-2 variant "Omicron" B1.1.529 was first identified in South Africa in November 2021. Given the large number of mutations in Omicron's spike protein compared to the original Wuhan strain, its binding efficacy to the ACE2 receptor and its potential to escape antibodies are in the spotlight. Recently, we presented an *ab initio* quantum mechanical model to characterize the interactions of spike protein's Receptor Binding Domain (RBD) with select antibodies and ACE2 variants. The model identified weak links among the residues constituting interactions with the human ACE2 receptor (hACE2), and also enabled us to characterize *in silico* mutated RBDs to identify potential Variants of Concern (VOC). In particular, we focused on the role of RBD residue 484 in the interaction of the Delta variant with ACE2 and neutralizing antibodies (nAbs). In this report, we apply our model to the Omicron VOC, and characterize its interaction pattern with hACE2. Our results show that (i) binding affinity with hACE2, compared to Delta, is considerably increased, possibly contributing to increased infectivity. (ii) The interaction pattern between B1.1.529 and hACE2 differs from previous variants by shifting the hot-spot interaction residues on hACE2, and potentially affecting nAbs efficacy. (iii) A K mutation in the RBD residue 484 can further improve Omicron's binding of hACE2 and evasion of nAbs. Finally, we argue that a library of hot-spots for point-mutations can predict binding interaction energies of complex variants.

1 Introduction

As a new development in the SARS-CoV-2 pandemic, the Omicron variant, lineage B1.1.529, was first identified in South Africa on November 9th 2021 [1]. Omicron presents 32 mutations in the spike protein, 15 in the Receptor Binding Domain (RBD) alone (residues 319-541). Epidemiological data collected over the month of November in South Africa suggest that the infectivity of Omicron is higher than Delta's, the most transmissible variant to date. In light of this, on November 26th 2021, the Technical Advisory Group on SARS-CoV-2 Virus Evolution (TAG-VEO) advised Omicron to be defined as a Variant of Concern (VOC) [1]. To assess the risk of a decrease in effectiveness of public health measures, including vaccines and therapeutics, researchers

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470748; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

are now focusing their attention on the ability of known antibodies to neutralize Omicron. The unprecedented high number of mutations in the spike protein is expected to facilitate Omicron's evasion of neutralizing antibodies (nAbs) effective on the Wuhan and Delta variants. A mechanistic characterization of the interaction between Omicron and select ligands could inform whether there is a need to identify novel nAbs or optimize those available. Such a characterization can also guide the development of future SARS-CoV-2 vaccines.

In a recent contribution [2], we presented a model to study the interaction of the SARS-CoV-2 spike RBD with ligands of clinical relevance through quantum mechanical (QM) calculations of the full protein structure. The model was applied to experimental crystal structures of the RBD of the Wuhan strain (WT) with the ACE2 receptor both in human (hACE2) and in the bat host *Rhinolophus macrotis* (macACE2). Our model agnostically identified the E484 residue as the weakest link in the hACE2 interaction with the human host, while being well adapted to the bat receptor macACE2. We also studied the impact of the E484K mutation, characteristic of the highly infectious Beta variant, on such interaction patterns, by generating *in silico* mutated virtual RBD structures and comparing the relative differences in binding energy with respect to the RBD-hACE2 complex of the Wuhan and Delta variant. Despite the conceptual simplicity of the approach, and the implicit assumptions in modeling virtual structures, we have shown that our method is largely consistent with existing data and can complement high throughput experimental approaches.

In this report, we investigate the Omicron spike RBD by mechanistically characterizing its interaction with hACE2, in comparison with Wuhan and Delta's RBDs, via a QM study that quantifies the relative binding energy of RBD-receptor assemblies. We identify changes in the interaction network among the residues participating in the interaction–hereafter we refer to these residues as hot-spots (HS). This distribution of HS residues as an observable represents a useful descriptor to quantify the "distance" between the binding phenotypes of Omicron and Wuhan/Delta, and is potentially relevant for characterizing the binding of the Omicron variant to nAbs. Our model decomposes the interaction into the contributions of each variant-defining mutation, and predicts how SARS-CoV-2 evolution can further explore the chemical space through additional single point mutations. In line with our previous contribution [2], we show that our model predicts that a mutation to lysine in the 484 RBD position can further improve the viral spike's binding to hACE2.

2 Methods

Computational approach We implement the full QM model enabled by the BigDFT computer program suite [3]. This approach uses large-scale calculations on supercomputers, based on the Density Functional Theory, to extract the systems' electronic density matrix. These calculations allow us to investigate inter-molecular interactions. The approach employed here is identical to the one described in our previous contribution [2], and has also been employed to investigate the interaction of SARS-CoV-2's main protease with natural peptidic substrates to design peptide inhibitors [4].

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. In particular, we draw a contact map to identify relevant chemical interactions between the spike RBD and the interacting molecules (e.g. various receptors) considered in this study. The strength of the inter-residue interaction is quantified by the Fragment Bond Order (FBO) [5]. FBO is calculated using the electronic structure of the system in

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470748; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

proximity of a given residue. Such an approach has been previously described in detail [6]. Briefly, we use the FBO to identify residues that have a chemical interaction, namely the amino acids of the counter-ligand that share a non-negligible bond–above a set threshold–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 HS interactions.

Once the chemical connection between amino acids is identified, we assign to each residue its contribution to the binding interaction between the two subsystems. Such interaction terms can be calculated from the output of the BigDFT code, and can be split into two parts: (i) an electrostatic attraction/repulsion term, defined from the electron distributions of each of the fragments, which has a long-range character (even far apart, two fragments may still interact); (ii) an attractive term provided by the chemical binding between the two fragments, which is non-zero only if the electronic clouds of the fragments superimpose (short-range). This term is correlated to 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 a mechanistic *ab initio* representation of the RBD-ligand interactions as the final output.

Crystal structures and generation of virtual structures for mutants We base our structural model on crystallographic structures from the RCSB database [7](PDB entries 6M0J). A pH of 7 is assigned for the protonation of histidines and other titratable residues, based on the PDBFixer tool in OpenMM [8]. Virtual structures are generated by imposing point mutations on the original structure in OpenMM. Structure relaxations are performed by optimizing the crystal geometry with the OpenMM package using the AMBER FF14SB force field [9]. 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 representation among possible conformations of the system.

