Prediction of Peptide-MHC II Interaction Affinity

Anton V. Tsukanov * Stepan A. Epifantsev Stepan A. Epifantsev

Anna A. Zaikina Kseniia D. Shalygina

Abstract

We present two novel deep learning architectures for predicting peptide-MHC class II binding affinities using ESM-2 protein language model embeddings. The first model employs a cross-attention mechanism to capture dynamic interactions between peptide and MHC residues, while the second utilizes graph attention networks (GAT) to model structural relationships. Through comprehensive interpretability analysis, we demonstrate that both architectures learn biologically meaningful binding patterns, including appropriate attention to polymorphic β-chain residues and core peptide anchors. Notably, our investigation reveals that incorporating full protein context during embedding generation provides more informative representations than using isolated peptide sequences alone. While the models achieve competitive performance (0.600 AUC for cross-attention), this work primarily contributes: (1) an interpretable framework for analyzing peptide-MHC interactions, (2) evidence for the importance of structural context in binding prediction, and (3) open-source implementations of attention-based and graph-based approaches. These findings provide foundation for developing more sophisticated structure-aware predictors in immunoinformatics.

1 Introduction

Predicting the binding affinity of peptides to MHC II is critical for the development of novel therapeutic approaches, including neoantigenic vaccines in oncology Moore and Nishimura [2020], Waldman et al. [2020], Lin et al. [2023] and tolerance induction strategies in autoimmune diseases. In vaccine development, these tools help select viral or bacterial peptides likely to trigger protective T-cell responses, accelerating new vaccine design You et al. [2022]. In cancer immunotherapy, models identify tumor-specific mutant peptides (neoantigens) that bind a patient's MHC alleles, guiding personalized vaccine and T-cell therapy trials Ott et al. [2017]. For autoimmune disease research, binding predictions reveal self-peptides that may be presented by risk-associated MHC variants, shedding light on triggers of self-reactive T-cell attacks Jensen et al. [2018]. In transplantation, in silico screens estimate donor peptide fragments that could activate recipient T-cells, improving donor–recipient matching and reducing rejection risk McKeever et al. [2021]. Drug developers also use these predictions to remove unwanted T-cell epitopes from therapeutic proteins, lowering immunogenicity Su et al. [2024]. Finally, advanced models such as residue–residue pair encodings offer broad, allele-agnostic accuracy, supporting both practical applications and fundamental mapping of the diverse human immunopeptidome Wang et al. [2024].

MHC II molecules are heterodimers of α- and β-chains forming a peptide-binding groove in which antigenic peptides are anchored through conserved hydrogen bonds and interactions with deep pockets (P1, P4, P6, P9). In contrast to MHC I, binding peptides are typically longer than 13 amino acids with a cortical region of about 9 residues, with both anchoring residues in the pockets Rammensee

*Additional information

Student research report. Not peer-reviewed.

[1995] and flanking sequences that enhance binding through nonspecific interactions Rademaker et al. [2025], Conant and Swanborg [2003] playing key roles in affinity.

The difficulty in accurately predicting affinity is due to several factors: the high polymorphism of the amino acid composition of the pockets between different MHC II alleles Rammensee [1995], the conformational flexibility of the molecule itself, which affects binding kinetics Sadegh-Nasseri et al. [2010], as well as the absence of tight constraints on peptide length and the significant contribution of flanking sequences to the stability of the complex Knapp et al. [2009], Ferrante [2013]. These features create significant computational difficulties in modeling peptide-MHC II interactions compared to more strictly deterministic MHC I systems.

2 Related works

To date, a wide range of computational methods have been developed to predict the binding affinity of peptides to MHC class II molecules using a variety of approaches. Traditional methods that are not based on deep learning (e.g., Racle and Gfeller [2024]) analyze solely the amino acid sequence of the peptide, identifying characteristic motifs associated with binding for specific MHC II alleles. These approaches rely on identifying conserved patterns in known ligands and assessing their presence in the peptides under investigation. Newer methods using deep learning (You et al. [2022], Wang et al. [2024]) demonstrate significantly higher prediction accuracy (evaluated by the ROC AUC metric) compared to classical approaches. Their key advantage is that they take into account not only the peptide sequence, but also the features of the MHC II molecule, for which a pseudo-sequence of 34 amino acid residues forming the peptide-binding groove is used (although these residues are arranged nonlinearly in the native protein structure).

