Benchmarking Methods and Tools for an End-to-end Antibody Design Framework: a Proof-of-concept

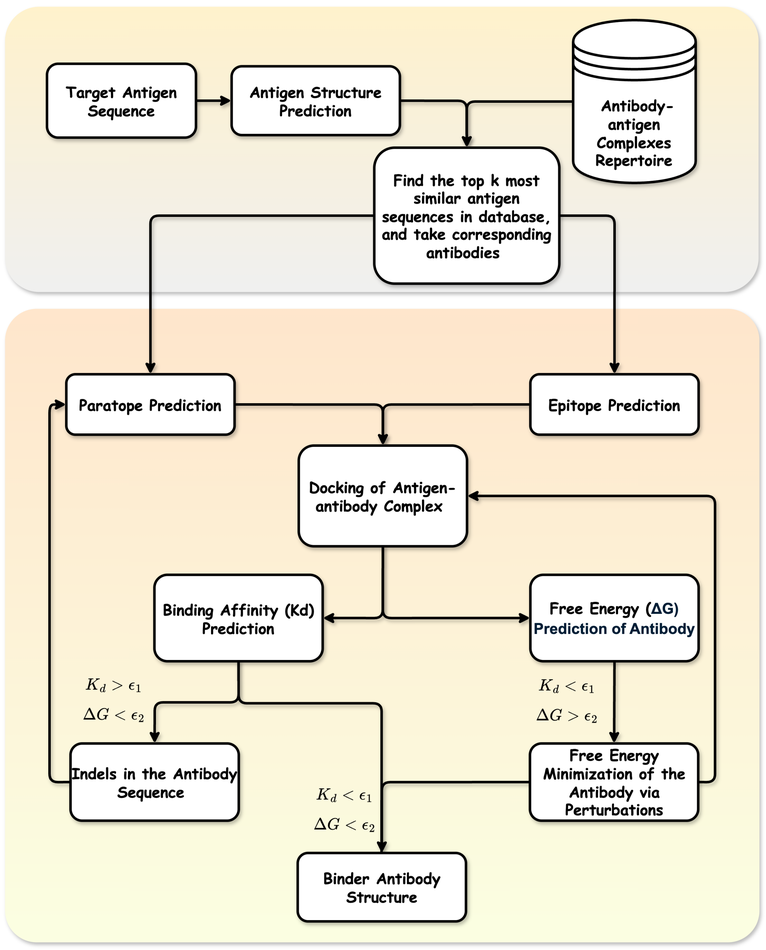
Mansoor Ahmed, Spencer VonBank, Belle Divine, and Murray Patterson

Introduction

- Antibodies are essential tools in therapeutics, diagnostics, and biomedical research
- The conventional wet-lab-based approaches are costly, time-intensive, and experimentally demanding
- We aim to benchmark state-of-the-art methods and tools into a unified pipeline that can design antibodies for any given target antigen sequence

Problem Formulation

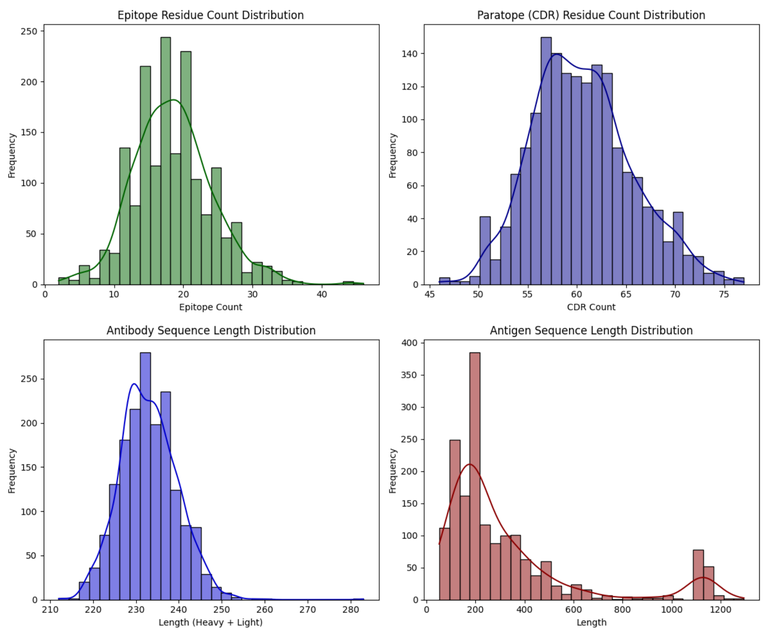
Given a target antigen sequence $S_{\text{ag}} = (a_1, a_2, ..., a_n), a_i \in \{A, R, N, ...\}$, predict the high affinity binder antibody structure $X_{\text{ab}=}\{(x_i, y_i, z_i)\}_{i=1}^{n} \in \mathbb{R}^{3n}$