3 Results

Our objective is to shed light on the mechanisms leading to Omicron as a complex variant that has undergone several mutations, with the perspective of informing the effort to anticipate the emergency of new VOCs. We base our analysis on the same rationale employed in our previous contribution [2] and focus on the HS residues that drive the RBD-ACE2 interaction.

As we focus on the S1 part of the RBD, the point mutations we impose to detail Omicron include: K417N, N440K, G446S, S477N, T478K, E484A, Q493K, G496S, Q498R, N501Y, and Y505H. This is a relatively high number of mutations, some of which have already appeared in other variants (for instance the N501Y in the Alpha VOC, and T478K in Delta). We have purposely excluded the RBD mutations which are contextually away from the interface and with no strong electrostatic character, namely: G339D, S371L, S373P, S375F.

We analyze modifications to the RBD-hACE2 interaction energies induced by the simultaneous application of the Omicron-defining mutations, and point out how novel HS interaction patterns in the hACE2 sequence emerge. We then identify the role of each mutation in the interaction rearrangement and categorize the mutations according to their stabilizing power inside and outside the interface, the HS region to which they belong, and their "modularity", intended to represent the independence of each mutation (characterized as whether the combined effects of mutations are additive).

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470748; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

3.1 Omicron has a higher interaction binding than Delta with rearranged HS residues

We compare the pattern of interacting HS residues of the Omicron variant with those of the Wuhan and the Delta variants. In Fig. 1 and 2, we show how Omicron has much stronger binding enthalpy than both the Wuhan and the Delta variants, which until now had shown the strongest value among the VOCs in our model. Interestingly, the interaction patterns highlight an Omicron-specific binding arrangement with hACE2. The interface residues of the spike at hACE2 differ from those of the other known variants, presumably due to the significant structural change of the Omicron, a consequence of the large number of mutations. In the network of interacting residues (Fig. 3), we can identify two main HS regions on the hACE2 side when interacting with the WT RBD. The first, which we call HS region A, highlighted by an ellipse, involves residues 37-38, 41-42 and 353. The second, HS region B, is mainly related to residues D30, K31, and H34-E35. Conversely, the Omicron network displays a modified arrangement. We thus define region A', differing from A through the removal of E37, the inclusion of K353, and especially the addition of D355 as the new pivotal point. The HS region B' shows an increased relevance of the residue E35, and the removal of residue D30.

Remarkably, this interface rearrangement leads to a small *decrease* of the chemical/short-range contribution of the RBD-hACE2 interaction. However, the increase in the long-range contribution via the electrostatic term counter-balances such reduction, and strongly favors the Omicron binding overall.

Comparing the interaction networks of Omicron and Wuhan with hACE2 (Fig. 3), we observe that on the RBD side the E484A mutation in Omicron pulls position 484 off the interface, compared to Wuhan. The opposite happens with N501Y. Also, Omicron's K417N compromises one of the strongest interactors of the Wuhan and Delta variants. No other residue replaces K417 at the interface, with the exception of a negligible contribution by Y453. However, such interaction loss is largely compensated by Q493K, which increases its interface strength and further adds electrostatic attraction, particularly to the E35 residue.

Our model had previously highlighted E35 as a feature in the spike binding specific to the hACE2 receptor, in contrast to the K31 residue for binding to macACE2 in *Rhinolophus macrotis* [2]. On the hACE2 side, Omicron's mutations exclude the K31 position from the HS interaction residues. This happens in the HS group B'.

On the HS group A', the interface mutation Q498R further improves the binding, and the combined action of Y505H and N501Y fosters the appearance of a new HS residue, D355, to take over the role that had consistently been played by K353 [10] across other known variants.

We also highlight N440K, which increases the electrostatic binding. Such a mechanism is similar to the impact of the T478K mutation which is a major contributor to the increased binding of the Delta variant.

In our model, the nature of Omicron's binding improvement over WT is largely electrostatic. This can be seen as a further adaptation of the SARS-CoV-2 Spike protein to the electric field generated by hACE2. Although we have not explicitly examined the interaction of Omicron RBD with nAbs, the fact that the interaction pattern with the host receptor is largely modified may constitute a concern, as it is plausible that the interaction pattern with nAbs will also be affected.

3.2 Chemical space explored by the virus

In the absence of experimental crystal structures, we employed a structural model of the Omicron RBD generated *in silico* from the consensus crystal structure of the Wuhan

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470748; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

