

Scripts and Notebooks used in the computational protocol in "Unraveling the Molecular Basis of Substrate Specificity and Halogen Activation in Vanadium-Dependent Haloperoxidases"

Symmetry Relax

A symmetry definition file was created using the biological assembly of 5lpc. The symmetry protocol was run with the monomer modeled in AlphaFold2.

A detailed tutorial to set up a structure for symmetry can be found at <https://new.rosettacommons.org/demos/latest/tutorials/Symmetry/Symmetry>. To run the symmetry relax protocol, you have to

- create a symmetry definition file
- adapt the variables in `sym_relax.options` and `sym_relax.command`
- run:

```
#use input pdb with B and C as symmetric chains to A
$ROSETTA3/src/apps/public/symmetry/make_symmdef_file_denovo.py -m NCS -p
input.pdb -a A -i B,C \
> out.sym
#change input pdb and symmetry in sym_relax.command
bash sym_relax.command
```

Cartesian ddG

A detailed tutorial to Rosetta cartesian ddG can be found on <https://www.rosettacommons.org/docs/latest/cartesian-ddG>. To run the cartesian ddG protocol, you have to

- relax your input structure in cartesian space
- create mutfiles based using the Rosetta syntax in https://www.rosettacommons.org/docs/latest/application_documentation/analysis/ddg-monomer
- run the ddG application on the relaxed structure

```
#enter the correct paths in cartrelax.command
bash cartrelax.command
#use your relaxed output and personalized mutfiles in cartddg.command
bash cartddg.command
```

The output will be a scorefile with the ddG upon mutation. Depending on the size the energetic frustration of the structure more relaxed structures should be produced and the low scoring ones selected for ddG calculations.

ProteinMPNN

The `pmpnn_complex.ipynb` jupyter notebook provides an example on how to

- set up the PMPNN model
- select residues near a site that are surface exposed
- scan each of the selected residues and calculate conditional probabilities
- plot the probability for each position in a sequence logo