

Unraveling the stability landscape of mutations in the SARS-CoV-2 receptor-binding domain

Mohamed Raef Smaoui (✉ mohamed.smaoui@ku.edu.kw)

Kuwait University <https://orcid.org/0000-0002-8891-7926>

Hamdi Yahyaoui

Kuwait University

Research Article

Keywords: COVID-19, Spike, Mutation, Energy Landscape, SARS-CoV-2, receptor-binding domain

Posted Date: August 17th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-59058/v1>

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

[Read Full License](#)

Version of Record: A version of this preprint was published at Scientific Reports on April 28th, 2021. See the published version at <https://doi.org/10.1038/s41598-021-88696-5>.

Unraveling the stability landscape of mutations in the SARS-CoV-2 receptor-binding domain

M. R. Smaoui* and H. Yahyaoui

Computer Science Department, Kuwait University, State of Kuwait

**Correspondence to: mohamed.smaoui@ku.edu.kw*

August 7, 2020

Abstract

The interaction between the receptor-binding domain of the SARS-CoV-2 spike glycoprotein and the ACE2 enzyme is believed to be the entry point of the virus into various cells in the body, including the lungs, heart, liver, and kidneys. The current focus of several therapeutic design efforts explore attempts at affecting the binding interaction between the two proteins to limit the activity of the virus and disease progression. In this work, we analyze the stability of the spike protein under all possible single-point mutations in the receptor-binding domain and computationally explore mutations that can affect the binding with the ACE2 enzyme. We unravel the mutation landscape of the receptor region and assess the toxicity potential of single and multi-point mutations, generating insights for future vaccine efforts on potential mutations that might further stabilize the spike protein and increase its infectivity. We developed a tool, called SpikeMutator, to construct full atomic protein structures of the mutant spike proteins and shared a database of 3,800 single-point mutant structures. We analyzed the recent 65,000 reported spike sequences across the globe and observed the emergence of stable multi-point mutant structures. Using the landscape, we searched through 7.5 million possible 2-point mutation combinations and report that the (R355D K424E) mutation produces one of the strongest spike proteins that therapeutic efforts should investigate for the sake of developing an effective vaccine.

keywords: COVID-19; Spike; Mutation; Energy Landscape; SARS-CoV-2; receptor-binding domain

1 Introduction

The outbreak of a respiratory illness in Wuhan, China on December 19th, 2019 has created a global public emergency and the spread of a new coronavirus disease (COVID-

19) [1,2]. The virus causing COVID-19 was termed SARS-CoV-2 [3] and shares sequence and structural similarity to the Middle East respiratory syndrome coronavirus (MERS-CoV) [4] and the severe acute respiratory syndrome coronavirus (SARS-CoV) [5–7]. As of August 6th, 2020, the COVID-19 pandemic has infected 18,956,836 people and killed over 710,053 worldwide (<https://www.worldometers.info/coronavirus/>). The pandemic has caused economic shutdown in many countries, fears of a global recession, restricted travel, closure of educational institutions, and mental distress on a global level [8–10]. With the drastic widespread and impact of SARS-CoV-2, many countries are still fighting against the first wave of the virus and fear the devastation of future waves [11].

SARS-CoV-2 contains a lipid envelope bilayer, with attached spike and membrane proteins, surrounding a stranded RNA genome of the virus [12]. Similar to MERS-CoV and SARS-CoV, the spike proteins in SARS-CoV-2 mitigate the attachment and binding of the virus to cell receptors and facilitate the release and entry of the viral genome into host cells [13–15]. The binding to host cells occurs at the receptor-binding domain (RBD) [16] in the S1 subunit of the spike protein. The SARS-CoV-2 RBD recognizes the angiotensin-converting enzyme 2 (ACE2) located in the lungs, heart, kidney and intestines as its host receptor [2,17]. Impeding the function of spike proteins has been the target of several antibodies, vaccines, and inhibitors [18–21].

Recent findings have reported the occurrence of sequence mutations in the SARS-CoV-2 spike protein, including some in the RBD [22]. The impact of the mutations on disease progression and biochemical phenotypes in COVID-19 still remains unknown. However, mutations in the spike protein and RBD of SARS-CoV and MERS-CoV are believed to have played a significant evolutionary factor in the transmission of the virus from bats to humans and influenced binding potentials to host receptors [23–25]. Mutations increasing affinity to virus receptors are ubiquitous [26,27], and exploring how mutations in the RBD region of SARS-CoV-2 impact the spike protein would aid in the design of better inhibitors and potential vaccines [28].

Vaccines targeting the RBD region will have to account for possible natural mutations that could influence the spike protein’s stability and tweak its dynamics with the ACE2 receptor. Mutations altering the RBD conformation have been recently shown to allow SARS-CoV-2 to elude antibody treatments and resist therapy [29]. It is hence imperative to study the possible mutations that could occur in the RBD and the impact they might have on COVID-19 progression. In this study, we computationally explore the mutation landscape of the RBD region and pinpoint mutations expressing strong binding potentials. We developed a tool, called SpikeMutator, to generate all possible single-point spike mutant trimer structures and map their free energies [30,31] to assess the affect of mutations on structure stability. We analyzed the current isolated spike sequences in the GISAID database [22] against the energy landscape and found evidence of accumulated mutations increasing the spike’s structural stability. To the best of our knowledge, this is the first work that aims to study the mutation landscape of the SARS-CoV-2 RBD region.

2 Results & Discussion

In this work, we used the cryo-EM SARS-CoV-2 spike glycoprotein structure (PDB ID 6VXX) reported by Walls et al. [32] as a reference model on which we assessed mutations. Table 1 reports the amino acid sequence of the receptor-binding domain contained in this structure and highlights areas with missing residue information. Table S1 of the supplementary material reports the entire spike sequence. In this structure, the receptor-binding domain (RBD) is located between amino acids 331 and 524, inclusive. Figure 1 presents a schematic image of the spike protein as a single and trimer structure. Three single structures aggregate to form a trimer conformation that binds to the ACE2 enzyme at the RBD interface.

We used *SpikeMutator* to mutate each amino acid in the RBD region to the 19 other canonical amino acids and generated a database of the 3D conformations of all possible spike trimer mutants. Every trimer structure contained 1 mutation that was simultaneously applied to each of the three aggregated spike proteins. We analyzed the free energies using Equation 1 for each mutant trimer and report the results in Figs 3-4. Figure 3 plots a 3D mutation energy landscape of the receptor-binding domain. One axis in the landscape represents the amino acids positioned at residue 331 through 524, the second axis plots the 20 canonical amino acids, and the third axis captures the free energy the mutant structure defined by the two other axes. Lower energy values correlate with favorable mutations that stabilize the RBD and can potentially improve binding with the ACE2 enzyme. Higher energy values suggest mutations that increase instability in the domain and potentially lower rates of binding with the ACE2 enzyme.

Figure 2 provides a 2D projection of the landscape and makes it easier to visualize the energies across different areas of the RBD. We observe that a mutation including a negatively charged, polar and hydrophilic amino acid such as aspartic acid (D) or glutamate (E) would increase the stability of the receptor-binding domain. A positively charged, polar, and hydrophilic amino acid such as arginine (R) introduces some of the most unfavorable mutations in the region 419-434, located relatively inside of the spike

Table 1: SARS-COV-2 Receptor-Binding Domain Sequence

```

331-----340-----360-----380-----401-----
NITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSP TKLNDLCFTNVYADSFVIRGDEVRQI
-----420-----440-----460-----481-----
APGQTGK IADYNYKLPDDFTGCVIAWNSNNLDSK--GNYNYLYR-----KPFERDI-----YF
-----500-----520--
PLQSYGFQPTN-VGYQPYRVVLSFELLHAPATV

```

The sequence of the receptor-binding domain as reported in the 6VXX PDB structure ranges from residue position 331-524, inclusive. Missing amino acids in the 3D structure are marked as ‘-’

timer and far from the solvent accessible surface. The trimer 6VXX contains an arginine (R) at position 355. The landscape plot suggests that this is an unstable residue position and that any other amino acid mutation at this position would increase the stability of the domain region. Hence, it is highly likely to observe a mutation at this position, which has actually been reported in a sequence from the UK on GISAID [22].

