Engenharia Computacional de Proteínas

Instrutores: Roberto Lins, Danilo Coêlho, Elton Chaves
Tutorial originalmente elaborado por Matheus Ferraz; revisado por Danilo Coêlho e Roberto Lins

Tackling the infectivity potential of SARS-CoV-2 variants through computational chemistry

The fight against COVID-19 pandemic has brought the scientific community together into a common purpose. As part of the response, numerous computational chemists and biophysicists have gathered their efforts to gain insight into the molecular aspects of the SARS-CoV-2 infection. To this end, a number of computational tools have aided the understanding of how the viral molecules interact with each other and with the human host, allowing for a broader picture of interventions to mitigate the disease, as well as the search for small molecules that may inhibit viral proteins.

Calculating binding strength of protein-protein interaction upon amino acid mutations

At a molecular perspective, the entry of the virus in the host cell is mediated by the spike (S) protein, a homotrimer glycoprotein found in the SARS-CoV-2 virion. The S protein is known as the primary antibody target in the context of a natural infection and mediates the membrane fusion and receptor recognition of the virus through binding the angiotensin converting enzyme (ACE2) receptor in the host cell (Figure 1A). The S protein is comprised of two subunits: S1 and S2. The S1 subunit at the N-terminal region is responsible for virus attachment and contains the receptor-binding domain (RBD), which directly binds to ACE2 (Figure 1C). During the fusion, the S protein undergoes large conformational changes, especially around the RBD, in which two major conformations are observed, as from cryo-EM results: the down state (shielded from receptor binding), and the up state (receptor accessible) (Figure 1B). A plethora of neutralizing antibodies targeting the RBD of SARS-CoV-2 and related coronaviruses have been identified.[1]

