ImmunoStruct: Integration of protein sequence, structure, and biochemical properties for immunogenicity prediction and interpretation

Kevin Bijan Givechian[1,2] **João Felipe Rocha**[3] **Edward Yang**[3]
Chen Liu[3] **Kerrie Greene**[2,4] **Rex Ying**[3] **Etienne Caron**[4]
Akiko Iwasaki[4,5] **Smita Krishnaswamy**[1,3]

[1]Department of Genetics, Yale University
[2]Medical Scientist Training Program, Yale University
[3]Department of Computer Science, Yale University
[4]Department of Immunobiology, Yale University
[5]Howard Hughes Medical Institute, Chevy Chase

Correspondence: akiko.iwasaki@yale.edu, smita.krishnaswamy@yale.edu.

Abstract

Epitope-based vaccines are promising therapeutic modalities for infectious diseases and cancer, but identifying immunogenic epitopes is challenging. The vast majority of prediction methods are sequence-based, and do not incorporate wide-scale structure data and biochemical properties across each peptide-MHC (pMHC) complex. We present ImmunoStruct, a deep-learning model that integrates sequence, structural, and biochemical information to predict multi-allele class-I pMHC immunogenicity. By leveraging a multimodal dataset of ∼27,000 peptide-MHC complexes that we generated with AlphaFold, we demonstrate that ImmunoStruct improves immunogenicity prediction performance and interpretability beyond existing methods, across infectious disease epitopes and cancer neoepitopes. We further show strong alignment with *in vitro* assay results for a set of SARS-CoV-2 epitopes. This work also presents a new architecture that incorporates equivariant graph processing and multi-modal data integration for the long standing task in immunotherapy.

1 Introduction

Epitope-based vaccines have shown promise in the clinic across infectious diseases and cancers such as melanoma, bladder carcinoma, glioblastoma, lung adenocarcinoma, and pancreatic cancer [4, 6, 8]. However, only a subset of patients respond to a limited number of epitopes, and validating epitope immunogenicity remains resource-intensive [3, 4, 6]. While patients with observed CD8+ T-cell-mediated responses can exhibit encouraging clinical outcomes, improved tools are still needed for ranking epitope immunogenicity to enhance candidate selection. For instance, in previous analyses only 2-6% of predicted neoantigen peptides were shown to actually result in true patient derived T-cell immunogenicity [3, 10, 21, 39]. Additionally, recent cancer neoantigen vaccine trials have predominantly used transcript expression and affinity-based methods for filtering and identifying patient-specific epitopes [27]. Therefore, an important challenge remains in developing methods to distinguish immunogenic peptides from those that are not.

Class-I immunogenicity determinants are complex and involve multiple factors, including antigen processing, peptide-MHC (pMHC) binding, pMHC presentation, pMHC-TCR recognition, and CD8+ T-cell activation [24]. The same peptide can also show varying CD8+ T-cell activation based on its paired MHC allele [34]. This is partly because structural differences in the peptide-MHC complex also influence binding cooperativity [37, 40]. Despite the development of numerous

Massachusetts Institute of Technology - Molecular Machine Learning Conference 2024.

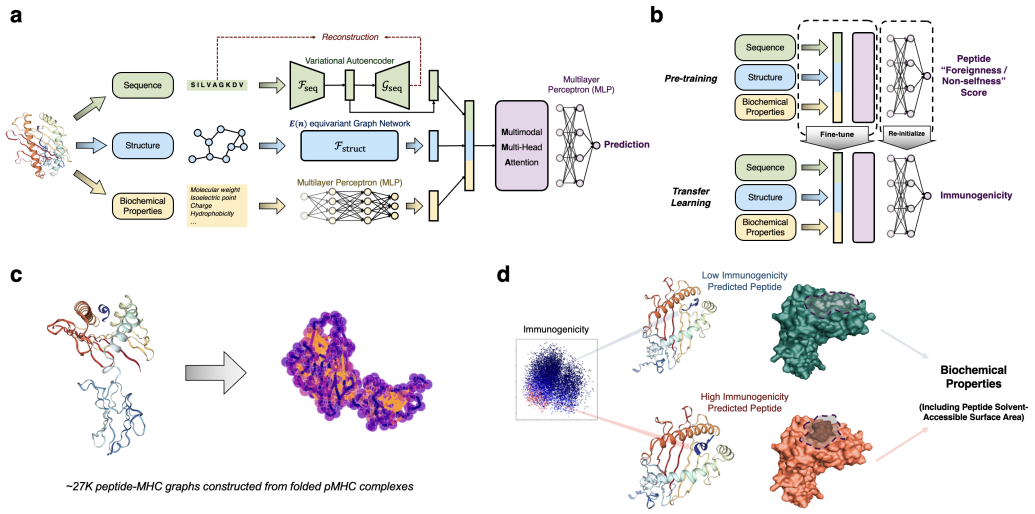

Figure 1: Overview of ImmunoStruct architecture and training. **(a)** ImmunoStruct separately encodes three modalities and uses the fused multimodal representations for downstream prediction. **(b)** During transfer learning, the model is pre-trained to predict peptide foreignness and further tuned to predict immunogenicity.**(c)** Graph construction from AlphaFold2 output. **(d)** Biochemical property computation from pMHC structure and sequence.

pMHC deep-learning models like NetMHCpan-4.1 [25], MHCnuggets [31], MHCflurry-2.0 [22], and BigMHC [1], most still rely mainly on peptide and MHC sequence data without incorporating multi-allelic structural information and biochemical information [39] that are proven to be critical for pMHC-TCR synapse recognition [28]. To bridge this gap, we first curated a multi-allele dataset of pMHC sequences, biochemical properties, and pMHC structures that we generated from our high-throughput AlphaFold2 pipeline [9]. Leveraging this multi-modal dataset, we developed ImmunoStruct to effectively integrate sequence information, structural data, and biochemical properties across pMHCs to improve immunogenicity prediction and interpretation.