Using the 65,000 spike protein sequences collected internationally and made available at GISAID [22], we found 3,405 spike sequences that diverged from the 6VXX reference spike sequence. 2,491 of those sequences contained missing readings in the receptor-binding domain and were excluded from further analysis. Twenty-one sequences presented indels (insertions and deletions) that might have affected the progression or severity of the disease. Countries reporting these sequences with indels are Australia, China, India, Israel, Qatar, UK, and USA. Interestingly, among these countries Qatar was the only country to report a spike protein with an insertion resulting in the longest known spike mutant with 196 amino acids. It is not known whether or not this longer sequence

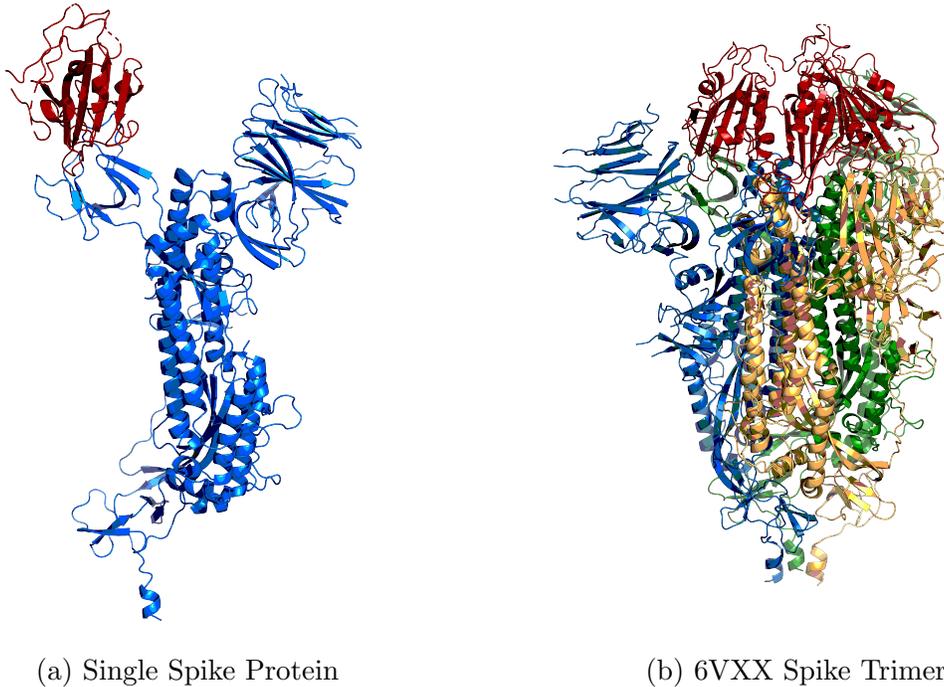


Figure 1: 3D Structure of the Spike Protein. A single spike protein (a) in blue contains a receptor-binding region highlighted in red. Three single spike proteins colored blue, orange and green aggregate to form a trimer structure as shown in (b) with a receptor-binding domain colored red that can interact with an ACE2 enzyme.

contributed to the country’s low COVID-19 mortality rate (0.1%).

The remaining 894 sequences had equal length to the 6VXX spike sequence but exhibited one or more point mutations. We report in Table 2 the top country origins of these sequences and the ΔE values generated using our mutation landscape and Equation 2. Table S2 of the supplementary material is an extension of Table 2 and contains the complete list of countries. We observe that all the countries in Table 2 report mutations that both stabilize and destabilize the receptor-binding domain, which could explain the varying severity of disease symptoms and health conditions reported in those countries [33].

Table 3 lists the spike mutations that appear worldwide ranked in order of the number of reporting countries. Mutation V367F was reported in 12 countries and appeared in 51 sequences. Our energy landscape suggests that this mutation is favorable and further stabilizes the receptor-binding domain ($\Delta E < 0$). Other mutations exhibit positive ΔE values and could lead to different binding potentials with the ACE2 enzyme, affecting the rate of disease, incubation periods and patient symptoms. Table S3 of the supplementary material reports the data for all 894 sequences. Table S4 of the supplementary material presents the point mutations reported in each country as of July 2020.

Out of the 894 sequences with mutations, 26 sequences exhibited 2 or more simultaneous point mutations. Table 4 reports these multiple-point mutations. It is interesting

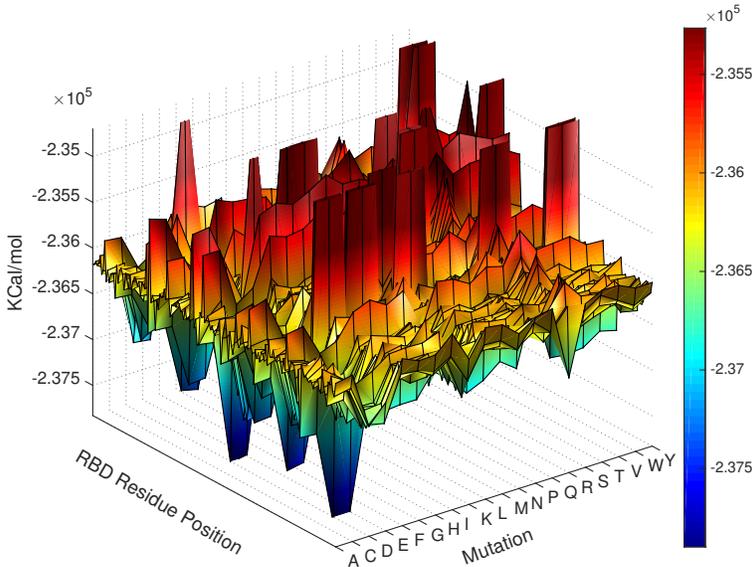
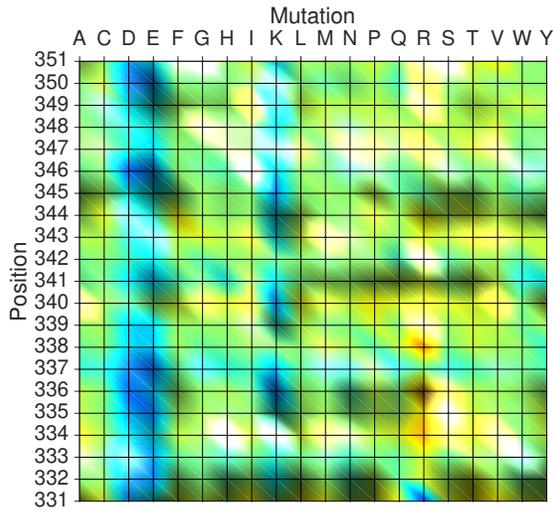
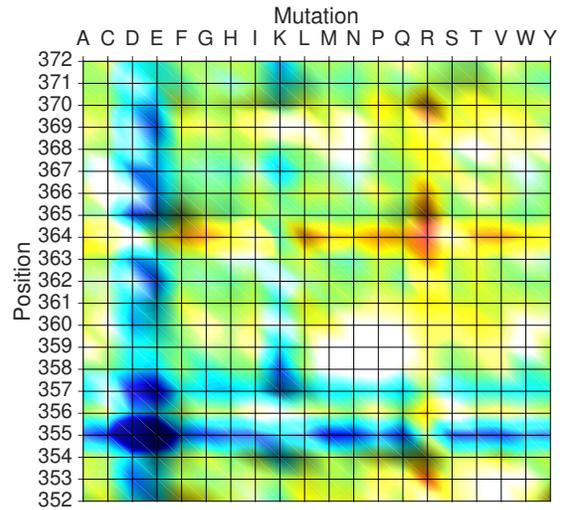


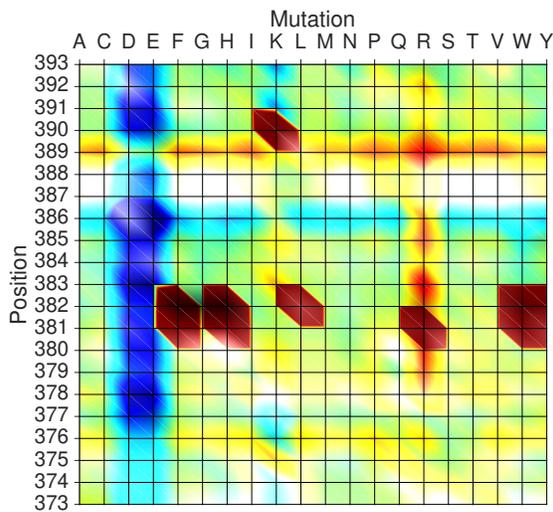
Figure 2: Mutation Landscape of the receptor-binding domain in SARS-CoV-2. Energy values are in kcal/mol and are computed by Eq. 1. Each x,y coordinate represents a mutation x at a position y. The z-values are the free energies of the mutated structures. Lower energies (blue) correspond to increased stability.



