2019 Spring @ https://qdata.github.io/deep2Read/

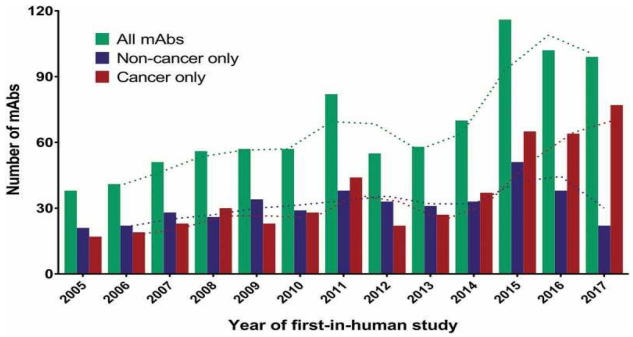
# Attentive cross-modal paratope prediction

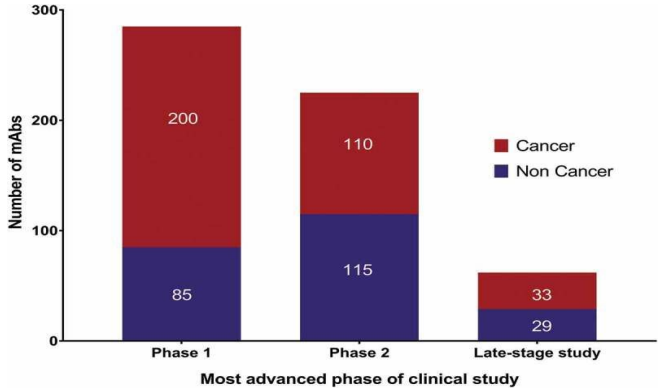
Andreea Deac, Petar Velickovic, Pietro Sormanni

Presented By Eli Draizen

# Background

- Antibodies are easier to create than small molecules
  - Created via phage display, ‘evolution in the lab’
  - More variable
- Antibodies as drugs are becoming more popular
- The challenge becomes creating binding site sequences in Antibody