2 Methods

2.1 The ImmunoStruct Model

ImmunoStruct integrates protein sequence, structure, and biochemical information for peptide-MHC immunogenicity prediction (Figure 1a). It takes in three input modalities: 1) peptide-MHC sequence , 2) peptide-MHC 3D structure given by AlphaFold2 (Figure 1c), and 3) biochemical properties (Figure 1d). These three modalities are separately modeled and subsequently merged into a multimodal embedding. The multimodal embedding is then passed to a multilayer perceptron (MLP) to produce a single scalar value for the prediction of immunogenicity. Technical details on data curation and processing are included in Appendix A.

Sequence modeling Peptide-MHC sequences are encoded by a variational autoencoder (VAE) [12], mapping to a 32-dimensional latent space (Figure 1a) and reverse-mapping to the input dimension to perform sequence reconstruction. This architecture choice is helpful for learning a compact and meaningful latent space that is semantically continuous [12, 33].

Structure modeling Peptide-MHC structures are processed using a graph convolutional network (GCN) [13] with $E(n)$ equivariant graph neural network (EGNN) [29] layers that are equivariant with respect to 3D rotation, translation, reflection, and permutation. The EGNN layers help capture spatial relationships and interactions within the complex. These layers generate node representations that are refined through a multi-head self-attention [35] module to produce a 64-dimensional structural embedding (Figure 1a).

Biochemical property modeling Given the known biochemical involvement of interactions within the pMHC-TCR synapse, and the importance of surface-area accessibility of the loaded peptide for immunogenicity [28, 37], we used the structural data for each pMHC sample to calculate the solvent-accessible surface area (SASA) of the pMHC to include as a component of the biochemical features (Figure 1d). The biochemical properties also included properties of each peptide, such as molecular weight, isoelectric point, charge, and hyrophobicity [20, 39]. We summarized these properties into two biochemical meta-properties (Figure 1a), and an MLP is used to further extract an 8-dimensional embedding from these meta-properties.

Feature combination via multimodal multi-head attention The 32-dimensional sequence embedding and the 64-dimensional structural embedding, along with the 8-dimensional biochemical property embedding, are concatenated to form a 104-dimensional representation vector (Figure 1a). This multimodal representation undergoes further processing with a multi-head attention layer, which consists of eight self-attention heads applied directly to the multimodal representation vector. This is an application of multi-head attention [35] on multimodal embedding vectors, and to the best of our knowledge, this is the first time it has been done in the immunogenicity prediction task.

2.2 Training, Optimization, and Interpretability

Training paradigm ImmunoStruct is trained using a transfer learning paradigm [43] (Figure 1b). After weight initialization, it is first pre-trained to predict foreignness, which has been shown to be associated with immunogenicity due to self-tolerance mechanisms [26, 39]. After that, it is trained to predict immunogenicity. All modules are warm-started from the pre-trained weights, with the exception of the final MLP which is re-initialized to accommodate the change in task (Figure 1b). The transfer learning process uses patterns between foreignness and epitope recognition [2, 26] to improve the performance on the main immunogenicity task that is different but related to the pre-training task.

Optimization objective During training, ImmunoStruct is simultaneously optimized by a prediction loss $\mathcal{L}_{\text{pred}}$, a protein sequence reconstruction loss $\mathcal{L}_{\text{recon}}$, and a normal distribution regularization on the latent space $\mathcal{L}_{\text{normal}}$. The loss function is given by Eqn (1), where the loss components are weighted by their respective coefficients $\lambda_1, \lambda_2, \lambda_3$.

$$\mathcal{L} = \lambda_1 \cdot \mathcal{L}_{\text{pred}}(\hat{Y}, Y) + \lambda_2 \cdot \mathcal{L}_{\text{recon}}(\hat{X}, X) + \lambda_3 \cdot \mathcal{L}_{\text{normal}}(\hat{Z}) \tag{1}$$

The prediction loss $\mathcal{L}_{\text{pred}}$ guides the model to predict the target accurately, while the reconstruction loss $\mathcal{L}_{\text{recon}}$ ensures that the learned representations carry sufficient information to faithfully reconstruct the input [14–16]. Additionally, the normal distribution regularization $\mathcal{L}_{\text{normal}}$ encourages the latent space of the VAE to approximate a standard normal distribution with zero mean and unit variance [12].

