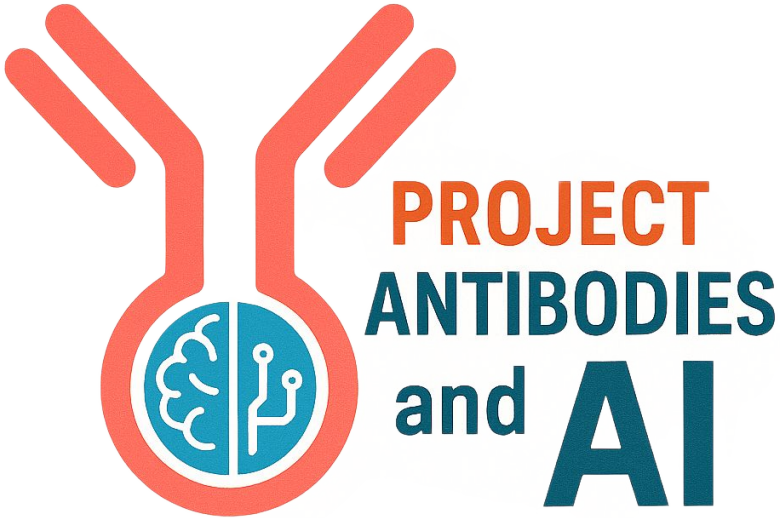
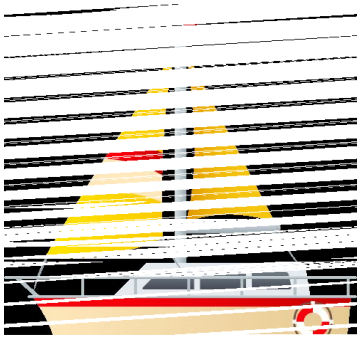


Goals & Progress

Team

Aamir "Moh"	**Hussain**
Eric	**Olson**
Jessica	**Adjoyi**
Michael	**Wu**
Rebecca	**Lind**
Todd	**Smith**
Wengang	**Zhang**

Day 1 - Jetstream maintenance



Days 2,3 - focus on coding, data

Two Grand Challenges: Disease & Antibody Engineering

Universal Diagnostic:

Given an antibody/TCR sequence can we say what antigen it binds

Universal Design:

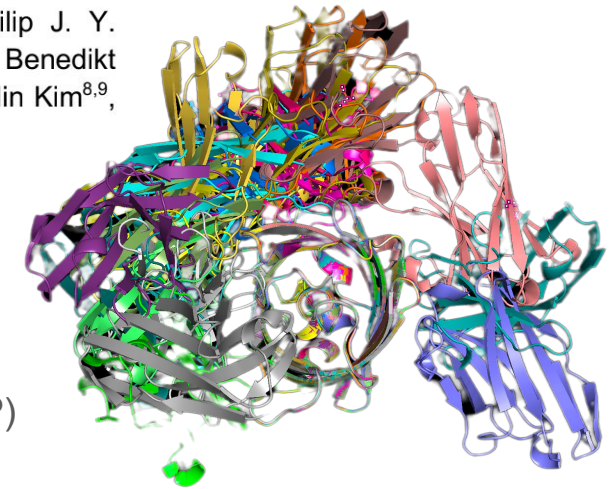
Given an epitope can we design an antibody *de novo*

RFAntibody

Atomically accurate de novo design of single-domain antibodies

Nathaniel R. Bennett[‡1,2,3], Joseph L. Watson[*‡1,2], Robert J. Ragotte[‡1,2], Andrew J. Borst[‡1,2], Déjenaé L. See[#1,2,4], Connor Weidle[#1,2], Riti Biswas[1,2,3], Ellen L. Shrock[1,2], Philip J. Y. Leung[1,2,3], Buwei Huang[1,2,4], Inna Goreshnik[1,2,5], Russell Ault[6,7], Kenneth D. Carr[2], Benedikt Singer[1,2], Cameron Criswell[1,2], Dionne Vafeados[2], Mariana Garcia Sanchez[2], Ho Min Kim[8,9], Susana Vázquez Torres[1,2,10], Sidney Chan[2], David Baker[*1,2,5]