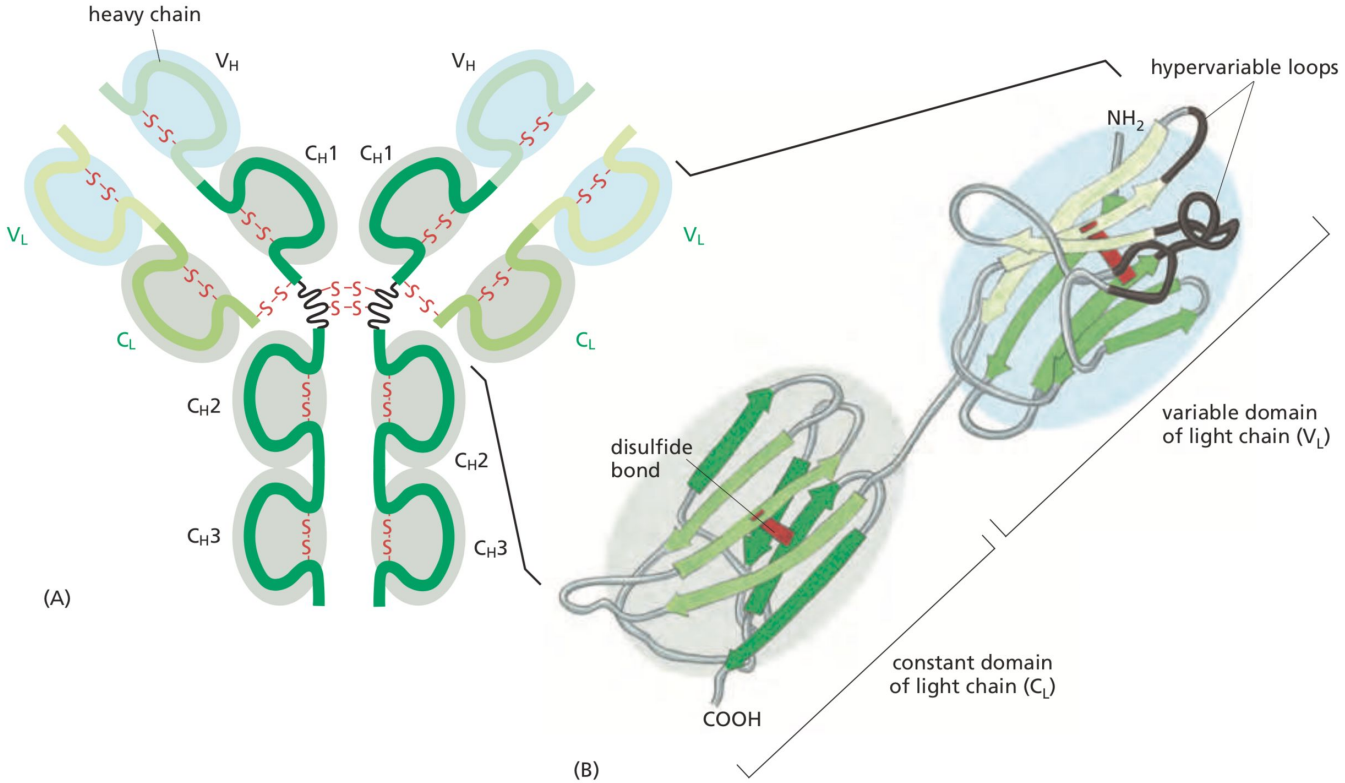


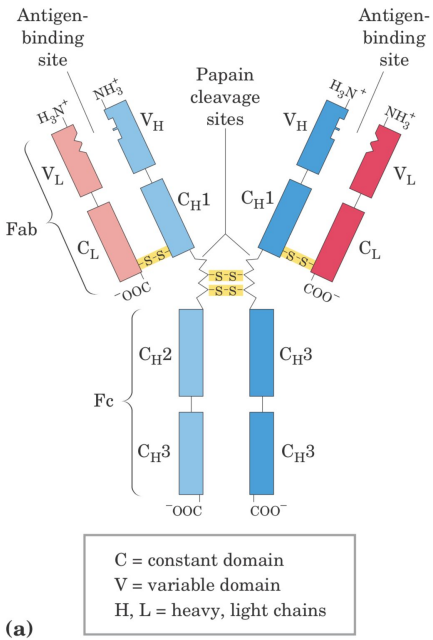




As of November 2018

Kaplon and Reichert. mAbs. 2019 Feb-Mar; 11(2): 219–238.

