Progress Report

Natália Teruel

September 2020

1 Introduction

Protein engineering is the conception of new polypeptides through chemical modifications for several applications. Synthetic protein structures and functions can be designed by computational approaches, that have been used to identify mutations that change ligand-binding affinity [1], function [2] and stability [3].

Previously, the Najmanovich group has developed the normal mode analysis (NMA) method ENCoM, an Elastic Network Contact Model that employs a potential energy function that includes a pairwise atom-type non-bonded interaction term and thus makes it possible to consider the effect of the specific nature of amino-acids on dynamics. NMA methods can be used to explore protein vibrations around an equilibrium conformation by mean of calculating the eigenvectors and eigenvalues associated to different normal modes. ENCoM performs better than existing NMA methods with respect to traditional applications of NMA methods and was the first to predict the effect of mutations on protein stability and function [4].

Among the many features of proteins that can be useful for specific purposes are its thermal stability and its solubility. Thermal stability predictions based on ENCoM were solely based on predicted vibrational entropy changes (ΔSvib). As good as the accuracy of our predictions were, better other methods [5], our bootstrapped root mean squared error (RMSE) on the prediction of experimental $\Delta\Delta$G were even better when combined with the enthalpy-based predictions of FoldX and Dmutant.

Several mechanisms have been found to affect the accessibility of conformational states that lead to aggregation. Among them, altered surface charge distributions and solvent exposure of hydrophobic residues [6] due to increases in flexibility. Our strategy is to introduce mutations that increase the stability and do not increase its flexibility as measured by the ΔSvib of each residue. We can also assess from generated conformational ensembles if mutations expose hydrophobic patches or produce drastic changes in surface charge distributions. Unlike existing methods for the prediction of solubility, we believe that this represents a first step in designing soluble proteins based on sound mechanistic considerations from studying protein aggregation. Planned applications could include both medical and industrial purposes.

The coronavirus pandemic has emerged as a major and urgent issue for health sciences to overcome. The infection process, as well as the antibodies recognition of the virus and the pharmaceutical advances on vaccines target mostly one of its proteins, the Spike glycoprotein [7, 8, 9]. The entry receptor for SARS-CoV-2 is the human cell-surface protein angiotensin converting enzyme 2 (ACE2), as well as for other lineages of human coronaviruses. It makes the Spike protein an extremely important study subject to understand SARS-CoV evolution [10]. Spike and ACE2 binding rely on Spike to be on its the open conformation, in which its Receptor Binding Domains (RBDs) are open, as shown in Figure 1. This fact lead us to think about the possibility of Spike's dynamics to be relevant to virulence.

2 Objectives

The methodological goals of my project are two-fold. First, we are interested in improving the prediction of thermal stability with the addition of an enthalpy-based component. This will permit us to optimize the code and finetune ENCoM to predict $\Delta\Delta$Gs more accurately, which is particularly important since ENCoM is unique among protein engineering software for its capacity of modulating stability while controlling the effect of mutations on dynamics. Like thermal stability, solubility is another general biophysical property of proteins of importance in any protein engineering project. The second goal of my project is to expand ENCoM for the prediction of solubility. It is now understood that solubility is intimately related to stability. Given the right conditions any protein can be made to aggregate [12] as the this is an additional thermodynamic state accessible to different extents from the native state depending on the protein and the environment. Thus, increasing the stability of the folded state may decrease the probability of aggregation [13].

During the project, a new methodological goal appeared. Our model was constructed based on a calculation for Svib that was based on a rigid-rotator approximation for harmonic oscillator [14]. However, we found a new formula specifically for the Svib calculation [15], and it has some new terms - associated with physical constants - that could help the precision of out results if used as a scale factor. ENCoM is a pseudo-physical universe, so the actual physical values for the variables would not be appropriate to use. Therefore, we would need to adjust the regression work to find the best thermodynamic β (β = h/KT) to use as the scale factor.

Due to the COVID-19 pandemic and the urge for a better understanding of its infection mechanisms, the study of the Spike protein dynamic applying the tools that our group develops became a priority.

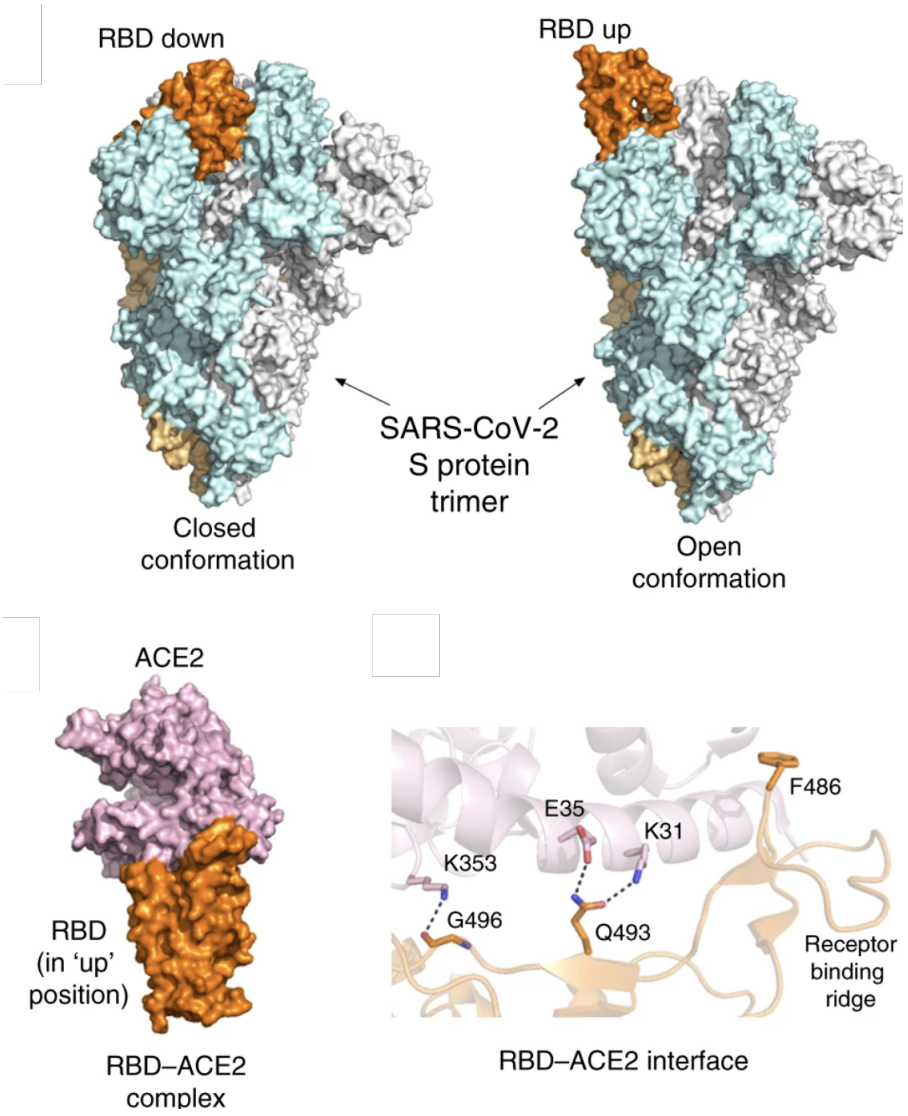


Figure 1: Representation of SARS-CoV-2 Spike protein mechanism, adapted from Zhang, Y., Kutateladze, T.G., 2020 [11]. Side views of the spike protein trimer in a closed conformation (left, PDB 6vxx) and open conformation (right, PDB 6vyb). Three protomers are colored light cyan, gray, and light orange. Buried in the closed state RBD (orange) from one of the protomers (light orange) swings up and is ready to bind ACE2 in the open state. Side view of the RBD-ACE2 complex (PDB 6m0j). Zoom in view of the interface of the RBD-ACE2 complex (PDB 6vw1). Dashed lines indicate salt bridges observed in the SARS-CoV-2 complex that are absent in the corresponding SARS-CoV complex.

3 Method results - ENCoM studies

3.1 Cleaning dataset

The experimental data on the mutations comes partly from the popmusic dataset. Not all the proteins on this dataset were used for different reasons. 1aon was too large to be used on FoldX - the time for repairing was absurd. For other proteins the clean and repair steps have changed the position of the mutations; 2imm, 1rg8 and 1h7m were manually corrected and 1rtp and 3pgk were taken out. Therefore the full dataset used for the regression with FoldX results had 476 mutated proteins.