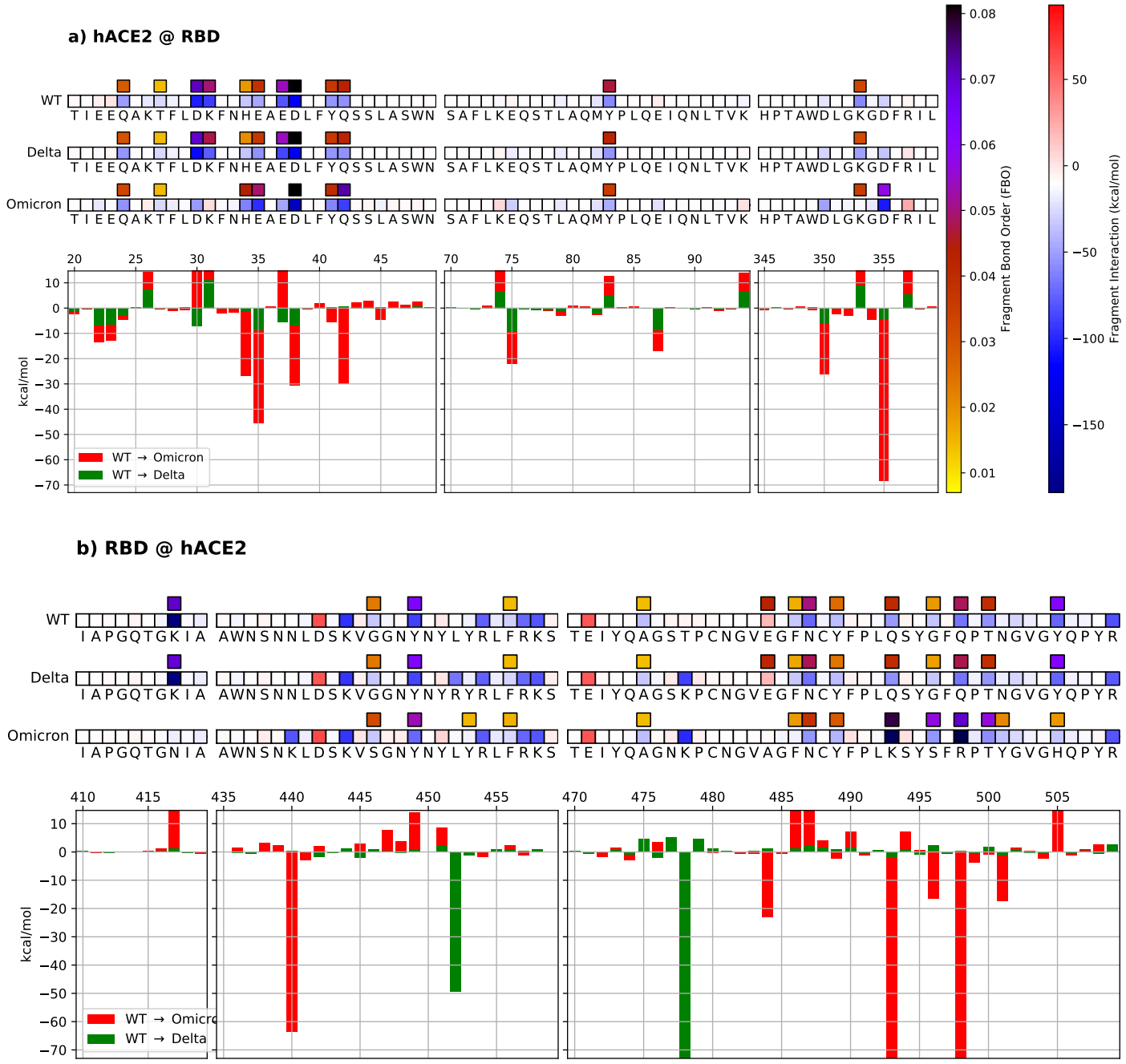

Fig 1. Mechanistic characterization of spike-hACE2 binding suggests that Omicron spike has stronger hACE2 binding that the Delta variant. Data are plotted on hACE2 (panel a) and the different viral spikes (panel b). Namely, we consider the Wuhan (WT), Delta, and Omicron spikes. Amino acids are represented by the corresponding letters and numbered on the histogram's horizontal axis. Their overall effect on the ligand is indicated by red/repulsive and blue/attractive squares; the energy scale is identical throughout all figures. Interface residues are highlighted by bars on top of the interaction values, and darker colors indicate stronger interface values. Histograms underneath the sequences show the relative change in binding energy (green: Delta compared to Wuhan; red: Omicron compared to Delta). Bottom right histograms represent the overall binding energy of hACE2 with the Wuhan, Delta, and Omicron variants, partitioned into chemical or electrostatic contributions.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470748; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

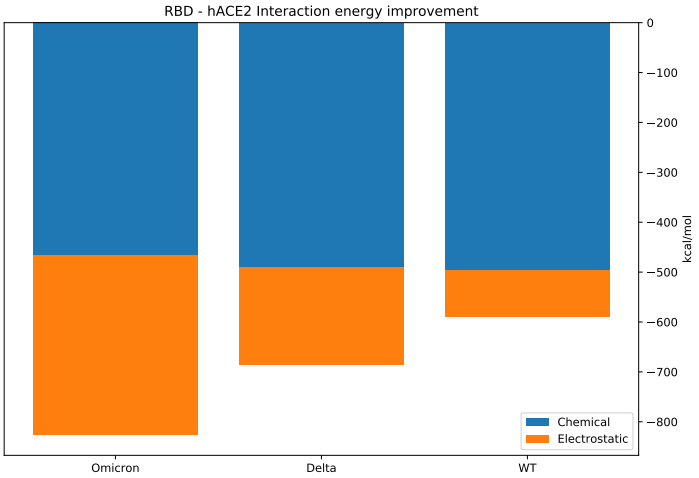

Fig 2. Overall binding energy of hACE2 with the Wuhan, and our model of Delta and Omicron variants, partitioned into chemical or electrostatic contributions.

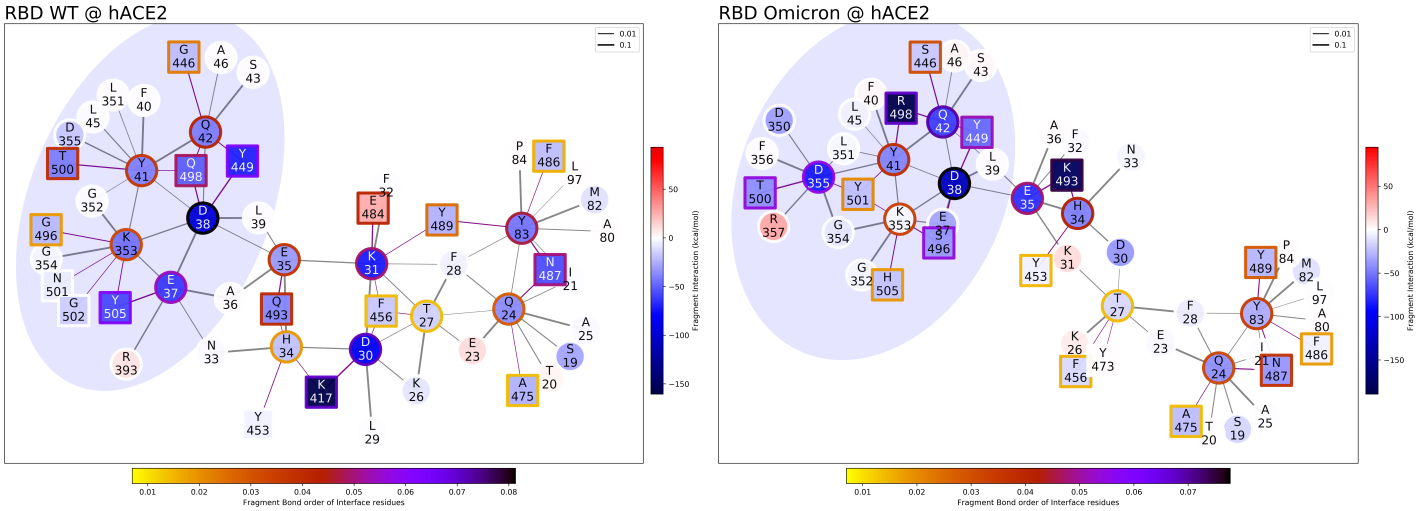

Fig 3. Comparison of Interaction networks between Wuhan and Omicron. Bonds are purple when inter-molecular or black when intra-molecular. Bond thickness is related to the strength of the FBO between residues, and the overall FBO strength is indicated by the color of the node frame. Graph nodes are represented in red (repulsive) and blue (attractive) based on the effect on their counterpart. Residues at the binding interface are highlighted by a yellow outline. The faint ellipse region highlights the group of residues that define what we called HS A (A'), for WT (Omicron). The remaining residues belong to the HS B (B') groups, respectively. The interaction graph of the Delta variant is similar to the WT.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470748; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