However, existing models have considerable potential for improvement. In particular, none of the reviewed approaches utilize state-of-the-art protein language models (such as ESM), which provide state-of-the-art representations of protein sequences. The integration of ESM embeddings could significantly improve the accuracy of predictions because: These embeddings contain rich information about structural and functional properties of proteins. Allow contextualized representations for each amino acid residue Particularly important for MHC II pseudo-sequence analysis, where traditional methods treat amino acids as a simple linear sequence, ignoring their spatial arrangement and local chemical properties.

3 Methods

3.1 The dataset preparation

For model training and validation, we used a pre-processed dataset from the study Jensen et al. [2018], which included a predefined split into training and validation samples. The raw data contained information on MHC class II α- and β-chain alleles, peptide amino acid sequences, and normalized IC50 values (IC50-normal between 0 and 1) Jensen et al. [2018] training quantifying binding affinity. A separate data sample Cheng et al. [2021] was used for independent testing, strictly avoiding overlap of peptides and allele combinations with the training and validation samples. The amino acid sequences of the corresponding MHC II alleles were extracted from the dataset provided in the article Jensen et al. [2018].

3.2 Model architectures

Two modern language models were used ESM-C Hayes et al. [2025] to generate protein embeddings. Models were trained in regression format using the MSELoss loss function to predict continuous IC50-normal values. The quality of the models was evaluated using three key metrics: Pearson correlation coefficient to assess the linear relationship between predicted and experimental IC50-normal values ROC-AUC and accuracy for binary binding classification (IC50-normal threshold = 0.496, where values below the threshold correspond to no binding) Optimization was performed using AdamW with a learning rate of 3×10- and a Decai weight of 1×10-. All experiments were implemented in Python using PyTorch and PyTorch Lightning frameworks. The full code for data processing and model training is available in the open source GitHub repository: https://github.com/ubercomrade/airi. Training was performed over 100 epochs using early stopping to prevent overtraining.

3.2.1 Cross-attention model

The first proposed approach (ProteinPeptideInteractionModel) is designed to predict the binding affinity between peptides and MHC class II (MHCII) molecules, a critical step in understanding immune response mechanisms. The architecture is biologically motivated by the nature of peptide-MHCII interactions, where the binding site of the MHCII protein accommodates peptide fragments, and their binding affinity is determined by both sequence-specific features and structural compatibility. The model processes ESMC embeddings of the MHCII protein (34 residues) and the candidate peptide (21 residues) through separate linear projections, transforming them into a shared latent space. A cross-attention mechanism then models the interactions between the peptide (query) and the MHCII protein (key and value), allowing the model to focus on specific residues in the binding groove that contribute most to the interaction. This is analogous to the way certain anchor residues in the peptide and polymorphic regions in the MHCII molecule dominate binding specificity. Following cross-attention, the model aggregates latent representations by averaging over peptide positions, capturing a global view of the interaction while reducing sensitivity to positional noise. The final prediction combines the attention-refined peptide representation with the original peptide embedding, ensuring that both the contextualized interaction features and intrinsic peptide properties contribute to the binding score. This design reflects the biological principle that peptide-MHCII binding depends not only on direct interactions but also on the peptide's intrinsic propensity to adopt a bound conformation. The use of dropout and ReLU activation introduces robustness, preventing overfitting to spurious correlations in the training data. Together, these components provide a computationally efficient yet biologically plausible framework for predicting peptide-MHCII interactions.

3.2.2 GATConv model

The GNNModel is a graph neural network designed to predict peptide-MHC class II (MHCII) binding affinity by explicitly modeling the structural and relational features of the protein-peptide interaction system. The architecture leverages Graph Attention Networks (GATConv) to capture residue-level interactions, reflecting the biological reality that peptide binding is influenced not only by sequence composition but also by the spatial and topological arrangement of residues in the MHCII binding site. The input to the model is a graph where nodes represent amino acid residues (from both the MHCII protein and the peptide), and edges encode structural or sequence-based relationships. The model first processes these residues through two GATConv layers, each followed by layer normalization and dropout for stable training. The use of multi-head attention (with 2 heads) allows the network to dynamically weigh the importance of neighboring residues, mimicking the way certain amino acids in the binding site (e.g., anchor pockets in MHCII) disproportionately influence peptide binding.