For the analysis of the regression with MD results, the dataset used was of all the single and double mutations for molecules smaller than 1000 amino acids, with the exception of 1mjcTRP52A.

3.2 Entropy and enthalpy linear regressions

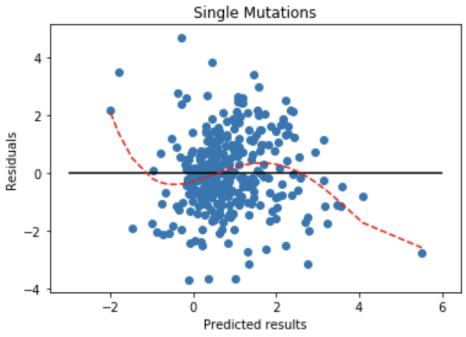
345 samples

coefficients [0.80225739 0.26081602]

Mean Absolute Error: 0.991274366778288

Mean Squared Error: 1.6570306438065676

Root Mean Squared Error: 1.287257023211203



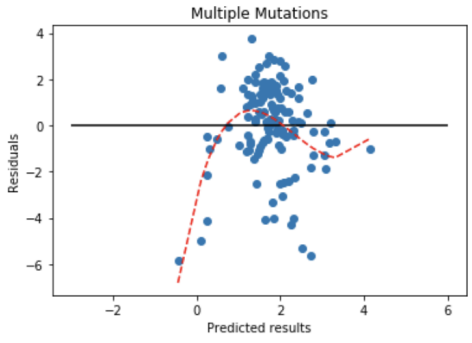
132 samples

coefficients [0.05612055 0.14488931]

Mean Absolute Error: 1.389907821467884

Mean Squared Error: 3.426225308861557

Root Mean Squared Error: 1.8510065664015234



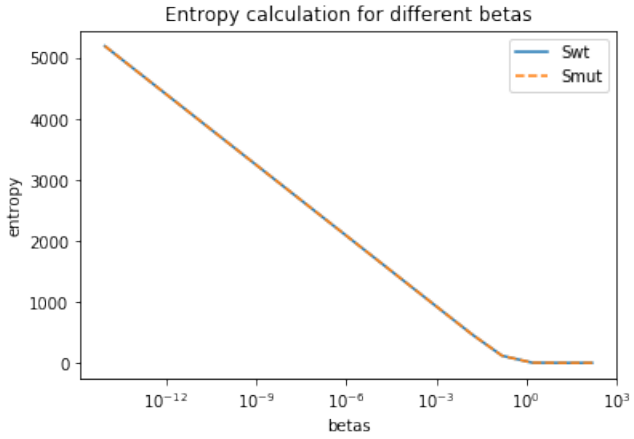
The results for the linear regressions were not very good, with the results for the Multiple Mutations being worse than the ones for Single Mutations, even though the sample was way smaller. It could point out to the fact that even

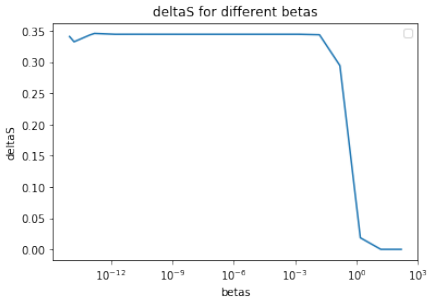
with results combined with the FoldX results and different coefficients, adding more mutations makes our predictions less accurate.

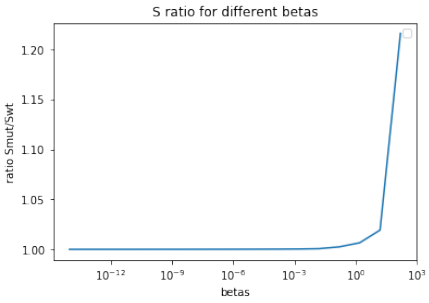
To reach better results and/or to understand the limitations of our method a little bit more, the regressions were redone to subsets of the single mutations dataset, according to the residues it mutated from and the residues it mutated to. The subset analyses gave us smaller errors, specially when dis-considering mutations involving aromatic residues (not shown).

3.3 Defining beta range

To understand which range of betas should be tested, some previous analyses were done using the new formula. For all the beginning of the range showed in the following images, we have the values calculated changing, but not the delta between them. Only for the last ones the values that we have show an increased delta.







It suggests that measuring the errors for the regression for betas over 1 may give us more interesting results. When trying to calculate the entropy for betas above 1, only for betas [1, 10, 100] we had results. For higher betas the entropy calculations go to zero. For smaller values the calculation crashes.

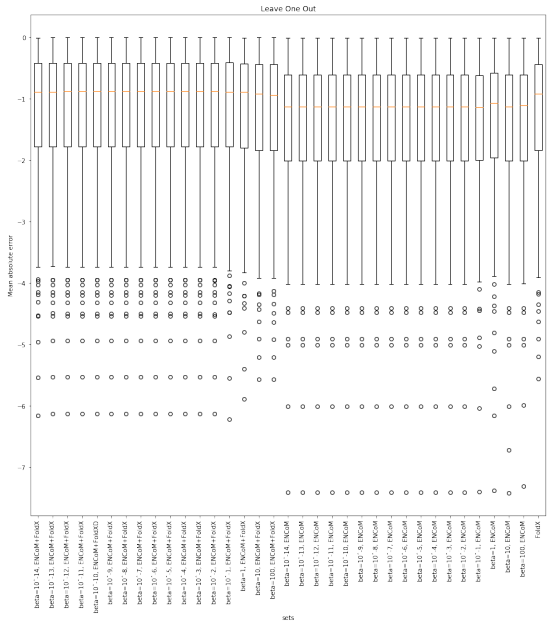


Figure 3: Leave One Out for all betas and possible regressions with FoldX results

3.4 Beta optimization

Due to previous attempts, the range of beta to be tested for ENCoM predictions is from 10^{-14} to 10^{2}.

The cross validation done is Leave One Out (regression and loo from sklearn), and we are comparing the results for the regression of ENCoM+Foldx for all betas, just ENCoM for all betas and just FoldX; and also ENCoM+Molecular Dynamics(MD) for all betas, just ENCoM for all betas and just MD. These were done separately because of the different datasets, as described in the last section.

The mean absolute error was the best between the scoring methods from the cross val package. The error results should be considered for their absolute values.

So here we can see that for the regression with FoldX the best betas are between 10^{-14} and 10^{-2}. For betas above this range the results get closer to the prediction using only FoldX results.

The regression with MD showed worse results, and for the best betas, around 1, the error is barely equal for only the ENCoM prediction and the association with MD mean potential.

