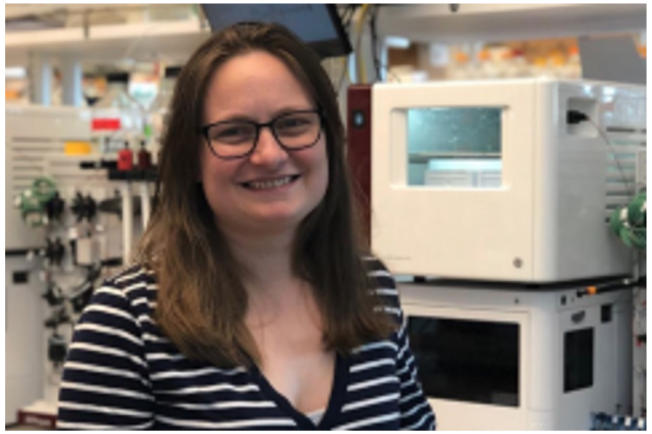
VIB training
Introduction to computational protein design

VUB-VIB Center for structural biology

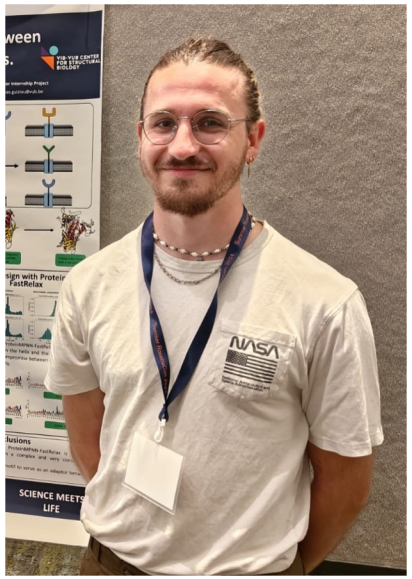
11/09/2024

Anastassia Vorobieva, Thomas Guiziou, Andrei Sokolovskii



Anastassia Vorobieva
Group leader VIB
Assistant professor VUB

Anastassia.vorobieva@vib.be



Thomas Guiziou
PhD student VUB-VIB

Thomas.guiziou@vib.be

Andrei Sokolovskii
PhD student VIB-VUB

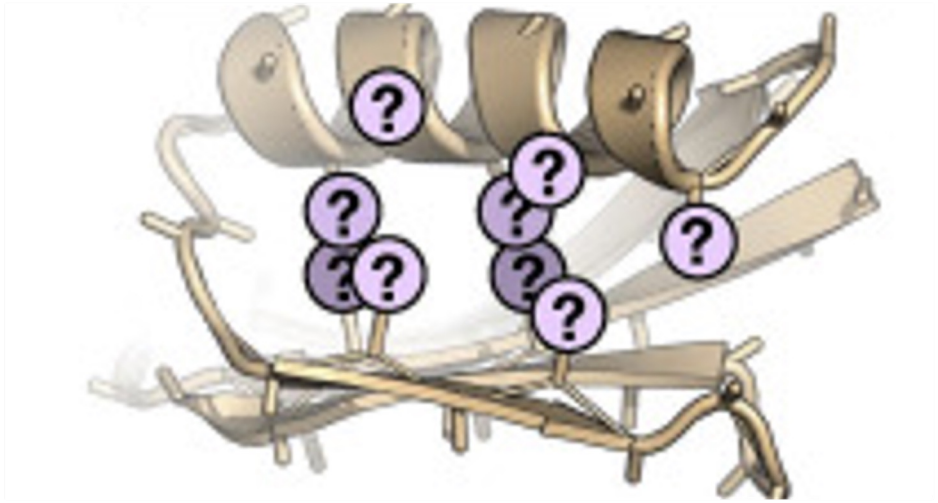
andrei.sokolovskii@vib.be

Training agenda

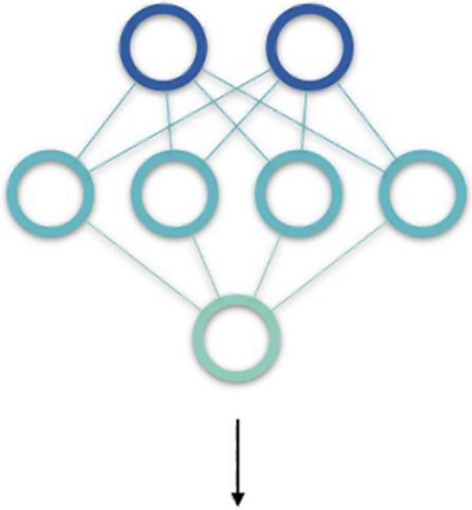
	DAY 1	DAY 2
9:30 – 10:55	Introduction: selecting a PDB template	Introduction to protein design
10:55 – 11:15	Break	Break
11:15 – 12:45	Preparing the PDB	Building backbones with RFDiffusion
12:45 – 13:45	Lunch	Lunch
13:45 – 15:10	Site saturation mutagenesis	Sequence design with ProteinMPNN
15:10 – 15:30	Break	Break
15:30 – 17:00	Combinatorial mutagenesis with PyRosetta	Filtering designs with ColabFold

Structure *vs* sequence-based protein design

Structure-based design



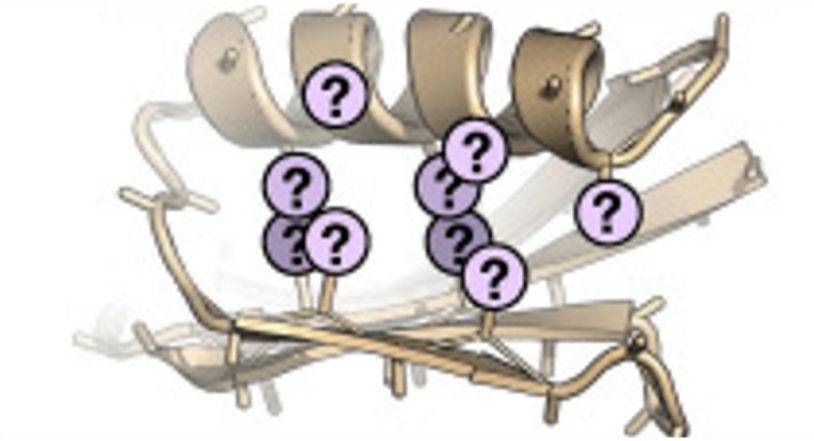
Direct sequence generation (e.g. LLMs)



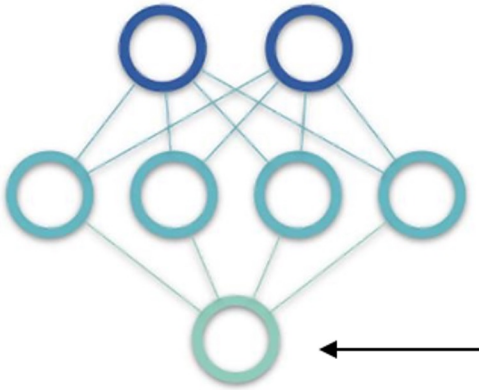
…MALKIPTHNHM…
…VFRDCEWS…
…WYIOPMNVGTDEW…

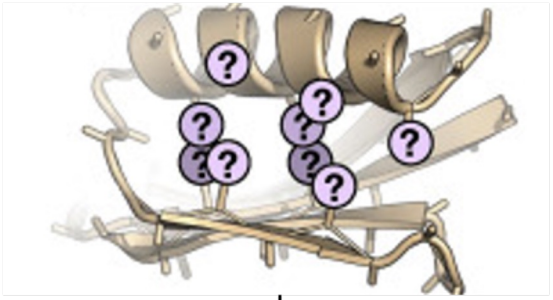
Structure *vs* sequence-based protein design

Structure-based design



Direct sequence generation (e.g. LLMs)

