
BIOPHYSICS - ENERGY ANALYSIS EXERCISE

Alba Jiménez, Aisha Shah, Cristina Artero

Table of contents

Table of contents	1
Introduction	1
Procedure	2
STEP 1: Fixing the protein structure for our analysis	2
STEP 2: Interface residues	5
STEP 3: Alanine scanning	6
STEP 4: Images by pymol	8
STEP 5: Most relevant sequence variants	10

Introduction

The aim of this report is to evaluate the relative contribution of interface residues to the interaction energy in the protein-protein complex between the Receptor Binding Domain (RBD) of SARS-CoV-2 Spike protein and its receptor the Angiotensin Converting Enzyme (ACE2).

Moreover, we need to identify known SARS-CoV-2 Spike Variants and evaluate their effect on ACE2 Binding. Replace the appropriate residue with the variant and reevaluate the interaction energy.

Procedure

STEP 1: Fixing the protein structure for our analysis

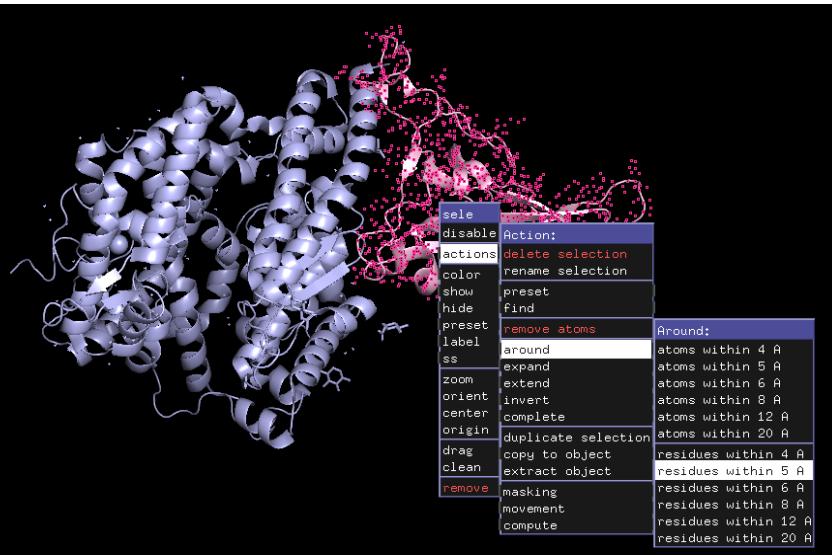
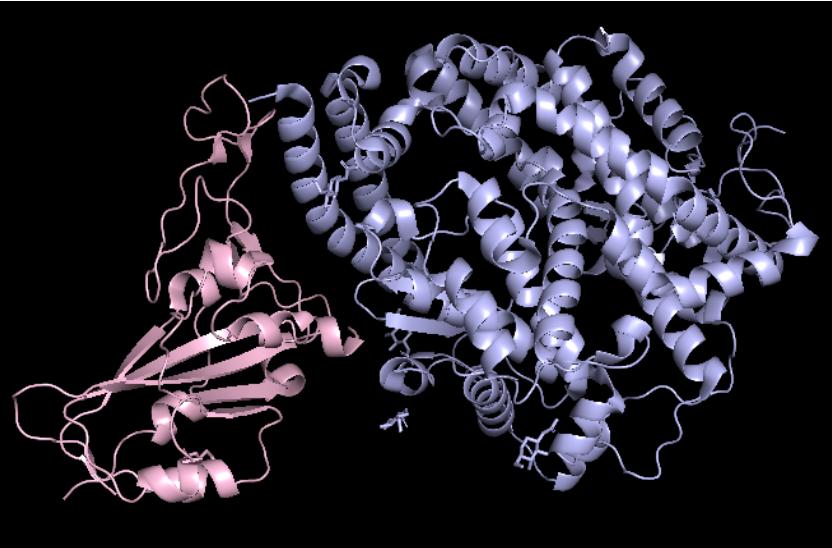
The interface between two given chains is defined by a list of residues on both of them that have at least one atom below a given distance. Therefore, to understand the interface of our desired structure, we downloaded the PDB structure (6M0J) in PDB and CIF format. As mentioned before, the structure contains 2 chains (A and E), which respectively represent ACE2 and Spike protein. Moreover, the crystal contains Zn^{2+} and Cl^- ions and NAG groups.

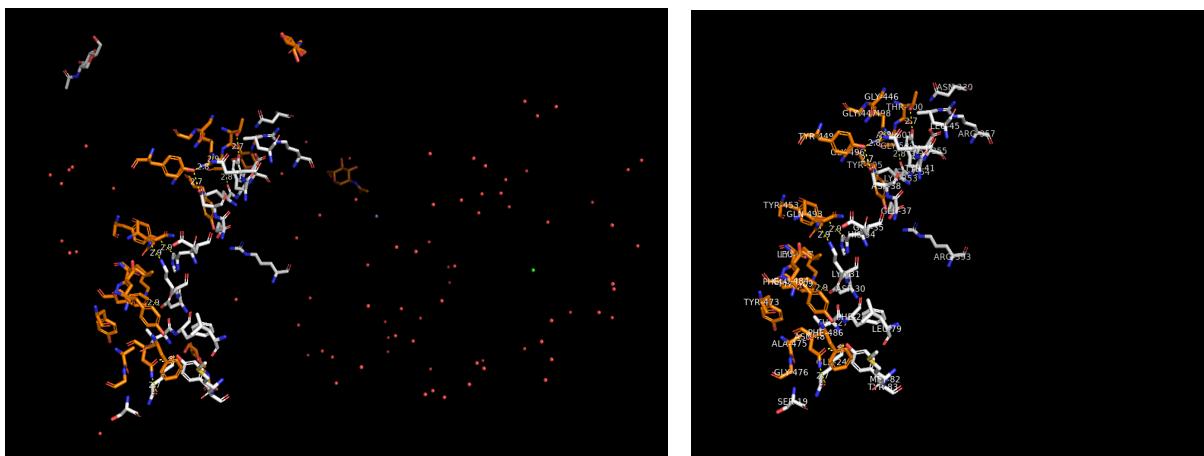
Afterwards, to better visualize the 3D crystal structure, we resorted to Pymol.