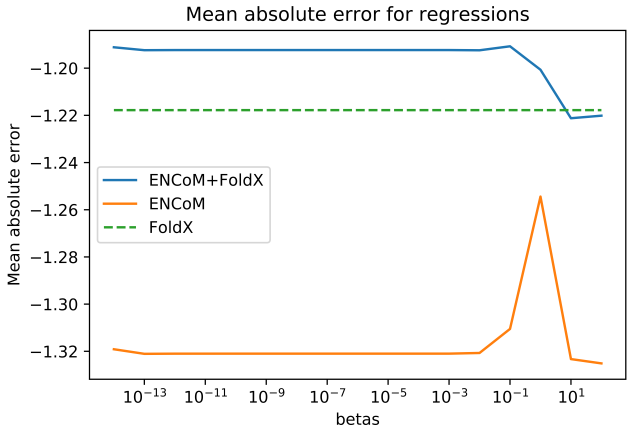


Figure 4: The mean value for all the groups described above in function of beta

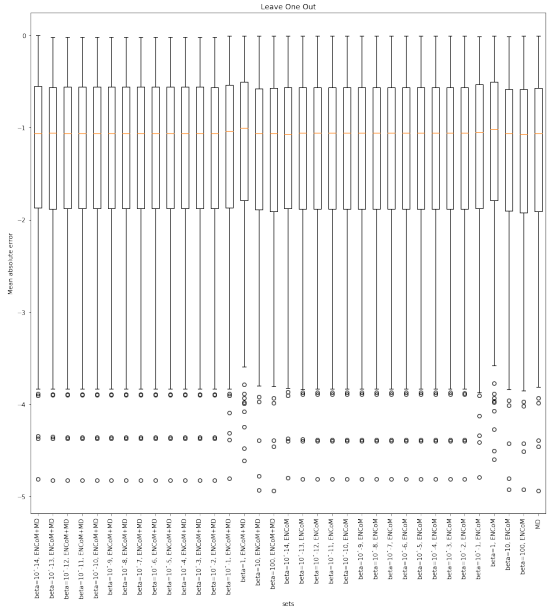


Figure 5: Leave One Out for all betas and possible regressions with MD results

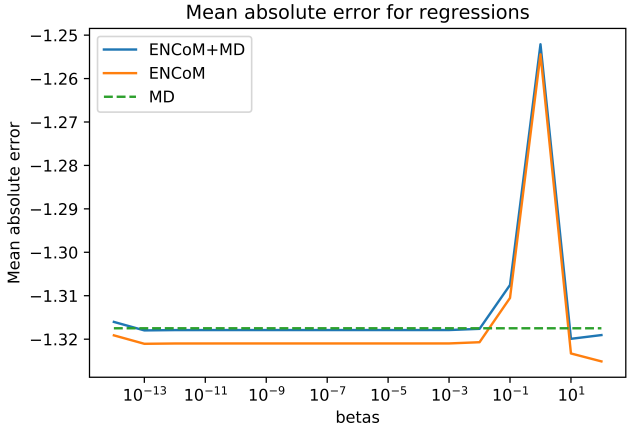


Figure 6: The mean value for all the groups described above in function of beta

4 Application results - Spike dynamics

4.1 Mutants dataset

I have been working considering the 13741 sequences of the protein available in May 08 according to the data from COVID-19 Viral Genome Analysis Pipeline, enabled by data from GISAID [16]. Currently this number is of 75409 sequences (September 09), but I have not checked the mutations associated to the new sequences; I chose not to work with it to maybe be able to use these new mutations as a validation set of a possible pipeline to predict future mutated structures.

Table 1 shows these mutations dis-considering the ones in positions out of the range of the reconstructed Protein Data Base (PDB) structure (positions 5, 8 and 1263). From this table, I have only performed the substitution mutations, and they were done for the full protein and for the 3 chains. The "corrected position" column is about the position of the mutation after structure reconstruction with Modeller.

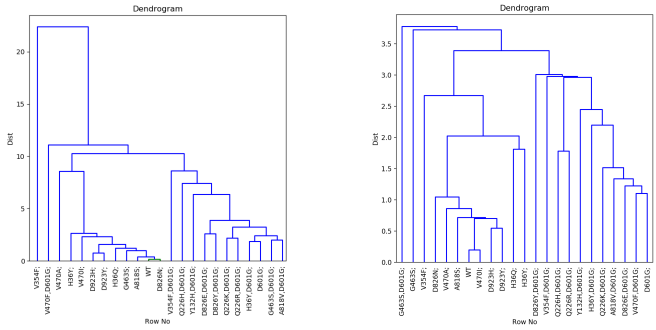
The chosen structures to perform these analyses were 6vxx and 6vyb due to good resolution and to the fact that this is a publication containing both closed and open Spike conformations [17].

4.2 Dynamic Signature analyses

According to the Dynamic Signature clustering, it seems that the effects that the single mutation D614G has are more relevant than any other mutation with the exception of the ones that happen in the binding area, which could explain its increased virulence [18]. Also, this dynamic characteristics seem to be very specific and hard to obtain with random mutations. Actually, clustering 100aa

length pieces of the dynamic signature vectors, we can clearly see that this point mutation is the most relevant for the flexibility characteristics of Spike from around position 250 to around position 750, which includes part of N-Terminal Domain (NTD) and all RBD (not shown).

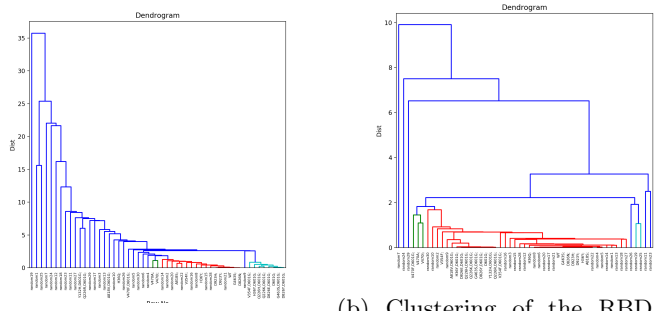
Mutations according to data	Corrected positions
H49Q;	H36Q;
H49X,D614X;	H36X,D601X;
H49X,D614G;	H36X,D601G;
H49Y;	H36Y;
H49Y,D614G;	H36Y,D601G;
Y145H,D614G;	Y132H,D601G;
Q239H,D614G;	Q226H,D601G;
Q239K,D614G;	Q226K,D601G;
Q239R,D614G;	Q226R,D601G;
Q239X,D614G;	Q226X,D601G;
V367F;	V354F;
V367X;	V354X;
V367F,D614G;	V354F,D601G;
G476S;	G463S;
G476S,D614G;	G463S,D601G;
V483A;	V470A;
V483F,D614G;	V470F,D601G;
V483I;	V470I;
V483X,D614G;	V470X,D601G;
A831S;	A818S;
A831V,D614G;	A818V,D601G;
D839E,D614G;	D826E,D601G;
D839N;	D826N;
D839X;	D826X;
D839X,D614X;	D826X,D601X;
D839Y,D614G;	D826Y,D601G;
D614G;	D601G;
D936H;	D923H;
D936X;	D923X;
D936Y;	D923Y;



(a) Clustering of closed structure mutants

(b) Clustering of open structure mutants

Figure 7: Dynamic signature clustering of the WT and 22 mutants. The results were done with 6vxx and 6vyb PDB files, and the structures were reconstructed and optimized with Modeller and mutated with FoldX



(a) Clustering with random generated mutants

(b) Clustering of the RBD dynamic signature with random generated mutants

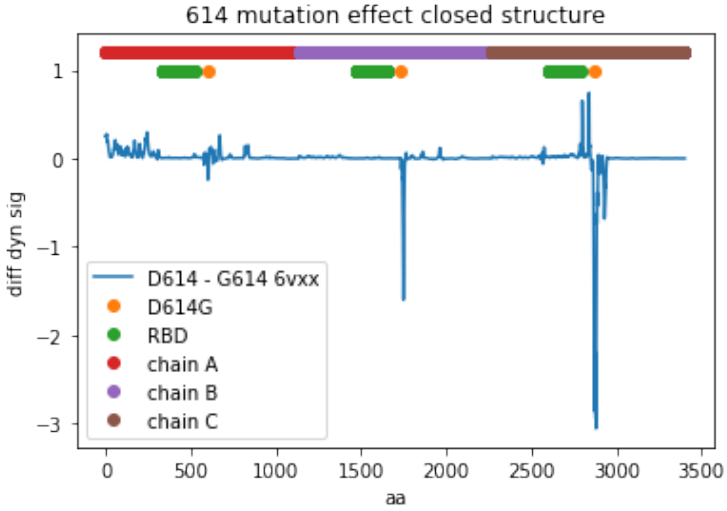
Figure 8: Dynamic signature clustering of the WT, 22 mutants from the database and 30 random generated mutants. The results were done with 6vxx PDB file, chain B, the structure was reconstructed and optimized with Modeller and mutated with FoldX

4.3 D614G mutation

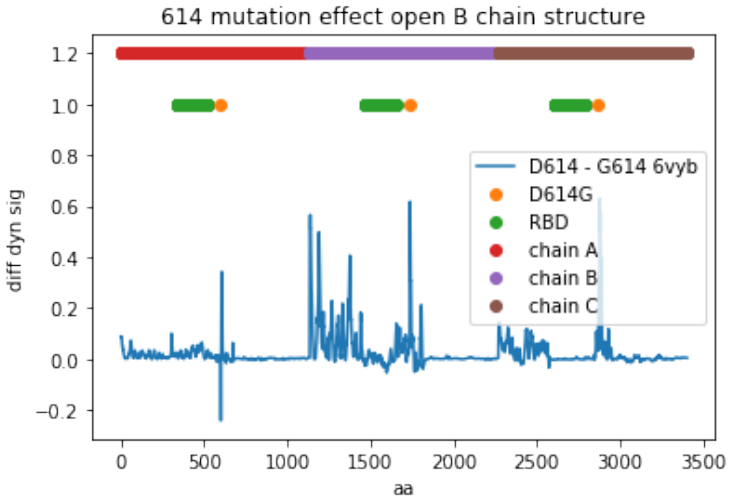
When checking the difference between WT D614 and mutant G614 dynamic signatures we can see that for the closed conformation the pattern tends to negative, indicating that this mutation makes the closed state more flexible specially in the position of the mutation. For the open B chain conformation the pattern is very much positive for the open RBD, the same chain NTD and the adjacent chain NTD also, indicating that this mutation makes these areas of the open conformation more rigid. This specific result will be the main base

for the hypothesis created with the following results.

