Article

Prediction and Evolution of the Molecular Fitness of SARS-CoV-2 Variants: Introducing SpikePro

Fabrizio Pucci [1,2] and Marianne Rooman[1,2,*]

[1] Computational Biology and Bioinformatics, Université Libre de Bruxelles, 1050 Brussels, Belgium; fapucci@ulb.be
[2] Interuniversity Institute of Bioinformatics in Brussels, 1050 Brussels, Belgium
* Correspondence: Marianne.Rooman@ulb.be

Abstract: The understanding of the molecular mechanisms driving the fitness of the SARS-CoV-2 virus and its mutational evolution is still a critical issue. We built a simplified computational model, called SpikePro, to predict the SARS-CoV-2 fitness from the amino acid sequence and structure of the spike protein. It contains three contributions: the inter-human transmissibility of the virus predicted from the stability of the spike protein, the infectivity computed in terms of the affinity of the spike protein for the ACE2 receptor, and the ability of the virus to escape from the human immune response based on the binding affinity of the spike protein for a set of neutralizing antibodies. Our model reproduces well the available experimental, epidemiological and clinical data on the impact of variants on the biophysical characteristics of the virus. For example, it is able to identify circulating viral strains that, by increasing their fitness, recently became dominant at the population level. SpikePro is a useful, freely available instrument which predicts rapidly and with good accuracy the dangerousness of new viral strains. It can be integrated and play a fundamental role in the genomic surveillance programs of the SARS-CoV-2 virus that, despite all the efforts, remain time-consuming and expensive.

Keywords: viral fitness; SARS-CoV-2; COVID-19; spike protein variants; deep mutagenesis; protein stability; protein binding affinity; immune escape

1. Introduction

Despite mitigation measures put in place around the world to slow down the fast spreading of the SARS-CoV-2 virus, the COVID-19 viral pandemic continues to have global devastating effects, with more than 150,000,000 people infected and 3,200,000 deaths at the end of April 2021 [1]. Lots of efforts and resources have been devoted in the last year to develop vaccines and new therapeutics in response to the SARS-CoV-2 infection [2,3]. Several vaccines such as mRNA-1273 [4], BNT162b2 [5], AZD1222 [6], Sputnik V [7], Ad26 [8] and NVX-CoV2373 [9] have proven to be safe and efficacious against the viral agent and have recently been approved by the regulatory agencies for emergency use. Thanks to these developments, large-scale vaccine administration is now ongoing throughout the world.

Moreover, while the pathogenic mechanisms of the viral infection are not totally clear [10], effective therapeutic agents have been developed. For example, neutralizing antibodies (nAbs) targeting the viral spike protein or human convalescent plasma have been employed in clinical practice by passively transferring them to patients [11–14]. Although with varying degrees of effectiveness, these therapies generally tend to improve the disease conditions and to reduce viral load especially if administered in early phases of the disease.

The increase in viral immunity at the population level due to infection, vaccination or passive immunization via nAbs clearly results in a stronger selection pressure on the SARS-CoV-2 virus [15,16]. This causes the emergence of new variants of the virus which are able to escape from the immune response. Lots of computational and experimental

studies are currently focusing on the understanding of these escape mechanisms in the SARS-CoV-2 viral infection [17–21] and on setting up SARS-CoV-2 immune surveillance of the world's population to track and eventually limit the spreading of potentially escaping variants [22–27].

However, the prediction of how SARS-CoV-2 evolves under this selective pressure is far from obvious. Indeed, even though SARS-CoV-2 has a moderate mutation rate compared to other RNA viruses due to its more accurate replication [28], tracking viral dynamics in the huge space of possible variant combinations (including also deletions and insertions) under the influence of human immunity makes predictions highly challenging. Extensive large-scale monitoring of SARS-CoV-2 evolution and host immunity will help to better understand these issues [28].

In this paper, we performed an extensive computational analysis of the mutational mechanisms that lead to the emergence of SARS-CoV-2 strains with increased fitness, with the aim to better understand the molecular mechanisms that drive viral adaptation and escape from the human immune system. We performed in silico mutagenesis experiments and predicted the impacts of mutations in the spike protein on its stability and on its affinity for nAbs and for the angiotensin-converting enzyme 2 (ACE2), known to be the SARS-CoV-2 entry point into the cells. We validated these predictions on viral variants for which experimental, epidemiological or clinical data have been obtained, and especially on the variants that are emerging and rapidly spreading to become prevalent genotypes. Our predictions are of utmost importance to help monitor the future evolutionary dynamics of SARS-CoV-2 and to identify the emergent strains whose spread will have to be limited via either the design of new vaccines or new mitigation measures.

2. Methods

2.1. Spike Protein Structures

The spike protein or S-protein of the SARS-CoV-2 virus (Uniprot code P0DTC2) is a homotrimeric glycoprotein attached to the viral membrane. It can adopt two forms, a closed and an open form. The transition between these forms increases the solvent exposure of the protein's receptor-binding domain (RBD), which encompasses residues 333–526 and mediates the fusion of the membranes of the virus and of the host's cells.

The 3-dimensional (3D) structures of the two forms have been experimentally resolved by cryo-electron microscopy (cryo-EM) and are deposited in the Protein DataBank (PDB) [29]. The closed form, with PDB code 6VXX, has a resolution of 2.80 Å [30], and the open form, 6VYB, has a resolution of 3.20 Å. These structures have thus quite a low resolution and do not contain all the residues of the spike protein. To obtain structures of the closed and open forms without missing residues, we modelled the complete amino acid sequence using the PDB structures 6VXX and 6VYB as templates and the homology modelling webserver SWISS-MODEL [31].

More accurate structures, resolved by X-ray crystallography, are available for the RBD of the spike protein. We used the PDB structure 6M0J [32] for this region, which contains the RBD bound to ACE2, with a resolution of 2.45 Å.

Furthermore, we set up a dataset of spike protein/nAb complexes taken from [33], referred to as $\mathcal{D}^{nAb}$. We used the following selection criteria:

- Human monoclonal nAbs generated in response to SARS-CoV-2 infection;
- nAbs targeting the spike protein;
- nAbs/spike protein complexes available in the PDB, with X-ray structure of resolution ≤ 3.2 Å.

$\mathcal{D}^{nAb}$ contains 31 structures of nAbs/spike protein complexes, listed in the GitHub repository github.com/3BioCompBio/SpikeProSARS-CoV-2. These nAbs exclusively target the RBD of the spike protein, and are assumed to mimic the diversity of the human immune B-cell repertoire.

2.2. Spike Protein Stability

To compute the change in folding free energy upon point mutations in the spike protein, we used the PoPMuSiC algorithm [34], which is based on the 3D structure of the target protein and a combination of statistical mean-force potentials. These potentials use a coarse-grained representation of protein structures and were derived from frequencies of association between sequence and structure motifs observed in a non-redundant set of well-resolved 3D structures, which were transformed into free energies using the inverse Boltzmann law. They take into account implicitly the effect of the solvent and thus drastically reduce the computational cost of the algorithms that use them. For further details about PoPMuSiC and its energy functions, see [34].

We applied PoPMuSiC to the modelled structures of the open and closed forms of the full spike protein (obtained using 6VYB and 6VXX as templates), and to the experimental structure of the RBD domain (6M0J), which are described in Section 2.1, and used it to compute the effect of all possible single-site mutations on the spike protein stability. More precisely, PoPMuSiC provided the change in folding free energy $\Delta\Delta G_i$ caused by each mutation i in each of the three spike protein structures considered. The $\Delta\Delta G_i^{\mathrm{S}}$ value used in what follows was obtained with the following rules: for mutations of residues in the RBD, we considered the $\Delta\Delta G_i$ based on the 6M0J structure of the RBD; for mutations of other residues, we averaged the predicted $\Delta\Delta G_i$'s obtained from the 6VYB- and 6VXX-based models.

2.3. Spike Protein/ACE2 Binding Affinity

For the changes in binding affinity upon single-site mutations, we used the BeAtMuSiC predictor [35], which is a linear combination of free energy values predicted by PoPMuSiC on the protein complex and on the separate partners. We applied BeAtMuSiC to predict the effect of all possible single-site mutations in the viral spike protein on its binding affinity for the ACE2 receptor of the host, which allows entry of SARS-CoV-2 virus into cells. For this purpose, we considered the X-ray structure 6M0J of the RBD/ACE2 receptor complex (see Section 2.1) as input, and computed the change in binding free energy $\Delta\Delta G_i^{\mathrm{ACE2}}$ of the RBD/ACE2 complex upon mutations i in the RBD. Mutations in the spike protein outside the RBD were assumed to have no effect on ACE2 binding, even though they might play a role due to long-range effects [36].

2.4. Spike Protein/nAb Binding Affinity

The changes in binding affinity between the spike protein and the 31 nAbs from the $\mathcal{D}^{\mathrm{nAb}}$ set (see Section 2.1) caused by all possible point mutations in the spike protein were also estimated using BeAtMuSiC [35]. We computed the effect of each mutation i on the binding affinity $\Delta\Delta G_i^{\mathrm{nAb}}(p)$ of each nAb/spike protein complex p, and computed their mean value over the 31 complexes from $\mathcal{D}^{\mathrm{nAb}}$:

$$\overline{\Delta\Delta G_i^{\mathrm{nAb}}} = \frac{1}{n_i} \sum_{p=1}^{n_i} \Delta\Delta G_i^{\mathrm{nAb}}(p) \tag{1}$$

where n_i is the number of structures that include the mutation i. Indeed, the structures of the nAb/spike protein complexes do not cover exactly the same region of the spike protein.

