ImmunoFoundation: A Multimodal Foundation Model for Immunogenicity Prediction and Peptide Optimization

João Felipe Rocha[*1] **Hiren Madhu**[*1] **Jenny Yongjia Liu**[1] **Apurva Mishra**[1] **Chen Liu**[1] **Rishabh Anand**[1] **Rex Ying**[1] **Smita Krishnaswamy**[1]

Abstract

Peptide immunogenicity, whether a peptide presented by an MHC molecule elicits a T-cell response, is central to designing vaccines, cancer immunotherapy, and therapeutic proteins. Existing tools rely on a single modality, such as peptide sequences or peptide–MHC interactions, and often ignore the T-cell response that depends on the TCR–peptide–MHC complex (TCR-pMHC) and its three-dimensional structure. The scarcity of labeled TCR-pMHC data with known structures makes it difficult to build a model that captures how all components of the TCR-pMHC contribute to immunogenicity. However, a *foundation model* of TCR-pMHCs can learn transferable representations across components, which can be adapted to immunogenicity, binding, and TCR specificity tasks, even with limited labeled data. We introduce **ImmunoFoundation**, a self-supervised multimodal backbone for protein-complex representation, fine-tuned for peptide–MHC immunogenicity. The model couples an ESM-2 sequence encoder with a Graph Transformer over structure, fused via cross-modal attention. Pretraining follows a curriculum that progressively introduces structural inductive bias. **ImmunoFoundation** outperforms prior multimodal class-I predictors on cancer neoepitope and infectious-disease tasks.

1 Introduction

Accurate immunogenicity prediction is foundational to translational immunology, driving the rational design of protein therapeutics with reduced immune-related adverse effects, such as cancer neoepitopes for personalized vaccines (Wells et al., 2020) and biologics for autoimmune diseases.

[1]Yale University. [*]Equal Contribution. Correspondence to: João Felipe Rocha <joaofelipe.rocha@yale.edu>.

Proceedings of the 43^{rd} International Conference on Machine Learning, Seoul, South Korea. PMLR 306, 2026.

Immunogenicity is driven by the T-cell response, which depends on interactions between T-cell receptors (TCR), an antigen peptide (epitope), and a major histocompatibility complex (MHC), which together form the TCR-peptide–MHC complex (TCR-pMHC) (Pishesha et al., 2022). These interactions are governed by amino acid sequence, three-dimensional structure, and biochemical properties. An accurate *in silico* ranking of peptides likely to elicit a T-cell response would reduce the number of candidate cancer neoepitopes derived to an experimentally tractable subset.

Most existing tools target the MHC class I pathway and operate on pMHC sequences alone: NetMHCpan (Hoof et al., 2009), MHCflurry (O'Donnell et al., 2020), MHCnuggets (Shao et al., 2020), BigMHC (Albert et al., 2022), PRIME2.0 (Schmidt et al., 2021). By construction, sequence-only models cannot capture structural determinants of immunogenicity such as the three-dimensional peptide conformation, peptide accessibility to the TCR, or the conformational stability of the TCR-pMHC complex. Structure-aware models that capture these features are bottlenecked by the scarcity of peptide–MHC structures, but high-accuracy structure prediction at scale (Jumper et al., 2021; Abramson et al., 2024) and the proteome-scale expansion of AlphaFoldDB to protein complexes (Han et al., 2026) is removing this bottleneck. To our knowledge, ImmunoStruct (Givechian et al., 2025) is the only published multimodal peptide–MHC immunogenicity predictor that jointly processes sequence, structure, and biochemical descriptors via explicit fusion. However, it is restricted to the MHC class I pathway for immunogenicity and was trained supervised on 26,000 labeled peptide–MHC complexes, and thus cannot leverage the expanding pool of unlabeled structures (Varadi & Velankar, 2023; Han et al., 2026).

A complementary approach targets TCR-peptide recognition directly (Montemurro et al., 2021; Springer et al., 2020) but lacks MHC structural context. These limitations share a root cause: immunogenicity is a downstream readout of a more general object, the TCR-pMHC complex. Closing the gap calls for a foundation model that learns complex representations from unlabeled structures – spanning TCR recognition and MHC class I and II pathways – and can be

specialized to immunogenicity.