Mutating the position 614 to every other residue, we see that we cannot not have the the effects on the open structure with none of them expect from Glutamine, which is similar to Aspartate. But we can see that some other residues have a similar effect as Glycine, such as Proline and Treonine.



(a) D614G mutation effect closed structure



(b) D614G mutation effect open structure

Figure 9: Dynamic effects of the D614G mutation on each residue of the protein.

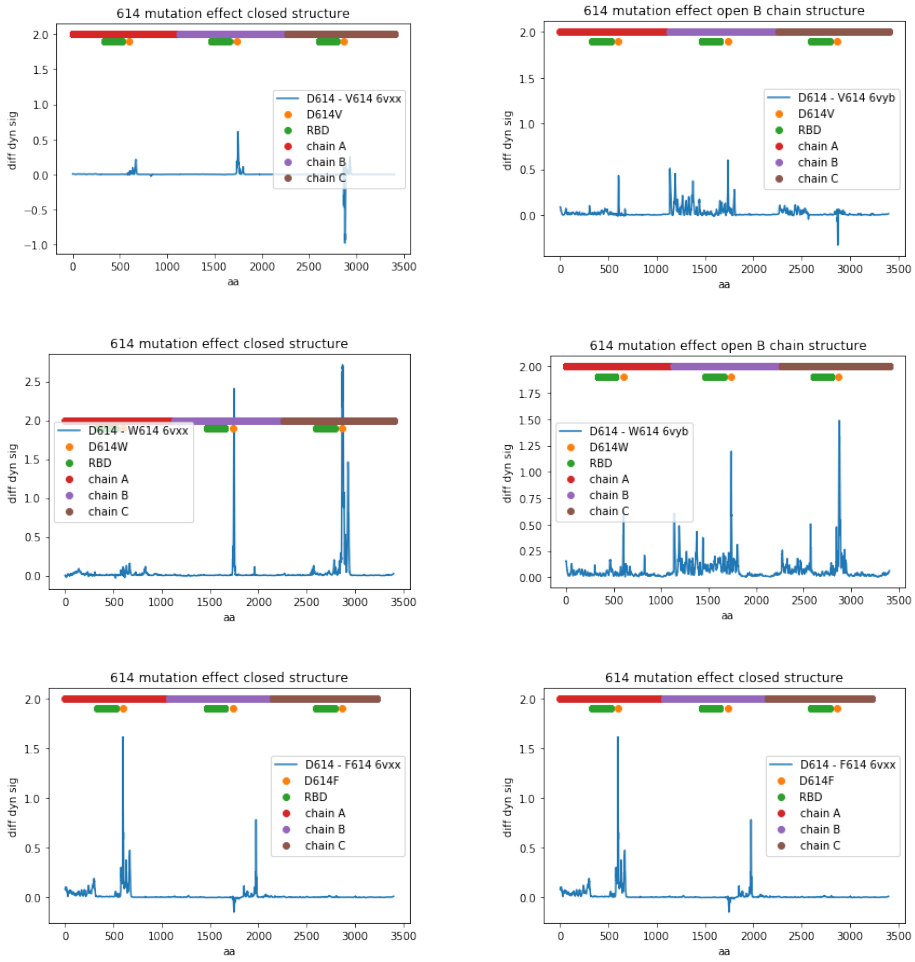


Figure 10: Effects of mutations on 614, part I.

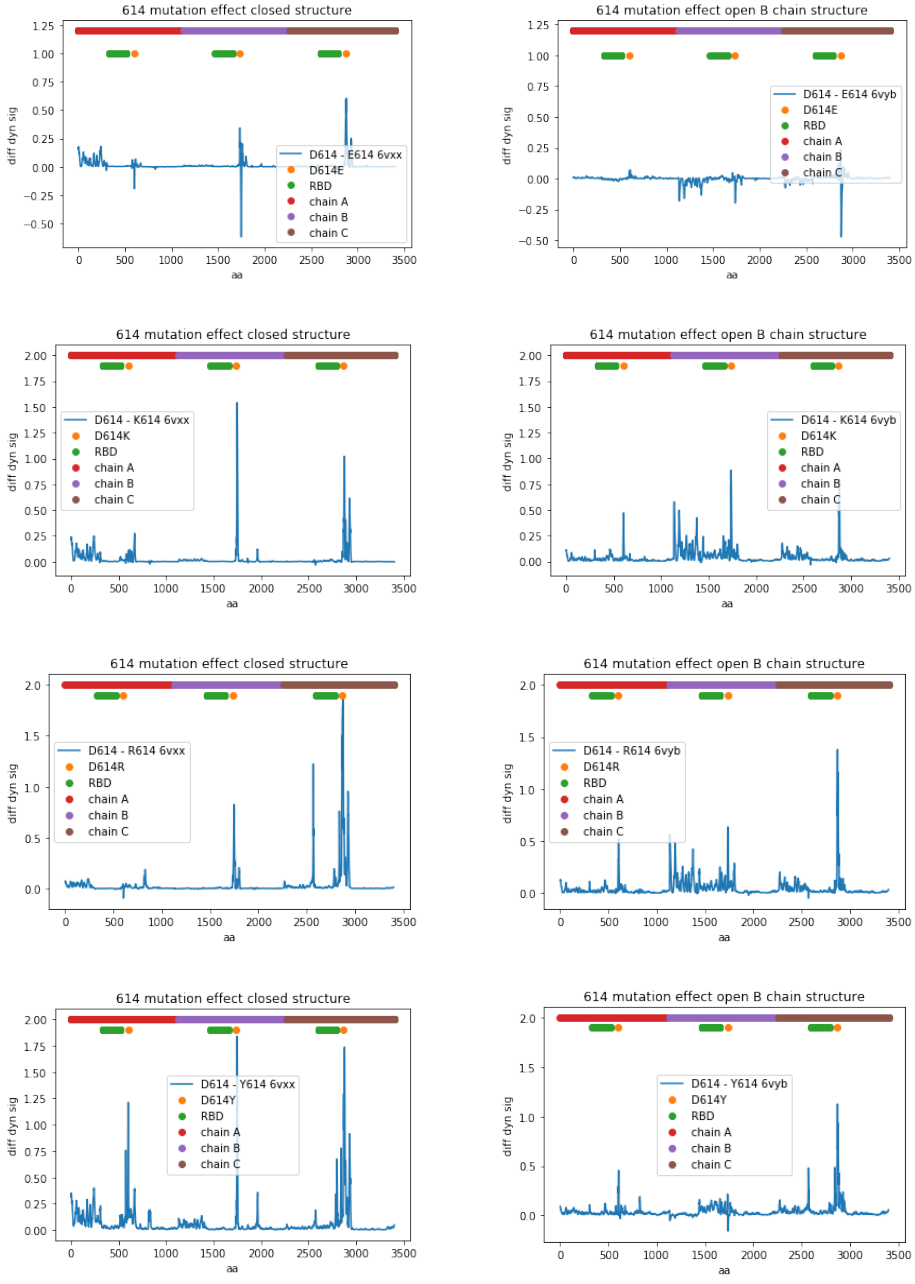


Figure 11: Effects of mutations on 614, part II.