2.5. SARS-CoV-2 Fitness

Viral fitness is a parameter related to how efficiently the virus produces infectious progeny [37]. Despite this simple definition, characterizing fitness quantitatively is very challenging [38], since it is a fairly complex function of different features among which the viral inter-host transmissibility, its infectivity and its ability to escape from the host's immune response [39]. In this paper, we estimated the global fitness Φ_i of a variant i of

the SARS-CoV-2 virus on the basis of a simplified model which only takes into account the spike protein. More precisely, we defined it as a product of three fitness contributions:

$$\Phi_i = \phi_i^{\mathrm{S}} \times \phi_i^{\mathrm{ACE2}} \times \phi_i^{\mathrm{nAb}} \tag{2}$$

where ϕ^{S}, ϕ^{ACE2} and ϕ^{nAb} represent the relative propensities of the mutant virus to be transmitted between hosts, to infect the host, and to escape from the host's immune system. These propensities are assumed to be higher for spike protein variants i that are stabler [40] ($\Delta\Delta G_i^{\mathrm{S}} < 0$), that have greater binding affinity for the ACE2 receptor [41] ($\Delta\Delta G_i^{\mathrm{ACE2}} < 0$), and that have lower binding affinity for nAbs ($\overline{\Delta\Delta G}_i^{\mathrm{nAb}} > 0$), respectively. We thus defined the fitness contributions ϕ_i^{S} and ϕ_i^{ACE2} of a mutation i to be a positive decreasing function of $\Delta\Delta G_i^{\mathrm{S}}$ and $\Delta\Delta G_i^{\mathrm{ACE2}}$, respectively, and ϕ_i^{nAb} a positive increasing function of $\overline{\Delta\Delta G}_i^{\mathrm{nAb}}$. More precisely:

$$
\begin{aligned}
\phi_i^{\mathrm{S}} &= \exp\left[-\max\left(\frac{\Delta\Delta G_i^{\mathrm{S}}}{\mathrm{kcal/mol}}, \beta^{\mathrm{S}} \right) + \mu^{\mathrm{S}} \right] \\
\phi_i^{\mathrm{ACE2}} &= \exp\left[-\max\left(\frac{\Delta\Delta G_i^{\mathrm{ACE2}}}{\mathrm{kcal/mol}}, \beta^{\mathrm{ACE2}} \right) + \mu^{\mathrm{ACE2}} \right] \\
\phi_i^{\mathrm{nAb}} &= \exp\left[\min\left(\frac{\overline{\Delta\Delta G}_i^{\mathrm{nAb}}}{\mathrm{kcal/mol}}, -\beta^{\mathrm{nAb}} \right) - \mu^{\mathrm{nAb}} \right]
\end{aligned}
\tag{3}
$$

where μ^{S}, μ^{ACE2}, μ^{nAb}, β^{S}, β^{ACE2} and β^{nAb} are parameters. The choice of the ϕ-functions and parameters is justified as follows:

- Mutations i that strongly destabilize the spike protein ($\Delta\Delta G_i^{\mathrm{S}} \gg 0$ kcal/mol) or its binding to ACE2 ($\Delta\Delta G_i^{\mathrm{ACE2}} \gg 0$ kcal/mol), or that stabilize its binding with nAbs ($\overline{\Delta\Delta G}_i^{\mathrm{nAb}} \ll 0$ kcal/mol) have a fitness close to zero.
- Mutations that stabilize the spike protein ($\Delta\Delta G_i^{\mathrm{S}} < 0$ kcal/mol) or its binding to ACE2 ($\Delta\Delta G_i^{\mathrm{ACE2}} < 0$ kcal/mol), or that destabilize its binding to nAbs ($\overline{\Delta\Delta G}_i^{\mathrm{nAb}} > 0$ kcal/mol) have an evolutionary advantage and a fitness higher than one.
- To avoid excessively high fitness values, we cut the exponential growth of the ϕ-functions for $\Delta\Delta G_i = \beta$, with $\beta = \beta^{\mathrm{S}} = \beta^{\mathrm{ACE}} = \beta^{\mathrm{nAb}}$ chosen to be -1, similarly to what has been proposed in [42].
- The folding free energy changes predicted by PoPMuSiC have been shown to be biased towards destabilizing mutations [43,44]. To correct for this effect, the μ^{S} parameter was chosen to be equal to 0.5. The changes in binding free energy predicted by BeAtMuSiC have an analogous bias, as they are constructed from PoPMuSiC scores. Following the BeAtMuSiC construction detailed in [35], a bias in the PoPMuSiC energy value of 0.5 kcal/mol results in a bias in the BeAtMuSiC energy value of 0.19 kcal/mol. We thus fixed $\mu^{\mathrm{S}} = 0.50$ and $\mu^{\mathrm{ACE}} = \mu^{\mathrm{nAb}} = 0.19$.
- We set by definition the fitness value of the wild-type equal to one: $\phi_0^{\mathrm{S}} = \phi_0^{\mathrm{ACE2}} = \phi_0^{\mathrm{nAb}} = 1$.

Note that PoPMuSiC and BeAtMuSiC are implicitly based on the approximation that variants do not impact too strongly on the target protein structure. We thus neglect here large conformational rearrangements in the spike protein and possible effects of allosteric communication.

The global viral fitness, which takes into account multiple mutations in the spike protein, is defined as the product of the fitness values of all point mutations i as:

$$\Phi = \prod_i^m \Phi_i \tag{4}$$

where m corresponds to the total number of mutations in the spike protein relative to the wild-type strain. Note that, in doing so, we considered the mutations as independent and discard possible epistatic effects.

3. Results

3.1. Computational Pipeline

In its viral evolution, SARS-CoV-2 and our immune system are constantly engaged in what is known as a cat-and-mouse game, where SARS-CoV-2 attempts to increase its fitness by increasing the inter-host transmissibility, the infectivity of the host and/or the escape from the host's immune response. To quantitatively describe the viral fitness landscape, we developed a simplified model in which we focused on the spike protein. This protein, which protrudes from the virus surface, is a crucial component of the infection, as its binding to the ACE2 receptor of the host mediates the entry of the virus into the host's cells. The binding affinity of the spike protein for ACE2 has thus been related to SARS-CoV-2 infectivity [41]. The stability properties of the spike protein itself are another key element in the viral infection which has been related to the viral transmissibility between hosts [40].

Moreover, the spike protein is a major inducer of the host's immune response [19,27]. We mimicked the effect of the immune system on the SARS-CoV-2 virus through a set of 31 nAb/spike protein complexes contained in the dataset $\mathcal{D}^{\mathrm{nAb}}$ (see Section 2.1). We observed that these nAbs target exclusively the RBD of the spike protein and that the epitopes cover almost the entire RBD surface, as shown in Figure 1. It is interesting to note that the epitopes targeted by the majority of these nAbs are situated in or close to the RBD region that binds to ACE2, as seen from comparing Figure 1a and Figure 1b; these epitopes are thus likely to be immunodominant. A recent investigation suggests that RBD-binding antibodies are the major contributors of the neutralizing activity in convalescent human plasma [19,27]. This justifies our approximation of considering the nAbs of the set $\mathcal{D}^{\mathrm{nAb}}$ as representative of the immune response.
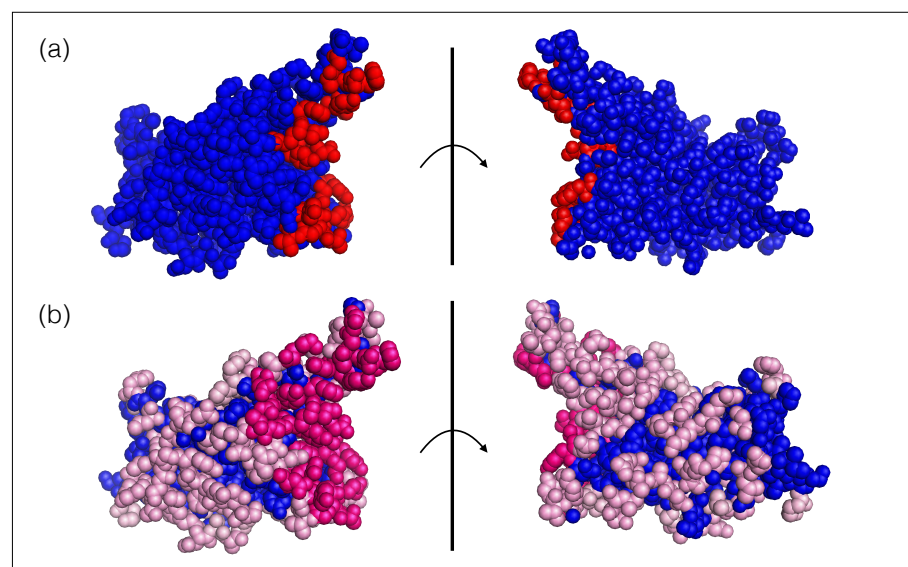
Figure 1. Receptor binding domain of the SARS-CoV-2 spike protein (PDB code 6M0J). The two views are related by a 180° rotation with respect to the plane (shown as a vertical line) representing the ACE2 binding interface. (**a**) The ensemble of residues that bind to ACE2 are colored in red spheres and the other residues are in blue. (**b**) The ensemble of epitope residues targeted by at least one nAb and less then ten nAbs of the $\mathcal{D}^{\mathrm{nAb}}$ set are in light pink spheres, those targeted by ten or more nAbs are in dark pink spheres and the other, non-epitope, residues are in blue. The list of epitope residues and ACE2 binding sites can be found in our GitHub repository (https://github.com/3BioCompBio/SpikeProSARS-CoV-2/blob/main/Structures/Epitope.dat, accessed on 10 April 2021).

