

Protein Design Project

The aim of this project is to make you familiar with the different methods of the protein design pipeline we have encountered throughout this course. You will try to design proteins against a chosen target. The proposed list of targets is non-exhaustive and you can also *pick your own protein of interest*.

The Deliverables:

- 2 page report containing the biological background of the target, why protein design is interesting here and your methodology - make sure to cite all the methods!
- 10 min presentation with the same content as the report, focus on figures instead (structure of target, protein design pipeline, results, structure of the target in complex with the binder)
- The sequence of your best 10 designed binders based on at least 2 orthogonal validation metrics (e.g. ColabFold pAE and Boltz' score where you pick the best) as fasta format

Potential targets:

PDB-IDs: 3ZTJ (H3), 3MJG (PDGFR), 4OGA (IR), 5U8R (IGF1R), 2GY7 (TIE2), 1XIW (CD3 δ), 3KFD (TGF β), 4O3V (VirB8) - Source: <https://www.nature.com/articles/s41586-022-04654-9>

KIT (PDB:2E9W), NRP1 (PDB:2QQI), PDGFR α (PDB:7LBF) - Source: rcsb.org ALK-2 (UniProt:Q04771), ALK-3 (UniProt:P36894), FCRL5 (UniProt:Q96RD9)

SARS-CoV-2 proteins (COVID-19 research): Spike glycoprotein: PDB IDs: 6VSB, 6VXX; Main protease: PDB ID: 5R7Y; NSP12 (RNA-dependent RNA polymerase): PDB ID: 6XEZ; NSP16 (2'-O-methyltransferase): PDB ID: 6W4H; ORF8: PDB ID: 7JTL

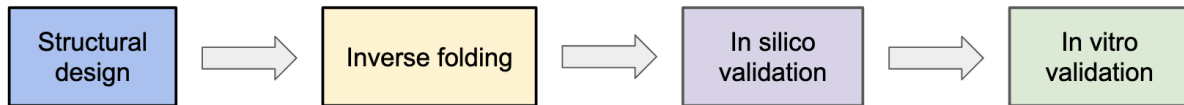
Cancer-related proteins: Bromodomain-containing protein 4 (BRD4): UniProt ID: O60885, PDB ID: 8RX0; Von Hippel-Lindau disease tumor suppressor (VHL): UniProt ID: P40337, PDB ID: 8RX0;

Ubiquitin-proteasome system: Ubiquitin-conjugating enzyme E2 R1 (UBE2R1): UniProt ID: P49427, PDB ID: 8RX0; E3 ubiquitin-protein ligase RBX1: UniProt ID: P62877, PDB ID: 8RX0

Source: <https://www.biorxiv.org/content/10.1101/2024.10.11.617496v2.full>

The Protein Design Pipeline - available methods

As we have already learned, the commonly used Protein Design pipeline consists of the following steps:



As we are limited in compute and time only a selected subset of these tools is available to us. Nonetheless it is important to explore most of the different tools so it makes the most sense for you. Again this list is non-exhaustive and if you find other methods feel free to use them! In my personal experience the models on the huggingface were much easier to use than on Google Colab as those would depend on the session on the browser being active - additional to a non-EPFL Google account. When using the huggingface some models may be inactive and it can take some minutes to rebuild them.

Available methods

Pipelines (combining two or more methods):

ColabDesign:

https://colab.research.google.com/github/sokrypton/ColabDesign/blob/main/rf/examples/diffusion_ori.ipynb

ProteinMPNN+ESM: <https://huggingface.co/spaces/simonduerr/ProteinMPNNEsm>

ProteinMPNN+ColabFold: <https://huggingface.co/spaces/simonduerr/ProteinMPNN>

BindCraft (AF2-backpropagation):

<https://colab.research.google.com/github/martinpacesa/BindCraft/blob/main/notebooks/BindCraft.ipynb>

Structural design

RFDiffusion:

https://colab.research.google.com/github/sokrypton/ColabDesign/blob/main/rf/examples/diffusion_foldcond.ipynb - (requires login with a non-EPFL Google account)

ProteinGenerator: https://huggingface.co/spaces/merle/PROTEIN_GENERATOR

EvoDiff: <https://huggingface.co/spaces/colbyford/evodiff>

Inverse Folding

CARBonAra (made by me - feel free to give feedback):

<https://huggingface.co/spaces/LBM-EPFL/CARBonAra>

ProteinMPNN: <https://huggingface.co/spaces/simondurr/ProteinMPNN>

ESM-IF:

https://colab.research.google.com/github/facebookresearch/esm/blob/main/examples/inverse_folding/notebook_multichain.ipynb

In silico validation:

ColabFold (enhanced AlphaFold2):

<https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>

AlphaFold3: <https://golgi.sandbox.google.com/>

Chai-1: <https://lab.chaidiscovery.com/>

Boltz-1: <https://huggingface.co/spaces/simondurr/boltz-1>

ESM-Fold:

<https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/ESMFold.ipynb>

AlphaFold-Multimer:

https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/beta/AlphaFold2_complexes.ipynb

OmegaFold:

<https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/beta/omegafold.ipynb>