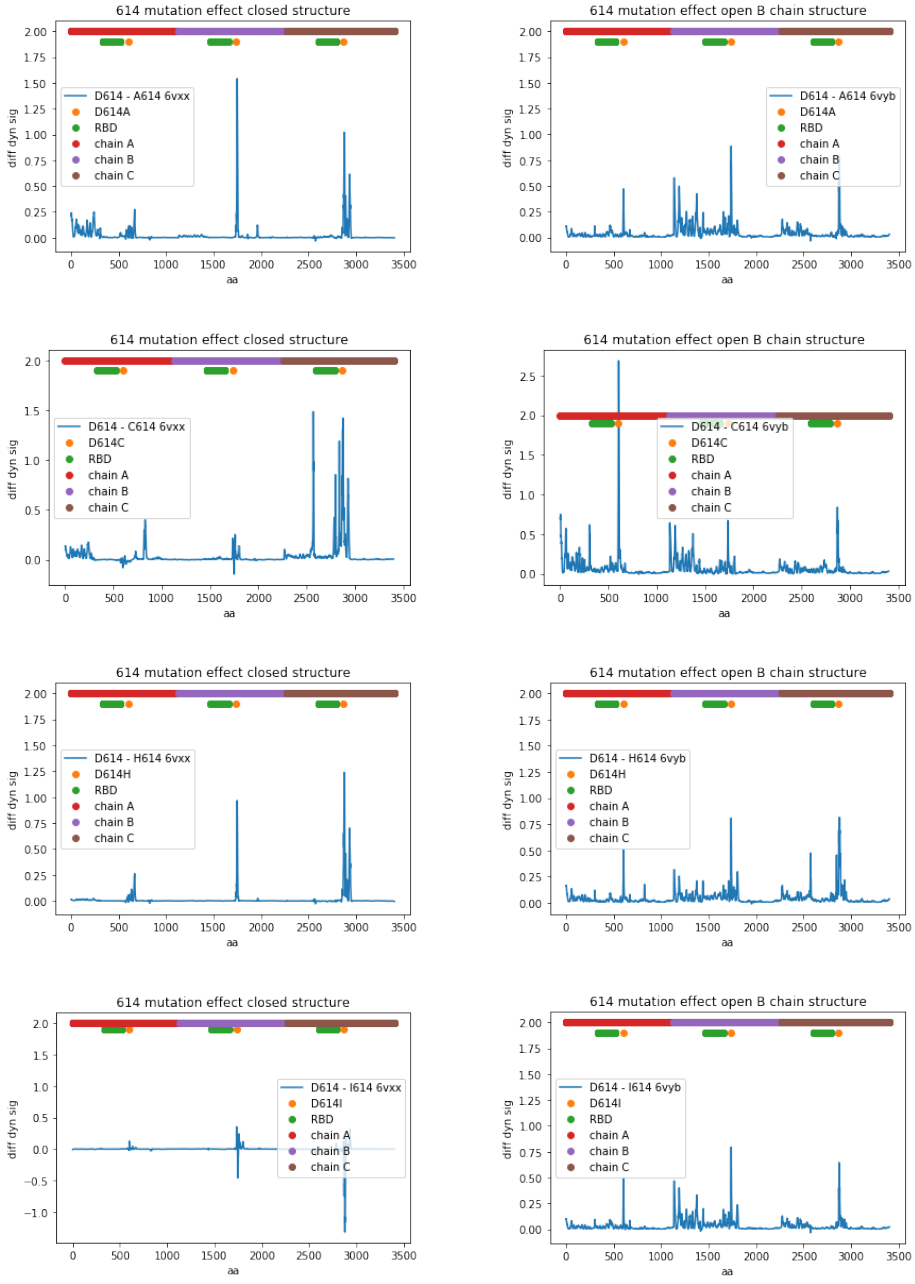


Figure 12: Effects of mutations on 614, part III.

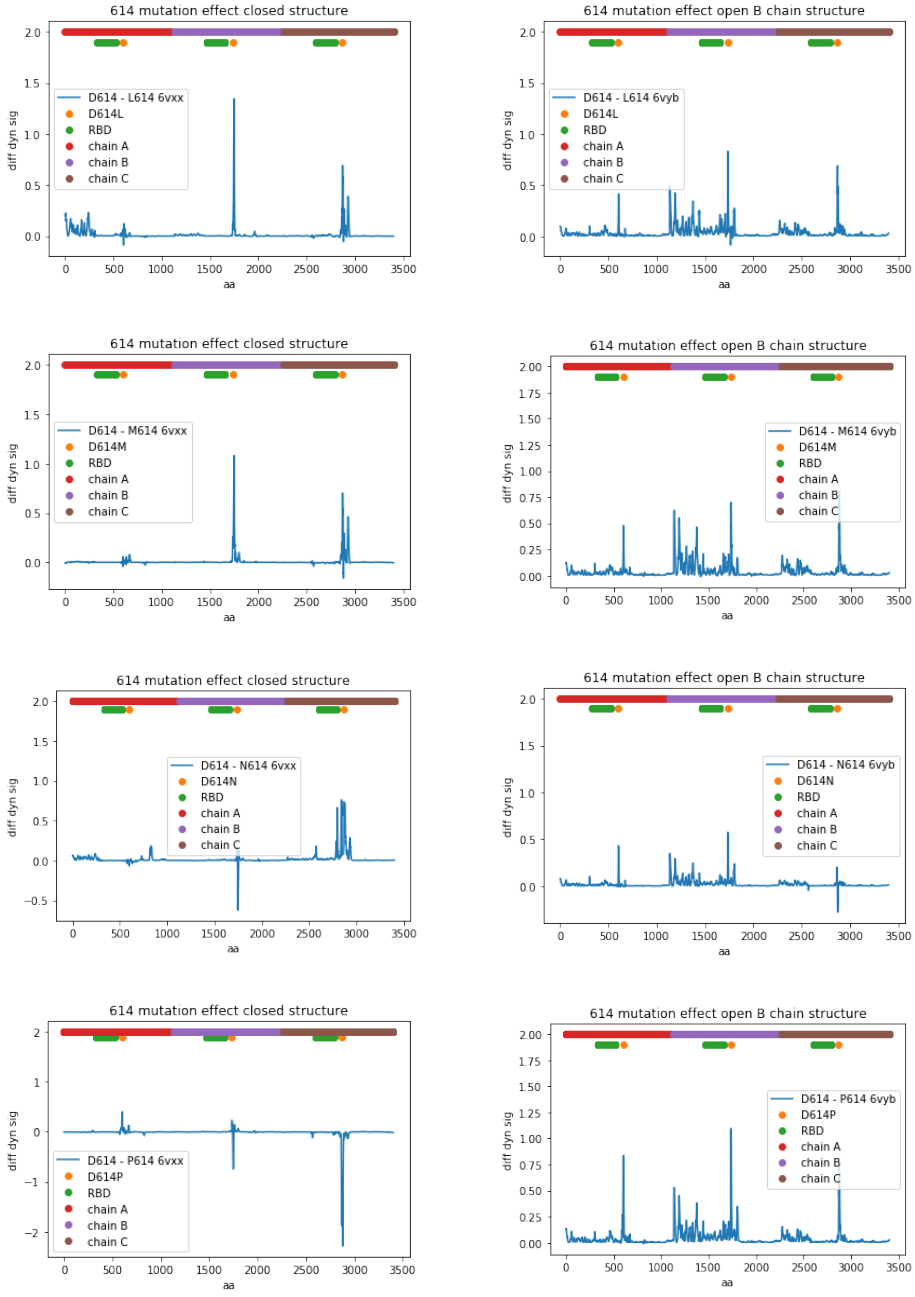


Figure 13: Effects of mutations on 614, part IV.

