REPORT

Energy analysis exercise 2023-24

Spike RBD-ACE2: Protein-protein interface analysis

BIOPHYSICS (2023/24)

Irene Acero I pousa

Josep Gelpi Buchaca

Alberto Meseguer Donlo

Alessandra Bonilla Salon, Alexandra Caus Carrascosa,

Paula Martín Cabezas and Irene Porta Bravo

Table of Contents

I. INTRODUCTION

Biophysics is the point of collision between physics and biology and plays an important role in unraveling the mysteries of life at the molecular level. It employs quantitative methods and principles from physics to understand the structure and dynamics of biological systems. As a sign of this, protein interfaces have been understood a little bit more and this knowledge could be extrapolated to other areas.

Protein interfaces are understood as the inner space where molecules interact and it holds the key to fundamental biological processes. The energy dynamics within these interfaces govern the stability and functionality of proteins. Exploring it contributes in comprehending molecular recognition, signaling pathways, and the overall behavior of biological macromolecules. In the pursuit of understanding protein interfaces, the focus often narrows down to individual residues within these molecular landscapes. Each amino acid residue contributes to the overall energy of the interface, influencing the binding affinity, stability and specificity of protein-protein interactions.

The primary objective of the present report is developing a code for interface residues calculation, mainly focused on the case of Spike RBD-ACE2. As a consequence, this could automate a laborious process, if it is done by hand, and decrease the error rate.

II. BACKGROUND

The interaction between the Spike Receptor-Binding Domain (RBD) and ACE2 is a crucial molecular process central to the infection mechanism of SARS-CoV-2, the virus responsible for COVID-19. Positioned on the viral surface, the Spike protein houses the RBD, specifically binding to the human angiotensin-converting enzyme 2 (ACE2) receptor on host cells. This binding serves as the initial step for the virus to enter host cells, initiating membrane fusion and subsequent infection. The specificity and strength of the Spike RBD-ACE2 interaction are pivotal factors determining the virus's ability to infiltrate target cells. As a consequent,, this interaction emerges as a prime target for therapeutic interventions, drawing intense research focus to comprehend and disrupt the viral entry mechanism**.**

III. METHODOLOGY AND PROCEDURE

In order to accomplish our main goal, the approach is the following:

- Determine amino acid residues that form the interface between the complex components (use 6m0j)
- Determine interface residues contributions.
- Determine the contribution to the stability of the complex by mimicking a traditional Ala-scanning experiment, i.e. replacing each residue in turn by Ala and evaluating the changes in the complex interaction energy.

Likewise, the procedure is divided into four steps to maintain the order.

Step 0: Preparation

This step encompassed from the obtention of the structure to the prior modifications on it, in order to fix possible issues in the protein complex and ensure more accurate calculations a posteriori.

Firstly, it was required to download the structure from the PDB database (*Image 1*). We searched for *6M0J*[1] and saved the file in PDB format. Before it, we had to take into account the protein structure of the composition of the Biological unit; as it was the case since the biological assembly and the asymmetric unit are the same, we did not need to remove additional chains.

Following this, the next step involved the removal of heteroatoms (*Image 2*). In order to enhance the specificity and relevance of the analysis concerning the direct interactions between the protein partners at the binding interface, a strategic decision is often made to eliminate heteroatoms from the evaluation of protein-protein interaction energy. This is particularly applicable in the context of the RBD-ACE2 complex, enabling a more focused and computationally efficient exploration of the critical residues involved in the interaction. To accomplish this, two distinct methods were employed. The initial method utilized Pymol[2], involving a series of commands that facilitated the removal of heteroatoms in three straightforward steps.

The second was done by us using Python and Biopython commands. While trying to execute the script provided to complete this part of the project, we encountered certain problems in different lines of code. Therefore, we decided to attempt removing the heteroatoms in our own way using our programming and Biopython knowledge. We thought a lot about how to do it, and we opted for an idea: to search within all the residues of both chains for those that were heteroatoms and eliminate them. However, our idea was not progressing because we couldn't find a way to eliminate objects present in another, and we also couldn't figure out how to access the different heteroatoms. Finally, with Alberto's help, we managed to find a way to eliminate them. We had to change our approach; now, instead of deleting the heteroatoms from the chain, we would add the residues that didn't contain heteroatoms to a newly created chain. This new chain would then be added to a new model, and this model to a structure. The way we identified residues that were not heteroatoms was by examining their ID. If there was a blank space at position 0, it meant that these were not heteroatoms, and therefore, they had to be added to the new chain. We followed this process for both chains in our protein. Finally, when we had the structure created, we saved it in a file called “6m0j_new” with "pdb" format to visualize it in Pymol, and this has been the result *(Image 3*).

