

Structural-bioinformatics analysis of SARS-CoV-2 variants reveals higher hACE2 receptor binding affinity for Omicron B.1.1.529 spike RBD compared to wild-type reference.

Vedat Durmaz
Katharina Köchl
Amit Singh
Michael Hetmann
Lena Parigger
Andreas Krassnigg
Daniel Nutz
Alexander Korsunsky
Centina König
Lee Chang
Marius Krebs
Riccardo Bassetto
Tea Pavkov-Keller
Verena Resch
Karl Gruber
Georg Steinkellner (✉ georg.steinkellner@innophore.com)
Christian C. Gruber (✉ christian.gruber@innophore.com)

Research Article

Keywords: COVID-19, SARS-CoV-2, coronavirus, spike protein, structural variant monitoring, alpha, beta, gamma, delta, omicron, variant of concern, molecular dynamics, Catalophore Halo, sequence genome analysis, linear interaction energy model, binding affinity prediction

Posted Date: December 8th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1153124/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

To date, more than 263 million people have been infected with SARS-CoV-2 during the COVID-19 pandemic. In many countries, the global spread came in several pandemic waves characterized by the emergence of new SARS-CoV-2 variants. Here, we report on a sequence- and structural-bioinformatics analysis by which we estimate the impact of amino acid exchanges on the affinity of the SARS-CoV-2 spike receptor-binding domain (RBD) to the human receptor hACE2. This is carried out by qualitative electrostatics and hydrophobicity analysis as well as through molecular dynamics simulations used for the development of a highly accurate linear interaction energy (LIE) binding affinity model that was calibrated on a large set of experimental binding energies. For the newest variant of concern (VOC), B.1.1.529 Omicron, our Halo difference point cloud studies reveal the largest impact on the RBD binding interface compared to any other VOC. Moreover, according to our LIE model, Omicron achieved a substantially higher ACE2 binding affinity than the wild-type and in particular the highest among all VOCs except for Alpha and therefore requires special attention and monitoring. Using this prediction model we provide early structural insight and binding properties before experimentally determined complex structures and binding affinity data become available in the upcoming months.

Introduction

In the COVID-19 pandemic, more than 263 million people have been infected with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) as reported by WHO in December 2021¹. So far, the global spread was distributed over several pandemic waves in many countries²: Each wave was characterised by the emergence of SARS-CoV-2 variants with some becoming dominant in regional endemic outbreaks³ or on a global level⁴⁻⁶. Whereas the timeline of the waves and the distribution of variants may appear to be asynchronous, several variants emerged that significantly dominated the global pandemic event, namely the variants of concern (VOCs), Alpha, Beta, Gamma and Delta^{7,8}.

While preparing potential drug targets⁹ of the SARS-CoV-2 proteome¹⁰ for our ultra-large-scale virtual drug screening^{11,12}, we started monitoring the emerging diversity of the SARS-CoV-2 genome landscape on a structural-bioinformatics level in order to keep design and strategy for the screening aligned with the actual pandemic situation. Recently, we reported on a then emerging variant carrying a particular amino acid exchange (S477N or S477G) in a highly flexible region of the spike receptor binding domain (RBD). We showed that this kind of exchange results in increased flexibility and increased affinity¹³ for the human receptor angiotensin-converting enzyme 2 (hACE2)^{14,15}. It is reported elsewhere that an increased spike-hACE2 affinity correlates with higher infectiousness¹⁶⁻¹⁸.

Still, indications for higher affinity or infectiousness do not automatically carry over to the identification of dominant SARS-CoV-2 variants, due to the complex dynamics of a pandemic situation. For example, the notable missense single mutation S477N/G embedded in the [variant under monitoring \(VUM\)](#) B.1.526 Iota did not become a major variant and spread only locally, not globally. In particular, local outbreaks e.g.

in New York City and New York State¹⁹ were characterized by domination of this variant, before the Delta variant became dominant there.

In the framework of our virus.watch project in cooperation with Amazon Web Services' Diagnostic Development Initiative we continue the structural monitoring of emerging drug- and disease-relevant SARS-CoV-2 mutations. Overall, the average number of amino acid exchanges per sequence, collected over a large set of available sequences, has increased dramatically in the past year. We present a more detailed analysis below.

Following the path of the S477G/N mutation, the currently widely discussed VOC B.1.1.529 Omicron raised our interest shortly after its genome became available on November 8th, 2021. It also contains the S477N mutation, which is the first time that this particular amino acid exchange is present in an official VOC.

The emerging SARS-CoV-2 VOC Omicron was first reported during an endemic outbreak in Botswana and South Africa (hCoV-19/SouthAfrica/NICD-N21441/2021, GISAID ID EPI_ISL_6913995)²⁰. It features a high number of mutations throughout the viral genome, 39 of which cause changes in the amino acid sequence of the spike protein²¹⁻²³. Within the spike RBD there are 15 amino acid replacements including our previously characterized mutation S477N. Thus, VOC Omicron comprises a three times higher number of mutations compared to the next most varied RBD contained in VUM B.1.640²¹.

A structural-bioinformatics analysis of this variant, in particular its expected binding mode and predicted affinity, is clearly needed. Such a claim is more than reasonable, both based on the extraordinary amount of RBD amino acid exchanges in this variant, as well as due to the presence of single mutations known to increase the binding affinity.

Furthermore, since the spike protein is also recognized by the immune system as a primary antigenic target to neutralize the virus, a close inspection of relevant changes in the interaction pattern is crucial. This is true in particular for, e.g., deepening the understanding of biomolecular recognition of antibodies that is the basis for many of today's approved Covid-19 vaccines^{24,25}. Concrete examples are Comirnaty (BNT162b2) by BioNTech SE, or Spikevax (mRNA-1273) by ModernaTX, Inc., who announced the development of an updated booster vaccine within the next few months. In the context of potential antibody evasion²⁶, vaccine manufacturers rely also on structural insights and consider them highly relevant, e.g., BioNTech has stated to closely monitor the emerging genomic diversity²⁷. In another line of application, several biotherapeutics currently in development, e.g., a recombinant human soluble hACE2-based decoy^{28,29}, rely on an effective binding to the spike protein at the surface of the virion.

Herein, we report the results of our structural-bioinformatics approach regarding the influence of these amino acid exchanges on the affinity of the spike RBD to the human receptor hACE2. Following our analysis of the spike-RBD sequence diversity and an analysis regarding changes in electrostatics and hydrophobicity, we develop and employ a linear interaction energy (LIE) model calibrated on experimental

binding data as well as umbrella-sampling molecular dynamics simulations. Our analysis focuses on primary effects of observed amino acid exchanges. Alterations of the glycosylation pattern or effects on the oligomerization of spike proteins were not considered at this stage. This approach delivers an early structural insight and binding-affinity estimates before experimental complex structures and experimental binding data will become available within the next months.

Results

In order to enable a comprehensive structural analysis of emerging partial or complete SARS-CoV-2 genomes, we follow a three-phase approach. In the first phase, we analyze new genomes (e.g. from GISAID or sequences directly provided by associated laboratories) for sequences with amino acid exchanges in regions of SARS-CoV-2 protein structures that are potentially relevant for drug-binding or cell-uptake.