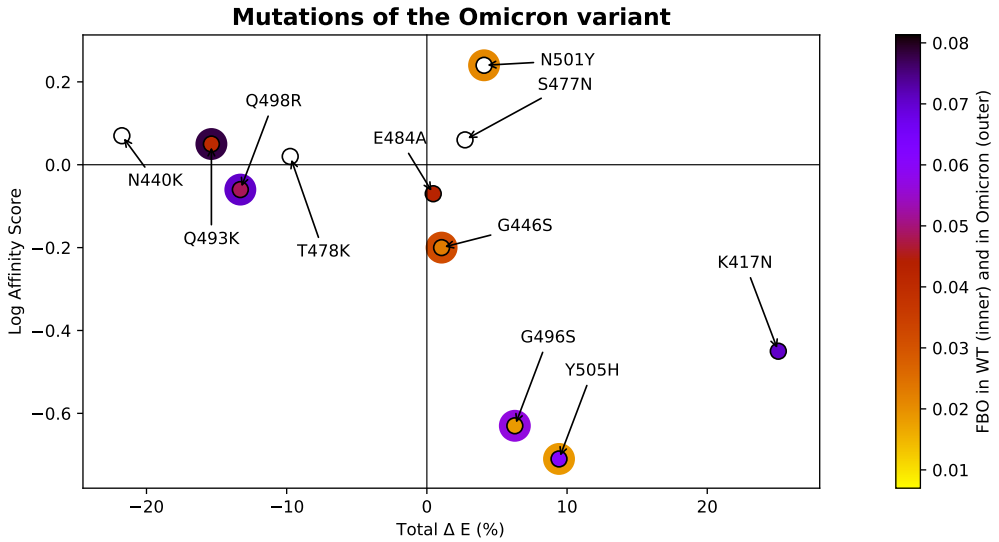

Fig 4. Comparison of the experimental binding affinity and the *ab initio* binding energies of the point-mutations characterizing the Omicron S1 RBD. The color of each point in the inner (/outer) region quantifies the FBO of the WT (/Omicron) residues when interaction with hACE2. White color indicates off-interface residues.

RBD (see Methods). Such a procedure enables us to study the impact of each of the point-mutation defining the variant by generating their virtual structure through the same method. The impact of a given point-mutation can also be studied experimentally. In our previous work [2], we pointed out that results from our model can complement experimental affinity data obtained from existing high throughput random mutation screening experiments for the Wuhan spike protein [11] and hACE2 [12]. We represent such comparison for the mutations that define the Omicron variant (Fig. 4). We also characterize each mutation by the magnitude of its FBO in the original Wuhan structure to verify whether or not it counted as an interface residue. This approach only has an informative character, as the two quantities (interaction energy on a 3D structure versus experimental binding affinity) can only be compared indirectly. Nonetheless, most of the data are in qualitative agreement, with the N501Y residue being the main exception (as already discussed in our previous contribution).

We also see that among Omicron mutations, residues with a strong interface character (K417 and Y505) are detrimental to binding, which confirms Omicron's trait of altering the contact residues. We had also previously pointed out the role of the E484 residue of the Wuhan strain, which resides at the interface with hACE2 but destabilizes the binding because of its electrostatic repulsion of the E35 hACE2 residue. In contrast, in Omicron, E484A mutation removes such electrostatic repulsion, at the cost of placing residue 484 off-interface.

3.2.1 Network interaction analysis

The plot in Fig. 4 enables us to split the mutations into three main groups. The mutations N440K, T478K, and S477N are off-interface both in WT and Omicron, and can therefore be interpreted as "Interface-Neutral". Their impact is, by construction, purely electrostatic, which explains the limited variation imposed by S477N mutation. Their interaction network is essentially identical to that of the WT, as the interface residues are unchanged by those mutations.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470748; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

We also identify mutations that we group together as "Interface-Destabilizing", and that have been lost in the Omicron interface. Together with the already discussed E484A, this group includes Y505H and K417N. Taken independently, their impact on the hACE2 binding is dramatic; in particular, the last two mutations compromise some of the strongest interaction sites on the WT RBD (Fig. 5 (networks in left part). Y505H destabilizes the assembly next to the hACE2 E37 residue (HS group A), whereas K417N detaches the D30 HS residue on HS group B.

A third group of Omicron mutations is defined by "Interface-Stabilizing" residues; namely, amino acids that reinforce or generate a chemical bond at the Omicron RBD-hACE2 interface. This includes G496S, G446S, N501Y, but particularly Q493K and Q498R, which reinforce the binding next to regions destabilized by the other mutations. Q493K and Q498R contribute to the evolution of the interaction patterns (Fig. 5). Q493K stabilizes the E35 hACE2 site (group B'), similar to the E484K mutation of the WT RBD, as already discussed in the previous contribution [2]. Q498R further stabilizes the D30-Q42 HS (group A') in hACE2.

It should be noted that for stabilizing mutations, the improved binding is not only of chemical nature. The above-mentioned charge-shift mutations have a strong electrostatic stabilizing power, which further consolidates the interface rearrangements. As anticipated in the previous section, this fact, combined with the destabilizing function of the mutations belonging to the previous group, generates a new interaction network defined by new HS residues of hACE2 @ Omicron. K353 is penalized (the G496S mutation is now the only one stabilizing it) in favor of the D355 residue, which emerged in HS group A', because of the combined effects of N501Y and Y505H in attracting Y41, and partially releasing K353. E35 emerged to second D30-K31, because of the combined effects of Q493K and K417N, together with the removal of the weak link through E484A. Contextually, Q498R reinforces the link at D38. Such an interface rearrangement is further improved by T478K, already a feature of the Delta variant, and N440K, which increase the electrostatic energy.