Firstly, we colored the SARS-CoV-2 Spike protein (pink) and the ACE2 receptor chain (purple) in different colors to be able to differentiate them, as we can see in the attached image.

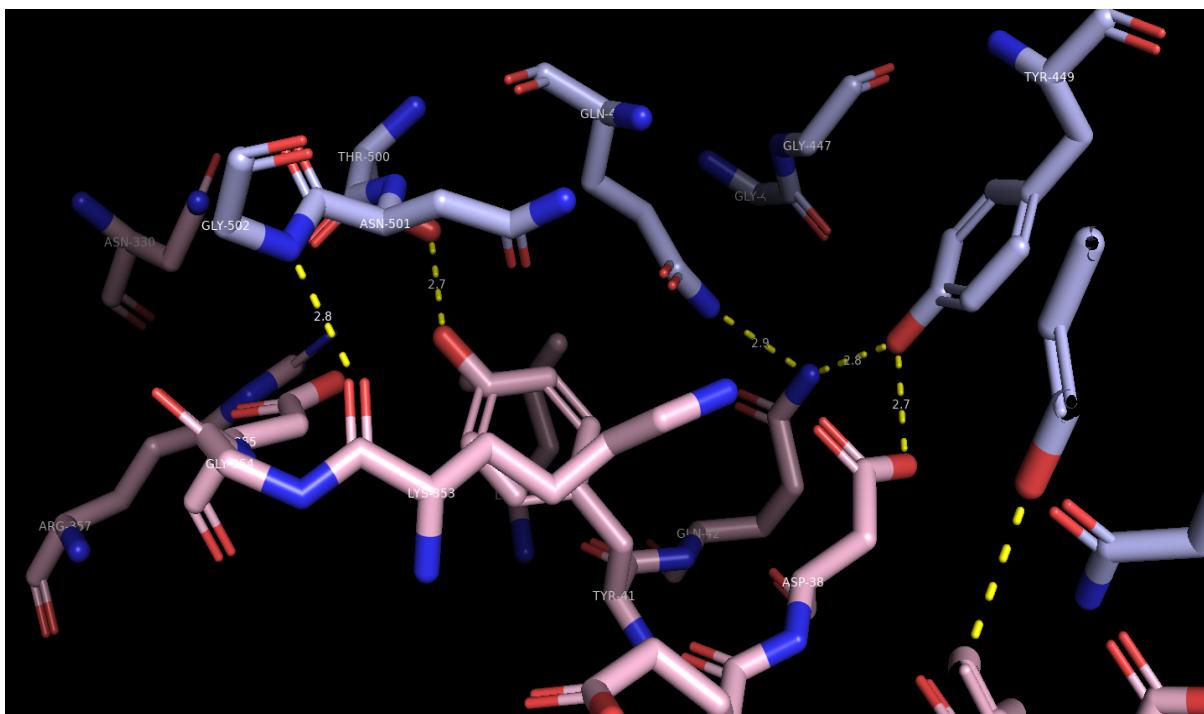
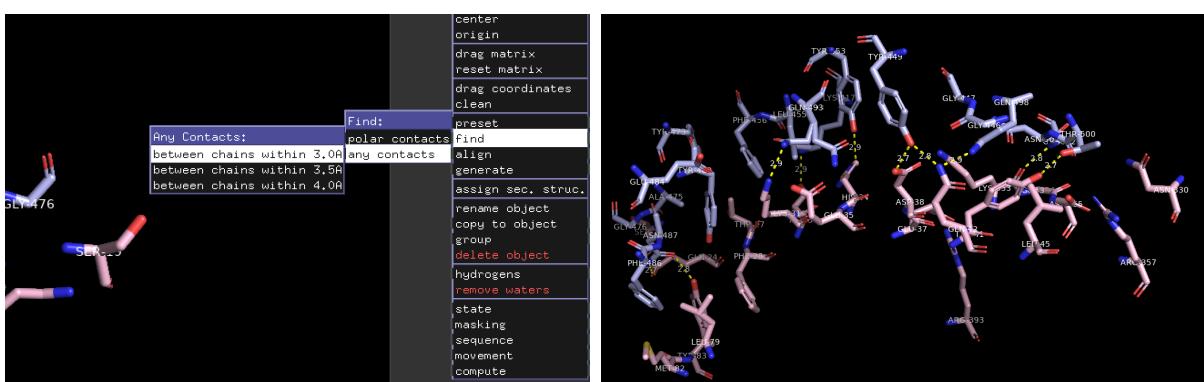
Secondly, we selected all residues within 5 Å for both chains. That way, we would obtain a selection of the residues that play a role in the interface interaction of both proteins.

From the remaining selection (see images in the next page), we can see in the picture on the left that there are some residual molecules such as water, single Zn^{2+} and Cl^- ions and NAG groups, that should be hidden to better visualize the interface residues, this time colored in orange and white.





Finally, we displayed the distance between atoms participating in any contact closer than 3 Å, and labelled those distances. We also labeled the amino acids to be able to check if the results from our scripts were accurate or not. In the pictures below we can see some examples of the interactions. The image below shows perfectly a zoom of the obtained results, but with different colors.



Next, we created a script to define the list of interface residues on each chain. We named it “list_res.py”.

Our script was designed to receive 2 input elements: the PDB file and a cut-off distance.

We run the script with 2 different pdb files. The other parameter (distance) remained the same (5). The files were the pdb file obtained from PDB (before fixing) and the file obtained after fixing it with the biobb_structure_checking module.

Below, we can find as output the lists of interface residues for each chain when running the script with both files.