In phase two, we look for sequences where one or more mutations are found that have not yet been investigated in our structural-analysis pipeline. In such a case, the sequence is submitted to a structure-modeling workflow, whose results can be used to predict an influence caused by the respective amino acid exchanges on active-site regions or cavities that are potentially relevant for drug-development. In the case of the spike protein, on whose RBD we focus herein, relevant models are analysed using our Catalophore-Halo technology (defined and illustrated below).

In phase three, we compare any modified RBD's Halo to the wild-type Halo. If such a Halo comparison shows a substantial change, we expose the new variant to a LIE-based molecular dynamics modelling pipeline in order to predict the corresponding change in binding affinity.

SARS-CoV-2 sequence analysis of spike-RBD diversity

By December 6th 2021, 5,799,116 SARS-Cov-2 genome sequences and 5,649,261 spike-protein sequences were available at GISAID²¹. For the sequence analysis of the SARS-CoV-2 RBD diversity we extracted 5,173,253 RBD sequences by filtering the spike-protein sequences for the RBD flanking residues "TSNF" (position 315-318) and "NFNG" (position 542-545) as well as for correct size of 223 amino acids. Within these sequences we identified a total number of 7,700,325 amino acid exchanges, which amounts to 0.67 % of all residues. Compared to August 2020, when 185 mutations were identified in 73,042 spike-RBD sequences¹³, this occurrence per sequence of mutations shows a 588-fold increase, highlighting the progressive evolution of this virus^{30,31}. The accumulation of mutations was significantly higher in the receptor binding motif (RBM), the 72 amino acid long sequence within the RBD, which mediates contacts with hACE2¹⁵: With a total of 7,451,124 amino acid exchanges, meaning 2.00 % of all residues within all of the collected RBM sequences, the relative number of mutations within the RBM increased 1,547-fold compared to August 2020, where 68 amino acid exchanges were found in 73,042 RBM sequences¹³.

Out of the 5,173,253 considered RBD and RBM sequences, 85.11 and 84.67 %, respectively, contain at least one amino acid exchange, with 99.48 % of all mutated RBDs bearing mutations within the RBM. The mean number of mutations within the RBD and RBM sequences in this analysis are 1.75 ± 0.57 and 1.70 ± 0.49 , respectively. With 15 amino acid substitutions located in the RBD, ten of which occur in the RBM, the recently discovered VOC Omicron stands out from previous variants by an at least three times higher number of mutations in this region²¹ (see also Figure 1b).

Referring to a list of 143 representative SARS-CoV-2 spike-protein sequences available at GISAID²¹, six of the amino acid exchanges within the Omicron RBD co-occur in other RBD variants, namely K417N (present in seven variants including VOC Beta), N440K (three variants), S477N (four variants), T478K (75 variants, including VOC Delta), E484A (two variants) and N501Y (33 variants including VOCs Alpha, Beta and Gamma). Their potential influences on one or more properties such as protein stability and flexibility as well as binding and sensitivity to antibodies have been evaluated, whereby exchanges N440K, S477N, T478K and N501Y were reported to possibly enhance binding affinity to hACE2^{13,32-36}.

The remaining nine mutations within the Omicron RBD (G339D, S371L, S373P, S375F, G446S, Q493R, G496S, Q498R and Y505H) do not occur in any of the 143 representative SARS-CoV-2 spike-protein sequences²¹. Nevertheless, possible influences of these amino acid exchanges have already been studied *in silico* or *in vitro*, with G339D, Q493R and Q498R being predicted to enhance binding to hACE2^{35,37,38}.

Visualization of amino acid exchanges using difference Halos

The influence of respective amino acid exchanges in the spike RBDs of VOCs Alpha, Beta, Delta and Omicron regarding crucial properties can be visualized by Catalophore Halos. Halos are multivariate physicochemical-property fields composed of points aligned to an equidistant grid. Herein, we show electrostatics (lower triangle, Figure. 2a) and hydrophobicity (upper triangle, Figure 2a) of the spike-RBD Halo. This depiction makes it easier to determine at a first glance, whether or not mutations have a noticeable influence on the RBD-hACE2 interface field and to decide whether or not a comprehensive structural analysis will be prioritized in such cases.

In order to spot the most prominent differences between variants, we chose to present the electrostatics and the hydrophobicity Halo point clouds (out of the 15 calculated physico-chemical properties), represented as a difference Halo matrix (see Figure 2). Differences were calculated by subtracting Halo 2 (columns of Figure 2a) from Halo 1 (rows). Regarding electrostatics and hydrophobicity, the Omicron Halo - among all the other variants plotted here - shows the most eye-catching differences compared to the wild-type Halo (Figure 2c and 2b). Interestingly, the Halos of the Beta and Gamma variants appear to be more different from the wild type regarding electrostatics than the Delta variant (Figure 2a), whereas the Alpha variant shows almost no difference regarding electrostatics and hydrophobicity.

By visual comparisons of the Halo cloud points regarding differences in property distribution of electrostatics and hydrophobicity, Omicron seems to be akin to a combination of Beta and Delta with

additional changes in the region where the mutation N440K is located (Figure 2b). The large difference in the Halo between the Omicron and the Delta variants is explained by the mutations G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, E484A, Q493R, G496S, Q498R, N501Y, Y505H only present in Omicron and the absence of L452R while only sharing T478K. At the binding interface, the new arginine replacing glutamine at Omicron spike position 493 yields a new strong hydrogen bond with ACE2 Glu35 which is clearly reflected by the dark blue area (increase in positive charge) of the Omicron-wildtype difference Halo as illustrated by Figure 2c. These depictions allow a rapid qualitative overview of the extent of physico-chemical changes at the RBD-hACE2 binding region. As a consequence, they quickly inform a decision on which variants to prioritize for further resource-intensive structural and dynamic analysis, e.g., calculating the binding affinities to hACE2 using the LIE prediction model we developed for this purpose.

LIE-model development and optimization

In order to achieve an as accurate as possible predictive model, we evaluated various simulation-time ranges and replicate numbers. For each combination of these two optimization parameters, we calculated two sets of binding energies ΔG , once by least-squares model fitting and once through leave-one-out cross-validation (LOOCV). The latter value set can safely be considered as a true prediction since none of the variants was included in the training set used for its evaluation.

We evaluated agreements between experimental and predicted ΔG values of each run-parameter combination by calculating corresponding mean absolute errors (MAE) given in kJ/mol and coefficients of determination (R^2). According to the last prediction column in Supporting Table 1 (visualized in Figure 3a), the highest prediction accuracy and lowest error was achieved with 50 replicates of 500 ps molecular dynamics (MD) simulation trajectories, where R^2 amounts to 0.79 and 0.71 with an average error of less than 1.9 and 2.1 kJ/mol for model fitting and cross-validation, respectively. Using MD simulations longer than 500 ps (1 and 2 ns) actually reduced the predictive accuracy during cross-validation in terms of both quality measures which are, by the way, strongly correlated ($r^2=0.92$). As expected, the highest accuracy was obtained with the highest number of replicates (50).

