Spike RBD-ACE2 Protein-protein interface analysis

The evaluation of the relative contribution of interface residues to the interaction energy in a protein-protein complex.

Authors: Alma de Maria López Sebastià, Lucia Navarro Torres, Izan Torrero Salas

Introduction

The rapid expansion of large-scale sequencing initiatives has generated an unprecedented volume of protein variant data, increasing the need for systematic methods to assess the structural and energetic consequences of amino-acid substitutions. Understanding how sequence variation influences protein-protein interactions has become essential for fields ranging from molecular medicine to biotechnology.

The interaction between the Receptor-Binding Domain (RBD) of the SARS-CoV-2 Spike protein and the human Angiotensin-Converting Enzyme 2 (ACE2) is a paradigmatic example of a biologically critical protein-protein interface. Binding between these two components is a necessary event for viral entry, and mutations in the RBD that alter ACE2 affinity have played a fundamental role in shaping the infectivity and epidemiological behavior of emerging SARS-CoV-2 variants. Quantitative evaluation of the energetic contribution of individual interface residues can therefore provide valuable insights into the determinants of complex stability and the molecular basis of viral adaptation.

This project aims to perform a detailed computational analysis of the RBD-ACE2 interface using the crystallographic structure 6M0J[1]. The work focuses on defining interface residues, computing interaction energies under relevant approximations, carrying out an alanine-scanning-like evaluation, and assessing the energetic impact of mutations present in major SARS-CoV-2 variants. The analysis integrates Python-based tools, the biobb_structure_checking module, NACCESS surface-area calculations, and FoldX energy estimations, allowing a multipronged examination of residue contributions to complex

[1] Lan, J., Ge, J., Yu, J. et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 581, 215-220 (2020). PDB ID: 6M0J.

stability. The results obtained will be critically discussed and contextualized in terms of residue chemistry, interaction networks, and known viral mutations.

Objectives

The objectives of this report are fourfold. First, we aim to identify and characterize the interface residues within the RBD-ACE2 complex. Second, we will quantify the contribution of individual residues to the interaction energy between both components. Third, an alanine-scanning will be performed computationally in order to evaluate the impact of side-chain removal on interface stability. Fourth, an evaluation of the energetic effects of mutations present in major SARS-CoV-2 variants will be performed.

Once we finish, the results obtained from our computational workflow scripts will be compared with those generated by FoldX to evaluate consistency and methodological performance.

Materials and Software

For the following analysis we used several softwares and python packages.

Python	● Bio.PDB in order to read and store the pdb files and extract the sequence as well as some needed information. ● biobb_structure_cheking to fix and predict all the necessary adjustments that we need to do to the original pdb file. Changes such as adding hydrogens or removing alternative locations. ● Numpy and math libraries allow us to do the needed equations to compute dielectric, epsilons and more. ● Matlplotlib , pandas and seaborn were used to plot the results and allow to see the results more clearly
Software	● We used NACESS in order to compute the surface area of the complex and both chains. ● Foldx was used to compare and ensure the results were well directed in an approximated way.

Methodology

Structure Retrieval and Preparation

The first step involved locating the appropriate protein structure in the PDB and confirming that the selected biological assembly matched the functional oligomeric form. Heteroatoms and chains not required for the analysis were removed. The structure was then refined using the *biobb_structure_checking* module, which added missing side chains, hydrogen atoms and atom charges, ensuring the system was suitable for subsequent energy calculations.

Interface Definition

An initial visual inspection of the protein-protein interface was performed in PyMOL to determine a suitable distance threshold that captured all relevant contact residues. An additional 1-2 Å was added to include neighbouring residues likely participating in the interface. Using this distance cutoff, a Python script was implemented to generate the list of interface residues for each chain, defining the interface for further analysis.

Interaction Energy Calculation

Interaction energies between the two chains were calculated as the difference between the total energy of each chain in the bound complex and in the unbound state. It was assumed that no significant structural rearrangements occur upon binding, allowing bonded terms and intra-chain nonbonded interactions to be omitted, considering only inter-chain contributions.

Under this approximation, the interaction energy for the A-E complex was defined as:

$$\Delta G^{A-E} = \Delta G^{A-E}_{elect} + \Delta G^{A-E}_{vdw} + \Delta G^{A-E}_{solv} - \Delta G^{A}_{solv} - \Delta G^{E}_{solv}$$

Solvation terms were calculated from atomic solvent-accessible surface areas (ASA) for all atom types. Calculations were performed using both the full set of residues and only those identified as part of the interface. The interface was redefined as necessary to capture as much of the interaction energy as possible, and the distance cutoff was modified accordingly. Summary plots and tables were produced to illustrate the key energetic contributions and emphasise the main stabilising forces, such as electrostatic interactions, van der Waals forces, hydrogen bonds, salt bridges and hydrophobic contacts.