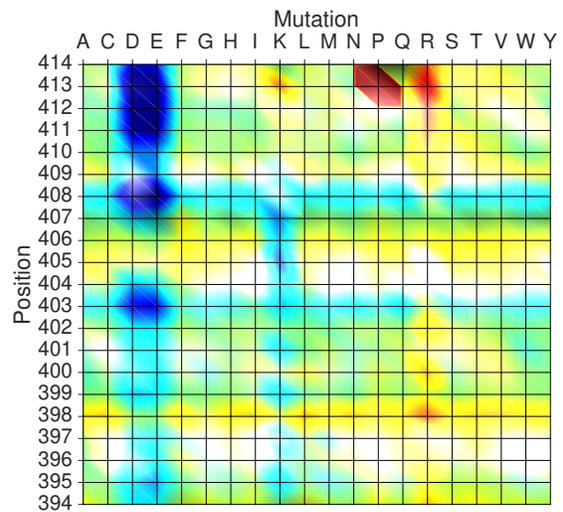
(a) Residue Positions 331-351



(b) Residue Positions 352-372



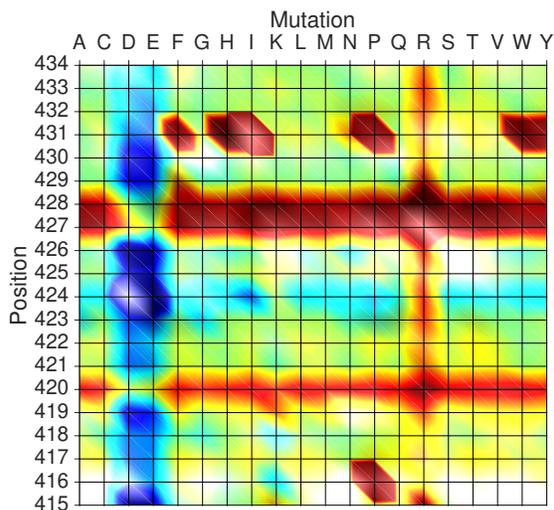
(c) Residue Positions 373-393



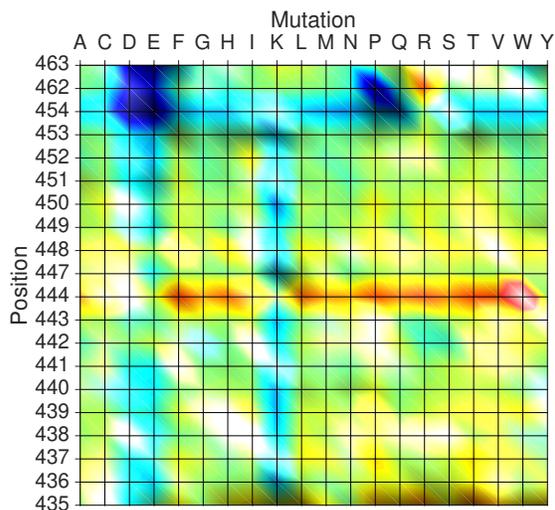
(d) Residue Positions 394-414

Figure 3: 2D Projection of the Mutation Landscape Representing Positions 331-414

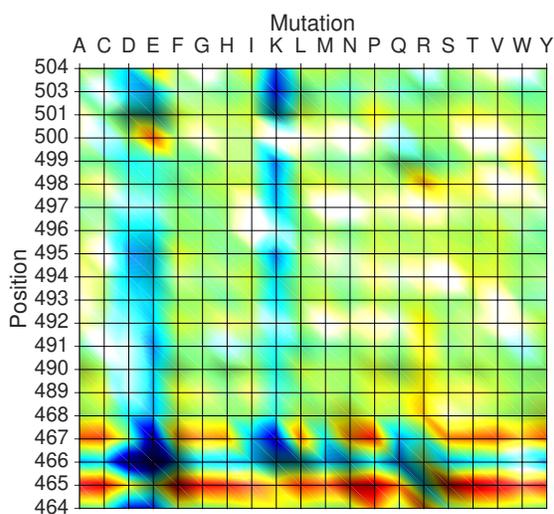
to note that all recent 2, 3, and 4-point mutations have a $\Delta E < 0$, suggesting a mutation drive to further stabilize the receptor-binding domain and potentially increase infectivity. The 2-point mutation (V367F G413V) that was reported in Spain appears to have evolved from the widely spread mutated sequence (V367F) reported in 12 countries. No other sequence reported a G413V mutation. Although we do not know the order in which



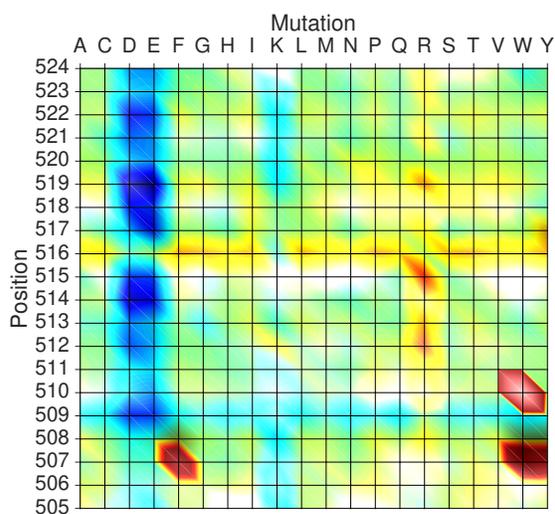
(a) Residue Positions 415-434



(b) Residue Positions 435-463



(c) Residue Positions 464-504



(d) Residue Positions 505-524

Figure 4: 2D Projection of the Mutation Landscape Representing Positions 415-524

the other mutations have appeared, it is theoretically possible for a sequence to undergo a destabilizing mutation first and then after some time experience a strong stabilizing one that brings the structure to a more overall stable conformation. The sequence with the 3-point mutation (Q506H P507S Y508N) could have undergone its first destabilizing mutation (Q506H $\Delta E > 0$) and then experienced two stabilizing single point mutations

(P507S $\Delta E < 0$) and (Y508N $\Delta E < 0$). If this is the case, then other spike sequences that have become mild or less dangerous because of destabilizing mutations could potentially experience future mutations that cause a regain in toxicity potential, and cause a periodic increase and decrease in COVID-19 symptoms.

We report in Table S5 of the supplementary material two sequences from China with 6 and 7-point mutations reported in 2019. We have not included these in our analysis as it was strange for that number of mutations to occur early on during the pandemic. A potential vaccine for SARS-CoV-2 or molecular therapeutic, that can inhibit the binding between the receptor-binding domain and the ACE2 enzyme, would need to work on different spike mutants that have started to appear and spread throughout different countries. It is not feasible to create a vaccine tailored to work on each mutant. However, if a vaccine can show good results on one of the most stable mutant structures, then it is possible that it will also show good results on many less stable ones.

In light of this, we used the energy values of the landscape to explore the $\Delta\tilde{E}$ values given by Equation 4 of all possible 7.5 million 2-point mutations. Table 5 reports the top 30 most stable 2-point mutations and Table S6 of the supplementary material reports the top 1000. Our analysis show that the 2-point mutation (R355D K424E) contributes strong structural stability to the spike protein and should be tested against potential vaccines and inhibitors.