The loss components are shown in Eqn (2), (3), (4) and (5). $\hat{Y}$ and Y are, respectively, the set of predicted and ground truth values to be predicted (foreignness or immunogenicity). $\hat{X}$ and X are, respectively, the set of predicted and ground truth protein sequences. $\hat{Z}$ is the set of latent embedding vectors at the latent space of the VAE, while Z follows the normal distribution $\mathcal{N}(0, \mathbf{I})$. N is the batch size and $|\hat{X}| = |X| = |\hat{Y}| = |Y| = |\hat{Z}| = |Z| = N$.

$$\mathcal{L}^{\text{bin}}_{\text{pred}}(\hat{Y}, Y) = \text{BCE}(\hat{Y}, Y) = \frac{1}{N} \sum_{(\hat{y}, y) \in (\hat{Y}, Y)} -\Big(w_{\text{pos}}\, y \log(\hat{y}) + (1 - y) \log(1 - \hat{y}) \Big) \tag{2}$$

$$\mathcal{L}^{\text{cont}}_{\text{pred}}(\hat{Y}, Y) = \text{MSE}(\hat{Y}, Y) = \frac{1}{N} \sum_{(\hat{y}, y) \in (\hat{Y}, Y)} \Big(\hat{y} - y \Big)^2 \tag{3}$$

$$\mathcal{L}_{\text{recon}}(\hat{X}, X) = \text{MSE}(\hat{X}, X) = \frac{1}{N} \sum_{(\hat{x}, x) \in (\hat{X}, X)} \Big(\hat{x} - x \Big)^2 \tag{4}$$

$$\mathcal{L}_{\text{normal}}(\hat{Z}) = D_{\text{KL}}(\hat{Z} || Z) = \frac{1}{N} \sum_{(\hat{z}, z) \in (\hat{Z}, Z)} \Big(\hat{z} \log \frac{\hat{z}}{z} \Big) = \frac{1}{N} \sum_{\mathcal{N}(\mu, e^{\epsilon}) \sim \hat{Z}} \Big(\mu^2 + e^{\epsilon} - \epsilon - 1 \Big) \tag{5}$$

It should be noted that the prediction loss comes in two alternative forms: When the prediction target is binary (i.e., for immunogenicity), it uses the binary cross entropy loss (Eqn (2)); When the prediction target is continuous (i.e., for foreignness), it uses the mean squared error loss (Eqn (3)). The binary cross entropy loss is implemented using BCEWithLogitsLoss in PyTorch [23], where w_{pos}

is the positive class weighting coefficient. Following the official recommendation from PyTorch, we set w_{pos} as the ratio of negative to positive sample counts.

For numerical stability, the normal distribution regularization is implemented using the reparameterization trick [12], where two separate linear layers in the neural network are respectively used to produce the mean μ and log variance ϵ of the latent representations. Then, the KL divergence can be computed using summations, subtractions, and exponents, without involving divisions (Eqn (5)).

Interpretability of latent embedding Interpretability is enabled through the analysis and visualization of the VAE latent space (for sequence), the EGNN structural learned attention weights overlaid onto the pMHC (for structure), and the multimodal embeddings (for combined analysis of all modalities). Our end-to-end approach aims to capture the sequence-structural organization critical for peptide-MHC immunogenicity prediction.

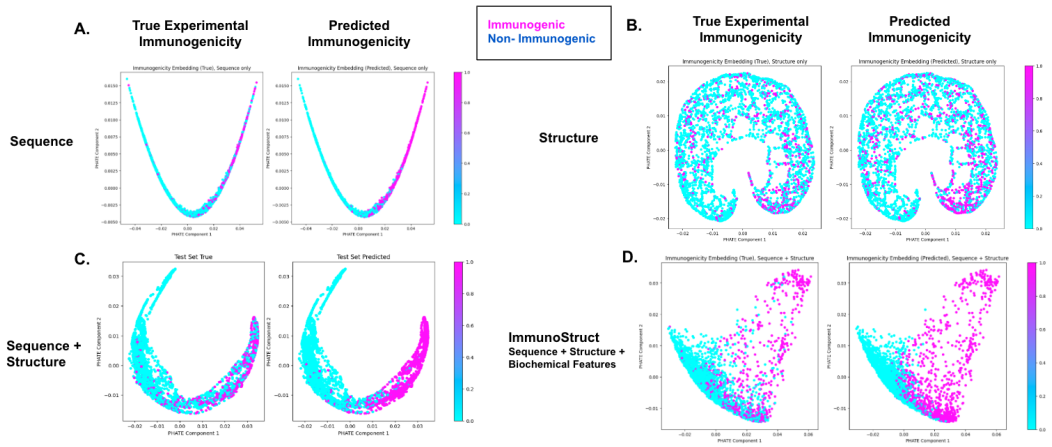

Figure 2: ImmunoStruct embedding visualizations between true and predicted immunogenicity scores for (A) the sequence-only model, (B) the structure-only model, (C) the sequence + structure model and (D) the ImmunoStruct model which includes sequence, structure, and biohemical properties.

Table 1: Ablation on the model components on IEDB data. All experiments were performed under 5 random seeds, and the mean and standard deviation values are displayed. For each metric, the best candidate is marked in **green** and the runner-up is marked in **blue**.
Bchems: biochemical properties, MMA: multimodal multi-head attention mechanism.