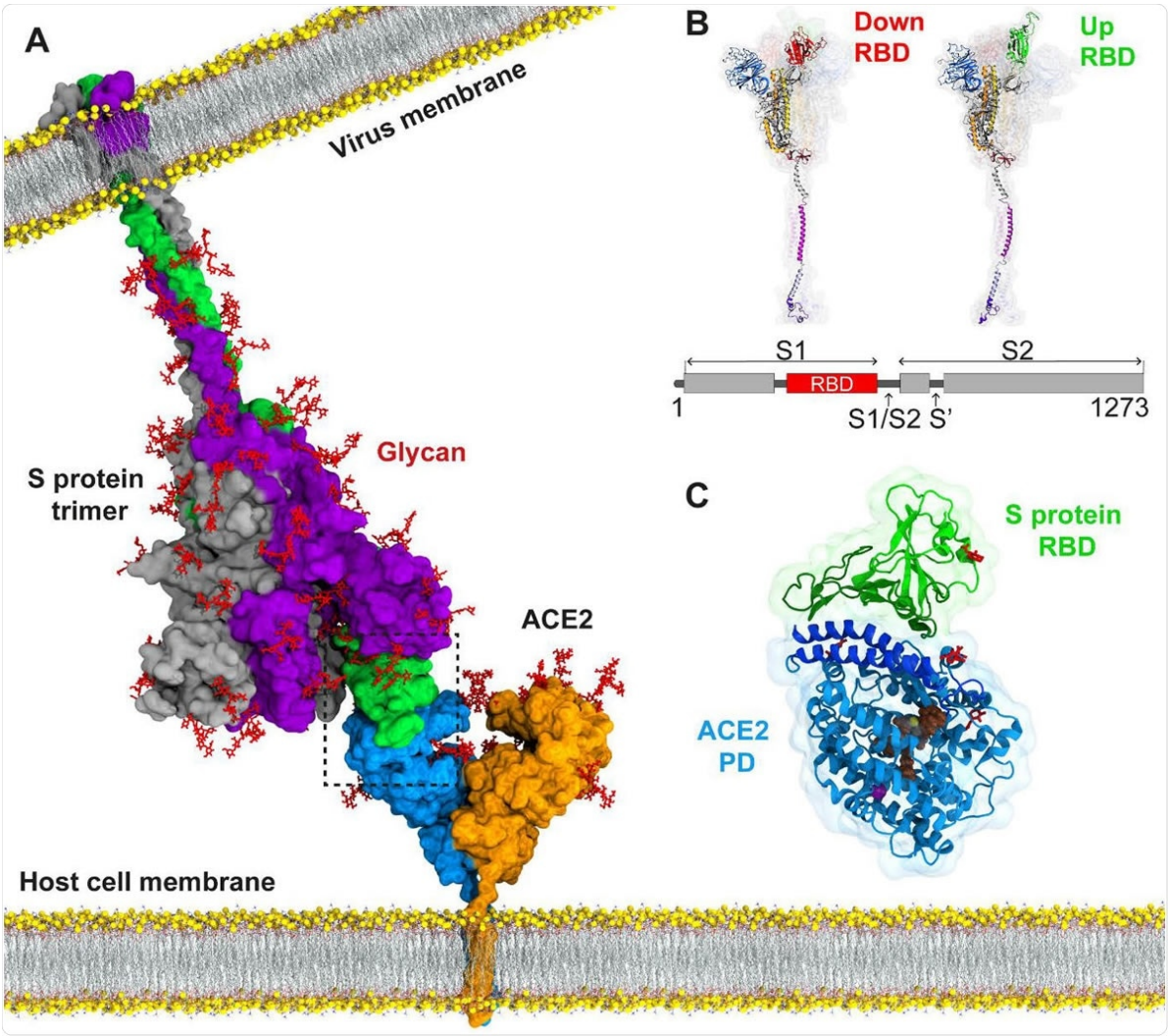

Figure 1. Model for the binding between SARS-CoV-2 S protein and the human ACE2 receptor. Full-length S protein complexed with ACE2. S protein is a homotrimer (green, purple, gray), incorporated to the viral membrane. Source: Taka et al. (2020).[2]

The ability of coronaviruses to infect humans is associated with their binding strengths to the ACE2 protein. It has been reported a number of SARS-CoV-2 mutations that influences the viral infectivity or increase on transmission rate. To better understand the physical-chemistry interplay of how these mutations impact at a molecular level, in this tutorial, we will calculations to assess the stability and binding affinity upon mutations.

Figure 2 shows the evolution of SARS-CoV-2 mutations in the RBD until June of 2022. Red lines represent mutations that strengthen the infectivity, while blue lines represent the mutations that weak the infection potency. This result suggests that evolution shapes novel variants more infectious.

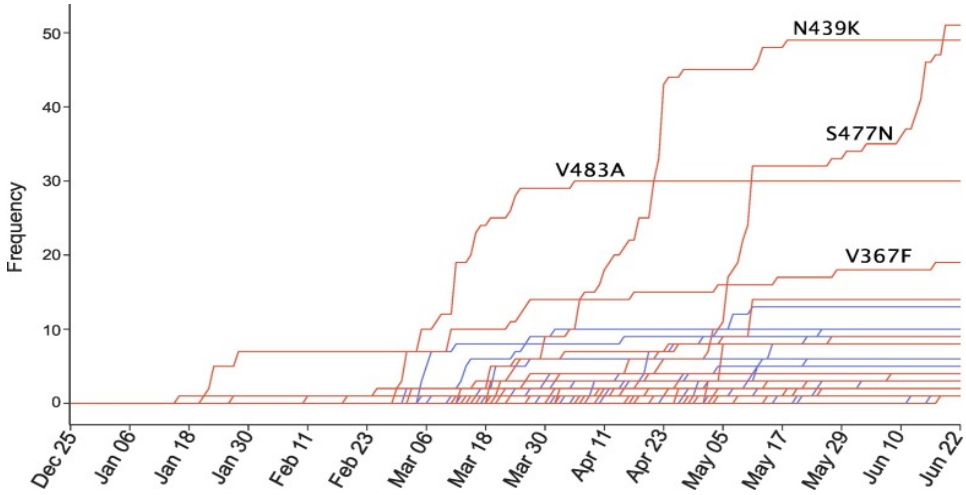

Figure 2 - Time evolution of SARS-CoV-2 mutations.

Let's analyze how the most prevalent mutation (N439K) impacts the thermodynamics stability and of binding of the RBD. In addition, we will evaluate the mutations Q493A, and N501Y, the latter was recently observed in the United Kingdom raising concern worldwide regarding its increased transmissibility.

Download FoldX (This part may be skipped)

Download the file **foldxLinux64.tar_.gz** and copy it in a directory of your choice (**the path to this final directory must be known for the next steps)**.

Extract the file.

FoldX does not require installation! The program itself is a self-contained binary containing the executable and the rotabase.txt.

Set the foldx binary into your bash environment by typing:

```
echo "alias foldx=<path_to_foldx_dir>/foldx" >> ~/.bashrc
```

where **<path_to_foldx_dir>** must be **replaced** by the path to the directory where you extracted the binary earlier.

Finally, type:

```
source ~/.bashrc
```

Type `foldx` and make sure everything is working fine.