The 2-point and 3-point mutations that countries have reported are still far from the most stable conformations possible. It appears that the virus still has millions of possible

Table 2: Top reporting countries for Spike sequences on GISAID

Reporting Origin	Submitted Sequences	% of total Sequences	Sequences with mutations	min ΔE	avg ΔE	max ΔE
United Kingdom	30,494	47.25	582	-681	-268	6,966
USA	12,912	20.01	67	-1,094	182	7,498
Australia	2,052	3.18	28	-375	267	6,966
Spain	1,784	2.76	7	-344	114	600
Netherlands	1,582	2.45	30	-366	-7	157
India	1,563	2.42	26	-416	-20	130
Canada	1,018	1.58	4	-497	-343	64
China	977	1.51	28	-457	43	593
Belgium	921	1.43	3	32	118	286
Switzerland	726	1.12	4	-74	-46	7
Portugal	702	1.09	11	-37	102	116
Denmark	655	1.01	2	20	51	83
Singapore	638	0.99	3	84	88	91
Japan	627	0.97	2	-509	-217	74

candidate mutations to select from to increase the receptor-binding domain’s stability and potentially become more toxic and lethal.

Table 3: Occurrences of Spike sequence mutations reported on GISAID. Mutation aPb denotes the change of amino acid a at position P to amino acid b .

Mutation	Occurrence	Reporting Countries	ΔE (kcal/mol)	Stabilization Rank (%)
V367F	51	12	-73.64	23.78
P384S	7	6	-37.3	31.48
A520S	23	6	21.4	52.71
S494P	12	6	83.93	73.35
P384L	21	5	31.53	57.05
V382L	7	5	6966.32	99.82
N439K	422	4	-417.5	6.92
P463S	8	4	-43.74	26.74
A348S	8	4	68.61	69.42
A344S	20	4	78.22	71.42
R408I	12	3	-328.02	10.62
S359N	3	3	-68.6	24.71
Y508H	4	3	-33.91	32.22
I468V	6	3	28.91	56.03
H519Q	4	3	66.69	68.92
F490L	4	3	90.57	75.60
A352S	7	3	128.22	83.08
A411S	3	3	143.49	85.17
A435S	3	3	156.87	86.46
G413E	2	2	-1093.77	0.43
R403K	9	2	-375.13	8.37
R346K	2	2	-366.34	8.77
G504D	2	2	-344.41	9.91
N354K	2	2	-340.27	10.15
G339D	4	2	-324.91	10.71
P521S	3	2	-109.11	20.00
Y505H	2	2	-104.54	20.28
Q414K	3	2	-37.51	31.42
A372T	14	2	-35.55	31.85
I402V	12	2	-2.36	42.80

Table 4: Multiple-point mutations in the Spike protein reported on GISAID

m-point Mutation	Occurrence	Reporting Country	Date	$\Delta\tilde{E}$ (kcal/mol)
R509K V510L	1	China	Feb 2020	-561.68
A411S C432F	1	China	Feb 2020	184.5
G404K D405C	1	India	Mar 2020	-411.52
D467V I468F	1	Australia	Mar 2020	558.79
N388T S399P	1	Taiwan	Mar 2020	132.0
P491L H519Q	1	Malaysia	Mar 2020	-118.41
G339D E340K	3	Canada	Apr 2020	-837.15
V367F G413V	1	Spain	Apr 2020	-35.96
N439K S494P	1	UK	Jun 2020	-362.94
P507H Y508N	1	India	Jun 2020	-182.07
N354D D364Y	1	China	Jan 2020	-92.46
S399A V407G I410V	1	India	Mar 2020	-55.45
Q506H P507S Y508N	1	India	Jun 2020	-118.12
F347I A348P T430N G431S	1	India	Jun 2020	-114.30

Table 5: Top Stabilizing 2-point mutations in the receptor-binding domain

Mutation	$\Delta\tilde{E}$
R355D K424E	-3,418.13
R355E K424D	-3,389.30
R355E K424E	-3,357.42
R355D K386D	-3,202.04
K386D K424D	-3,182.71
R355D K386E	-3,168.83
K386D K424E	-3,150.83
K386E K424D	-3,149.50
R355E K386D	-3,141.33
K386E K424E	-3,117.62
R355E K386E	-3,108.12
R355D P426D	-2,998.42
R355D R466E	-2,995.17
R355D F429D	-2,985.36
K424D P426D	-2,979.09
K424D R466E	-2,975.84
K424D F429D	-2,966.03
R355D F429E	-2,963.68
R355D R466D	-2,950.92
K424E P426D	-2,947.21
K424D F429E	-2,944.35
K424E R466E	-2,943.96
R355D G413D	-2,939.45
R355E P426D	-2,937.71
R355E R466E	-2,934.46
K424E F429D	-2,934.15
K424D R466D	-2,931.59
R355E F429D	-2,924.65
G413D K424D	-2,920.12
K424E F429E	-2,912.47

3 Methods

To assess the effect of mutations in the receptor-binding domain on the stability of the spike protein, we selected the cryo-EM SARS-CoV-2 spike glycoprotein trimer structure (PDB ID 6VXX) [32] as a 3D blueprint model on which we will perform mutations.

3.1 Constructing Spike Mutations

The spike protein attached to the membrane of SARS-CoV-2 is a trimer structure. Three copies of the spike protein aggregate to form the receptor-binding domain at their terminal region. This region binds to the ACE2 enzyme in various cells. We devised a new algorithm, called *SpikeMutator* (Algorithm 1), to create 3D atomic models for all possible single-point mutations in the receptor-binding domain (positions 331-524) and calculate the free energies of each structure. Each of the 194 amino acids in this region is mutated into the 19 other possible canonical amino acids using SCWRL4 [34], producing 3,880 single-point mutation possibilities. Every mutation is applied to each of the three chains in the trimer structure separately and the three structures are subsequently joined and run through an energy minimization step to relax any steric clashes. For each mutated structure, *SpikeMutator* calculates the Lennard-Jones (LJ), Coulomb, and Solvation energies.

3.2 Computing Spike Structure Energies

The free energy of a protein molecule in a solvent medium is correlated with its structural stability. In general, increased stability promotes better binding potential with other molecules. Lower stability can indicate weaker binding potential. We calculate the free energy E_m^K of a mutant spike structure with a mutation m at position K by computing the LJ, Coulomb, and solvation energy values of the mutant trimer using Equation 1,

$$E_m^K = LJ_m^K + Coul_m^K + S_m^K \quad (1)$$

where LJ_m^K is the Lennard-Jones potential, $Coul_m^K$ is the Coulomb energy, and S_m^K measures the solvation energy resulting from the contact of the trimer surface with water molecules for the spike structure with mutation m at position K . The LJ and $Coul$ terms measure the electrostatic potential and charges between the atoms of the trimer structure computed after undergoing an energy minimization step to reduce any steric clashes introduced in the mutation phase.

Low free energy values E indicate increased stability in the receptor-binding domain, and improved overall stability in the spike protein. Conversely, high E values can indicate increased instability in the receptor-binding domain, and lower overall stability in the spike trimer structure.

Algorithm 1 Single-point mutations simulation algorithm

```
1: function SpikeMutator(TrimerModel, Sequence)
2:   MutationList  $\leftarrow$  List(“ARNDCQEGHILKMFPSTWYV”)
3:   RBD  $\leftarrow$  [331, 332, ..., 524]
4:   SpikeA  $\leftarrow$  TimerModel.chain1
5:   SpikeB  $\leftarrow$  TimerModel.chain2
6:   SpikeC  $\leftarrow$  TimerModel.chain3
7:   Landscape  $\leftarrow$  [ ]
8:   for pos in RBD do
9:     for m in MutationList do
10:      newSequence  $\leftarrow$  mutate(Sequence, pos, m)
11:      MutantA  $\leftarrow$  mutateStructure(SpikeA, newSequence)
12:      MutantB  $\leftarrow$  mutateStructure(SpikeB, newSequence)
13:      MutantC  $\leftarrow$  mutateStructure(SpikeC, newSequence)
14:      Spike  $\leftarrow$  join(MutantA, MutantB, MutantC)
15:      Spike  $\leftarrow$  prepareMDSetup(Spike)
16:      Spike  $\leftarrow$  energyMinimizeMD(Spike)
17:      CoulombE  $\leftarrow$  Spike.calculateCoulombEnergy()
18:      LJE  $\leftarrow$  Spike.calculateLJEnergy()
19:      SolvationE  $\leftarrow$  calculateSolvationEnergy(Spike)
20:      Landscape.add( [ pos, m, Spike, CoulombE, LJE, SolvationE ] )
21:     end for
22:   end for
23:   return Landscape
24: end function
```

The E values generated by *SpikeMutator* are displayed in Fig 2. The figure describes an energy landscape of all possible single-point mutations in the receptor-binding domain. The amino acid positions of the domain are plotted on the x-axis, the 20 possible mutations on the y-axis, and the E value produced by a (position, mutation) pair makes up the z-axis.