The SARS-CoV-2 fitness, defined qualitatively on the basis of its efficiency to produce infectious progeny [37], is complicated to define quantitatively [38]. It depends in a complex manner on a series of features, among which the transmissibility of the virus between hosts, the infectivity of the host and the ability of the virus to escape from the host's immune response [39]. We thus approximated the global SARS-CoV-2 fitness Φ as a product of three fitness contributions ϕ^S, ϕ^{ACE2} and ϕ^{nAb}, which describe the transmissibility, infectivity and escape features, respectively (Equations (1)–(4)). In addition, we focused exclusively on the impact of variants of the spike protein. We estimated the three fitness contributions ϕ^S, ϕ^{ACE2} and ϕ^{nAb} in terms of the change in folding free energy upon mutation of the spike protein ($\Delta\Delta G^S$), the change of its binding affinity for ACE2 ($\Delta\Delta G^{ACE2}$) and the change of its binding affinity for a set of nAbs ($\overline{\Delta\Delta G}^{nAb}$), using statistical physics-based approaches and more specifically the PoPMuSiC [34] and BeAtMuSiC [35] algorithms, as detailed in Sections 2.2–2.4. Note that the effect of multiple mutations on the fitness were considered as independent and thus simply multiplied (Equation (4)).

In order to identify mutations in the spike protein that increase or decrease the SARS-CoV-2 transmissibility or infectivity, or that facilitate or block the escape from the protective immunity elicited by the infection, we constructed a computational pipeline of three steps, schematically represented in Figure 2, in which we estimated $\Delta\Delta G^S$ and ϕ^S, $\Delta\Delta G^{ACE2}$ and ϕ^{ACE2}, and $\overline{\Delta\Delta G}^{nAb}$ and ϕ^{nAb}. Using this pipeline, we performed large-scale computational mutagenesis experiments, in which we introduced basically all point mutations in the spike protein and predicted their effect on viral fitness. In what follows, we confronted these predictions with a large series of available experimental, epidemiological and clinical data on the SARS-CoV-2 infection and evolution.

$$\Phi = \phi^S \times \phi^{ACE2} \times \phi^{nAb}$$
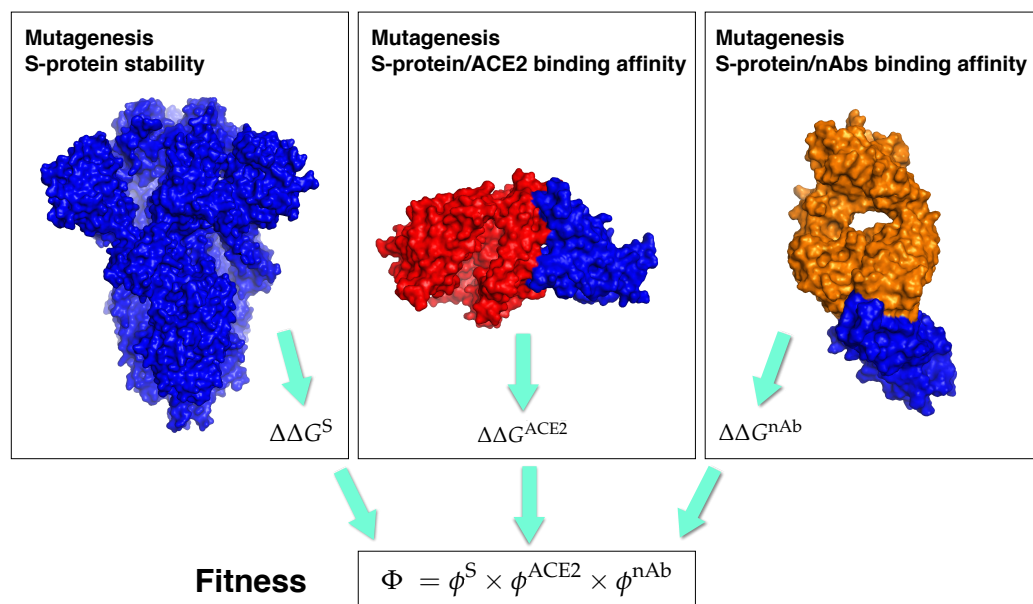
Figure 2. Schematic representation of the three steps of our computational pipeline: in silico mutagenesis experiments to compute the change in stability of the spike protein, and its change in binding affinity for ACE2 and for the 31 nAbs from $\mathcal{D}^{nAb}$. The spike protein is in blue, ACE2 in red and the antigen-binding fragment of a nAb in orange. The structures used for the pictures on the left, center and right have the PDB codes 6VXX, 6M0J and 7B3O, respectively.

Our prediction pipeline, called SpikePro, is freely available as an easy-to-use c++ program, which needs a variant spike protein sequence in fasta format as input. It outputs the sequence alignment with the reference spike protein (Uniprot code P0DTC2), the list of all point mutations introduced and the predicted overall viral fitness Φ. It can be downloaded from github.com/3BioCompBio/SpikeProSARS-CoV-2.

### 3.2. Spike Protein Stability and SARS-Cov-2 Transmissibility

We performed a large in silico mutagenesis experiment to study the influence of mutations on spike protein stability and thus on inter-host viral transmissibility [40]. Using PoPMuSiC [34], we computed the change in folding free energy $\Delta\Delta G_i^S$ of all possible single-site mutations i in the spike protein, and the corresponding fitness contribution ϕ_i^S defined in Equation (4).

As a first check of our method, we analyzed the relation between the predicted $\Delta\Delta G_i^S$ values for all point mutations in the RBD domain and the measured effects of these variants on spike protein expression [45]. These measurements were done using a yeast surface display platform, in which protein expression was quantitatively determined at large scale via flow cytometry. Even though protein expression and stability are only partially correlated, we found a good Pearson correlation coefficient of -0.51 (p-value $< 10^{-200}$) between the measured expression and the predicted $\Delta\Delta G_i^S$ values, which can be considered as the first validation of our approach.

To analyze the relation between stability predictions and epidemiological data, we compared the computed spike protein stability changes $\Delta\Delta G_i^S$ with the observed mutation rate R_i. We estimated R_i as the number of occurrences of each point mutation i in the set of about 7.8×10^5 SARS-CoV-2 spike protein sequences collected in the GISAID database [46], divided by the number of residues in the spike protein. We analyzed R_i as a function of the predicted $\Delta\Delta G_i^S$ values for all possible mutations i in the whole spike protein. As seen in Figure 3a, the majority of mutations that became dominant during the evolutionary trajectory show a slight increase in the spike protein stability, with $\Delta\Delta G_i^S$ between -1 and 0 kcal/mol. A smaller number of dominant variants have their stability slightly decreased with $\Delta\Delta G_i^S$ between 0 and 1 kcal/mol. Outside of this free energy interval, the rate R_i almost vanishes.

Moreover, we found a very good agreement between the predicted fitness ϕ_i^S and the R_i rate, as seen in Figure 3b. Indeed, variants that are predicted to be fitter than the wild type protein, and especially variants i with $\phi_i^S > 2$, have a high R_i rate, which means that they circulate a lot and became fixed during viral evolution. We will deepen this point in Sections 3.6 and 3.7.

It is important to underline that we did not fit any parameters of our model on the SARS-CoV-2 data. Thus, this prediction as well as all the predictions presented in the following sections are truly blind predictions.
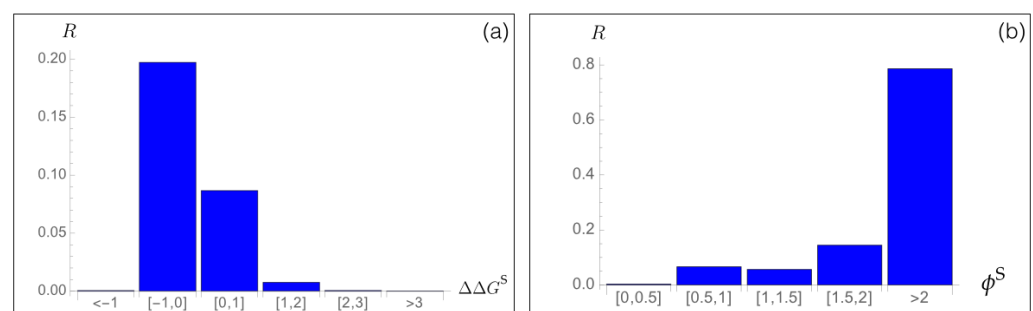
Figure 3. Mutation rate R_i of SARS-CoV-2 spike protein variants i observed in the GISAID database [46] as a function of (**a**) the predicted change in folding free energy $\Delta\Delta G_i^S$ (in kcal/mol) and (**b**) the predicted fitness contribution ϕ_i^S.