---

[1] RCSB Protein Data Bank (2020). 6M0J: Crystal structure of SARS-CoV-2 spike receptor-binding domain bound with ACE2. Retrieved from https://www.rcsb.org/structure/6M0J

[2] Pymol is a molecular visualization system utilized for creating detailed three-dimensional representations of molecular structures. Widely employed in structural biology, it facilitates the visualization and analysis of complex biomolecular complexes, including proteins and nucleic acids.

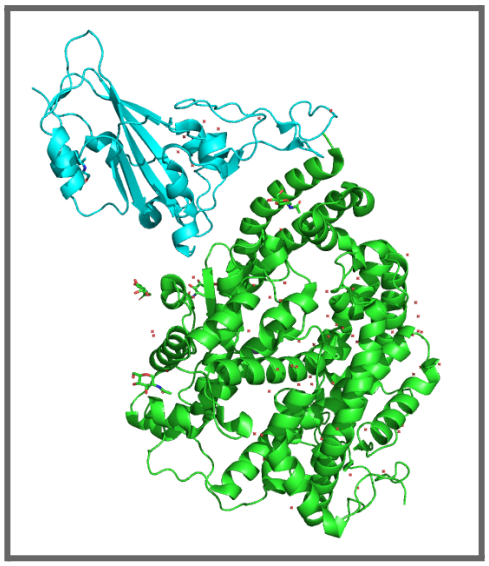

Image 1. Before any change

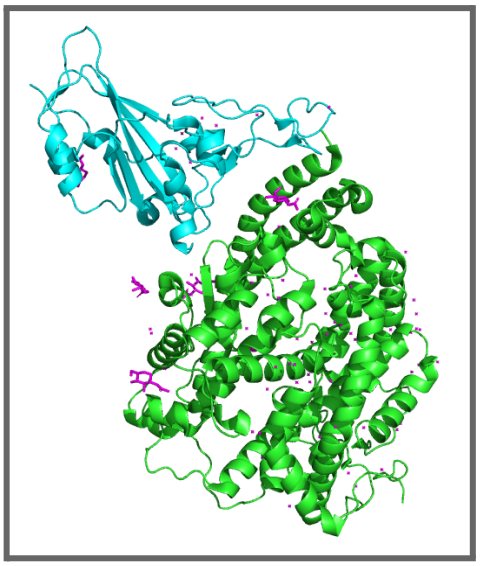

Image 2. Selected heteroatoms

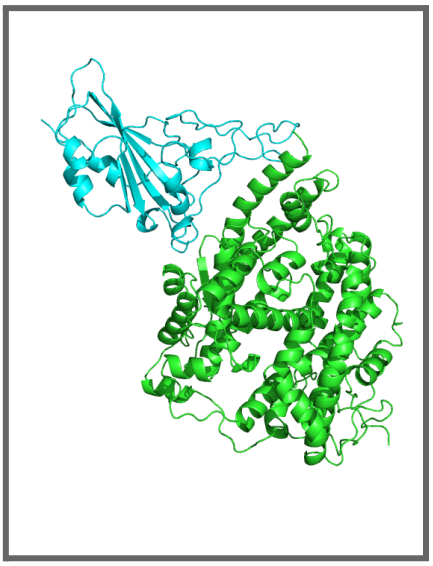

Image 3. 6m0j_new without heteroatoms

With this part finished, we had to perform a quality checking on the structure. To do so, we followed the github steps that we were given but using the file that we created while removing the heteroatoms. We used our own file to detect if we had done something wrong or if we had forgotten to remove any heteroatom. The result is represented by the image below. With this file, we continued our project.