Each mutation can be characterized by the difference in energy between its final mutated state and its initial non-mutated state. This change of energy at a position K to some amino acid m can be captured by Equation 2,

$$\Delta E_m^K = E_m^K - E^{(0)} \quad (2)$$

where E_m^K is the free energy of the spike protein with amino acid m at position K and $E^{(0)}$ is the free energy of the initial non-mutated spike. Negative ΔE values suggest mutations that increase stability and positive ΔE values suggest mutations that are potentially destabilizing.

3.3 Solvation Energy using Dipolar water solvent

The Solvation term is computed using a fast and detailed dipolar water model that solves the dipolar nonlinear Poisson-Boltzmann-Langevin equation using the AQUASOL subroutine [35]. More precisely, the solvation energy in *SpikeMutator* is computed by Equation 3.

$$S = F_{(p_0, C_{dip})} - F_{(0,0)} - \left(k_B T \frac{\ln(1 - N_A C_{dip} a^3)}{N_A C_{dip} a^3} \right) \int_{solvent} d\mathbf{r} \rho_{dip}(\mathbf{r}) \quad (3)$$

where $F_{(p_0, C_{dip})}$ defines the free energy of the system defined at dipoles of moment values p_0 and concentration C_{dip} , $F_{(0,0)}$ the free energy of the system with solvent concentration set to zero, a^3 is the lattice grid size volume of the solvent, k_B is the Boltzmann constant, T temperature in Kelvin, and r is the surface definition of the solvent-accessible surface probe.

The Solvation energy is calculated with the following setup: atomic charges and radii assigned with PDB2PQR using CHARMM force field at neutral pH. A grid of 257 points per edge spaced by 1 Å, a temperature of 300K, and a solvent accessible surface with an Rprobe of 1.4 Å. All hydrogen-bonds were optimized. We used a trilinear interpolation protocol for projection of fixed charges on the grid, a lattice grid size for the solvent: $a = 2.8$ Å, solvent made of dipoles of moment $p_0 = 3.00$ D at a concentration of $C_{dip} = 55$ M. No salt was added to the solution and small ions were used to equilibrate the system when needed. The electrostatic potential was set to zero at the boundaries, and the stopping criteria for residual was sent to: 1.10^{-6} (when possible). Although using an implicit water model is more computationally expensive, the dipolar model provides more accurate energies and accounts for more realistic atomic interactions.

3.4 Energy Minimization & Molecular Dynamics

The process of mutating an amino acid of a structure into another amino acid can often introduce steric clashes with neighbouring residues in its spacial vicinity. After performing a point mutation in each of the three spike chains, *SpikeMutator* joins the three structures back into their initial conformation. Prior to calculating the energy of this new structure, we perform a short energy minimization (EM) run using the GROMACS 2019.3 molecular dynamics package [36] to relax the structure and remove any severe steric clashes. The setup of the run is as follows: molecules were prepared in a cubic box (with a minimum distance of 35 Å from any edge of the box to any atom) and neutralized with chloride ions and modeled using the AMBER99SB-ILDN [37] force field along with the TIP-3P water model. We used a cutoff of 10 Å for van der Waals and short range electrostatic interactions, and calculated long range electrostatic interactions using a particle mesh Ewald sum [38, 39]. 2000 EM steps were performed for every structure using a steepest descent algorithm and the Verlet cut-off scheme.

3.5 Multiple-point mutations

The number of possible mutations grows exponentially as we consider higher-order point mutations. In the receptor-binding domain, there are 7.5 million possible combinations of 2-point mutations, 9.6 billion possible combinations of 3-point mutations, and 9.2 trillion possible 4-point mutations. Since it is not feasible to efficiently calculate the ΔE values of these multiple-point mutations, we can estimate the change in energy caused by multiple-point mutations by summing the individual ΔE values produced by each single-point mutation to produce an estimate change in energy given by Equation 4,

$$\Delta\tilde{E} = \sum_{K \in \text{dom}(\mathcal{M})} \Delta E_{\mathcal{M}(K)}^K \quad (4)$$

where \mathcal{M} is a map between spike amino acid positions and mutations, $\text{dom}(\mathcal{M})$, returns the positions of the requested mutations, $\mathcal{M}(K)$ returns the desired amino acid mutation at position K .

3.6 Supercomputer Resources

To perform our simulations, we resorted to the cloud-computing system provided by Amazon Web Services (AWS). Constructing the energy landscape was possible with the utilization of 24 EC2 machines with 16 CPUs each running in parallel for several weeks. The SpikeMutator algorithm and a copy of the GROMACS 2019.3 molecular dynamics package [36] was installed on each machine. Generating the 7.5 million possible combinations of 2-point mutations and ranking their energies was also made possible with these machines on AWS.

4 Conclusion

We studied in this work the stability of the spike protein under all possible single-point mutations in the receptor-binding domain and explored mutations that can influence structural stability and affect binding with the ACE2 enzyme. We devised a tool, called SpikeMutator, to construct full atomic protein structures of the mutant spike proteins and generated a database of all possible single-point mutant trimer structures. We observed that the sequences isolated from COVID-19 patients exhibited some mutations that both increased and decreased the spike’s structural stability. Out of the 7.5 million possible 2-point mutation combinations, we found that the (R355D K424E) mutation produces one of the most stable spike proteins and should be included in the testing of possible vaccines and molecular inhibitors of SARS-CoV-2.

Our future work will be dedicated to the elaboration of a mutation model that captures the transitions between different spike mutation states in order to detect multiple-point mutations that can cycle between low to high energy states. This would potentially provide some empirical evidence on how mutations can manifest different clinical symptoms.

Database Availability

The 3D conformations of all 3,880 SARS-CoV-2 spike trimer mutants can be accessed at <http://spikemutator.com>. The SpikeMutator program is available for academic use and can be downloaded entirely from the same website.

Author Contributions

MRS and HY designed the study. MRS developed code and performed experiments. MRS and HY analyzed the results. MRS and HY read and approved the final manuscript.

Competing Interests

The authors declare no potential conflict of interest with respect to the research, authorship, and publication of this article.

References

- [1] D. S. Hui, E. I Azhar, T. A. Madani, F. Ntoumi, R. Kock, O. Dar, G. Ippolito, T. D. Mchugh, Z. A. Memish, C. Drosten, A. Zumla, and E. Petersen, “The continuing 2019-ncov epidemic threat of novel coronaviruses to global health - the latest 2019 novel coronavirus outbreak in wuhan, china,” *Int J Infect Dis*, vol. 91, pp. 264–266, 02 2020.
- [2] P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, H.-D. Chen, J. Chen, Y. Luo, H. Guo, R.-D. Jiang, M.-Q. Liu, Y. Chen, X.-R. Shen, X. Wang, X.-S. Zheng, K. Zhao, Q.-J. Chen, F. Deng, L.-L. Liu, B. Yan, F.-X. Zhan, Y.-Y. Wang, G.-F. Xiao, and Z.-L. Shi, “A pneumonia outbreak associated with a new coronavirus of probable bat origin,” *Nature*, vol. 579, pp. 270–273, 03 2020.
- [3] Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, “The species severe acute respiratory syndrome-related coronavirus: classifying 2019-ncov and naming it sars-cov-2,” *Nat Microbiol*, vol. 5, pp. 536–544, 04 2020.
- [4] A. M. Zaki, S. van Boheemen, T. M. Bestebroer, A. D. M. E. Osterhaus, and R. A. M. Fouchier, “Isolation of a novel coronavirus from a man with pneumonia in saudi arabia,” *N Engl J Med*, vol. 367, pp. 1814–20, Nov 2012.
- [5] N. S. Zhong, B. J. Zheng, Y. M. Li, Poon, Z. H. Xie, K. H. Chan, P. H. Li, S. Y. Tan, Q. Chang, J. P. Xie, X. Q. Liu, J. Xu, D. X. Li, K. Y. Yuen, Peiris, and Y. Guan, “Epidemiology and cause of severe acute respiratory syndrome (sars) in guangdong, people’s republic of china, in february, 2003,” *Lancet*, vol. 362, pp. 1353–8, Oct 2003.
- [6] A. A. Rabaan, S. H. Al-Ahmed, S. Haque, R. Sah, R. Tiwari, Y. S. Malik, K. Dhama, M. I. Yattoo, D. K. Bonilla-Aldana, and A. J. Rodriguez-Morales, “Sars-cov-2, sars-cov, and mers-cov: A comparative overview,” *Infez Med*, vol. 28, no. 2, pp. 174–184, 2020.
- [7] A. Grifoni, J. Sidney, Y. Zhang, R. H. Scheuermann, B. Peters, and A. Sette, “A sequence homology and bioinformatic approach can predict candidate targets for immune responses to sars-cov-2,” *Cell Host Microbe*, vol. 27, pp. 671–680.e2, 04 2020.
- [8] M. Nicola, Z. Alsafi, C. Sohrabi, A. Kerwan, A. Al-Jabir, C. Iosifidis, M. Agha, and R. Agha, “The socio-economic implications of the coronavirus pandemic (covid-19): A review,” *Int J Surg*, vol. 78, pp. 185–193, Jun 2020.