Before fixing

Chain A : {'LEU A79', 'GLY A354', 'GLU A35', 'THR A27', 'LYS A353', 'PHE A28', 'LEU A45', 'ASP A30', 'GLN A24', 'LYS A31', 'ASP A355', 'ARG A393', 'HIS A34', 'GLN A42', 'ARG A357', 'GLU A37', 'TYR A83', 'MET A82', 'SER A19', 'ASN A330', 'TYR A41', 'ASP A38'}

Chain E: {'PHE E456', 'TYR E505', 'PHE E486', 'TYR E489', 'GLU E484', 'ASN E501', 'TYR E473', 'GLN E498', 'GLY E502', 'GLN E493', 'TYR E449', 'GLY E447', 'LEU E455', 'THR E500', 'ASN E487', 'GLY E446', 'GLY E476', 'TYR E453', 'GLY E496', 'LYS E417', 'ALA E475'}

After fixing

Chain A: {'TYR A83', 'ASP A355', 'HIE A34', 'SER A19', 'ARG A357', 'LEU A79', 'ASP A38', 'LYS A31', 'ALA A386', 'LEU A351', 'GLY A354', 'ASP A30', 'LEU A45', 'THR A27', 'LYS A353', 'ALA A387', 'THR A324', 'GLY A352', 'TYR A41', 'MET A82', 'ASN A330', 'ARG A393', 'GLN A325', 'GLU A37', 'GLN A42', 'GLY A326', 'PHE A28', 'GLU A35', 'GLN A24', 'ILE A21'}

Chain E: {'GLY E502', 'LEU E455', 'PHE E490', 'TYR E495', 'TYR E453', 'GLN E493', 'GLY E476', 'GLY E496', 'GLY E446', 'GLN E498', 'VAL E445', 'ASN E501', 'GLY E485', 'PRO E499', 'VAL E503', 'PHE E486', 'TYR E505', 'SER E477', 'LYS E417', 'GLY E504', 'TYR E449', 'TYR E489', 'PHE E497', 'ASN E487', 'GLY E447', 'ALA E475', 'ARG E403', 'THR E500', 'GLU E484', 'PHE E456', 'TYR E473'}

The residues that appear at the “after fixing” lists but not in “before fixing” ones (also notice that they are not written in bold) are the result of adding H-bonds when running the biobb_structure_checking module from the Jupyter notebook script.

STEP 2: Interface residues

Firstly, we splitted the structure in its two chains. We decided to use the previous script and add a few lines in the end to receive the PDB file, separate both chains and create different PDB files to store them by using the module PDBIO. We would like to highlight that the file from which we generated the chains was the fixed one that was created using biobb_structure_checking module. The files were named as A.pdb (for ACE2) and E.pdb (for Spike protein).

Secondly, we wrote a script that computed the calculation of ΔG^{A-B} . If we recall the formula:

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

That script, named “Func_energy_calculation.py”, received different inputs: the fixed PDB file, the fixed pdbqt, the PDB files of the chains and a cutoff distance (which takes into account the atoms within a range of the specified distance in angstroms to evaluate the atoms which have interactions between them).

There, we declared a function to compute the interaction energy. That function calls other functions that separately calculate the electrostatics, the Van der Waals and the solvation energies of A, B and AB. Then, by using the formula from above, it returns the interaction energy.

In our case, we obtained that, if we use as input 6m0j_fixed.pdb, 6m0j_fixed.pdbqt, A.pdb, E.pdb and 5 as cutoff distance, the energies are:

$\Delta G_{\text{elect}}^{A-B}$	4.310102656876709 kcal/mol
$\Delta G_{\text{vdw}}^{A-B}$	-72.42825491635966 kcal/mol
$\Delta G_{\text{Solv}}^{A-B}$	-0.7666480000000162 kcal/mol
ΔG_{Solv}^A	-24.435952 kcal/mol
ΔG_{Solv}^B	27.783634000000003 kcal/mol
ΔG^{A-B}	-72.23248225948298 kcal/mol

STEP 3: Alanine scanning

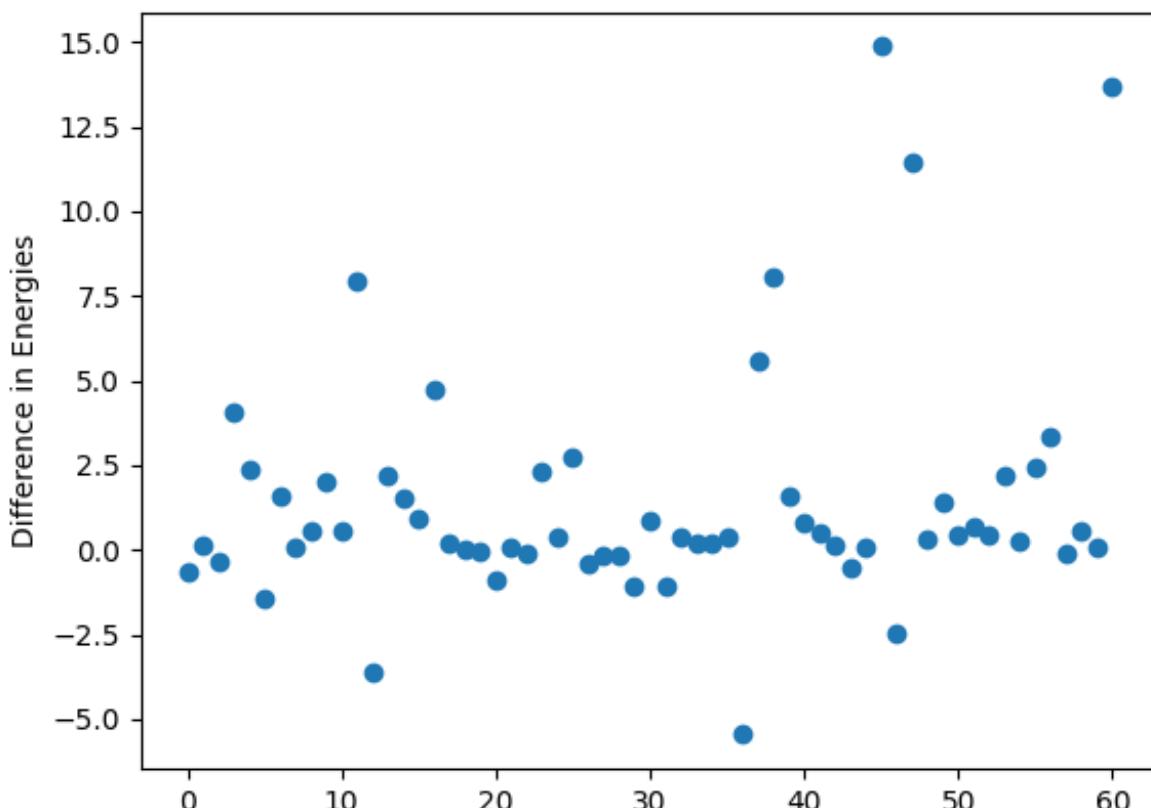
To determine the effect of replacing each interface residue by Alanine in the overall ΔG^{A-B} , we wrote a function which received as an input the same as the function to compute the interaction energy.