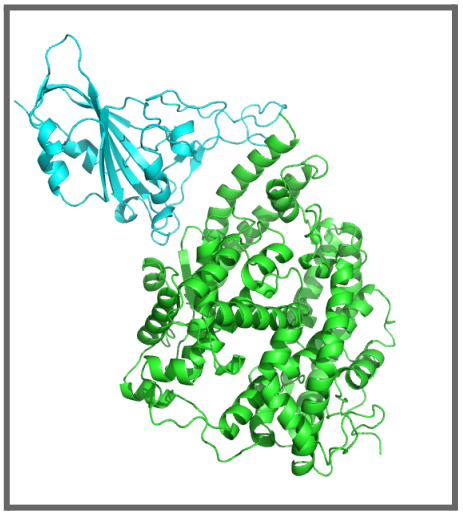

Image 4. 6m0j_fixed

Step 1: Identifying interface residues

In this section, our focus was on locating the residues situated at the interface of the protein complex. To achieve this, we will start with a visual identification using the Pymol interface. In the following images, it is evident that the RBD-ACE2 complex is composed of two chains (*Image 5*, colored in cyan and green). Subsequently, we manually selected those residues that appear to be in contact with the other chain (represented in pink in the image). This was the area where we identified the residues that form bonds.

Image 5. RBD-ACE2 complex with highlighted contact region of its chains

Next, we displayed the polar contacts (represented with dashes in *Image 6*) and once again manually selected those residues that are linked by a bond. As a final step, we measured the distances between each pair of selected residues, obtaining information that would serve as input in the code we would use.

We observed that the distances range between 2 to 3.5 Å, prompting us to opt for a distance of 4.3 Å as the input. This value ensures the inclusion of all residues comprising the interface.

Image 6. Bonds in the interface

In our project, we had developed two Python scripts to delineate the interface residues between protein chains, each employing a distinct methodology to approach the task.

OPTION CODE 1:

Python function **get_interface_residues** which is designed to identify interface residues between two specified chains of a protein within a given distance cutoff. The process involves:

- Loading the protein structure from the PDB file, identifying the atoms (coordinates) from two specified chains (chain1_id and chain2_id)
- Implementing a neighbor search to detect atoms within a set distance, signaling an interface interaction.
- Generating a collection of residue numbers that define the interface .

The output from OPTION CODE 1 is a printed interface with residue numbers for Chain A and Chain E of a protein structure file, which in our case is 6m0j_fixed.pdb.

OPTION CODE 2:

An alternative python function **get_interface** which is designed to identify interface residues between protein chains. This function:

- Skips hydrogen atoms to reduce computation time, only considering heavier atoms for interface detection.
- Employs the NeighborSearch module to find all pairs of atoms from different chains that are within the specified distance.
- Instead of lists, this approach uses sets for each chain to store unique residue objects that represent the interface residues. Sets are more efficient than lists when the order of items is not relevant and when duplicate entries are not necessary.
- This function returns a dictionary where chain identifiers are keys, and the associated values are sets of residue objects that form the interface. (interface[ch.id])

The output from OPTION CODE 2 provides a structured data collection (dictionary) containing residue objects for each chain that are part of the protein interface, based on the specified distance cutoff.

Both parts of the code perform essentially the same task, but use slightly different approaches to achieve it. Both start with a protein structure and a distance as input, and return residues at the interface of specific chains. The main difference lies in the implementation of neighbor search and how the results are stored.

List of interface residues on each chain obtained in OPTION CODE 1:

Interface residues in Chain A: [386, 387, 393, 19, 21, 24, 27, 28, 30, 31, 34, 35, 37, 38, 41, 42, 45, 324, 325, 326, 330, 79, 82, 83, 351, 352, 353, 354, 355, 357]

Interface residues in Chain E: [403, 417, 445, 446, 447, 449, 453, 455, 456, 473, 475, 476, 477, 484, 485, 486, 487, 489, 490, 493, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505]

Step 2: Energy calculation

In this step of our analysis, we have conducted a detailed assessment of interaction energies between protein residues at the interface of a complex, likely involving the Spike RBD-ACE2 complex as a case study. This quantitative analysis is pivotal in understanding the energetics of protein-protein interactions which are essential for the functionality of biological systems and can also be critical targets for therapeutic interventions.