Components				Transfer Learning?	ROC AUC	PR AUC	mean PPVn
Structure	Sequence	Bchems	MMA				
✓					0.775 (0.012)	0.442 (0.022)	0.433 (0.020)
	✓				0.842 (0.006)	0.553 (0.014)	0.430 (0.009)
✓	✓				0.840 (0.007)	0.547 (0.024)	0.444 (0.028)
	✓	✓			0.842 (0.006)	0.554 (0.020)	0.419 (0.017)
✓	✓	✓			0.840 (0.006)	0.547 (0.018)	0.440 (0.032)
✓	✓	✓	✓		**0.860** (0.013)	**0.615** (0.045)	**0.462** (0.015)
✓				✓	0.805 (0.006)	0.497 (0.022)	0.414 (0.027)
	✓			✓	0.845 (0.007)	0.568 (0.024)	0.427 (0.017)
✓	✓			✓	0.844 (0.005)	0.569 (0.014)	0.440 (0.035)
	✓	✓		✓	0.844 (0.006)	0.561 (0.017)	0.442 (0.024)
✓	✓	✓		✓	0.840 (0.006)	0.554 (0.028)	0.441 (0.029)
✓	✓	✓	✓	✓	**0.882** (0.005)	**0.696** (0.020)	**0.514** (0.021)

3 Results

3.1 ImmunoStruct Benefits from Multimodal Learning

We first validate that ImmunoStruct benefits from all three input modalities – sequence, structure, and biochemical information. By visualizing the latent space embeddings using PHATE [18], we observed that the latent space of ImmunoStruct with all three modalities exhibited the clearest organization when comparing the true vs predicted labels (Figure 2). This observation is quantitatively confirmed by the immunogenicity prediction ablation study, as shown in Table 1. The ablation study also demonstrated that transfer learning is generally beneficial compared to training from scratch.

3.2 ImmunoStruct Improves Immunogenicity Prediction on Infectious Diseases

Next, we compared ImmunoStruct with existing methods, including Prime-2.0 [7], NetMHCpan [25], MHCnuggets [31], MHCflurry [22], BigMHC [1], and DeepNeo [11]. Performances are reported on ROC AUC, PR AUC and mean PPVn. ROC AUC and PR AUC are composite metrics that reflect the threshold-agnostic aggregated classification performance on {sensitivity, specificity} and {precision, recall}, respectively. Mean PPVn is a widely used and clinically important metric [1, 22, 25, 31] that measures the precision of several top-ranked pMHC candidates predicted in a setting similar to epitope ranking for clinical trial implementation. For all metrics, ImmunoStruct established a new state-of-the-art in immunogenicity prediction, as shown in Table 2.

3.3 ImmunoStruct Improves Immunogenicity Prediction on Human Cancer Neoepitopes

The clinical potential of therapeutic neoeptiope based vaccines has been increasing in recent years [4, 6, 8], so we next sought to explore the prediction performance of ImmunoStruct in the setting of

Table 2: Comparison of immunogenicity prediction methods on **IEDB data**. All experiments were performed under 5 random seeds, and the mean and standard deviation values are displayed. For each metric, the best candidate is marked in **green** and the runner-up is marked in **blue**.

Model	ROC AUC	PR AUC	mean PPVn
Prime-2.1 [7]	0.538 (0.012)	0.212 (0.031)	0.207 (0.028)
NetMHCpan [25]	0.537 (0.027)	0.214 (0.020)	0.220 (0.018)
DeepNeo [11]	0.767 (0.032)	0.438 (0.024)	0.411 (0.017)
MHCnuggets [31]	0.546 (0.023)	0.246 (0.019)	0.233 (0.031)
MHCflurry [22]	0.577 (0.021)	0.260 (0.026)	0.273 (0.028)
BigMHC-EL [1]	0.588 (0.018)	0.269 (0.016)	0.290 (0.024)
BigMHC-IM [1]	0.684 (0.028)	0.319 (0.048)	0.333 (0.034)
$\text{BigMHC}_{\text{retrained}}$ [1]	**0.793** (0.013)	**0.462** (0.027)	**0.445** (0.009)
ImmunoStruct (ours)	**0.882** (0.005)	**0.696** (0.020)	**0.514** (0.021)

Table 3: Comparison of immunogenicity prediction methods on **cancer neoepitopes**. All experiments were performed under 5 random seeds, and the mean and standard deviation values are displayed. For each metric, the best candidate is marked in **green** and the runner-up is marked in **blue**.

Model	ROC AUC	PR AUC	mean PPVn
Prime-2.1 [7]	0.645 (0.026)	0.259 (0.023)	0.295 (0.029)
NetMHCpan [25]	0.559 (0.049)	0.210 (0.037)	0.211 (0.045)
DeepNeo [11]	0.535 (0.016)	0.261 (0.053)	0.222 (0.070)
MHCnuggets [31]	0.530 (0.014)	0.210 (0.017)	0.175 (0.023)
MHCflurry [22]	0.658 (0.015)	0.304 (0.032)	0.329 (0.030)
BigMHC-EL [1]	0.632 (0.034)	0.248 (0.039)	0.245 (0.042)
BigMHC-IM [1]	**0.771** (0.040)	**0.373** (0.062)	**0.357** (0.036)
$\text{BigMHC}_{\text{retrained}}$ [1]	0.682 (0.012)	0.310 (0.020)	0.325 (0.030)
ImmunoStruct (ours)	**0.770** (0.038)	**0.410** (0.078)	**0.374** (0.139)