It is instructive to analyze the localization of the variants fixed through viral evolution in the 3D structure of the spike protein. The mean values of R_i in the core (solvent accessibility <20%), in partially buried regions (20–50%) and at the surface (>50%) are equal to 0.06, 0.06 and 0.23, respectively. This indicates that variants that became fixed are mainly situated in solvent-exposed regions, where they can play a key role in modulating binding with other biomolecules. Variants in buried or partially buried regions are less

often observed, as these areas are more constrained from a structural point of view and are usually not involved in function.

We also compared our stability predictions with results of molecular dynamics and nanomechanical simulations, which identified three protein segments as strongly contributing to the RBD stability, i.e., (A348–A352), (F400–R403) and (N450–R454) [47]. We found that the value of $\Delta\Delta G_i^S$ averaged over all possible point mutations inserted in these three segments is equal to (1.0, 2.1, 1.4) kcal/mol, respectively, and the corresponding fitness contribution ϕ_i^S to (0.7, 0.5, 0.6). The strongly destabilizing effects of mutations predicted in these three regions indicate that they are particularly optimized for stability, in agreement with [47]. This result further supports the ability of our approach to properly capture the stability properties of the spike protein.

3.3. Spike Protein/ACE2 Binding Affinity and SARS-CoV-2 Infectivity

We analyzed here the impact of variants on the binding of the spike protein with the ACE2 receptor. For all possible point substitutions i in the spike protein, we computed the change in binding affinity of the spike protein/ACE2 complex, $\Delta\Delta G_i^{ACE2}$, using the BeAtMuSiC program [35]. Based on the $\Delta\Delta G_i^{ACE2}$ values, we estimated the ϕ_i^{ACE2} viral fitness (Equation (4)), aimed at modeling infectivity. Indeed, a higher binding affinity between the spike protein and ACE2 results in a higher efficiency of virus entry into the host's cells [41], which in turn leads to an increase in SARS-CoV-2 infectivity.

We compared the predicted binding free energy values $\Delta\Delta G_i^{ACE2}$ with the experimentally characterized binding properties of thousands of variants introduced in the RBD of the spike protein using a yeast surface display platform, in which binding to ACE2 was quantitatively determined via flow cytometry [45]. Such deep mutagenesis scanning techniques are excellent tools to estimate biophysical quantities on a large scale. However, even though the average accuracy is reasonably good, the measured quantities are often noisy [48].

A good agreement was found between the computed $\Delta\Delta G_i^{ACE2}$ values and the large-scale measured binding affinity properties, with a Pearson's correlation coefficient of -0.46 (p-value $< 10^{-240}$). This result can be considered as very good, especially as not only the computed but also the experimental values have limited accuracy. It clearly underlines the quality of our prediction approach.

3.4. Spike Protein/nAb Binding Affinity and Immune Escape

Immune evasion is the well-known mechanism used by viruses to evade from the immune system of its host, thus making its replication and spreading more efficient [49]. This mechanism involves a series of strategies such as spontaneous mutations that result in the inactivation of nAbs [50] or in the inhibition of pattern-recognition receptors initiating signalling pathways [51].

To represent the diversity of the B-cell receptor repertoire and to mimic the effect of the human immune response, we considered the set $\mathcal{D}^{nAb}$ of more than 30 nAbs, of which the 3D structures with the RBD of the spike protein were experimentally resolved (see Section 2.1). We performed a large-scale in silico mutagenesis experiment by introducing all possible point mutations i in the RBD of the spike protein and by computing with BeAtMuSiC [35] the resulting change in binding free energy $\overline{\Delta\Delta G_i}^{nAb}$ averaged over all spike protein/nAb complexes that contain the mutation, as well as their associated fitness contribution ϕ_i^{nAb} (see Equations (1)–(4)). With this procedure, we identified key spike protein variants that are likely to either help or destroy the neutralization activity of the nAbs.

In a first stage, we performed validation tests on BeAtMuSiC's $\Delta\Delta G_i^{nAb}$ predictions. We compared them with deep mutagenesis scanning data measuring the impact of mutations in the RBD on their escape fractions from two nAbs, REGN10933 and REGN10987, which are often administrated as a cocktail to COVID-19 patients [52]. The escape fractions were estimated using a high-throughput yeast-surface-display platform, in which folded

RBDs were expressed on the yeast cell surface and the fraction of cells that express mutant RBDs and that are bound to nAbs was measured [20]. Per-mutant escape fraction values close to zero indicate that the variant protein is bound to nAbs while values close to one indicate that it is not.

The structures of the complexes formed by the spike protein and REGN10933 or REGN10987 nAbs were recently resolved (PDB code 6XDG). They target two different structural epitopes in the RBD of the spike protein. We did not include these structures in our set $\mathcal{D}^{nAb}$ as they were resolved via cryo-EM technique at only 3.9 Å of resolution. We predicted the changes in binding affinity $\Delta\Delta G_i$ of the two spike protein/nAb complexes caused by all RBD mutations i for which experimental escape fractions were available. Despite the low resolution of the 3D structures, we found good Pearson correlation coefficients of 0.48 (p-value $< 10^{-28}$) and 0.43 (p-value $< 10^{-12}$) between the per-mutant escape fractions and the computed changes in affinity $\Delta\Delta G_i$ for REGN10933 and REGN10987 nAbs, respectively.

In a second stage, we estimated the fitness contributions ϕ_i^{nAb} of all possible mutations i in the spike protein's RBD on the basis of the predicted changes in binding free energy for the set of 31 good-resolution nAbs/spike protein complexes collected in $\mathcal{D}^{nAb}$. We made here and in what follows the strong approximation that these 31 nAbs represent the diversity of the human nAb repertoire. To validate this model, we compared the estimated fitness contributions ϕ_i^{nAb} with a series of data obtained from in vivo experiments aimed to study the viral escape from nAbs.

We started by considering the set of 22 variants of the spike protein for which the neutralizing activity of six nAbs has been experimentally tested in terms of the relative degree of resistance (in %) of the growth of each mutant virus in the presence or in the absence of each of these nAbs [17]; we considered the average percentage over the six nAbs tested. Low percentages identify variants that escape much more from nAbs than the wild type virus and high percentages, variants that only weakly affect the wild-type spike protein/nAbs affinity. We predicted correctly 18 out of the 22 variants as having ϕ^{nAb} fitness values greater than one; the last four variants have $\phi^{nAb} \sim 0.9$. Detailed results are reported in Table 1 for the five variants shown to have the broadest in vitro neutralizing spectrum [17]. Our results reproduce quite well the in vitro trends: variants that are likely to escape from at least some nAbs tend to have fitness values larger than one. Note, moreover, that the antibodies tested in [17] are different from the nAbs of our $\mathcal{D}^{nAb}$ set. Because of that, we did not expect such a good match between the experiments and our predictions. This result suggests that the set $\mathcal{D}^{nAb}$ is truly representative of the antibody repertoire neutralizing the SARS-CoV-2 virus.

Table 1. List of the five variants of the spike protein RBD which have the broadest in vitro neutralizing spectrum, as measured in [17]. Their measured average resistance to six nAbs compared to the wild-type are given, as well as their fitness ϕ^{nAb} predicted on the basis of the 31 nAbs from the $\mathcal{D}^{nAb}$ set.

| Variants	Resistance to nAbs	ϕ^{nAb}
S349A	35%	1.1
G446A	37%	1.4
G447A	41%	1.6
N448A	26%	1.5
E484A	44%	1.1

The response to the viral infection drastically depends on the ensemble of nAbs present in the host, given that each nAb behaves differently with respect to wild-type and variant strains. In agreement with this, the predicted change in binding free energy $\Delta\Delta G_i^{nAb}$ is found to strongly depend on the considered variant and nAb/spike protein complex, as clearly seen in Figure 4. Remember that it is the average $\overline{\Delta\Delta G_i}^{nAb}$ over all the nAbs that is used to define the fitness contribution ϕ^{nAb} and thus the overall immune escape ability.
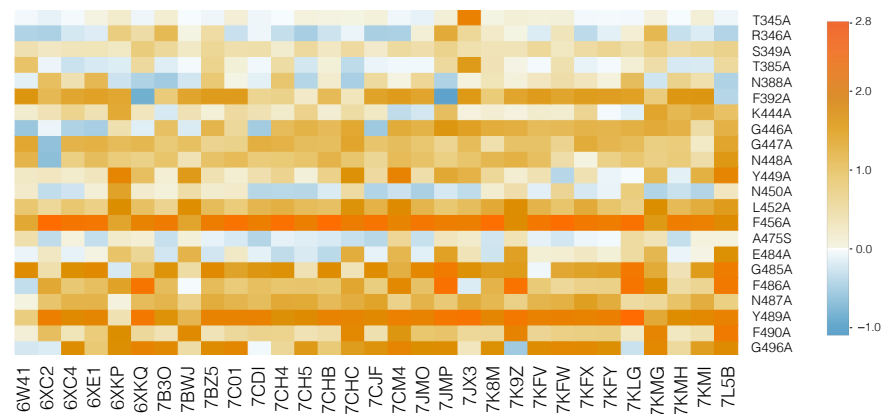
Figure 4. HeatMap of the predicted $\Delta\Delta G_i^{\mathrm{nAb}}$ values for each of the 31 nAb/spike protein complexes from the set $\mathcal{D}^{\mathrm{nAb}}$. The color scale is shown on the right (in kcal/mol). Light blue corresponds to variants that slightly stabilize the complex and orange, to mutations that destabilize the complex. The most destabilizing mutations are likely to lead the virus to escape from the immune system.

