

ENERGY ANALYSIS

EXERCISE 2023-24

MARÍA LÓPEZ
OSCAR CONTRERAS
STEPHANIE PADILLA
JANA MORENO

INDEX

1. Case Study.....	2
2. Strategy.....	2
3. Methodology.....	2
a. Preparation.....	2
b. Step 1: Interface Residues.....	3
c. Step 2: Interaction Energy.....	3
d. Step 3: Ala-Scanning Experiment.....	5
e. Step 4: PyMOL Images.....	8
4. Conclusion.....	13

1. Case Study

In this case study, our focus is on the interaction between the Receptor Binding Domain (RBD) of the SARS-CoV-2 Spike protein and its receptor, Angiotensin Converting Enzyme (ACE2). The formation of this complex is a crucial step in the viral infection process, and it serves as a primary target for many vaccines available in the market, trying to block this binding. Our detailed analysis will be done on the structure 6m0j sourced from PDB, which contains only the RBD domain from Spike protein.

This study aims to analyze and quantify the influence of specific interface residues on the interaction energy within the RBD-ACE2 protein-protein complex. This type of analysis is essential for understanding molecular details of interactions.

2. Strategy

The project aims to assess the contribution of individual amino acid residues to the stability of the complex. The strategy involves identifying interface residues using the 6m0j structure and conducting Ala-scanning experiments to replace each residue with alanine, evaluating changes in interaction energy. Additionally, known SARS-CoV-2 Spike variants will be considered, and their effects on ACE2 binding will be assessed by replacing the corresponding residues and reevaluating interaction energy. Optionally, the analysis will be repeated using FoldX for validation, but we didn't do it. The structural and biological implications of these substitutions will be interpreted, contributing insights into the functional consequences of individual amino acid variations in the RBD-ACE2 complex.

3. Methodology

a. Preparation

As a first step, we have to clean the data. In order to do this, we have to take different factors into account, such as the fact that this structure consists of two chains (A and E), the first being larger than the second.

As a next step, we need to remove all heteroatoms, which are atoms (small molecules as ions) in a macromolecular structure that are not considered part of the biopolymer.

Finally, we use the biobb_structure_checking module to perform a quality check on the structures, and add the missing side chains, hydrogen atoms, and atom charges.

Having already cleaned up the entire structure, we will get a file called *6m0j_fixed.pdb*.

Incident Report:

Issue: Challenges in Obtaining and Preparing the PDB Structure

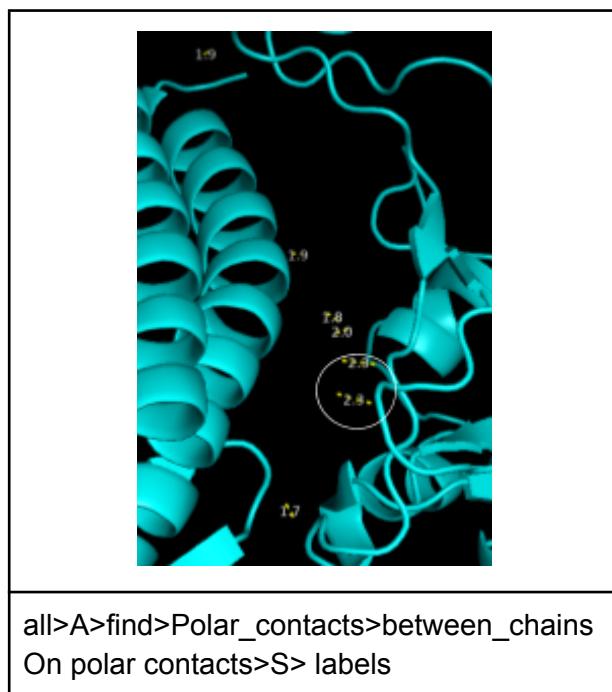
Our project encountered hurdles while using BioPython to manage the Protein Data Bank (PDB) structure for the Receptor Binding Domain (RBD) of the SARS-CoV-2 Spike protein. Specifically, attempts to add or append elements to the Model() or Chain() functions within BioPython faced limitations, impeding the essential structure preparation process.

Solution: To overcome this obstacle, we opted for an alternative approach, incorporating modified code obtained from a GitHub repository. This external code, designed for refining protein structures, offered the flexibility required for efficient data manipulation. The integration process involved adapting the GitHub code to seamlessly align with our existing BioPython workflow, ensuring a smooth execution and validating the final structure's suitability for subsequent analysis.