Antibody Repertoire - Expolaratory Data Analysis

We curate an antigen-antibody repertoire using the AsEP dataset, a high-quality benchmark of 1,723 antibody–antigen complexes from AbDb. Each sample includes:

- Paired PDB complex structure and torch tensors containing graph and sequence data
- Residue-level masks for CDRs, antigen surface, and epitope sites



Antigen Structure Prediction

We employ **AlphaFold** for antigen structure prediction (a state-of-the-art deep learning model developed by DeepMind). We compute the following metrics to benchmark its performance:

- PTM (Predicted TM-score): A confidence metric estimating the overall structural accuracy.
- PAE (Predicted Aligned Error): Provides residue-level error estimates between different parts of the structure.

Size Range	Runtime	PTM	PAE
50–100 AA	~1min	0.83	0.76
200–300 AA	~2min	0.81	0.79
800–1200 AA	~10min	0.74	0.85

Table: Performance metrics across different antigen sizes.

Antibody Repertoire Search via Sequence Alignment

Size Range	Runtime (1 Iteration)
50–100 AA	~2 sec
200–300 AA	~3 sec
800–1200 AA	~5 sec

Table: PSI-BLAST runtimes across different sequence lengths (using single iteration).

Antibody Binding Site (Paratope) Prediction

We benchmark the following tools for paratope prediction:

- **ParaGraph** is a graph-based deep learning model that predicts paratopes using antibody 3D structures as input, capturing residue-level spatial features through geometric message passing.
- **ParaSurf** identifies paratope residues using physicochemical surface properties of the antibody, without requiring the antigen.
- **ParaAntiProt** leverages protein language model (PLM) embeddings from AntiBERTy and structural encoders to predict paratopes directly from antibody sequences.

Table: Paratope prediction performance comparison.

Method	AUC ROC	F1	PR ROC	MCC
ParaGraph	0.902	0.632	0.534	0.621
ParaSurf-Paragraph Dataset	0.927	0.654	0.545	0.634
ParaSurf	0.932	0.661	0.555	0.639
ParaAntiProt	0.921	0.622	0.567	0.623

Antigen Binding Site (Epitope) Prediction

We developed a new graph-based method and benchmarked it with the baseline for epitope prediction:

Table: Performance comparison of our approach and baseline epitope prediction methods on a held-out test set. Best values in each column are in bold.

Method	MCC	Prec.	Recall	AUCROC	F1
Ours	**0.263**	**0.281**	**0.457**	**0.650**	**0.348**
WALLE	0.210	0.235	0.258	0.635	0.145
EpiPred	0.029	0.122	0.142	—	0.112
ESMBind	0.016	0.106	0.090	0.506	0.064
MaSIF-site	0.037	0.125	0.114	—	0.128

Docking

Before performing computational docking, we develop a transformation function to bring the antigen and antibody in close proximity by minimizing the stochasticity in the docking module:

Synthetic Distance Matrix Construction

To pre-position the antibody paratope and antigen epitope to avoid local minima docking positions, we construct a synthetic distance matrix:

$$D = \begin{bmatrix} D_{\text{ab}} & d \cdot \mathbf{1} \\ d \cdot \mathbf{1}^{\top} & D_{\text{ag}} \end{bmatrix}$$

- D_{ab}: pairwise Cα distances within the paratope
- D_{ag}: pairwise Cα distances within the epitope
- d: user-defined inter-group distance (e.g., 30 Å)
- $\mathbf{1}$: matrix of ones, setting all inter-group distances to d

Multidimensional Scaling (MDS)