We introduce **ImmunoFoundation**, a self-supervised multimodal backbone for protein-complex representation, fine-tuned for peptide–MHC (pMHC) immunogenicity. The architecture pairs a frozen ESM-2 encoder (Lin et al., 2023) with a Graph Transformer over predicted structure, fused through cross-modal attention. The central methodological choice is how structural inductive bias is introduced: rather than imposing the full multi-chain prior up front, we introduce it progressively. Stage 1 pretrains the structure track on AlphaFoldDB monomers (Varadi & Velankar, 2023) under masked coordinate autoencoding, with a sequence-side masked language modeling head atop the frozen ESM-2 encoder. Stage 2 introduces the complex inductive bias via continued pretraining on the 1.8M complexes of AlphaFoldDB-complex (Han et al., 2026), learning interface geometry and inter-chain co-arrangement. The cross-modal attention module is then fine-tuned on labeled immunogenicity data.

Our contributions are: (i) **ImmunoFoundation**, the first self-supervised foundation backbone for protein complexes, specialized to immunogenicity; (ii) a two-stage pretraining curriculum that progressively introduces a multi-chain inductive bias, exploiting the recently released proteome-scale AlphaFoldDB-complex set; (iii) an evaluation against the previous best multimodal MHC class I predictor on infectious-disease and cancer neoepitope benchmarks. The backbone is designed to admit a third curriculum stage on peptide–MHC-specific structures and downstream extensions to MHC class II and TCR-pMHC complexes, providing a route to rational design of TCRs.

2 Method

ImmunoFoundation is a multimodal encoder for protein complexes, fine-tuned for pMHC immunogenicity after a curriculum that progressively introduces structural inductive bias (Figure 1). The model has two parallel encoders – for sequence and structure – that are fused at the fine-tuning stage by a cross-modal attention module. Pretraining is self-supervised; the immunogenicity head is supervised.

2.1 Architecture

Sequence track. Let $a = (a_1, \ldots, a_L)$ denote a protein sequence of length L, where $a_i \in \{1, \ldots, 20\}$ is a canonical amino acid. The track consists of a frozen ESM-2 (650M) encoder f_{ESM} (Lin et al., 2023), a trainable transformer adaptor g_θ, and a trainable language-modeling head h_ϕ. ESM-2 produces per-residue embeddings, $f_{\text{ESM}}(a) \in \mathbb{R}^{L \times d_{\text{ESM}}}$, further projected through the adaptor to hidden dimension d: $\mathbf{z}^{\text{seq}} = g_\theta(f_{\text{ESM}}(a)) \in \mathbb{R}^{L \times d}$. The LM head is used only as an auxiliary pretraining target; the downstream pipeline consumes $\mathbf{z}^{\text{seq}}$ from the adaptor, not the LM head logits.

Structure track. The structure track operates on a 2D residue-level k-nearest neighbors graph derived from AlphaFold3-predicted 3D structure. Each residue i has node features $\mathbf{h}_i = [\mathbf{z}_i^{\text{seq}} \| \mathbf{x}_i]$, where $\mathbf{x}_i \in \mathbb{R}^3$ is the 3D coordinates of the amino acid's C_α atom and $\|$ is a concatenation. The track uses a Graph Transformer (Ying et al., 2021; Dwivedi & Bresson, 2020) with N_s blocks, each performing multi-head self-attention over residue node tokens producing refined per-residue embeddings, $\mathbf{z}^{\text{str}} \in \mathbb{R}^{L \times d}$.