A corresponding scatter plot of predicted vs. experimental binding energies is depicted in Figure 3c. The ranking of variants in our training set, which comes from experimental measurements, is very well reflected by our model predictions, in particular comparing for instance the SARS-CoV-2 wild type (WT) with Alpha (green dots in Figure 3c). In addition, the distribution of experimental ΔG values in terms of mean and standard deviation (-55.76 ± 4.7 kJ/mol) agrees with those of the fitting (-55.83 ± 3.0 kJ/mol) and cross-validation procedure (-55.78 ± 4.0 kJ/mol) extraordinarily well, which was achieved with the optimal fitted weights, $w^{vdw}=0.758$ and $w^{elec}=0.028$ (see below).

For this large difference in weights between the Van der Waals and the electrostatic contributions, two explanations are natural. On the one hand, spike-hACE2 binding affinities might mainly be affected by hydrophobic rather than electrostatic interactions and, on the other hand, the computed electrostatic

interaction term might just not correlate well with experimental findings. However, having applied this predictive model to an unpublished internal set of 21 hACE2 variants in complex with the wild type for which EC50 values were available, we obtained a high prediction accuracy of $R^2=0.64$. Prior to that, EC50 values had been converted to binding free energies in analogy to K_D values.

Considering the number of replicates needed for a sufficient convergence of the energies and consequently of affinity-based variant rankings, the plot in Figure 3b indicates reasonable stability already from 30 replicates on, although MAE and R^2 are still slightly worse compared to 50 replicates (Table S1). We can safely assume a continuous growth of the model accuracy with additional replicates. However, with regard to the required run time (currently 2.7 days on a Amazon AWS C6i.16xlarge instance), 50 replicates seem to be a reasonable choice if we put more emphasis on accuracy. For this convergence analysis we determined the fraction of variant pairs associated with a swap in their energy order while increasing the number of replicates. All frequencies in percent were obtained via division by the total number of variant pairs $n=43^2=1849$. In more urgent cases, 30 replicates would take around one and a half days to run at a slightly lower accuracy.

Another major quality aspect of binding-affinity models is related to the predicted ranking of VOCs, as we strive for high consensus between predicted and experimental top-N variants. Figure 3d illustrates the fraction of variants among the top N predicted ones that are also included in the top N experimental variants of the training set. We compared the performance of an LIE model A) trained with 50 replicates and applied to predictions on the basis of 50 replicates as well, B) same as A) but using 30 replicates for both steps, and C) a mix of A) and B), namely model training with 50 but predictions on 30 replicates. Finally, a non-empirical prediction model published by Singh et al.¹³ in 2020 and based on umbrella sampling along with weighted-histogram analysis was applied to the same training set of 43 variants. As expected, the highest consensus, especially in the range of up to top 5, was obtained by models trained with 50 replicates, models A) and C) corresponding to the green and red plots (where green is mainly hidden behind the red line). Regarding predictions of training-set items, 80-100 % of the top 3-5 are in agreement with experimental top 3-5 candidates. Considerably less consensus was obtained with LIE model B) based on 30 training replicates yielding around 60-80 % at top 3-5 and, in particular, by the umbrella sampling method with 30-50 % consensus.

LIE application to VOCs

Using our final predictive model calibrated with the optimal weights

$$\Delta G = 0.758 \Delta E^{vdw} + 0.028 \Delta E^{elec} ,$$

we estimated binding affinities for VOCs that emerged during the past year: Alpha, Beta, Gamma, Delta, and most recently Omicron. However, we have to bear in mind that our structural variant models only comprise the spike RBD rather than the entire spike trimer. Table 2 shows not only our calculated binding free energies along with K_D values derived at 310 K temperature, but, for the purpose of rank comparison,

also experimental values for WT, Alpha, Beta, and Gamma recently determined by Barton et al. through surface plasmon resonance³².

Variant	Experimental binding affinities				LIE model prediction			
	K_D	K_D^*	K_D^*	ΔG	ΔG	K_D	K_D	$\Delta\Delta G$
	[nM]	[nM]	ratio	[kJ/mol]	[kJ/mol]	[nM]	ratio	[kJ/mol]
WT	74.4	1.70	1	-52.0	-53.1	1.11	1	0
Alpha	7.0	0.16	10.6	-58.1	-58.1	0.16	6.8	-4.9
Beta	20.0	0.46	3.7	-55.4	-53.2	1.07	1.0	-1.2
Gamma	13.5	0.31	5.5	-56.4	-54.2	0.74	1.5	-2.1
Delta	<i>n.d.</i>				-53.5	0.96	1.2	-1.5
Delta plus	<i>n.d.</i>				-52.3	1.56	0.7	-0.2
Omicron	<i>n.d.</i>				-56.8	0.27	4.1	-4.7

Table 1. SARS-CoV-2 VOCs binding free energies and dissociation constants predicted by LIE model and compared to K_D values determined through surface plasmon resonance by Barton et al.³² K_D^* refers to experimental K_D scaled to the value range of our training set using the corresponding WT K_D (1.7 nM) as a reference. K_D ratio relates all predicted as well as Barton’s measured (and scaled) K_D values to the corresponding wild-type reference K_D , whereas $\Delta\Delta G$ was calculated as the predicted binding-energy difference from the predicted WT energy (-50.9 kJ/mol).

For comparability reasons on the Gibbs energy level, Barton’s dissociation constants have been scaled to the value range of Zahradnik’s data/training set using the WT as an anchor point for which Zahradnik had published a K_D value of 1.7 nM. This scaling gives an additional column K_D^* determined according to equation

$$K_D^* = K_D^{Barton} \frac{K_{D,WT}^{Zahradnik}}{K_{D,WT}^{Barton}} = K_D^{Barton} \frac{1.70 \text{ nM}}{74.4 \text{ nM}} .$$

The K_D ratio column simply relates predicted K_D values to the WT reference K_D of 1.7 nM indicating how many times more strongly a particular mutant binds than the wild type. A final column with $\Delta\Delta G$ represents deviations of predicted binding energies from the WT binding energy.

For VOC binding affinity predictions, 240 replicates have been produced and analysed. According to these results in Table 1 and Figure 3c, the predicted order of dissociation constants for WT, Alpha, Beta and Gamma exactly reflects experimental findings^{32,39}. In particular, with a 6.8-fold decrease compared to the wild-type dissociation constant, the most outstanding binding affinity among all VOCs (as far as

experimental results are published) was predicted for Alpha. This is also in line with lab data^{32,39} featuring around 10-fold binding-affinity increase and reports⁴⁰ stating, for the UK between October and November 2020, a 70%–80% rise in transmissibility with respect to the wild type reference.