Mar. 2024



20 Nanobody designs with Green Fluorescent Protein (GFP)

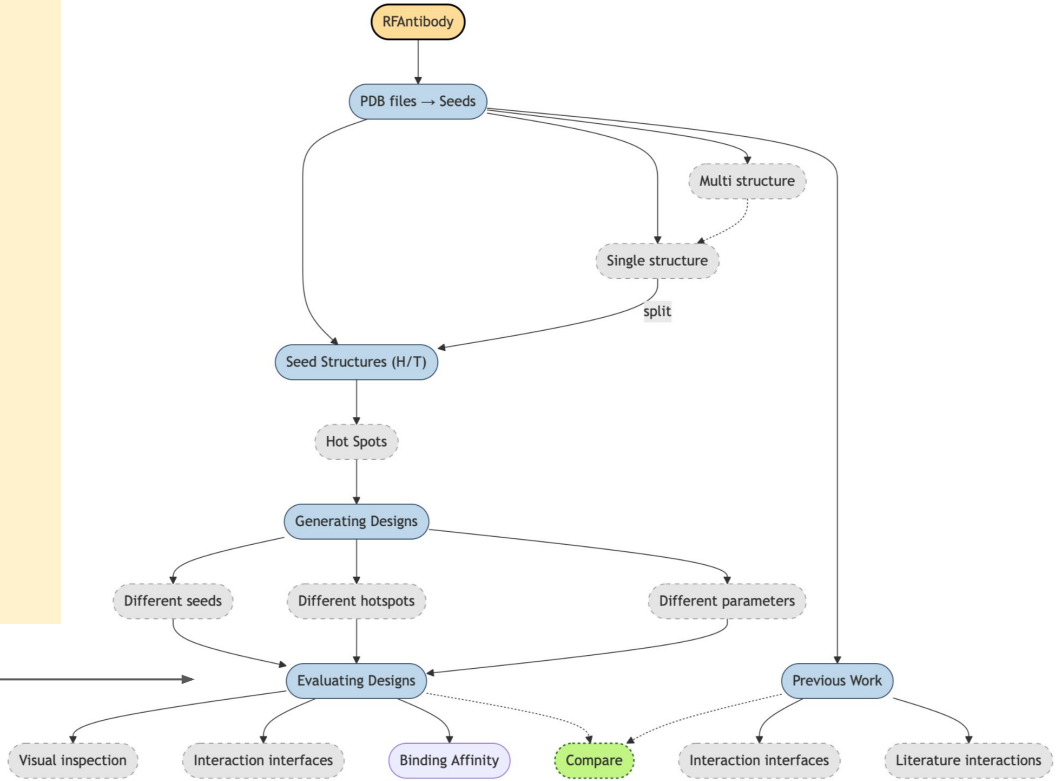
Big Question: Of 100 designs, which ones are good?

- 23 Nanobody:FP PDB files
- Literature values
- Tools for identifying interactions
- Tools to view designed structures
- Tools to estimate binding affinities
- AI tools: coding, processing, data extraction

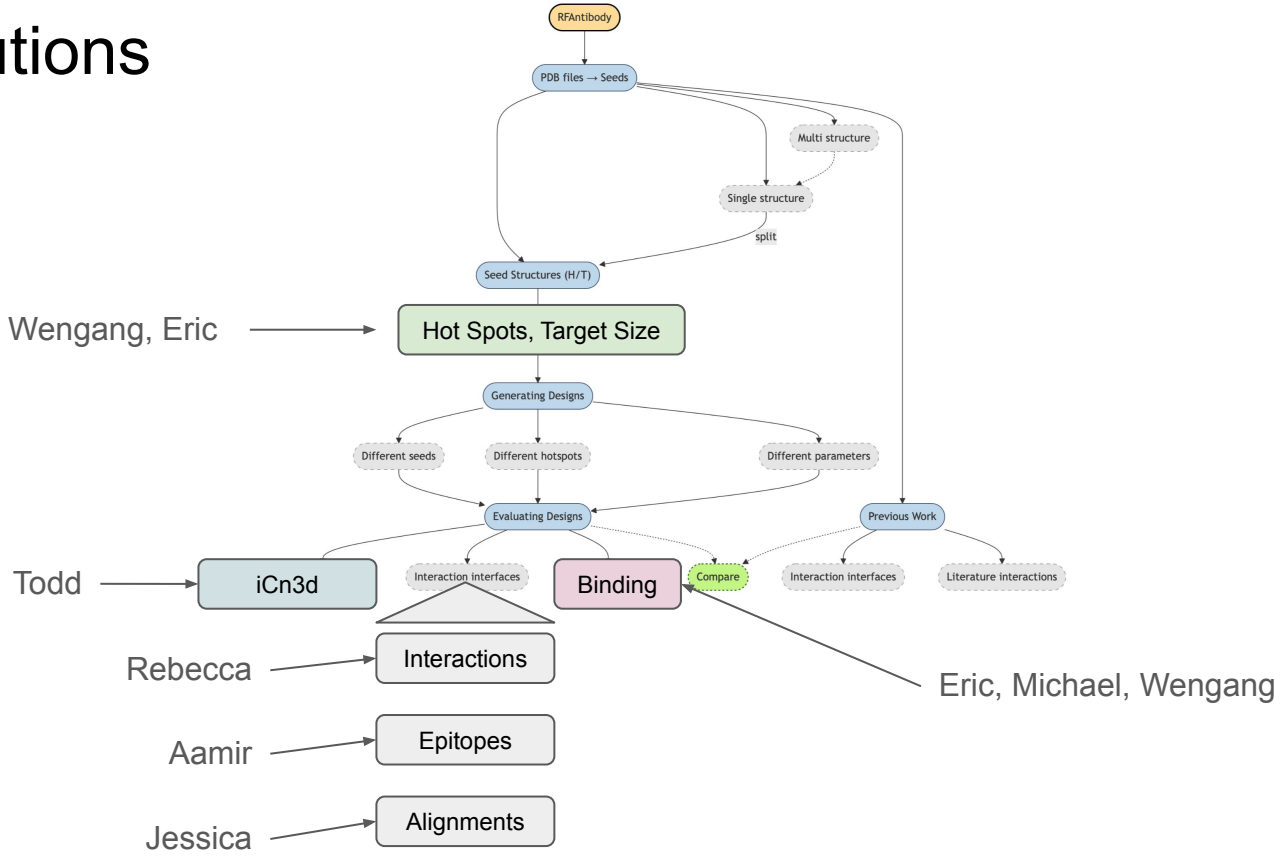
Work:

- Build an assessment pipeline that used affinity prediction as a key measure
- Test RFAntibody parameters

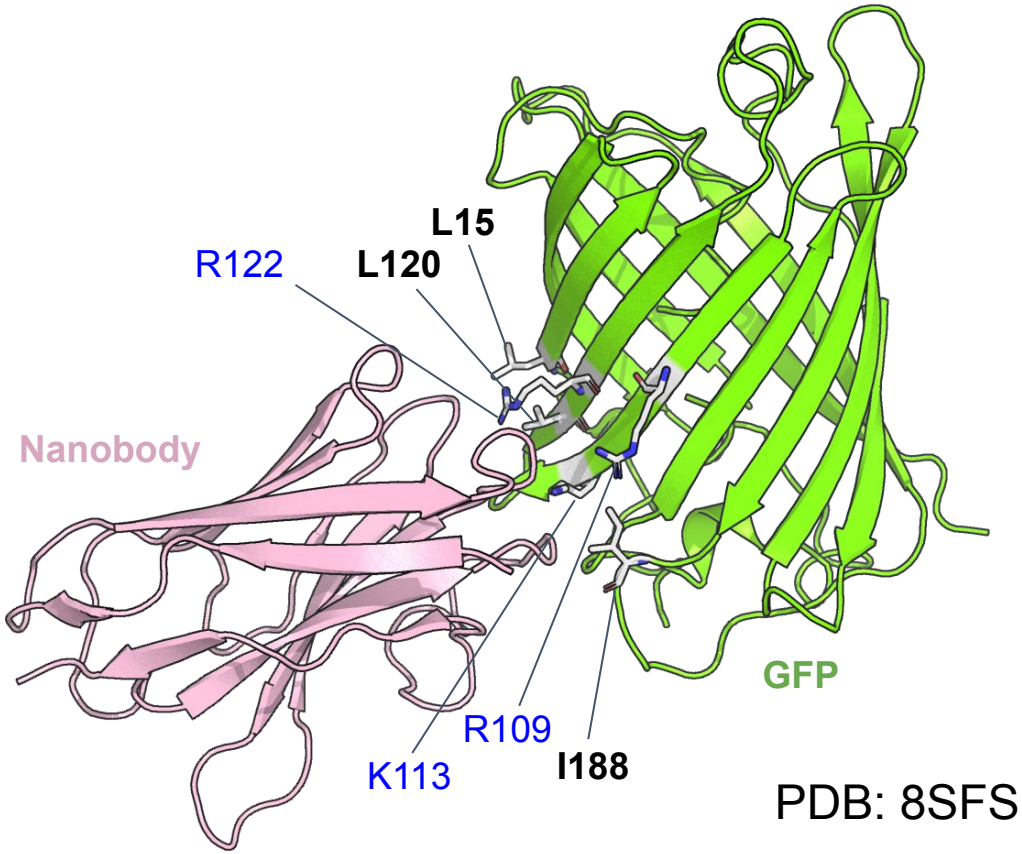
Critical Aspect



Contributions

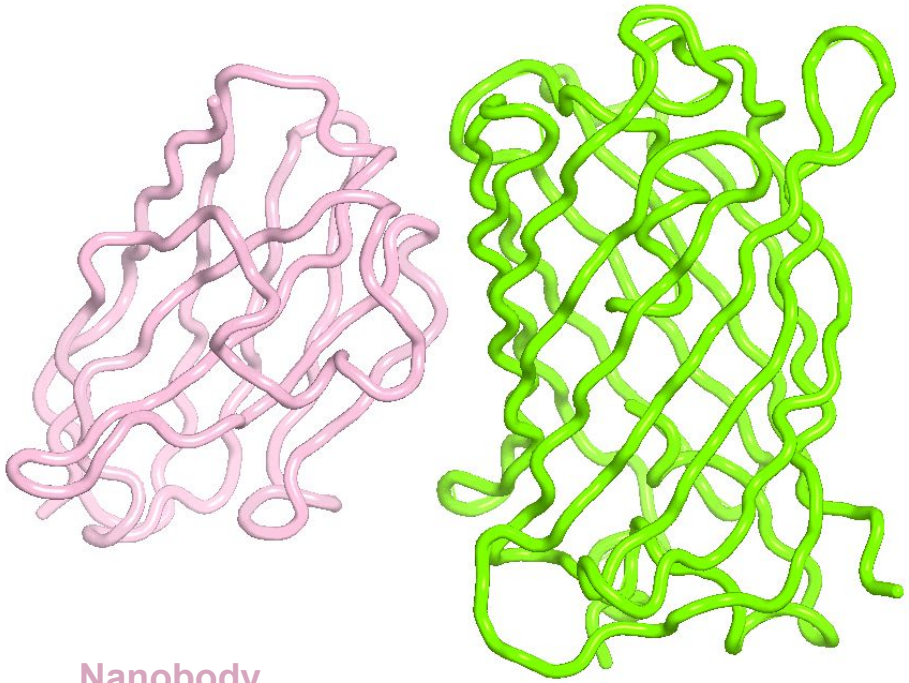


Alanine scanning identifies hot spot residues on GFP



PDB: 8SFS

Diffusion process for nanobody design