In this case, we took into account that Alanines have atoms that all other residues will have (except glycine). So, we calculated the energies of those atoms in common and stopped when an uncommon atom appeared.

We printed the energy of the complex when each residue was mutated to Alanine. Then, we stored all those energies in a dictionary to see the energies of each mutated amino acid in a sorted way. Afterwards, we stored in a list the difference of each energy and the original interaction energy to be able to plot later $\Delta\Delta G$ and to better visualize which mutations produce bigger changes.

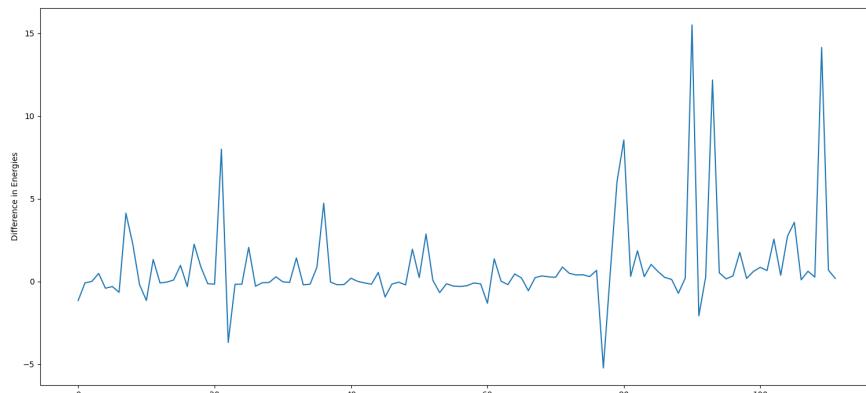
We concluded that residues TYR.A41, TYR.E453, LEU.E455, PHE.E456, PHE.E486, TYR.E489 and TYR.E505 produce the biggest changes in the structure. Precisely, residues TYR.A41, LEU.E455, PHE.E456, PHE.E486, TYR.E489 and TYR.E505 contribute to less stability. On the other hand TYR.E453 and LEU.455 are less destabilizing than PHE.486 and TYR.E505 (Table given below).

We stated that since the lower the energy the better for stability. That means that residues that have a higher $\Delta\Delta G$ are contributing to less stability whereas having negative $\Delta\Delta G$ contributes to more stability. That is important because it defines which amino acids are important for the stability of the complex and for SARS-CoV-2 to have a higher infectivity and define more virulent variants.



Number	Residue ID	Interaction energy	$\Delta\Delta G$
486	PHE.E486 (non-polar)	-57.37	14.867
505	TYR.E505 (polar)	-58.56	13.673
489	TYR.E489 (polar)	-60.8	11.429
456	PHE.E456 (non-polar)	-64.1906	8.0419
41	TYR.A41 (polar)	-64.27	7.9666
455	LEU.E455 (non-polar)	-66.6346	5.5979
453	TYR.E453 (polar)	-77.685	-5.453

If we run the same script but with a cutoff distance of 8, we will get more results but the energies of the residues from previous tests are the same. This shows that there is consistency in our code. In particular, for a cutoff distance of 8, the relevant residues are:



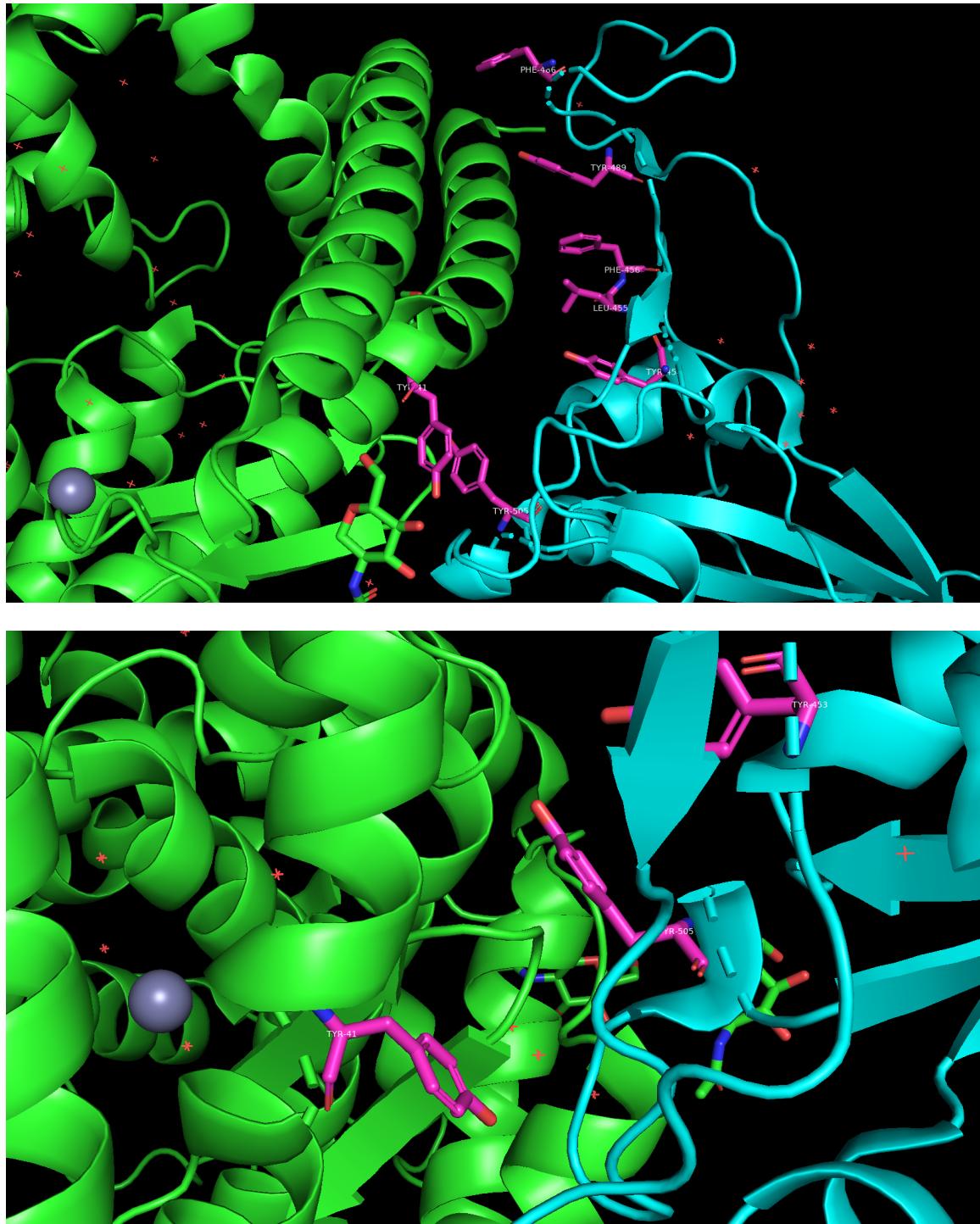
Number	Residue ID	Interaction energy	$\Delta\Delta G$
486	PHE.E486 (non-polar)	-68.33113359101328	15.497
505	TYR.E505 (polar)	-69.69061262636986	14.137
489	TYR.E489 (polar)	-71.65971714598007	12.168
456	PHE.E456 (non-polar)	-75.28014470977638	8.5478
41	TYR.A41 (polar)	-75.84	7.9904
455	LEU.E455 (non-polar)	-77.80510576666208	6.0228
453	TYR.E453 (polar)	-89.05260802719378	-5.225