Cross-modal fusion. The two tracks are fused by a cross-modal attention module. Given per-residue embeddings $\mathbf{z}^{\text{seq}}$ and $\mathbf{z}^{\text{str}}$, fused tokens are produced by bidirectional cross-attention,

$$\begin{aligned}\mathbf{z}^{\text{fused}} = &\,\text{MHA}\big(\mathbf{Q} = \mathbf{z}^{\text{seq}}, \mathbf{K} = \mathbf{z}^{\text{str}}, \mathbf{V} = \mathbf{z}^{\text{str}}\big) \\ &+ \text{MHA}\big(\mathbf{Q} = \mathbf{z}^{\text{str}}, \mathbf{K} = \mathbf{z}^{\text{seq}}, \mathbf{V} = \mathbf{z}^{\text{seq}}\big)\,,\end{aligned}$$

followed by feed-forward layers. A learned `[CLS]` token is prepended at this stage, and its post-fusion output serves as the complex-level representation. The fusion module is instantiated only at the supervised fine-tuning stage.

2.2 Pretraining curriculum

The two tracks are pretrained independently under separate self-supervised objectives. The sequence track is trained with masked language modeling (MLM) and the structure track with masked autoencoding (MAE) on 3D coordinates. The tracks do not share parameters during pretraining. The cross-modal fusion module is instantiated at fine-tuning. The sequence-side objective trains the LM head, h_ϕ, to recover masked residues from the masked sequence embedding,

$$\mathcal{L}_{\text{MLM}} = -\mathbb{E}_{a,\mathcal{M}} \sum_{i \in \mathcal{M}} \log p_\phi\big(a_i \,\big|\, g_\theta(f_{\text{ESM}}(\tilde{a}))\big)\,, \qquad (1)$$

where $\tilde{a}$ is the masked sequence and $\mathcal{M}$ the masked positions (15% mask rate). ESM-2 weights stay frozen while only θ and ϕ are updated. The structure-side objective trains a small decoder h_ξ to recover the 3D C_α coordinates of a masked subset $\mathcal{C}$ of nodes from the encoder output,

$$\mathcal{L}_{\text{MAE}} = \mathbb{E}_{G,\mathcal{C}} \frac{1}{|\mathcal{C}|} \sum_{i \in \mathcal{C}} \big\| h_\xi\big(\mathbf{z}_i^{\text{str}}\big) - \mathbf{x}_i \big\|_2^2\,, \qquad (2)$$

updating the structure track parameters ψ and decoder parameters ξ.

Masking. At each step, surface residues—those with fewer than 12 neighboring residues within 8 Å—are masked, and their C_α coordinates are zeroed at the input. The encoder must reconstruct the masked positions from the remaining residues, which provide longer-range structural context.

The curriculum introduces structural inductive bias progressively: Stage 1 teaches the structure track the fold of a single

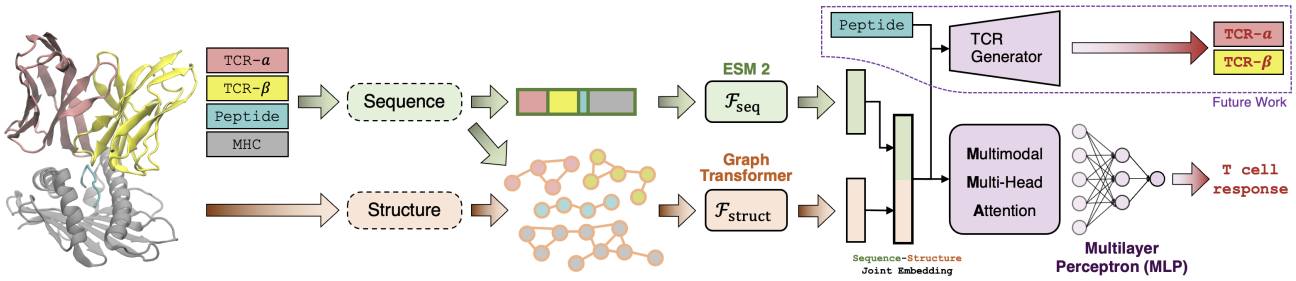

Figure 1. **ImmunoFoundation** pipeline. The sequence track is a frozen ESM-2 encoder followed by a trainable transformer adaptor and a language-modeling head trained with masked language modeling. The structure track is a Graph Transformer over k-nearest-neighbor residue graphs derived from predicted three-dimensional coordinates, trained with masked coordinate autoencoding of surface residues.

chain; Stage 2 extends this to multi-chain arrangements by changing the data distribution while keeping the objective fixed. This staging mirrors the AlphaFold to AlphaFold-Multimer trajectory (Evans et al., 2021) and follows the principle that progressive (vs upfront) inductive bias eases optimization and improves transfer (Bengio et al., 2009). The masking criterion, loss weights, and optimizer settings are shared across stages. Stage 1 weights initialize Stage 2.