Alanine-Scanning Analysis

To quantify the contribution of individual interface residues to the complex stability, an alanine-scanning approach was applied. For each residue, the effect of replacing the side chain with that of alanine was assessed by removing all additional side-chain atoms (since alanine represents the minimal backbone-substituent, except for glycine). The resulting changes in ΔG^{A-E} were plotted to highlight residues that strongly destabilize the interface, and the results were interpreted in relation to the physicochemical nature of the original amino acids.

Structural Visualization

PyMOL images were generated to highlight key residues and their interactions. These visualizations provided structural context to the energetic analyses and facilitated identification of residues critical for binding.

Variant Analysis

The main sequence variants that influence the receptor-binding domain (RBD) were identified, including the mutations found in the Alpha, Beta and Delta SARS-CoV-2 strains. The effects of these naturally occurring mutations on the interaction energy were evaluated. Mutated structures were generated using either PyMOL mutagenesis tool or *biobb_structure_checking*, and the interaction energies were recalculated to evaluate how these evolutionary substitutions influence the stability of the molecular interface.

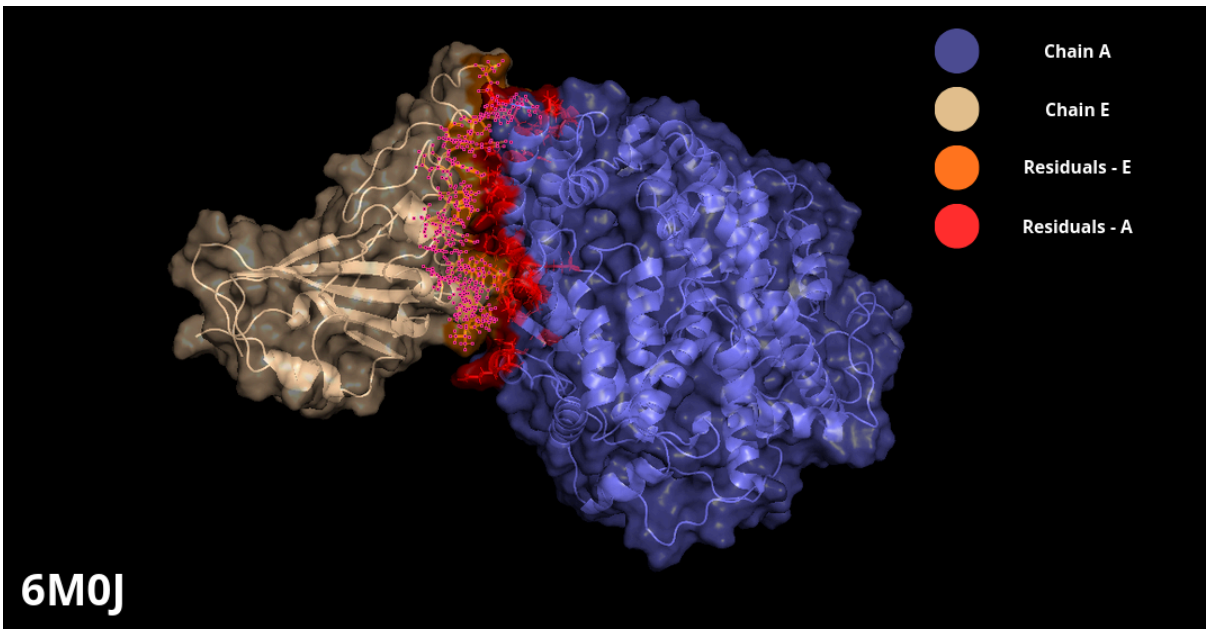
FoldX Calculations

FoldX is a computational tool designed to estimate the energetic impact of mutations and structural changes in protein complexes by using an empirical force field that evaluates contributions such as van der Waals interactions, electrostatics and solvation.

Results

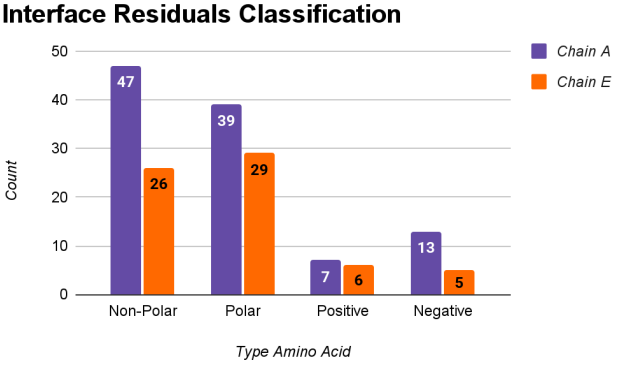
Before interface definition, the structural data were prepared using a standardized preprocessing workflow. The target complex was retrieved from the PDB, and its biological assembly was examined to retain only the chains involved in the functional assembly. All heteroatoms were removed to prevent artefactual contacts. Structural quality was then

evaluated, and missing side chains, hydrogen atoms, and atomic charges were added using the biobb_structure_checking module, resulting in a fully consistent, simulation-ready model.