After graph propagation, the model separates the node embeddings into protein and peptide representations using a binary mask. A global mean pooling operation aggregates these embeddings into fixed-size vectors, summarizing the overall structural features of the MHCII molecule and the peptide independently. These representations are then concatenated and passed through a fully connected network to produce the final binding score. This design mirrors the biological principle that peptide-MHCII affinity arises from both the global compatibility of the peptide with the binding site and the local interactions between key residues. The inclusion of dropout and ReLU activations ensures generalization, preventing overfitting to noise in the training data while maintaining interpretability. By combining graph-based feature learning with attention mechanisms, the model effectively captures the complex interplay of structural and biochemical factors governing peptide-MHCII interactions.

4 Results

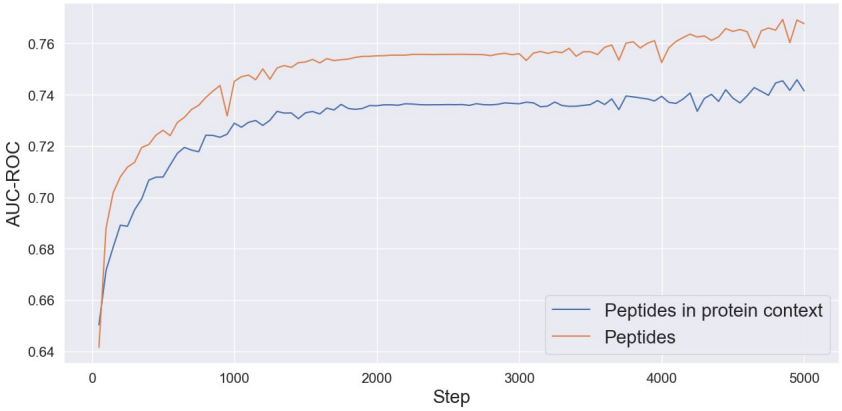

Figure 1: Validation curves during the train for the models based on embeddings received purely from the peptides sequences (Peptides) and from amino acids retrieved from the protein that was processed to obtain peptide (Peptides in protein context)

We trained our models and evaluated their predictive accuracy using the ROC AUC metric, comparing their performance against established models tested on the same benchmark dataset. The results are presented in Table 1. No significant improvement in accuracy was observed. We hypothesize that this outcome may be attributed to the low informational content of peptide embeddings in our initial approach.

To address this limitation, we adopted an alternative testing and inference strategy: instead of using isolated peptide embeddings, we generated context-aware epitope embeddings by incorporating the full protein environment. To achieve this goal we produced embeddings for the entire protein sequence and extracted the relevant epitope regions from these embeddings.

We then compared our original model (trained on context-independent peptide embeddings) against a variant trained on context-aware embeddings. The results, illustrated in Figure 7, demonstrate the impact of structural and sequential context on epitope representation.

Table 1: Model metrics on the IC50 test dataset

Model	AUC ROC
PPIModel	0.600
GNNModel	0.550
NetMHCIIpan-3.2	0.678
PUFFIN	0.688
DeepMHCII	0.693
NetMHCIIpan-4.0_BA	0.681
RPEMHC	0.707

To bridge our computational predictions with biological insights, we performed model interpretability analyses using Principal Component Analysis (PCA) on model embeddings and attention map visualization from the multi-head attention blocks.

