Engenharia Computacional de Proteínas

Instrutores: Roberto Lins, Danilo Coêlho, Elton Chaves

Tutorial elaborado por Carlos Cruz e Roberto Lins, inspirado no tutorial disponível no site: https://github.com/RosettaCommons/RFdiffusion

Generating a novel antibody mimetic using RosettaFold diffusion against myeloid cell leukaemia 1, as a cancer therapy strategy

The power of antibody mimetics

“The design of antibody mimetics holds great promise for revolutionizing therapeutic interventions by offering alternatives to conventional antibody therapies. Structure-based computational approaches have emerged as indispensable tools in the rational design of those molecules, enabling the precise manipulation of their structural and functional properties.” Chaves et al., FEBS Open, 2024, https://doi.org/10.1002/2211-5463.13855. Read the review article to find out more about the advantages of antibody mimetics and the different strategies to design them.

How does RFdiffusion work to generate an antibody mimetic?

Also, according to Chaves et al., “RFdiffusion is an updated version of RoseTTAFold that uses denoising diffusion probability models (DPPMs) to generate low-resolution backbone models, and then use the ProteinMPNN network to subsequently design sequences encoding these structures. Binders are designed similarly to creating photo-realistic images from textual instructions, resulting in novel proteins with higher binding potential and experimental success. The denoising process acts on a random sample of residue backbone, through an interactive DL-based design workflow, which disrupts coordinates toward true proteins, by minimizing the mean square error (MSE) to design the sequences. The authors demonstrated that RFDifusion is capable of generating binders for proteins used as target context, by selecting input residues in the target chain (defined as hotspots) to which the designed chain binds.

Application background

According to researchers from Astrazeneca: “Disruption of the normal apoptotic function is often observed in cancer, where cell death is avoided by

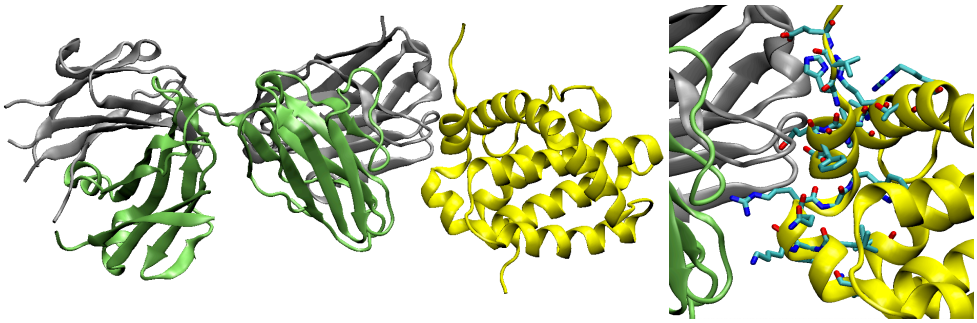
the overexpression of anti-apoptotic proteins of the Bcl-2 (B-cell lymphoma 2) family, including Mcl-1 (myeloid cell leukaemia 1). This makes Mcl-1 a potential target for drug therapy, through which normal apoptosis may be restored by inhibiting the protective function of Mcl-1." Luptak, et al., Acta Cryst. D 75: 1003-1014, 2019 (https://doi.org/10.1107/S2059798319014116). As part of this effort, they have solved antibody fragments against Mcl-1 (PDB ID: 6QB6).

Hands on using the RFdiffusion colab

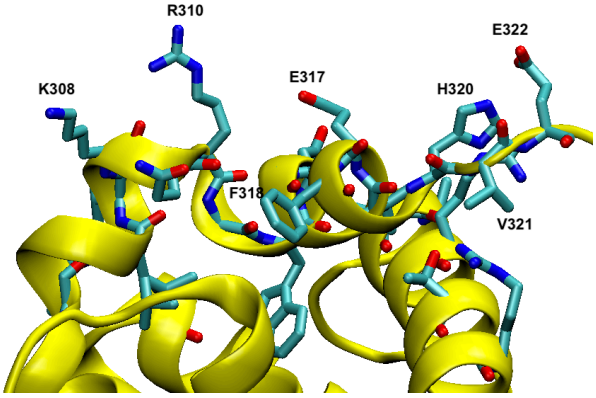
1. Access: https://colab.research.google.com/github/sokrypton/ColabDesign/blob/v1.1.1/rf/examples/diffusion.ipynb

2. First setup your environment, so that the RFdiffusion and its dependencies are locally installed by clicking on ▶ icon under "setup RFdiffuion". Wait until it completes the installation.