With the curated structure in place, the interface between chains A and E was identified by selecting residues involved in at least one interatomic contact below a visually optimized distance threshold. Following inspection in PyMOL, this cutoff was expanded by 1-2 Å to include neighboring residues that contribute indirectly to the interaction surface.

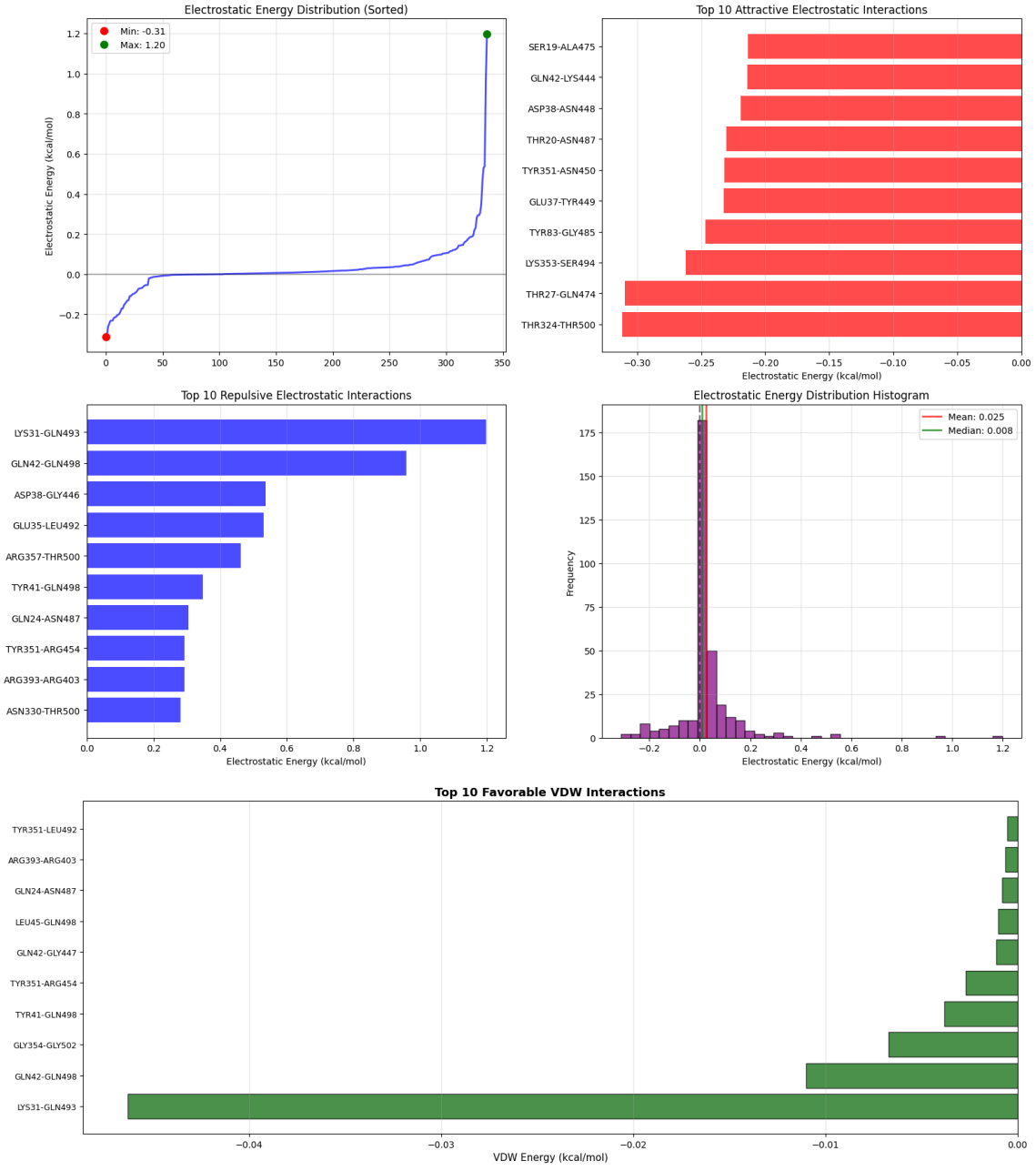
Analysis of amino acid composition at the interface reveals notable differences in the chemical properties of the two chains. Chain A contains a high proportion of non-polar residues (47 in total), suggesting that hydrophobic packing plays a major role in stabilizing its contribution to the interaction. In contrast, chain E presents a smaller hydrophobic core and a relatively higher proportion of polar and special residues, which may introduce flexibility or localized structural constraints at the interface. Both chains contain comparable numbers of positively and negatively charged residues, indicating that electrostatic interactions are present but are not the dominant stabilizing force compared to hydrophobic forces. Overall, the interface appears to rely on hydrophobic contacts from chain A combined with a more heterogeneous, polar environment from chain E, producing a complementary surface that supports a stable and specific protein-protein interaction.