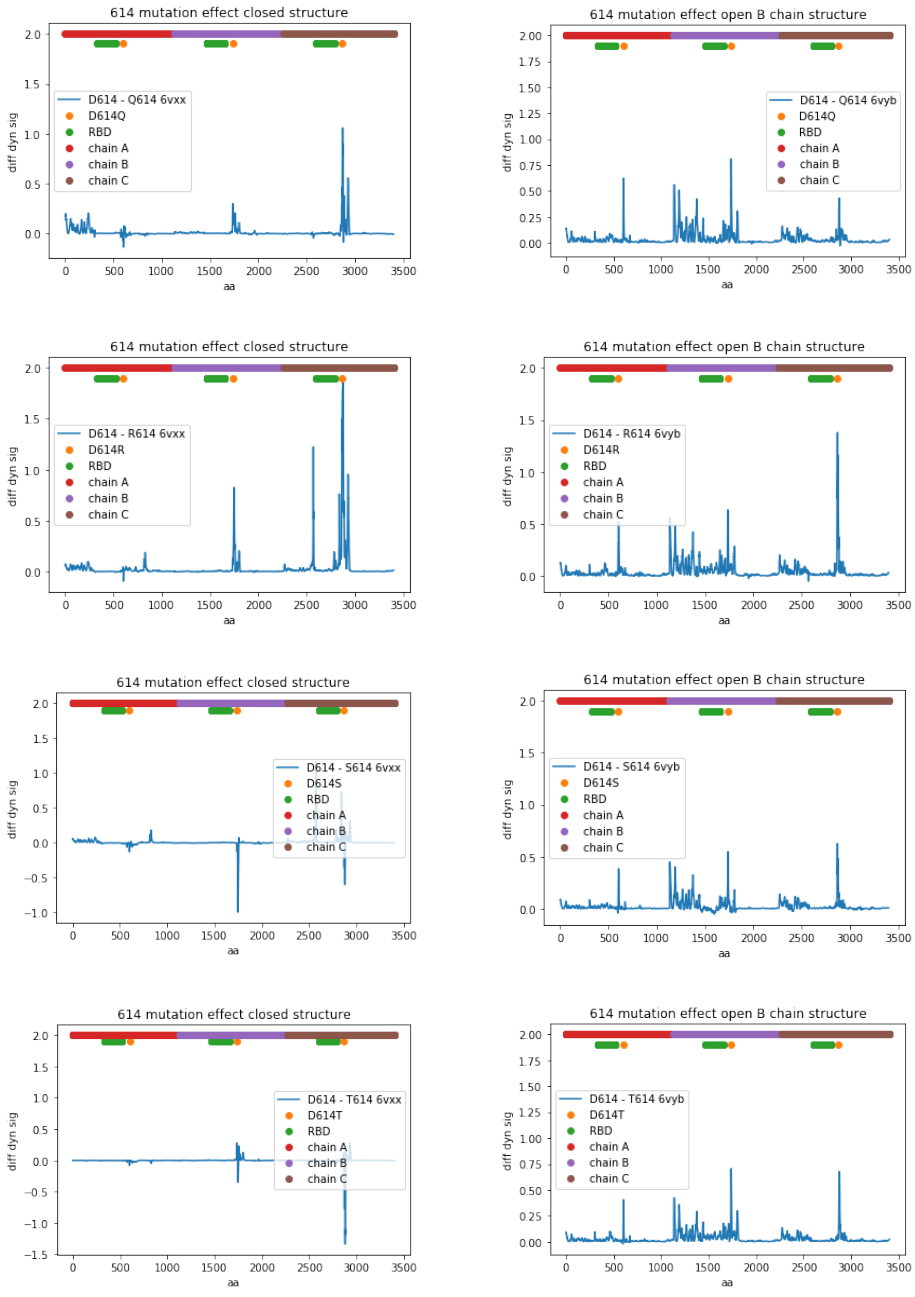


Figure 14: Effects of mutations on 614, part V.

4.4 SARS-CoV and SARS-CoV-2

The comparison was also done with structures from the "old corona virus" SARS-CoV Spike protein. The 5x58 and 5x5b PDB files [19] were chosen for this comparison because among all available possible open and closed pairs of structures from SARS-CoV Spike, these were the ones with best alignment scores to the structures that we were using for the SARS-CoV-2 Spike.

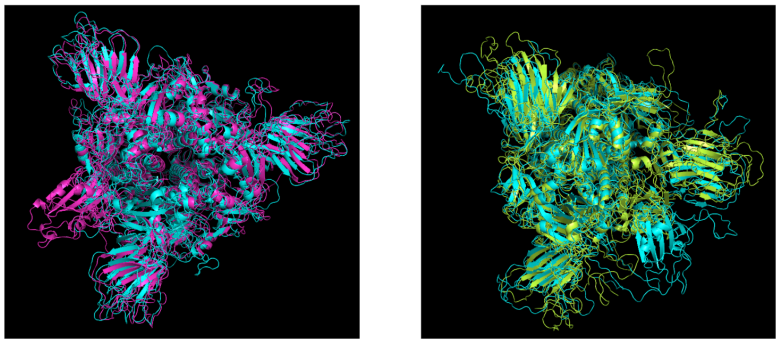
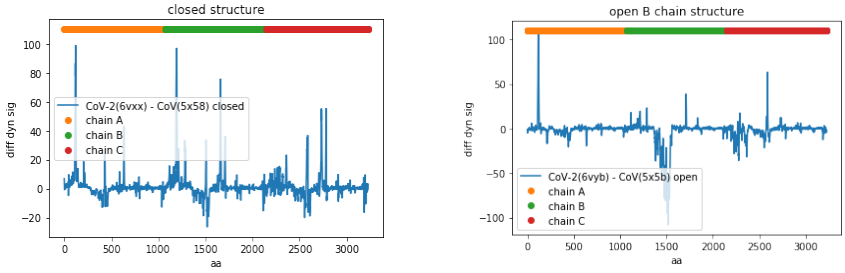


Figure 15: Representative image of the chosen pairs of structures for SARS-CoV Spike and SARS-CoV-2 Spike

They were also reconstructed and optimized with Modeller, and had their dynamic signature calculated. Only the aligned residues were considered, adjusting the open chains to superpose (chain A from 5x5b and chain B from 6vyb), and the difference between the dynamic signatures was calculated.

Considering the results we had for the D614 and G614 comparison and considering that the SARS-CoV was less virulent than the SARS-CoV-2, according to our hypothesis, the results for CoV-2 - CoV would have the opposite pattern from the previous D614 - G614, which happened, as can be seen in Figure 16.



(a) Closed structure comparison

(b) Open structure comparison

Figure 16: Dynamic signature comparison. The results were done with 5x58, 6vxx and 5x5b, 6vyb PDB files, and the structures were reconstructed and optimized with Modeller and aligned with Blast

4.5 New mutants generation

4.5.1 Vibrational entropy analysis

Comparing the order that I would put the mutations at 614 for every residue from less to more virulent according to my hypothesis, the delta Svib calculated with ENCoM follows almost the same order. It could help with the analysis of the protein with several possible mutations since it is a little less computationally demanding and an easier way to interpret results.

Figure 17: Closed structure dSvib heat map

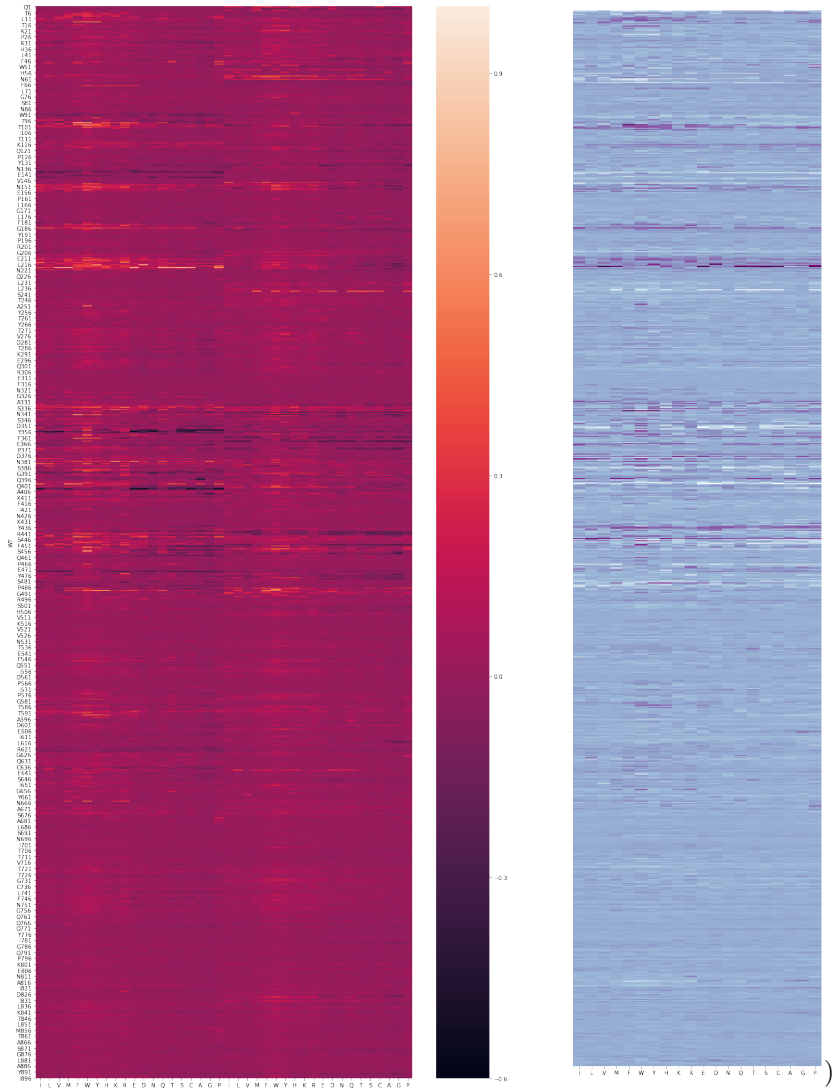
So according to the hypothesis that a less virulent mutant would be less flexible than WT when closed and more flexible than WT when open, since it would favor open conformation occupancy and favor binding and infection, we should look for mutants with lighter representation for the closed structure (positive WT_{Svib} - $Mutant_{Svib}$) and darker for the open structure (negative WT_{Svib} - $Mutant_{Svib}$)

