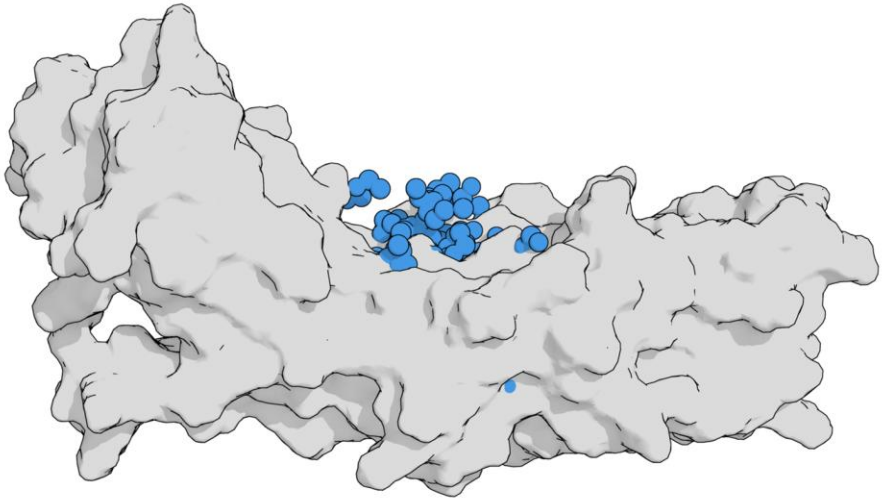
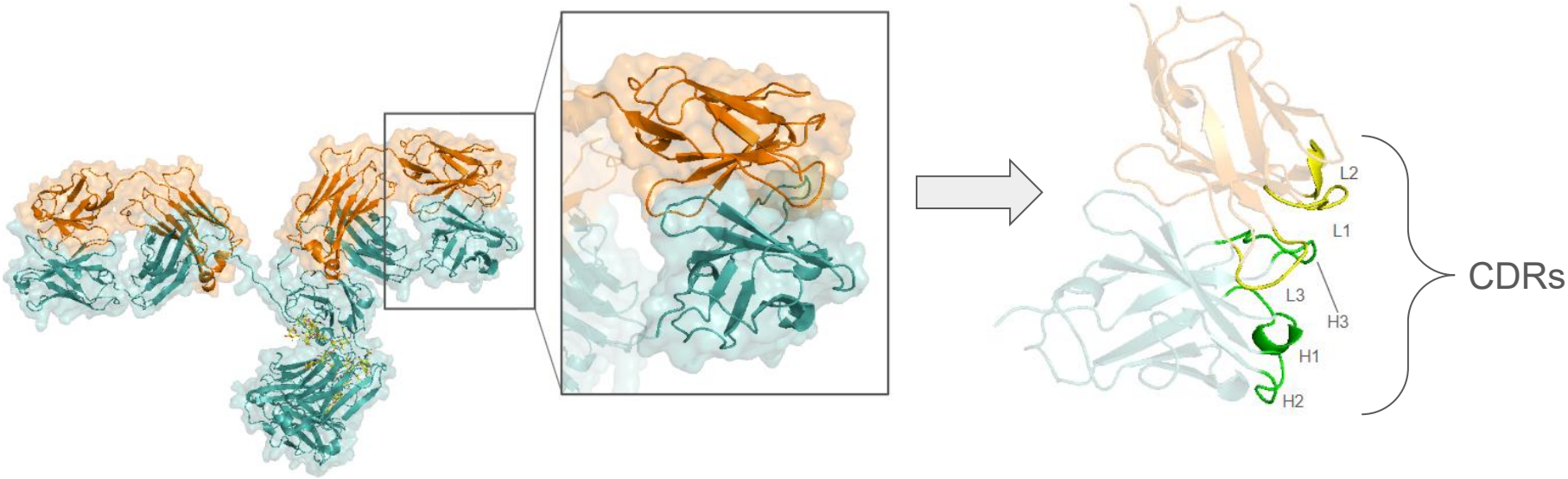