Stage 1: Monomers. Inputs are single chains from the AlphaFoldDB human proteome and Swiss-Prot subsets (Varadi & Velankar, 2023). The model learns folded-chain geometry at the residue, neighborhood, and global levels.

Stage 2: Dimers. Inputs are multi-chain complexes from the proteome-scale AlphaFoldDB-complex set of 1.8M high-confidence predicted complexes (Han et al., 2026). The multi-chain inductive bias enters through the data rather than the objective: the kNN graph is built over the entire complex, so attention routes information across chain boundaries. Reconstructing residues located at interfaces uses cross-chain context that monomer pretraining cannot provide.

Stage 3: Closing the pMHC distribution gap. In Stage 2, the multi-chain signal is mediated by data and graph topology rather than a pMHC-specific objective. Our pretraining corpus, AlphaFoldDB-complex, is dominated by globular protein complexes. However, pMHCs have a peptide-in-groove geometry that is underrepresented in this corpus. We plan to address this gap with a third curriculum stage on a pMHC-specific structural corpus, the natural lever for pushing the backbone closer to the downstream task.

2.3 Fine-tuning

For the IEDB pMHC immunogenicity, inputs are the peptide and MHC sequences, and an AlphaFold3 predicted structure. The cross-modal attention module and a binary classification head are jointly trained on the fused token, $\mathcal{L}_{\text{ft}} = \text{BCE}\big(y,\, \sigma\big(\text{Head}(\mathbf{z}^{\text{fused}}_{\texttt{[CLS]}})\big)\big)$, where $y \in \{0, 1\}$ is the experimentally derived immunogenicity label.

For the cancer neoepitope task, the fine-tuning loss augments $\mathcal{L}_{\text{BCE}}$ with a cancer-specific contrastive term $\mathcal{L}_{\text{CW}} = \mathcal{L}_{\text{sim}} + \mathcal{L}_{\text{ind}}$, yielding $\mathcal{L}^{\text{CW}}_{\text{ft}} = \mathcal{L}_{\text{BCE}} + \lambda_{\text{CW}} \cdot \mathcal{L}_{\text{CW}}$. This term is applied to cancer neoepitope–wildtype pairs, encouraging the model to separate immunogenic neoepitopes from their non-immunogenic wildtype counterparts in the fused representation space (details on Appendix B).

The pretrained encoders, the fusion module, and the head are all trained at the same learning rate; ESM-2 is frozen throughout. Following the evaluation setup of Givechian et al. (Givechian et al., 2025), we use peptides of length 8–11 and mirror ImmunoStruct's data splits, drawing from IEDB infectious-disease epitopes and CEDAR cancer neoepitopes.

3 Experiments

This section describes the data, training schedule, and compute used to train **ImmunoFoundation**. The downstream evaluation protocol mirrors (Givechian et al., 2025) so that gains are attributable to the pretraining curriculum rather than to differences in benchmark construction.

Pretraining data and schedule

Stage 1 (monomers). From AlphaFoldDB (Varadi & Velankar, 2023) (human proteome and Swiss-Prot), we retain chains of length 50–1024 with mean pLDDT $\geq$ 70 to remove low-confidence predictions, yielding ∼580K structures. Stage 1 is trained for 5 epochs.

Stage 2 (complexes). From AlphaFoldDB-complex (Han et al., 2026) release of 1.8M, we retain complexes with mean interface pLDDT $\geq$ 70. Stage 2 initializes from Stage 1 weights and is trained for 7 epochs.

Setup. Both stages use AdamW (lr $= 1 \times 10^{-4}$, weight decay 1×10^{-6}), per-device batch 32 with gradient accumulation 4 over 4 GPUs via DDP (effective batch 512); full hyperparameters in Appendix C.