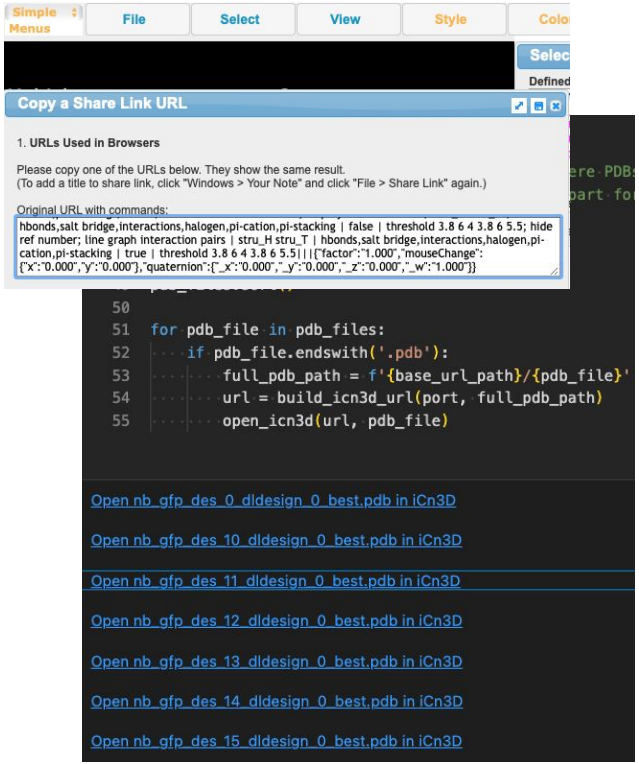


How do we evaluate the *de novo* nanobody design?

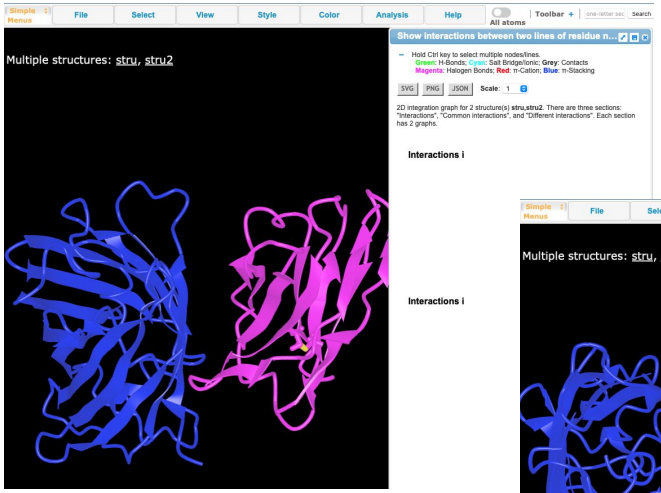
- The stock RFantibody recommends **geometry-based** filtering strategy.
- We propose using **physics-based** affinity prediction methods, such as MM-GBSA and PRODIGY, to improve the filtering process.