We apply classical **Multidimensional Scaling (MDS)** to embed this distance matrix into 3D space:

$$X = \text{MDS}(D)$$

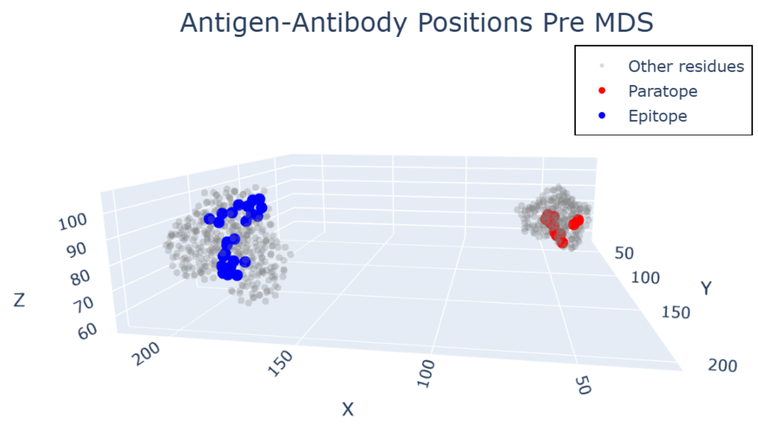
This generates a 3D target configuration for both epitope and paratope.

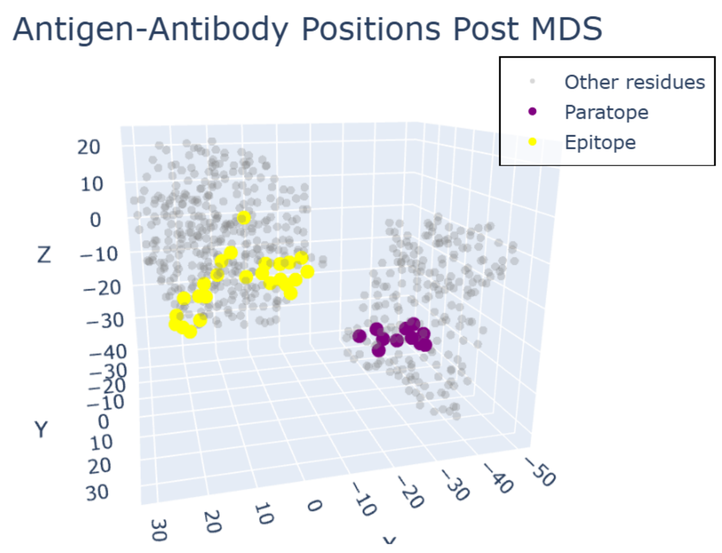
Rigid-Body Superposition via Kabsch Algorithm

To align the original paratope and epitope coordinates to their MDS-embedded targets, we compute the optimal rigid-body transformation using the **Kabsch algorithm**:

$$\min_{R,t} \sum_i \|R\mathbf{p}_i + t - \mathbf{q}_i\|^2$$

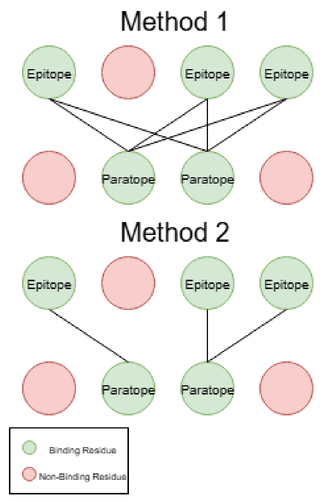
- $\mathbf{p}_i$: original Cα coordinates
- $\mathbf{q}_i$: MDS-embedded coordinates
- $R \in SO(3)$: optimal rotation matrix
- $t \in \mathbb{R}^3$: optimal translation vector





Due to the uncertainty of whether we will have individual residue level epitope to paratope binding information, we propose 2 separate methods to guide docking using Rossetta:

- Method 1: Involves all epitope values being mapped to all paratope values, and each pair carrying a increase in total docking score if the Cα of the 2 is within 4.5A.
- Method 2: Will incorporate a increase in docking score if the Cα are within 4.5 angstroms, but the residue pairs will be only the exact binding combination from previous information.



Method	Epi P	Epi R	Para P	Para R
Method 1	0.41	0.42	0.45	0.67
Method 2	0.40	0.49	0.42	0.71

Binding Affinity Prediction

We employ CSM-Ab, which is a graph-based deep learning model to predict the binding free energy of a docked complex.

The binding free energy (ΔG) is directly related to binding affinity K_d via the equation:

$$\Delta G = -RT \ln(K_d)$$

where R is the gas constant and T is the absolute temperature. This relationship allows predicted free energy values to be interpreted in terms of binding strength.