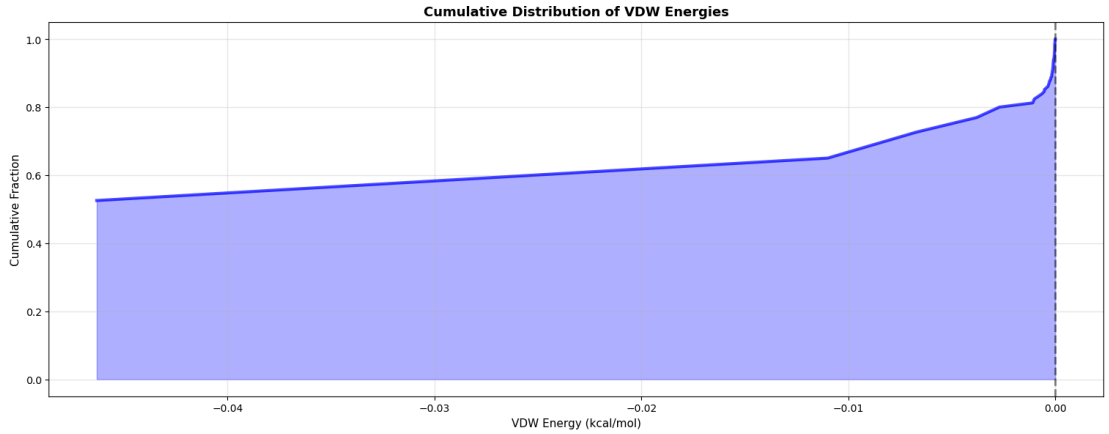


Afterwards, we computed the interaction energies between the two chains. The calculations included the electrostatic, van der Waals, and solvation contributions. The results are summarized in the following table, with the total interaction energy presented in the final column. A cutoff of 10 Å was used to obtain these results.

ΔG^{A-E}_{elect}	ΔG^{A-E}_{vdw}	ΔG^{A-E}_{solv}	ΔG^{A}_{solv}	ΔG^{E}_{solv}	ΔG^{A-E}
-16.287	-8.667	27.928	8.424	33.193	**-39.4589**

Subsequently, we prepared a table highlighting the most relevant interaction energies and identified the specific interactions across the interface that contribute to them, classifying each by interaction type.

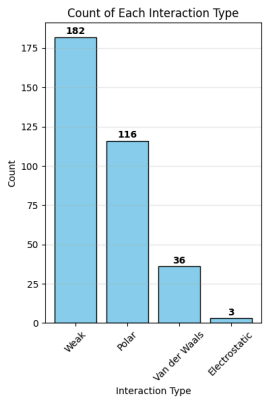


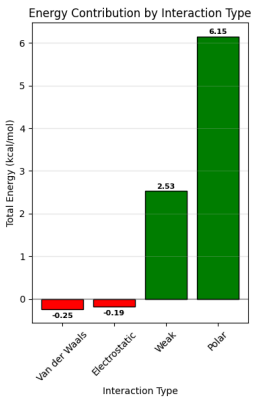


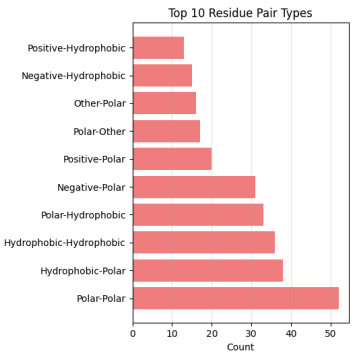
Afterwards, we plotted the interactions in the interface with more total energy. We can see how the two interactions that are favorable are having higher electrostatics than van der Waals. This match with the idea that the interactions are Mixed but weak because the values that we are getting are low.