Improvements: For future initiatives involving PDB structure acquisition and preparation, proposed enhancements encompass refining BioPython documentation to provide clearer guidance on structure manipulation. Additionally, fostering the creation of community support forums, standardized integration tools, comprehensive biological data repositories, and user-friendly interfaces could collectively enhance the efficiency of these processes.

b. Step 1: Interface Residues

The purpose of step 1 was to determine which residues on both chains have at least one atom below a given distance. We begin by inspecting the structure with Pymol and choosing the most appropriate distance. We started with the largest value, which in our case is 2.3Å. After adding 2Å, to take into account the adjacent residues, the sum is 4.3Å ($2.3 + 2 = 4.3\text{Å}$). This resulting value will be used in the following Python script.



The next step is to create a Python script that defines the interface residues on each chain. We will obtain two sorted lists, the first one being the interface residues for Chain A, and the second being the interface residues for Chain E.

```
Interface residues for Chain A: ['19', '24', '27', '28', '30', '31', '330', '34', '35', '353', '354', '355', '357', '37', '38', '393', '41', '417', '42', '445', '446', '447', '449', '45', '453', '455', '456', '473', '475', '476', '484', '486', '487', '489', '490', '493', '496', '498', '500', '501', '502', '505', '79', '82', '83']  
Interface residues for Chain E: ['19', '24', '27', '28', '30', '31', '330', '34', '35', '353', '354', '355', '357', '37', '38', '393', '41', '417', '42', '445', '446', '447', '449', '45', '453', '455', '456', '473', '475', '476', '484', '486', '487', '489', '490', '493', '496', '498', '500', '501', '502', '505', '79', '82', '83']
```

c. Step 2: Interaction Energy

The methodology for evaluating interaction energy between protein chains involves considering the bound (complex) and unbound (isolated chains) states. Two key approximations are made: assuming the 3D structure remains consistent between states,

neglecting changes in bonded and non-bonded terms within the same chain, and considering solvation energies for all atom types based on Accessible Surface Area (ASA) values. The interaction energy is calculated using electrostatic and van der Waals energies, as well as solvation energies in the bound and unbound states for each residue. The total interaction energy is obtained by summing these values using a Python script (step 2 in the final script). The selection of interface residues is crucial, and adjustments may be made based on the captured energy. The methodology offers a simplified yet practical approach, requiring validation against experimental or computational data for robustness.

For this exploration, we have had to find each parameter in order to solve this equation. We are interested in finding the difference in energies in between a complex binding between two proteins.

$$\Delta G^{A-B} = \Delta G_{\text{elect}}^{A-B} + \Delta G_{\text{vdw}}^{A-B} + \Delta G_{\text{Solv}}^{A-B} - \Delta G_{\text{Solv}}^A - \Delta G_{\text{Solv}}^B$$

We evaluate the variance of energy in the interface by iterating through the structure how it varies the energy between the two binding proteins. The total difference in electrostatic energies, Van der Waals and the solvation of the complex minus the solvation of each individual protein.

```
WARNING: atom not found in library ( GLY:OXT )
WARNING: atom not found in library ( GLY:OXT )
Interaction energy for all residues
Total Elec Int.      :   -5.2649
Total Vdw Int.       :  -173.0042
Total Solv AB        : -516.5464
Total Solv           A: -420.4202
Total Solv           E: -102.2662
DGintAB-A-B         : -172.1290
\R -5.2649 -173.0042 -516.5464 -420.4202 -102.2662 -172.1290
Interaction energy based in interface residues only
Total Elec Int.      :    4.5784
Total Vdw Int.       : -137.6728
Total Solv AB        :   -1.4860
Total Solv           A: -25.9451
Total Solv           E:  19.4183
DGintAB-A-B         : -128.0536
\R  4.5784 -137.6728   -1.4860   -25.9451   19.4183 -128.0536
```

From this output we can see that the total change in $\Delta\Delta G$ is not that different when computing it for the interaction of all residues and when computing it for the interaction of the residues in the interface only. Which means that most of the energetic importance is due to the binding in the interface.

Incident Report:

Error:

We had some warnings when trying to traverse the library on the aaLib.lib, that was not detected since those residues were written as “SERN” and “ASPC”.

```
WARNING: atom not found in library ( SER:H1 )
WARNING: atom not found in library ( SER:H2 )
WARNING: atom not found in library ( SER:H3 )
WARNING: atom not found in library ( ASP:OXT )
WARNING: atom not found in library ( THR:H1 )
WARNING: atom not found in library ( THR:H2 )
WARNING: atom not found in library ( THR:H3 )
WARNING: atom not found in library ( GLY:OXT )
None
```

Solution:

We used a manual tip, we changed those that were the terminals for SER or ASP, this way it was detectable. We couldn't solve them all the way, but we reduced those terminals.