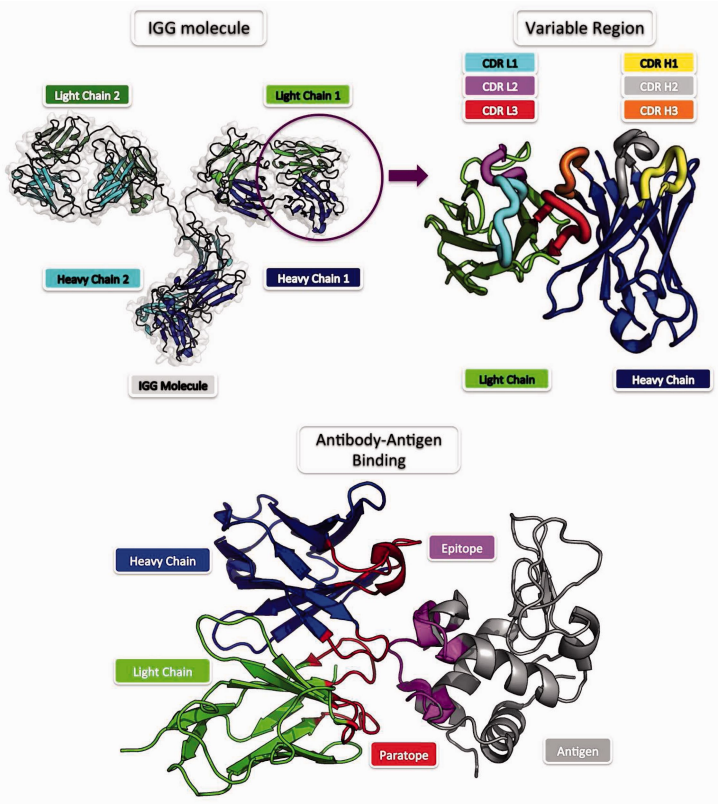
# Antibody structure




The Cell, Alberts; Biochemistry, Lehninger

# Definitions

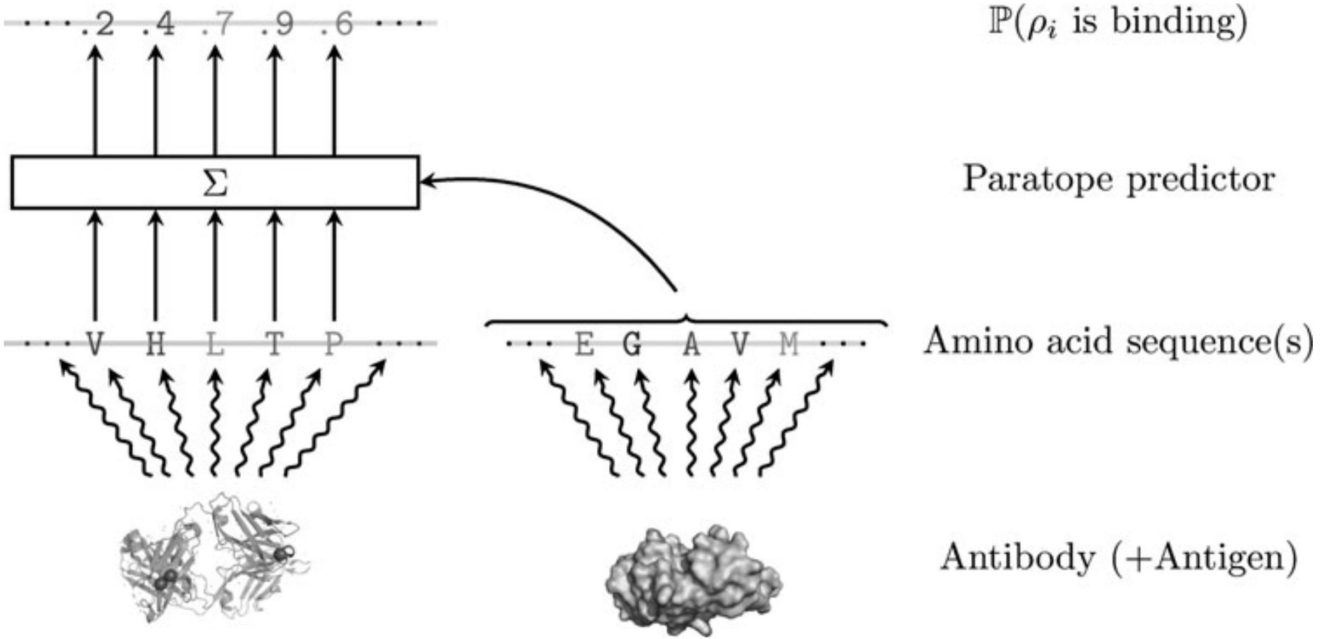



- **Paratope:** Binding site on antibody
  - On average 40–50 residues, while only around 18–19 residues are in contact with antigen
- **Epitope:** Binding site on antigen
- **Antigen:** Target protein (E.g. from parasite)