For the new SARS-CoV-2 Omicron spike variant, our studies reveal a similarly remarkable 4.1-fold increase of binding affinity compared to the wild type. In terms of binding strength it therefore ranges between Alpha and Gamma. A significantly lower binding affinity, somewhere between Beta and WT, was predicted for Delta and Delta plus. Moreover, Delta plus (having K417N in addition to the amino acid exchanges found in Delta) was predicted to have a lower affinity than Delta, which is supported by experimental results revealing weaker binding due to the presence of K417N^{39,41}. In addition, the authors allude to an increase of immune evasion caused by this mutation. It should be noted that Omicron as well includes this mutation possibly explaining its lower binding affinity compared to Alpha and thereby indicating a strong tendency to immune evasion for Omicron. The relative order of WT, Delta and Omicron is also in line with Docking results recently published by Kumar et al.⁴², although energy magnitudes and therefore differences given in that article translate to a K_D ratio of 10 million for Omicron compared to the WT.

Discussion

Following the trade-off model of virulence⁴³, the SARS-CoV2 virus, like other viruses⁴⁴, is constantly evolving⁴⁵ by modulating the rate of infectious transmission, higher virulence, and higher virus production in order to improve viral fitness. In this context, viral immune evasion is an evolutionary strategy to allow for the coexistence of viruses and their hosts⁴⁶ as shown with artificial polymutant SARS-CoV-2 spike-protein pseudotypes that resisted antibody neutralization to a similar degree as circulating VOCs.⁴⁷

Typically, viruses adapt to a specific host. However, when hosts fluctuate in time or space, generalist viruses may evolve as well⁴⁸. It is hypothesized that during this process, intermediate virulence maximizes pathogen fitness as a result of a trade-off between virulence and transmission.⁴⁹ Although it is often speculated, that virulence of viruses will finally decrease over time, the pandemics of our past, such as SARS in 2003 and flu in 1918-20, 1957, 1968 and 2009, decayed due to other reasons and not because the viruses evolved to cause milder symptoms.⁵⁰

During the ongoing COVID-19 pandemic it is therefore required to continuously adapt mitigation strategies, drugs, and vaccines in order to mitigate the impact during the early phase of the establishment of this virus. The world has employed massive sequencing programs (based on infected patients, animal, or large-scale sewage monitoring samples) that continuously provide information on changes in the viral genome. While this ensures an early detection of altered genomes, it by principle lacks information about the changed characteristics of the emerging variants. Therefore, only further and detailed investigations like structural analysis and the prediction of the biological impact of occurring mutations allow us to be

one step ahead of the virus by helping to forecast altered virulence, transmissibility, or immune-evasion potential.

Given the enormous number of emerging genomes, a fast tool such as Catalophore Halos can help to rapidly identify changes in RBD/hACE2 interface fields and guide a decision of which variants to simulate in full atomistic detail. Based on the LIE method, we have developed an empirical binding-affinity estimator that predicts remarkably accurate binding energies for spike-RBD-hACE2 complexes at moderate wall-clock run times, especially using massive cloud computing facilities. With the necessary caution, this technique helps to raise flags to indicate potential higher infectiousness. The remarkable precision of our SARS-CoV-2 spike-RBD-hACE2 binding-affinity model is mainly due to the high number of replicates used to achieve a satisfactory convergence in binding energies and both a solid and sufficiently large training set from one source, thanks to Zahradnik et al.³⁴ With a mean absolute error of around 2 kJ/mol our predictions are in the same error range as experimental binding-energy methods.⁵¹

Binding affinity is usually measured and indicated by the equilibrium dissociation constant (K_D), which is used to evaluate and classify the strength of biomolecular interactions. The smaller the K_D value, the greater the binding affinity of the ligand for its target. The actual K_D value relevant to a concrete biological situation depends on the physiological environment, e.g., salt concentration, temperature, or pH. Therefore, both the measured and modeled absolute K_D values are only valid within the observation range, making the comparison between unrelated data sources complex. Thus, we reported relative changes in K_D values that clearly show a structural, biologically explainable increase in the binding affinity of all VOCs, particularly pronounced for the new VOC Omicron.

Many of our computational results associated with model training as well as application to SARS-CoV-2 VOCs are in remarkable agreement with experimental findings. For instance, the relative order of binding-affinity predictions for spike VOCs are exactly in line with experimental observations^{32,39,41} of WT, Alpha, Beta, and Gamma, though, Beta was predicted very close to the wildtype. Moreover, the outstanding binding strength associated with Alpha as well as the decrease in binding affinity due to the immune-escape mutation K417N is also perfectly reflected. Since, according to our findings, Omicron achieved a substantially higher increase in binding affinity than all other VOCs except for Alpha and, in addition to that, contains a mutation associated with immune evasion, this new variant requires our special attention and monitoring.

In summary, our LIE prediction model allows us to estimate high-level binding affinities between the spike RBD and hACE2 regions. However, although we found the model fairly accurate when compared to experimental values, predicted affinities should be interpreted as binding "trends" instead of absolute K_D values. As noted elsewhere¹⁶⁻¹⁸, increased binding affinity often leads to increased infectivity of SARS-CoV-2. Given the experience with the global wave of VOC Delta, it is therefore now necessary to closely monitor the spread and impact of Omicron in the coming months. Initial medical reports⁵² from South Africa⁵³, Botswana, and Europe⁵⁴ with a relatively small, non-representative sample of patients revealed a

changed clinical picture with a comparatively mild course of the disease⁵⁵. As population immunity increases, either through infection or vaccination, steady modulation of immune-evading mutations might contribute to a permanent establishment of SARS-CoV-2, wherein Omicron apparently shows the potential of playing a significant role.

Methodology

Analysis of spike-RBD diversity

For the analysis of RBD diversity, all spike-protein sequences that were available by December 6th 2021 at GISAID²¹ were downloaded in FASTA format. Processing and analysis of the sequences was performed employing in-house tools in Python. The spike-RBD sequences (residues 319-541) were extracted by splitting each spike-protein sequence after the amino acid motif “TSNF” and before “NFNG”, which refer to the RBD flanking regions. Only sequences consisting of 223 residues (the length of the RBD) were accepted, i.e., insertions and deletions were not considered. Each residue of the retrieved RBD sequences was compared with the respective residue in the reference RBD²¹. Every mismatch, except for low-quality or sequencing error residues (indicated by “X”), was counted as one mutation. For the analysis of spike-RBD diversity, residues 437-508 were considered.

Preparation of spike-RBD-hACE2 structures

A 2.47-Å-resolution crystal structure of the SARS-CoV-2 spike RBD bound to hACE2 (PDB-Code: 6m0J) was used as a starting structure. Employing the molecular modeling and MD simulation package Yasara⁵⁶, conformational stress was removed by steepest descent energy minimization followed by simulated annealing (timestep 2 fs, atom velocities scaled down by 0.9 every 10th step). For this purpose, we selected the AMBER14 force field⁵⁷ applying an 8 Å force cutoff. SARS-CoV-2 spike-RBD variant sequences were constructed based on the SARS-CoV-2 spike RBD starting sequence from the wild-type lineage. Mutant structures were built using homology modelling by implementing Yasara with a maximum of five alignment variations per template and not more than 50 conformations tried per loop. The minimized starting structure served as a template. The final input files contained residues 333-526 of the respective SARS-CoV-2 RBD and residues 19-615 of hACE2 coordinating one zinc ion.