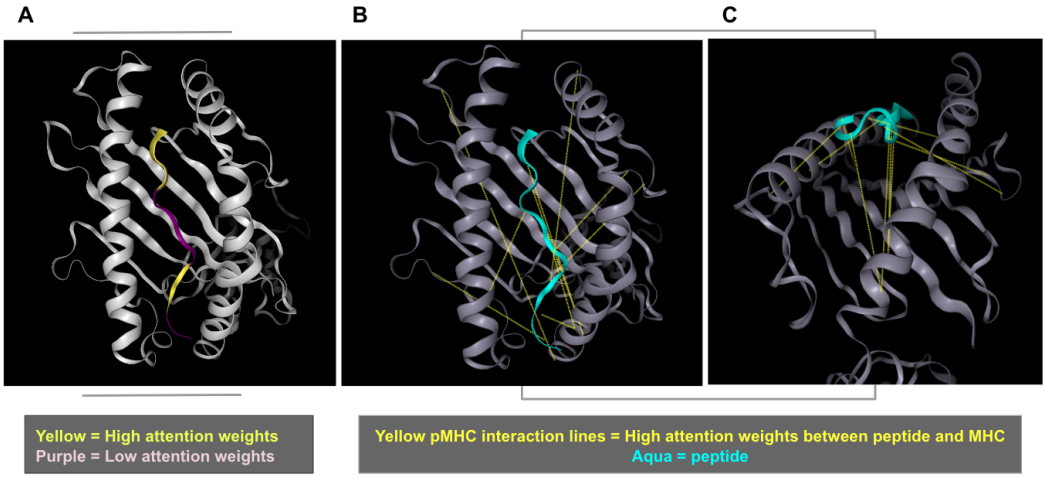

Figure 3: Visualization of structural regions that are influential to predicted immunogenicity. Immunogenicity learnred attention weights are shown for the peptide itself (top view, panel A) and for the peptide-MHC interactions from a top view (panel B) and side view (panel C).

cancer neoepitope immunogenicity. Single mutation changes to a protein can lead to presentation of a peptide that only differs by one amino acid from its wild-type form, which may make distinguishing immunogenic neoepitopes challenging [5, 37]. Upon extending ImmunoStruct to this data setting, we observed superior performance compared to existing methods (Table 3).

Moreover, we validated ImmunoStruct in a cohort of cancer patients comprising a variety of different tumor types and immunotherapeutic treatments [5]. In comparison to tumor-mutational burden, we found ImmunoStruct predictions were able to stratify patient survival more effectively for both overall survival and progression-free survival without any re-training ($p < 0.05$, data not shown).

3.4 Exploring ImmunoStruct Predictions by Visualizing pMHC Attention

To next explore which parts of the pMHC structure contribute most to immunogenicity prediction, we extracted the ImmunoStruct attention weights from the structural multi-head attention layers. To date and to our knowledge, there are no structural graph attention-based deep learning methods from wide-scale pMHC structure data for immunogenic biostructural interpretation. We therefore built out a visualization module that highlights critical regions of the pMHC complex that influence the predicted immunogenicity, providing insights into structural determinants of epitope recognition.

Figure 3 shows examples of attention visualizations. An example immunogenic 9-mer peptide was AlphaFolded into the HLA-A:A0201 protein. The top angle shows highly weighted positions at position 2, 8, and 9, two of which have been previously been shown [37] to be important for peptide presentation and immunogenicity (Figure 3A). To next visualize the top pMHC interaction weights rather than the peptide positions themselves, we overlaid the node-to-node weights from a top view (Figure 3B) and side view (Figure 3C). This showed both local and distant node-to-node relationships, providing a way to explore ImmunoStruct prediction and suggest potential bio-spatial interactions for future exploration.

3.5 ImmunoStruct Predictions Align with Experimental Immunogenicity Assay Results

Given the challenge of immunogenicity prediction and importance of further validation, we aligned our predictions with ELISpot IFNg measurements that we generated *in vitro* for a set peptides in combination with one of five HLA alleles. ImmunoStruct successfully predicted 15 of 19 immunogenicities (ROC AUC = 0.780, PR AUC = 0.941, and accuracy = 79.0%), providing an additional validation of the model (Figure 4A). For each allele, one plate is shown to visualize IFNg production as an example plate for a given pMHC (Figure 4B). Details of the *in vitro* immunogenicity validation are described in Appendix B.

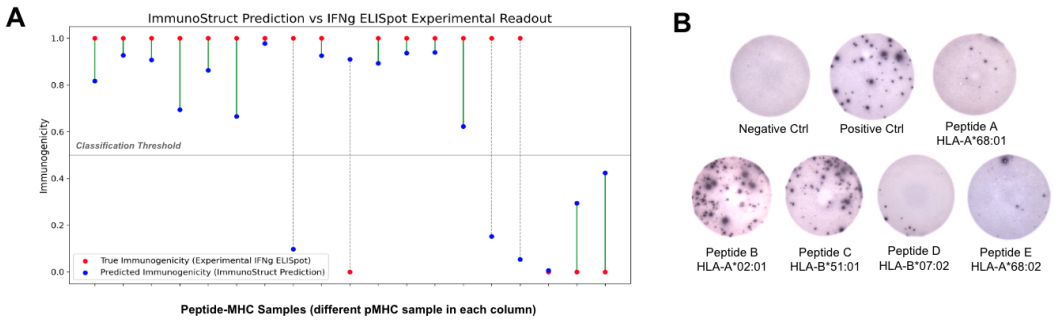

Figure 4: Comparison of ImmunoStruct-predicted immunogenicity and experimental IFNg ELISpot readouts across different peptide-MHC (pMHC) samples. (A) Lollipop plot visualization where each column represents a unique epitope-MHC pair, with red dots indicating the experimentally measured immunogenicity (IFNg ELISpot) and blue dots showing the predicted immunogenicity from ImmunoStruct. Vertical lines connect paired values for each sample, with green lines for correct predictions and dashed gray lines for incorrect predictions. (B) ELISpot plate images (one example plate from each allele validated).