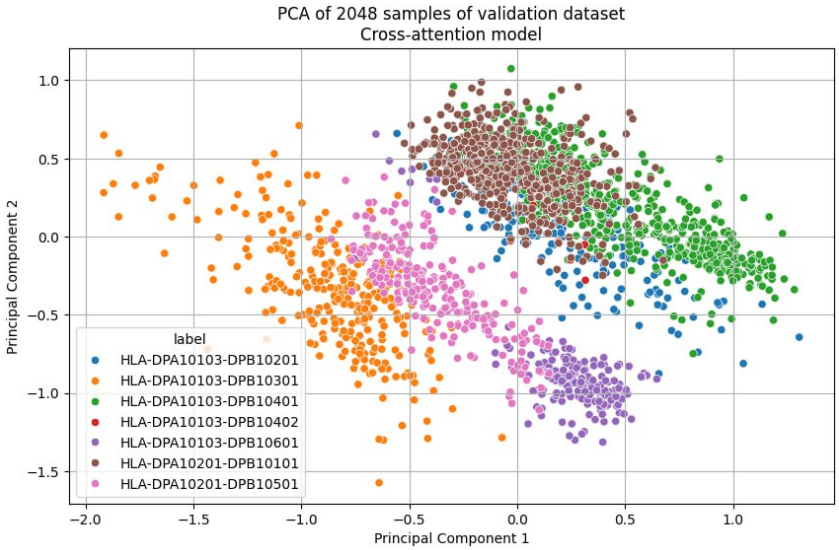

Figure 2: Principal component analysis (PCA) of the first two principal components for the cross-attention model embeddings, colored by HLA allele

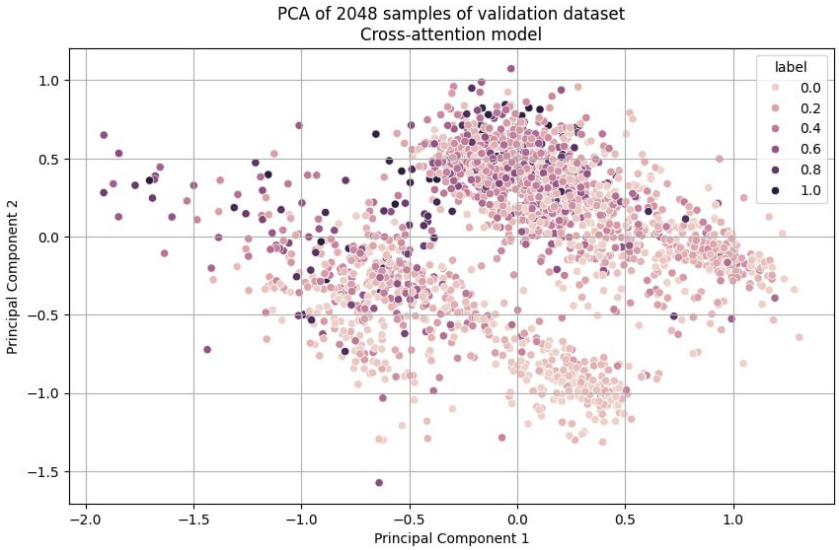

Figure 3: PCA of the first two principal components for the cross-attention model embeddings, colored by binding score (IC50)

First, we examined the latent representations by projecting embeddings onto their two principal components. The PCA of cross-attention model embeddings revealed a strong separation by MHC type, with no apparent correlation to IC50 binding scores in Figures 1 and 2. This suggests that the model's representations are dominated by MHC-specific features rather than peptide-specific ones, likely due to the greater structural diversity across MHC alleles compared to the peptide sequences and the longer context of protein. To address this imbalance, we introduced a gated cross-attention mechanism, which reweights the contributions of protein and peptide embeddings by the extra fully-connected layer. After this adjustment, the resulting embeddings exhibited a smooth gradient of IC50 values without distinct clustering by HLA type, indicating that the model successfully integrated peptide information into its binding affinity predictions.

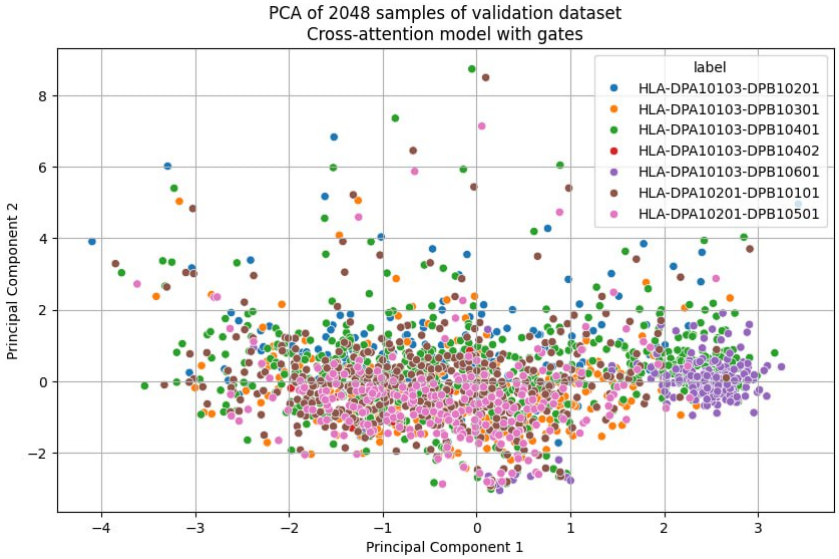

Figure 4: PCA of the first two principal components for the gated cross-attention model embeddings, colored by HLA allele