After the aaLib.lib file its modified, we obtain just this, which we accepted.

```
WARNING: atom not found in library ( GLY:OXT )
None
```

d. Step 3: Ala-Scanning Experiment

To determine the effect of replacing each interface residues with Ala in the overall ΔG_{A-B} , to do it we follow these steps:

1. As a first step, we must prepare all the necessary items for the following steps: import libraries, define classes, load parameter files, and load the database structure using NACCESS.
2. The second step is to calculate the interaction energies (electrostatic and VdW) for each residue, along with solvation energies based on accessible surface area (ASA).
3. Next, we have to calculate interface-based energies, in which interface residues are used to recalculate interaction and solvation energies.
4. The script then performs Ala scanning by changing each interface residue to Ala and calculates the changes in electrostatic, solvation AB, solvation A, and total interactions.
5. Finally, the $\Delta\Delta G$ values for Ala mutations on interface residues are plotted using matplotlib.

Those steps can be reflected in step 3 of the Python script.

```
# Store the calculated values
residue_ids = []
bar_values = []

print("Ala Scanning: DDGs for X->Ala mutations on interface residues")
for ch in st[0]:
    for res in ch.get_residues():
        if args.cutoff_dist > 0 and res not in interface[ch.id]:
            continue
        residue_id_str = residue_id(res)
        print(
            '{:11} {:11.4f}{:11.4f}{:11.4f}{:11.4f}'.format(
                residue_id(res),
                - Intelec[res] + elec_ala[res],
                - Intvdw[res] + Intvdw_ala[res],
                - IntsolvAB[res] + IntsolvAB_ala[res],
                - IntsolvA[res] + IntsolvA_ala[res],
                - Intelec[res] + elec_ala[res] - Intvdw[res] + Intvdw_ala[res] - IntsolvAB[res] + \
                    IntsolvAB_ala[res] - IntsolvA[res] + IntsolvA_ala[res]
            )
        )
    )
```

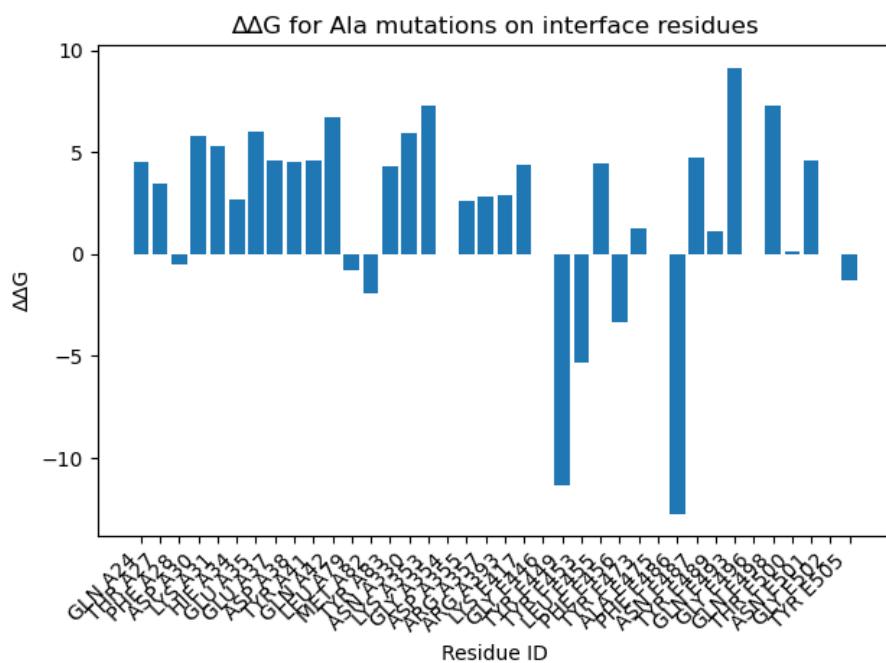
Overall, this is the code we implemented to find out the changes in energies when substituting each residue in the interface for an Alanine.

The output of the code:

residue	electro	vdw	solv_AB	solvA	total
---------	---------	-----	---------	-------	-------