Again we can notice and the most destabilizing one is PHE.E486 and TYR.E505 are the most destabilizing ones.

STEP 4: Images by pymol

After identifying the most relevant residues, we can open Pymol to better visualize them.

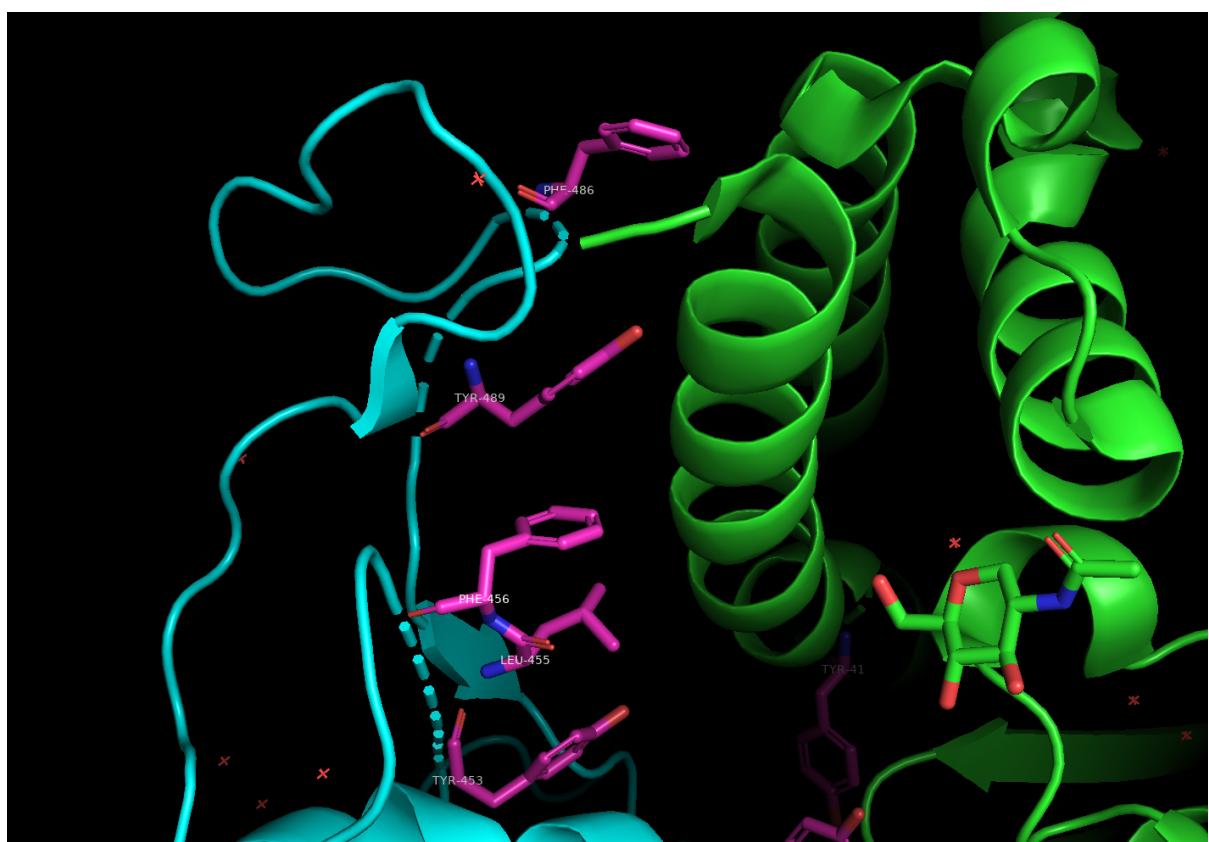
Now, in the photo below we can see ACE2 in green, Spike protein in light blue and the most relevant amino acids colored in pink.



Notice that TYR.A41 and TYR.E505 are facing each other. They establish polar contacts (since both amino acids are polar) which may be important for the structure. Losing one of them would mean a remarkable interaction.

In the next photo, we see PHE.E486, TYR.E489, PHE.E456, LEU.E455 and TYR.E453. Check that they all are facing an alpha-helix of ACE2. Losing all except TYR.E453 would also cause a great loss in stability. On the other hand, TYR.E453 could be mutated to Alanine and would not affect the structure that much.