We also validated our fitness predictions ϕ^{nAb} against the large-scale experimental estimation of the immune escape fractions of about 2000 variants, averaged over a set of 17 nAbs [53]; note that these nAbs are not in the set $\mathcal{D}^{\mathrm{nAb}}$. We found a reasonably good overall Pearson correlation coefficient of 0.29 (*p*-value $< 10^{-25}$) between ϕ^{nAb} and measured escape fractions. Looking in more detail, the residues whose mutations most affect nAb binding belong to two regions of the RBD: the 443–450 and 484–490 loops that are situated at both sides of the ACE2 binding interface [53]. Using our set $\mathcal{D}^{\mathrm{nAb}}$ of nAbs, we predicted the second region as potentially leading to immune escape with a ϕ^{nAb} value of 1.6. The nAb escaping capability is predicted to be weaker for the first region, with $\phi^{\mathrm{nAb}} = 1.1$.

A 3D representation of the per-residue fitness contributions ϕ_i^{nAb} in the RBD of the spike protein, averaged over all possible mutations at each position, is shown in Figure 5. This figure is very useful to identify residues whose mutation is likely to lead to the escape of the virus from the $\mathcal{D}^{\mathrm{nAb}}$ set of nAbs.
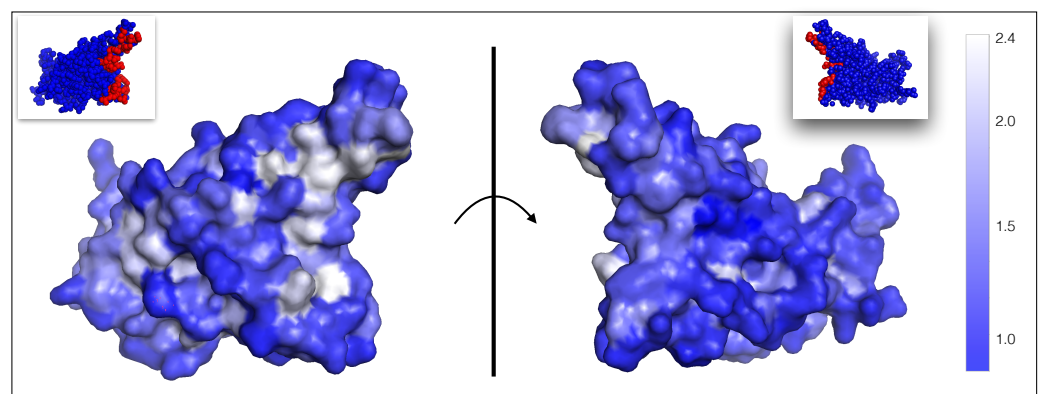
Figure 5. Average per-residue ϕ^{nAb} fitness contributions related to the ability of the virus to escape from the immune system, mapped onto the 3D structure of the spike protein RBD (PDB code 6M0J). As shown in the color bar on the right, residues whose mutation lead to highest average ϕ^{nAb} and thus to viral escape from nAbs are shown in white; residues with fitness values of one or lower and not allowing viral escape are in dark blue. The ϕ^{nAb} fitness values for all residues in the RBD are given in our GitHub repository (https://github.com/3BioCompBio/SpikeProSARS-CoV-2/blob/main/phi_nAb.dat, accessed on 10 April 2021). The ACE2 binding interface is shown in red in the two small pictures at the top (see also Figure 1). The left and right pictures are related by a 180° rotation with respect to the vertical plane (shown as a line) representing the ACE2 binding interface.

3.5. Immune Escape from Polyclonal Human Sera

We examined to what extent our method reproduces the impact of variants on the neutralizing activity of polyclonal human sera. Note that such activity depends on a wide range of factors among which inter-patient variability and time since infection [53]. Our computational approach is obviously unable to capture all intricate dependencies but instead, we expect it to detect general trends.

We used deep mutagenesis scanning data from [53], in which the escape fractions of about 2000 single-site RBD variants were assessed on the neutralizing activity of plasma samples taken from 17 SARS-CoV-2-infected individuals, at different time points after infection. We calculated the correlation between the escape fraction for each variant averaged over the patients and post-infection time points and the predicted fitness contributions ϕ^{nAb} computed from the $\mathcal{D}^{\text{nAb}}$ set of nAbs. We obtained a reasonably good Pearson correlation coefficient of 0.35 (p-value $< 10^{-42}$) between the predicted and measured quantities.

Only few residues appear to contribute substantially to the escape mechanisms, when averaged over the whole plasma sample collection. Indeed, only 23 residues have an average escape fraction greater than 3%. Our predictions for these residues are in very good agreement with experiments: we obtained an average per-residue ϕ^{nAb} equal to 1.5. Residue F456 shows almost perfect agreement: it has the highest measured escape fraction, and also has the highest predicted ϕ^{nAb} value, equal to 2.2. Almost all substitutions at that position are predicted to strongly impact on the binding in the majority of spike-protein/nAbs complexes analyzed.

Finally, it is interesting to compare the measured immune escaping fractions in polyclonal plasma discussed in this section with the experimentally characterized escape fractions in the set of nAbs studied in [53] and discussed in Section 3.4. We found that their linear correlation coefficient is equal to 0.4 (p-value $< 10^{-16}$), which indicates there are differences between the tested cocktail of nAbs and serum plasma. Possible explanations include the scarcity of the tested antibodies in the polyclonal plasma, or the subdominance of the epitopes they target [53].

3.6. Overall Variant Fitness, Transmissibility, Infectivity and Immune Escape

We focused on five SARS-CoV-2 variants most frequently observed worldwide, as reported in the GISAID database [46] in March 2021, and predicted their fitness; the results are shown in Table 2.

The most frequently observed spike protein variant involves the substitution of aspartic acid at position 614 into glycine, situated outside the RBD. This variant quickly became dominant after its appearance in early 2020 [40,54]. We correctly predicted a substantial increase in fitness for this variant with respect to wild type, which is driven by an increased stability of the spike protein ($\phi^{\text{S}}_{\text{D614G}} = 3.7$). We hypothesize that this stabilization leads to a higher person-to-person viral transmissibility, as also suggested in [40,54,55] and observed in vivo [55]. In the latter study, a stabilization of the spike protein was measured upon D614G substitution via a strengthening of the S1–S2 subunit interactions, where S1 is the receptor binding subunit containing the RBD and S2 is the membrane fusion subunit. In contrast, this variant was shown to alter neither the binding of the spike protein to ACE2 nor the antibody neutralization, as it is situated outside the RBD [55]. We also correctly reproduced this result, with fitness values of $\phi^{\text{ACE2}}_{\text{D614G}} = 1.0 = \phi^{\text{nAb}}_{\text{D614G}}$ (Table 2). The overall predicted fitness is thus $\Phi_{\text{D614G}} = 3.7$.

Two other variants, A222V and P681H, show similar albeit less pronounced trends. Our results predict an increase in transmissibility ($\phi^{\text{S}}_{A222V} \approx \phi^{\text{S}}_{P681H} \approx 2.0$), but to a lesser extent than D614G. Experimental data are in agreement with the weaker impacts of these variants on the spike protein fitness and the viral transmissibility compared to D614G [56,57]. The A222V variant has been related to the large outbreaks in Europe in early summer 2020, while P681H is associated to the so-called UK lineage (B.1.1.7) that appeared in UK in late 2020 and is now becoming dominant in Europe in the current outbreaks.

Finally, N501Y is also a widely spread variant appearing in all major lineages, i.e., UK (B.1.1.7), Brazilian (P.1) and South African (B.1.351) lineages. We predicted this variant as having a high overall fitness Φ due to a combination of increased fitness contributions $\phi^{\mathrm{S}}_{N501Y}$ and $\phi^{\mathrm{ACE2}}_{N501Y}$, but a $\phi^{\mathrm{nAb}}_{N501Y}=1$. In other words, we predicted this variant to be more transmissible and infectious than the wild type but to have no impact on the response of the human immune system. More precisely, we predicted N501Y as improving the stability of the spike protein RBD and its binding affinity for ACE2; the latter property is also suggested by another computational study [58]. No clinical data suggest that N501Y is able to evade the post-vaccination immune response [59]; this tends to support our prediction results.

Table 2. The five most widely observed variants and their predicted fitness. Occurrences refer to their number of occurrences in the GISAID database [46], ϕ^{S}_i, ϕ^{ACE2}_i, and ϕ^{nAb}_i to the fitness contributions of the variants i related to the stability of the spike protein, its binding affinity for ACE2 and its escape propensity from the host's immune system, respectively, and Φ_i to the total fitness.

| Variants	Occurrences	ϕ^{S}_i	ϕ^{ACE2}_i	ϕ^{nAb}_i	Φ_i
D614G	96%	3.7	1.0	1.0	3.7
A222V	19%	2.0	1.0	1.0	2.0
P681H	19%	1.6	1.0	1.0	1.6
N501Y	18%	2.1	1.4	1.0	2.9

3.7. Viral Evolution and Overall Fitness

We applied our prediction pipeline to analyze SARS-CoV-2 evolution, focusing on the spike protein. We started by predicting the viral fitness Φ of all the SARS-CoV-2 strains collected in the GISAID database [46] from December 2019 until March 2021, which amounts to about 7.8×10^5 strains. We subdivided the strains according to the month of collection and computed the per-month average of viral fitness. The results are reported in Figure 6a as a function of time. Clearly, we predict an increase in the viral fitness since the beginning of the infection in December 2019, in agreement with epidemiological results. This result demonstrates once again the quality of our computational pipeline.