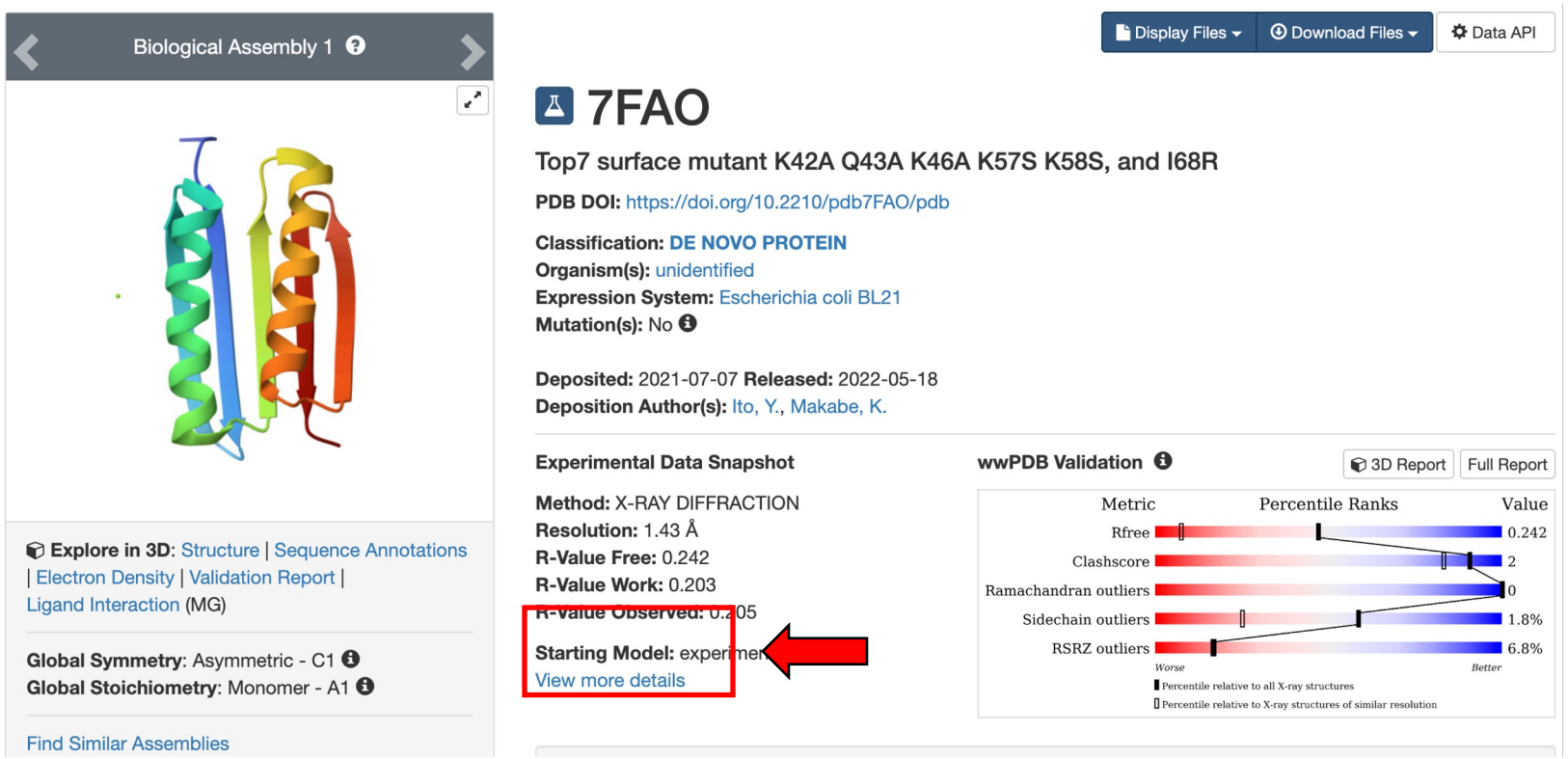


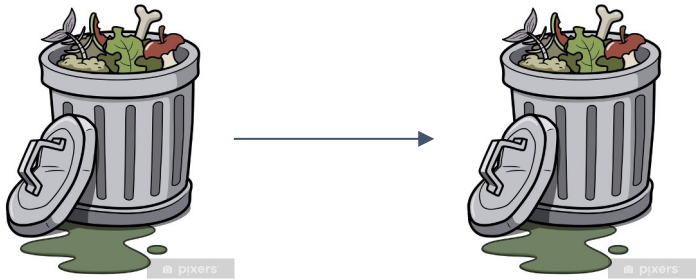


…MALKIPTHNHM…
…VFRDCEWS…
…WYIOPMNVGTDEW…

DAY 1: Predicting the effect of mutations

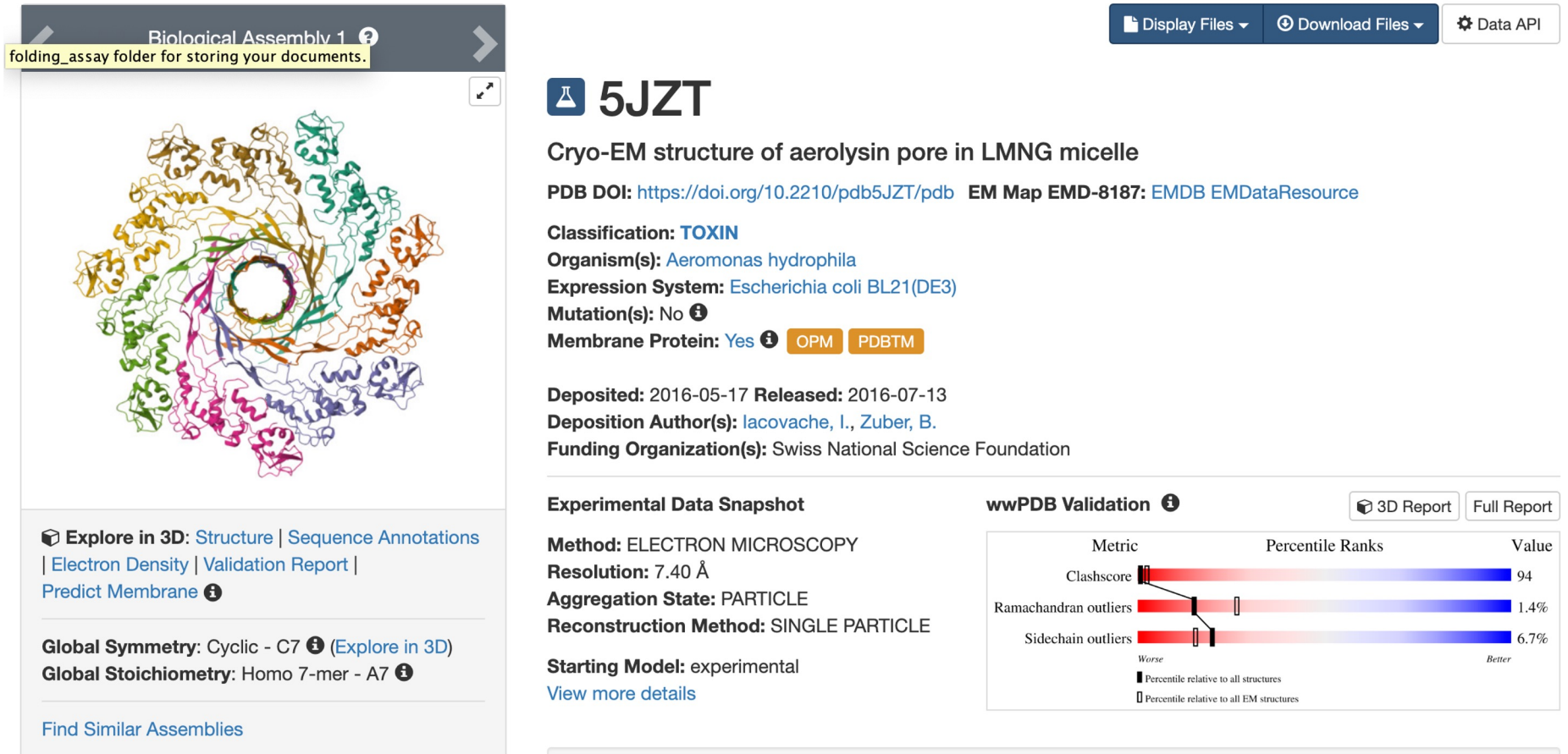
1. Selecting your template PDB structure





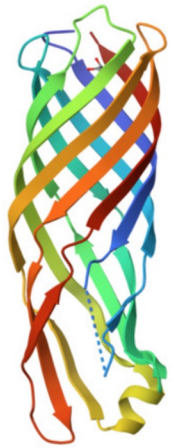
EXERCISE: download and visualize the PDB structures to find the issues

Problem: Low resolution structure



Problem: Poor refinement

Biological Assembly 1



Explore in 3D: Structure | Sequence Annotations | Electron Density | Validation Report | Ligand Interaction (GOL) | Predict Membrane

Global Symmetry: Asymmetric - C1
Global Stoichiometry: Monomer - A1

Find Similar Assemblies

Display Files | Download Files | Data API

2F1V

Outer membrane protein OmpW

PDB DOI: https://doi.org/10.2210/pdb2F1V/pdb

TMB12_3_nb3

Classification: MEMBRANE PROTEIN
Organism(s): Escherichia coli K-12
Expression System: Escherichia coli
Mutation(s): No
Membrane Protein: Yes OPM PDBTM MemProtMD

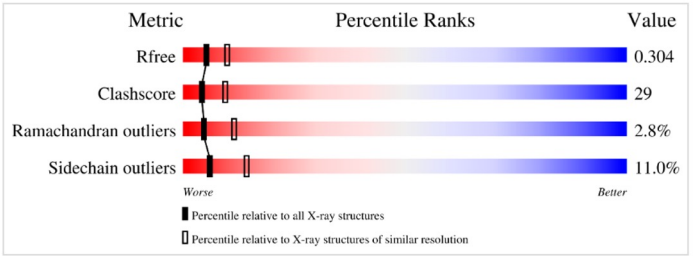
Deposited: 2005-11-15 **Released:** 2006-01-24
Deposition Author(s): van den Berg, B.

Experimental Data Snapshot

Method: X-RAY DIFFRACTION
Resolution: 2.70 Å
R-Value Free: 0.314
R-Value Work: 0.292
R-Value Observed: 0.293

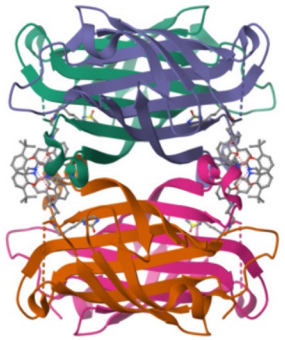
Starting Model: experimental
View more details

wwPDB Validation 3D Report | Full Report



Problem: Missing loops, alternative conformations (b-factor)

Biological Assembly 1



Explore in 3D: Structure | Sequence Annotations | Validation Report | Ligand Interaction (UFU)

Global Symmetry: Dihedral - D2 (Explore in 3D)
Global Stoichiometry: Homo 4-mer - A4

Find Similar Assemblies

Biological assembly 1 assigned by authors and generated by PISA (software)

Biological Assembly Evidence: gel filtration

Macromolecule Content

Display Files | Download Files | Data API

8QEX

Streptavidin variant with a cobalt catalyst for CH metal-catalyzed hydrogen-atom-transfer (M-HAT)

PDB DOI: https://doi.org/10.2210/pdb8QEX/pdb

Classification: METAL BINDING PROTEIN
Organism(s): Streptomyces avidinii
Expression System: Escherichia coli BL21(DE3)
Mutation(s): No

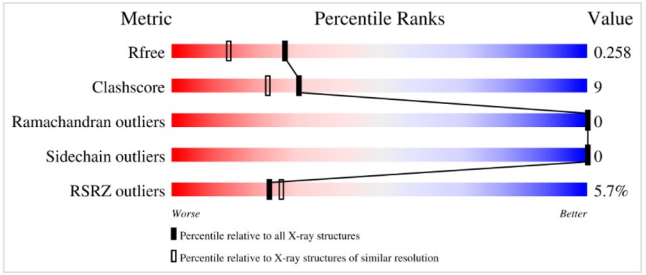
Deposited: 2023-09-01 **Released:** 2024-07-31
Deposition Author(s): Jakob, R.P., Chen, D., Ward, T.R.
Funding Organization(s): Not funded