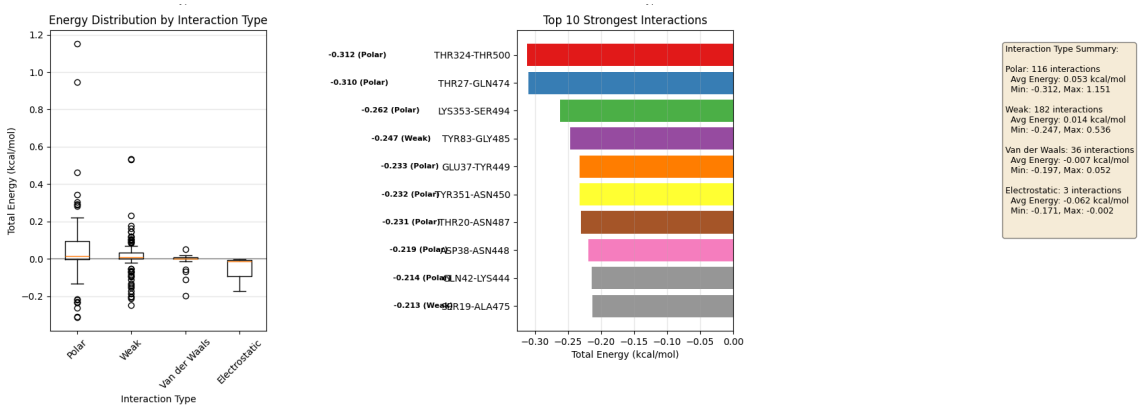
	Rank	ACE2_Residue	RBD_Residue	Electrostatic_Energy	VDW_Energy	Total_Energy	Favorable	Interaction_Type	Strength
0	1	LYS31	GLN493	1.198	-0.046	1.151	No	Mixed Weak	Moderate
1	2	GLN42	GLN498	0.958	-0.011	0.947	No	Mixed Weak	Weak
2	3	ASP38	GLY446	0.536	-0.000	0.536	No	Mixed Weak	Weak
3	4	GLU35	LEU492	0.530	-0.000	0.530	No	Mixed Weak	Weak
4	5	ARG357	THR500	0.461	-0.000	0.461	No	Mixed Weak	Weak
5	6	TYR41	GLN498	0.349	-0.004	0.345	No	Mixed Weak	Weak
6	7	THR324	THR500	-0.312	-0.000	-0.312	Yes	Mixed Weak	Weak
7	8	THR27	GLN474	-0.310	-0.000	-0.310	Yes	Mixed Weak	Weak
8	9	GLN24	ASN487	0.304	-0.001	0.303	No	Mixed Weak	Weak
9	10	ARG393	ARG403	0.294	-0.001	0.293	No	Mixed Weak	Weak

And the following plot makes a summary of the classification of the interactions we are analyzing. We easily see how the energy contribution by interaction type and the count of each interaction type is clearly dominated by the Weaks and Polars. The top 10 residue pair shows how the top 3 were Polar-Polar interactions, Polar-Hydrophobic and Hydrophobic-Hydrophobic. For the top 10 strongest interactions (favorable means energy < 0) we can see the top 10 how they are mostly polar.

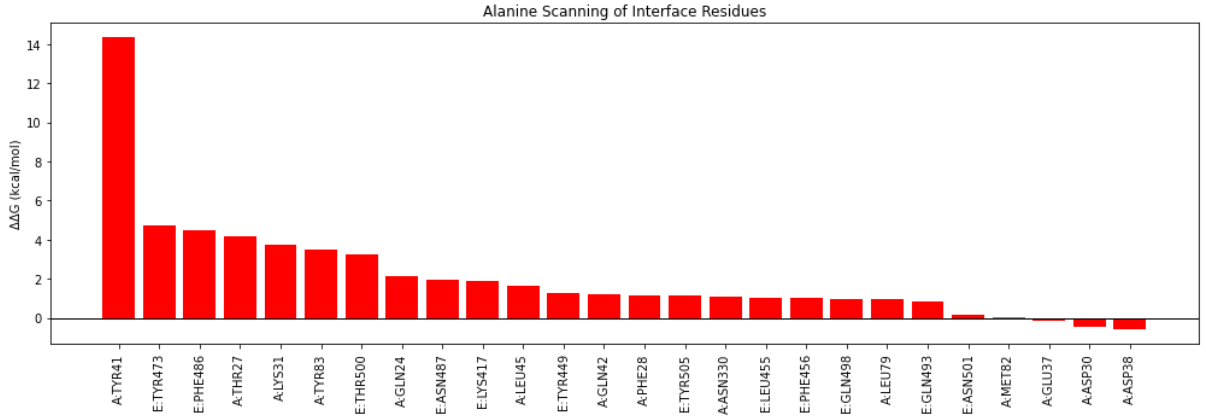








To further investigate, we assessed the impact of replacing each interface residue with alanine on the overall interaction energy previously computed, the ΔG^{A-E}. We hypothesized that alanine substitution at interface hotspots would increase ΔG^{A-E} due to the loss of electrostatic contacts and van der Waals packing, while the compensatory solvation gain would be smaller. For non-critical residues, the changes were expected to be modest or negligible. The following plots summarize the stability changes across the interface. We want to highlight that since the alanine side chain is part of all other residues except glycine, there is no need to perform the replacement explicitly.