Note that to predict the future evolution of the fitness Φ, it is necessary to take into account different parameters such as the varying repertoire of human nAbs and the effect of vaccination. While the fitness contributions ϕ^{S} and ϕ^{ACE2} are expected to reach a plateau when the spike protein sequence becomes optimal for stability and for binding to ACE2, the cat-and-mouse game played by the virus and its host leads the host to continuously adapt its B-cell repertoire to the new variants of the virus, so that ϕ^{nAb} certainly increases with respect to the old nAbs, but not with respect to the new nAbs. In total, the overall fitness Φ is expected to plateau after some time, or at least to increase less.

We analyzed in more detail the evolution of the partial distribution function of the per-month averaged fitness in Figure 6b. In January 2020, the population was dominated by the wild type strain whose fitness Φ is by definition equal to one. The effect of the D614G spike protein variant with a predicted $\Phi \sim 4$ started to be observed from May 2020, while in October of the same year, additional mutations with $\Phi = 2.0$, such as A222V, started to be fixed in the population, leading to a further increase in Φ. In March 2021, the distribution became dominated by new variants, i.e., UK, South-African and Brazilian variants, with a much higher fitness than both the wild-type and D614G strains.
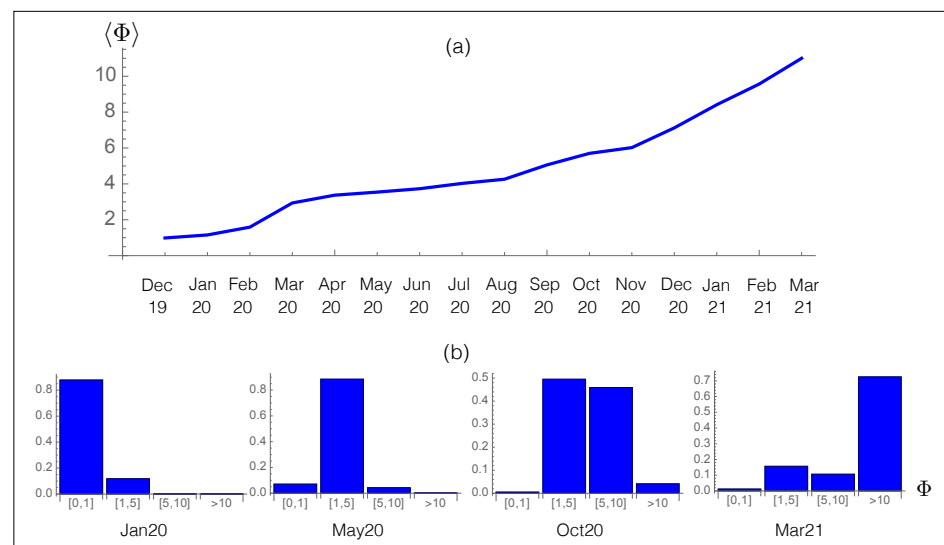
Figure 6. Time evolution of the predicted overall fitness Φ. (**a**) Average fitness $\langle\Phi\rangle$ per month for the SARS-CoV-2 strains collected from the GISAID database [46] as a function of time; (**b**) Probability distribution of Φ for the SARS-CoV-2 strains collected from GISAID during a given month.

Finally, we carefully checked that our large-scale mutagenesis predictions are not biased towards high fitness values. Indeed, such bias could potentially cause a trivial increase in fitness upon evolution and lead to erroneous interpretations. To check this, we created 2×10^6 viral strains by inserting either three or five random mutations in the wild-type spike protein and assumed that they became fixed with probability one independently of their fitness value; the number of random mutations was chosen based on the average number of single variants per strain in the GISAID database [46] which is between three and four. We then computed the fitness Φ for all these random variant strains. On the other hand, we plotted the Φ distribution of the real variant strains observed in the GISAID database. The fitness distribution of the two simulated strains and of the real viral strain are completely different, as shown in Figure 7. Indeed, when three or five random mutations are inserted in the spike protein, the Φ distributions have a median value of 0.32 and 0.12, respectively; moreover, 79% and 86% of the mutated strains have a lower overall fitness Φ than the wild-type virus. In contrast, the distribution of real strains has a median of 4.8 and basically all the strains (99.5%) have a predicted fitness higher than the wild type. In summary, random strains show a fitness distribution that peaks at a value of zero, in contrast to the real viral strains whose fitness distribution is much more extended, reaching values of more than 10. This confirms that the high fitness values that we predict are not due to unwanted biases of any kind, and that one cannot obtain fitness values as high as those of variants observed in real strains just by considering random mutations. This last analysis further supports the unbiased nature and validity of our computational approach.
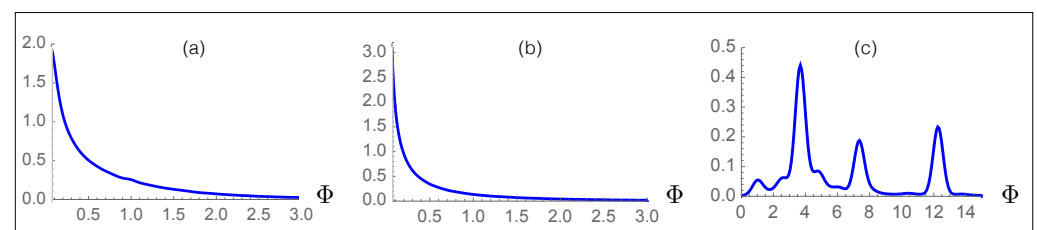
Figure 7. Probability distribution of the predicted fitness Φ for 10^6 simulated viral strains obtained by inserting (**a**) three and (**b**) five random mutations in the spike protein; (**c**) probability distribution function of the predicted fitness Φ for all the strains collected in the GISAID database.

4. Conclusions

Here we set up and validated SpikePro, a simple computational model that predicts the impact of spike protein variants on the SARS-CoV-2 fitness and more specifically, on inter-host viral transmissibility, infectivity of the host and ability of escaping from the host's immune system. Moreover, the program is easy to use and can be freely downloaded from github.com/3BioCompBio/SpikeProSARS-CoV-2. SpikePro allows identification, with good accuracy and in a few seconds, of new SARS-CoV-2 variants with high fitness which need to be closely monitored by health agencies.

Although innovative diagnostic strategies [60] and genomic surveillance schemes [61,62] have recently been introduced to mitigate SARS-CoV-2 spreading, which for example integrate rapid large-scale viral genome sequencing with epidemiological analyses [61], such approaches remain time-consuming and require substantial amounts of human and financial resources to be implemented. SpikePro has thus a central role to play in the genomic surveillance programs of the new SARS-CoV-2 strains, especially in the near future with the growing number of people vaccinated and thus the larger selective pressure on the virus [63].

We thoroughly analyzed and validated SpikePro on a wide series of experimental, epidemiological and clinical data available. Despite the simplicity of the model, the approximations made, and the absence of parameters that were fitted to optimize the accuracy of the predictions, the SpikePro pipeline reproduces well the collected data. Whether the validation is performed on large-scale mutagenesis data, nAb cocktails or polyclonal human sera, whether the comparison involves the fitness of the spike protein, of the spike protein/ACE2 complex, or of a series of spike protein/nAb complexes, the results are good with correlation coefficients in the 0.3 to 0.5 range.

In addition, SpikePro predicts a high overall fitness value for the frequently occurring variants such as the UK, Brazilian or South-African variants and correctly identifies the main fitness contributions. It also reproduces well the overall fitness evolution of the SARS-CoV-2 virus over the past pandemic year.

It has to be emphasized that the SpikePro model, besides being able to reproduce known results, has a true prediction potential in describing and interpreting the effect of new spike protein variants that could be fixed in the near future and the future SARS-CoV-2 evolution, owing to the physical description of the fitness in terms of free energy contributions, which are estimated using the well-known structure-based PoPMuSiC and BeAtMuSiC predictors [34,35].

Despite the progress we made towards a better understanding of the molecular mechanisms underlying the SARS-CoV-2 fitness, we made some approximations in the construction of our model which we will try to relax in future studies. In particular, we did not take into account possible amino acid deletions or insertions in the spike protein, although they have been proven to influence viral fitness. For example, the deletions Δ69–70 and Δ144–145 in the N-terminal domain of the spike protein, found in different lineages, have been associated to an increase in human-to-human viral transmission and to altered antigenicity [21,64]. It would also be interesting to take into account epistatic effects. Indeed, while more and more variants become fixed, interactions between them are expected to become non-negligible. The model should also be extended to other proteins of the SARS-CoV-2 virus such as the non-structural protein 1 (Nsp1) which also contributes to immune evasion [65], rather than considering the spike protein only. Finally, when more nAbs/spike protein complexes will be resolved at high resolution, they will enrich our set $\mathcal{D}^{\text{nAb}}$ and better describe the B-cell receptor repertoire. Considering a weighted combination of the effects of RBD variants on all nAbs depending on different factors such as time and vaccination status would further improve our method in mimicking the immune response and its temporal evolution.