- I decided to work with position 1-900 (according to the reconstructed structures) since the end of the glycoprotein is its core and it is highly conserved;
- All mutants for these positions were done with FoldX for both open and closed states (34200 mutants);

We can see interesting patterns according to the positions of the mutants. Remembering the presented hypothesis, we would be searching for opposite patterns for closed and open states, where

– Less virulent mutants would be lighter for the closed state and darker for the open state; darker in the difference matrix;

– More virulent mutants would be darker for the closed state and lighter for the open state; lighter in the difference matrix.



(a) Left half for closed structure, right half for open structure.

(b) Difference between open and closed.

Figure 18: Heat map for positions 14 to 913 when with corrected positions.

The region of mutations around positions 403 and 404 would be possible examples of more virulent mutants. The region of mutations between positions 452 and 458 would be possible examples of less virulent mutants, as well as between positions 224 and 234, in the NTD.

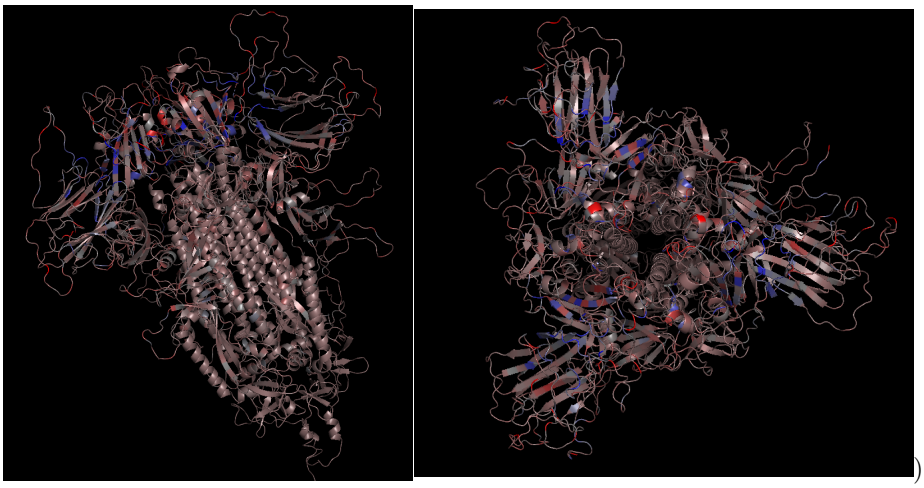


Figure 19: Spike structure colored according to the results, with red representing positions where mutations would make the mutant more virulent, and blue representing positions where mutations would make the mutant less virulent. This was made with median values for each position.

4.5.2 Mutation ranking

If sorting the mutants according to the difference between open and closed states (Figure 18b), we can find the top candidates for more virulent and less virulent mutants.

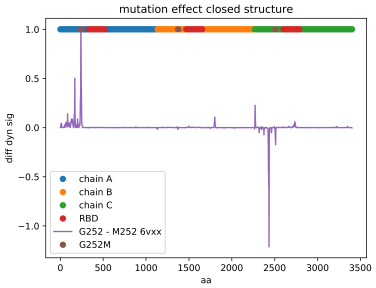
The positions pointed out in these selections are specially associated to some specific areas of the structure, that seem to be located in the interface between different RBDs and between RBDs and NTDs. This observation can be linked with other published studies about the dynamic relation between the RBD and the adjacent NTD [20].

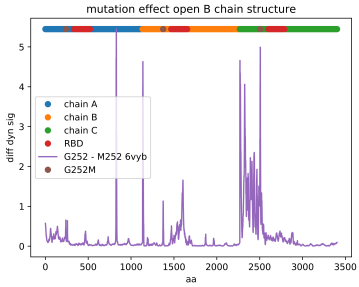
The dynamic signatures were calculated for these 100 mutants (50 top values, 50 bottom values), and the deltas for the dynamic signature were observed as previously described. Here we are looking specially for patterns that are similar to the one found to D614 - G614 when it comes to more virulent candidates and patterns more similar do the CoV-2 - CoV when it comes to the less virulent ones. Not all of these 100 mutants give us these patterns, but several of them do.

In Figures 20 and 21 we can see some of these candidates. We can see clearly in the profile differences that these mutations affect the flexibility of either RBD or NTD or both.

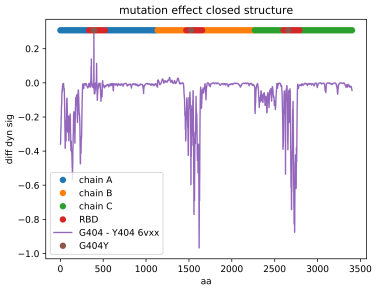
Top virulent mutants	Reconstructed positions	Corrected positions
1	K404D	K417D
2	K404P	K417P
3	Y356E	Y369E
4	K404C	K417C
5	K404E	K417E
6	G239M	G252M
7	K404I	K417I
8	Y408G	Y421G
9	G239S	G252S
10	G239T	G252T
11	D454W	D467W
12	Y476M	Y489M
13	G391Y	G404Y
14	G391N	G404N
15	G59W	G72W
16	Q396A	Q409A
17	G403Y	G416Y
18	G239P	G252P
19	G239C	G252C
20	D454Y	D467Y
21	S148F	S161F
22	G239Q	G252Q
23	S148I	S161I
24	E452Y	E465Y
25	G239D	G252D
26	G391W	G404W
27	G239W	G252W
28	L355A	L368A
29	Y356V	Y369V
30	R21Y	R34Y
31	Y356N	Y369N
32	T60F	T73F
33	P478H	P491H
34	N488W	N501W
35	G239E	G252E
36	E452W	E465W
37	E452A	E465A
38	L355C	L368C
39	G239H	G252H
40	I455T	I468T
41	Q1F	Q14F
42	G491I	G504I
43	Q1H	Q14H
44	G391H	G404H
45	L355N	L368N
46	L355S	L368S
47	E452S	E465S
48	L355P	L368P
49	R390N	R403N
50	G403S	G416S

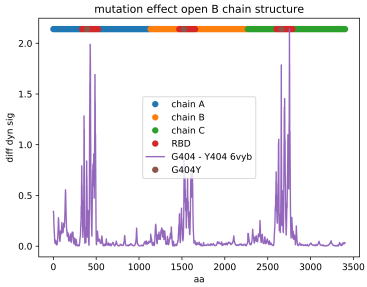
Least virulent mutants	Reconstructed positions	Corrected positions
1	G219V	G232V
2	G219E	G232E
3	R342F	R355F
4	G219Q	G232Q
5	G219S	G232S
6	P217D	P230D
7	G219C	G232C
8	D98F	D111F
9	G219P	G232P
10	R342Y	R355Y
11	G219M	G232M
12	G219T	G232T
13	F451L	F464L
14	I218P	I231P
15	D98W	D111W
16	P217Y	P230Y
17	R342W	R355W
18	N381K	N394K
19	S456W	S469W
20	G400M	G413M
21	L442W	L455W
22	K100W	K113W
23	A359W	A372W
24	L442M	L455M
25	F451V	F464V
26	N152R	N165R
27	G400T	G413T
28	F451H	F464H
29	G400F	G413F
30	G400P	G413P
31	G400I	G413I
32	G489H	G502H
33	N381E	N394E
34	Y383I	Y396I
35	P371I	P384I
36	R453W	R466W
37	A251W	A264W
38	G219I	G232I
39	R441G	R454G
40	F334I	F347I
41	K100Y	K113Y
42	G489Y	G502Y
43	N381G	N394G
44	A14W	A27W
45	D215R	D228R
46	Y383W	Y396W
47	G219H	G232H
48	D98R	D111R
49	G400S	G413S;
50	S362F	S375F





(a) 6th more virulent candidate





(b) 13th more virulent candidate

Figure 20: Two good examples of more virulent candidates.