3.2.2 Additive role of a point-mutation

We now introduce another criteria to classify the Omicron point-mutations. Here, we would like to assess whether the effect of point mutations can be combined in an additive fashion. For this, we represent the per-residue *discrepancy* of the binding energy gain of the Omicron variant with respect to the sum of all the contributions that are brought by each of the point mutations separately (Fig. 6). We highlight the regions of the assembly which deviate from the additivity as the null hypothesis; i.e. $\Delta E \equiv E_O - E_{WT} \simeq \sum_{\text{mut}\in O} (E_{\text{mut}} - E_{WT})$, where we may interpret each mutation as additive. In hACE2, we see that the above mentioned HS residues D30, E35, D355, and Q42, deviate from modularity, due to interface rearrangements. Such analysis, performed on the RBD side, highlights the role of each of the mutations (in the assumptions of our model): S477N, T478K, and E484A exhibit an almost entirely modular character, together with Q493K, Y505H and, to some extent, K417N. The other mutations are less modular and cluster together in a different network. In particular, the largest deviation from modularity is ascribed to the glycine-serine mutations G446S and G496S.

According to these information, we use our model to analyze the role of mutations of group A and B separately. On top of that, we know that off-interface enhancing mutations may happen, as this may have lead to increased SARS-CoV-2 infectivity. Such enchancing mutations include the modular T478K, as well as N440K.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470748; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

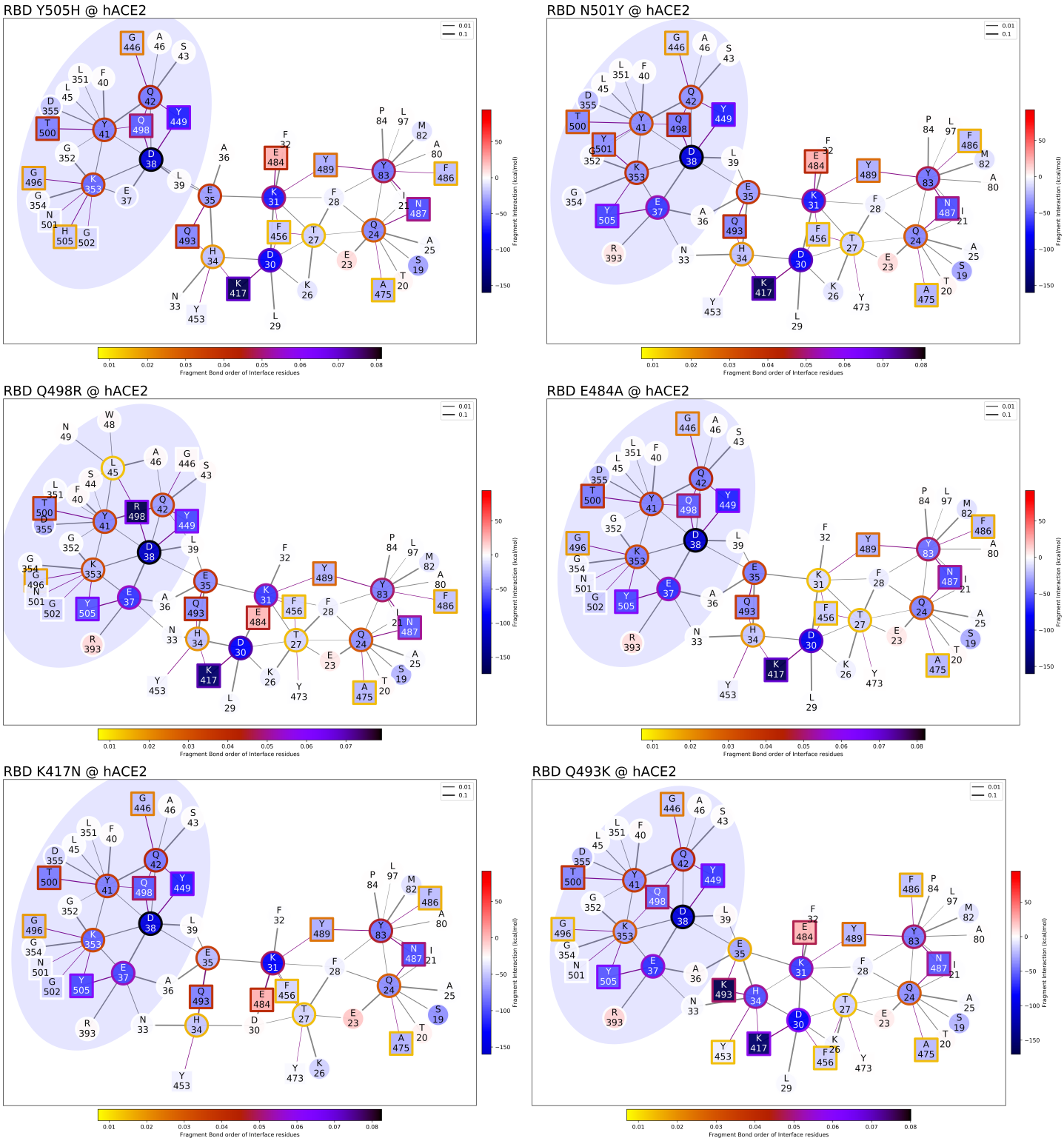

Fig 5. Notable interaction networks of introducing individual point mutations of the Omicron variant into the WT strain. On the left side are interface-destabilizing mutations, and on the right side interface-stabilizing mutations.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470748; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Point Mutation	Interface	Δ E wrt WT	Modular	hACE2 chemical HS	HS Group
N440K	Off-Interface	Stabilizing (22 %)	No	N/A	None
T478K	Off-Interface	Stabilizing (10 %)	Yes	N/A	None
S477N	Off-Interface	Neutral (-2%)	Yes	N/A	None
G496S	Stabilizing	Neutral (1%)	No	Q42	A
Q498R	Stabilizing	Stabilizing (13%)	No	D38	A
N501Y	Stabilizing	Neutral (-4%)	No	Y41, K353	A
G446S	Stabilizing	De-Stabilizing (-6%)	No	K353	A
Q493K	Stabilizing	Stabilizing (15%)	Yes	H34, E35	B
E484A	De-Stabilizing	Neutral (0%)	Yes	Releases K31 (E35)	B
Y505H	De-Stabilizing	De-Stabilizing (-9%)	Yes	Releases E37 (K353)	A
K417N	De-Stabilizing	De-Stabilizing (-25%)	Yes*	Releases D30 (H34)	B