Molecular dynamics simulations

Again using Yasara, each spike-RBD-hACE2 complex was centered in a cuboid simulation box under periodic boundary conditions with a solute-wall distance of 5 Å on every side. This box was filled with explicit solvent molecules of the TIP3P model⁵⁸, approximating 0.997 g/mol water density and with 0.9 % sodium chloride ions as well as additional ions for system neutralization. Protein-structure topologies and energetics were parameterized according to the Amber14 force field. The system pressure was set to 1 bar using the Manometer1D setting of Yasara, while temperature coupling to 310 K was achieved by velocity rescaling as described by Krieger et al.⁵⁹ Long-range electrostatic interactions were treated

through Particle-Mesh Ewald⁶⁰ summation (PME) with an 8 Å cutoff. Our protocol consisted of a steepest-descent energy and successive simulated-annealing minimization step (following Yasara's standard minimization protocol) and a final production run of 500 ps (respectively 2 ns in case of model optimization) per replicate using a step size of 2 fs for intramolecular and 4 fs for intermolecular forces. Afterwards we extracted intermolecular interaction-energy contributions, E^{vdw} and E^{elec} , caused by Van der Waals and electrostatic forces, respectively. In addition, hACE2 in its dissociated state underwent the same procedure from simulation-box generation up to production MD in order to have both bound and unbound states available for the development of a predictive model.

Linear interaction energy model development

From the various MD-based approaches to binding-affinity estimation, the empirical linear interaction energy (LIE) method developed in the 90s by Åqvist et al.^{61,62} has provided a remarkable trade-off between predictive accuracy and computational effort.^{51,63–65} In contrast to complex methods mimicking thermodynamic reaction paths represented by high degrees of decomposition and numbers of long trajectories, it exclusively relies on short simulations of a bound/associated (PL) and an unbound/dissociated state (L) of the ligand as depicted by the embedded graphic in the upper left corner of Figure 3a. Starting from the linear-response approximation⁶⁶, the developers pointed out a strong relationship between the Gibbs free energies of binding on the one side and, on the other side, average (denoted by angle brackets $\langle \rangle$) differences ΔE^{vdw} and ΔE^{elec} of the ligand's interaction energies with its surrounding atoms, namely protein and solvent atoms in the bound and solely solvent atoms in the unbound case.

$$\Delta G = \frac{1}{2} \left[\langle E^{elec} \rangle_{PL} - \langle E^{elec} \rangle_L \right] + \alpha \left[\langle E^{vdw} \rangle_{PL} - \langle E^{vdw} \rangle_L \right].$$

The coefficient alpha was calibrated with respect to a small training set of protein-ligand complexes. Subsequent LIE studies revealed significantly better correlations when both coefficients, α and β , were treated as empirical parameters, possibly extended by further features, and fitted to available training sets with known binding affinities^{63,67,68}

$$\Delta G = \beta \left[\langle E^{elec} \rangle_{PL} - \langle E^{elec} \rangle_L \right] + \alpha \left[\langle E^{vdw} \rangle_{PL} - \langle E^{vdw} \rangle_L \right].$$

Our final LIE model of the spike-hACE2 binding indeed makes use of two empirical parameter weights w^{vdw} and w^{elec} (used instead of α and β and having omitted the angle brackets for the sake of clarity):

$$\Delta G = w^{vdw} \Delta E^{vdw} + w^{elec} \Delta E^{elec} . \quad (1)$$

Since we are dealing with two interacting proteins rather than a protein-ligand system, we arbitrarily considered hACE2 as a ligand molecule simulated in complex with the spike RBD as well as freely in

solvent. It must also be noted that due to the huge size of the hACE2 ligand and in order to reduce the impact of noise, only interactions of hACE2 amino acids within a 5 Å environment of spike RBD (brown colored area of hACE2 object in embedded graphic of Figure 3a) were evaluated upon energy computation in the bound as well as unbound case.

For model development and evaluation we employed a sufficiently large training set of 43 spike-RBD variants including the wild type (PDB ID 6M0J) in complex with hACE2 for which Zahradnik et al. had recently published K_D values obtained through yeast surface display titration³⁴. The number of mutations per variant ranges from one to seven, with a maximum distance to hACE2 amounting to 2.5 nm. Using a linear regression model

$$\underbrace{\begin{pmatrix} \Delta G_1^{\text{exp}} \\ \vdots \\ \Delta G_{43}^{\text{exp}} \end{pmatrix}}_y \approx \underbrace{\begin{pmatrix} \Delta \langle E_1^{\text{vdw}} \rangle & \Delta \langle E_1^{\text{elec}} \rangle \\ \vdots & \vdots \\ \Delta \langle E_{43}^{\text{vdw}} \rangle & \Delta \langle E_{43}^{\text{elec}} \rangle \end{pmatrix}}_A \underbrace{\begin{pmatrix} w^{\text{vdw}} \\ w^{\text{elec}} \end{pmatrix}}_x$$

along with least-squares fitting

$$x = (A^T A)^{-1} A^T y$$

interaction energy weights of the empirical model were fitted to experimental ΔG^{exp} values derived from K_D values at a temperature of $T=310$ K and using the gas constant R :

$$\Delta G = RT \ln K_D \quad (2)$$

Gibbs free energies of binding for new variants were then predicted by summing up the two weighted interaction-energy contributions according to Equation (1) and translated back to K_D values by inverting Equation (2).

Our predictive model was validated and optimized through leave-one-out cross-validation, ensuring that none of the 43 variants were inside the training set used to predict its binding energy. Mean absolute errors and squared Pearson correlation coefficients served as measures for model accuracy. In order to achieve a satisfactory binding-energy convergence and increase model accuracy, we produced and analysed 50 replicates of 2 ns MD trajectories and averaged the two interaction-energy terms.

Catalophore-Halo analysis

A Catalophore Halo is a multivariate property field composed of a collection of points in Cartesian space discretized onto an equidistant grid annotated with currently 19 physicochemical and statistical properties (e.g. electrostatics, hydrophobicity, flexibility, potential energies, hydrogen-bonding potential, or dissolvability) that are projected by a bio-molecule into its surroundings⁶⁹. In other words, a Halo point

cloud describes the spatial distribution of physical properties induced by an adjacent (bio)molecule. From the moment of its calculation, it is entirely independent of the underlying biomolecular structure.

For example, electrostatic potential influences are visible, either through charged surface spots or helical dipole moments⁷⁰⁻⁷² that are hardly captured by sequence or structure patterns. The foundation for this approach lies in the Ligsite algorithm⁷³ applied to identify protein cavities and modified to calculate properties of these enclosed spaces (thus the term "cavities"). Compared to Catalophore cavities⁷⁴ used for enzyme discovery and drug design⁷⁵, Halos are constructed at the proximity to the protein surface and are a priori unrestricted in size (within the limits of the entire protein's surface size) by principle.