The procedure employed in this step includes the following:

- **Calculation of Interaction Energies:** We utilized computational methods to calculate various interaction energies, such as electrostatic, van der Waals, and solvation energies, for the residues at the protein interface. These calculations were performed with precision to determine the contributions of individual residues to the overall stability of the complex.

- **Evaluation of Stability:** By analyzing the total interaction energy and the solvation energies for the complex as well as for the individual chains, we gauged the stability of the protein complex. Negative values in solvation energies indicate a release of energy upon complex formation, which typically contributes to the stability of the protein-protein interaction.

- **Determination of Free Energy Change (ΔG):** The overall change in free energy was computed for the interaction between the two chains, providing insights into the favorability of the protein complex formation under physiological conditions. A negative ΔG value is indicative of a stable and energetically favorable interaction.

- **Analysis of Results:** The calculated interaction energies revealed critical residues that significantly contribute to the interface stability. Electrostatic and van der Waals interactions suggest complementarity and stability of the interface, while solvation energies reflect the energetics of complex formation and the isolated chains.

Interaction energies for a set of interface residues

```
Interaction energy based in interface residues only
Total Elec Int.      :      8.9322
Total Vdw Int.       :   -146.4477
Total Solv AB        :     -6.7017
Total Solv         A:     -29.2811
Total Solv         E:      19.6783
DGintAB-A-B          :   -134.6145
The total interaction energy of the interface residues is -134.61452624957326
```

Image 7. Output obtained for the interaction energy of the interface residues.

- **Total Electrostatic Interaction (Total Elec Int.):** The electrostatic interaction energy is positive (+8.9322), which might suggest repulsive interactions or less favorable electrostatic complementarity at the interface. However, in the context of the entire system, this can be counterbalanced by other forces.

- **Total Van der Waals Interaction (Total Vdw Int.):** The van der Waals interaction energy is negative (-146.4477), indicating favorable, stabilizing interactions. This is expected, as van der Waals forces generally contribute to the stability of the interface through non-covalent bonds.

- **Total Solvation Energy for AB (Total Solv AB):** This value is negative (-6.7017), suggesting that the formation of the complex between A and B is associated with a release of solvation energy, which can contribute to the overall stability of the complex.

- **Total Solvation Energy for A (Total Solv A):** This value is negative (-29.2811), indicating the energy required to solvate chain A when it is isolated.

- **Total Solvation Energy for E (Total Solv E):** In contrast to the other solvation energies, this one is positive (+19.6783), suggesting that solvation contributes positively to the energy of chain E when it is isolated, which could be due to a gain in solvation upon disassociation of the complex or a higher degree of exposure to solvent.

- **Overall Change in Free Energy (DGintAB-A-B):** The change in free energy for the interaction is negative (-134.6145), indicating that the interaction is favorable and the complex is likely to be stable under physiological conditions. This is the most critical value as it reflects the net gain or loss of energy when the two chains come together to form the complex.

- **The Total Interaction Energy of the Interface Residues:** The provided number (-134.6145264957326) corroborates the free energy change, further implying a stable and energetically favorable interaction at the interface of the protein complex.

Interaction energy between protein residues in the complex (CHAIN A and E)

```
Interaction energy of residues in the complex
Total Elec Int & Vdw Int.:   -176.7422
Total Solv AB       :   -507.1030
Total Solv         A:   -398.5663
Total Solv         E:   -108.5366
DGintAB-A-B         :   -176.7422
```

Image 8. Output obtained for the interaction energy of the whole complex.

- **Total Electrostatic and Van der Waals Interaction (Total Elec Int & Vdw Int):** The sum of electrostatic and van der Waals interactions is negative (-176.7422) which might suggest that the interactions between the residues are favorable, as negative values indicate a thermodynamically stable interaction.

- **Total Solvation Energy for AB (Total Solv AB) :** The total change in solvation energy upon complex formation (AB) is -507.1030. This large negative value indicates that the complex formation is associated with a significant reduction in solvation energy, which typically contributes to the stability of the complex.

- **Total Solvation Energy for A (Total Solv A):** The solvation energy for chain A alone is -398.5663. This value refers to the energy required to solvate chain A when it is not in a complex with chain E.

- **Total Solvation Energy for E (Total Solv E):** The solvation energy for chain E alone is -108.5366. Similarly, this value refers to the energy required to solvate chain E when it is isolated.