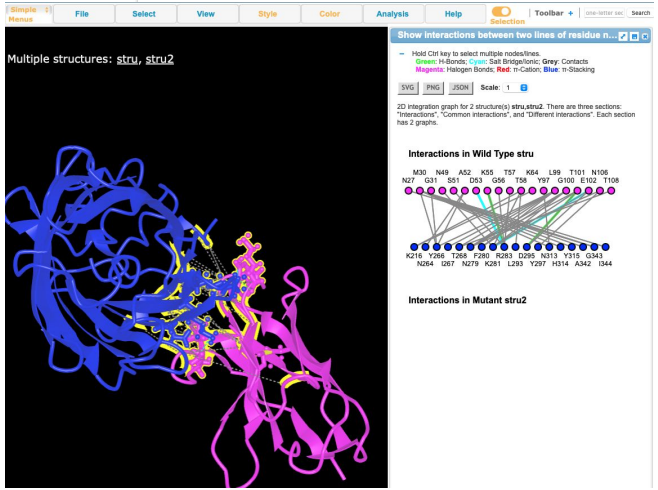
iCn3D - How do you easily view designs

1. Use “shareable link” to create clickable URLs



2. Clickable URLs create tab, control click to step through





🧪 Epitope Analysis of Nanobody Designs

```python
for pdb_file in pdb_files:
    if pdb_file.endswith('.pdb'):
        full_path = os.path.join(pdb_dir, pdb_file)
        cmd = ["node", "epitope_local.js", full_path, "H", "T"]
        result = subprocess.run(cmd, capture_output=True, text=True)

        if result.returncode == 0:
            lines = result.stdout.strip().split('\n')
            for line in lines:
                parts = [p.strip() for p in line.split(',')]
                if len(parts) == 7:
                    all_data.append({
                        'Design': pdb_file,
                        'Chain': parts[1],
                        'Residue': int(parts[2]),
                        'Residue_AA': parts[3],
                        'Partner_Chain': parts[4],
                        'Partner_Residue': int(parts[5]),
                        'Partner_Residue_AA': parts[6]
                    })
                else:
                    print(f"Skipping unexpected line in {pdb_file}: {line}")
        else:
            print(f"Error running epitope_local on {pdb_file}:\n{result.stderr}")

# ✅ Convert to DataFrame
df = pd.DataFrame(all_data)
```

	Design	Chain	Residue	Residue_AA	Partner_Chain	Partner_Residue	Partner_Residue_AA
0	nb_gfp_des_0_dldesign_0_best.pdb	stru_H	27	D	stru_T	234	K
1	nb_gfp_des_0_dldesign_0_best.pdb	stru_H	28	L	stru_T	234	K
2	nb_gfp_des_0_dldesign_0_best.pdb	stru_H	29	S	stru_T	234	K
3	nb_gfp_des_0_dldesign_0_best.pdb	stru_H	30	L	stru_T	132	D
4	nb_gfp_des_0_dldesign_0_best.pdb	stru_H	30	L	stru_T	234	K
...	...	...	...	...	...	...	...
838	nb_gfp_des_9_dldesign_0_best.pdb	stru_H	107	W	stru_T	316	Q
839	nb_gfp_des_9_dldesign_0_best.pdb	stru_H	107	W	stru_T	335	F
840	nb_gfp_des_9_dldesign_0_best.pdb	stru_H	107	W	stru_T	337	T
841	nb_gfp_des_9_dldesign_0_best.pdb	stru_H	109	K	stru_T	258	N
842	nb_gfp_des_9_dldesign_0_best.pdb	stru_H	109	K	stru_T	259	S

🎯Objective:

- **automate the extraction and comparison of predicted epitopes.**
- **using a tool called epitope_local to analyze a folder (gfp_rf2) containing 20 nanobody designs.**

🐞Challenges:

- **Output from the epitope tool wasn't always structured consistently.**

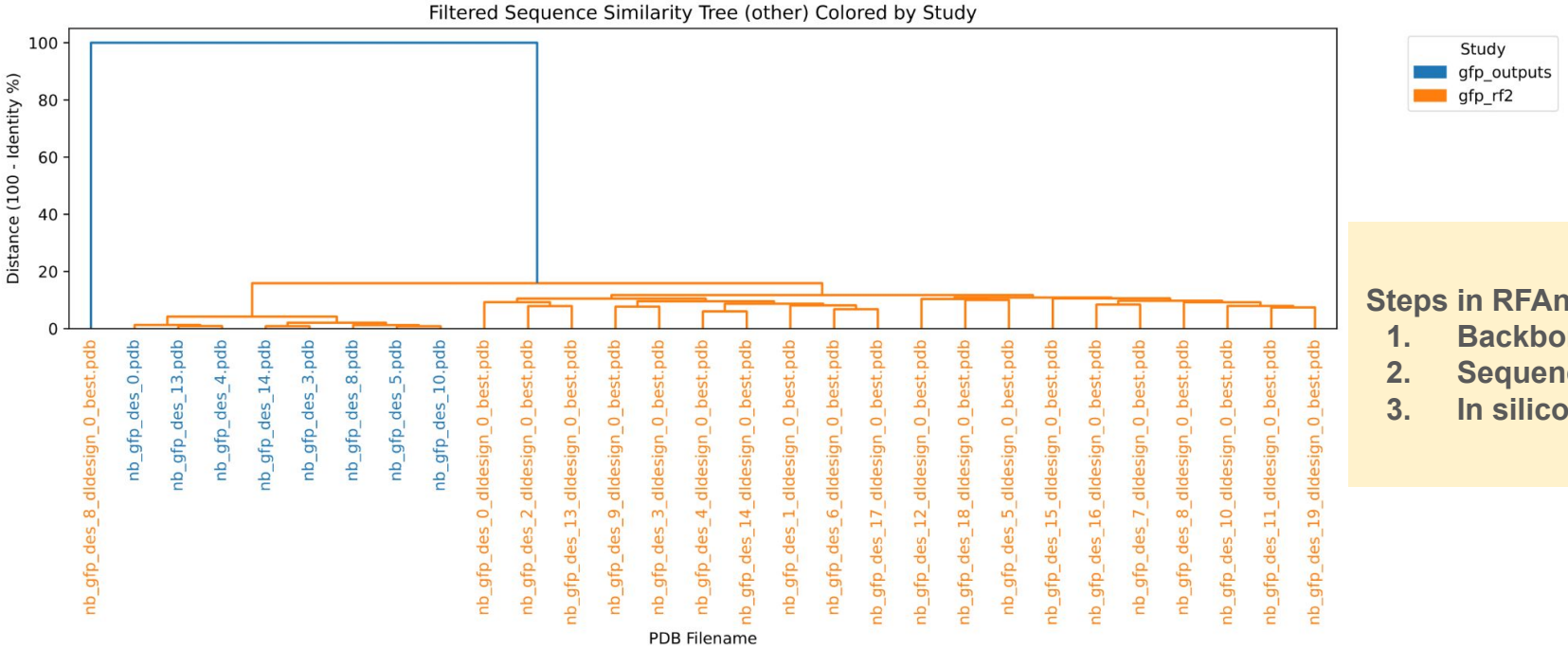
🌱Future Directions:

- **Visualize and compare residue interaction patterns across designs.**
- **Identify the most promising designs for experimental validation.**

Comparison of generated sequences via Dendroplot

Objective: Modify existing code to generate dendroplot comparing gfp_outputs (first step result) vs gfp_rf2 (result of last two steps).

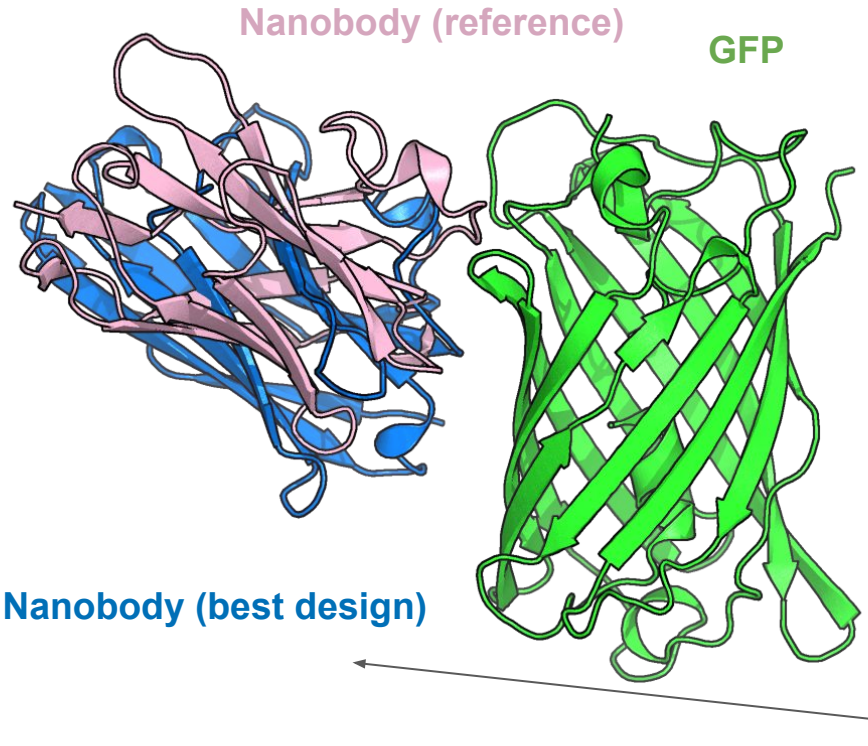
Outcome: Confirmed increased variability of nanobodies from second step



Steps in RFAntibodies

1. **Backbone design**
2. **Sequence design**
3. **In silico filtering**

Physics-based binding affinity evaluation methods identified best design



Physics-based (MM-GBSA) binding affinity evaluation: ΔG = -26 kcal/mol

Predicting antibody affinity from antibody:target structure

Prodigy - Predict the binding affinity of protein-protein complexes from structural data