Esmaielbeiki, Krawczyk, *et al*. *Brief Bioinform*. 2015

# Problem Statement

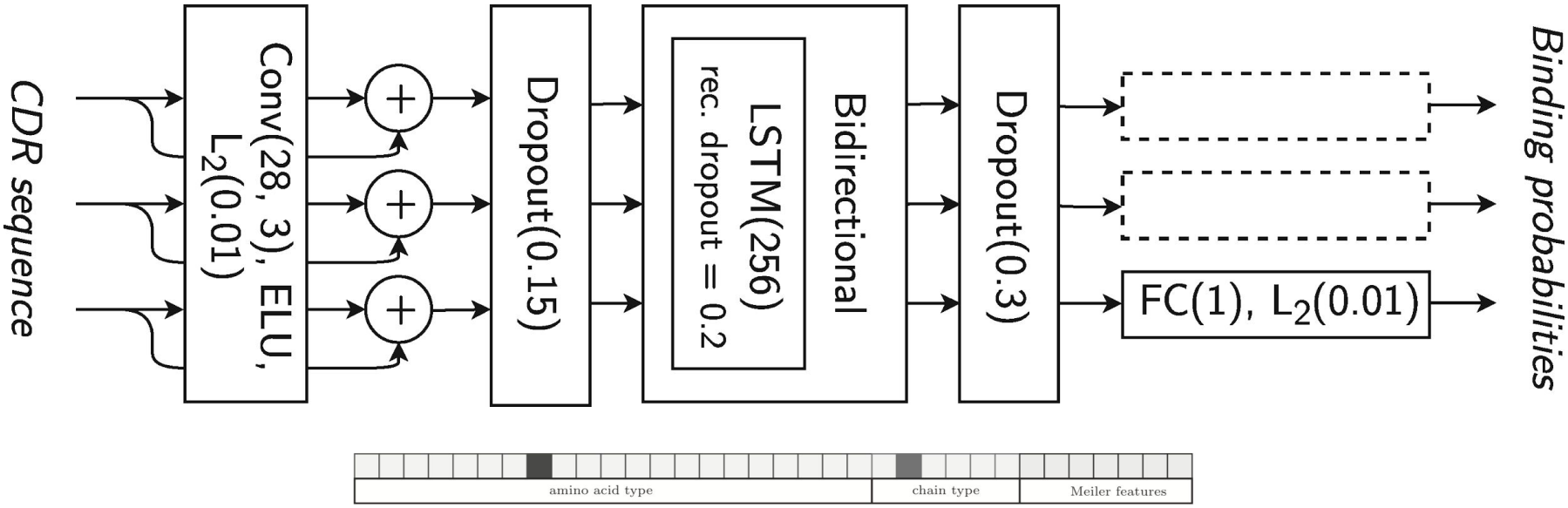
**Goal: Given a CDR and antigen sequence, predict if it will bind an antigen**



# Previous Approaches

- Template-based: sequence or structural homology from alignments of to one side of interface with Ab-Ab structures
- Residue propensities of known Ab-Ag complexes
  - Probability distributions based on a Kolmogorov-Smirnov test (Ab-i-Patch)

# Parapred


CDR sequence
Conv(28, 3), ELU, L2(0.01)
Dropout(0.15)
Bidirectional
LSTM(256)
rec. dropout = 0.2
Dropout(0.3)
FC(1), L2(0.01)
Binding probabilities
amino acid type
chain type
Meiler features
1-hot residue features

# Fast-ParaPred

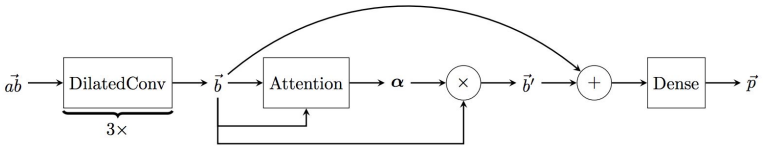
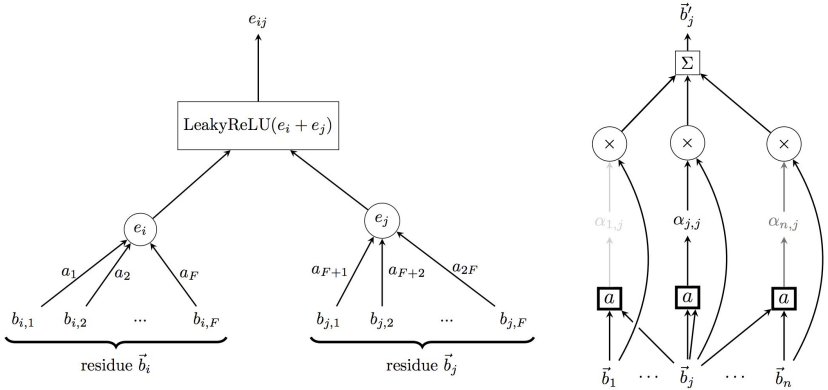



*Figure 1.* The Fast-Parapred architecture.