Fine-tuning data. Following ImmunoStruct, downstream tasks predict immunogenicity on infectious-disease epitopes (Vita et al., 2019) and CEDAR neoepitopes (Koşaloğlu-Yalçın et al., 2023), detailed in Ap-

pendix A.

4 Results

This section reports peptide–MHC immunogenicity prediction performance against the ImmunoStruct evaluation protocol (Givechian et al., 2025) and ablations over the pretraining curriculum.

Immunogenicity prediction. We evaluate on the IEDB infectious-disease ($n = 3{,}500$) and CEDAR cancer-neoepitope ($n = 208$) test sets of (Givechian et al., 2025), mirroring their splits and reporting ROC AUC (AUROC) and mean PPV_n. Following (Albert et al., 2022), we treat mean PPV_n as the primary metric for this task, as it directly reflects the precision of top-ranked predictions that matters most in practical immunogenicity screening. All results are averaged over five random seeds; we report sample standard deviation for AUROC and standard error of the mean for mean PPV_n (in parentheses in Tables 1 and 2). Baselines are PRIME-2.1 (Schmidt et al., 2021), NetMHCpan-4.1 (Hoof et al., 2009), MHCnuggets (Shao et al., 2020), MHCflurry-2.0 (O'Donnell et al., 2020), DeepNeo (Kim et al., 2023), BigMHC (EL, IM, and retrained variants) (Albert et al., 2022), NeoaPred (CEDAR only) (Jiang et al., 2024), and the multimodal ImmunoStruct itself. Tables 1 and 2 summarize performance on both cohorts.

Table 1. Immunogenicity prediction performance on the IEDB infectious-disease test set ($n = 3{,}500$). Best value per metric is bolded.

Method	AUROC	PPV_n
PRIME-2.1	0.538 ± 0.012	0.207 ± 0.013
NetMHCpan	0.537 ± 0.027	0.220 ± 0.008
MHCnuggets	0.546 ± 0.023	0.233 ± 0.014
MHCflurry	0.577 ± 0.021	0.273 ± 0.013
DeepNeo	0.767 ± 0.032	0.411 ± 0.008
BigMHC-EL	0.588 ± 0.018	0.290 ± 0.011
BigMHC-IM	0.684 ± 0.028	0.333 ± 0.015
$\text{BigMHC}_{\text{retrained}}$	0.793 ± 0.013	0.445 ± 0.004
ESM2	0.764 ± 0.014	0.421 ± 0.035
ImmunoStruct	0.882 ± 0.005	0.514 ± 0.009
ImmunoFoundation (ours)	**0.886** ± 0.007	**0.608** ± 0.008

Table 2. Immunogenicity prediction performance on the CEDAR cancer-neoepitope test set ($n = 280$). Best value per metric is bolded.

Method	AUROC	PPV_n
PRIME-2.1	0.645 ± 0.026	0.295 ± 0.013
NetMHCpan	0.559 ± 0.049	0.211 ± 0.020
MHCnuggets	0.530 ± 0.014	0.175 ± 0.010
MHCflurry	0.658 ± 0.015	0.329 ± 0.013
DeepNeo	0.535 ± 0.016	0.222 ± 0.031
BigMHC-EL	0.632 ± 0.034	0.245 ± 0.019
BigMHC-IM	0.771 ± 0.040	0.357 ± 0.016
$\text{BigMHC}_{\text{retrained}}$	0.682 ± 0.012	0.325 ± 0.013
NeoaPred	0.556 ± 0.016	0.292 ± 0.036
ImmunoStruct	0.771 ± 0.024	0.365 ± 0.057
ImmunoFoundation (ours)	0.733 ± 0.047	**0.409** ± 0.035

Ablations. To assess the contribution of each pre-training stage, we performed ablations in which individual stages are removed from the curriculum (Table 4). Each stage contributes meaningfully to the final performance, with no single stage being redundant. As we can see in the Table, adding training Phase 2 is beneficial for downstream immunogenecity prediction.