https://github.com/haddocking/prodigy

prodigy tfrc_rf2/tfrc_15_dldesign_0_best.pdb --selection H,L T

prodigy -q tfrc_rf2/ --selection H,L T

```
[+] Parsed structure file tfrc_15_dldesign_0_best (3 chains, 397 residues)
[+] No. of intermolecular contacts: 55
[+] No. of charged-charged contacts: 9.0
[+] No. of charged-polar contacts: 6.0
[+] No. of charged-apolar contacts: 11.0
[+] No. of polar-polar contacts: 2.0
[+] No. of apolar-polar contacts: 14.0
[+] No. of apolar-apolar contacts: 13.0
[+] Percentage of apolar NIS residues: 39.00
[+] Percentage of charged NIS residues: 24.67
[++] Predicted binding affinity (kcal.mol-1):    -10.0
[++] Predicted dissociation constant (M) at 25.0°C:  4.7e-08
```

```
tfrc_3_dldesign_0_best    -5.653
tfrc_8_dldesign_0_best    -5.895
tfrc_12_dldesign_0_best   -6.811
tfrc_10_dldesign_0_best   -5.454
tfrc_7_dldesign_0_best    -5.643
tfrc_1_dldesign_0_best    -9.179
tfrc_18_dldesign_0_best   -9.570
tfrc_2_dldesign_0_best    -5.359
tfrc_4_dldesign_0_best    -6.845
tfrc_14_dldesign_0_best   -9.342
tfrc_13_dldesign_0_best   -8.428
tfrc_6_dldesign_0_best    -6.126
tfrc_11_dldesign_0_best   -7.392
tfrc_15_dldesign_0_best   -9.986
tfrc_9_dldesign_0_best    -6.329
tfrc_5_dldesign_0_best    -6.011
tfrc_19_dldesign_0_best   -9.721
tfrc_17_dldesign_0_best   -6.385
tfrc_16_dldesign_0_best   -5.365
tfrc_0_dldesign_0_best    -8.252
```

Plan - do comparisons using structures with know affinity for assessment of accuracy, add step to pipeline to characterize output from RFAntibody

Issues - May not show good agreement with known values, look into this more and compare to other software

Prodigy -Predicting antibody affinity from antibody:target structure

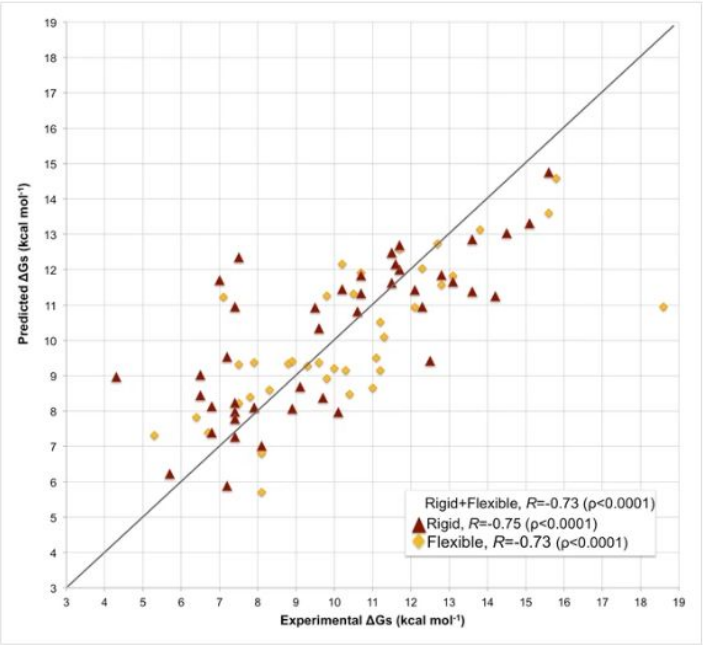
Anna Vangone, Alexandre MJJ Bonvin (2015) Contacts-based prediction of binding affinity in protein–protein complexes eLife 4:e07454

Li C. Xue, João Pglm Rodrigues, Panagiotis L. Kastritis, Alexandre Mjj Bonvin, Anna Vangone, PRODIGY: a web server for predicting the binding affinity of protein–protein complexes, *Bioinformatics*, Volume 32, Issue 23, December 2016, Pages 3676–3678,
https://doi.org/10.1093/bioinformatics/btw514
http://milou.science.uu.nl/services/PRODIGY

https://github.com/haddocking/binding-affinity-benchmark
Protein-protein binding affinity benchmark