References

1. Dong, E.; Du, H.; Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **2020**, *20*, 533–534. [CrossRef]
2. Sanders, J.M.; Monogue, M.L.; Jodlowski, T.Z.; Cutrell, J.B. Pharmacologic treatments for coronavirus disease 2019 (COVID-19): A review. *JAMA* **2020**, *323*, 1824–1836. [CrossRef]
3. Le, T.T.; Andreadakis, Z.; Kumar, A.; Román, R.G.; Tollefsen, S.; Saville, M.; Mayhew, S. The COVID-19 vaccine development landscape. *Nat. Rev. Drug Discov.* **2020**, *19*, 305–306. [CrossRef]
4. Baden, L.R.; El Sahly, H.M.; Essink, B.; Kotloff, K.; Frey, S.; Novak, R.; Diemert, D.; Spector, S.A.; Rouphael, N.; Creech, C.B.; et al. Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *N. Engl. J. Med.* **2020**, *384*, 403–416. [CrossRef]
5. Polack, F.P.; Thomas, S.J.; Kitchin, N.; Absalon, J.; Gurtman, A.; Lockhart, S.; Perez, J.L.; Pérez Marc, G.; Moreira, E.D.; Zerbini, C.; et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *N. Engl. J. Med.* **2020**, *383*, 2603–2615. [CrossRef] [PubMed]
6. Voysey, M.; Clemens, S.A.C.; Madhi, S.A.; Weckx, L.Y.; Folegatti, P.M.; Aley, P.K.; Angus, B.; Baillie, V.L.; Barnabas, S.L.; Bhorat, Q.E.; et al. Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: An interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. *Lancet* **2021**, *397*, 99–111. [CrossRef]
7. Logunov, D.Y.; Dolzhikova, I.V.; Shcheblyakov, D.V.; Tukhvatulin, A.I.; Zubkova, O.V.; Dzharullaeva, A.S.; Kovyrshina, A.V.; Lubenets, N.L.; Grousova, D.M.; Erokhova, A.S.; et al. Safety and efficacy of an rAd26 and rAd5 vector-based heterologous prime-boost COVID-19 vaccine: An interim analysis of a randomised controlled phase 3 trial in Russia. *Lancet* **2021**, *397*, 671–681. [CrossRef]
8. Sadoff, J.; Le Gars, M.; Shukarev, G.; Heerwegh, D.; Truyers, C.; de Groot, A.M.; Stoop, J.; Tete, S.; Van Damme, W.; Leroux-Roels, I.; et al. Interim Results of a Phase 1–2a Trial of Ad26. COV2. S Covid-19 Vaccine. *N. Engl. J. Med.* **2021**, *384*, 1824–1835. [CrossRef] [PubMed]
9. Keech, C.; Albert, G.; Cho, I.; Robertson, A.; Reed, P.; Neal, S.; Plested, J.S.; Zhu, M.; Cloney-Clark, S.; Zhou, H.; et al. Phase 1–2 trial of a SARS-CoV-2 recombinant spike protein nanoparticle vaccine. *N. Engl. J. Med.* **2020**, *383*, 2320–2332. [CrossRef]
10. Tsatsakis, A.; Calina, D.; Falzone, L.; Petrakis, D.; Mitrut, R.; Siokas, V.; Pennisi, M.; Lanza, G.; Libra, M.; Doukas, S.G.; et al. SARS-CoV-2 pathophysiology and its clinical implications: An integrative overview of the pharmacotherapeutic management of COVID-19. *Food Chem. Toxicol.* **2020**, *146*, 111769. [CrossRef]
11. Bhimraj, A.; Morgan, R.L.; Shumaker, A.H.; Lavergne, V.; Baden, L.; Cheng, V.C.C.; Edwards, K.M.; Gandhi, R.; Muller, W.J.; O'Horo, J.C.; et al. Infectious Diseases Society of America guidelines on the treatment and management of patients with COVID-19. *Clin. Infect. Dis.* **2020**. [CrossRef]
12. Weinreich, D.M.; Sivapalasingam, S.; Norton, T.; Ali, S.; Gao, H.; Bhore, R.; Musser, B.J.; Soo, Y.; Rofail, D.; Im, J.; et al. REGN-COV2, a Neutralizing Antibody Cocktail, in Outpatients with Covid-19. *N. Engl. J. Med.* **2020**, *384*, 238–251. [CrossRef] [PubMed]
13. Jiang, S.; Zhang, X.; Yang, Y.; Hotez, P.J.; Du, L. Neutralizing antibodies for the treatment of COVID-19. *Nat. Biomed. Eng.* **2020**, *4*, 1134–1139. [CrossRef] [PubMed]
14. Joyner, M.J.; Carter, R.E.; Senefeld, J.W.; Klassen, S.A.; Mills, J.R.; Johnson, P.W.; Theel, E.S.; Wiggins, C.C.; Bruno, K.A.; Klompas, A.M.; et al. Convalescent plasma antibody levels and the risk of death from covid-19. *N. Engl. J. Med.* **2021**, *384*, 1015–1027. [CrossRef]
15. Alcami, A.; Koszinowski, U.H. Viral mechanisms of immune evasion. *Trends Microbiol.* **2000**, *8*, 410–418. [CrossRef]
16. Williams, T.C.; Burgers, W.A. SARS-CoV-2 evolution and vaccines: Cause for concern? *Lancet Respir. Med.* **2021**, *9*, 333–335. [CrossRef]
17. Ku, Z.; Xie, X.; Davidson, E.; Ye, X.; Su, H.; Menachery, V.D.; Li, Y.; Yuan, Z.; Zhang, X.; Muruato, A.E.; et al. Molecular determinants and mechanism for antibody cocktail preventing SARS-CoV-2 escape. *Nat. Commun.* **2021**, *12*, 469. [CrossRef] [PubMed]

18. Weisblum, Y.; Schmidt, F.; Zhang, F.; DaSilva, J.; Poston, D.; Lorenzi, J.C.; Muecksch, F.; Rutkowska, M.; Hoffmann, H.H.; Michailidis, E.; et al. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *Elife* **2020**, *9*, e61312. [CrossRef]

19. Greaney, A.J.; Starr, T.N.; Gilchuk, P.; Zost, S.J.; Binshtein, E.; Loes, A.N.; Hilton, S.K.; Huddleston, J.; Eguia, R.; Crawford, K.H.; et al. Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. *Cell Host Microbe* **2021**, *29*, 44–57. [CrossRef]

20. Starr, T.N.; Greaney, A.J.; Addetia, A.; Hannon, W.W.; Choudhary, M.C.; Dingens, A.S.; Li, J.Z.; Bloom, J.D. Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. *Science* **2021**, *371*, 850–854. [CrossRef]

21. McCarthy, K.R.; Rennick, L.J.; Nambulli, S.; Robinson-McCarthy, L.R.; Bain, W.G.; Haidar, G.; Duprex, W.P. Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. *Science* **2021**, *371*, 1139–1142. [CrossRef]

22. Tegally, H.; Wilkinson, E.; Giovanetti, M.; Iranzadeh, A.; Fonseca, V.; Giandhari, J.; Doolabh, D.; Pillay, S.; San, E.J.; Msomi, N.; et al. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv* **2020**. [CrossRef]

23. Andreano, E.; Piccini, G.; Licastro, D.; Casalino, L.; Johnson, N.V.; Paciello, I.; Dal Monego, S.; Pantano, E.; Manganaro, N.; Manenti, A.; et al. SARS-CoV-2 escape in vitro from a highly neutralizing COVID-19 convalescent plasma. *bioRxiv* **2020**. [CrossRef]

24. Wibmer, C.K.; Ayres, F.; Hermanus, T.; Madzivhandila, M.; Kgagudi, P.; Oosthuysen, B.; Lambson, B.E.; De Oliveira, T.; Vermeulen, M.; Van der Berg, K.; et al. SARS-CoV-2 501Y. V2 escapes neutralization by South African COVID-19 donor plasma. *Nat. Med.* **2021**, *27*, 622–625. [CrossRef]

25. Wang, Z.; Schmidt, F.; Weisblum, Y.; Muecksch, F.; Barnes, C.O.; Finkin, S.; Schaefer-Babajew, D.; Cipolla, M.; Gaebler, C.; Lieberman, J.A.; et al. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature* **2021**, *592*, 616–622. [CrossRef] [PubMed]

26. Collier, D.A.; De Marco, A.; Ferreira, I.A.; Meng, B.; Datir, R.; Walls, A.C.; Kemp, S.A.; Bassi, J.; Pinto, D.; Fregni, C.S.; et al. SARS-CoV-2 B.1.1.7 escape from mRNA vaccine-elicited neutralizing antibodies. *medRxiv* **2021**. [CrossRef]

27. Thomson, E.C.; Rosen, L.E.; Shepherd, J.G.; Spreafico, R.; da Silva Filipe, A.; Wojcechowskyj, J.A.; Davis, C.; Piccoli, L.; Pascall, D.J.; Dillen, J.; et al. Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. *Cell* **2021**, *184*, 1171–1187. [CrossRef]