We get a large positive value for phenylalanine, it might be because it is polar and has a larger size than Alanine so by mutating it to alanine we will lose some vdw interactions so this change will be destabilizing.



STEP 5: Most relevant sequence variants

First, we need to find the PDB files for the 3 variants. In PDB, we found the following ids:

7EKF → alpha
7EKG → beta
7V8B → delta

Then, we repeated the whole process from before:

1. Use biobb_structure_checking to fix all of them and generate pdbqt files.
2. Consequently, we splitted their chains into single PDB files.

We have to recognise that delta variants's PDB file could not be passed through the biobb_checking module so we decided to focus on the other 2.

We obtained the following results when cutoff distance is 5:

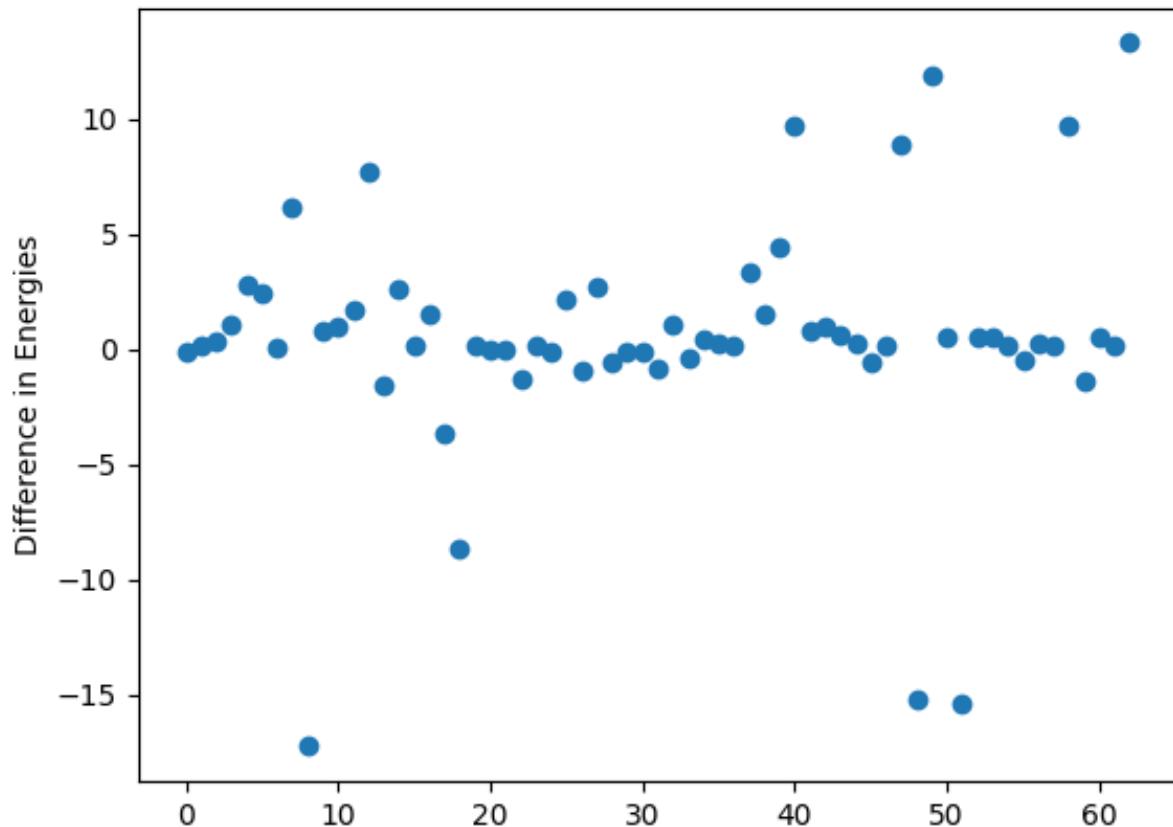
Alpha: $\Delta G^{A-B} = -47$ kcal/mol

Beta: $\Delta G^{A-B} = -74.31$ kcal/mol

We can see that the Alpha variant has the lowest energy.

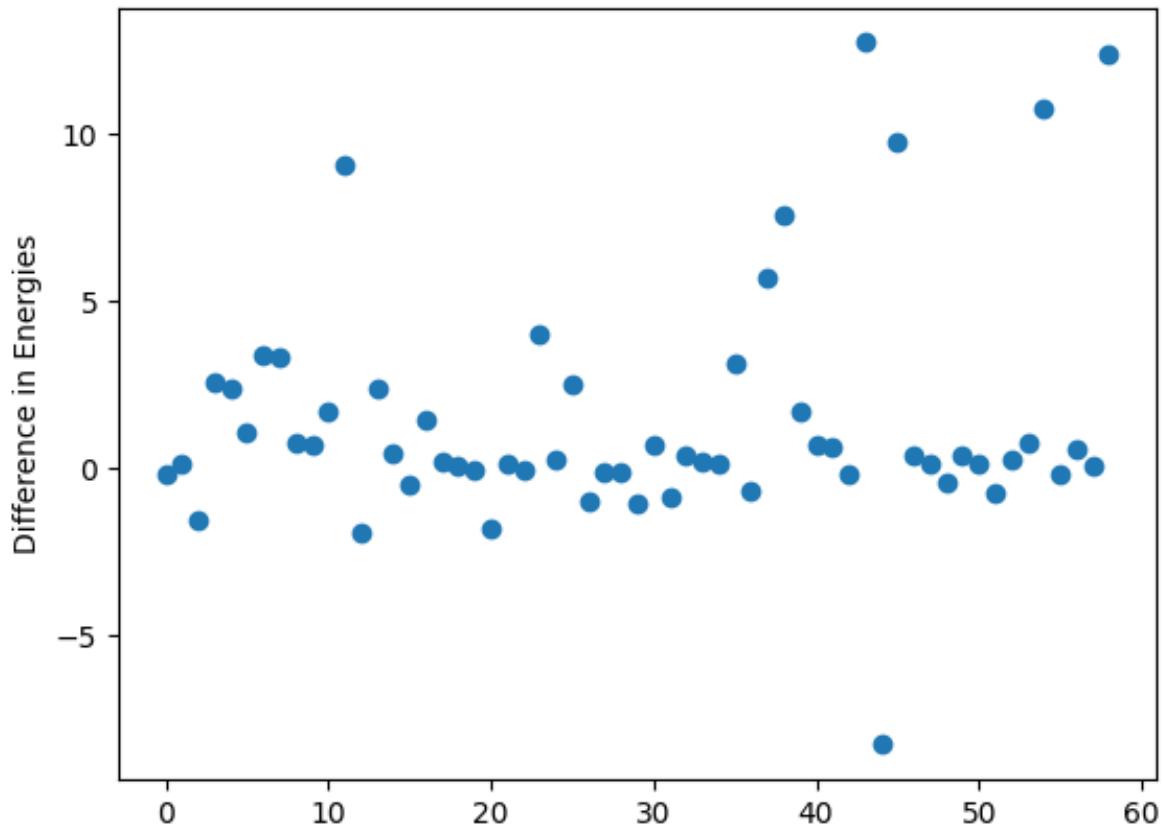
Furthermore, we can find the plots for the $\Delta\Delta G$ for alpha and beta variants when running and ala-scanning.