Ala Scanning: DDGs for X->Ala mutations on interface residues					
GLN A24	-0.4486	0.8239	1.6975	2.4856	4.5584
THR A27	0.0955	2.8535	0.7679	-0.2643	3.4526
PHE A28	0.0407	1.3121	-0.4867	-1.4038	-0.5378
ASP A30	-1.4475	2.2174	1.6563	3.3476	5.7737
LYS A31	-1.0577	3.7652	0.0922	2.4986	5.2984
HIE A34	0.0803	-2.0503	1.6662	3.0157	2.7119
GLU A35	0.1796	1.3301	1.8874	2.6278	6.0249
GLU A37	0.4639	1.9786	0.9107	1.2671	4.6202
ASP A38	-0.2732	1.5836	0.8174	2.4061	4.5339
TYR A41	0.1712	6.0405	0.0000	-1.6475	4.5641
GLN A42	-0.3585	0.8485	0.6767	5.5626	6.7293
LEU A79	-0.0087	1.0152	-0.6568	-1.1270	-0.7774
MET A82	0.0060	1.4767	-1.4961	-1.8871	-1.9005
TYR A83	0.2481	4.1978	0.0025	-0.1313	4.3171
ASN A330	-0.0970	1.7424	0.9007	3.3847	5.9308
LYS A353	-0.8902	5.5884	0.1560	2.4628	7.3170
GLY A354	0.0000	0.0000	0.0000	0.0000	0.0000
ASP A355	0.6012	1.8800	0.0379	0.0682	2.5873
ARG A357	-0.3497	1.4214	0.4835	1.2802	2.8353
ARG A393	-0.4306	0.6750	0.9676	1.6621	2.8742
LYS E417	1.1607	1.5647	0.3601	1.3150	4.4005
GLY E446	0.0000	0.0000	0.0000	0.0000	0.0000
TYR E449	-0.8691	0.6386	-5.4993	-5.5999	-11.3297
TYR E453	-0.0611	-5.0079	-0.4708	0.2208	-5.3190
LEU E455	0.2311	4.9844	0.0000	-0.7840	4.4315
PHE E456	-0.0479	3.4312	-0.8367	-5.8768	-3.3303
TYR E473	-0.0542	1.0003	0.4993	-0.1828	1.2625
ALA E475	0.0000	0.0000	0.0000	0.0000	0.0000
PHE E486	-0.0627	5.6375	-4.3741	-13.9374	-12.7366
ASN E487	-0.1344	0.7049	0.6103	3.5211	4.7020
TYR E489	0.1018	7.7467	-1.6383	-5.0943	1.1160
GLN E493	-0.3094	2.0971	2.3015	5.0541	9.1433
GLY E496	0.0000	0.0000	0.0000	0.0000	0.0000
GLN E498	-0.3399	3.8386	-0.0110	3.8272	7.3149
THR E500	-0.9379	1.8938	-0.5208	-0.3066	0.1285
ASN E501	0.2641	3.3650	0.0578	0.9234	4.6103
GLY E502	0.0000	0.0000	0.0000	0.0000	0.0000
TYR E505	0.0147	7.5597	-1.6400	-7.2549	-1.3205

The most important column in the output of the code is the last column that calculates the differences of energy ($\Delta\Delta G$) between the energy of residues by interface and mutated residues for ALA, taking into account that the distance in Interface is that of 4.3.



This graph represents the difference in energy when mutated in a barplot, where we can compare the mutations in the residues.

Incident Report:

The energies we obtained from the Ala scanning were different than those we should have, we consulted Irene for this part and rechecked our code, in the end those differences were most likely due to the cleaning of the PDB structure in the preparation step.

```
WARNING: atom not found in library ( GLY:OXT )
<Atom OXT>
Interaction energy based in interface residues only
Total Elec Int.      :    4.3352
Total Vdw Int.       : -136.2065
Total Vdw Int.       : -136.2065
Total Solv AB        :   -4.9106
Total Solv          A:  -25.6235
Total Solv          E:   16.7311
DGintAB-A-B         : -127.8895
\#R 6m0j      4.3352 -136.2065   -4.9106   -25.6235   16.7311  -127.8895
Ala Scanning: DDGs for X->Ala mutations on interface residues
GLN A24      -0.4655  -0.5397   1.8081   2.5306   3.3336
THR A27      0.0938   3.1875   0.7692   -0.2673   3.7832
PHE A28      0.0531   1.3642   -0.5991  -1.6069  -0.7887
ASP A30      -1.3923   2.5027   1.7551   3.4265   6.2921
LYS A31      -1.0000   5.0141   2.2068   1.7994   8.0203
HIE A34      0.0885   0.5239   1.6992   3.0397   5.3514
GLU A35      -0.0381   0.1963   1.8335   2.7878   4.7794
GLU A37      0.4491   2.0987   0.9054   1.5457   4.9989
ASP A38      -0.2653   1.4971   0.8910   2.5722   4.6951
TYR A41      0.1661   5.2696   0.0000  -1.5778   3.8579
GLN A47      -0.1170   2.1764   1.4494   6.7775   0.7888
```

This is what we should have obtained, but the difference in $\Delta\Delta G$ total of Irene's code and of our code, is not that big. Since, for interaction energy based on the interface between residues only for her is -127.8895 and ours is -128.0536.