47 structure files with affinity measures and references

Figure 3

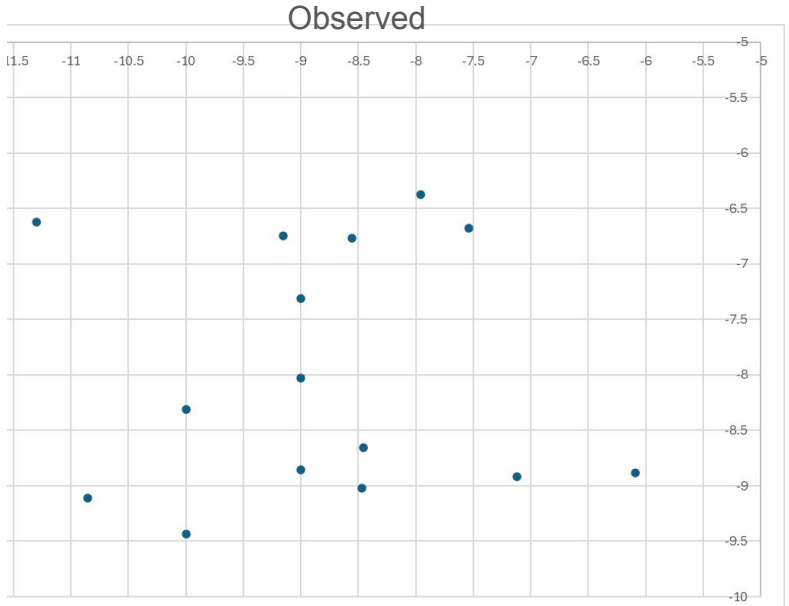


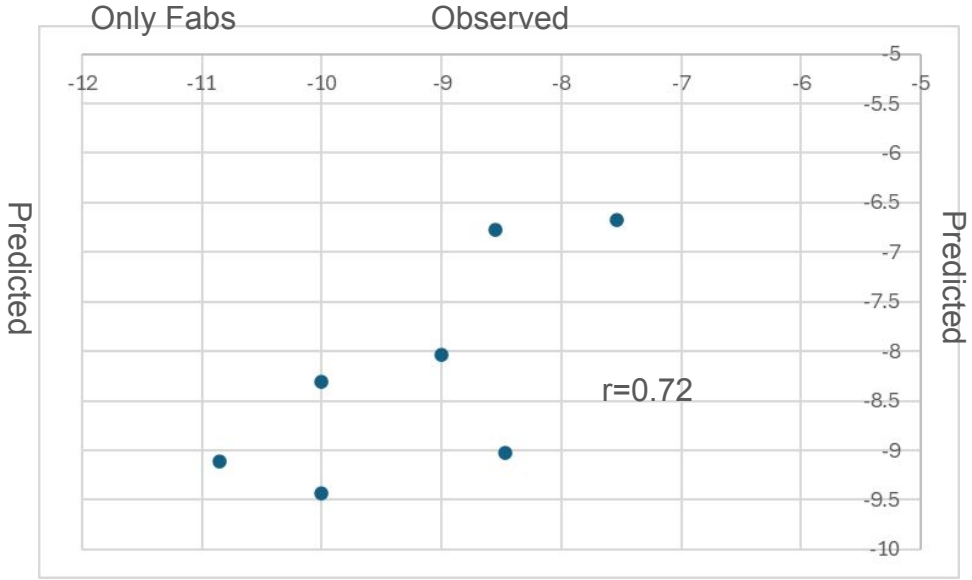
Scatter plot of predicted vs experimental binding affinities.

How well does Prodigy perform?

Selected 15 antibody:target structures from literature and affinity benchmark set

Compared Prodigy predicted dissociation constant and measured values





Ranking of GFP binders - predicted binding affinity

8sfs_0_dldesign_0_best	-9.225
8sfs_7_dldesign_0_best	-8.586
8sfs_8_dldesign_0_best	-7.766
8sfs_14_dldesign_0_best	-8.772
8sfs_10_dldesign_0_best	-8.418
8sfs_2_dldesign_0_best	-10.153
8sfs_19_dldesign_0_best	-8.822
8sfs_17_dldesign_0_best	-9.054
8sfs_4_dldesign_0_best	-10.960
8sfs_6_dldesign_0_best	-8.296
8sfs_18_dldesign_0_best	-8.205
8sfs_15_dldesign_0_best	-6.810
8sfs_5_dldesign_0_best	-8.877
8sfs_16_dldesign_0_best	-6.058
8sfs_1_dldesign_0_best	-8.140
8sfs_3_dldesign_0_best	-8.393
8sfs_9_dldesign_0_best	-9.799
8sfs_11_dldesign_0_best	-7.348
8sfs_13_dldesign_0_best	-7.579
8sfs_12_dldesign_0_best	-8.218

kd ranges from 90nM to 36uM

To do -
Generate table with additional information
Examine contacts predicted by Prodigy - use “--contact_list” and “--pymol_selection” to explore and compare with other approaches

--contact_list : Output a list of contacts
--pymol_selection : Output a script to highlight the interface

Summary and Next Steps

- Created a foundation for running RFantibody and evaluating designs
  - Explored hot spots and chain size
  - Evaluation - start on visual inspection, interaction data, epitopes, sequence analyses
  - Evaluation - examined different binding affinity programs
  - Learned about parsing output -> Python, PANDAS Data Frames and further manipulation
- Next Steps
  - Pull parts together to create pipelined analyses
  - Develop ways to go deeper into the data
  - Continue design evaluation approaches
  - One “order” the sequences and test in the laboratory
  - Evaluate Prodigy more deeply - different formats of antibodies, comparison with other methods, larger set of structures, confirm the values from literature

Key Takeaways:

- The stock RFantibody recommends **geometry-based** filtering strategy.
- We propose using **physics-based** affinity prediction methods, such as MM-GBSA and PRODIGY, to improve the filtering process.