Since not all Halo areas are significant, the relevance of a specific point is scored by a threshold function that pinpointly evaluates the physico-chemical properties and derives the required thickness of the Halo above the surface. For known protein-protein complexes like the spike-RBD-hACE2 complex, the space occupied by hACE2 is taken as a template for cropping the spike-RBD Halo using a 5 Å radius cutoff in addition. It is important to note that, on purpose, the counterpart protein (in this case hACE2) is not influencing the properties of the spike-RBD Halo.

Halo point clouds become particularly powerful when comparing the distribution of properties at the binding interface of similar molecules that are likely/able to form a complex with the same target molecule such as the spike-RBD variants with respect to hACE2. Thus, so-called difference Halos particularly highlight regions of significantly differing property values when comparing two given Halos. Difference Halos covering the binding interface of the spike VOCs were generated as follows: first, a correspondence list of point pairs of the two underlying clouds was created taking into account only pairs of points with a distance less than 0.5625 Å. A difference cloud was then generated by placing a new point at the geometric center of each pair of two corresponding points. Finally, property values for new points were calculated as the difference of the values associated with the corresponding pair of two original points, thereby subtracting values of the mobile point cloud from the fixed point cloud.

Computational details

The simulation performance was optimized in cooperation with Amazon Web Services (AWS), who supplied the necessary cloud infrastructure in the framework of the diagnostic development initiative. We used clusters of AWS Elastic Computing (EC2) x86 instances of the C5 und C6i families running 64-bit Amazon Linux 2 with AMI Kernel 5.10, e.g. c6i.8xlarge (32 virtual CPUs), c5.4xlarge (16 vCPUs) and c6i.16xlarge (64 vCPUs) equipped with Intel Xeon Scalable processors. With this configuration the simulation of each variant was completed within three to six days. Initial proof-of-principle simulations were executed with GROMACS 2021.2 on ARM-based AWS Graviton instances c6g.8xlarge (32 vCPUs).

Data availability

Publicly available datasets were analyzed in this study. This data can be found here: <https://www.gisaid.org/>. Input and final structure files as well as Pandas Dataframes of interaction

energies exported as Python Pickle files generated within this work are available for download at <https://doi.org/10.6084/m9.figshare.17129771>.

Declarations

Acknowledgments

Technical and infrastructure support was provided by the Amazon Web Services Diagnostic Development Initiative (DDI). The computational results presented in this manuscript have been natively produced in massive cloud computing facilities provided by Amazon Web Services within DDI, project nr. "CC ADV 00502188 2021 TR" entitled "virus.watch/SARSCoV-2". Financial support was provided by the Austrian Science Fund (FWF) through the doc.funds project DOC-46 "Catalox", the Doctoral Academy Graz of the University of Graz, the Austrian Centre of Industrial Biotechnology (Austrian Research Promotion Agency, FFG, project nr. 872161) in the Next Generation Bioproduction project nr. 92017 and of the Austrian Research Promotion Agency General Programme funding scheme project nr. 41404876 "VirtualCure - Rapid Development of an Automated & Expandable In-silico High-Throughput Drug Repurposing Screening Pipeline". Catalphore is a registered trademark (AT 295631) of Innophore GmbH. Calculations were carried out using the software described in the methods section embedded in the Catalphore™ Drug Solver platform with a non-commercial open-science license granted by Innophore GmbH. Initial spike models were generated within the FASTCURE consortium (<https://fastcure.net/>). We thank Hanna Lindermuth for reading and editing the manuscript. We thank all researchers who shared SARS-CoV-2 genome sequences in GISAID. A GISAID acknowledgment table containing sequence data used in this study is available at DOI 10.6084/m9.figshare.17129771.

Author Contributions

V.D. developed and parameterized the LIE model, performed MD simulations and analysis and drafted the manuscript with input from all authors. K.K. contributed in optimizing and analysing the LIE model, predicted the variant complex structures and contributed in data analysis and literature research. A.S. performed MD simulations, contributed to MD data analysis. M.H. contributed to analysis of data, contributed to structure preparations and prepared Halo figures. An.K. contributed to analysis of data, revised the manuscript, advised and contributed to data preparation. D.N., C.K. and K.K. wrote software to generate Halos and difference Halos. D.N., Al.K. contributed to the Halo data preparation, set up containers for cloud computing and gave technical support for using the AWS cloud resources. L.P. performed data preparation and sequence analysis, contributed to programming and analysis of sequence and genome data as well as to structural data analysis. C.K. contributed to the programming of difference cloud comparisons for Halo data calculation and preparation. L.C., M.K., R.B. gave technical, project and setup advice for resources in the AWS cloud for the MD simulations. T.P. gave structural advice and structural biology input for data analysis. V.R. contributed to data preparation, created and enhanced visualizations, contributed in evaluating, preparing and visualizing the data. K.G. gave structural and scientific advice, contributed in evaluating, preparing, and visualizing the data.

C.C.G. , G.S. contributed in evaluating, preparing and interpreting the data, designed, managed and supervised the project. All authors edited the manuscript to its final form.

Declaration of interests

V.D., K.K., A.S., An.K., D.N., Al.K., L.P., C.K., V.R. report working for Innophore. L.C., M.K., R.B., report working for Amazon Web Services, a company that also provides cloud computing services. K.G., G.S., C.C.G. report being shareholders of Innophore, an enzyme and drug discovery company. Additionally, G.S. and C.C.G. report being managing directors of Innophore. The research described here is scientifically and financially independent of the efforts in any of the above mentioned companies and open-science.

Additional Information

Supplementary information

Supporting_information.docx: A pdf document consisting of a supplementary table.

Competing financial interests: The authors declare no competing interests.

References

1. WHO. Coronavirus disease (COVID-19) pandemic, <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>. (2021).
2. Hale, T. *et al.* Government responses and COVID-19 deaths: Global evidence across multiple pandemic waves. *PLOS ONE* **16**, e0253116 (2021).
3. Martin, J. *et al.* Tracking SARS-CoV-2 in Sewage: Evidence of Changes in Virus Variant Predominance during COVID-19 Pandemic. *Viruses* **12**, (2020).
4. Vaughan, A. Delta to dominate world. (2021).
5. McCallum, M. *et al.* Molecular basis of immune evasion by the delta and kappa SARS-CoV-2 variants. *Science* eabl8506 (2021).
6. Yu, F., Lau, L.-T., Fok, M., Lau, J. Y.-N. & Zhang, K. COVID-19 Delta variants—Current status and implications as of August 2021. *Precis. Clin. Med.* (2021).
7. Parums, D. V. Revised World Health Organization (WHO) terminology for variants of concern and variants of interest of SARS-CoV-2. *Med. Sci. Monit. Int. Med. J. Exp. Clin. Res.* **27**, e933622-1 (2021).
8. Thye, A. Y.-K. *et al.* Emerging SARS-CoV-2 variants of concern (VOCs): an impending global crisis. *Biomedicines* **9**, 1303 (2021).
9. Open for outbreaks. *Nat. Biotechnol.* **38**, 377–377 (2020).