Experimental Data Snapshot

Method: X-RAY DIFFRACTION
Resolution: 1.90 Å
R-Value Free: 0.254
R-Value Work: 0.229
R-Value Observed: 0.230

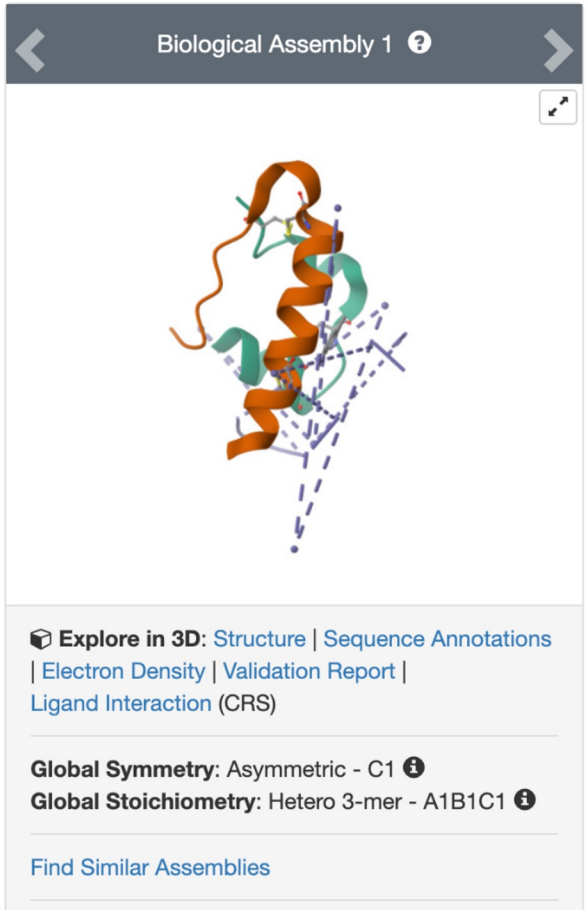
Starting Model: experimental
View more details

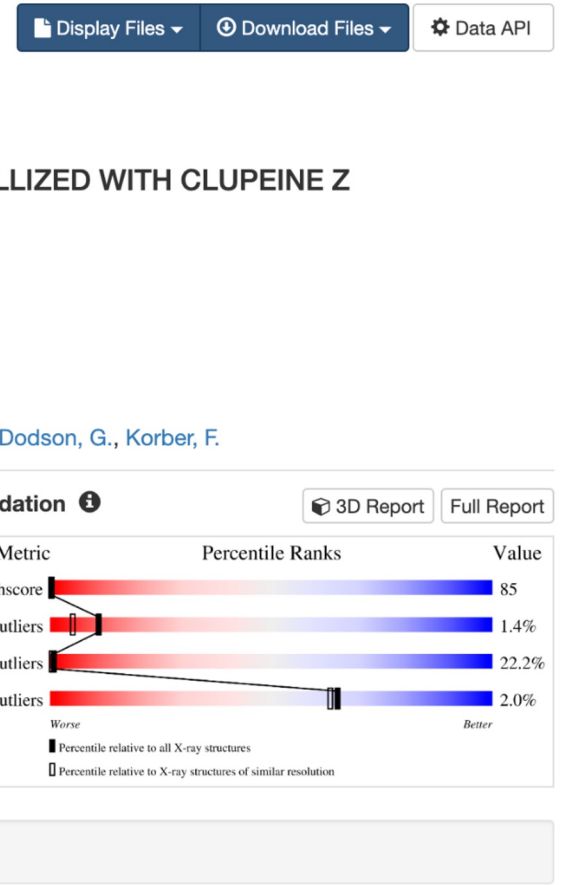
wwPDB Validation 3D Report | Full Report



Ligand Structure Quality Assessment

Problem: Ligand modelling

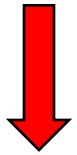




This time, download the PDB and the MTZ map. Look at the unknown ligands (UNK) in Pymol

2. Preparing the PDB structure

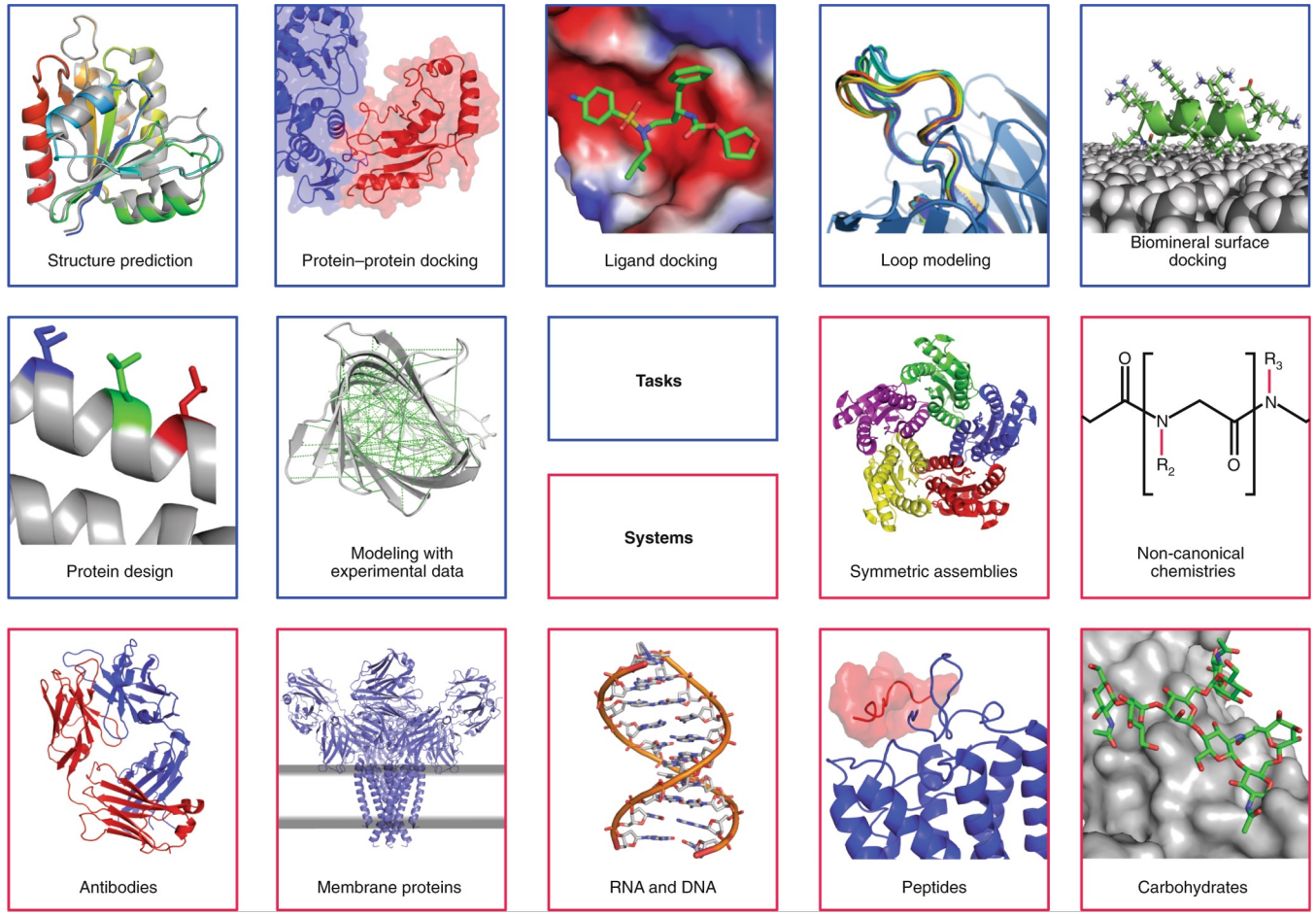
A. Clean PDB



```
ATOM   8178  CD2 LEU A 207     135.703 156.448 175.146  1.00 59.65           C
ATOM   8179  N   LEU A 208     138.542 160.482 176.163  1.00 64.39           N
ATOM   8180  CA  LEU A 208     139.558 160.550 177.204  1.00 64.39           C
ATOM   8181  C   LEU A 208     139.523 161.898 177.917  1.00 64.39           C
ATOM   8182  O   LEU A 208     140.107 162.065 178.988  1.00 64.39           O
ATOM   8183  CB  LEU A 208     140.944 160.306 176.607  1.00 64.39           C
ATOM   8184  CG  LEU A 208     141.162 158.912 176.018  1.00 64.39           C
ATOM   8185  CD1 LEU A 208     142.520 158.825 175.340  1.00 64.39           C
ATOM   8186  CD2 LEU A 208     141.011 157.842 177.088  1.00 64.39           C
ATOM   8187  OXT LEU A 208     138.909 162.851 177.438  1.00 64.39           O
TER    8188      LEU A 208
HETATM 8189  O1' LMT E 301     138.158 131.617 150.769  1.00 20.00           O
HETATM 8190  C1  LMT E 301     137.918 131.876 152.147  1.00 20.00           C
HETATM 8191  C2  LMT E 301     137.062 133.128 152.260  1.00 20.00           C
HETATM 8192  C3  LMT E 301     137.071 133.678 153.676  1.00 20.00           C
HETATM 8193  C4  LMT E 301     136.376 135.025 153.714  1.00 20.00           C
HETATM 8194  C5  LMT E 301     135.935 135.347 155.125  1.00 20.00           C
HETATM 8195  C6  LMT E 301     135.530 136.801 155.206  1.00 20.00           C
HETATM 8196  C7  LMT E 301     136.343 137.543 156.242  1.00 20.00           C
HETATM 8197  C8  LMT E 301     135.478 138.063 157.375  1.00 20.00           C
HETATM 8198  C9  LMT E 301     136.194 137.846 158.690  1.00 20.00           C
HETATM 8199  C10 LMT E 301     135.627 138.700 159.797  1.00 20.00           C
HETATM 8200  C11 LMT E 301     135.963 138.052 161.121  1.00 20.00           C
HETATM 8201  C12 LMT E 301     135.929 139.084 162.215  1.00 20.00           C
```