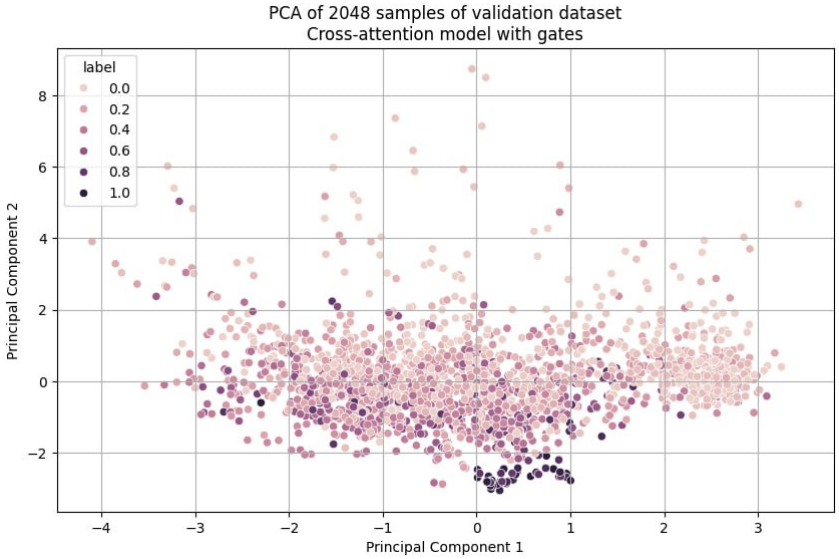

Figure 5: PCA of the first two principal components for the gated cross-attention model embeddings, colored by binding score (IC50)

Next, we analyzed attention maps to understand how the model processes interactions between MHC and peptide residues. The multi-head attention patterns revealed several biologically plausible behaviors in Figure 5. The model consistently allocated greater attention to residues in the beta chain of MHC compared to the alpha chain. This aligns with known immunology, as the beta chain exhibits higher polymorphism across HLA-DR alleles. Attention weights sharply distinguished between padded regions and the central peptide residues, suggesting the technical nuances connected to the peptide sequence padding and model recognition of the binding cleft's structural constraints. The middle residues—likely positioned within the groove—received the highest attention, consistent with their role in stabilizing peptide-MHC interactions.

In the gated cross-attention variant in Figure 6, certain heads (e.g., Head 8) exhibited smoother attention distributions across the peptide, mitigating overemphasis on specific positions. This refinement indicates that the gating mechanism helps the model better balance local and global peptide features.

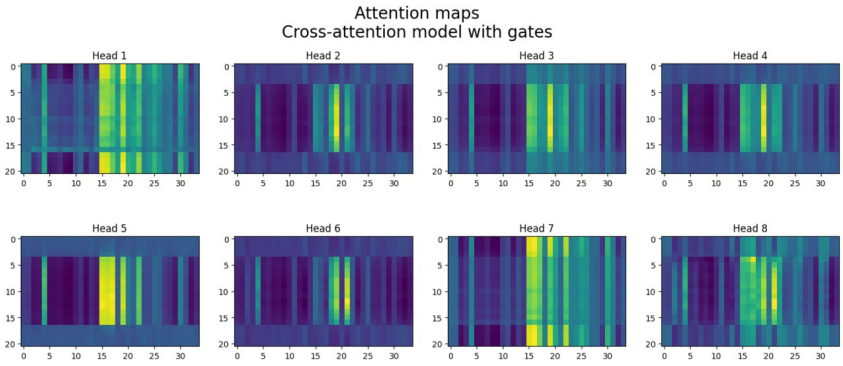

Figure 6: Attention maps across individual heads in the cross-attention model. Brighter regions indicate higher activation weights, reflecting residues with stronger influence on the interaction

