Supplementary Materials

Title

Benchmarking Sequence-Based and AlphaFold-Based Methods for pMHC-II Binding Core Prediction: Distinct Strengths and Consensus Approaches

Authors

Soobon Ko[1], Honglan Li[2], Hongeun Kim[2,3], Woong-Hee Shin[3], Junsu Ko[2], Yoonjoo Choi[1*]

[1] Combinatorial Tumor Immunotherapy MRC, Chonnam National University Medical School, Jeollanam-do, 58128, Republic of Korea

[2] Arontier Inc., Seoul, 06735, Republic of Korea

[3] Department of Biomedical Informatics, Korea University College of Medicine, Seoul, 02841, Republic of Korea

Correspondence

Yoonjoo Choi

Department of Microbiology and Immunology, Chonnam National University Medical School, Jeollanam-do, 58128, Republic of Korea

The National Immunotherapy Innovation Center

Tel: +82-62-379-8477

Email: kalicuta@gmail.com

Figure S1. Precision–Recall (PR) and ROC curves for sequence-based prediction models. Precision–Recall (left) and ROC (right) of NetMHCIIpan-4.3 (red) and DeepMHCII (orange) on the benchmark dataset. For NetMHCIIpan-4.3, the %Rank was used as the continuous binding score across all alleles, reflecting the predicted binding likelihood output by the model. For DeepMHCII, normalized affinity scores were used, with the default binder threshold (IC_{50} < 500 nM) corresponding to a score of approximately 0.426. Dashed horizontal lines indicate the random baseline corresponding to positive-class prevalence.

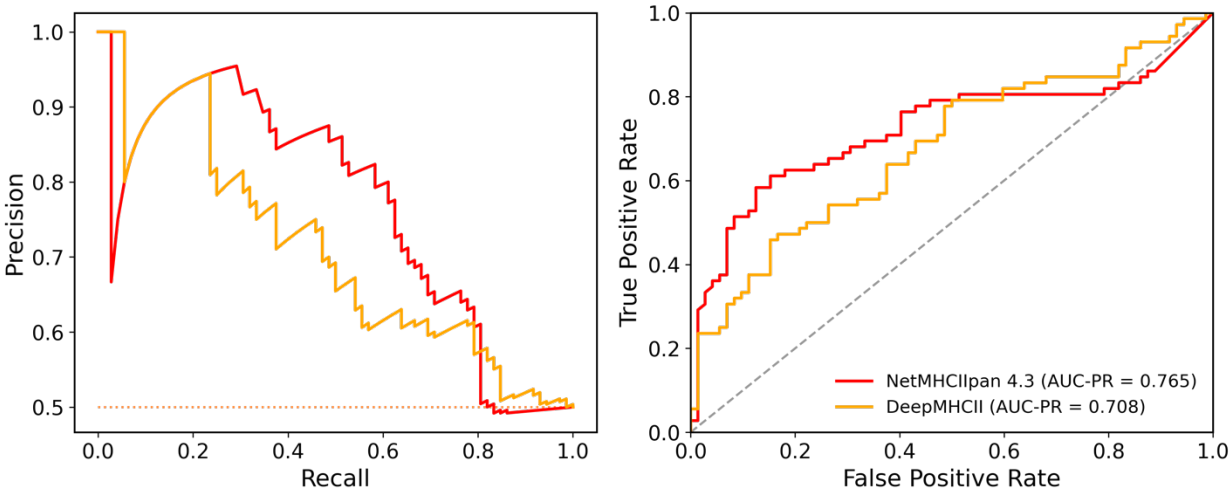

Figure S2. Representative mispredicted pMHC-II complexes illustrating structural causes of core misclassification across all prediction methods. Each example corresponds to an experimentally verified binder misclassified by both sequence-based and structure-based approaches. Red segments denote the true binding core (experimental reference), and slate regions indicate flanking residues predicted within the groove. **(A)** 6DFS, chimeric. The peptide length is moderate (15 residues in length), but the predicted core is shifted relative to the true register. **(B)** 3T0E, human. An extended Gly-rich region (GGG) forms a flexible loop protruding beyond the canonical 9mer core. **(C)** 3O6F, human. The peptide contains AEG-induced helix formation and GSGG loops, disrupting proper core anchoring. **(D)** 6DFW, chimeric. Both EGA-mediated helical turns and GGG loops lead to partial peptide insertion and core misalignment.