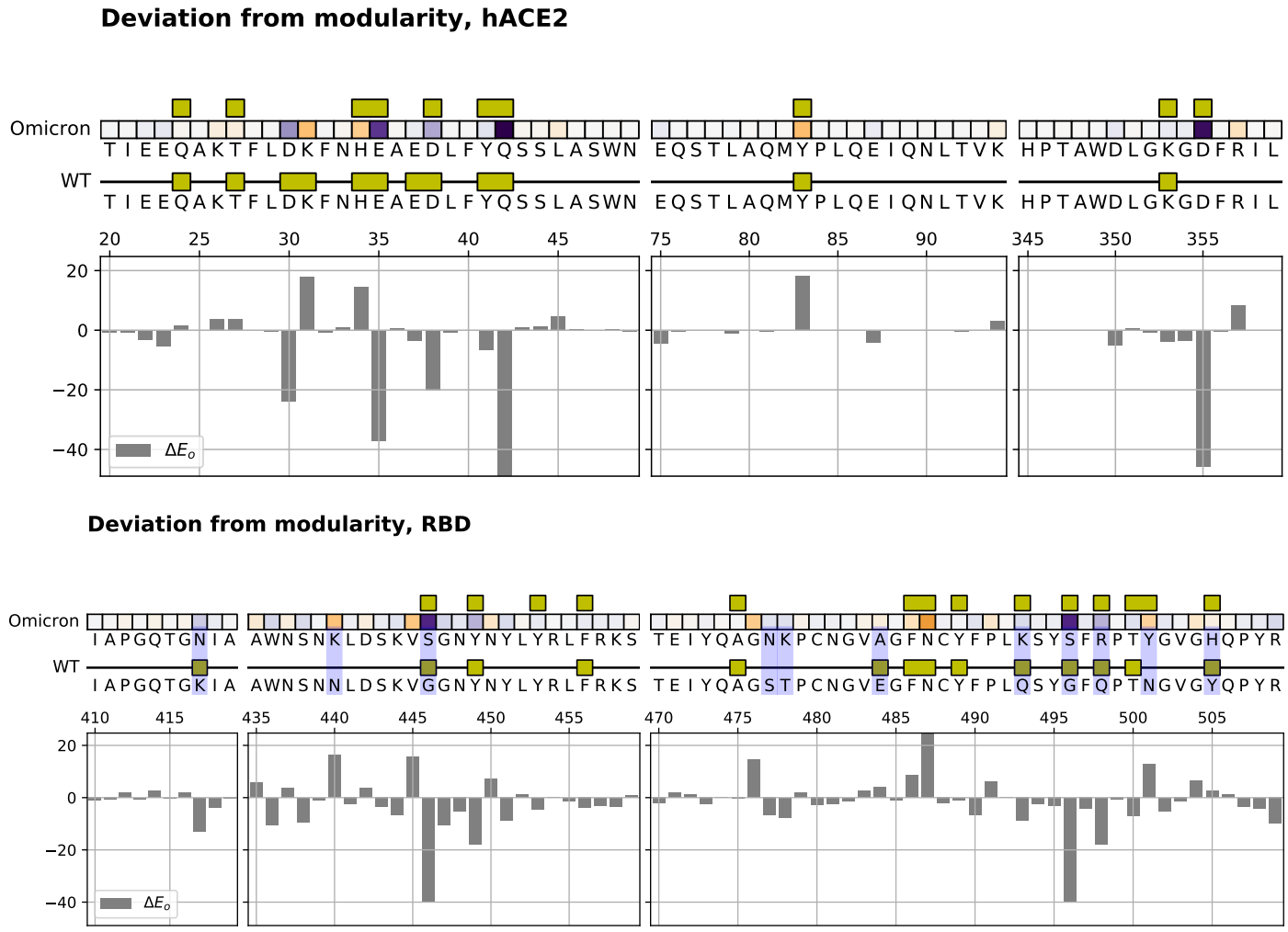

Fig 6. Difference between the interaction energy gain of the Omicron structure with respect to the sum of the contributions provided by each of the point-mutation separately. Mutated residues are highlighted with light blue. A mutation will be assumed as modular if the ΔE_o value is negligible. This criterion singles out S477N, T478K, E484A, Y505H, Q493K, and, considering its large interaction in Delta, also K417N.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470748; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

3.3 Hypothesis on Omicron's spike evolutionary trajectory

The presence of destabilizing mutations bring about the possibility that the Omicron variant has acquired some of its mutations in a certain order. To examine this, we proceed to analyze the mutations in the group B/B'. The K417N mutation seems so destabilizing that it should be accompanied by a mutation stabilizing the same HS group (B), namely Q493K, *and* likely by other off-interface mutations, as the combination of K417N and Q493K (both classified as modular) still has a destabilizing effect (-10%). We then speculate on the viral evolutionary trajectory to have acquired a first group of mutations: Q493K, K417N and N440K (or maybe T478K). The E484A mutation would release the weak link in the region of group B, therefore we include it in the same group.

Another strongly destabilizing mutation is Y505H, which has a modular character and can therefore be accompanied by another stabilizing mutation involving the same HS group, like Q498R. The non-modular nature of this mutation also suggests that other modifications may play a role, like G446S and G496S, which apparently seem neutral mutations for the WT but can restore strength to the HS group (as it can be seen from their interface-stabilizing character). This modification may have happened around a N501Y preexisting variant, like Alpha for instance.

The mutation S477N possibly does not belong to the same category of mutations as K417N and Y505H, since it appears neutral and modular based on our analysis.

3.4 A mutation to lysine in residue 484 is predicted to enhance Omicron's binding affinity

We impose on the simulated Omicron structure two mutations that define previous variants of concern: L452R from Delta, and A484K from Beta and Gamma mutations (we here denote the mutation into lysine as A484K because we assume it as based on the Omicron variant). The aim of this analysis is to verify whether the theoretical acquisition of these mutations in Omicron would further improve its binding to hACE2. Results are presented in Fig.7. The theoretical acquisition of L452R by Delta is expected not to improve hACE2 binding as its electrostatic contribution is washed out in the interaction. Conversely, A484K is predicted to have a dramatic stabilizing effect via the novel involvement of residue E75, available to bind a lysine. We caution, however, that the degree of approximation is, in this case, the highest we have attempted, in relation to the virtual crystal structure (the Omicron one) we have employed as a reference.