- **Overall Change in Free Energy (ΔGintAB-A-B):** The overall free energy change (ΔG) for the interaction is -176.7422. This is the critical value as it represents the net change in free energy upon binding of the two chains (A and B) together. A negative ΔG value indicates that the formation of the complex is favorable under the given conditions.

Step 3: Replacing each interface residues by alanine

The task involves assessing the impact of substituting each residue within the interface with Alanine (Ala) on the overall ΔG_{A-B}. Subsequently, a plot of the results would be generated, emphasizing the residues that play a more significant role in interface stability.

Changing each residue at the interface to alanine in the calculation of the change in Gibbs free energy (ΔG) for the A-B interaction (ΔG^{A-B}) can have several effects on the stability and strength of the interaction between the two molecules A and B.

1. **Reducing Interaction:** Alanine is a small, nonpolar amino acid. If we replace a residue at the interface with alanine, specific interactions are likely to be lost, especially if the original residue contributed larger interactions, such as hydrogen bonds or ionic interactions. This could result in a reduction in the affinity of the A-B interaction.

2. **Changes in Solubility:** Alanine is hydrophobic, so if we replace a polar residue at the interface with alanine, we change the solubility of the protein or protein complex. This could affect the overall stability of the structure.

3. **Impact on Conformation:** Alanine does not have bulky side chains or particular functional groups. If the original residue contributed significantly to the three-dimensional structure or interface conformation, the change to alanine could affect structural stability.

4. **Effect on Dynamics:** Changing residues at the interface to alanine could alter the dynamics of the A - B interaction. Alanine is more flexible than some larger amino acids, so it could affect the stiffness or flexibility of the interface.

In general, the final specific effect will depend on the original residue we are replacing, as well as the nature of the A-B interaction and the associated biological function. By performing these types of mutations we understand the contribution of individual residues to stability and molecular interaction.

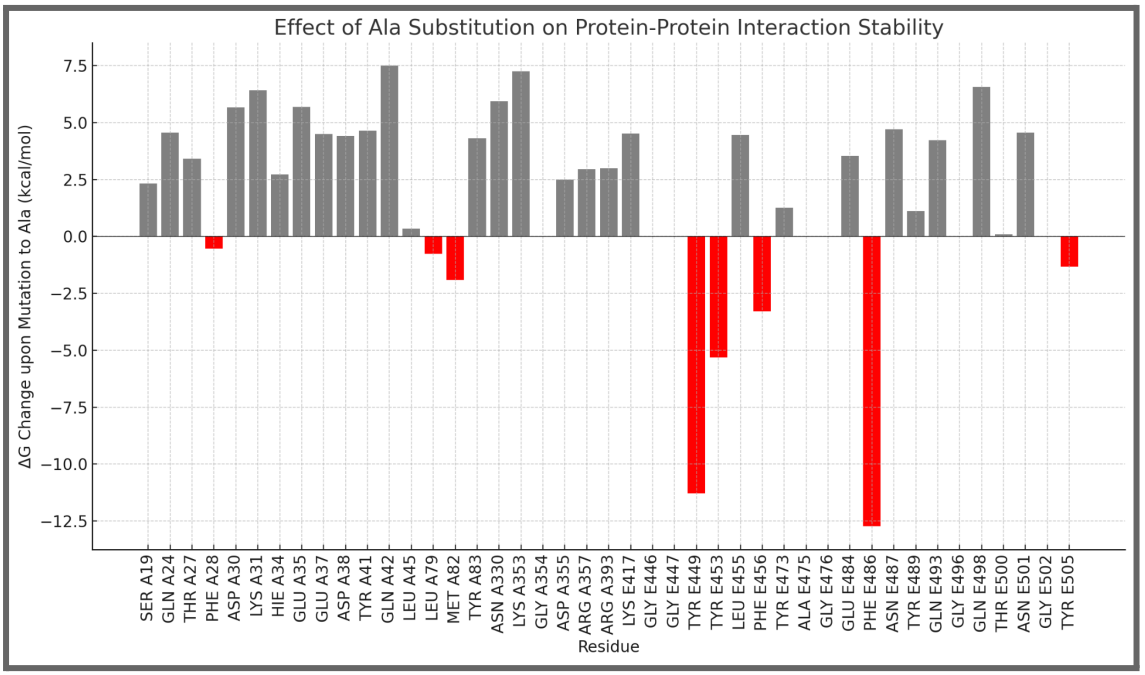

Image 9. Bar plot of the effect on the stability of alanine substitution for each interface residue.