On the other hand, there are some $\Delta\Delta G$ total of the residues that are affected by this changes more dramatically, yet those who should worsen the $\Delta\Delta G$ total and those who make it better, they remain the same residues, so even if the analysis is not that exact it is useful to see the type of interaction changes.

e. Step 4: PyMOL Images

To make the images in pymol we have done step by step the following pipeline:

Step 1. Upload the “6m0j_fixed.pdb” structure into pymol.

Step 2. Color by chain:

>util.cbc

Step 3. Make selections of the chains using the sequence (by hand), in this case ChainA, chainE.

Step 4. Obtaining the interface:

4.1. residues of interface in ChainA:

>select interface1, byres ChainA within 4.3 of ChainE

4.2. To find the interactions on ChainE:

Action > find > polar contacts > with any atoms

color them magenta

4.3. Show lines for ChainA and ChainE

4.4. for easier visualization

color by atoms

Step 5. Obtaining distances between the interactions on the interface:

5.1. Select the interactions between chains manually

5.2. Finding out distances by retracing each of the interactions measuring each of the interactions:

Wizard > Measurement > Distances

Select Merge with previous -> to obtain all in one object

Step 6. Take pictures of the interface

ALA SCANNING

Step 1. select interested amino acids, those representative from the residue list provided in step 3.

Step 2. color it

Step 3. label residue

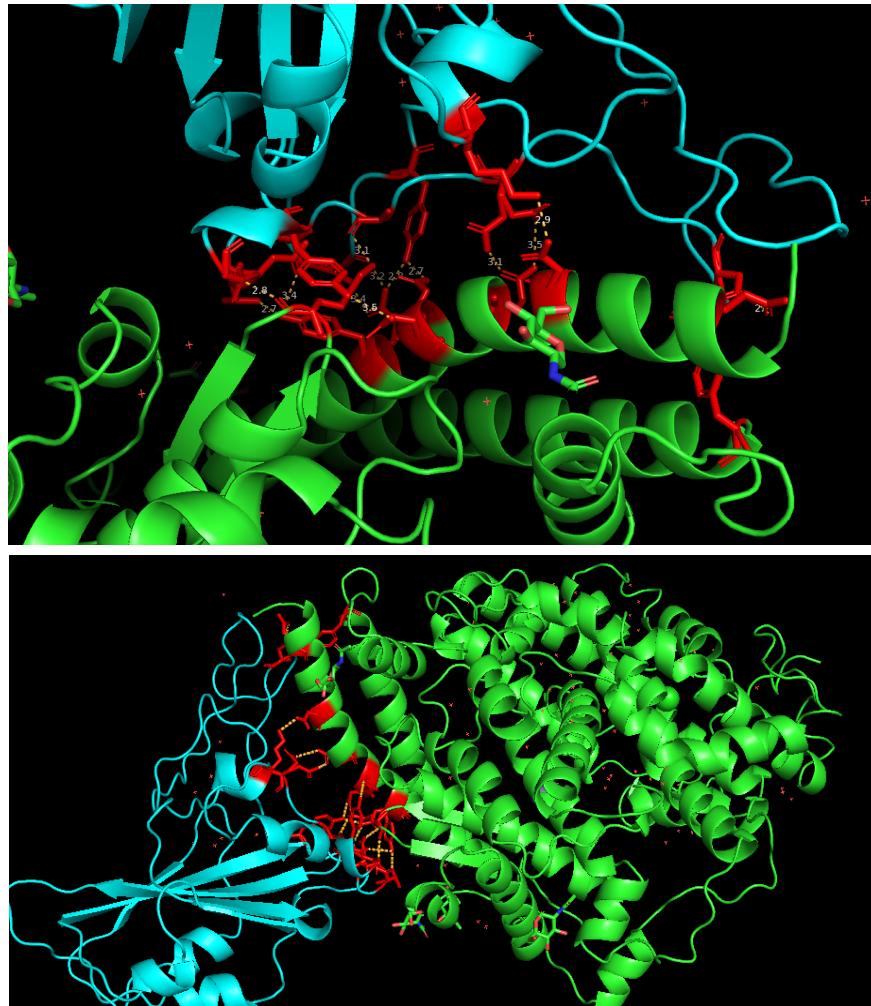
Step 4. Wizard > Mutagenesis > click on no mutation and change it into ALA