Binding free energy of the best docked native complexes ranges anywhere from -5kcal/mol to -15 kcal/mol

Size Range	Estimated G(kcal/mol)
50–100 AA	11.42
200–300 AA	10.58
800–1200 AA	12.31

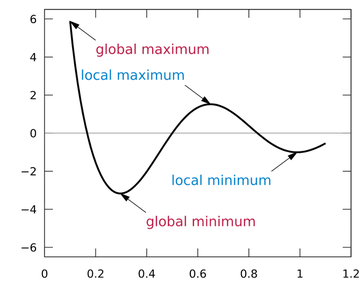
Table: PSI-BLAST runtimes across different sequence lengths (using single iteration).

CDR Region Prediction

CDR boundaries are defined based on a fixed antibody numbering scheme (e.g., Kabat, Chothia, or IMGT).

Free Energy Estimation & Minimization

- We employ a molecular dynamics mechanics tool (OpenMM) using the Amber relax from AlphaFold2 to estimate the free energy ΔG of the antibody
- If $\Delta G > \epsilon_2$, we perform free energy perturbation in the antibody by running Monte-Carlo simulations to bring the antibody to its lowest energy state



Indels (Substitutions, Deletions, Insertions) in Antibody Sequence

- When the antibody has an acceptable free energy $\Delta G < \epsilon_1$ and low binding affinity $K_d < \epsilon_2$, we perform sequence mutations to optimize the antibody to be a suitable binder for the given antigen
- We employ IPRO +/- (iterative protein redesign and optimization procedure) to perform amino acid substitutions, insertions, and deletions in the sequence
- The resultant antibody with $\Delta G < \epsilon_1$ and $K_d < \epsilon_2$ is a potential binder antibody for the target antigen
- This process is repeated for all k candidate antibodies and produces the refined antibodies as output

References

[1] Stephen F. Altschul et al. "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs". In: *Nucleic Acids Research* 25.17 (Sept. 1997), pp. 3389–3402. ISSN: 0305-1048. DOI: 10.1093/nar/25.17.3389. eprint: https://academic.oup.com/nar/article-pdf/25/17/3389/3639509/25-17-3389.pdf. URL: https://doi.org/10.1093/nar/25.17.3389.

[2] Sidhartha Chaudhury, Sergey Lyskov, and Jeffrey J. Gray. "PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta". In: *Bioinformatics* 26.5 (Jan. 2010), pp. 689–691. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btq007. eprint: https://academic.oup.com/bioinformatics/article-pdf/26/5/689/48860623/bioinformatics_26_5_689.pdf. URL: https://doi.org/10.1093/bioinformatics/btq007.

[3] Lewis Chinery et al. "Paragraph—antibody paratope prediction using graph neural networks with minimal feature vectors". In: *Bioinformatics* 39.1 (Nov. 2022), btac732. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btac732. eprint: https://academic.oup.com/bioinformatics/article-pdf/39/1/btac732/48448850/btac732.pdf. URL: https://doi.org/10.1093/bioinformatics/btac732.

[4] Mahmood Kalemati et al. "ParaAntiProt provides paratope prediction using antibody and protein language models". In: *Scientific Reports* 14 (Nov. 2024). DOI: 10.1038/s41598-024-80940-y.

[5] Chunan Liu et al. *AsEP: Benchmarking Deep Learning Methods for Antibody-specific Epitope Prediction*. 2024. arXiv: 2407.18184 [cs.LG]. URL: https://arxiv.org/abs/2407.18184.

[6] Yoochan Myung, Douglas E V Pires, and David B Ascher. "CSM-AB: graph-based antibody–antigen binding affinity prediction and docking scoring function". In: *Bioinformatics* 38.4 (Nov. 2021), pp. 1141–1143. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btab762. eprint: https://academic.oup.com/bioinformatics/article-pdf/38/4/1141/49007587/btab762.pdf. URL: https://doi.org/10.1093/bioinformatics/btab762.

[7] Angelos-Michael Papadopoulos et al. "ParaSurf: a surface-based deep learning approach for paratope–antigen interaction prediction". In: *Bioinformatics* 41.2 (Feb. 2025), btaf062. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btaf062. eprint: https://academic.oup.com/bioinformatics/article-pdf/41/2/btaf062/61813119/btaf062.pdf. URL: https://doi.org/10.1093/bioinformatics/btaf062.