RFantibody: A pipeline for *de novo* antibody CDR design

April 15, 2026

Institute for Protein Design
University of Washington

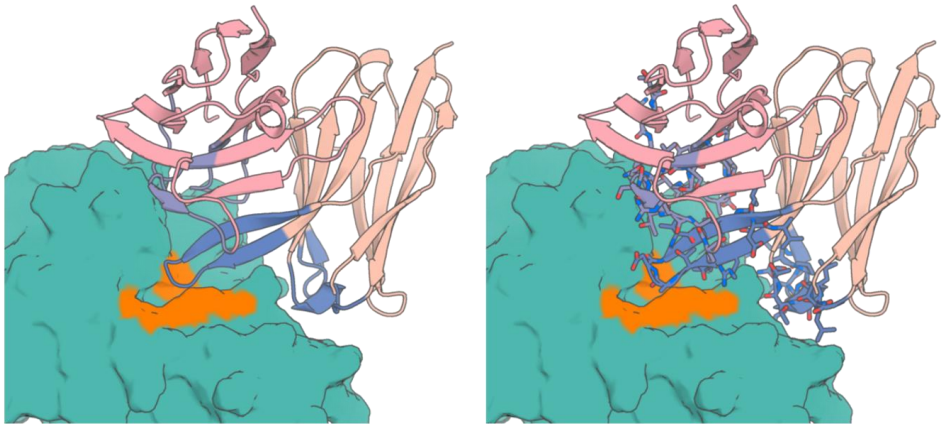
The RFantibody pipeline is an *in silico* pipeline for the design of conventional antibody and camelid-derived single-domain antibody (VHH) CDRs



- The complementarity-determining regions (CDRs) of antibodies form their primary antigen-binding site
- Each chain–heavy (teal) and light (orange)–contains 3 CDRs
- The combined diversity of these CDRs may allow for high affinity and specificity against a broad variety of targets

The RFantibody pipeline is an *in silico* pipeline for the design of conventional antibody and camelid-derived single-domain antibody (VHH) CDRs

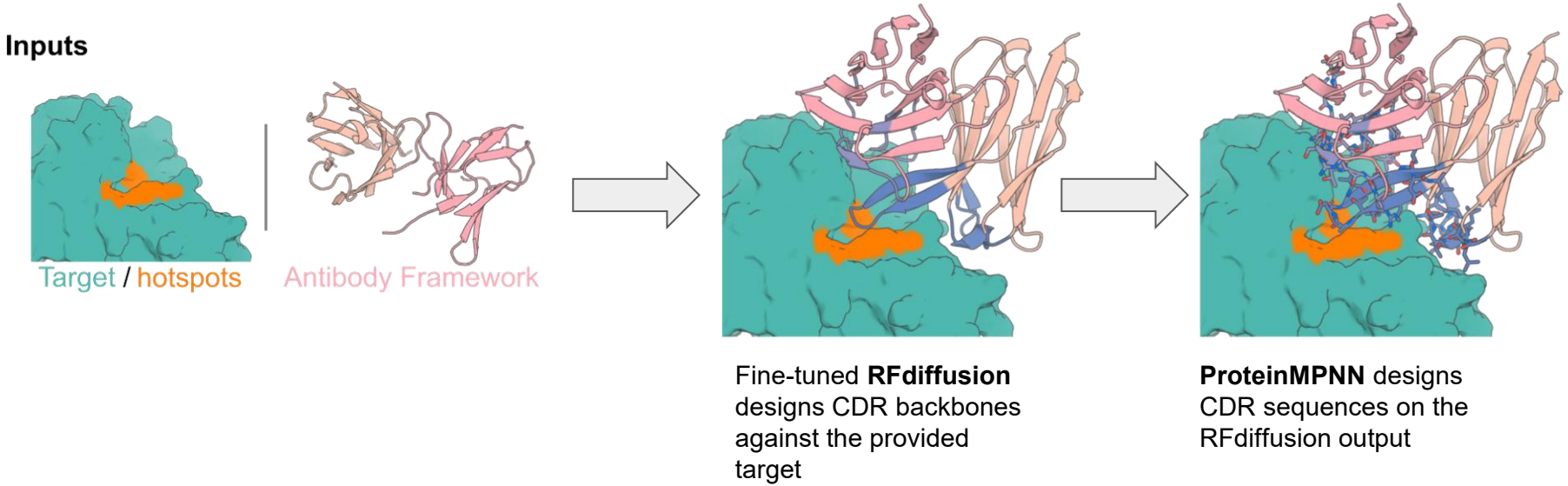
- Original RFantibody was first released in 2024
  - New version anticipated for release later this year
- Based on the first iterations of RFdiffusion and ProteinMPNN, fine-tuned to design **Chothia-defined CDRs** of antibodies and VHHs against protein antigens