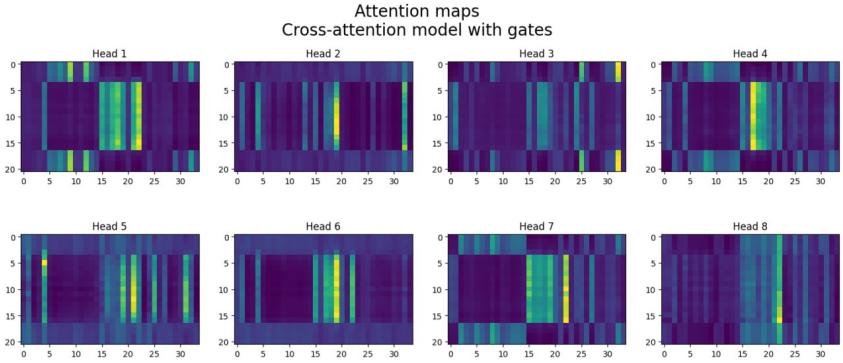

Figure 7: Attention maps across individual heads in the gated cross-attention model. Brighter regions denote higher activation weights, demonstrating the gating mechanism's effect on attention distribution

Together, these interpretability analyses confirm that our models capture biologically meaningful patterns—prioritizing polymorphic MHC regions and recognizing binding cleft constraints.

5 Conclusion and Future Work

This study systematically evaluated deep learning approaches for predicting peptide-MHC class II binding affinity through two novel architectures leveraging ESM embeddings. While our cross-attention and GATConv models achieved competitive but not superior performance compared to existing tools (ROC AUC 0.550-0.600 vs 0.681-0.707 in benchmarks), the work provides significant methodological and biological insights. The interpretability analyses revealed that our models successfully capture fundamental biological principles of MHC II interactions, including the predominant role of β-chain polymorphisms and the importance of central peptide anchors, as evidenced by attention map visualizations and embedding space analyses.

The demonstration that context-aware embeddings (derived from full protein sequences) influence prediction quality suggests several promising research directions. Future work should prioritize three key advancements: First, integrating explicit 3D structural constraints through graph networks or geometric deep learning could better model the peptide-MHC binding groove's conformational

flexibility. Second, developing dynamic context-weighting mechanisms may help balance local peptide features with global protein environment information. Third, multi-task learning frameworks that jointly predict binding affinity, TCR recognition, and peptide processing could capture broader immunological relationships.

From a biological perspective, the models' autonomous discovery of known MHC II binding principles through attention patterns supports using interpretable machine learning as a discovery tool in immunology. This approach could be extended to identify novel binding motifs or allelic interaction patterns in understudied MHC variants. The released open-source implementation provides a foundation for these future studies.

While challenges remain in handling MHC II's peptide length variability and conformational diversity, this work establishes that combining advanced protein language models with biologically inspired architectures can yield interpretable predictors. As the field progresses toward clinical applications like personalized vaccine design and tolerance induction therapies, such transparent models will be crucial for building trust and understanding in computational immunology tools.

References

Jun Cheng, Kaïdre Bendjama, Karola Rittner, and Brandon Malone. Bertmhc: improved mhc–peptide class ii interaction prediction with transformer and multiple instance learning. *Bioinformatics*, 37 (22):4172–4179, 2021.

Stephanie B Conant and Robert H Swanborg. Mhc class ii peptide flanking residues of exogenous antigens influence recognition by autoreactive t cells. *Autoimmunity reviews*, 2(1):8–12, 2003.

Andrea Ferrante. For many but not for all: how the conformational flexibility of the peptide/mhcii complex shapes epitope selection. *Immunologic research*, 56:85–95, 2013.

Thomas Hayes, Roshan Rao, Halil Akin, Nicholas J Sofroniew, Deniz Oktay, Zeming Lin, Robert Verkuil, Vincent Q Tran, Jonathan Deaton, Marius Wiggert, et al. Simulating 500 million years of evolution with a language model. *Science*, page eads0018, 2025.

Kamilla Kjaergaard Jensen, Massimo Andreatta, Paolo Marcatili, Søren Buus, Jason A Greenbaum, Zhen Yan, Alessandro Sette, Bjoern Peters, and Morten Nielsen. Improved methods for predicting peptide binding affinity to mhc class ii molecules. *Immunology*, 154(3):394–406, 2018.