4 Conclusion

Our work highlights the first deep learning model which uses large-scale multi-allelic AlphaFold-generated structural and biochemical pMHC data for an application to the immunogenicity prediction task, across both infectious disease and cancer neoepitopes. In addition, the multi-head attention layers enable immunogenic motif hypothesis generation and machine-guided interpretation of pMHC immunogenicity. With the advent of mRNA vaccines, cancer neoantigen vaccines, as well as immunogenicity bio-therapeutic toxicity challenges, this work also has applications to protein-based therapeutic discovery. Moreover, it may provide a framework for future protein-based prediction tasks that may benefit from integrating predicted structural and biochemical data for downstream applications.

References

[1] B. A. Albert, Y. Yang, X. M. Shao, D. Singh, K. N. Smith, V. Anagnostou, and R. Karchin. Deep neural networks predict class i major histocompatibility complex epitope presentation and transfer learn neoepitope immunogenicity. *Nature machine intelligence*, 5(8):861–872, 2023. 2, 5

[2] V. P. Balachandran, M. Łuksza, J. N. Zhao, V. Makarov, J. A. Moral, R. Remark, B. Herbst, G. Askan, U. Bhanot, Y. Senbabaoglu, et al. Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. *Nature*, 551(7681):512–516, 2017. 3

[3] A.-M. Bjerregaard, M. Nielsen, V. Jurtz, C. M. Barra, S. R. Hadrup, Z. Szallasi, and A. C. Eklund. An analysis of natural t cell responses to predicted tumor neoepitopes. *Frontiers in immunology*, 8:1566, 2017. 1

[4] E. Blass and P. A. Ott. Advances in the development of personalized neoantigen-based therapeutic cancer vaccines. *Nature reviews Clinical oncology*, 18(4):215–229, 2021. 1, 5

[5] A. Borch, I. Carri, B. Reynisson, H. M. G. Alvarez, K. K. Munk, A. Montemurro, N. P. Kristensen, S. A. Tvingsholm, J. S. Holm, C. Heeke, et al. Improve: a feature model to predict neoepitope immunogenicity through broad-scale validation of t-cell recognition. *Frontiers in Immunology*, 15:1360281, 2024. 6, 11

[6] T. Fan, M. Zhang, J. Yang, Z. Zhu, W. Cao, and C. Dong. Therapeutic cancer vaccines: advancements, challenges, and prospects. *Signal Transduction and Targeted Therapy*, 8(1):450, 2023. 1, 5

[7] D. Gfeller, J. Schmidt, G. Croce, P. Guillaume, S. Bobisse, R. Genolet, L. Queiroz, J. Cesbron, J. Racle, and A. Harari. Improved predictions of antigen presentation and tcr recognition with mixmhcpred2. 2 and prime2. 0 reveal potent sars-cov-2 cd8+ t-cell epitopes. *Cell Systems*, 14 (1):72–83, 2023. 5

[8] M. Gupta, A. Wahi, P. Sharma, R. Nagpal, N. Raina, M. Kaurav, J. Bhattacharya, S. M. Rodrigues Oliveira, K. G. Dolma, A. K. Paul, et al. Recent advances in cancer vaccines: challenges, achievements, and futuristic prospects. *Vaccines*, 10(12):2011, 2022. 1, 5

[9] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, et al. Highly accurate protein structure prediction with alphafold. *nature*, 596(7873):583–589, 2021. 2, 11

[10] P. D. Katsikis, K. J. Ishii, and C. Schliehe. Challenges in developing personalized neoantigen cancer vaccines. *Nature Reviews Immunology*, 24(3):213–227, 2024. 1

[11] J. Y. Kim, H. Bang, S.-J. Noh, and J. K. Choi. Deepneo: a webserver for predicting immunogenic neoantigens. *Nucleic acids research*, 51(W1):W134–W140, 2023. 5

[12] D. P. Kingma. Auto-encoding variational bayes. In *International Conference on Learning Representations*, 2014. 2, 3, 4

[13] T. N. Kipf and M. Welling. Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations*, 2017. 2

[14] C. Liu, M. Amodio, L. L. Shen, F. Gao, A. Avesta, S. Aneja, J. C. Wang, L. V. Del Priore, and S. Krishnaswamy. Cuts: A deep learning and topological framework for multigranular unsupervised medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 155–165. Springer, 2024. 3

[15] C. Liu, D. Liao, A. Parada-Mayorga, A. Ribeiro, M. DiStasio, and S. Krishnaswamy. Diffkillr: Killing and recreating diffeomorphisms for cell annotation in dense microscopy images. *arXiv preprint arXiv:2410.03058*, 2024.