10. Hodgson, J. The pandemic pipeline. *Nat. Biotechnol.* **38**, 523–532 (2020).
11. Gorgulla, C. *et al.* A multi-pronged approach targeting SARS-CoV-2 proteins using ultra-large virtual screening. *iScience* **24**, 102021 (2021).
12. Gruber, C. C. & Steinkellner, G. Wuhan coronavirus 2019-nCoV—what we can find out on a structural bioinformatics level. *Innophore GmbH Austria* DOI: 10.6084/m9.figshare.11752749.v3, 24044224 Bytes (2020).
13. Singh, A., Steinkellner, G., Köchl, K., Gruber, K. & Gruber, C. C. Serine 477 plays a crucial role in the interaction of the SARS-CoV-2 spike protein with the human receptor ACE2. *Sci. Rep.* **11**, 4320 (2021).
14. Shang, J. *et al.* Cell entry mechanisms of SARS-CoV-2. *Proc. Natl. Acad. Sci.* **117**, 11727 (2020).
15. Shang, J. *et al.* Structural basis of receptor recognition by SARS-CoV-2. *Nature* **581**, 221–224 (2020).
16. Chen, C. *et al.* Computational prediction of the effect of amino acid changes on the binding affinity between SARS-CoV-2 spike RBD and human ACE2. *Proc. Natl. Acad. Sci.* **118**, e2106480118 (2021).
17. Augusto, G. *et al.* In vitro data suggest that Indian delta variant B.1.617 of SARS-CoV-2 escapes neutralization by both receptor affinity and immune evasion. *Allergy* all.15065 (2021) doi:10.1111/all.15065.
18. Ozono, S. *et al.* SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity. *Nat. Commun.* **12**, 848 (2021).
19. Annavajhala, M. K. *et al.* Emergence and expansion of SARS-CoV-2 B.1.526 after identification in New York. *Nature* **597**, 703–708 (2021).
20. Amoako DG, Everatt J, Scheepers C, Mohale T, Ntuli N, Mahlangu B, Mnguni A, Ismail A, Bhiman JN, Wolter N. *GISAID, Collected 2021-11-08, Originating laboratory: Lancet Laboratories Submitting laboratory: National Institute for Communicable Diseases of the National Health Laboratory Service.*
21. Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance* **22**, (2017).
22. Implications of the emergence and spread of the SARS-CoV-2 B.1.1.529 variant of concern (Omicron) for the EU/EEA. 7 (2021).
23. Hadfield, J. *et al.* Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121–4123 (2018).
24. Untersmayr, E. *et al.* Immunologically relevant aspects of the new COVID-19 vaccines—an ÖGAI (Austrian Society for Allergology and Immunology) and AeDA (German Society for Applied Allergology)

position paper. *Allergo J. Int.* **30**, 155–168 (2021).

25. Chaudhary, J. K. *et al.* Insights into COVID-19 Vaccine Development Based on Immunogenic Structural Proteins of SARS-CoV-2, Host Immune Responses, and Herd Immunity. *Cells* **10**, (2021).
26. Nabel Katherine G. *et al.* Structural basis for continued antibody evasion by the SARS-CoV-2 receptor binding domain. *Science* **0**, eabl6251.
27. Schrörs, B. *et al.* Large-scale analysis of SARS-CoV-2 spike-glycoprotein mutants demonstrates the need for continuous screening of virus isolates. *PLOS ONE* **16**, e0249254 (2021).
28. Zhang, H., Penninger, J. M., Li, Y., Zhong, N. & Slutsky, A. S. Angiotensin-converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: molecular mechanisms and potential therapeutic target. *Intensive Care Med.* **46**, 586–590 (2020).
29. Monteil, V. *et al.* Human soluble ACE2 improves the effect of remdesivir in SARS-CoV-2 infection. *EMBO Mol. Med.* **13**, e13426 (2021).
30. Singh, D. & Yi, S. V. On the origin and evolution of SARS-CoV-2. *Exp. Mol. Med.* **53**, 537–547 (2021).
31. Wu, A. *et al.* One year of SARS-CoV-2 evolution. *Cell Host Microbe* **29**, 503–507 (2021).
32. Barton, M. I. *et al.* Effects of common mutations in the SARS-CoV-2 Spike RBD and its ligand, the human ACE2 receptor on binding affinity and kinetics. *eLife* **10**, e70658 (2021).
33. Gan, H. H., Twaddle, A., Marchand, B. & Gunsalus, K. C. Structural Modeling of the SARS-CoV-2 Spike/Human ACE2 Complex Interface can Identify High-Affinity Variants Associated with Increased Transmissibility. *J. Mol. Biol.* **433**, 167051 (2021).
34. Zahradník, J. *et al.* SARS-CoV-2 variant prediction and antiviral drug design are enabled by RBD in vitro evolution. *Nat. Microbiol.* **6**, 1188–1198 (2021).
35. Starr, T. N. *et al.* Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* **182**, 1295-1310.e20 (2020).
36. Suleman, M. *et al.* Bioinformatics analysis of the differences in the binding profile of the wild-type and mutants of the SARS-CoV-2 spike protein variants with the ACE2 receptor. *Comput. Biol. Med.* **138**, 104936 (2021).
37. Fang, S. *et al.* In silico prediction of new mutations that can improve the binding abilities between 2019-nCoV coronavirus and human ACE2. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 1–1 (2021) doi:10.1109/TCBB.2021.3058265.

38. Xue, T. *et al.* Single point mutations can potentially enhance infectivity of SARS-CoV-2 revealed by in silico affinity maturation and SPR assay. *RSC Adv.* **11**, 14737–14745 (2021).
39. Bayarri-Olmos, R. *et al.* Functional Effects of Receptor-Binding Domain Mutations of SARS-CoV-2 B.1.351 and P.1 Variants. *Front. Immunol.* **12**, 4145 (2021).
40. Leung, K., Shum, M. H., Leung, G. M., Lam, T. T. & Wu, J. T. Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. *Eurosurveillance* **26**, 2002106 (2021).
41. Laffeber, C., de Koning, K., Kanaar, R. & Lebbink, J. H. G. Experimental Evidence for Enhanced Receptor Binding by Rapidly Spreading SARS-CoV-2 Variants. *J. Mol. Biol.* **433**, 167058 (2021).
42. Kumar, S., Thambiraja, T. S., Karuppanan, K. & Subramaniam, G. Omicron and Delta Variant of SARS-CoV-2: A Comparative Computational Study of Spike protein. *bioRxiv* 2021.12.02.470946 (2021) doi:10.1101/2021.12.02.470946.
43. Messenger, S. L., Molineux, I. J. & Bull, J. J. Virulence evolution in a virus obeys a trade off. *Proc. R. Soc. Lond. B Biol. Sci.* **266**, 397–404 (1999).
44. Morse, S. S. & Schluederberg, A. Emerging viruses: the evolution of viruses and viral diseases. *J. Infect. Dis.* **162**, 1–7 (1990).
45. Callaway, E. The coronavirus is mutating - does it matter? *Nature* **585**, 174–178 (2020).
46. Vossen, M. T., Westerhout, E. M., Söderberg-Nauclér, C. & Wiertz, E. J. Viral immune evasion: a masterpiece of evolution. *Immunogenetics* **54**, 527–542 (2002).
47. Schmidt, F. *et al.* High genetic barrier to SARS-CoV-2 polyclonal neutralizing antibody escape. *Nature* (2021) doi:10.1038/s41586-021-04005-0.
48. Elena, S. F., Agudelo-Romero, P. & Lalic, J. The Evolution of Viruses in Multi-Host Fitness Landscapes. *Open Virol. J.* **3**, 1–6 (2009).
49. Blanquart, F. *et al.* A transmission-virulence evolutionary trade-off explains attenuation of HIV-1 in Uganda. *Elife* **5**, e20492 (2016).
50. Ryan, J. R. Past pandemics and their outcome. *Pandemic Influenza Emerg. Plan. Community Prep. Boca Raton CRC Pr Llc* 3–23 (2008).
51. Foloppe, N. & Hubbard, R. Towards predictive ligand design with free-energy based computational methods. *Curr. Med. Chem.* **13**, 3583–3608 (2006).
52. Callaway, E. & Ledford, H. How bad is Omicron? What scientists know so far. *Nature* d41586-021-03614-z (2021) doi:10.1038/d41586-021-03614-z.