B. Preparing the PDB for Rosetta calculations

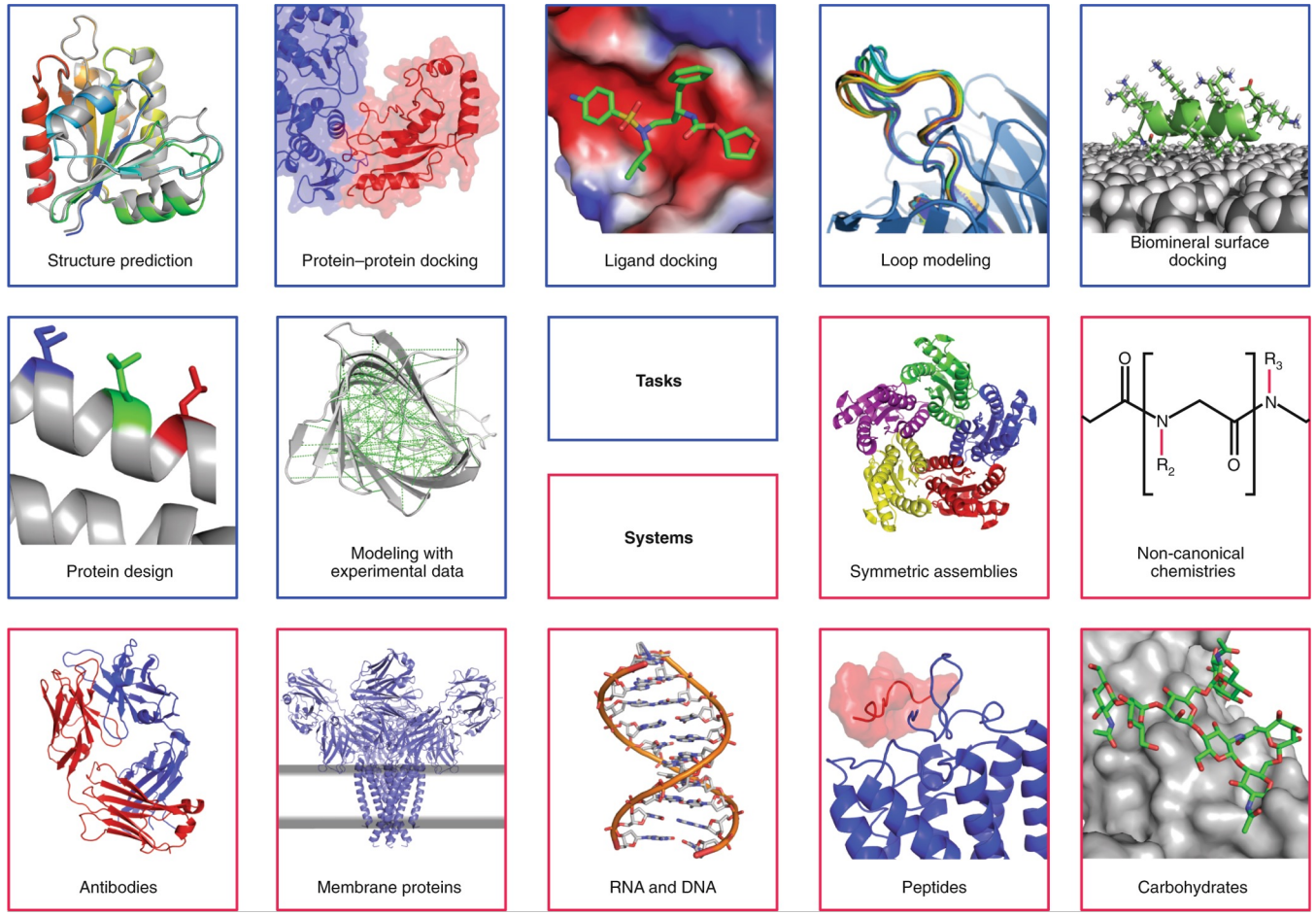
What is Rosetta ?



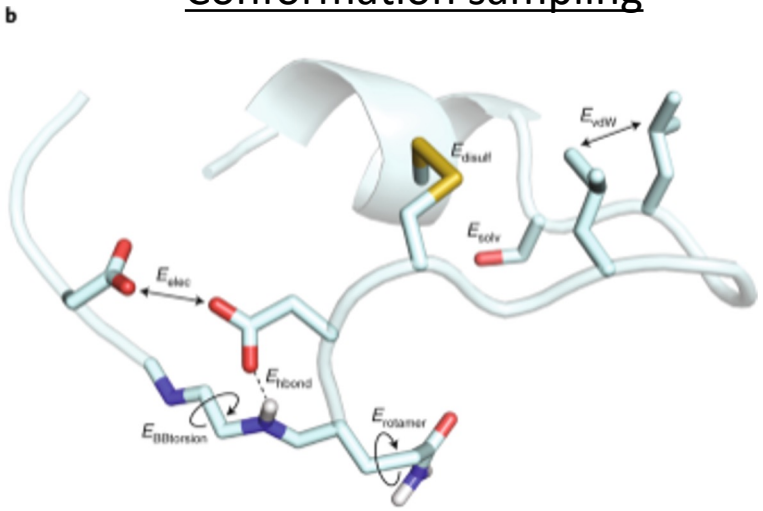
https://doi.org/10.1038/s41592-020-0848-2

B. Preparing the PDB for Rosetta calculations

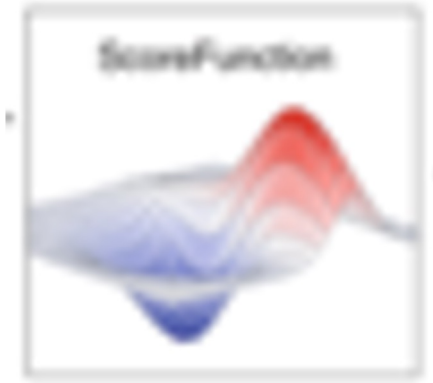
What is Rosetta ?



Conformation sampling



Scoring

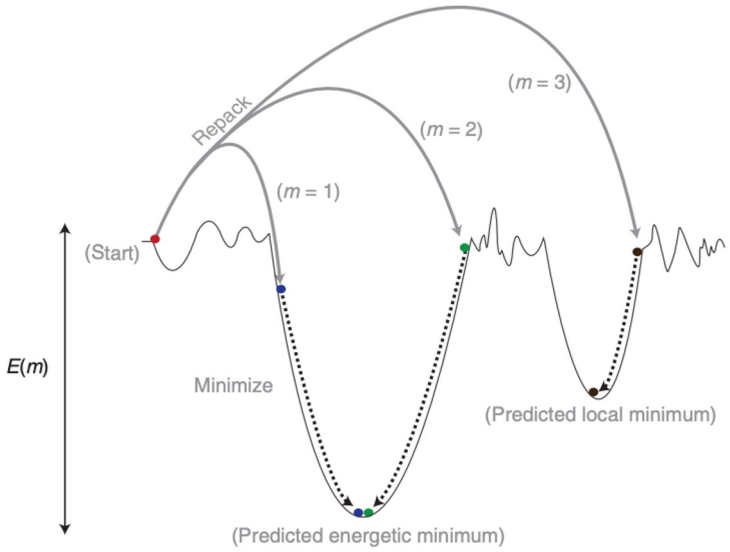


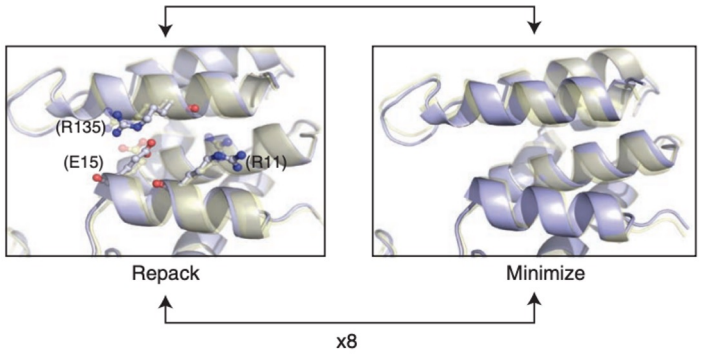
E_{vdW}	Lennard–Jones for attractive or repulsive interaction
E_{hbond}	Hydrogen bonding allows buried polar atoms
E_{elec}	Electrostatic interaction between charges
E_{disulf}	Disulfide bonds between cysteines
E_{solv}	Implicit solvation model penalizes buried polar atoms
$E_{BBtorsion}$	Backbone torsion preferences from main-chain potential
$E_{rotamer}$	Side-chain torsion angles from rotamer library
E_{ref}	Unfolded state reference energy for design