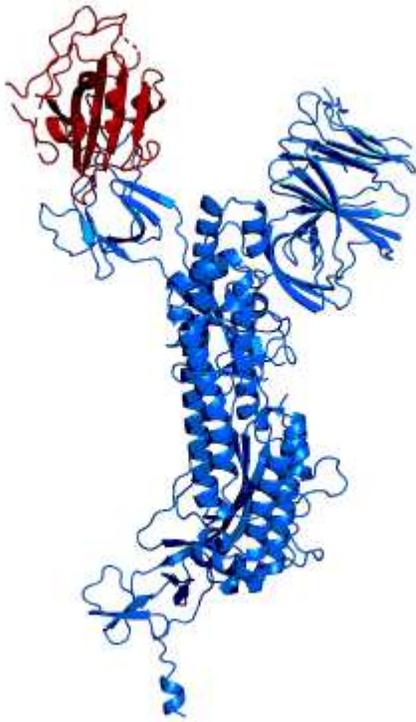
- [9] R. M. Viner, S. J. Russell, H. Croker, J. Packer, J. Ward, C. Stansfield, O. Mytton, C. Bonell, and R. Booy, “School closure and management practices during coronavirus outbreaks including covid-19: a rapid systematic review,” *Lancet Child Adolesc Health*, vol. 4, pp. 397–404, 05 2020.
- [10] J. Torales, M. O’Higgins, J. M. Castaldelli-Maia, and A. Ventriglio, “The outbreak of covid-19 coronavirus and its impact on global mental health,” *Int J Soc Psychiatry*, vol. 66, pp. 317–320, 06 2020.
- [11] S. Xu and Y. Li, “Beware of the second wave of covid-19,” *Lancet*, vol. 395, pp. 1321–1322, 04 2020.
- [12] C. A. de Haan, M. Smeets, F. Vernooij, H. Vennema, and P. J. Rottier, “Mapping of the coronavirus membrane protein domains involved in interaction with the spike protein,” *J Virol*, vol. 73, pp. 7441–52, Sep 1999.
- [13] S. Liu, G. Xiao, Y. Chen, Y. He, J. Niu, C. R. Escalante, H. Xiong, J. Farmar, A. K. Debnath, P. Tien, and S. Jiang, “Interaction between heptad repeat 1 and 2 regions in spike protein of sars-associated coronavirus: implications for virus fusogenic mechanism and identification of fusion inhibitors,” *Lancet*, vol. 363, pp. 938–47, Mar 2004.
- [14] P. Verdecchia, C. Cavallini, A. Spanevello, and F. Angeli, “The pivotal link between ace2 deficiency and sars-cov-2 infection,” *Eur J Intern Med*, vol. 76, pp. 14–20, 06 2020.
- [15] Y. He, J. Li, S. Heck, S. Lustigman, and S. Jiang, “Antigenic and immunogenic characterization of recombinant baculovirus-expressed severe acute respiratory syndrome coronavirus spike protein: implication for vaccine design,” *J Virol*, vol. 80, pp. 5757–67, Jun 2006.
- [16] A. Bonavia, B. D. Zelus, D. E. Wentworth, P. J. Talbot, and K. V. Holmes, “Identification of a receptor-binding domain of the spike glycoprotein of human coronavirus hcov-229e,” *J Virol*, vol. 77, pp. 2530–8, Feb 2003.
- [17] I. Hamming, W. Timens, M. L. C. Bulthuis, A. T. Lely, G. J. Navis, and H. van Goor, “Tissue distribution of ace2 protein, the functional receptor for sars coronavirus. a first step in understanding sars pathogenesis,” *J Pathol*, vol. 203, pp. 631–7, Jun 2004.
- [18] L. Du, Y. He, Y. Zhou, S. Liu, B.-J. Zheng, and S. Jiang, “The spike protein of sars-cov—a target for vaccine and therapeutic development,” *Nat Rev Microbiol*, vol. 7, pp. 226–36, Mar 2009.

- [19] L. Du, W. Tai, Y. Yang, G. Zhao, Q. Zhu, S. Sun, C. Liu, X. Tao, C.-T. K. Tseng, S. Perlman, S. Jiang, Y. Zhou, and F. Li, “Introduction of neutralizing immunogenicity index to the rational design of mers coronavirus subunit vaccines,” *Nat Commun*, vol. 7, p. 13473, 11 2016.
- [20] L. Du, Y. Yang, Y. Zhou, L. Lu, F. Li, and S. Jiang, “Mers-cov spike protein: a key target for antivirals,” *Expert Opin Ther Targets*, vol. 21, pp. 131–143, Feb 2017.
- [21] L. A. Jackson, E. J. Anderson, N. G. Roupael, P. C. Roberts, M. Makhene, R. N. Coler, M. P. McCullough, J. D. Chappell, M. R. Denison, L. J. Stevens, A. J. Pruijssers, A. McDermott, B. Flach, N. A. Doria-Rose, K. S. Corbett, K. M. Morabito, S. O’Dell, S. D. Schmidt, P. A. Swanson, 2nd, M. Padilla, J. R. Mascola, K. M. Neuzil, H. Bennett, W. Sun, E. Peters, M. Makowski, J. Albert, K. Cross, W. Buchanan, R. Pikaart-Tautges, J. E. Ledgerwood, B. S. Graham, J. H. Beigel, and mRNA-1273 Study Group, “An mrna vaccine against sars-cov-2 - preliminary report,” *N Engl J Med*, Jul 2020.
- [22] Y. Shu and J. McCauley, “Gisaid: Global initiative on sharing all influenza data - from vision to reality,” *Euro Surveill*, vol. 22, 03 2017.
- [23] S. K. P. Lau, P. C. Y. Woo, K. S. M. Li, Y. Huang, H.-W. Tsoi, B. H. L. Wong, S. S. Y. Wong, S.-Y. Leung, K.-H. Chan, and K.-Y. Yuen, “Severe acute respiratory syndrome coronavirus-like virus in chinese horseshoe bats,” *Proc Natl Acad Sci U S A*, vol. 102, pp. 14040–5, Sep 2005.
- [24] Chinese SARS Molecular Epidemiology Consortium, “Molecular evolution of the sars coronavirus during the course of the sars epidemic in china,” *Science*, vol. 303, pp. 1666–9, Mar 2004.
- [25] Y. Yang, C. Liu, L. Du, S. Jiang, Z. Shi, R. S. Baric, and F. Li, “Two mutations were critical for bat-to-human transmission of middle east respiratory syndrome coronavirus,” *J Virol*, vol. 89, pp. 9119–23, Sep 2015.
- [26] N. R. Meyerson and S. L. Sawyer, “Two-stepping through time: mammals and viruses,” *Trends Microbiol*, vol. 19, pp. 286–94, Jun 2011.
- [27] A. Demogines, M. Farzan, and S. L. Sawyer, “Evidence for ace2-utilizing coronaviruses (covs) related to severe acute respiratory syndrome cov in bats,” *J Virol*, vol. 86, pp. 6350–3, Jun 2012.
- [28] T. Kar, U. Narsaria, S. Basak, D. Deb, F. Castiglione, D. M. Mueller, and A. P. Srivastava, “A candidate multi-epitope vaccine against sars-cov-2,” *Sci Rep*, vol. 10, p. 10895, 07 2020.