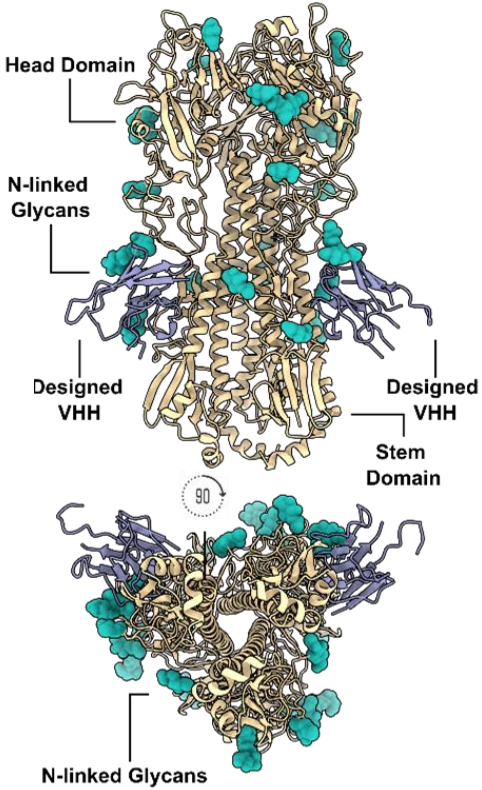
RFdiffusion-Antibody

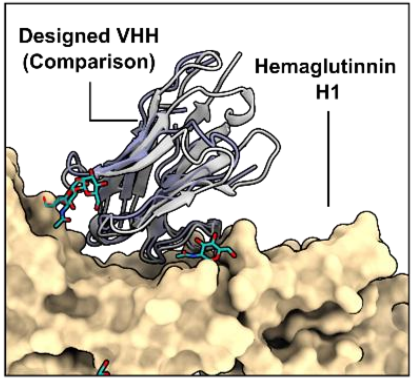
ProteinMPNN

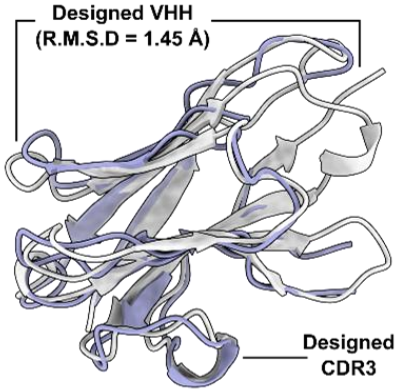
The RFantibody pipeline is an *in silico* pipeline for the design of conventional antibody and camelid-derived single-domain antibody (VHH) CDRs



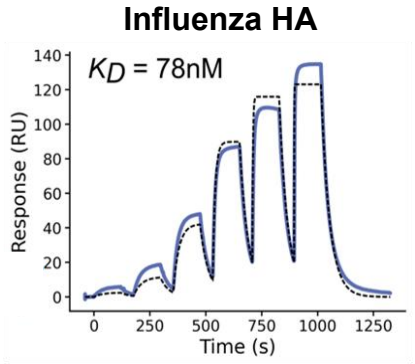
RFantibody has been used to successfully design nanobodies to a specified epitope

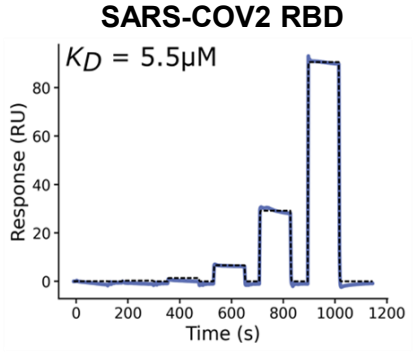


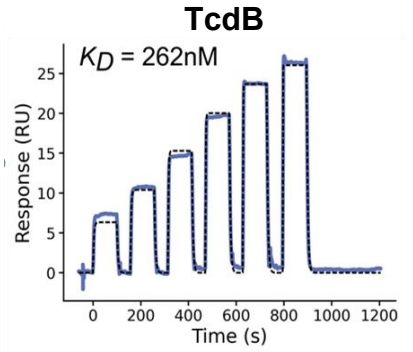




Flu HA (strain A/USA:Iowa/1943 H1N1)