This plot illustrates the change in interaction energy (ΔG) resulting from the mutation of interface residues to alanine in a protein-protein complex. The residues are displayed along the x-axis, and the corresponding ΔG changes upon mutation to alanine are represented by the bars on the y-axis.

Red bars indicate a significant decrease in stability upon mutation to alanine (negative ΔG change), highlighting these residues as particularly important for the stability of the interface. These residues often have side chains that are involved in specific interactions, such as hydrogen bonds, ionic interactions, or hydrophobic contacts, that contribute to the overall binding energy of the complex.

Gray bars, especially those above the zero line, suggest that mutation to alanine does not significantly destabilize the complex, or may even stabilize it, indicating these residues are less critical for the interface's stability.

- **Large Negative ΔG (Critical Residues):** Residues with a large negative ΔG upon mutation to alanine are likely to have side chains that play a significant role in the binding interface. These could be residues that are involved in key interactions, such as salt bridges or hydrogen bonds, which are lost when replaced with the smaller side chain of alanine.
- **Small or Positive ΔG (Less Critical Residues):** These residues might not be directly involved in binding or may contribute less to the binding energy. The side chains of these residues may be solvent-exposed or not optimally positioned for interaction with the binding partner. Their replacement with alanine might not disrupt critical interactions, and in some cases, it might even remove steric clashes or create a more favorable interaction environment.

Step 4: Visualization of relevant residues and interactions

After plotting each residue replaced by alanine in a graph, we selected those that are significant and impact the stability of the protein complex. For this step, we utilized Pymol to visualize these regions more effectively.

Those are the amino acids and their position: TRY E449, TRY E453, PHE E456, PHE E486.

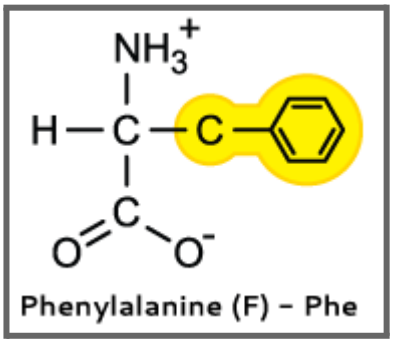

Image 10. Phenylalanine structural formula

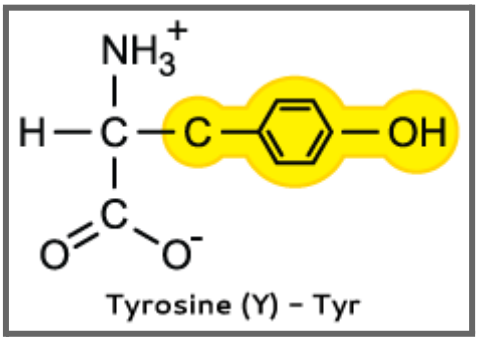

Image 11. Tyrosine structural formula

Tyrosine (polar amino acid) and phenylalanine (non-polar) share a structural similarity in the phenyl part. However, tyrosine has an additional hydroxyl group in its side chain. When either of these residues is mutated to alanine, the additional functional group (hydroxyl in the case of tyrosine) is eliminated and replaced with a simple methyl in alanine.

These mutations can affect the stability of a protein because the side chains of tyrosine and phenylalanine are often crucial in forming specific interactions, such as hydrogen bonds or hydrophobic interactions, at the protein binding interface. By changing these residues to alanine, which has a simple and non-polar side chain, these specific interactions can be disrupted, potentially leading to a decrease in the stability of the protein binding interface.

