

# Energy analysis exercise 2025-26

## Spike RBD-ACE2 Protein-protein interface analysis

**Objective:** To evaluate the relative contribution of interface residues to the interaction energy in a protein-protein complex. Case RBD-ACE2

**Background:** The growing amount of genomics data given by the abundance of large-scale sequencing projects is releasing large amounts of protein sequence variants. Most proteins are known to work forming complexes either permanent or transient with other macromolecules. In most cases such complexes are key for the regulation of the activity of the proteins involved. There is an increasing interest in understanding the effect of sequence variants on the stability of such complexes. As an initial approach to that analysis, we pretend to evaluate the relative contribution of the amino acid residues forming the protein-protein interface in the interaction energy between the complex components.

As a case study we will concentrate on the complex between the Receptor Binding Domain (RBD) of SARS-Cov-2 Spike protein and its receptor, the Angiotensin Converting Enzyme (ACE2). The formation of this complex is a key step in the viral infection. Most vaccines in the market try to block this binding process. Analysis will be done on the structure **6m0j** from the PDB. This structure contains only the RBD domain from Spike.

**Strategy:** We will evaluate the contribution of individual amino acid residues through the following steps:

1. Determine which amino acid residues form the interface between the complex components (use 6m0j)
2. Determine the contribution to the stability of the complex using two strategies:
  - a. Evaluate contribution of each residue of the interface to the interaction energy of the complex (Corresponds to the  $\Delta G$  of moving one chain from solution to the complex)
  - b. by mimicking a traditional Ala-scanning experiment, i.e. replacing each residue in turn with Ala and evaluating the changes in the complex interaction energy.
3. 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.
4. Perform the same analysis using FoldX (<http://foldxsuite.crg.eu/>), and compare results

## Methodology:

### Step 1.

The interface between can be defined by a list of residues on both chains that have at least one atom below a given distance.

1. Using pymol inspect visually the structure and choose a suitable distance in the way that all contact residues are included. Add 1-2 Å to that distance so the adjacent residues are also considered.
2. Prepare a python script to define the list of interface residues on each chain
3. Setup the initial protein structure as necessary
  1. Obtain the required structure from the PDB.
  2. Check at PDB which is the composition of a “Biological assembly”. Remove all chains but those involved in the assembly, if necessary
  3. Remove all heteroatoms.
  4. Perform a quality checking on the structures, and add missing side-chains, hydrogen atoms and atom charges (use CMIP settings and prepare a PDBQT file), using the [biobb structure checking](#) module

### Step 2

Interaction energy between chains corresponds to the difference between the total energy of each chain in the bound state (the complex) and the unbound state (the two chains are isolated in solution). To simplify the calculation, we will take the following approximations:

- We will assume that 3D structure does not change between bound and unbound states, hence bonded terms and non-bonded terms between atoms belonging to the same chain will not change and will not be considered.
- We will assume that the effects of heteroatoms not related to the interface are negligible
- We will consider solvation energies obtained from ASA values for all atom types (not only hydrophobic ones).

With these assumptions prepare a python script or notebook to evaluate the Interaction energy among chains. Interaction energy between components of an A-B complex will come from the following:

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

This can be done with all residues or only with the interface residues. Check that the interface from Step 1 represents most of the energy involved. If this is not the case modify the cut-off distance accordingly.

Prepare the necessary plots to show the results, and a table with the most relevant interactions energies.

Identify the interactions across the interface responsible for such energies and classify the type of interaction.

*Optional:*

- Derive and implement an algorithm to define the interface from the interaction energy instead of distance based.
- Replace the surface-only solvation by the PBSA approach (Poisson Boltzman for the electrostatic solvation and surface for the hydrophobic one)

### **Step 3**

Determine the effect of replacing each interface residue by Ala in the overall  $\Delta G^{A-B}$ , and make a plot of the results obtained, highlighting those residues that are more relevant for the stability of the interface. Discuss the results obtained in relation to the nature of the amino acids involved (hint, as Ala sidechain is part of all others, except Gly, there no need to do the replacement, just take into account the different atoms)

### **Step 4**

Prepare images with pymol of the interface highlighting the relevant residues and interactions.

### **Step 5**

Find the most relevant sequence variants for RBD (alfa, beta, and delta SARS-Cov-2 strains, etc.) and analyze the effect of the actual mutations on the energy. Build structure with the replaced sidechains, either using pymol or using biobb\_structure\_checking.