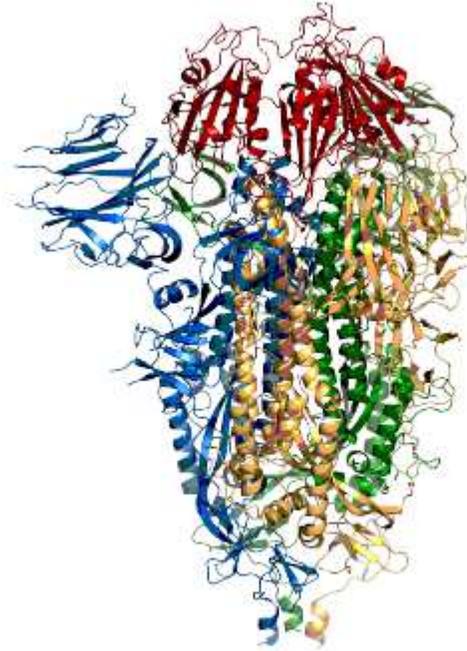
- [29] Y. Weisblum, F. Schmidt, F. Zhang, J. DaSilva, D. Poston, J. C. C. Lorenzi, F. Muecksch, M. Rutkowska, H.-H. Hoffmann, E. Michailidis, C. Gaebler, M. Agudelo, A. Cho, Z. Wang, A. Gazumyan, M. Cipolla, L. Luchsinger, C. D. Hillyer, M. Caskey, D. F. Robbiani, C. M. Rice, M. C. Nussenzweig, T. Hatziioannou, and P. D. Bieniasz, “Escape from neutralizing antibodies by sars-cov-2 spike protein variants,” *bioRxiv*, Jul 2020.
- [30] M. R. Smaoui and J. Waldispühl, “Complete characterization of the mutation landscape reveals the effect on amylin stability and amyloidogenicity,” *Proteins*, vol. 83, pp. 1014–26, Jun 2015.
- [31] M. R. Smaoui, C. Mazza-Anthony, and J. Waldispühl, “Investigating mutations to reduce huntingtin aggregation by increasing htt-n-terminal stability and weakening interactions with polyq domain,” *Comput Math Methods Med*, vol. 2016, p. 6247867, 2016.
- [32] A. C. Walls, Y.-J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, and D. Veessler, “Structure, function, and antigenicity of the sars-cov-2 spike glycoprotein,” *Cell*, vol. 181, pp. 281–292.e6, 04 2020.
- [33] M. Gao, L. Yang, X. Chen, Y. Deng, S. Yang, H. Xu, Z. Chen, and X. Gao, “A study on infectivity of asymptomatic sars-cov-2 carriers,” *Respir Med*, vol. 169, p. 106026, 08 2020.
- [34] G. G. Krivov, M. V. Shapovalov, and R. L. Dunbrack, Jr, “Improved prediction of protein side-chain conformations with scwrl4,” *Proteins*, vol. 77, pp. 778–95, Dec 2009.
- [35] P. Koehl and M. Delarue, “Aquasol: An efficient solver for the dipolar poisson-boltzmann-langevin equation,” *J Chem Phys*, vol. 132, p. 064101, Feb 2010.
- [36] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl, “Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit,” *Bioinformatics*, vol. 29, pp. 845–54, Apr 2013.
- [37] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw, “Improved side-chain torsion potentials for the amber ff99sb protein force field,” *Proteins*, vol. 78, pp. 1950–8, Jun 2010.
- [38] T. Darden, D. York, and L. Pedersen, “Particle mesh ewald - an $n \cdot \log(n)$ method for ewald sums in large systems,” *Journal of Chemical Physics*, vol. 98, pp. 10089–10092, 1993.

- [39] U. Essmann, L. Perera, M. Berkowitz, T. Darden, H. Lee, and L. Pedersen, “A smooth particle mesh ewald method,” *Journal of Chemical Physics*, vol. 103, pp. 8577–8593, 1995.

Figures



(a) Single Spike Protein



(b) 6VXX Spike Trimer

Figure 1

3D Structure of the Spike Protein. A single spike protein (a) in blue contains a receptor-binding region highlighted in red. Three single spike proteins colored blue, orange and green aggregate to form a trimer structure as shown in (b) with a receptor-binding domain colored red that can interact with an ACE2 enzyme.

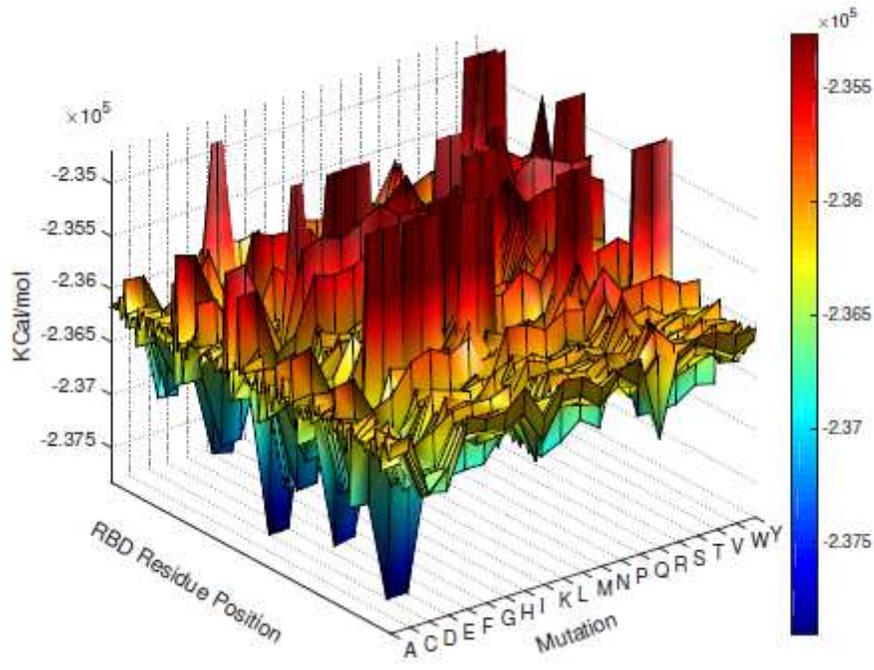
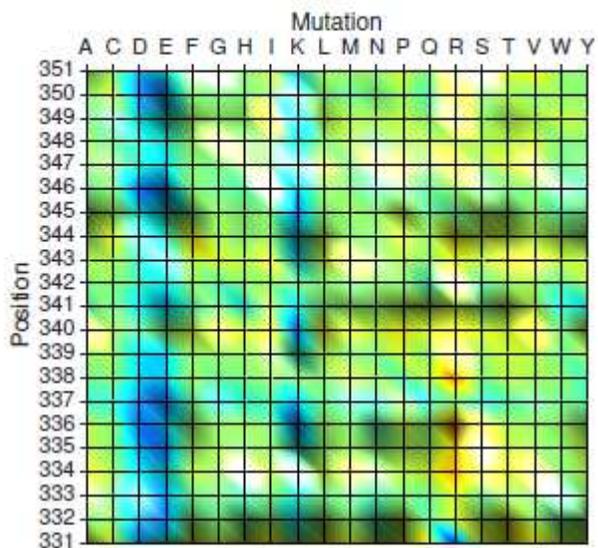
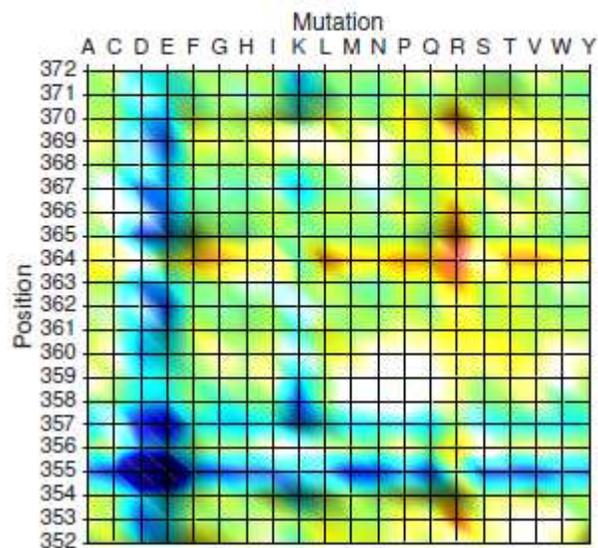