4 Discussion

Since the start of the COVID-19 pandemic, the emergence of SARS-CoV-2 variants has raised the concern, among researchers and healthcare providers, that the efficacy of nAbs, vaccines, therapies, and prophylaxis strategies could be compromised, especially because of adaptive changes in the viral spike protein. Omicron is the most recent of such variants, first identified in South Africa thanks to their highly efficient SARS-CoV-2 genotyping surveillance. At the time of writing this contribution, the World Health Organization has labeled Omicron a variant of concern (VOC); research is ongoing to characterize its potential effect on public health [1]. Omicron is suspected to be more transmissible than other variants in light of the steep rise of related cases in areas of South Africa. No information to date correlates Omicron to more severe symptoms and increased hospitalizations, compared to other variants. Limited information suggests that Omicron may cause an increased risk of reinfection [13]. Studies are ongoing to verify an impact of Omicron on the efficacy of vaccines, antigen tests, and current treatments [13].

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470748; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig 7. Characterization of the virtual mutation A484K on the Omicron variant Data are plotted on the ACE2 primary structure (a), and on the Wuhan spike RBD (b), when binding to the Omicron and the putative Omicron+L452R or A484K variants. 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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470748; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

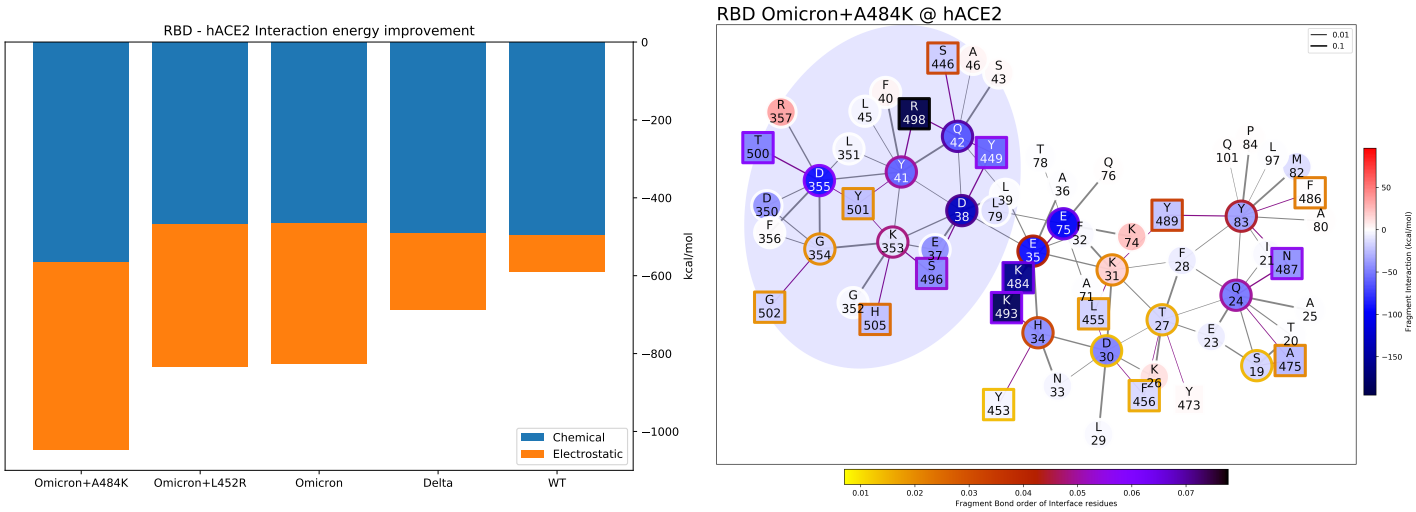

Fig 7. (Cont.) Overall binding energy of Wuhan, Delta, and Omicron compared to the theoretical variants Omicron+L452R and Omicron+A484K. To the right, the interaction network between hACE2 and the theoretical Omicron+A484K spike shows residue E75 being recruited at the interface, giving rise to another HG group (which may be labeled as B").

In this work, we perform *ab initio* QM modeling of an *in silico* generated structural representation of Omicron to mechanistically characterize its binding to the human ACE2 receptor (hACE2). We have already applied this approach studying experimental structures and *in silico* spike variants of SARS-CoV-2 interacting with select neutralizing antibodies, and with the ACE2 receptor of the human host and the bat *Rhinolophus macrotis* [2].

Our simulations show that Omicron has a higher binding energy to the human ACE2 than both the Wuhan strain and Delta variant. Interestingly, its large number of mutations seems to push Omicron to interact with hACE2 with a rearranged interaction pattern compared to previous variants of the Wuhan type strain, Delta included. We identify three main categories among the Omicron-defining mutations in the RBD: (i) interface-stabilizing mutations, which contribute to increase the overall binding energy; (ii) interface-destabilizing mutations, which decrease the overall binding energy; and (iii) off-interface mutations with an electrostatic energetic effect on binding. The *ab initio* QM simulation therefore highlights a substantial difference from predictions for the Delta variant which is expected to benefit exclusively from mutations of the third group. Our model also predicts that Omicron's spike acts on hACE2 with a different pattern of interacting residues compared to any other simulated variant. These differences cluster in two multi-residue HS regions, denominated A and B for Wuhan and A' and B' for Omicron. In particular, Omicron spike's forgoes the contact with E37 while establishing strong bonds with D355 and K353 in the A' cluster, while contextually tightening the contact with E35 and renouncing D30 in B'. The overall loss of short-range interaction is largely compensated by the long-range contribution of the novel electrostatic interactions, especially those enabled by 440K, 494K, and 498R, which ultimately lead to an increased total energy of binding. We argue that a similar identification of HS regions could be employed to identify nAbs against Omicron and/or to optimize available nAbs.

The analysis of Omicron's interaction with hACE2 predicts the total increase of binding energy to be actually higher than the one obtained via the sum of the individual contribution simulated for the correspondent point mutations (the purely

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470748; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

modular hypothesis). This is the consequence of the structural and electronic rearrangements at the interface. Taken individually, specific mutations such as K417N are highly detrimental to binding; we thus argue that they are likely to have originated together with, or after, other "redeeming" mutations affecting binding in the same region. Under this light, it is interesting to hypothesize plausible evolutionary trajectories of the spike, assuming the intermediate steps leading to Omicron displaying no worse than equal binding energy to hACE2 than the Wuhan spike. This is important, especially because intermediate evolutionary steps leading to Omicron are currently missing in the existing surveillance data. We advance one such hypothesis of an intermediate variant in the supplementary information of this paper.