https://doi.org/10.1038/s41592-020-0848-

B. Relaxing the PDB for Rosetta calculations

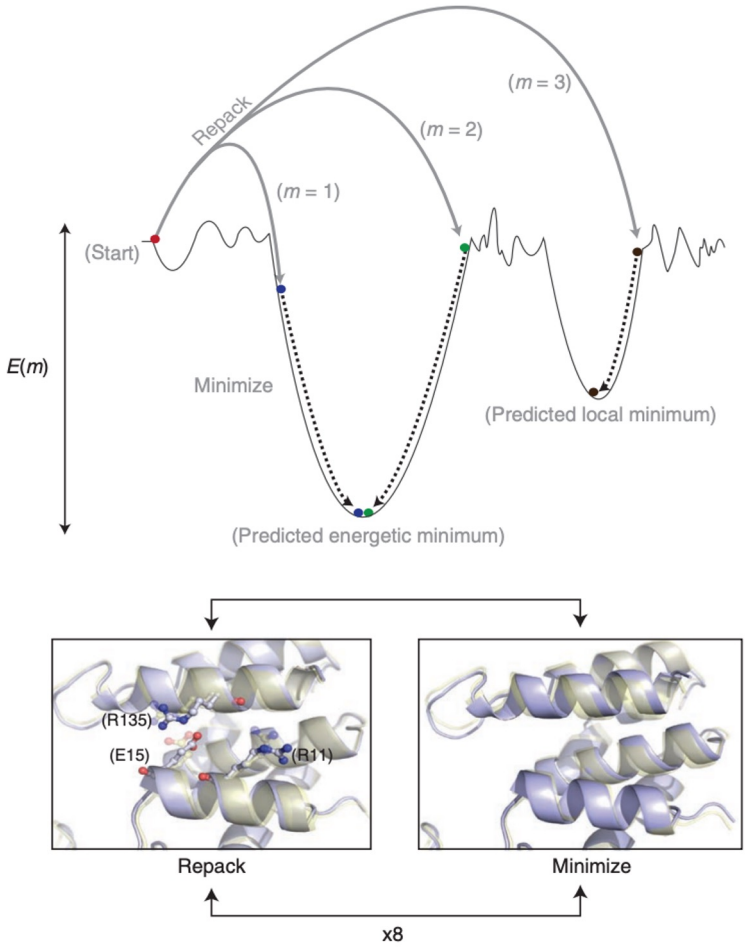
Rosetta FastRelax algorithm





B. Relaxing the PDB for Rosetta calculations

Rosetta FastRelax algorithm



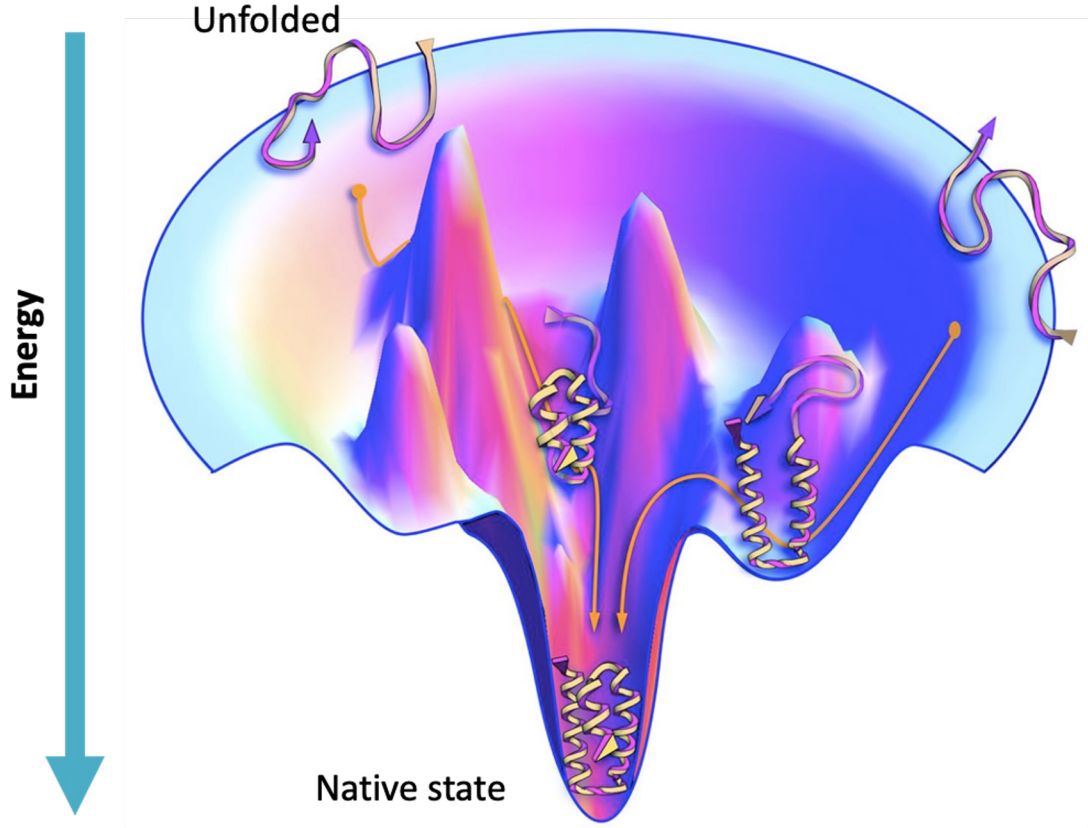
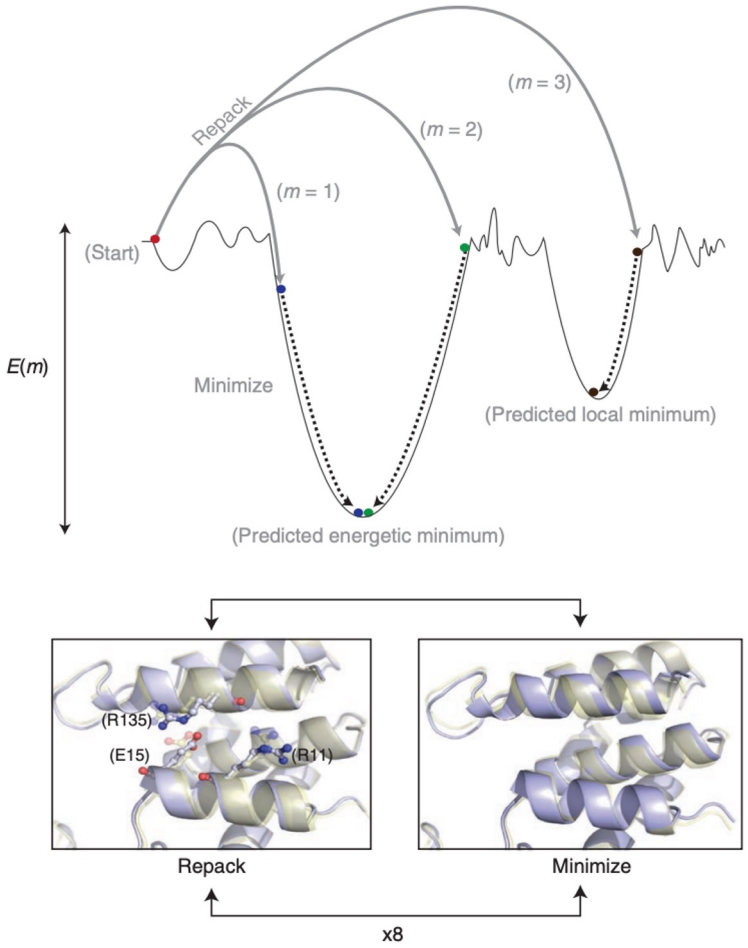
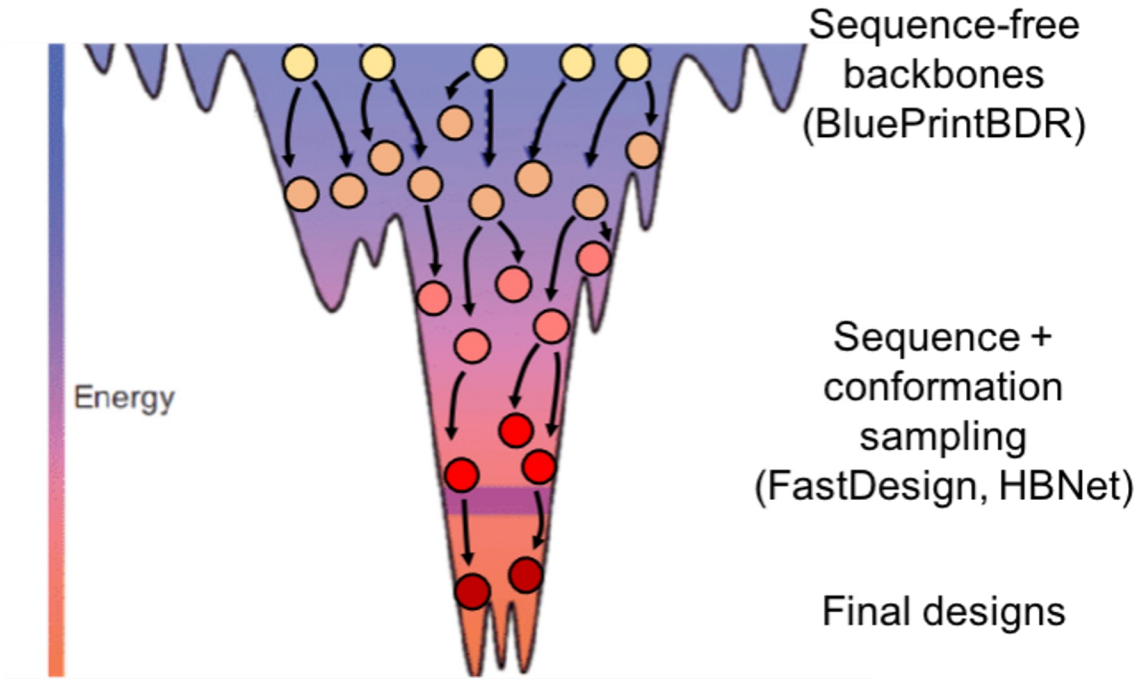


Image from: "The protein-folding problem, 50 years on." *science* 338, no. 6110 (2012): 1042-1046.