Alpha variant



ID	Residue ID and name	Interaction energy	$\Delta\Delta G$
31	LYS.A31	-40.8483	6.1546
34	HIE.A34	-64.2555	-17.2526
41	TYR.A41	-39.2956	7.70734
83	TYR.A83	-55.6352	-8.63229
456	PHE.B456	-37.282	9.7209
486	PHE.B486	-38.1218	8.88114
487	ASN.B487	-62.2145	-15.2116
489	TYR.B489	-35.1244	11.8785
493	GLN.B493	-62.3749	-15.372
501	TYR.B501	-37.327	9.67592
505	505 TYR.B505	-33.715	13.2879

Beta



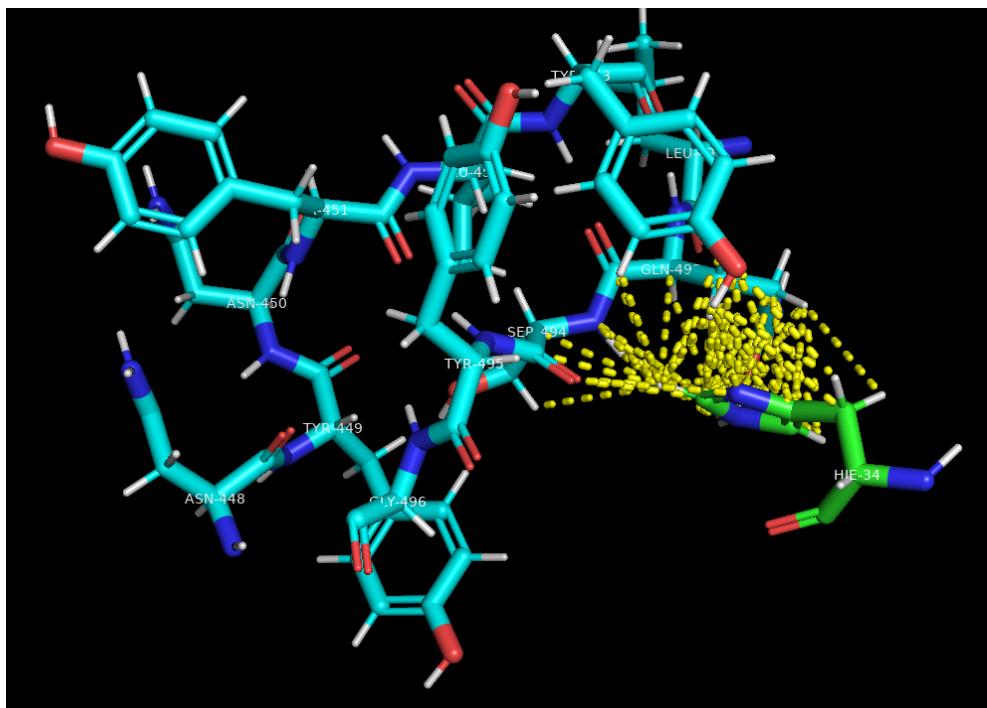
ID	Residue ID and Name	Interaction energy	$\Delta\Delta G$
41	TYR.A41	-65.2209	9.08632
353	LYS.A353	-70.2825	4.02472
455	TYR.A41	-68.6165	5.69078
486	PHE.B486	-61.5827	12.7246
487	ASN.B487	-82.5722	-8.26495
489	TYR.B489	-64.5726	9.73462
501	TYR.B501	-63.5362	10.7711
505	505 TYR.B505	-61.9562	12.3511

$\Delta\Delta G$				
Residue id and name (Chain E and B are the same they are labelled differently in different variants)	Cutoff distance 5			Cutoff distance 8
	alpha variant (7EKF)	beta variant (7EKG)	6m0j	6m0j
HIE.A34 (+)	-17.2526	3.28619	0.0758266	0.0894895
LEU.492 (non-polar)	Not present at cutoff 5	0.157489	Not present at cutoff 5	0.33656
GLN.B493 (polar)	-15.372	-0.423789	1.43091	1.75958
SER.B494 (polar)	0.486564	Not present at cutoff 5	Not present at cutoff 5	0.182607
TYR.B505	13.2879	12.3511	13.673	14.137
PHE.E486 (non-polar)	8.88	12.72	14.867	15.497