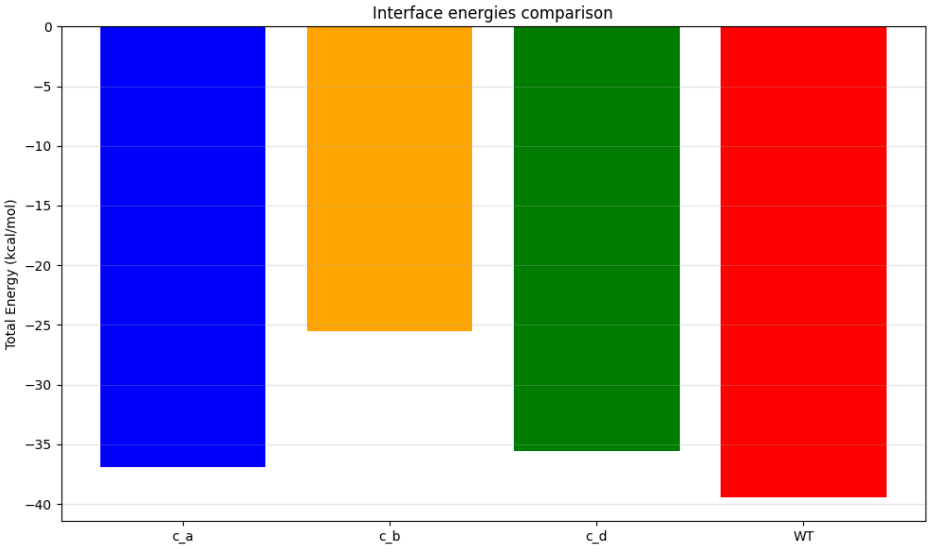


Most residues exhibited modest ΔΔG values, indicating that their mutation had limited energetic impact. In contrast, A:TYR41, showed a notably elevated ΔΔG value, of about 14 kcal/mol, which reflects its key role in stabilizing the protein-protein interface. The remaining residues contributed less substantially to the binding energy, suggesting a more peripheral or supportive role in the interaction.

We identified the most relevant sequence variants for the receptor-binding domain (RBD), including the Alpha, Beta, and Delta SARS-Cov2 strains, among others. For each variant, we analyzed the effect of the specific mutations on the interaction energy at the interface. Structural models were generated by replacing the side chains in PyMOL, ensuring proper geometry and atomic charges.

After repeating the whole energetic analysis that we previously conducted, we obtained the following table.

Variant	ΔG^{A-E}_{elect}	ΔG^{A-E}_{vdw}	ΔG^{A-E}_{solv}	ΔG^{A}_{solv}	ΔG^{E}_{solv}	ΔG^{A-E}
Alpha	*-16.943*	*-9.3145*	*41.677*	*23.460*	*28.859*	***-36.90***
Beta	*-6.466*	*-10.053*	*43.270*	*23.460*	*28.847*	***-25.55***
Delta	*-16.286*	*-8.666*	*41.664*	*23.460*	*28.847*	***-35.59***



Looking at the comparison we easily see how the variants Alpha and Delta are having an interface total energy similar or close enough to the WT than the beta one.

For additional checking we used the FoldX with the aim to explore the results and compare them. Beside the numbers making sense to us, we could have made errors on the parameters used that could bias our results and conclusions. Here are the results for all the files:

Variant	ΔG^{A-E}_{elect}	ΔG^{A-E}_{vdw}	ΔG^{A-E}_{solv}	ΔG^{A-E}
WT	***-2.39***	***-14.98***	***-18.16***	***-7.77***
Alpha	*-2.28*	*-15.63*	*-19.07*	***11.15***
Beta	*-1.43*	*-15.29*	*-18.76*	***11.15***
Delta	*-2.42*	*-14.98*	*-18.16*	***-7.82***