B. Relaxing the PDB for Rosetta calculations

Rosetta FastRelax algorithm





Let’s connect to your computing environment

1. Connect to the VUB Jupyter hub: **https://notebooks.hpc.vub.be/**

1. Sign-in with your VSC account and authorize Jupyterhub

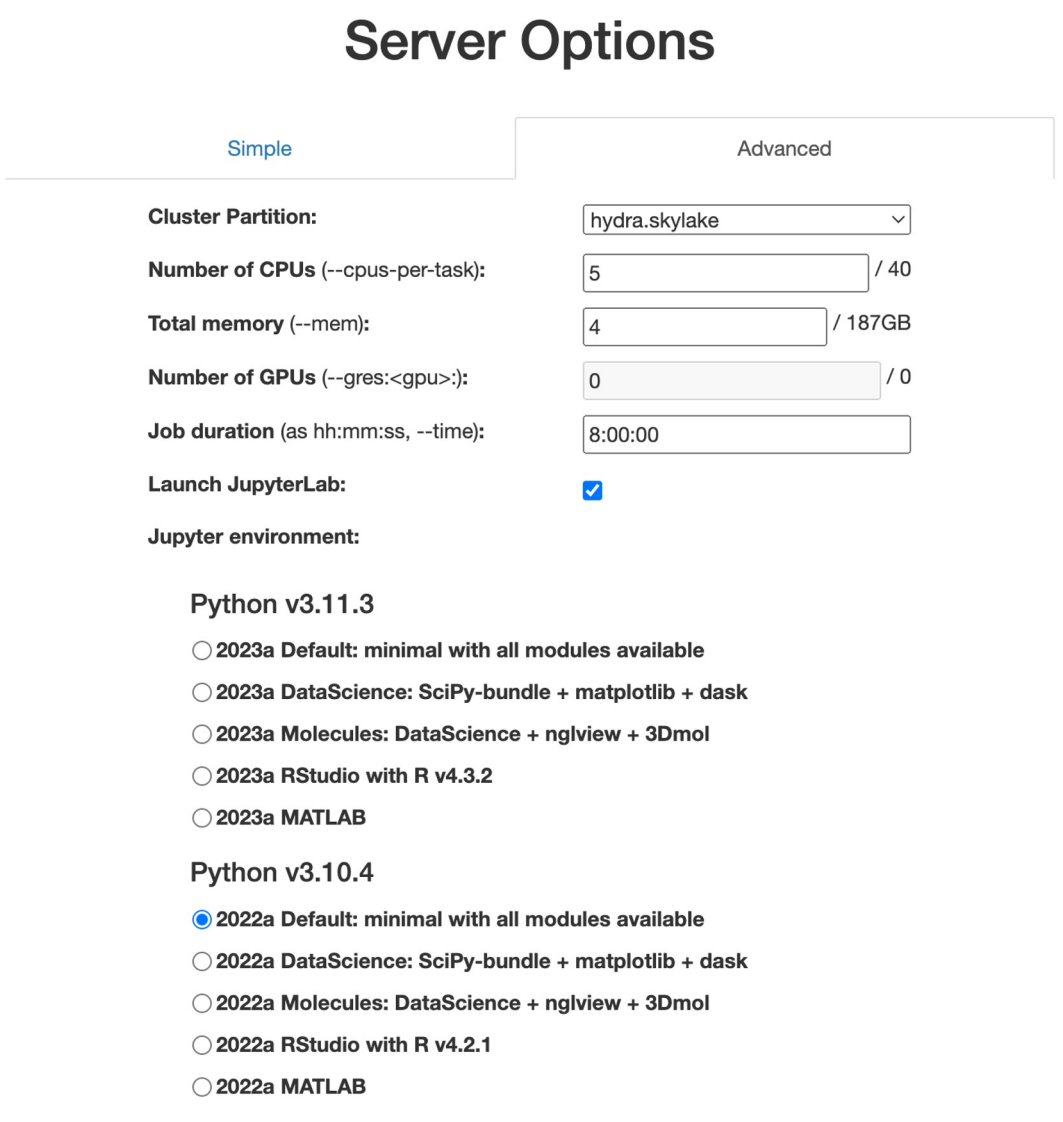
1. Spawn a server
   Reservation: keep empty(automatically *train_comp_prot_d1*)

1. Open a terminal window and create a new directory for the class
   mkdir training_day1

1. Go to the directory
   cd training_day1/

1. Download the course material:
   git clone https://github.com/AndreiSokolovskii/Protein_ddG_workshop.git
   cd Protein_ddG_workshop
2. Look for dependencies on the left and load:

   1) PyRosetta/4.release-384-gompi-2022a; 2) SciPy-bundle/2022.05-foss-2022a
   3) py3Dmol/2.0.1.post1-GCCcore-11.3.0; 4) Biopython/1.79-foss-2022a; matplotlib/3.5.2-foss-2022a
   5) Seaborn/0.12.1-foss-2022a 6) matplotlib/3.5.2-foss-2022a
1. Open the Jupyter notebook that was downloaded



You're fast … let's try something else: relax without coordinate constraints

```
xml = pyrosetta.rosetta.protocols.rosetta_scripts.XmlObjects.create_from_string("""
<ROSETTASCRIPTS>
    <SCOREFXNS>
    <ScoreFunction name="SFX1" weights="ref2015_cart">
       <Reweight scoretype="coordinate_constraint" weight="1.0"/>
    </ScoreFunction>
    </SCOREFXNS>
    <RESIDUE_SELECTORS>
    </RESIDUE_SELECTORS>
    <TASKOPERATIONS>
    </TASKOPERATIONS>
    <FILTERS>
    </FILTERS>
    <MOVERS>
       <AtomCoordinateCstMover name="coord_cst" />
       <FastRelax name="relax" cartesian="true" scorefxn="SFX1" />
    </MOVERS>
    <APPLY_TO_POSE/>
    <PROTOCOLS>
       <Add mover="coord_cst" />
       <Add mover="relax" />
    </PROTOCOLS>
</ROSETTASCRIPTS>
""").get_mover("ParsedProtocol")

working_dir = os.getcwd()
output_dir = dest
```

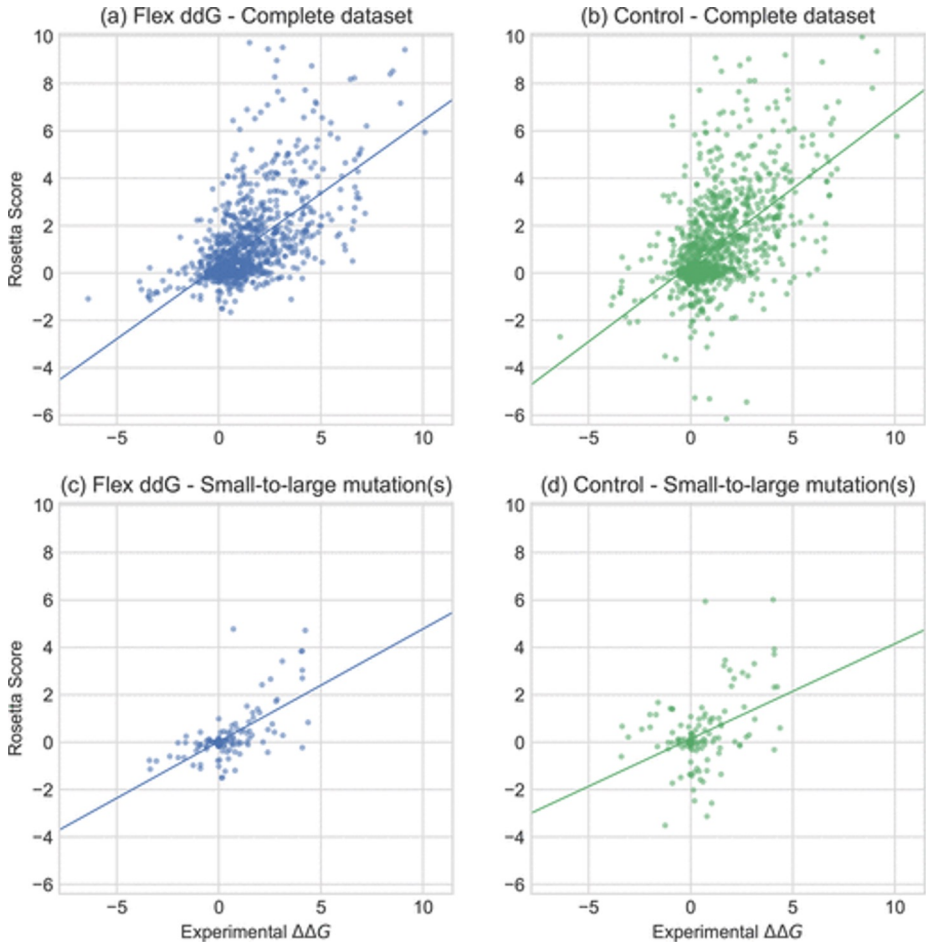
Comment out this line

Change **conf**'s *coord_cst* parameter to **False** in a cell below

Look at the models in pymol … what is the difference?

3. Running DDG calculations – Rosetta (energy model)

Rosetta's scorefunction was fitted to experimentally-determined DDGs



Benchmarks from Kortemme

Cartesian ΔΔG

Step 1: Relax Starting Model 20 Times

Step 2: Mutate Target Residue

Step 3: Repack Target Residue

Step 4: Minimize Target Residue And Its Neighbors

Repeat Until Converged (max 5)