## General hints

- You can prepare all functionality in a single script or (probably better), prepare a series of reusable scripts that run in a pipeline. Alternatively, you can prepare all steps in a Jupyter notebook
- For standalone scripts make sure that they have a proper command line in a way that can be re-used with different input PDBs or parameters.
- All output should be properly formatted and human readable.
- Prepare re-usable functions for calculating energy components
- Report preparation, comment on the following questions:
  1. Which residues form the complex interface
  2. Which is the correlation between the interface residues obtained from a distance criterium and those obtained from a energy criterium
  3. Plot energies (interaction, solvation, etc.) for the interface residues
  4. Is the Ala scanning experiment providing the same information as direct interaction energies?
  5. Can we identify specific interactions (electrostatic, Hbond, hydrophobic) that are relevant to the complex stability?
  6. How the direct analysis compares with the equivalent analysis done with FoldX?

Include plots and images as necessary.

Discuss the validity of the approximations assumed in step2

## Setup: Parameters and energy calculations

- Parameters for electrostatic, var der Waals and solvation can be taken from [https://github.com/jlgelpi/BioPhysics/tree/master/Energy\\_analysis\\_project/](https://github.com/jlgelpi/BioPhysics/tree/master/Energy_analysis_project/)
- The repo provide two python scripts:
  - forcefield.py: simple example interface to read VDW parameters
  - basic\_setup.py: example script to setup the system using PDB and PDBQT files as input
- Structure checking and management can be done with the biobb\_structure\_checking python module either from PiPy or BioConda distributions:
  - conda install -c bioconda biobb\_structure\_checking (preferred)
  - pip3 install biobb\_structure\_checking
  - Documentation at <https://biobb-structure-checking.readthedocs.io/en/latest/>
- Surface areas will be calculated using NACCESS though Biopython (instructions at Biophysics github repo)
- Electrostatic energies will be calculated using the Mehler-Solmajer dielectric.

## Equations

Vdw interaction:

$$E_{vdw_{ij}} = 4\epsilon_{ij} \left( \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right)$$

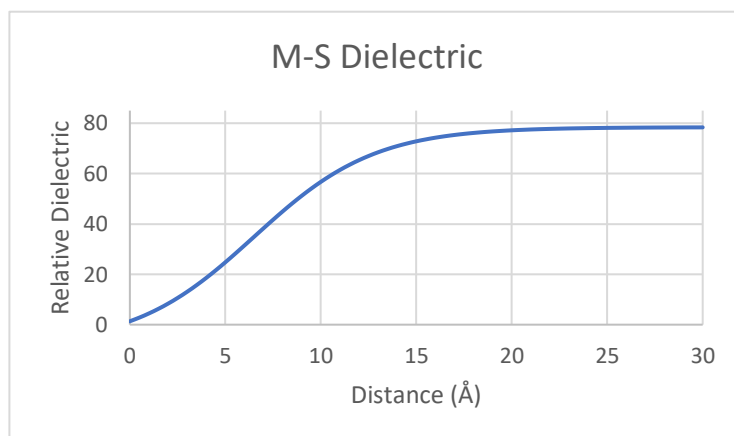
**Note:** Consider [Combination Rules for vdw parameters](#)

Electrostatic interaction:

$$E_{elec_{ij}} = 332.16 \frac{q_i q_j}{\epsilon r_{ij}}$$

Mehler-Solmajer Dielectric

$$\epsilon_r = \frac{86.9525}{1 - 7.7839 e^{-0.3153 r}} - 8.5525$$



Solvation:

$$\Delta G = \sum_{AtTypes} \sigma_i ASA_i$$

**Note:** We will use only this term for solvation, including also non-hydrophobic atoms

## Software and Materials

Modelling, visualization: PyMol, Chimera

Scripting: Python (v3), BioPython modules. Other python modules for graphics and maths. Text Editor. Jupyter notebook optional

Other software: NACCESS (copy and instructions at the repository)

Structure setup: Biobb\_structure\_checking python library.

Software Version Control: Git

## Evaluation

The exercise will be done in groups of **3 people** and will be evaluated from a written report (PDF, uploaded to Atenea) including the analysis results, plots and the necessary figures. Code or notebook should be uploaded or made available in a git repository.

The report should describe the operation performed, the results of the analysis and a critical discussion of the results (See hints above)

AI helper tools can be used to prepare code, but code and methodology should be fully explained in the report.