The assumption of realism for our methodology becomes bolder and bolder the more the viral spike differs from the available crystal structure of the Wuhan strain. Such differences are rapidly accumulating, and the Omicron variant is an example in this sense. Nonetheless, this *ab initio* approach has previously aligned with experimental and empirical observations. In particular, we have shown the potential to make predictions for SARS-CoV-2 related events, such as evasion of nAbs and the stabilizing effect of adaptive mutations [2]. In applying it to Omicron, we push ourselves further away from the available crystal structures into uncharted chemical space of *in silico* virtual structures. We expect the validity of our predictions to soon be evaluated in light of upcoming experimental datasets. In the meanwhile, we intend this paper as a proof-of-concept of what *ab initio* modeling can predict, in an effort to explain and ultimately anticipate SARS-CoV-2 spike evolution. Finally, we argue that the generality of *ab initio* predictions opens up opportunities for studying antibody escape routes and viral evolution, *inter alia*, and inform decision-making on related issues.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470748; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

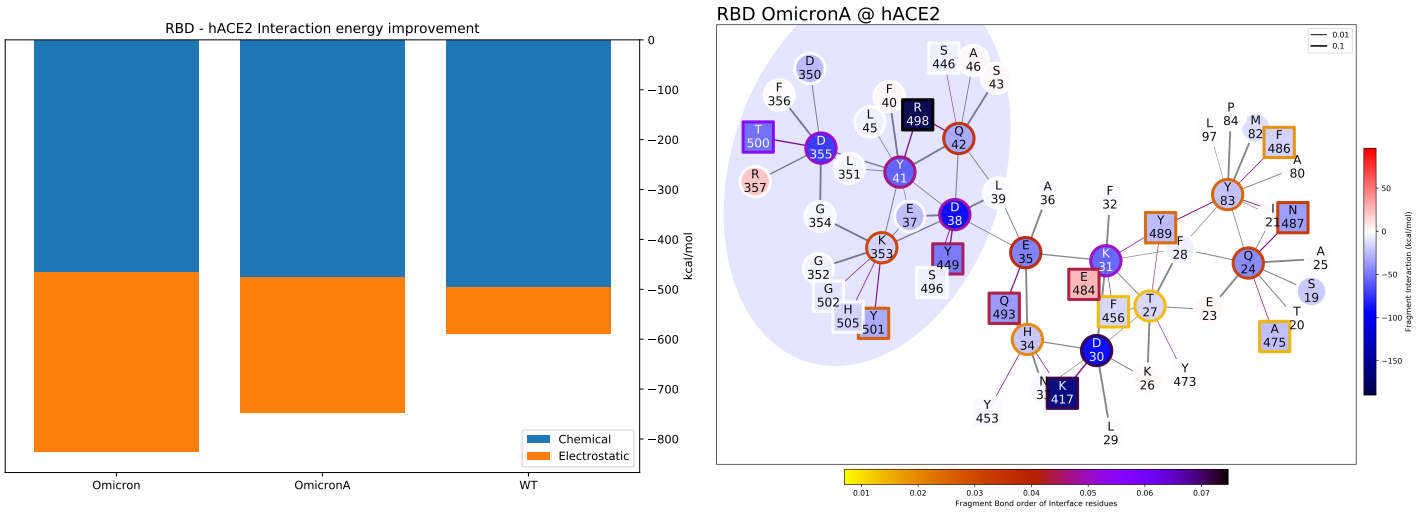

Fig A.1. Total binding energy of Wuhan and Omicron binding hACE2 compared to the theoretical intermediate variant OmicronA. OmicronA interaction network is shown to the right.

A Supporting information

A.1 Putative intermediate variant

We have tried to regroup the mutations which belong to the HS group A, namely G446S, G496S, N501Y, Y505H, Q498R, together with the ubiquitous T478K mutation, to verify if this "theoretical" intermediate mutation still exhibits the same features of our virtual Omicron variant, in the vicinity of HS group A. The figure below shows that with such a hypothetical variant the D355 HS emerges in group A', yet with some differences at the border of the region. In particular, the mutations N501Y and G496S seem to be structurally in competition, which may indicate a particular order in the way in which they have been acquired by the virus.

Acknowledgments

We acknowledge useful discussions with William Dawson, Viviana Cristiglio, Michel Masella, Lorenzo Fontolan, Karla Ilic Djurjic, 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).

Author Contributions

Conceptualization: MZ, LG, BM. Formal analysis: LG. Funding acquisition: LG, BM. Investigation: MZ, LG, MF, WJ, BM. Methodology: MZ, LG. Software: LG. Supervision: LG, MF, WJ, BM. Writing – original draft: MZ, LG, BM. Writing – review and editing: MZ, LG, MF, BM.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470748; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

1. World Health Organization Classification of Omicron (B.1.1.529): SARS-CoV-2 Variant of Concern https://www.who.int/news/item/26-11-2021-classification-of-omicron-(b.1.1.529)-sars-cov-2-variant-of-concern

2. Zaccaria M*, Genovese L*, Farzan M, Dawson W, Nakajima T, Johnson W, Momeni B. Anticipating future SARS-CoV-2 variants of concern through *ab initio* quantum mechanical modeling bioRxiv. 2021. doi: https://doi.org/10.1101/2021.11.25.470044

3. 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. doi:10.1039/c5cp00437c

4. 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. doi:10.1039/D1SC03628A

5. 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. doi:10.1021/ACS.JCTC.9B01152

6. Ratcliff LE, Dawson W, Fisicaro G, Caliste D, Degomme A, Videau B, et al. Flexibilities of Wavelets as a Computational Basis Set for Large-Scale Electronic Structure Calculations. J. Chem. Phys. 2020. 152: 194110.

7. RCSB PDB. 2021. https://www.rcsb.org/

8. OpenMM. 2021. https://github.com/openmm/pdbfixer

9. 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. doi:10.1021/ACS.JCTC.5B00255

10. Wan Y, Shang J, Graham R, Baric RS, Li F. Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. J Virol. 2020. 94: 1–9. doi:10.1128/jvi.00127-20

11. 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. doi:10.1016/j.cell.2020.08.012 l screens. Nature. 2020. 580: 663–668. doi:10.1038/s41586-020-2117-z

12. 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. 2020. 369: 1261–1265. doi:10.1126/SCIENCE.ABC0870

13. World Health Organization Update on Omicron https://www.who.int/news/item/28-11-2021-update-on-omicron