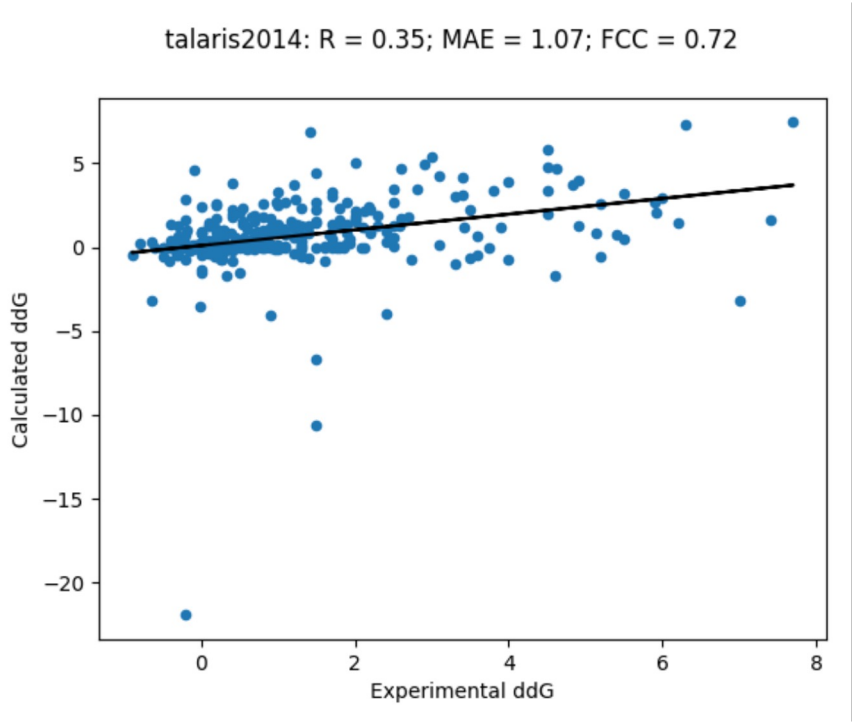
Repack target residue

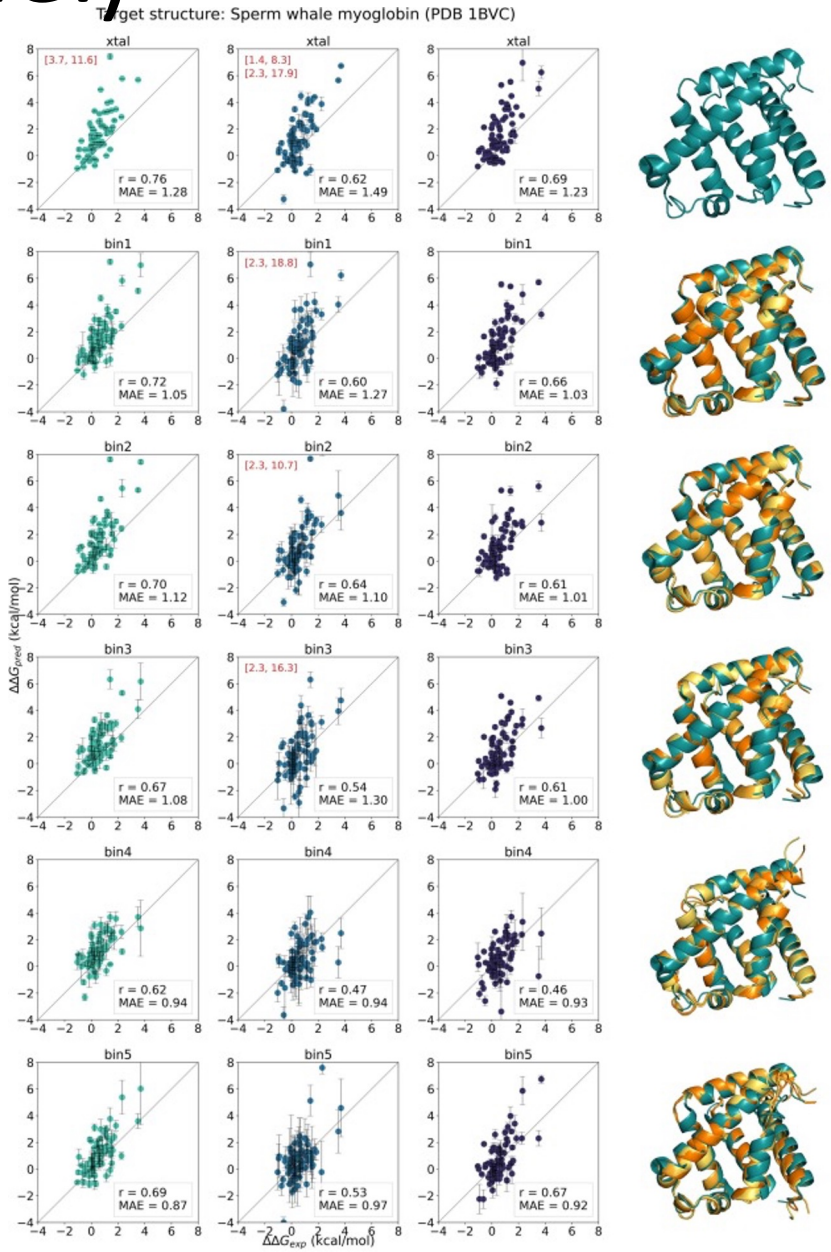
Minimize target residue and its neighbors

$$\Delta\Delta G = \frac{\sum_i MUT\_total\_score_i}{ddg\_iterations} - \frac{\sum_i WT\_total\_score_i}{ddg\_iterations}$$

3. Running DDG calculations – Rosetta (energy model)

Accurate predictions can be made for alanine scanning, from homology and AlphaFold models

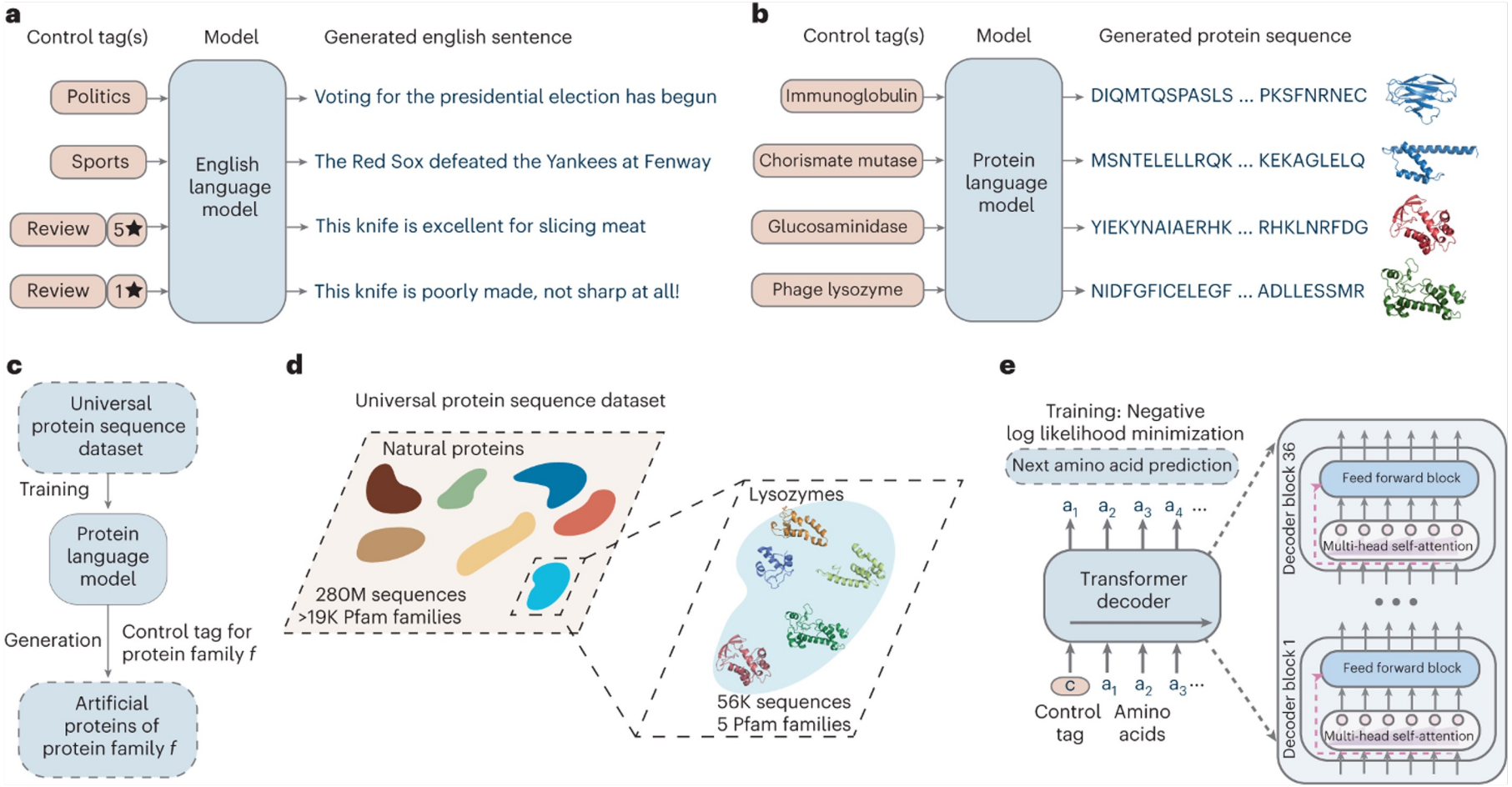




https://doi.org/10.1016/j.csbj.2022.11.048

Detailed benchmark: DOI: 10.3389/fbioe.2020.558247

3. Running DDG calculations – LLMs (evolutive models)



A notebook to run a basic LLM fitness prediction will be sent after the class (high memory requirements)

https://doi.org/10.1038/s41587-022-01618-2

Exercise 1: generate a DDG heatmap with Rosetta

Back to the Jupyter notebook … Let's start by mutating residues in a small peptide

Remember:

$$\Delta\Delta G = \frac{\sum_i MUT\_total\_score_i}{ddg\_iterations} - \frac{\sum_i WT\_total\_score_i}{ddg\_iterations}$$

Like the color palette? No? Then change it!

Exercise 2: generate an interaction_DDG heatmap

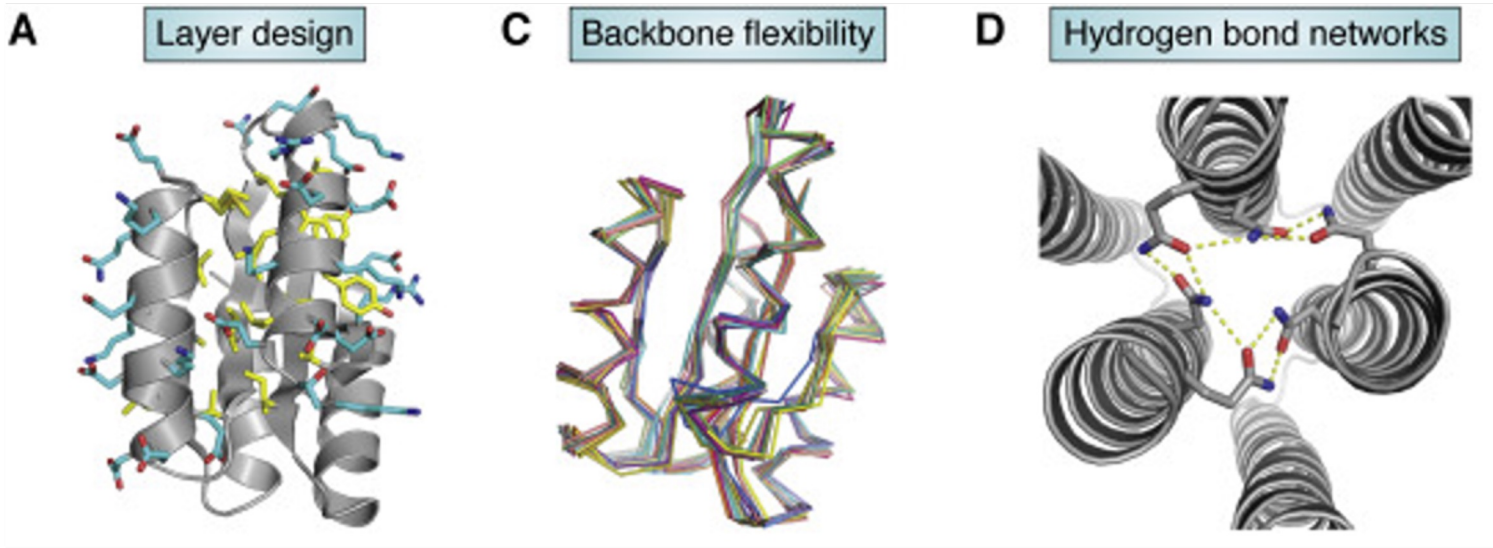
Let’s probe the interaction of the alpha-helical peptide BHRF1 with the Bim BH3 domain (PDB 2WH6)