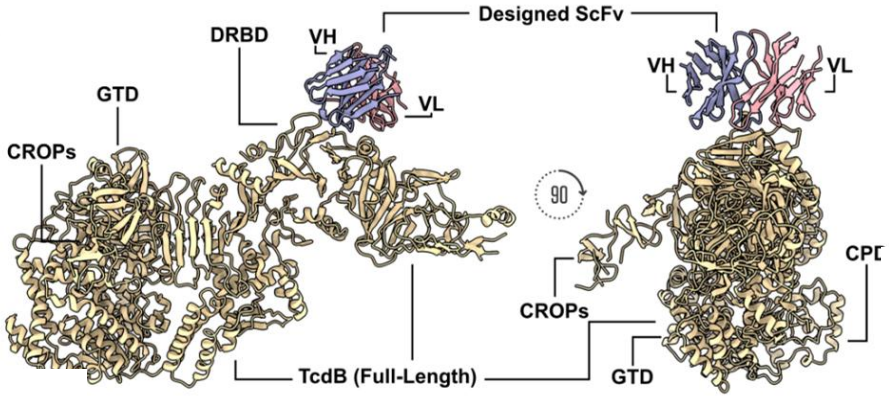


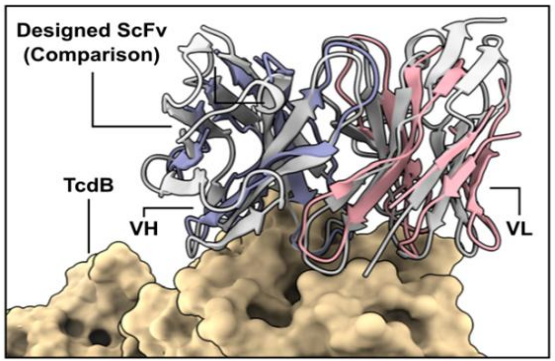

Andrew Borst

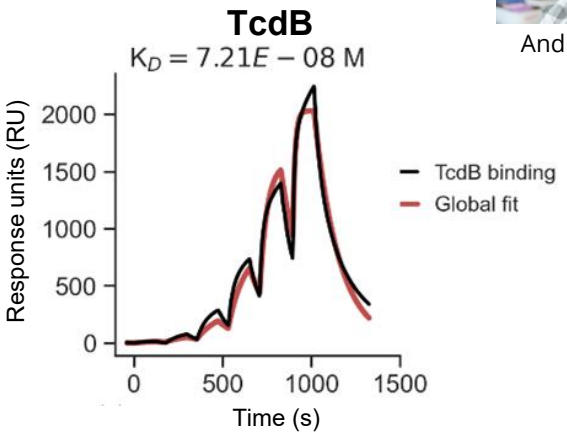

Connor Weidle

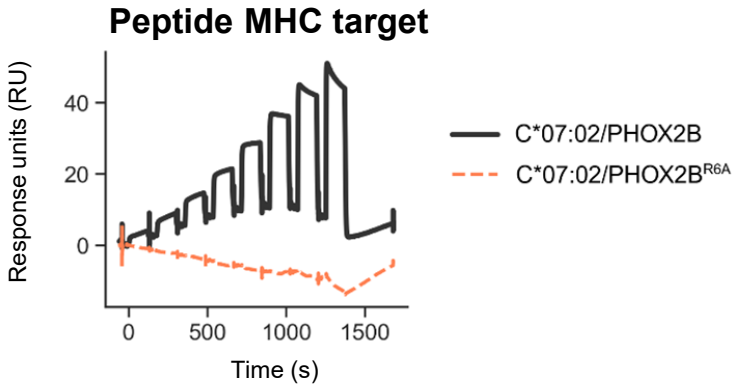

Rob Ragotte

Combinatorial shuffling of designed heavy and light chains have yielded epitope-specific scFvs










Andrew Borst


Connor Weidle


Rob Ragotte

Bennett et al. Atomically accurate de novo design of antibodies with RFdiffusion (bioRxiv, 2025)

RFantibody pipeline overview

Design structure

RFdiffusion: deep learning (DL) model trained to design the backbone of a protein or motif given the structure of a target or scaffold

Design sequence

ProteinMPNN: DL model trained to optimize an amino acid sequence fitting a provided protein backbone

Filter

RFantibody: DL model trained to predict the probability of an Ab binding an antigen given sequence and structure of the complex

RFantibody pipeline can alternatively be run with AlphaFold3 as binding oracle (recommended)

Design structure

RFdiffusion: deep learning (DL) model trained to design the backbone of a protein or motif given the structure of a target or scaffold

Design sequence

ProteinMPNN: DL model trained to optimize an amino acid sequence fitting a provided protein backbone

Filter

AlphaFold3: State-of-the-art protein-protein complex prediction model outputting structure with confidence metrics

Section Overview

I. Setup
II. Framework preparation
III. Target preparation
IV. Running RFdiffusion
V. Running ProteinMPNN
VI. Structure prediction and analysis

Setup

- We recommend running RFantibody on GPU nodes within a high-performance computing (HPC) cluster for library-scale projects. See the RFantibody README for instructions for installing RFantibody.