Step 5. label it again

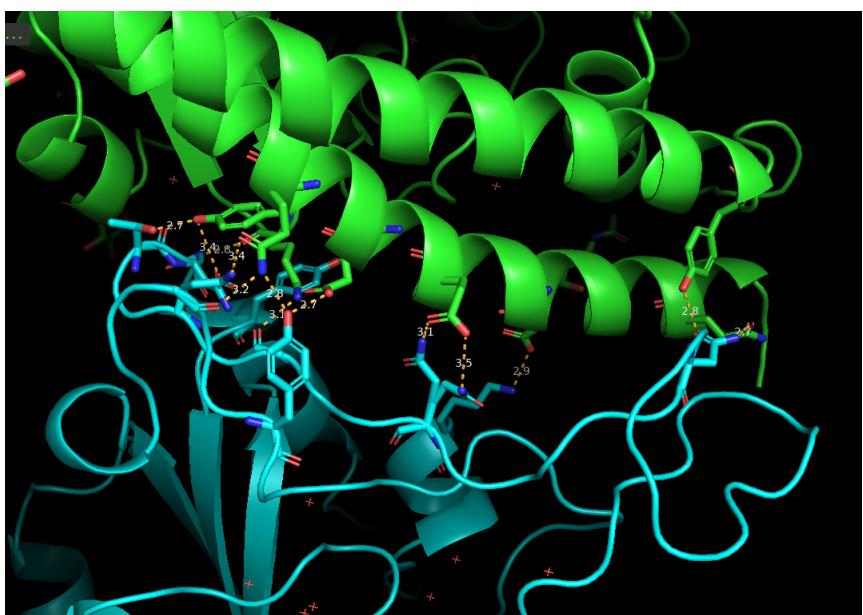
Step 6. find if it does the same interaction and the distance

Step 7. take picture

Here are some pictures of the interphase:



The first two images are highlighting the interphase. Achieved by following the previous steps.



On the 3rd image we can see the distances and interactions between the two proteins.

Provided List:

Ala Scanning: DDGs for X->Ala mutations on interface residues					
GLN A24	-0.4486	0.8239	1.6975	2.4856	4.5584
THR A27	0.0955	2.8535	0.7679	-0.2643	3.4526
PHE A28	0.0407	1.3121	-0.4867	-1.4038	-0.5378
ASP A30	-1.4475	2.2174	1.6563	3.3476	5.7737
LYS A31	-1.0577	3.7652	0.0922	2.4986	5.2984
HIE A34	0.0803	-2.0503	1.6662	3.0157	2.7119
GLU A35	0.1796	1.3301	1.8874	2.6278	6.0249
GLU A37	0.4639	1.9786	0.9107	1.2671	4.6202
ASP A38	-0.2732	1.5836	0.8174	2.4061	4.5339
TYR A41	0.1712	6.0405	0.0000	-1.6475	4.5641
GLN A42	-0.3585	0.8485	0.6767	5.5626	6.7293
LEU A79	-0.0087	1.0152	-0.6568	-1.1270	-0.7774
MET A82	0.0060	1.4767	-1.4961	-1.8871	-1.9005
TYR A83	0.2481	4.1978	0.0025	-0.1313	4.3171
ASN A330	-0.0970	1.7424	0.9007	3.3847	5.9308
LYS A353	-0.8902	5.5884	0.1560	2.4628	7.3170
GLY A354	0.0000	0.0000	0.0000	0.0000	0.0000
ASP A355	0.6012	1.8800	0.0379	0.0682	2.5873
ARG A357	-0.3497	1.4214	0.4835	1.2802	2.8353
ARG A393	-0.4306	0.6750	0.9676	1.6621	2.8742
LYS E417	1.1607	1.5647	0.3601	1.3150	4.4005
GLY E446	0.0000	0.0000	0.0000	0.0000	0.0000
TYR E449	-0.8691	0.6386	-5.4993	-5.5999	-11.3297
TYR E453	-0.0611	-5.0079	-0.4708	0.2208	-5.3190
LEU E455	0.2311	4.9844	0.0000	-0.7840	4.4315
PHE E456	-0.0479	3.4312	-0.8367	-5.8768	-3.3303
TYR E473	-0.0542	1.0003	0.4993	-0.1828	1.2625
ALA E475	0.0000	0.0000	0.0000	0.0000	0.0000
PHE E486	-0.0627	5.6375	-4.3741	-13.9374	-12.7366
ASN E487	-0.1344	0.7049	0.6103	3.5211	4.7020
TYR E489	0.1018	7.7467	-1.6383	-5.0943	1.1160
GLN E493	-0.3094	2.0971	2.3015	5.0541	9.1433
GLY E496	0.0000	0.0000	0.0000	0.0000	0.0000
GLN E498	-0.3399	3.8386	-0.0110	3.8272	7.3149
THR E500	-0.9379	1.8938	-0.5208	-0.3066	0.1285
ASN E501	0.2641	3.3650	0.0578	0.9234	4.6103
GLY E502	0.0000	0.0000	0.0000	0.0000	0.0000
TYR E505	0.0147	7.5597	-1.6400	-7.2549	-1.3205