- Substitute RNN with:
  - Dilated Convolutions
    - 3 layers 64,128,256 feats with rates 1,2,4
  - Self-attention layer
- **ab** is AntiBody CDR sequence
- **p** is classification of binding site or not

# AG-Fast-Parapred

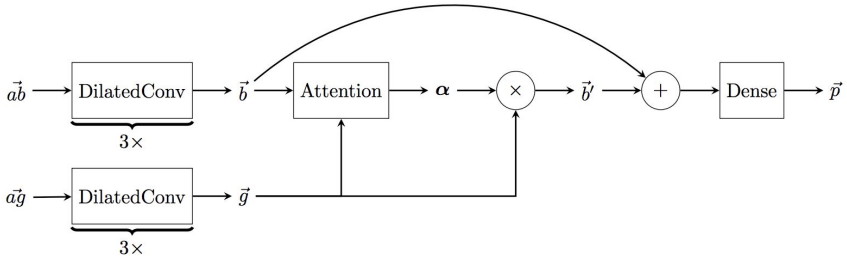
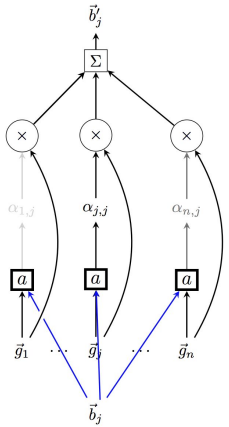



*Figure 4.* The AG-Fast-Parapred architecture.




- Include <u>all</u> residues from **A**nti**G**en
- Same Dilated Convolutions as with **ab**
- Update attention weights where **ab** is the query and **ag** as key/value:

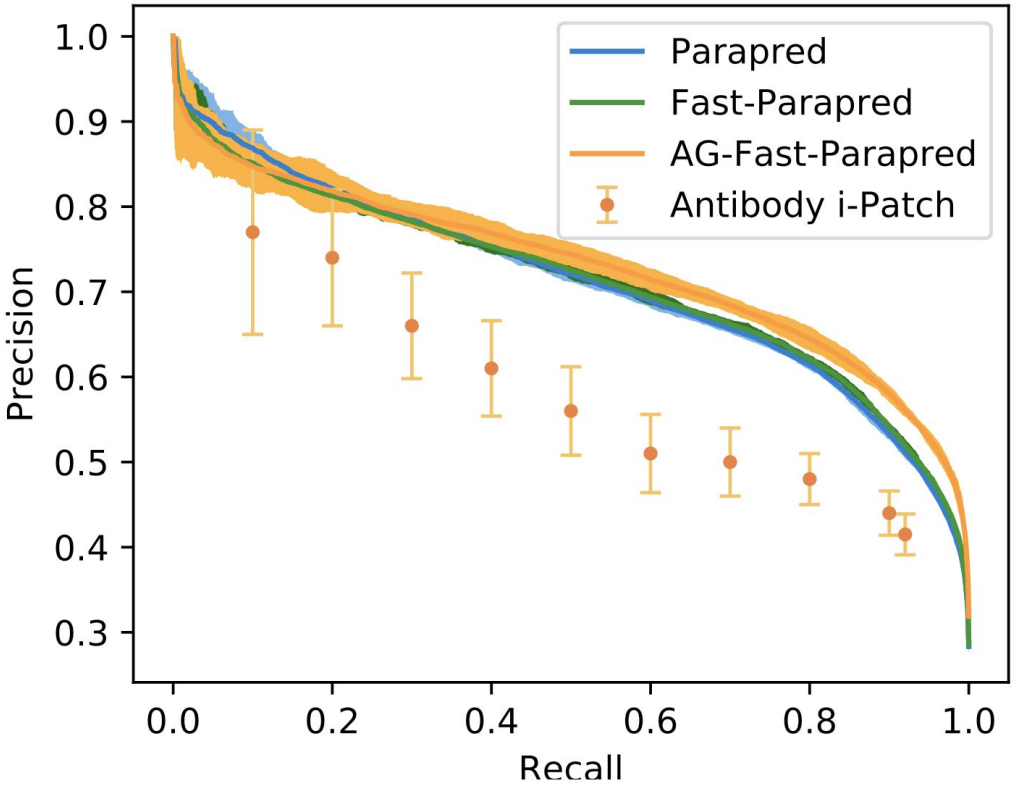
$$\alpha_{ij} = \frac{\exp\left(\text{LeakyReLU}\left(\vec{\mathbf{a}}^T[\mathbf{W}_1\vec{b}_i \| \mathbf{W}_2\vec{g}_j]\right)\right)}{\sum_{k\in\nu_i}\exp\left(\text{LeakyReLU}\left(\vec{\mathbf{a}}^T[\mathbf{W}_1\vec{b}_i \| \mathbf{W}_2\vec{g}_k]\right)\right)} \quad (5)$$