5 Conclusion

We present **ImmunoFoundation**, a multimodal foundation backbone for protein complexes specialized to pMHC immunogenicity. The central design choice is curricular: rather than imposing the multi-chain inductive bias up front, the model is pretrained first on AlphaFoldDB monomers and then on the proteome-scale AlphaFoldDB-complex set, before a cross-modal fusion module is fine-tuned on labeled pMHCs. On the cancer neoepitope and infectious-disease cohorts of the ImmunoStruct evaluation, **ImmunoFoundation** improves over the previous best multimodal class I baseline. Ablations show that pretraining stages effectively contribute and order of staging matters: complex pretraining without prior monomer pretraining does not recover the full gain. The result supports a more general claim implicit in the architecture: immunogenicity is a downstream readout of the TCR-pMHC protein complex, and a foundation backbone trained on the unlabeled structural corpus is a productive starting point for the family of protein complex-level tasks that have historically been bottlenecked by the scarcity of labeled data.

Future Work. 1) Stage 3 pretraining on pMHC structures. As discussed in Section 2.2, a third curriculum stage on a pMHC-specific structural corpus—combining experimental structures from the PDB with predicted pMHCs—is the natural lever for closing this distribution gap on AlphaFoldDB-complex and the most direct route to further gains on immunogenicity. **2) TCR recognition and antigen-conditioned design.** A natural extension is to add the TCR into the input complex as additional chains for TCRA and TCRB, followed by fine-tuning on TCR-pMHC binding labels. Coupled with a generative head over TCR CDR3 sequences conditioned on the fused TCR-pMHC representation, the same backbone supports antigen-conditioned TCR design, providing a route from immunogenicity prediction to TCR engineering for cancer immunotherapy and TCR-T cell therapy. **3) Other directions.** The architecture is class-agnostic, so class II MHC presentation is a direct extension that exploits shared MHC representations. The pretrained backbone is also a natural starting point for other complex-level tasks bottlenecked by labeled-data scarcity, such as protein–protein affinity and antibody–antigen binding.

References

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630 (8016):493–500, 2024.

Albert, B. A., Yang, Y., Shao, X. M., Singh, D., Smith, K. N., Anagnostou, V., and Karchin, R. Deep neural networks predict mhc-i epitope presentation and transfer learn neoepitope immunogenicity. *bioRxiv*, pp. 2022–08, 2022.

Bengio, Y., Louradour, J., Collobert, R., and Weston, J. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pp. 41–48, 2009.

Dwivedi, V. P. and Bresson, X. A generalization of transformer networks to graphs. *arXiv preprint arXiv:2012.09699*, 2020.

Evans, R., O'neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., et al. Protein complex prediction with alphafold-multimer. *biorxiv*, pp. 2021–10, 2021.

Givechian, K. B., Rocha, J. F., Liu, C., Yang, E., Tyagi, S., Greene, K., Ying, R., Caron, E., Iwasaki, A., and Krishnaswamy, S. Immunostruct enables multimodal deep learning for immunogenicity prediction. *Nature Machine Intelligence*, pp. 1–14, 2025.

Han, Y., Tsenkov, M. I., Venanzi, N. A., Bertoni, D., Cha, S., Chacon, A., Dietrich, N., Fomitchev, B., Goldtzvik, Y., Hsu, D., et al. Alphafold database expands to proteome-scale quaternary structures. *bioRxiv*, pp. 2026–03, 2026.

Hoof, I., Peters, B., Sidney, J., Pedersen, L. E., Sette, A., Lund, O., Buus, S., and Nielsen, M. NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics*, 61(1):1–13, 2009.

Jiang, D., Xi, B., Tan, W., Chen, Z., Wei, J., Hu, M., Lu, X., Chen, D., Cai, H., and Du, H. Neoapred: a deep-learning framework for predicting immunogenic neoantigen based on surface and structural features of peptide–human leukocyte antigen complexes. *Bioinformatics*, 40 (9):btae547, 2024.

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

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

Koşaloğlu-Yalçın, Z., Blazeska, N., Vita, R., Carter, H., Nielsen, M., Schoenberger, S., Sette, A., and Peters, B. The cancer epitope database and analysis resource (cedar). *Nucleic acids research*, 51(D1):D845–D852, 2023.

Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637): 1123–1130, 2023.