53. Pulliam, J. R. C. *et al.* Increased risk of SARS-CoV-2 reinfection associated with emergence of the Omicron variant in South Africa. *medRxiv* 2021.11.11.21266068 (2021)
doi:10.1101/2021.11.11.21266068.
54. Torjesen, I. Covid-19: Omicron may be more transmissible than other variants and partly resistant to existing vaccines, scientists fear. *BMJ* **375**, n2943 (2021).
55. Mallapaty, S. Omicron-variant border bans ignore the evidence, say scientists. *Nature* d41586-021-03608-x (2021) doi:10.1038/d41586-021-03608-x.
56. Krieger, E. & Vriend, G. New ways to boost molecular dynamics simulations. *J. Comput. Chem.* **36**, 996–1007 (2015).
57. Maier, J. A. *et al.* ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).
58. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
59. Krieger, E., Darden, T., Nabuurs, S. B., Finkelstein, A. & Vriend, G. Making optimal use of empirical energy functions: Force-field parameterization in crystal space. *Proteins Struct. Funct. Bioinforma.* **57**, 678–683 (2004).
60. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).
61. Åqvist, J., Medina, C. & Samuelsson, J. E. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* **7**, 385–391 (1994).
62. Hansson, T. & Åqvist, J. Estimation of binding free energies for HIV proteinase inhibitors by molecular dynamics simulations. *Protein Eng.* **8**, 1137–1144 (1995).
63. Nervall, M., Hanspers, P., Carlsson, J., Boukharta, L. & Åqvist, J. Predicting Binding Modes from Free Energy Calculations. *J. Med. Chem.* **51**, 2657–2667 (2008).
64. van Lipzig, M. M. H. *et al.* Prediction of ligand binding affinity and orientation of xenoestrogens to the estrogen receptor by molecular dynamics simulations and the linear interaction energy method. *J. Med. Chem.* **47**, 1018–1030 (2004).
65. Stjernschantz, E. *et al.* Are Automated Molecular Dynamics Simulations and Binding Free Energy Calculations Realistic Tools in Lead Optimization? An Evaluation of the Linear Interaction Energy (LIE) Method. *J. Chem. Inf. Model.* **46**, 1972–1983 (2006).
66. Lee, F. S., Chu, Z., Bolger, M. & Warshel, A. Calculations of antibody-antigen interactions: microscopic and semi-microscopic evaluation of the free energies of binding of phosphorylcholine

analogs to McPC603. *Protein Eng.* (1992) doi:10.1093/PROTEIN/5.3.215.

67. Vosmeer, C. R. *et al.* Towards Automated Binding Affinity Prediction Using an Iterative Linear Interaction Energy Approach. *Int. J. Mol. Sci.* **15**, 798–816 (2014).
68. Durmaz, V., Schmidt, S., Sabri, P., Piechotta, C. & Weber, M. Hands-off Linear Interaction Energy Approach to Binding Mode and Affinity Estimation of Estrogens. *J. Chem. Inf. Model.* **53**, 2681–2688 (2013).
69. Gruber, K., Steinkellner, G. & Gruber, C. Determining novel enzymatic functionalities using three-dimensional point clouds representing physico chemical properties of protein cavities, WO2014080005A1. (2020).
70. Adhya, L., Mapder, T. & Adhya, S. Role of terminal dipole charges in aggregation of α -helix pair in the voltage gated K⁺ channel. *Biochim. Biophys. Acta BBA - Biomembr.* **1828**, 845–850 (2013).
71. Hol, W. G. J., Halie, L. M. & Sander, C. Dipoles of the α -helix and β -sheet: their role in protein folding. *Nature* **294**, 532–536 (1981).
72. Sengupta, D., Behera, R. N., Smith, J. C. & Ullmann, G. M. The α Helix Dipole: Screened Out? *Structure* **13**, 849–855 (2005).
73. Hendlich, M., Rippmann, F. & Barnickel, G. LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. *J. Mol. Graph. Model.* **15**, 359–363 (1997).
74. Steinkellner, G. *et al.* Identification of promiscuous ene-reductase activity by mining structural databases using active site constellations. *Nat. Commun.* **5**, 1–9 (2014).
75. Prattes, M. *et al.* Structural basis for inhibition of the AAA-ATPase Drg1 by diazaborine. *Nat. Commun.* **12**, 3483 (2021).

Figures

Figure 1

Overview of structural models of the spike protein of SARS-CoV-2 including mutation locations compared to the wild type. One of the trimer chains is shown as a transparent surface with the RBD domain on top of the spike protein a: Delta variant, mutation locations are shown as yellow spheres b: Omicron variant, mutation locations are shown as blue spheres. The top box shows a zoom to the RBD with the calculated Halo point cloud of the RBD.

Figure 2

Surface representation of VOC Halos. a: Upper triangle: hydrophobicity difference Halos, difference values are colored from red -0.33 to blue +0.33, where white 0 means no difference. Lower triangle: electrostatics difference Halos, difference values are colored from red -1.0 to blue +1.0. Diagonal: binding interfaces between variant spike RBDs and hACE2 b: Hydrophobicity difference Halo between WT and Omicron variant. c: Electrostatics difference Halo between WT and Omicron variant. d: Binding interface between Omicron variant and hACE revealing an additional H bond between spike R493 and ACE2 E35.

Figure 3

LIE model development. a: Optimization of run parameters time range and number of replicates. b: Binding-energy convergence (rank-swap frequency and average cumulative energies with standard-deviation range) vs. number of MD replicates. c: Predicted (model fitting/cross-validation and VOC prediction) vs. experimental binding affinities of a 43 variants training set plus available VOCs calculated for hACE2 residues in a 5 Å vicinity of the spike RBD (brown colored region in embedded graphic). d: Consensus fraction between predicted and experimental top N variants

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [OmicronSIpreprint1.docx](#)