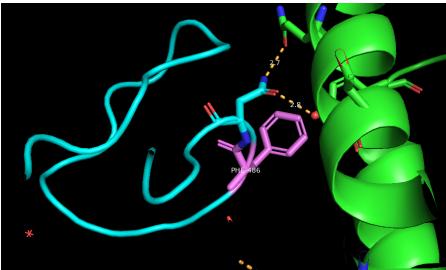
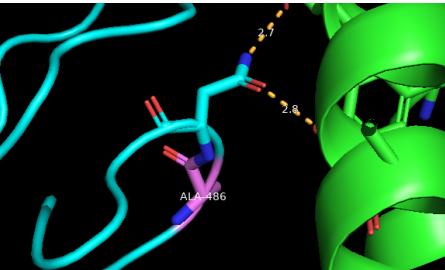
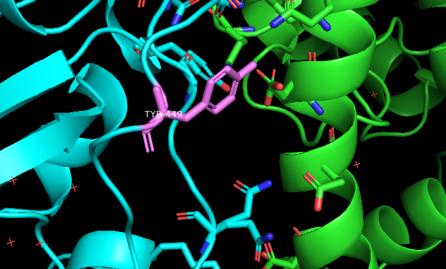
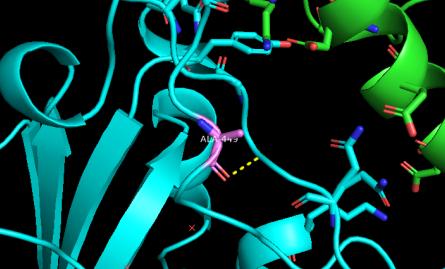
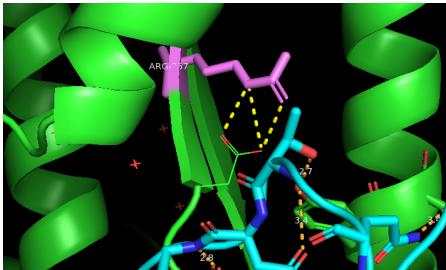
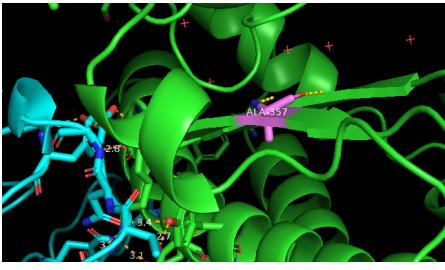
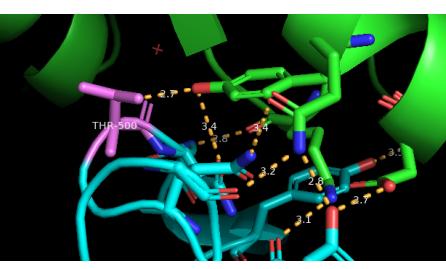
Annotations:

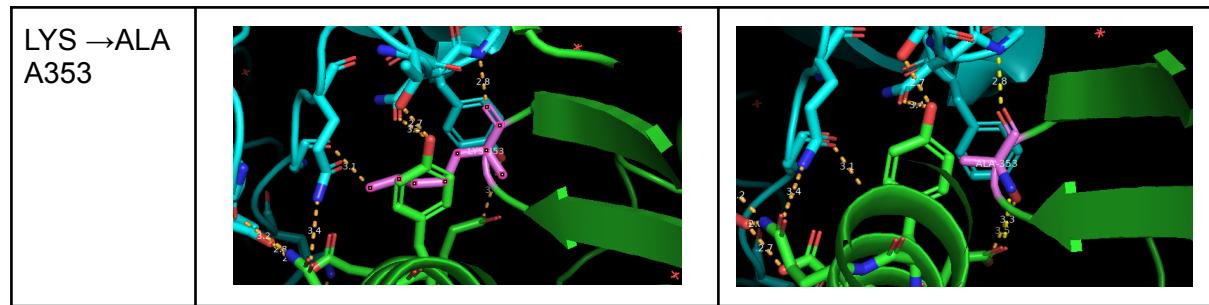
These are the cases we find most representative, having two mutations that would make $\Delta\Delta G$ better, one case that barely changes it and finally three cases that would make it worse with different variations.

There are also two cases where there would be no change in the variation of energy which would be changing Alanine for Alanine and as well as changing Alanine for Glycine. It makes sense because glycine is the one amino acid with the simplest sidechain, composed only by an atom of hydrogen.