1. Assessing the impact of the mutations in the folding stability of SARS-CoV-2 RBD

Before starting the tutorial, you must download the file **ECP_TUTORIAL_FoldX.tar.gz**. Extract it in a directory that you know the path and cd (change directory) into it in the Terminal command line.

1.1. Geometry-optimize the structure of the RBD.

Frequently, the structures in the PDB present structural inconsistencies, such as small steric clashes between the atomic radii, improper orientation of sidechains such as asparagine, glutamine or histidine, or even dihedral's angles outside of acceptable values. Thus, FoldX fixes these errors and lowers the global energy (ΔG). To this end, the Repair application is used.

In the Terminal go to folder **ECP_TUTORIAL_FoldX/task1** (use command **cd**)

Type the command below, as a single line, in the terminal (**do not copy from PDF, characters may be replaced, and an error message will be prompted!**):

```
foldx --command=RepairPDB --pdb=rbd.pdb --ionStrength=0.05 --pH=7
--vdwDesign=2 --pdbHydrogens=false --water=predict
```

Compare the total energy of the protein prior and subsequently to the Repair step:

Task 1

An output file **rbd_Repair.fxout** will be written.

Type in terminal:

```
head -2 rbd_Repair.fxout
```

The value for the total score of the initial structure is the first one accompanying the Starting Structure label:

```
Starting Structure  37.2203  -74.5226  -26.8149
```

Following, type in terminal:

tail -1 rbd_Repair.fxout

The value for the total score of the final and repaired structure is the one accompanying the last residue (PROE527).

```
PROE527  -5.42987 -83.8253 -47.9016
```

Answer the following questionnaire:

Fill the gaps:

1.1 Energy Prior to the repair application: _____ kcal/mol.

1.2 Energy Subsequently to the repair application: _____ kcal/mol.

1.3 Was the change in the total energy expected? Why?

2. Perform the mutations using the BuildModel application.

To perform the mutations, it's necessary to have a mutational list file (herein termed as individual_list.txt). This file contains a single line (*e.g.* NA439K, which means we are going to replace the residue N 439 from chain A to a K residue). To assure convergence, we will set –numberOfRuns=5, because some residues have many rotamers, and thus, it requires more calculations to achieve more reliable results.

Go to folder **ECP_TUTORIAL_FoldX/task2/N439K** (use command **cd** with appropriate destiny path)

Copy the repaired pdb from previous step, which is in the task1 directory:

```
cp ../../task1/rbd_Repair.pdb .
```

Now, perform the mutation by typing the following command, as a single line, in the terminal (**do not copy from PDF, characters may be replaced, and an error message will be prompted!**):

```
foldx --command=BuildModel --pdb=rbd_Repair.pdb
--mutant-file=individual_list.txt --ionStrength=0.05 --pH=7 --water=predict
--vdwDesign=2 --pdbHydrogens=false --numberOfRuns=5
```

This procedure will provide the free energy of the mutant (here, we will call it ΔG_{mut}).

The difference in free energy (ΔΔG) is given by ΔG_{mut}-ΔG_{wt} and can be calculated for each run by reading the file: "Raw_rbd_Repair.fxout". If you are curious, you can open it with a text editor (e.g., vi, emacs, pico).

However, we will use as the reference value the average of each run through the following file: Average_rbd_Repair.fxout. The values we are interested are the two first values, corresponding to the Standard deviation (SD) and the average ΔΔG, respectively.

Type in terminal:

head Average_rbd_Repair.fxout

The values for SD and ΔΔG are the two accompanying the rbd_Repair_1 label, respectively:

```
Pdb           SD          total_energy BackboneHbond SidechainHbond
rbd_Repair_1 0.019317      -0.31092 -0.620998  -0.62205
```

The impact of the mutation will be determined based on the value for the ΔΔG, as depicted below:

1. highly stabilising (ΔΔG < −1.84 kcal/mol);
2. stabilising (−1.84 kcal/mol ≤ ΔΔG < −0.92 kcal/mol);
3. slightly stabilising (−0.92 kcal/mol ≤ ΔΔG < −0.46 kcal/mol);
4. neutral (−0.46 kcal/mol < ΔΔG ≤ +0.46 kcal/mol);
5. slightly destabilising (+0.46 kcal/mol < ΔΔG ≤ +0.92 kcal/mol);
6. destabilising (+0.92 kcal/mol < ΔΔG ≤ +1.84 kcal/mol);
7. highly destabilising (ΔΔG > +1.84 kcal/mol).

Figure 3 - Free-energy variation classification.