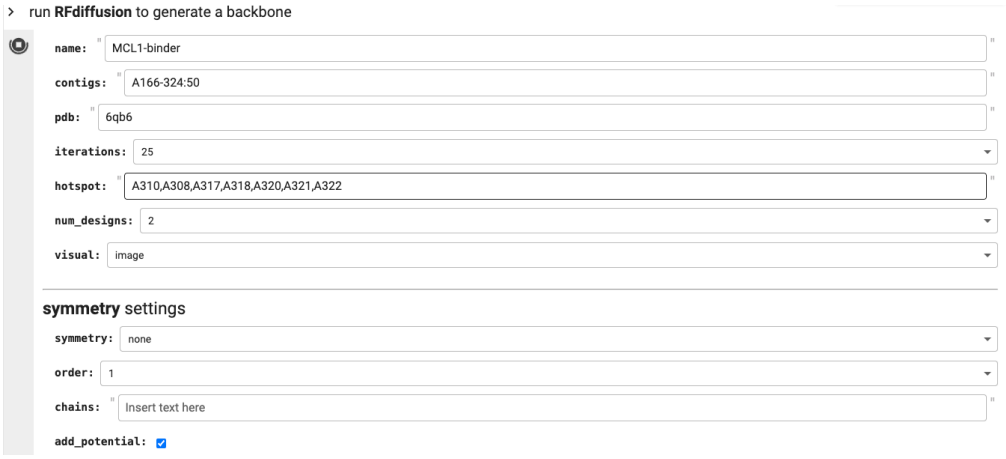
3. Before running the design, we need to inspect the antibody-antigen complex and define key-residues to guide our design. Below is a cartoon representation of the antibody chains (green and gray) complexed to Mcl-1 (yellow) (left) and their interaction interface (right), where interacting residues from the Mcl-1 protein are shown in stick model.



Look at those residues in detail in the figure below (figure is rotated by 90°). It is a mostly hydrophilic interaction surface. For protein design, it is crucial to include a mix of hydrophilic and hydrophobic interactions. Therefore, we will select the following highlighted residues (here called as hotspots) to guide the design of our antibody mimetic.



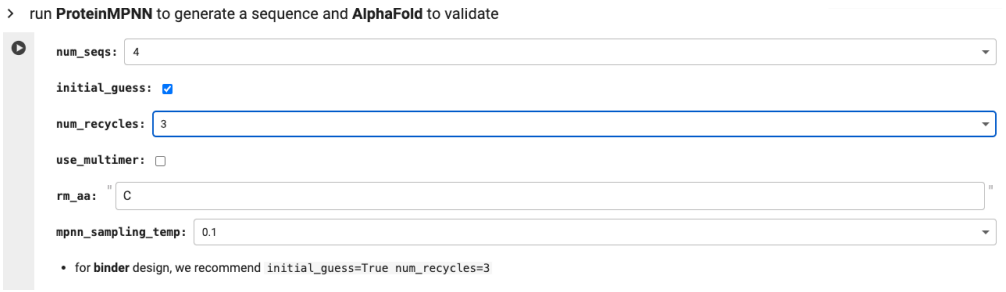
4. Let's now run RFdiffusion to iteratively design an antibody mimetic binder backbone around the previously defined Mcl-1 hotspots residues (binding site).



The commands above inform the software to design a 50-residues long miniproteins backbone around the previously selected residues. It will perform 25 iterations and generate 2 designs. The "add_potential" box is checked to ensure steric clashes are avoided during the design.

5. Once it is finished you can inspect the generated structures by running the "Display 3D structure" module.

6. It is now time to generate sequences for our backbones by running ProteinMPNN (details can be found in Dauparas, et al., Science, 378: 49-55, 2022, https://www.science.org/doi/10.1126/science.add2187) and subsequently validate them using AlphaFold (Jumper et al., Nature, 596, 583-589, 2021, https://www.nature.com/articles/s41586-021-03819-2). We will use the following parameters below to generate 4 sequences for each model.



7. After that, you can display the best results and download them for further triage of the most promising candidates.

8. Using your favorite visualization software (*e.g.,* PyMol, VMD, Chimera), download the results and inspect the interface interactions. Discuss the quality of the generated models.

Conclusion

This tutorial was intended to give you a brief introduction on the use of RFdiffusion. The method can also be used for other applications in protein design (please see Watson et al., Nature, 620: 1089-1100, 2023, https://www.nature.com/articles/s41586-023-06415-8). It is also important to keep in mind that RFdiffusion works best when used in stand-alone mode, locally installed in your computer server, as ideally you should run a larger number of iterations per design and should generate several thousand or even a few hundreds of thousand decoys to be triaged as promising candidates. Finally, visualization of the selected candidates is a crucial step before attempting experimental validation.