# Experiments

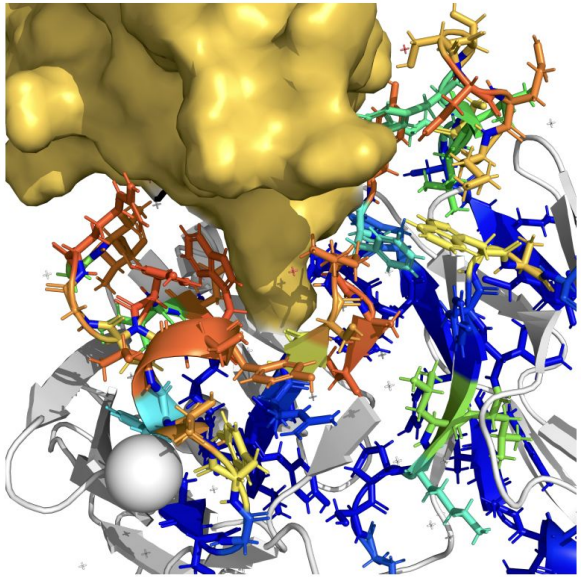
- Data collected from Structural Antibody Database (SAbDB)
  - All: 3522 pdb structures
  - Filtered: ~304 structures
- Results after 10 rounds of 10-fold Cross Validation:

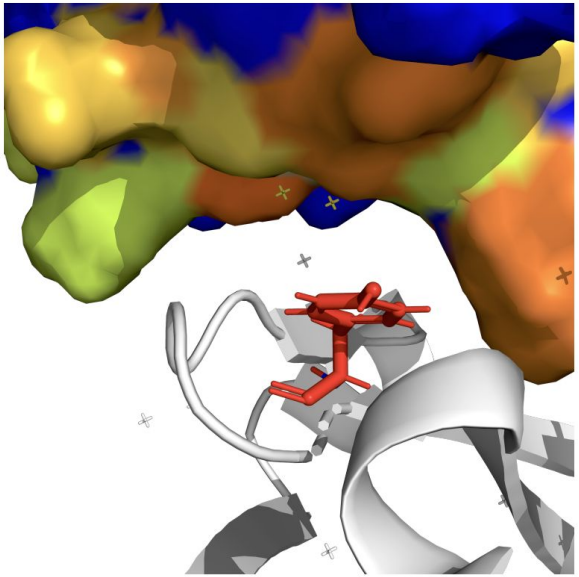
| Method | ROC AUC | MCC | Epoch time |
|---|---|---|---|
| **proABC** (Olimpieri et al., 2013) | 0.851 | 0.522 | — |
| **Parapred** (Liberis et al., 2018) | $0.880 \pm 0.002$ | $0.564 \pm 0.007$ | $0.190 \pm 0.019$s |
| **Fast-Parapred** (ours) | $0.883 \pm 0.001$ | $0.572 \pm 0.004$ | $\mathbf{0.085} \pm 0.015$s |
| **AG-Fast-Parapred** (ours) | $\mathbf{0.899} \pm 0.004$ | $\mathbf{0.598} \pm 0.012$ | $0.178 \pm 0.020$s |

# Results


Parapred
Fast-Parapred
AG-Fast-Parapred
Antibody i-Patch
Precision
Recall
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.0
0.2
0.4
0.6
0.8
1.0

# Results





Left: Ab-Ag complex.

Right: Normalized Ag attention weights

Warmer closers indicate higher probabilities and coefficients

# Conclusion

- Fast-Parapred outperforms current state of the art method and does not use any antigen sequence
- AG-Fast-Parapred outperfoms Fast-Parapred since it uses knowlege of antigen sequences, which maybe at the binding site.
- No 3D coordinates used

**Drawbacks:**

- A CDR sequence is needed, not a generative model
- Trained on a small dataset

# Relation to our project

1. Predict binding sites on target protein and potential superfamilies it my bind to using GraphNN
2. Only include those residues in AG-Fast-ParaPred

   ***or***

3. Find partner bind site with complimentary shape and charge
   a. Search through examples of target superfamily with interacting superfamily
4. Given structure of antibody without variable regions, design new variable region with similar geometry and charge to the partner binding site or
   a. Denoising Autoencoder/GAN