Now go to directories **ECP_TUTORIAL_FoldX/task2/N501Y** and **ECP_TUTORIAL_FoldX/task2/Q493A (**use command **cd** with appropriate respective destiny path**)** and repeat the procedure for the calculations of the ΔΔG of mutations N501Y and Q493A, respectively (**do not forget to copy rbd_Repair.pdb into each directory!)**. Indicate the impact of each mutation based on the average ΔΔG.

Task 2

An output file **Average_rbd_Repair.fxout** will be written for each calculation.

Type in terminal:

head Average_rbd_Repair.fxout

for each output file calculation.

2.1 Fill the following table based on ΔΔG values and classification on Figure 3. Which mutation do you expect to be advantageous?

Mutation	ΔΔG	SD	Impact
N439K			
Q493A			
N501Y			

3. Assessing the impact of the mutations in the binding affinity of SARS-CoV-2 RBD and ACE2 and a neutralizing antibody

Now let's evaluate the impact of the mutations in the binding free energy. The change in binding free energy reflects in the binding affinity.

Go to the directory named "**ECP_TUTORIAL_FoldX/task3/N439K**" (use command **cd** with appropriate destiny path).

The coordinates for the structure for the complex RBD-ACE2 is found in the "***complex_Repair.pdb***" file. **The structure of the complex is already repaired (since the file contains thousands of atoms, it would take a while to perform minimization).**

Now, perform the mutation N439K. First, make sure you are in the **ECP_TUTORIAL_FoldX/task3/N439K** directory.

Type the following command, as a single line, in the terminal (**do not copy from PDF, characters may be replaced and an error message will be prompted!**):

foldx --command=BuildModel --pdb=complex_Repair.pdb --mutant-file=individual_list.txt --ionStrength=0.05 --pH=7 --water=predict --vdwDesign=2 --pdbHydrogens=false --numberOfRuns=5

Upon calculation completion, retrieve the SD and the ΔΔG from the **Average_complex_Repair.fxout file,** as you did previosuly**:**

head Average_rbd_Repair.fxout

Now go to directories **ECP_TUTORIAL_FoldX/task3/N501Y** and **ECP_TUTORIAL_FoldX/task3/Q493A** **(**use command **cd** with appropriate respective path**)** and repeat the procedure for the calculations of the ΔΔG of mutations N501Y and Q493A, respectively.

Finally, go to the **ECP_TUTORIAL_FoldX/task3/K417N-E484K-N501Y-Ab** folder. The coordinates for the structure of the RBD-Antibody complex are found in the ***"complex_Ab_Repair.pdb"*** file. The structure for this complex is also already repaired, due to the high number of atoms in the system.

Perform the simultaneous mutation of three residues: **K417T, E484K and N501Y**. These are mutations that arose in the Brazilian gamma variant of concern of SARS-CoV-2, giving it an evolutionary advantage [3].

Type the following command, as a single line, in the terminal (**do not copy from PDF, characters may be replaced and an error message will be prompted!**):

foldx --command=BuildModel --pdb=complex_Ab_Repair.pdb
--mutant-file=individual_list.txt --ionStrength=0.05 --pH=7
--water=predict --vdwDesign=2 --pdbHydrogens=false
--numberOfRuns=5

Task 3

An output file named **Average_complex_Repair.fxout** or **Average_complex_Ab_Repair.fxout** will be written for each calculation.

Type in terminal:

head Average_complex_Repair.fxout (or **head Average_complex_Ab_Repair.fxout)**

for each output file calculation.

3.1 Fill the following table based on ΔΔG values and classification on Figure 3. Which mutation do you expect to be advantageous? Be careful on interpreting the result for RBD-Antibody complex.

Mutation	ΔΔG	SD	Impact
N439K			
Q493A			
N501Y			
K417N,E484K,N501Y			

References

(1) Ho, M. Perspectives on the development of neutralizing antibodies against SARS-CoV-2. *Antib Ther* **2020**, *3* (2), 109-114. DOI: 10.1093/abt/tbaa009.

(2) Taka, E.; Yilmaz, S. Z.; Golcuk, M.; Kilinc, C.; Aktas, U.; Yildiz, A.; Gur, M. Critical Interactions Between the SARS-CoV-2 Spike Glycoprotein and the Human ACE2 Receptor. *bioRxiv* **2020**, 2020.2009.2021.305490. DOI: 10.1101/2020.09.21.305490.

(3) Ferraz, M. V. F.; Moreira, E. G.; Coêlho, D. F.; Wallau, G. L.; Lins, R. D. Immune evasion of SARS-CoV-2 variants of concern is driven by low affinity to neutralizing antibodies. *Chemical communications* **2021**, 10.1039/D1CC01747K. DOI: 10.1039/D1CC01747K.