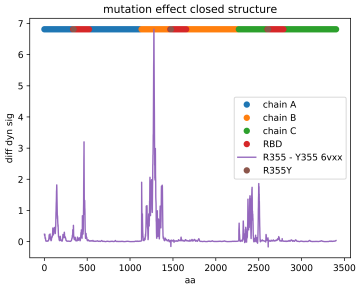
5 Future steps

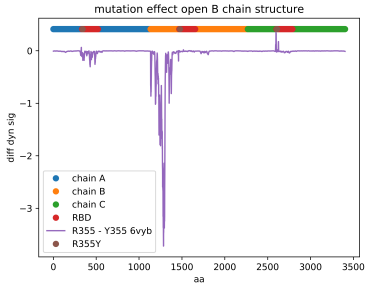
5.1 Following computational analyses

The immediate next step is to develop a script to perform a scoring function for the desired profiles for dynamic signature differences. Several of the top and bottom candidates follow the expected profile according to what was observed for SARS-CoV and for the D614G lineage, but not all of them. An automated function to evaluate it would be interesting for expanding the model back to big data analysis to select the best candidates for an experimental validation of the model.

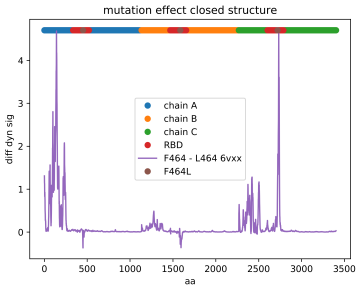
I'm also dedicating myself to look for the specific residue-residue interactions that explain the dynamic patterns that we can observe for some mutants. This type of analysis is time consuming due to the necessity of looking into each mutated structure in particular and searching for observable patterns.

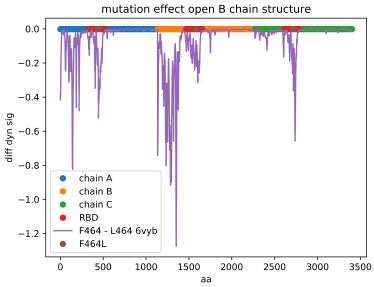
One extra computational analysis that I wish to do to test some of the hypotheses is Molecular Dynamics simulations. This is a technique extremely computationally consuming since it calculates physically the movement of each atom in each dt. However, to do it for some selected mutants could be useful to test the hypothesis that different rigidity or flexibility of the closed and open





(a) 10th less virulent candidate





(b) 13th less virulent candidate

Figure 21: Two good examples of less virulent candidates.

conformations affect the occupancy of these two states. It could also make the computational point of view richer and more complete by considering the role of the membrane and the sugars that are a part of the biological assembly of the protein.

5.2 Experimental validation

An extremely important step for the project is an experimental validation for this virulence model. To do so we count on collaborations with experimental laboratories. However, the experimental design for validation is something to take into consideration during the mutants selection.

The first aspect that was taken into consideration was the ethical problems associated with making gain of function mutations into viruses. For this reason specifically we are not only working with more virulent candidates, that would be the main focus considering predicting future dangerous lineages, but also less virulent candidates.

Techniques that measure both binding and viral RNA into host cells could be used for validation. The alternatives for the more virulent type of experimental testing would involve cloning the mutated protein, but without the transfer of viral RNA. There are some published experimental data, including all mutations

for all positions - like the ones performed in the presented simulations - for the RBD region [21], but this type of published data is about cloning only the RBD into generic membrane proteins of yeasts, and therefore it can only give information about the binding with ACE2 and none about the dynamic mechanism of Spike conformations.

5.3 Extension of virology models

If better developed and validated, the intention is to build an online server with these results to monitor new lineages based on each mutation epidemiological risk. It could help with the public health decisions based on local sequencing results.

Exploring dynamic analyses for other viral glycoproteins is also a part of the main goal. The lorger-term idea is to work with Hemagglutinin dynamics due to Influeza high mutation rates and pandemic potential and try to find the same idea of dynamic patterns, considering also the pH-related residues interactions. Working with arboviruses, such as Dengue, Zika and Chikungunya is also a distant but important future goal. From what I could find on computational approaches to viruses studies, this idea of dynamic analysis seems to be very much coherent with ongoing scientific discussions and especially feasible for large datasets. To be able to apply it to tropical diseases that are endemic for years (decades in the case of Dengue) would mean a lot for several scientific and personal reasons.

References

[1] Tinberg CE, Khare SD, Dou J, Doyle L, Nelson JW, et al. Nature. 501(7466):212–16, 2013.

[2] Jiang L, Althoff EA, Clemente FR, Doyle L, Rothlisberger D, et al. Science. 319(5868):1387–91, 2008.

[3] Zhang S-B, Wu Z-L. Bioresource Technology. 102(2):2093–96, 2011.

[4] Frappier V, Najmanovich RJ. PLoS Comput Biol. 10(4):e1003569, 2014.

[5] Frappier V, Chartier M, Najmanovich RJ. Nucleic Acids Res. 43(W1):W395–400, 2015.

[6] De Simone A, Dhulesia A, Soldi G, Vendruscolo M, Hsu S-TD, et al. Proc Natl Acad Sci USA. 108(52):21057–62, 2011.

[7] Pinto, D., Park, Y.-J., Beltramello, M., Walls, A.C., Tortorici, M.A., Bianchi, S., Jaconi, S., Culap, K., Zatta, F., De Marco, A., et al. Nature 583, 290–295, 2020.

[8] Rogers, T.F., Zhao, F., Huang, D., Beutler, N., Burns, A., He, W.-T., Limbo, O., Smith, C., Song, G., Woehl, J., et al. Science, 15 June 2020.

[9] Cao, Y., Su, B., Guo, X., Sun, W., Deng, Y., Bao, L., Zhu, Q., Zhang, X., Zheng, Y., Geng, C., et al. Cell 182, 73–84, 2020.

[10] Letko, M., Marzi, A., and Munster, V. Nat Microbiol 5, 562–569, 2020.

[11] Zhang, Y., Kutateladze, T.G. Nat Commun 11, 2920, 2020.

[12] Baldwin AJ, Knowles TPJ, Tartaglia GG, Fitzpatrick AW, Devlin GL, et al. J Am Chem Soc. 133(36):14160–63, 2011.

[13] Gsponer J, Vendruscolo M. Protein Pept. Lett. 13(3):287–93, 2006.

[14] Martin Karplus and Joseph N. Kushick. Macromolecules 14 (2), 325-332, 1981.

[15] Ma, B., Tsai, C. J., Nussinov, R. Biophysical journal, 79(5), 2739–2753, 2000.

[16] Elbe, S., and Buckland-Merrett, G. Global Challenges, 1:33-46, 2017.

[17] Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Cell. 181(2):281-292.e6, 2020.

[18] Korber, B., Fischer, WM., Gnanakaran, S., Celia, CL., Saphire, EO., Montefiori, DC. et al. Cell 182, 812–827, 2020.

[19] Yuan, Y., Cao, D., Zhang, Y. et al. Nat Commun 8, 15092, 2017.

[20] Melero R, Sorzano COS, Foster B, et al. Preprint. bioRxiv. 2020.07.08.191072, 2020.

[21] Starr TN, Greaney AJ, Hilton SK, et al. bioRxiv. 2020.06.17.157982. 2020.