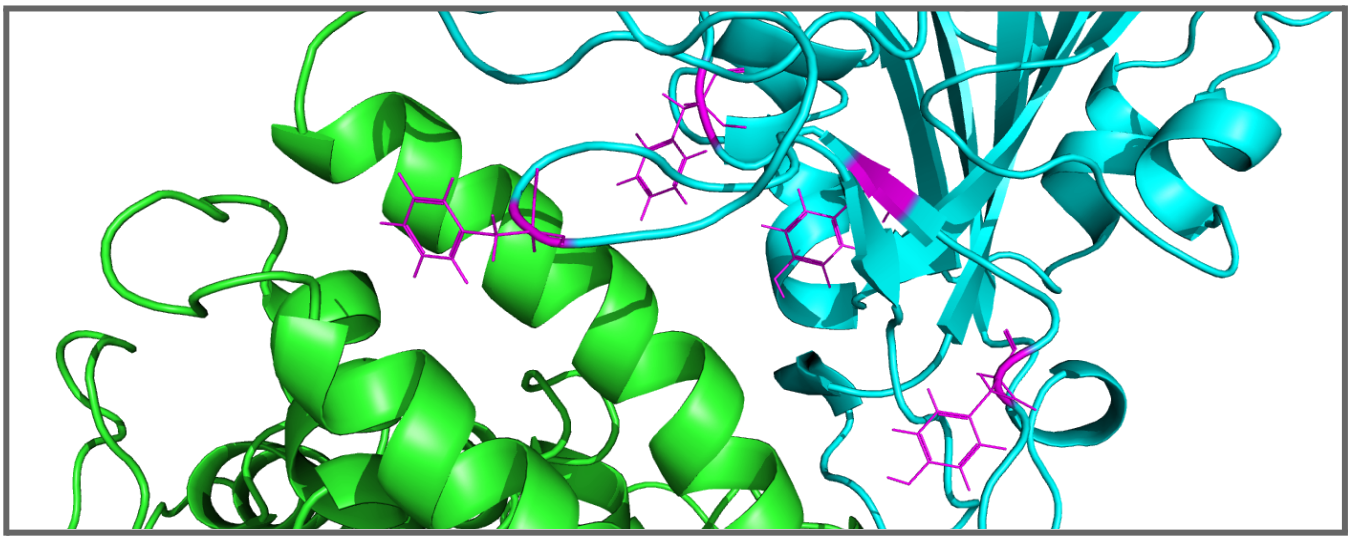

Image 12. Visualization of found significant residues.

IV. RESULTS

INTERFASE ENERGY VS TOTAL ENERGY

From the results obtained with the code, we observed that the interface energy is comparable to the total energy. If the interface energy significantly contributes to the total energy, it suggests that interactions in the binding region are critical for the overall stability of the protein or protein complex. In other words, the formation and maintenance of interactions in the interface have a substantial impact on the overall stability of the protein structure.

This observation may indicate that the interactions in the interface are specific and strong, underscoring the importance of this region for the biological function of the protein.

PLOT ANALYSIS

Analyzing the plot, it's crucial to focus on the residues exhibiting a more negative change in Gibbs energy—specifically, PHE E486, TYR E449, TYR E453, and PHE E456. These residues are more inclined to possess side chains that play a significant role in the binding interface. There has been a significant change of stability which means that the mutation of this residue to alanine has changed the amino acid, so it is very important. The importance of this mutation lies in the difference in size and complexity of the side chains. By replacing phenylalanine with alanine, the large and aromatic portion of the side chain is lost, potentially affecting local interactions and the three-dimensional structure of the protein at the binding interface. This could influence the stability of the protein in that specific area and have consequences for interactions with other molecules.

These might be residues participating in vital interactions, such as salt bridges or hydrogen bonds, which are disrupted when substituted with the smaller alanine side chain.

V. CONCLUSION

The integrated approach, combining experimental visualization, computational analysis, and theoretical mutation studies, provides a comprehensive understanding of the Spike RBD-ACE2 protein complex.

Identified interface residues, energy calculations, and alanine substitution analyses contribute valuable insights into the molecular interactions and stability of the complex. We found that if we exchange PHE E486, TYR E449, and TYR E453 with alanine, there is a crucial change in the stability of Spike RBD-ACE2. This is due to the phenyl group, which, due to its size and complexity, formed specific interactions that structurally affect the protein.

In summary, the presented steps form a robust frameworK for the detailed exploration of protein-protein interactions of the Spike RBD-ACE2 protein complex and we could develop a code capable of compute the energy interaction.