II. Antibody framework preparation

- **You will need an antibody structure** (one VH/VL pair or one VHH) to pass into RFdiffusion for your framework.
- For this workshop, the structure should already be **Chothia-numbered** for CDR identification. If you know the PDB ID for your antibody framework, a Chothia-numbered version can be easily downloaded from SAbDab.
- If the antibody is part of a complex: open in PyMOL, select your antibody chains, copy that to a new object, and save that object as a PDB.
    a. **If the VH/VL are part of a Fab or full IgG, it is strongly advised to truncate it down to the variable domains**

III. Target preparation and hotspot selection

1. **Crop your target protein** around the epitope you want to bind in PyMOL
2. Run the cleanup script. This script will remove heteroatoms, combine all chains into a single chain, and renumber residues so they're continuously numbered starting from 1

III. Target preparation and hotspot selection

What are hotspots? Hotspots are solvent-accessible residues on your target that you want the designed antibody to contact.

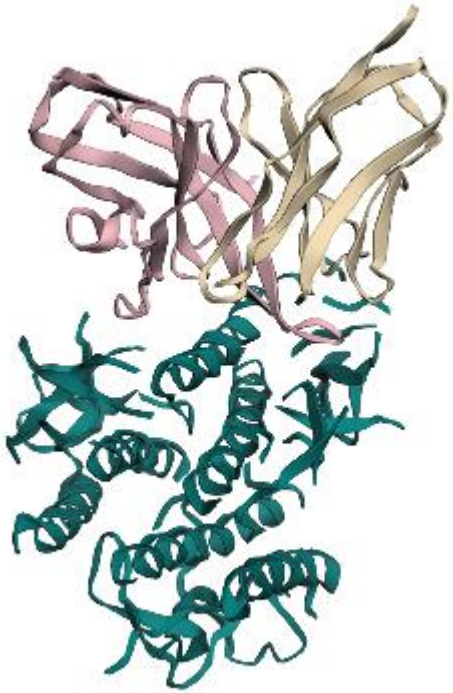
- When choosing hotspots, begin with 3 and try to arrange them in a triangle across the epitope. If your hotspots are too close together, they're redundant. If they're too far apart — so far that no single antibody could satisfy all of them at once — then they're not going to be meaningful constraints. The triangle approach hits a nice balance.

IV. Running RFdiffusion

1. **Specify the desired CDR loop lengths.** A typical approach is to keep a smaller length range for CDR H1 and H2, and leave more diversity — a wider range — for CDR H3, since CDR H3

IV. Running RFdiffusion

- Once RFdiffusion finishes, you'll see your output files. At this point, you can download them and inspect in PyMOL (at this stage, the designed backbones should have a glycine string when viewing the sequence)
- Target is **teal**, heavy chain is **pink**, light chain is **wheat**



V. Running ProteinMPNN

Next, prepare the hyperparameters for ProteinMPNN. **A few considerations**:

- Generating **more sequences per backbone** improves chances of a viable binder with that given structure
- **Increasing the temperature increases sequence diversity.** Increase no higher than 0.2-0.3 — higher than that can result in lower design quality
- **Omitting residues:** We strongly recommend omitting **cysteine** in antibody and VHH design to avoid potential sites of oxidation and inadvertent disulfide bonds. It may also be beneficial to omit **methionine** from the CDRs, since methionine is also prone to oxidation

VI. Structure prediction and analysis: when not to use AF3

AF3 confidence metrics are more predictive of binding than the RoseTTAFold-based oracle described in the RFantibody manuscript. Best to use AF3 **unless:**

- Your target is too large (over 5,000 amino acids) and can't be reasonably cropped
- Your target is a viral target or otherwise restricted by the AlphaFold server, and you don't have access to a local AF3 installation
- You've hit the server prediction limit (30 predictions per day per email address)

VI. Running AF3

1. Run the AF3 input generation script on your ProteinMPNN output
2. Head over to the AlphaFold3 web server, upload the JSON, and submit your predictions.

Alternately, you can run RF2 (or RF3) locally from the command line. (Using different input formats.)

VI. Structure prediction and analysis

Download your results from the AF3 server (or from the local run). You can visually inspect the predicted structures in PyMOL to get an initial sense of the dock and whether the CDR loops look reasonable.

There are 2 metrics you want to look at: **ipTM** and **pAE**:

ipTM (interface predicted TM-score) is a confidence metric for the interface between your antibody and target. Higher = more confident.

pAE (predicted aligned error) is a measure of positional uncertainty. Lower pAE values between chains suggest lower uncertainty/greater confidence in their relative positions.

You can also calculate the **RMSD** between your RFdiffusion backbone and the AF3 prediction. Low RMSD = greater agreement between diffusion output and predicted dock

Thank you!

If you have any questions, feel free to email dlsee@uw.edu or ritib@uw.edu