[16] C. Liu, K. Xu, L. L. Shen, G. Huguet, Z. Wang, A. Tong, D. Bzdok, J. Stewart, J. C. Wang, L. V. Del Priore, and S. Krishnaswamy. Imageflownet: Forecasting multiscale image-level trajectories of disease progression with irregularly-sampled longitudinal medical images. *arXiv preprint arXiv:2406.14794*, 2024. 3

[17] M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, and M. Steinegger. Colabfold: making protein folding accessible to all. *Nature methods*, 19(6):679–682, 2022. 11

[18] K. R. Moon, D. Van Dijk, Z. Wang, S. Gigante, D. B. Burkhardt, W. S. Chen, K. Yim, A. v. d. Elzen, M. J. Hirn, R. R. Coifman, et al. Visualizing structure and transitions in high-dimensional biological data. *Nature biotechnology*, 37(12):1482–1492, 2019. 5

[19] A. Nelde, T. Bilich, J. Heitmann, et al. Sars-cov-2-derived peptides define heterologous and covid-19-induced t cell recognition. *Nature Immunology*, 22(1):74–85, 2021. doi: 10.1038/s41590-020-00808-x. 11

[20] D. Osorio, P. Rondón-Villarreal, and R. Torres. Peptides: a package for data mining of antimicrobial peptides. *The R Journal*, 7(1):4–14, 2015. doi: 10.32614/RJ-2015-001. 3, 11

[21] P. A. Ott, S. Hu-Lieskovan, B. Chmielowski, R. Govindan, A. Naing, N. Bhardwaj, K. Margolin, M. M. Awad, M. D. Hellmann, J. J. Lin, et al. A phase ib trial of personalized neoantigen therapy plus anti-pd-1 in patients with advanced melanoma, non-small cell lung cancer, or bladder cancer. *Cell*, 183(2):347–362, 2020. 1

[22] T. J. O'Donnell, A. Rubinsteyn, and U. Laserson. Mhcflurry 2.0: improved pan-allele prediction of mhc class i-presented peptides by incorporating antigen processing. *Cell systems*, 11(1): 42–48, 2020. 2, 5

[23] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019. 3

[24] N. Pishesha, T. J. Harmand, and H. L. Ploegh. A guide to antigen processing and presentation. *Nature Reviews Immunology*, 22(12):751–764, 2022. 1

[25] B. Reynisson, B. Alvarez, S. Paul, B. Peters, and M. Nielsen. Netmhcpan-4.1 and netmhciipan-4.0: improved predictions of mhc antigen presentation by concurrent motif deconvolution and integration of ms mhc eluted ligand data. *Nucleic acids research*, 48(W1):W449–W454, 2020. 2, 5

[26] L. P. Richman, R. H. Vonderheide, and A. J. Rech. Neoantigen dissimilarity to the self-proteome predicts immunogenicity and response to immune checkpoint blockade. *Cell Systems*, 9(4): 375–382, 2019. 3

[27] U. Sahin, E. Derhovanessian, M. Miller, B.-P. Kloke, P. Simon, M. Löwer, V. Bukur, A. D. Tadmor, U. Luxemburger, B. Schrörs, et al. Personalized rna mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature*, 547(7662):222–226, 2017. 1

[28] K. Saotome, D. Dudgeon, K. Colotti, M. J. Moore, J. Jones, Y. Zhou, A. Rafique, G. D. Yancopoulos, A. J. Murphy, J. C. Lin, et al. Structural analysis of cancer-relevant tcr-cd3 and peptide-mhc complexes by cryoem. *Nature Communications*, 14(1):2401, 2023. 2, 3

[29] V. G. Satorras, E. Hoogeboom, and M. Welling. E (n) equivariant graph neural networks. In *International conference on machine learning*, pages 9323–9332. PMLR, 2021. 2

[30] A. Sette and J. Sidney. Nine major hla class i supertypes account for the vast preponderance of hla-a and -b polymorphism. *Immunogenetics*, 50(3):201–212, 1999. 11

[31] X. M. Shao, R. Bhattacharya, J. Huang, I. A. Sivakumar, C. Tokheim, L. Zheng, D. Hirsch, B. Kaminow, A. Omdahl, M. Bonsack, et al. High-throughput prediction of mhc class i and ii neoantigens with mhcnuggets. *Cancer immunology research*, 8(3):396–408, 2020. 2, 5

[32] M. Slota, J.-B. Lim, Y. Dang, and M. L. Disis. Elispot for measuring human immune responses to vaccines. *Expert Review of Vaccines*, 10(3):299–306, 2011. doi: 10.1586/erv.10.169. 11

[33] X. Sun, D. Liao, K. MacDonald, Y. Zhang, C. Liu, G. Huguet, G. Wolf, I. Adelstein, T. G. Rudner, and S. Krishnaswamy. Geometry-aware generative autoencoders for warped riemannian metric learning and generative modeling on data manifolds. *arXiv preprint arXiv:2410.12779*, 2024. 2

[34] F. E. Tynan, D. Elhassen, A. W. Purcell, J. M. Burrows, N. A. Borg, J. J. Miles, N. A. Williamson, K. J. Green, J. Tellam, L. Kjer-Nielsen, et al. The immunogenicity of a viral cytotoxic t cell epitope is controlled by its mhc-bound conformation. *The Journal of experimental medicine*, 202(9):1249–1260, 2005. 1

[35] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017. 2, 3

[36] R. Vita, S. Mahajan, J. A. Overton, S. K. Dhanda, S. Martini, J. R. Cantrell, D. K. Wheeler, A. Sette, and B. Peters. The immune epitope database (iedb): 2018 update. *Nucleic Acids Research*, October 24 2018. doi: 10.1093/nar/gky1006. 11

[37] J. K. Weber, J. A. Morrone, S.-g. Kang, L. Zhang, L. Lang, D. Chowell, C. Krishna, T. Huynh, P. Parthasarathy, B. Luan, et al. Unsupervised and supervised ai on molecular dynamics simulations reveals complex characteristics of hla-a2-peptide immunogenicity. *Briefings in Bioinformatics*, 25(1):bbad504, 2024. 1, 3, 6