Discussion

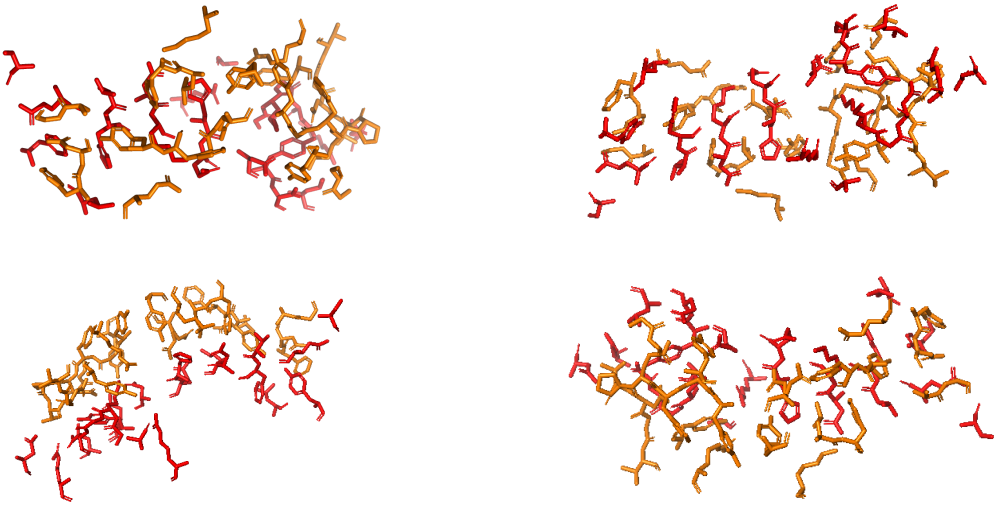
The preprocessing and interface definition workflow ensured that the interaction analysis was based on a structurally consistent and biologically relevant model. By curating the PDB structure, removing heteroatoms, restoring missing atoms, and standardizing protonation and charge states, we minimized artefacts that could obscure the true determinants of the interface. The strategy used to define the interface between chains A and E balanced geometric precision with biological realism. Starting from direct atomic contacts and expanding the distance cutoff allowed us to identify both primary contact residues and secondary contributors that stabilize the interaction indirectly. The resulting residue sets per chain indicate that the interaction is not limited to a small hotspot but spans a broad, chemically diverse surface. This diversity, including polar, charged, and aromatic residues, suggests a multifaceted binding mechanism in which electrostatic steering, hydrogen bonding, and van der Waals packing all play a role. Such an extended interface is characteristic of stable, functionally relevant protein-protein complexes rather than transient encounters, reinforcing the structural significance of the A-E interaction.

The interaction energy analysis, using a 10 Å cutoff, revealed a stabilizing association between chains A and E, with a total interaction energy of -39.46 kcal/mol. Both the electrostatic (-16.29 kcal/mol) and van der Waals (-8.67 kcal/mol) terms contribute favorably to complex formation, indicating that charge complementarity and close-range packing play a central role in stabilizing the interface. As expected for MM/PBSA calculations, the solvation component is positive (Gsolv A-E = 27.93 kcal/mol), reflecting the energetic cost associated with desolvating the interacting surfaces. This penalty is partially offset by the individual solvation energies of chains A and E, which together modulate the net desolvation effect. Overall, the balance of these contributions shows that the stability of the WT complex arises primarily from direct electrostatic and van der Waals interactions at the interface, which are sufficiently strong to overcome the moderate solvation penalty.

To assess the contribution of individual residues, we performed alanine scanning. Results reveal a heterogeneous energetic landscape across the interface. In particular, residue A:TYR41 showed a notably elevated ΔΔG value of approximately 14 kcal/mol, indicating that it plays a critical role in stabilizing the complex: This strong effect suggests that TYR41 is involved in key stabilizing interactions at the interface, such as hydrogen bonding, aromatic stacking, or electrostatic contacts. Mutation of this residue to alanine disrupts these interactions, leading to substantial destabilization. In contrast, most other residues exhibited modest or negligible ΔΔG values, consistent with roles in surface complementarity rather than dominant energetic contributions. This pattern, where a small number of energetic

hotspots are embedded within a broader supportive scaffold of residues identified by geometric proximity, highlights the partial overlap between distance-based and energy-based interface definitions. Such behavior is characteristic of biologically relevant protein-protein interfaces and may guide future mutagenesis or rational design efforts.

Finally, we visualized the interface in PyMOL, highlighting residues according to their contribution to interaction energy. Hydrogen bonds and salt bridges were depicted as sticks, allowing us to map the spatial distribution of key residues and visually confirm the correspondence with the energetic analysis from previous sections.