Montemurro, A., Schuster, V., Povlsen, H. R., Bentzen, A. K., Jurtz, V., Chronister, W. D., Crinklaw, A., Hadrup, S. R., Winther, O., Peters, B., et al. Nettcr-2.0 enables accurate prediction of tcr-peptide binding by using paired tcrα and β sequence data. *Communications biology*, 4(1): 1060, 2021.

O'Donnell, T. J., Rubinsteyn, A., and Laserson, U. MHCflurry 2.0: improved pan-allele prediction of MHC class I-presented peptides by incorporating antigen processing. *Cell Systems*, 11(1):42–48, 2020.

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

Schmidt, J., Smith, A. R., Magnin, M., Racle, J., Devlin, J. R., Bobisse, S., Cesbron, J., Bonnet, V., Carmona, S. J., Huber, F., et al. Prediction of neo-epitope immunogenicity reveals tcr recognition determinants and provides insight into immunoediting. *Cell Reports Medicine*, 2(2), 2021.

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

Springer, I., Besser, H., Tickotsky-Moskovitz, N., Dvorkin, S., and Louzoun, Y. Prediction of specific tcr-peptide binding from large dictionaries of tcr-peptide pairs. *Frontiers in immunology*, 11:1803, 2020.

Varadi, M. and Velankar, S. The impact of alphafold protein structure database on the fields of life sciences. *Proteomics*, 23(17):2200128, 2023.

Vita, R., Mahajan, S., Overton, J. A., Dhanda, S. K., Martini, S., Cantrell, J. R., Wheeler, D. K., Sette, A., and Peters, B. The immune epitope database (iedb): 2018 update. *Nucleic acids research*, 47(D1):D339–D343, 2019.

Wells, D. K., van Buuren, M. M., Dang, K. K., Hubbard-Lucey, V. M., et al. Key parameters of tumor epitope immunogenicity revealed through a consortium approach improve neoantigen prediction. *Cell*, 183(3):818–834, 2020.

Ying, C., Cai, T., Luo, S., Zheng, S., Ke, G., He, D., Shen, Y., and Liu, T.-Y. Do transformers really perform badly for graph representation? *Advances in neural information processing systems*, 34:28877–28888, 2021.

A Dataset details

IEDB infectious-disease peptide–MHCs. The primary fine-tuning corpus is approximately $[N_{\text{iedb}}]$ peptide–MHC pairs from infectious-disease assays in IEDB (Vita et al., 2019), with allele-stratified splits to limit allele leakage between train and test.

CEDAR cancer neoepitopes. Cancer-restricted neoepitopes from CEDAR (Koşaloğlu-Yalçın et al., 2023) ($[N_{\text{cedar}}]$ pairs), with sequence-similarity filtering against IEDB to suppress near-duplicate contamination.

SARS-CoV-2 ELISpot validation. An external set of $[N_{\text{covid}}]$ SARS-CoV-2 epitopes characterized by ELISpot, used as a held-out probe of out-of-distribution generalization.

Cancer-patient survival cohort. A retrospective cohort of $[N_{\text{cohort}}]$ patients with neoepitope predictions linked to overall survival; per-patient immunogenicity scores are aggregated and used as a covariate in survival analysis.

B Contrastive Loss

For the cancer neoepitope task, the fine-tuning loss augments the binary cross-entropy term $\mathcal{L}_{\text{BCE}}$ with a cancer-specific contrastive term $\mathcal{L}_{\text{CW}}$ applied to neoepitope–wildtype pairs. The contrastive term encourages the model to separate immunogenic neoepitopes from their non-immunogenic wildtype counterparts in the fused representation space. The full objective decomposes as

$$\mathcal{L}_{\text{ft}}^{\text{CW}} = \mathcal{L}_{\text{BCE}} + \lambda_{\text{CW}} \cdot \mathcal{L}_{\text{CW}}, \tag{3}$$

$$\mathcal{L}_{\text{CW}} = \mathcal{L}_{\text{similarity}} + \mathcal{L}_{\text{independence}}, \tag{4}$$

