

# Web-based Protein Design Cheat Sheet

## **Potential targets:**

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

KIT (PDB:2E9W), NRP1 (PDB:2QQI), PDGFR $\alpha$  (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

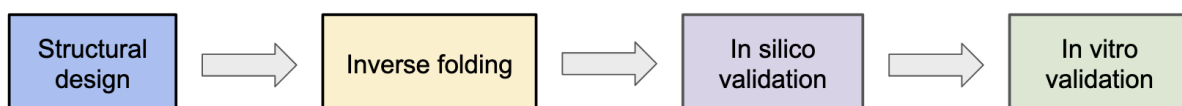
Source: <https://www.biorxiv.org/content/10.1101/2024.10.11.617496v2.full>

## **Example project**

Running one of the fully integrated pipelines (e.g. BindCraft, RFDiffusion on Colab) and then cross-validating it with AF3, Chai-1 or Boltz and comparing the metrics.

## **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 hugging face 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 hugging face some models may be inactive and it can take some minutes to rebuild them. When you generate results from Google Colab make sure to download them, they will not be stored in your Google Drive. A note to consider about BindCraft is that it will not converge if the metrics are not good enough so in order to obtain successful binders it is advised to pick very small proteins (or alternatively really easy targets - not in this list).

### **Available methods**

*Pipelines (combining two or more methods):*

ColabDesign:

[https://colab.research.google.com/github/sokrypton/ColabDesign/blob/main/rf/examples/diffusion\\_ori.ipynb](https://colab.research.google.com/github/sokrypton/ColabDesign/blob/main/rf/examples/diffusion_ori.ipynb)

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

BindCraft (AF2-backpropagation):

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

*Structural design*

AFDesign: <https://colab.research.google.com/github/sokrypton/ColabDesign/>

RFDiffusion (conditional diffusion possible):

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

ProteinGenerator (only unconditional, but you can write in the sequence):

[https://huggingface.co/spaces/merle/PROTEIN\\_GENERATOR](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](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/> on <https://huggingface.co/chaidiscovery/chai-1>

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](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>

BioEmu:

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