A comparison of the WT RBD-ACE2 interaction energies with those of the Alpha, Beta, and Delta variants reveals that the WT complex remains the most stable, with a total interaction energy of -39.46 kcal/mol. This strong stabilization arises from balanced electrostatic (-16.29 kcal/mol) and van der Waals (-8.67 kcal/mol) contributions, partially compensated by the expected desolvation penalty (Gsolv A-E = 27.93 kcal/mol). All three variants show weaker binding, with Alpha (-36.90 kcal/mol) and Delta (-35.59 kcal/mol) remaining relatively close to the WT, while Beta displays a markedly reduced affinity (-25.55 kcal/mol). The main differences across variants stem from shifts in the electrostatic term, which is strongly destabilized in Beta (-6.47 kcal/mol) compared to WT and the other variants, whereas van der Waals contributions remain similar. Solvation penalties are consistently higher in all variants (41-43 kcal/mol) relative to the WT, further contributing to their reduced stability. Overall, the data indicate that the mutations primarily modulate electrostatic interactions and solvation effects, leading to a cumulative decrease in binding strength relative to the WT.

The energetic differences can be directly linked to the specific mutations introduced in each variant. Alpha (N501Y) introduces a bulky aromatic residue at position 501, which increases local hydrophobicity and reduces the extent of desolvation, explaining its moderate destabilization (GA-E = -36.90). Beta (K417N, E484K, N501Y) combines three mutations

that alter both charge and polarity at key interface positions: K417N removes a positive charge, E484K introduces a new positive charge, and N501Y adds an aromatic ring. This combination disrupts the electrostatic landscape and hydration pattern more severely than in other variants, resulting in the weakest binding (GA-E = -25.55). Delta (L452R, T478K) introduces two positively charged residues at positions 452 and 478, which modifies the local electrostatic environment but preserves packing interactions, leading to an intermediate destabilization (GA-E = -35.59). Overall, these results demonstrate that even single-residue substitutions can significantly reshape the solvation and surface complementarity of the RBD-ACE2 interface, ultimately modulating binding affinity through subtle but cumulative energetic effects.

To further validate our computational workflow, we used FoldX as an independent method to re-evaluate the interaction energies of the WT RBD-ACE2 complex and its Alpha, Beta, and Delta variants. Although the values obtained with our own code were internally consistent and aligned with the expected energetic trends, there was still a possibility that parameter choices, structural preparation steps, or implementation details could introduce bias into our results. FoldX provides an empirical force-field-based framework with standardized parameters for electrostatics, van der Waals interactions, and solvation, making it a suitable tool for cross-checking our calculations.

Overall, the FoldX results serve as an external validation step, confirming that the general energetic patterns observed with our method are robust despite differences in absolute values. This comparison strengthens confidence in our conclusions regarding the impact of specific mutations on the RBD-ACE2 interface.

Conclusion

In summary, this study provides a comprehensive evaluation of how individual interface residues contribute to the interaction energy within the RBD-ACE2 complex, both in the wild-type and in key SARS-CoV-2 variants. Through rigorous structural preprocessing, distance-based interface definition, energetic decomposition, and alanine scanning mutagenesis, we identified the molecular determinants that most strongly stabilize the interaction. Although the interface is broad and chemically diverse, only a limited subset of residues functions as energetic hotspots, exerting a disproportionate influence on binding through hydrogen bonding, aromatic contacts, and electrostatic interactions. Extending this analysis to the Alpha, Beta, and Delta variants revealed that specific mutations can significantly reshape the energetic balance of the interface, primarily by altering solvation and surface complementarity rather than disrupting direct physical contacts. These

variant-induced changes highlight the sensitivity of protein-protein interactions to subtle sequence modifications and underscore the importance of local chemical environments in determining binding affinity. An independent analysis performed using FoldX yielded systematically different absolute energy values, reflecting methodological differences between approaches; however, it reproduced the same qualitative trends across variants. This consistency supports the robustness of the observed energetic patterns while emphasizing that FoldX serves as a qualitative basis rather than a direct quantitative validation. Overall, this integrative approach not only clarifies the energetic architecture of the RBD-ACE2 interface but also establishes a robust framework for predicting, comparing, and potentially modulating protein-protein interactions through targeted mutagenesis or rational design.

References

1. Lan, J., Ge, J., Yu, J. et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 581, 215-220 (2020). PDB ID: 6M0J.

2. BioExcel. biobb_structure_checking: A Python library for structure validation and repair. GitHub repository. Available at: https://github.com/bioexcel/biobb_structure_checking Accessed: [23/11/2025].

3. Schrödinger, L., & DeLano, W. (2020). PyMOL. Retrieved from http://www.pymol.org/pymol

Comments

The project was developed collaboratively by Alma de María López, Lucía Navarro and Izan Torrero. Izan Torrero was responsible for cleaning and preprocessing the data, while Alma de María López and Lucía Navarro were responsible for writing the report and documenting the procedures used. The steps following pre-processing and the final report were achieved through the coordinated efforts of all team members.