$$\mathcal{L}_{\text{similarity}}(Z_{\text{C}}, Z_{\text{W}}) = \frac{1}{N}\left(\sum_i \left((Z_{\text{C}} Z_{\text{W}}^\top)_{ii} - \mathbf{1}_i^{\text{C}}\right)^2 + \lambda_{\text{off-diag}} \sum_i \sum_{j \neq i} \left((Z_{\text{C}} Z_{\text{W}}^\top)_{ij}\right)^2\right), \tag{5}$$

$$\mathcal{L}_{\text{independence}}(Z_{\text{C}}, Z_{\text{W}}) = \frac{1}{D}\left(\sum_i \left((Z_{\text{C}}^\top Z_{\text{W}})_{ii} - 1\right)^2 + \lambda_{\text{off-diag}} \sum_i \sum_{j \neq i} \left((Z_{\text{C}}^\top Z_{\text{W}})_{ij}\right)^2\right), \tag{6}$$

where $Z_{\text{C}}, Z_{\text{W}} \in \mathbb{R}^{N \times D}$ are the batched fused `[CLS]` representations of N paired cancer and wildtype peptides, $\mathbf{1}_i^{\text{C}} \in \{0, 1\}$ indicates whether the i-th cancer peptide is immunogenic, and $\lambda_{\text{CW}}, \lambda_{\text{off-diag}}$ are weighting hyperparameters.

Intuitively, $\mathcal{L}_{\text{similarity}}$ operates on the $N \times N$ sample-wise cross-correlation matrix: diagonal entries are pulled toward $\mathbf{1}_i^{\text{C}}$, so immunogenic cancer peptides are pushed apart from their wildtype counterparts while non-immunogenic ones are kept close, and off-diagonal entries are decorrelated. Conversely, $\mathcal{L}_{\text{independence}}$ operates on the $D \times D$ feature-wise cross-correlation matrix, encouraging diagonal entries to be unit-correlated and off-diagonal entries to be decorrelated, which discourages redundant feature dimensions.

C Hyperparameters

Table 3 lists the full set of architecture and pretraining hyperparameters.

D Ablations

Table 4 reports the effect of removing the Phase-1 or Phase-2 pre-training stage, evaluated on the held-out immunogenicity benchmark. Removing Phase-2 from the Phase-1 backbone results in a consistent drop across both AUROC and PPV@n, indicating that Phase-2 pre-training contributes non-trivially beyond what is captured by the Phase-1 backbone alone. Only training on complexes (Phase-2 only) improves over Phase-1 performance. However, individual phases are not able to outperform the model that was trained on both Phase-1 and Phase-2.

Table 3. ImmunoFoundation architecture and pretraining hyperparameters.

Sequence track	
Backbone	ESM-2 (`esm2_t33_650M_UR50D`)
Backbone state	frozen
Representation layer	33 (final)
Backbone embedding dim (d_{ESM})	1280
Adaptor layers (g_θ)	10
Adaptor attention heads	8
Adaptor feed-forward dim	256
Output dim (d)	32
Structure track	
Architecture	Graph Transformer
Layers (N_s)	4
Attention heads	8
Feed-forward dim	256
Output dim (d)	32
Graph type	k-nearest neighbor (over C_α)
k	15
Pretraining objectives	
λ_{MLM}	1.0
λ_{MAE}	1.0
MLM mask rate	0.15
MAE masking criterion	residues with < 12 neighbors within 8 Å (surface)
MAE masking scope	applied identically in both stages
Masked coordinate fill	zero
Optimization	
Optimizer	AdamW
Learning rate	1×10^{-4}
Weight decay	1×10^{-6}
Batch size (per device)	32
Gradient accumulation steps	4
Max epochs	200
Devices	4 GPUs (DDP)
Train / validation split	0.9 / 0.1

Table 4. Ablation of pre-training phases. Mean ± standard deviation over 5 seeds on the IEDB test set.

Metric	Phase-1 & 2	Phase-1 only	Phase-2 only
AUROC	**0.886** ± 0.007	0.855 ± 0.002	0.864 ± 0.004
PPVn	**0.608** ± 0.008	0.538 ± 0.018	0.542 ± 0.002