28. Lauring, A.S.; Hodcroft, E.B. Genetic Variants of SARS-CoV-2—What Do They Mean? *JAMA* **2021**, *325*, 529–531. [CrossRef]

29. Berman, H.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.; Weissig, H.; Shindyalov, I.; Bourne, P. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [CrossRef]

30. Walls, A.C.; Park, Y.J.; Tortorici, M.A.; Wall, A.; McGuire, A.T.; Veesler, D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **2020**, *181*, 281–292. [CrossRef]

31. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; de Beer, T.A.P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296–W303. [CrossRef]

32. Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **2020**, *581*, 215–220. [CrossRef]

33. Raybould, M.I.; Kovaltsuk, A.; Marks, C.; Deane, C.M. CoV-AbDab: The coronavirus antibody database. *Bioinformatics* **2021**, *37*, 734–735. [CrossRef]

34. Dehouck, Y.; Grosfils, A.; Folch, B.; Gilis, D.; Bogaerts, P.; Rooman, M. Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: PoPMuSiC-2.0. *Bioinformatics* **2009**, *25*, 2537–2543. [CrossRef]

35. Dehouck, Y.; Kwasigroch, J.M.; Rooman, M.; Gilis, D. BeAtMuSiC: Prediction of changes in protein–protein binding affinity on mutations. *Nucleic Acids Res.* **2013**, *41*, W333–W339. [CrossRef]

36. Qiao, B.; Olvera de la Cruz, M. Enhanced binding of SARS-CoV-2 spike protein to receptor by distal polybasic cleavage sites. *ACS Nano* **2020**, *14*, 10616–10623. [CrossRef] [PubMed]

37. Wargo, A.R.; Kurath, G. Viral fitness: Definitions, measurement, and current insights. *Curr. Opin. Virol.* **2012**, *2*, 538–545. [CrossRef] [PubMed]

38. Orr, H.A. Fitness and its role in evolutionary genetics. *Nat. Rev. Genet.* **2009**, *10*, 531–539. [CrossRef]

39. Domingo, E. Mechanisms of viral emergence. *Vet. Res.* **2010**, *41*, 38. [CrossRef]

40. Baric, R.S. Emergence of a Highly Fit SARS-CoV-2 Variant. *N. Engl. J. Med.* **2020**, *383*, 2684–2686. [CrossRef] [PubMed]

41. Shang, J.; Ye, G.; Shi, K.; Wan, Y.; Luo, C.; Aihara, H.; Geng, Q.; Auerbach, A.; Li, F. Structural basis of receptor recognition by SARS-CoV-2. *Nature* **2020**, *581*, 221–224. [CrossRef]

42. Echave, J.; Jackson, E.L.; Wilke, C.O. Relationship between protein thermodynamic constraints and variation of evolutionary rates among sites. *Phys. Biol.* **2015**, *12*, 025002. [CrossRef]

43. Pucci, F.; Bernaerts, K.; Teheux, F.; Gilis, D.; Rooman, M. Symmetry principles in optimization problems : An application to protein stability prediction. *IFAC-PapersOnLine* **2015**, *48*, 458–463. [CrossRef]

44. Pucci, F.; Bernaerts, K.V.; Kwasigroch, J.M.; Rooman, M. Quantification of biases in predictions of protein stability changes upon mutations. *Bioinformatics* **2018**, *34*, 3659–3665. [CrossRef]

45. Starr, T.N.; Greaney, A.J.; Hilton, S.K.; Ellis, D.; Crawford, K.H.; Dingens, A.S.; Navarro, M.J.; Bowen, J.E.; Tortorici, M.A.; Walls, A.C.; et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* **2020**, *182*, 1295–1310. [CrossRef]

46. Elbe, S.; Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob. Chall.* **2017**, *1*, 33–46. [CrossRef] [PubMed]

47. Moreira, R.A.; Chwastyk, M.; Baker, J.L.; Guzman, H.V.; Poma, A.B. Quantitative determination of mechanical stability in the novel coronavirus spike protein. *Nanoscale* **2020**, *12*, 16409–16413. [CrossRef]

48. Fowler, D.M.; Stephany, J.J.; Fields, S. Measuring the activity of protein variants on a large scale using deep mutational scanning. *Nat. Protoc.* **2014**, *9*, 2267–2284. [CrossRef]

49. Beachboard, D.C.; Horner, S.M. Innate immune evasion strategies of DNA and RNA viruses. *Curr. Opin. Microbiol.* **2016**, *32*, 113–119. [CrossRef]

50. Doud, M.B.; Hensley, S.E.; Bloom, J.D. Complete mapping of viral escape from neutralizing antibodies. *PLoS Pathog.* **2017**, *13*, e1006271. [CrossRef]

51. Bowie, A.G.; Unterholzner, L. Viral evasion and subversion of pattern-recognition receptor signalling. *Nat. Rev. Immunol.* **2008**, *8*, 911–922. [CrossRef]

52. Hansen, J.; Baum, A.; Pascal, K.E.; Russo, V.; Giordano, S.; Wloga, E.; Fulton, B.O.; Yan, Y.; Koon, K.; Patel, K.; et al. Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail. *Science* **2020**, *369*, 1010–1014. [CrossRef]

53. Greaney, A.J.; Loes, A.N.; Crawford, K.H.; Starr, T.N.; Malone, K.D.; Chu, H.Y.; Bloom, J.D. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe* **2021**, *29*, 463–476. [CrossRef]

54. Hou, Y.J.; Chiba, S.; Halfmann, P.; Ehre, C.; Kuroda, M.; Dinnon, K.H.; Leist, S.R.; Schäfer, A.; Nakajima, N.; Takahashi, K.; et al. SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. *Science* **2020**, *370*, 1464–1468. [CrossRef] [PubMed]

55. Zhang, L.; Jackson, C.B.; Mou, H.; Ojha, A.; Peng, H.; Quinlan, B.D.; Rangarajan, E.S.; Pan, A.; Vanderheiden, A.; Suthar, M.S.; et al. SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. *Nat. Commun.* **2020**, *11*, 6013. [CrossRef] [PubMed]

56. Hodcroft, E.B.; Zuber, M.; Nadeau, S.; Crawford, K.H.D.; Bloom, J.D.; Veesler, D.; Vaughan, T.G.; Comas, I.; Candelas, F.G.; Stadler, T.; et al. Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *medRxiv* **2020**. [CrossRef]

57. Peacock, T.P.; Goldhill, D.H.; Zhou, J.; Baillon, L.; Frise, R.; Swann, O.C.; Kugathasan, R.; Penn, R.; Brown, J.C.; Sanchez-David, R.Y.; et al. The furin cleavage site of SARS-CoV-2 spike protein is a key determinant for transmission due to enhanced replication in airway cells. *bioRxiv* **2020**. [CrossRef]

58. Villoutreix, B.O.; Calvez, V.; Marcelin, A.G.; Khatib, A.M. In Silico Investigation of the New UK (B.1.1.7) and South African (501Y.V2) SARS-CoV-2 Variants with a Focus at the ACE2–Spike RBD Interface. *Int. J. Mol. Sci.* **2021**, *22*, 1695. [CrossRef]

59. Xie, X.; Liu, Y.; Liu, J.; Zhang, X.; Zou, J.; Fontes-Garfias, C.R.; Xia, H.; Swanson, K.A.; Cutler, M.; Cooper, D.; et al. Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K and N501Y variants by BNT162b2 vaccine-elicited sera. *Nat. Med.* **2021**, *27*, 620–621. [CrossRef]

60. Falzone, L.; Gattuso, G.; Tsatsakis, A.; Spandidos, D.A.; Libra, M. Current and innovative methods for the diagnosis of COVID-19 infection. *Int. J. Mol. Med.* **2021**, *47*, 1–23. [CrossRef]

61. Meredith, L.W.; Hamilton, W.L.; Warne, B.; Houldcroft, C.J.; Hosmillo, M.; Jahun, A.S.; Curran, M.D.; Parmar, S.; Caller, L.G.; Caddy, S.L.; et al. Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: A prospective genomic surveillance study. *Lancet Infect. Dis.* **2020**, *20*, 1263–1272. [CrossRef]

62. Deng, X.; Gu, W.; Federman, S.; Du Plessis, L.; Pybus, O.G.; Faria, N.R.; Wang, C.; Yu, G.; Bushnell, B.; Pan, C.Y.; et al. Genomic surveillance reveals multiple introductions of SARS-CoV-2 into Northern California. *Science* **2020**, *369*, 582–587. [CrossRef] [PubMed]

63. Burioni, R.; Topol, E.J. Assessing the human immune response to SARS-CoV-2 variants. *Nat. Med.* **2021**, *27*, 571–572. [CrossRef]

64. Ribes, M.; Chaccour, C.; Moncunill, G. Adapt or perish: SARS-CoV-2 antibody escape variants defined by deletions in the Spike N-terminal Domain. *Signal Transduct. Target. Ther.* **2021**, *6*, 164. [CrossRef]

65. Thoms, M.; Buschauer, R.; Ameismeier, M.; Koepke, L.; Denk, T.; Hirschenberger, M.; Kratzat, H.; Hayn, M.; Mackens-Kiani, T.; Cheng, J.; et al. Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. *Science* **2020**, *369*, 1249–1255. [CrossRef]