Figure 2

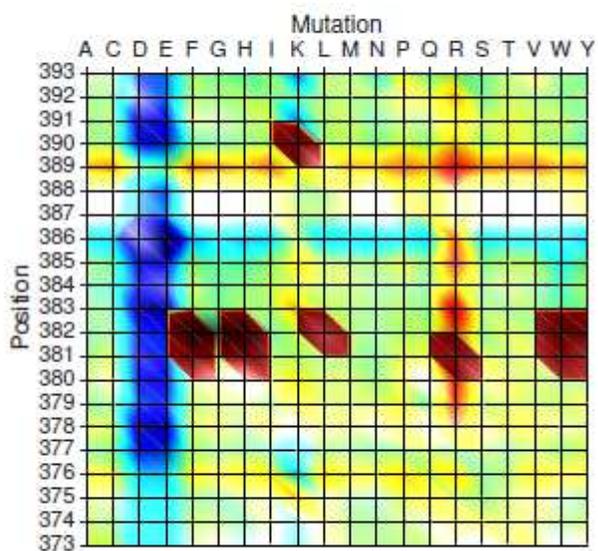
Mutation Landscape of the receptor-binding domain in SARS-CoV-2. Energy values are in kcal/mol and are computed by Eq. 1. Each x,y coordinate represents a mutation x at a position y. The z-values are the free energies of the mutated structures. Lower energies (blue) correspond to increased stability.



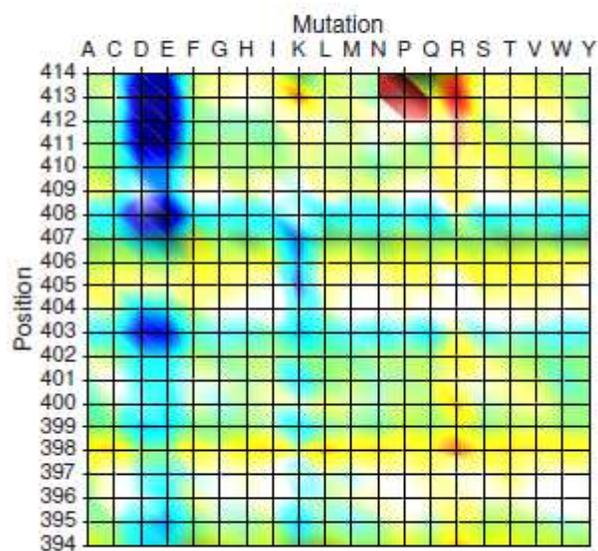
(a) Residue Positions 331-351



(b) Residue Positions 352-372



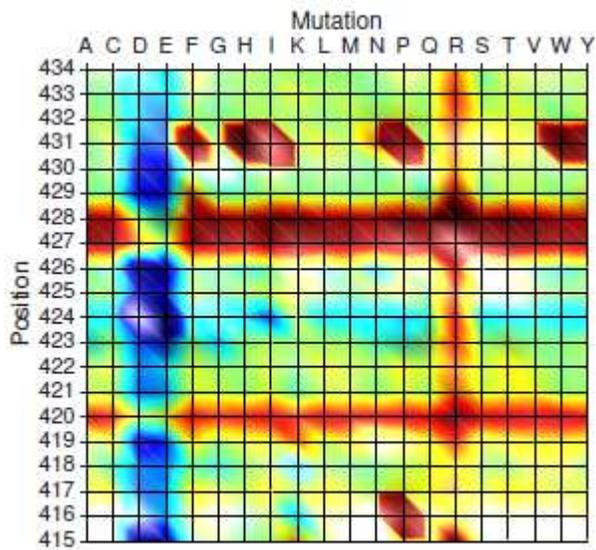
(c) Residue Positions 373-393



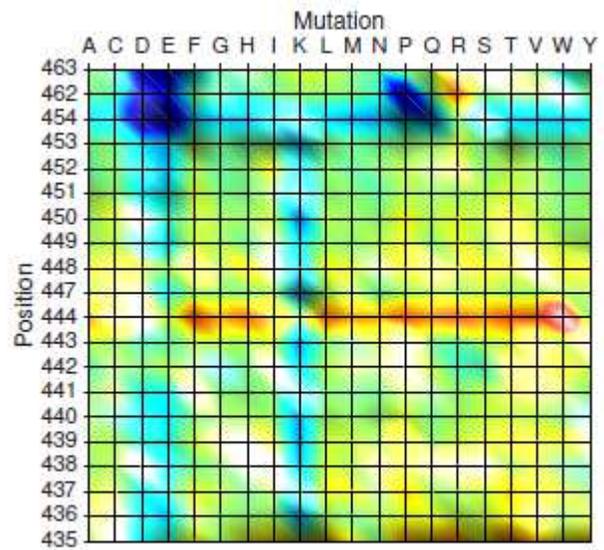
(d) Residue Positions 394-414

Figure 3

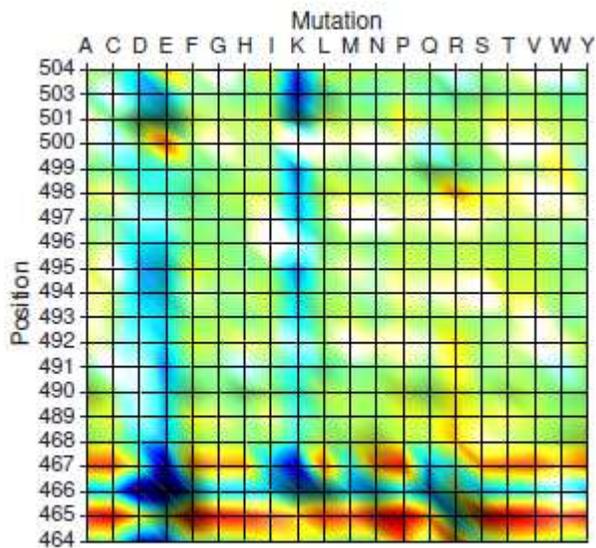
2D Projection of the Mutation Landscape Representing Positions 331-414



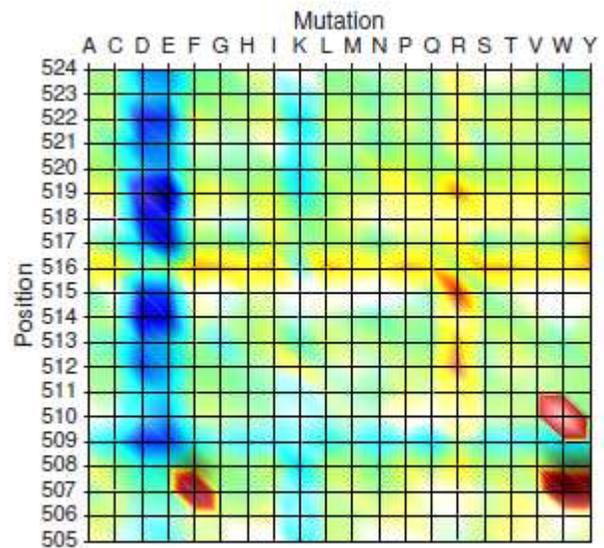
(a) Residue Positions 415-434



(b) Residue Positions 435-463



(c) Residue Positions 464-504



(d) Residue Positions 505-524

Figure 4

2D Projection of the Mutation Landscape Representing Positions 415-524

Supplementary Files

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

- [Covid19supp.pdf](#)