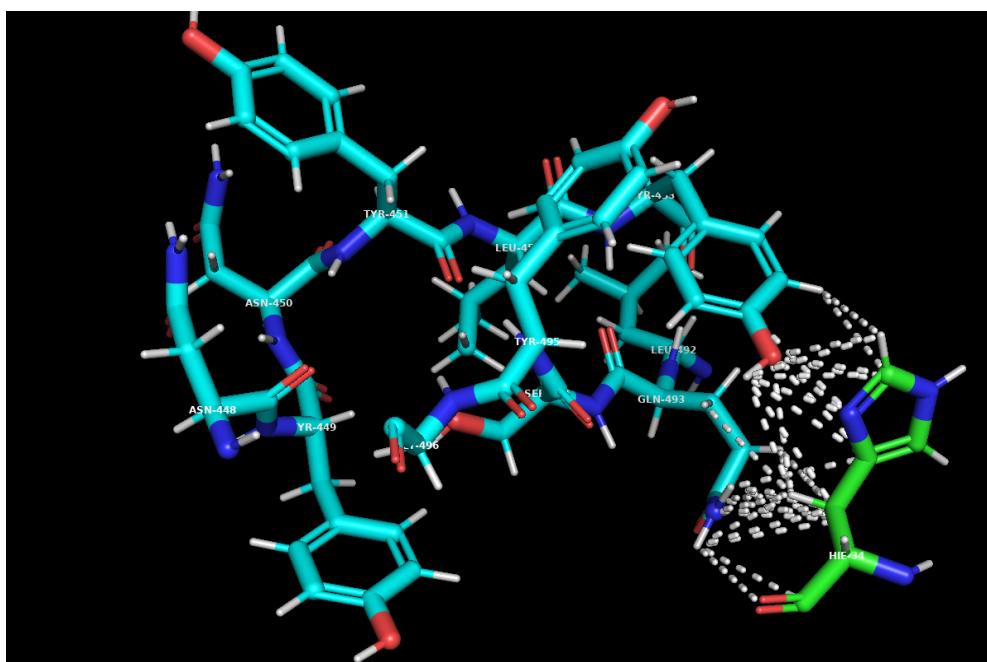
We can see that in all variants TYR B505, PHE.E486, TYR E489 are always highly destabilizing.

In 6m0j and beta variant, the residue Serine is not at a distance lower than or equal to 5 with any residues of chain A. Moreover, the change in energy (absolute) due to the energy contribution of Histidine 34 of chain A and GLN 493 of chain B is much much higher than the change of energy in other variants. This can be due to the conformation of of Histidine34 (green) as shown below:

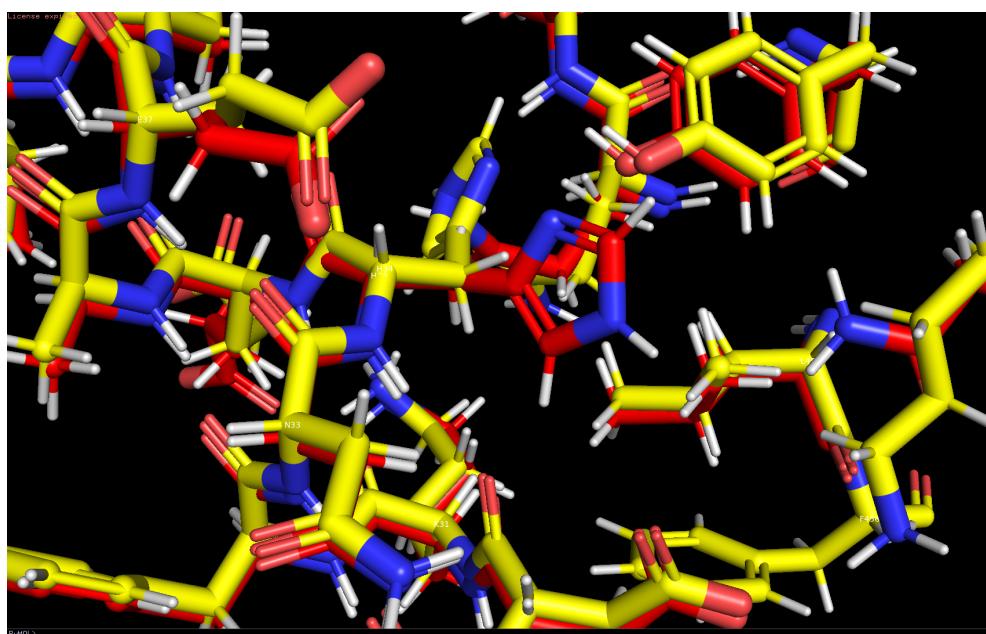
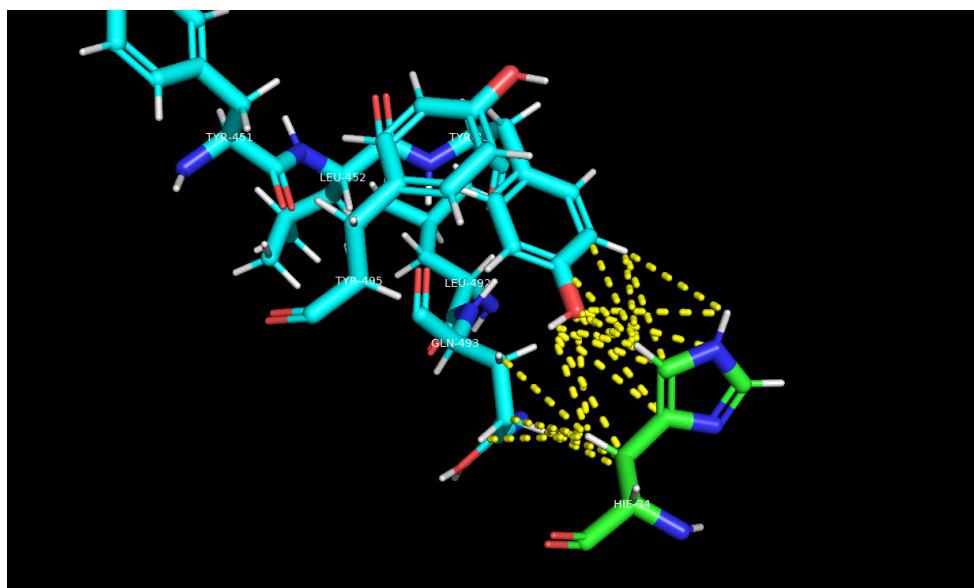
Alpha Variant



Beta Variant



6m0j



In this image we can see the overlapped chains of alpha (yellow) and beta (red) showing different conformations of HIE.A34.