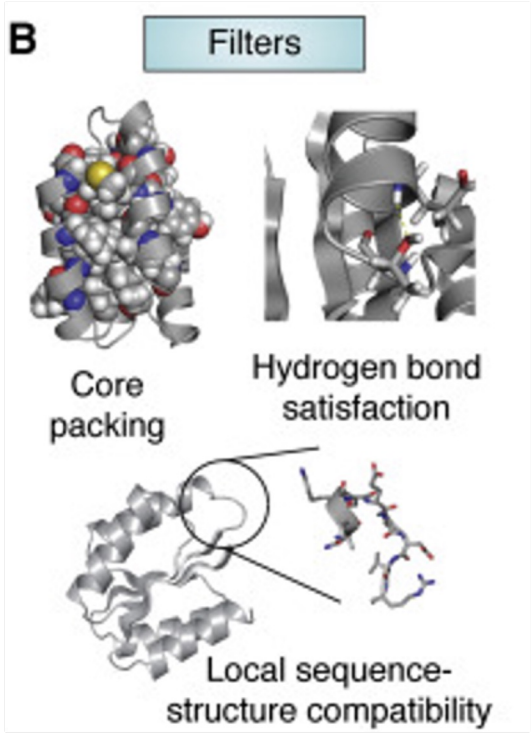
You’ll score the mutations in the complex then separately.

Here a different formula applies:

$$\Delta\Delta G_{interface} = \frac{\sum_{i=0}^{n} COMPLEX\_score_n}{n} - \frac{\sum_{i=0}^{n} SEPARARATE\_score_n}{n}$$

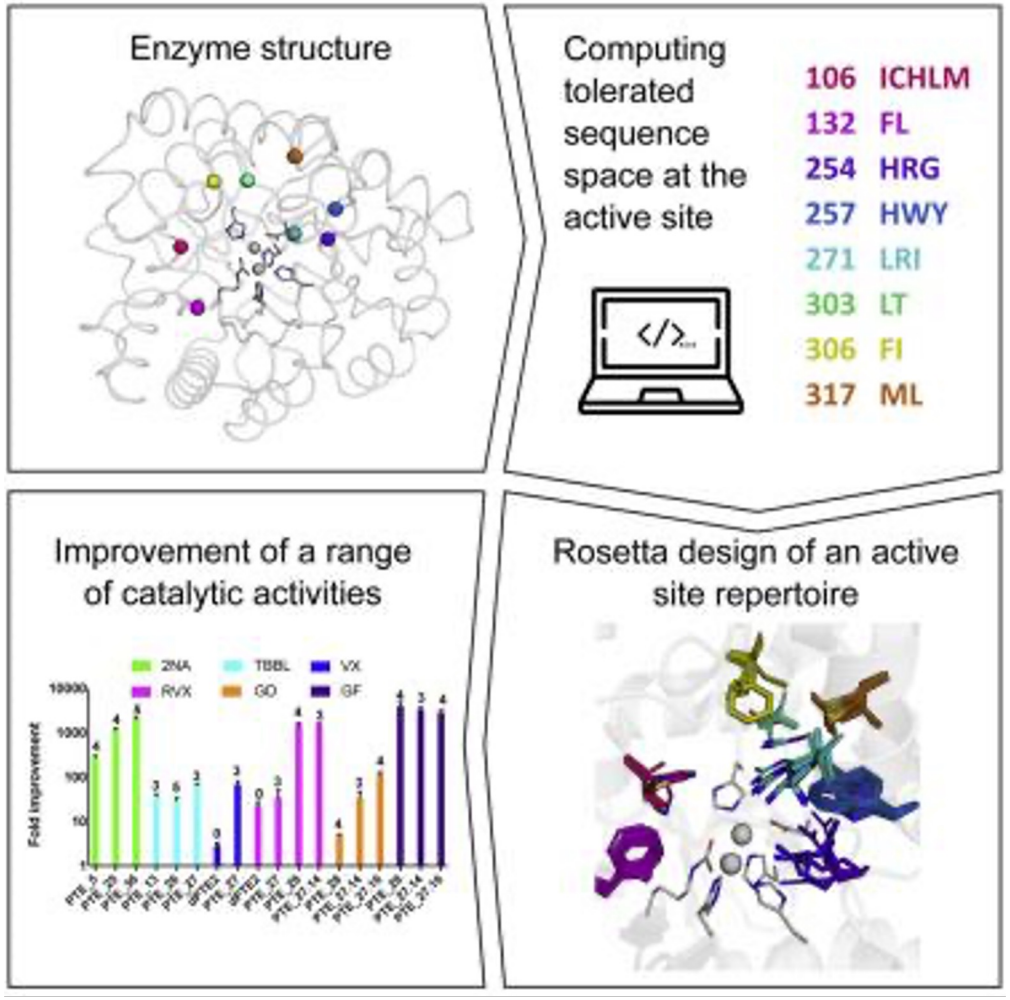
4. Running Combinatorial design – Rosetta





DOI: 10.1016/j.jbc.2021.100558

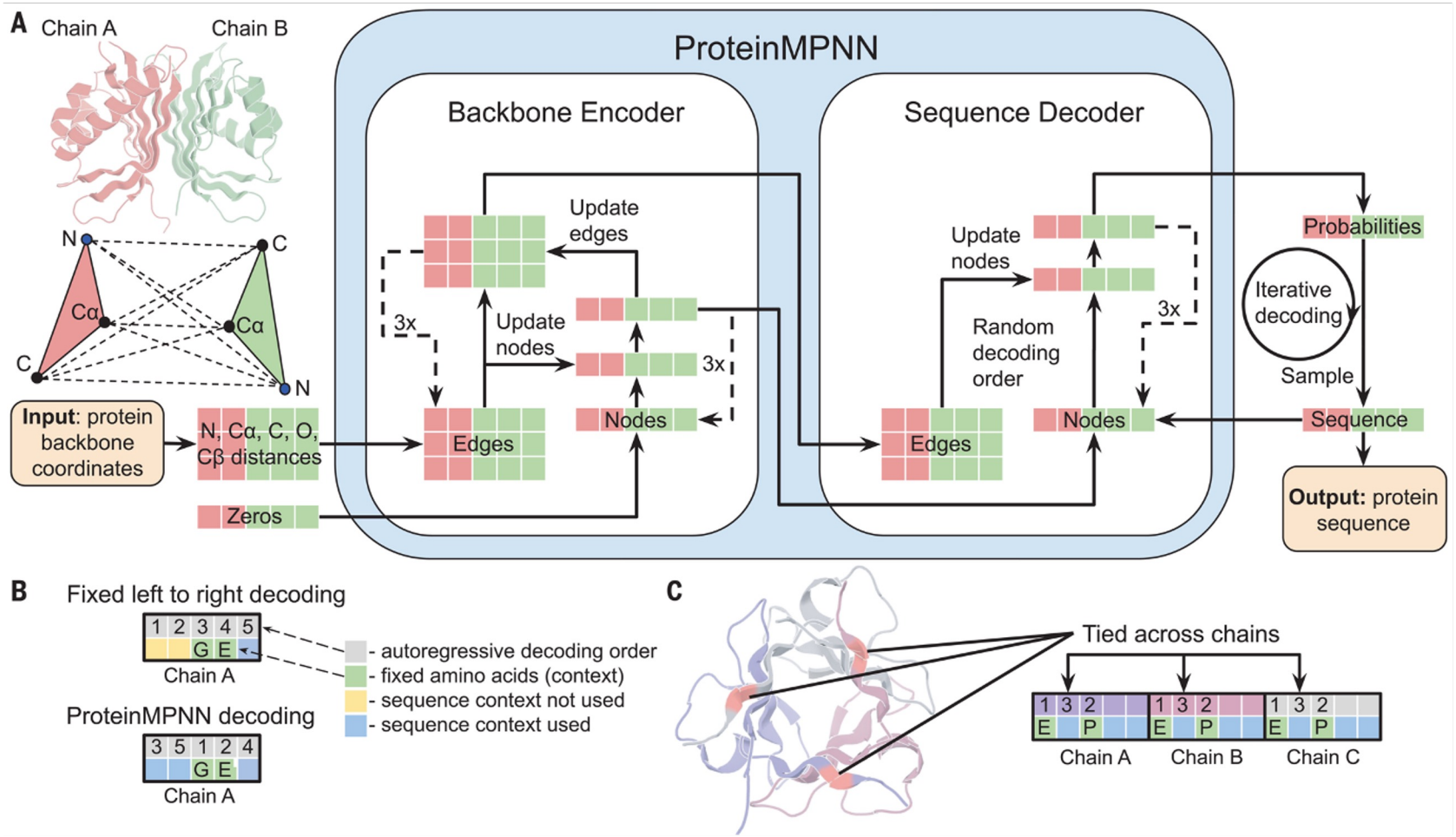
4. Running Combinatorial design - PROSS



Improves bacterial expression, stability and function of many eukaryotic (human) proteins

Paper: https://doi.org/10.1016/j.molcel.2018.08.033
Server: https://pross.weizmann.ac.il/

4. Running Combinatorial design - ProteinMPNN



DOI: 10.1126/science.add218

More about it tomorrow!

Exercise: Run combinatorial design calculations with Rosetta

DAY 1: Questions?

See you tomorrow!!