

## The 66th ASH Annual Meeting Abstracts

## POSTER ABSTRACTS

## 803. EMERGING TOOLS, TECHNIQUES, AND ARTIFICIAL INTELLIGENCE IN HEMATOLOGY

**DeepSurf2-FF: Introducing Force Fields in Deep Learning for Improving *in-Silico* Prediction of Bcr-Antigen Interactions in B Cell Lymphomas**

Angelos-Michael Papadopoulos<sup>1,2</sup>, Anastasia Iatrou<sup>3</sup>, Apostolos Axenopoulos<sup>4,1</sup>, Paolo Ghia<sup>5,6</sup>, Kostas Stamatopoulos<sup>3</sup>, Federico Alvarez, Prof, PhD<sup>2</sup>, Petros Daras<sup>1</sup>

<sup>1</sup>Information Technologies Institute, Centre for Research and Technology Hellas, Thessaloniki, Greece

<sup>2</sup>Universidad Politécnica de Madrid, Madrid, Spain

<sup>3</sup>Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece

<sup>4</sup>Aristotle University of Thessaloniki, Thessaloniki, Greece

<sup>5</sup>Medical School, Università Vita-Salute San Raffaele, Milano, Italy

<sup>6</sup>Division of Experimental Oncology, IRCCS Ospedale San Raffaele, Milano, Italy

The B cell receptor immunoglobulin (BcR IG) serves as a distinctive molecular marker for each B cell clone, facilitating interactions with antigens that subsequently impact clonal behavior. BcR signaling is crucial for every aspect of B cell physiology, while it plays a significant role in pathological conditions involving B cells, such as B cell malignancies and autoimmune disorders. Hence, understanding the structure of BcR IG complexed with its cognate antigenic epitopes is vital for unraveling the mechanisms underlying BcR-antigen interactions with far-reaching implications extending to the development of targeted immunotherapeutic approaches. Although studying actual protein crystals would be ideal, the crystallographic processes are notoriously lab-intensive and demanding. In order to overcome this limitation, we herein propose an innovative *in-silico* approach, employing 3D analysis of BcR-antigen interactions. The proposed DeepSurf2-FF extends the work of DeepSurf2.0 (Papadopoulos et al., 2023) by introducing the following innovative features: i) our deep learning model has been modified adopting a new state-of-the-art architecture; ii) a new set of non-geometric properties based on the force fields has been added to our model. The above extensions resulted in significant improvements in the prediction accuracy of our method. Our method was trained and tested on the same well-established paratope prediction benchmark used by PECAN (Pittala et al., 2020), which includes 460 BcR-antigen complexes. The dataset was divided into three sets: the training set (205 complexes), the validation set (103 complexes), and the test set (152 complexes) filtered to ensure that antibodies shared no more than 95% pairwise sequence identity. The dataset included only complexes with paired IG heavy and light chains, having a resolution below 3 Å, and protein antigens. Following previous methods, residues were labeled as binding if any heavy atom was within 4.5 Å of an antigen-heavy atom. Due to the nature of the paratope prediction task, binding site atoms are significantly fewer than non-binding atoms, resulting in a class imbalance. Our method was compared to the previous state-of-the-art Paragraph model (Chinery et al., 2023). The latter had set a high benchmark in paratope prediction, achieving remarkable results on this task. To further enhance its performance, the Paragraph model was also pre-trained on an expanded dataset of 1060 complexes from SabDab. In contrast, DeepSurf2-FF did not require pre-training on such an extensive dataset, since it achieved state-of-the-art results across all metrics, only by training on the official paratope prediction benchmark. DeepSurf2-FF and Paragraph were compared based on four key metrics; AUC-PR which evaluates the performance of imbalance datasets by focusing on the minority class (binding site atoms), providing insights into precision and recall trade-offs, AUC-ROC which measures the model's ability to discriminate between binding and non-binding atoms, F-score which balances the importance of false positives and false negatives and MCC which is a performance metric used to evaluate the quality of binary classifications. Compared to the Paragraph model, DeepSurf2-FF demonstrated significant improvements: a 17.82% increase in PR-AUC (0.820 vs. 0.696), a 4.6% increase in ROC-AUC (0.977 vs. 0.934), a 4.09% increase in F-score (0.713 vs. 0.685), and a 5.66% increase in MCC (0.691 vs. 0.654). Overall, through rigorous training and testing on a well-established paratope prediction benchmark, DeepSurf2-FF outperformed the previous state-of-the-art Paragraph model, demonstrating significant advancements across all evaluation metrics. This significant development highlights the potential of DeepSurf2-FF in enhancing our understanding of BcR-antigen interactions and facilitating future research in the field of (immuno)hematology, offering novel and profound insights into the natural history and management of B cell-related pathologies.

**Disclosures Ghia:** AstraZeneca: Consultancy, Research Funding; Beigene: Consultancy; Abbvie: Consultancy, Research Funding; BMS: Consultancy, Research Funding; Johnson & Johnson: Consultancy, Research Funding; Galapagos: Consultancy; Loxo@Lilly: Consultancy; MSD: Consultancy; Roche: Consultancy. **Stamatopoulos:** Bristol Myers Squibb: Honoraria; Roche: Research Funding; Novartis: Research Funding; Janssen: Honoraria, Research Funding; AstraZeneca: Honoraria, Research Funding; AbbVie: Honoraria, Research Funding; Lilly: Honoraria.

<https://doi.org/10.1182/blood-2024-204732>