[38] D. Weiskopf, M. A. Angelo, E. L. de Azeredo, J. Sidney, J. A. Greenbaum, A. N. Fernando, A. Broadwater, R. V. Kolla, A. D. De Silva, A. M. de Silva, et al. Comprehensive analysis of dengue virus-specific responses supports an hla-linked protective role for cd8+ t cells. *Proceedings of the National Academy of Sciences*, 110(22):E2046–E2053, 2013. 11

[39] D. K. Wells, M. M. van Buuren, K. K. Dang, V. M. Hubbard-Lucey, K. C. Sheehan, K. M. Campbell, A. Lamb, J. P. Ward, J. Sidney, A. B. Blazquez, et al. Key parameters of tumor epitope immunogenicity revealed through a consortium approach improve neoantigen prediction. *Cell*, 183(3):818–834, 2020. 1, 2, 3

[40] P. Wu, T. Zhang, B. Liu, P. Fei, L. Cui, R. Qin, H. Zhu, D. Yao, R. J. Martinez, W. Hu, et al. Mechano-regulation of peptide-mhc class i conformations determines tcr antigen recognition. *Molecular cell*, 73(5):1015–1027, 2019. 1

[41] F. Yang, K. Patton, T. Kasprzyk, B. Long, S. Gupta, S. J. Zoog, K. Tracy, and C. Vettermann. Validation of an ifn-gamma elispot assay to measure cellular immune responses against viral antigens in non-human primates. *Gene Therapy*, 29(1-2):41–54, 2022. doi: 10.1038/s41434-020-00214-w. 11

[42] N. Zaidi, M. Soban, F. Chen, H. Kinkead, J. Mathew, M. Yarchoan, T. D. Armstrong, S. Haider, and E. M. Jaffee. Role of in silico structural modeling in predicting immunogenic neoepitopes for cancer vaccine development. *JCI insight*, 5(17), 2020. 11

[43] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1):43–76, 2020. 3

Appendices

A Multimodal Dataset for Immunogenicity Prediction

Due to the scarcity of publicly available peptide-MHC structural data, we implemented a high-throughput AlphaFold2 pipeline to generate the largest pMHC structural dataset to our knowledge. This dataset ulmitately comprised of approximately 27,000 entries with paired experimental immunogenicity labels.

A.1 pMHC Sequence Data

In order to curate a dataset that would provide sufficient epitope sequence training data across alleles, we downloaded data from the immmune epitope database (IEDB) [36]. Peptide sequences were filtered to several criteria: (1) human sequences, (2) T-Cell experimental assay data (ELISpot, Tetramer staining), and MHC class-I restriction. To avoiding training across rare human alleles with small and often only negative sample sizes, we curated a dataset for the top 27 most common alleles in the human population ($\sim$ 97% human population coverage) as previously described [30, 38]. This provided a dataset of 27,000 peptide-MHC pairs with known positive or negative experimental epitope label for the IEDB dataset. Cancer neoepitope pMHC sequence data was obtained as previously described and filtered to the same allele set [5, 30, 36, 38].

A.2 pMHC AlphaFold2 Structure Data

The AlphaFold2 pipeline was implemented as proposed previously [9]. To run our high-throughout dataset of peptide-MHC sequence combinations, we made use of ColabFold [17], a wrapper on top of AlphaFold2 that enables highly customizable and easily useable protein folding operations. With ColabFold, we were able to generate a protein folding from peptide-MHC pairs with only a few lines of code and a shared set of configurations. This enabled us to distribute our protein generation across multiple hosts, drastically reducing the amount of time taken from several weeks to a couple of days. After the proteins were generated, we applied post processing to the structures to convert them to a deep learning model readable graph format containing hydrogen-bond interactions, ionic interactions, and hydrophobic interactions, as well as amino acid node features and their 3D positional information.

A.3 Biochemical Properties

Reasoning the potential benefit of including biochemical properties as additional features in our prediction task, we integrated biochemical properties into the ImmunoStruct prediction. The peptides package [20] was used to calculate all possible 84 biochemical properties for each peptide. The top 10 biochemical properties with the highest standard deviation were selected in addition to the pMHC solvent-accessible surface area score (SASA), computed as previously described [42]. These biochemical/biostructural properties were then grouped into two meta-property features (Bchem1 and Bchem2) to be used as additional features during training.

B pMHC *in vitro* Immunogenicity Validation

IFN-gamma-based enzyme-linked immunospot (ELISpot) assays were used to validate a set of SARS-CoV-2 epitopes. ELISpot assays provide a standardized measurement of T cell responses [32, 41]. Previously frozen human peripheral blood mononuclear cells (PBMCs) were thawed rapidly to 37°C and incubated overnight (Millipore, Massachusetts, U.S.A.). PMBCs were plated at 400,000 cells/well and co-incubated with 10 μg/mL of peptide for 48 hours at 37°C. ELISpot plates were analyzed using CTL Immunospot. Positive controls were PMBCs plated at 200,000 cells/well and co-incubated with Spike Glycoprotien peptide megapool at 1 μg/mL (JPT Peptide Technologies, Berlin, Germany). Negative controls were unstimulated PMBCs plated at 200,000 cells/well. Positive T cell responses were defined by greater or equal to the mean of negative controls + 3. Three additional true negative peptide-MHC samples validated through similar methods were identified through literature review to improve data balance [19].