Bernhard Knapp, Ulrich Omasits, Barbara Bohle, Bernard Maillere, Christof Ebner, Wolfgang Schreiner, and Beatrice Jahn-Schmid. 3-layer-based analysis of peptide–mhc interaction: In silico prediction, peptide binding affinity and t cell activation in a relevant allergen-specific model. *Molecular immunology*, 46(8-9):1839–1844, 2009.

Feng Lin, Emma Z Lin, Misa Anekoji, Thomas E Ichim, Joyce Hu, Francesco M Marincola, Lawrence D Jones, Santosh Kesari, and Shashaanka Ashili. Advancing personalized medicine in brain cancer: Exploring the role of mrna vaccines. *Journal of Translational Medicine*, 21(1):830, 2023.

Patricia McKeever, Natasa Stefanovic, Thomas W. H. Kay, Benjamin Lloyd, Simon Sneyd, and Paul G. Holt. Advances in computational prediction of peptide–mhc binding for transplantation compatibility. *Transplantation Reviews*, 35(2):100600, 2021.

Tamson V Moore and Michael I Nishimura. Improved mhc ii epitope prediction—a step towards personalized medicine. *Nature Reviews Clinical Oncology*, 17(2):71–72, 2020.

Patrick A. Ott, Zaheer Hu, Deniz B. Keskin, Sharada A. Shukla, Jiyu Sun, Daniel J. Bozym, Wai Hin Zhang, Angela Luoma, Anita Giobbie-Hurder, Louise Peter, Chihiro Chen, Michael Olive, Terry A. Carter, Shijun Li, David J. Lieb, Tal Eisenhaure, Eriselda Gjini, Janelle P. Stevens, William J. Lane, Ira K. Javeri, Kumar Nellaiappan, Adrian M. Salazar, Heather Daley, Michael Seaman, Michael Johansson, Sally B. Gabriel, Samuel Rodig, Daniel H. Barouch, Cheng Wu, Gad Getz, Kai Wucherpfennig, Harold Vogel, E. Frank Fritsch, and Thomas A. Chan. An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature*, 547(7662):217–221, 2017.

Julien Racle and David Gfeller. How to predict binding specificity and ligands for new mhc-ii alleles with mixmhc2pred. In *HLA Typing: Methods and Protocols*, pages 215–235. Springer, 2024.

Daniel T Rademaker, Farzaneh M Parizi, Marieke van Vreeswijk, Sanna Eerden, Dario F Marzella, and Li C Xue. Predicting reverse-bound peptide conformations in mhc class ii with pandora. *Frontiers in Immunology*, 16:1525576, 2025.

Hans-Georg Rammensee. Chemistry of peptides associated with mhc class i and class ii molecules. *Current opinion in immunology*, 7(1):85–96, 1995.

Scheherazade Sadegh-Nasseri, Sateesh Natarajan, Chih-Ling Chou, Isamu Z Hartman, Kedar Narayan, and AeRyon Kim. Conformational heterogeneity of mhc class ii induced upon binding to different peptides is a key regulator in antigen presentation and epitope selection. *Immunologic research*, 47:56–64, 2010.

Lingtao Su, Yan Yan, Bo Ma, Shiwei Zhao, and Zhenyu Cui. Gihp: Graph convolutional neural network based interpretable pan-specific hla-peptide binding affinity prediction. *Frontiers in Genetics*, 15:1405032, 2024.

Alex D Waldman, Jill M Fritz, and Michael J Lenardo. A guide to cancer immunotherapy: from t cell basic science to clinical practice. *Nature Reviews Immunology*, 20(11):651–668, 2020.

Xuejiao Wang, Tingfang Wu, Yelu Jiang, Taoning Chen, Deng Pan, Zhi Jin, Jingxin Xie, Lijun Quan, and Qiang Lyu. Rpemhc: improved prediction of mhc–peptide binding affinity by a deep learning approach based on residue–residue pair encoding. *Bioinformatics*, 40(1):btad785, 2024.

Ronghui You, Wei Qu, Hiroshi Mamitsuka, and Shanfeng Zhu. Deepmhcii: a novel binding core-aware deep interaction model for accurate mhc-ii peptide binding affinity prediction. *Bioinformatics*, 38(Supplement_1):i220–i228, 2022.