DeepSurf2.0: A Deep Learning Approach for Predicting Interactions of B Cell Receptors With Antigens

Angelos-Michail Papadopoulos[1], Anastasia Iatrou[2], Apostolos Axenopoulos[1], Andreas Agathangelidis[2,3], Kostas Stamatopoulos[2], Petros Daras[1]

[1]Information Technologies Institute, Centre for Research and Technology Hellas, Thessaloniki, Greece, [2]Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece, [3]School of Biology, University of Athens, Athens, Greece

INTRODUCTION

- **BcR IG**: Unique identifier for B cells, influencing their behavior and interactions with antigens.
- **Significance**: Vital in B cell homeostasis and implicated in conditions like B cell lymphomas.
- **Challenge**: Traditional crystallographic procedures are notoriously labor-intensive.
- **Solution**: Introducing **DeepSurf2.0**, a deep learning tool for predicting BcR-antigen interactions (PPI) and expediting protein-protein docking study.

AIM

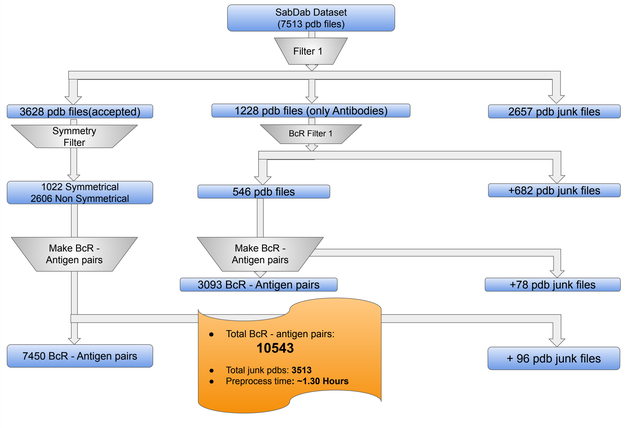
- Introduce and evaluate the efficacy of **DeepSurf2.0**, a deep learning-based computational tool, for predicting BcR-antigen interactions.
- Initiate a foundational advancement in docking algorithms by precisely identifying the receptor's binding region.
- Provide a publicly available BcR-antigen dataset derived from SAbDab[1].

METHOD

1. Utilized SAbDab, a comprehensive database of antibody structures from the Protein Data Bank (PDB).
2. Refined the dataset through specific filtering steps, resulting in **10,534 BcR-antigen pairs**.
3. Trained DeepSurf2.0 using 9,431 BcR-antigen pairs and evaluated its performance on a separate test set of 1,103 pairs.
4. Metrics used for evaluation: **DCA** (Distance between Predicted binding site center and nearest antigen Atom) and **OVR** (Intersection of real and predicted binding sites divided by their union).

RESULTS

1. Refine SAbDab dataset to create BcR-antigen pairs



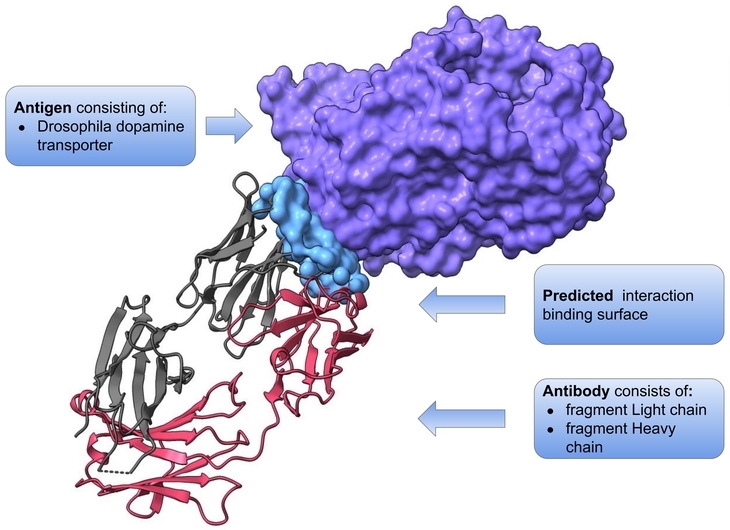
(fig 1) The **SAbDab** database was meticulously refined according to the following criteria:

- Retention of solely complete BcR IGs, specifically those possessing both heavy and light chains (*Filter 1*).
- Preservation of a singular biological assembly from multimeric protein complexes (*Symmetry Filter*).
- Exclusion of BcRs devoid of associated antigens (*BcR Filter 1*).
- Configuration of each BcR-antigen pair to encompass three distinct chains: one heavy and one light for the BcR, and a singular chain for the antigen (*Make BcR - Antigen pairs*).
- Total BcR - Antigen pairs: **10,534**.

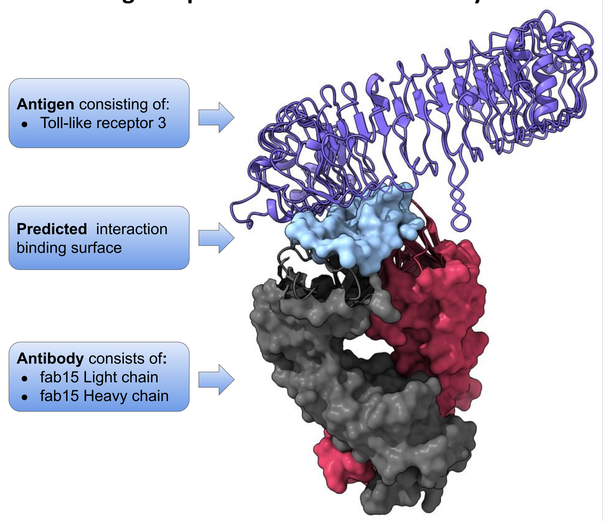
DeepSurf2.0 was tested on a separate unseen test set of 1,103 BcR-antigen pairs:

- **34 % DCA:**
  A "hit" (valid predicted binding site) was detected in 375/1,103 cases.
- **35.2.% OVR:**
  The overlap between the predicted and ground truth binding site: 35.2%.

2. Binding Site prediction on 4xnx antibody



3. Binding Site prediction on 3ulv antibody



The outcomes from **DeepSurf 2.0** cannot be directly equated with the prevailing state-of-the-art methods[2]. given its training and testing on a novel benchmark dataset sourced from SAbDab. Consequently, this presents a unique challenge in which a BcR may bind to multiple antigens, in contrast to the Docking Benchmark DB5.5[3], where there are currently only 103 BcR-antigen pairs.

(fig 2-3) DeepSurf2.0 accurately forecasts the binding sites of two receptors, specifically antibodies derived from the 4xnx and 3ulv entries in the Protein Data Bank (PDB).

CONCLUSIONS

DeepSurf2.0:

- Serves as a foundation for enabling subsequent docking algorithms to target the predicted interaction binding surface rather than the entire protein structure.
- This advancement underscores the transformative potential of deep learning within the realm of (immuno)hematology, holding the potential to provide novel insights into the pathogenesis and progression of B cell-related disorders.
- A new publicly available dataset from SAbDab is provided, posing a new challenge for PPI, particularly concerning BcR-Antigen instances. This dataset uniquely incorporates multiple antigens for each receptor and elevates the order of magnitude of the total pairings by a factor of 2, compared to the preceding benchmark DB5.5.

REFERENCES

1. **Dunbar J., et. al.** *Oxford University Press; 2014*
2. **Sharon S., et. al.** *The Protein Journal; 2023*
3. **Vreven T., et al.** *Journal of molecular biology; 2015*

CONTACT INFORMATION

Angelos-Michail Papadopoulos - angepapa@iti.gr
Petros Daras - daras@iti.gr
Kostas Stamatopoulos - kostas.stamatopoulos@certh.gr