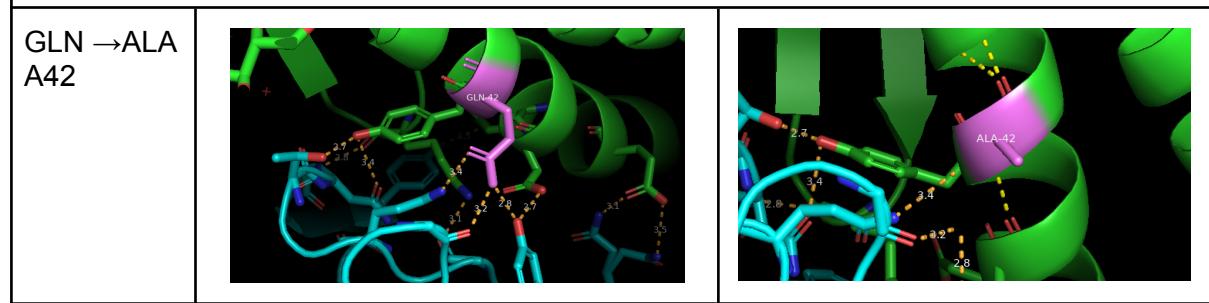


These are the study cases with representative photos of each case:

Mutation	Original	Mutated
PHE →ALA E486		
This is one of the most dramatic changes, most important, a $\Delta\Delta G$ total of -12.7366 mostly due to solvation. The change for ALA makes the solvation get better since PHE is more hydrophobic.		
TYR →ALA E449		
This change is also pretty dramatic, since the $\Delta\Delta G$ total is -11.3297, since the solvation also gets better. Most probably the same reason as the previous case but TYR has less intense hydrophobicity.		
ARG →ALA A357		
The change of ARG for ALA has a small $\Delta\Delta G$ total of 2.8353, which means that it gets a little worse. That is due to the protein being in a hydrophobic surrounding, that is because ARG has a positive polar side chain.		
THR →ALA E500		
This change has barely any significance at all, this is because the change in energies actually nullifies each other. Since the structures are very similar, the electrostatic and the solvation get better but the van der waals get worse.		

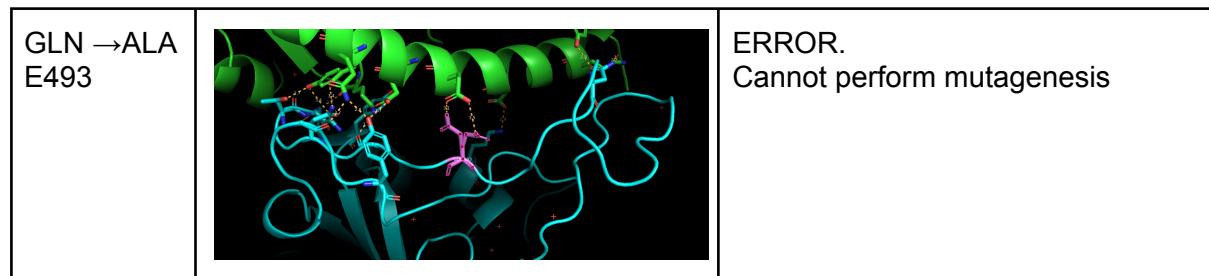


This is one of the most severe changes in terms of the $\Delta\Delta G$ total getting worse. Caused mainly because of the solvation and the van der waals interactions. Since the LYS makes more interactions than the ALA.



This is a similar case to the previous one, but it is in terms of solvation.

- Incident



This is the case in which $\Delta\Delta G$ total varies the most in terms of getting worse.

Sadly, this residue couldn't be explored in pymol, when trying to perform the mutagenesis on the GLN of the position 493 in the ChainE into Alanine there was this error raised:

You clicked /cleaned_6m0j/E/GLN'493/CA'A

Selected:

ExecutiveRMSPairs-Error: Atom counts between selection sets don't match (3 != 4).

Error:

4. Conclusion

In summary, our study extensively explored the complex molecular interactions between the Receptor Binding Domain (RBD) of the SARS-CoV-2 Spike protein and Angiotensin Converting Enzyme (ACE2). We aimed to understand how specific interface residues influence the interaction energy within the RBD-ACE2 protein-protein complex.

The insights we gained are crucial, offering a detailed understanding of the molecular dynamics governing the RBD-ACE2 interaction. These findings not only illuminate the fundamental aspects of this protein-protein complex but also establish a foundation for future research. The knowledge obtained is particularly important for the development of therapeutics and vaccines, targeting a critical step in the SARS-CoV-2 infection process.

Despite encountering challenges, our study's methodology has proven to be robust and adaptable. The framework we established for evaluating interface residues and conducting Ala-scanning experiments provides a guide for future investigations. This foundation is not only relevant to the specific context of viral infections but also paves the way for exploring protein-protein interactions in a broader sense. As we navigate the intricate realm of biological processes, the insights from this study offer valuable guidance for researchers across diverse fields.