TCR-ECHO: Evolutionary Cross-attention with Physicochemical Bias for Hierarchical TCR-Peptide Binding Prediction

Tapanmitra Ravi[1] **Irene M. Ong**[1,2,3,4]

[1]Department of Biostatistics & Medical Informatics, University of Wisconsin-Madison
[2]Department of Obstetrics and Gynecology, University of Wisconsin-Madison
[3]Center for Human Genomics and Precision Medicine, University of Wisconsin-Madison
[4]Carbone Comprehensive Cancer Center, University of Wisconsin-Madison

Abstract

Accurate prediction of T-cell receptor (TCR)-peptide binding specificity remains challenging due to vast immune receptor diversity and complex molecular recognition principles. We present a deep learning framework integrating evolutionary protein representations with physicochemical binding principles for TCR-epitope interaction prediction. Our approach employs separate ESM2 encoders for TCR CDR3β and peptide sequences, capturing evolutionary patterns from millions of proteins. The architecture introduces two key innovations: physicochemically-informed cross-attention incorporating Atchley factor biases to model molecular complementarity, and hierarchical contrastive learning operating at residue and interaction levels to structure binding-specific representations.

The bidirectional cross-attention mechanism models mutual recognition between binding partners, while Atchley factor integration provides physicochemical context—hydrophobicity, polarity, and structural propensities—governing biochemical interactions beyond sequence patterns. Hierarchical contrastive learning progressively refines representations from sequence patterns to interaction compatibility, creating peptide-specific clusters in TCR embedding space.

Comprehensive evaluation demonstrates superior performance across multiple scenarios. Systematic ablation studies confirm each component's importance, particularly for novel peptide generalization. The learned representations show clear biological interpretability, with TCRs binding identical peptides clustering in embedding space.This framework establishes foundations for computational immunology tools by integrating evolutionary information with molecular binding principles, which are informative for understanding TCR-peptide binding.

1 Introduction

T cells, which are part of the adaptive immune system, specifically recognize foreign antigens through recognition of tumor, bacterial or viral peptides presented by major histocompatibility complex class I (MHC) on affected cells, through its receptor (T cell receptor or TCR), to selectively destroy them [1, 2].

Experimental methods, such as peptide-MHC (pMHC) binding affinity assays and MHC-associated eluted ligand mass spectrometry, have been used to study these interactions [5, 6]. However, these techniques are costly, and labor-intensive. To address these limitations, various computational methods have been developed to predict MHC peptide binding and TCR-epitope recognition [3, 4, 7]. These methods range from simple motif-based approaches to more complex machine learning models, such as artificial neural networks, hidden Markov models, and regression models [8, 9, 10, 11, 12]. Recent advances in deep learning have led to significant improvements in protein-related tasks, such as protein structure prediction and protein-protein interaction prediction [13, 14]. Transformer-based models, in particular, have shown remarkable success in natural language processing and have been adapted for protein sequence analysis [15, 16, 17]. These models, pre-trained on large datasets of protein sequences, can capture complex patterns and relationships within the data, making them valuable for transfer learning in various downstream tasks [18, 19]. For example, Evolutionary Scale Modeling (ESM) [21] leverages evolutionary information in multiple sequence

alignments to predict protein structure, demonstrating direct inference of full atomic-level protein structure from primary sequence to create a protein language model.

Recently, models like PanPep [25] and VitTCR [23, 24] have shown promising results in predicting TCR-epitope interactions, demonstrating improvements over traditional methods. PanPep employs meta-learning approaches to handle novel epitopes, while VitTCR utilizes transformer architectures for sequence analysis. Additionally, EPACT[43] and ImmuneClip [41] have demonstrated the utility of contrastive learning for TCR-pMHC binding prediction using full-length sequences, with single-level contrastive objectives. While these models show promise, they primarily focus on sequence patterns without explicitly modeling the physicochemical basis of molecular recognition or adequately simulating the mutual interaction process between TCR and peptide during binding.

In this study, we present TCR-ECHO (Evolutionary Cross-attention with Hierarchical and Physicochemical Optimization), a novel deep learning framework that addresses the fundamental challenge of modeling TCR-peptide recognition. Our approach focuses on the CDR3β region, which exhibits the highest variability among TCR regions and serves as the primary determinant of epitope specificity [36, 37]. TCR-ECHO introduces physicochemically-informed cross-attention that explicitly models the mutual recognition process between binding partners—a critical aspect that existing methods fail to capture adequately. Drawing inspiration from multi-level contrastive approaches like MucLiPred [44], we implement a hierarchical contrastive learning framework operating at both residue and interaction levels, though our mechanisms are specifically designed for TCR-peptide binding. By combining evolutionary representations with molecular interaction principles, TCR-ECHO aims to computationally model the biological reality of immune recognition, providing both accurate predictions and interpretable insights into the features driving TCR-peptide specificity.

2 Methods

2.1 Architecture Overview

We developed a deep learning framework for predicting TCR-peptide binding that addresses the fundamental challenge of modeling molecular recognition between immune receptors and their targets. Our approach combines modern protein language models with domain-specific biological knowledge to capture both the sequence patterns and physicochemical interactions that govern binding specificity.

As illustrated in **Figure 1**, the model processes TCR CDR3β sequences and peptide sequences through a carefully designed pipeline. We focus on CDR3β regions as they represent the most variable portion of the TCR and contribute most significantly to peptide recognition. This choice also aligns with data availability as many TCR sequencing studies capture only the β chain. The architecture consists of five key components: pre-trained protein encoders, residue-level contrastive learning, physicochemically-informed cross-attention, interaction-level contrastive learning, and final classification.

2.2 Initial Sequence Encoding

For encoding protein sequences, we employ ESM (Evolutionary Scale Modeling) [21], a transformer-based protein language model pre-trained on millions of protein sequences from across the tree of life. ESM learns evolutionary patterns, structural motifs, and functional relationships directly from sequence data without requiring structural information. Its pre-training objective—masked amino acid prediction—enables it to capture both local and long-range dependencies in protein sequences, making it particularly suitable for understanding binding-competent sequence features.

We use separate ESM encoders for TCR and peptide sequences, recognizing that these molecules have evolved under different selective pressures and exhibit distinct structural properties. To adapt these general protein models to our specific binding prediction task, we employ Low-Rank Adaptation (LoRA) [45] on the top 5 transformer layers. This parameter-efficient fine-tuning strategy allows us to specialize the representations for TCR-peptide interactions

while preserving the valuable evolutionary knowledge captured during pre-training. During training, we randomly mask 10% of amino acids in both sequences. This masking strategy serves two purposes: it prevents the model from memorizing specific sequence patterns and encourages learning of more generalizable binding principles that remain robust to sequence variations.

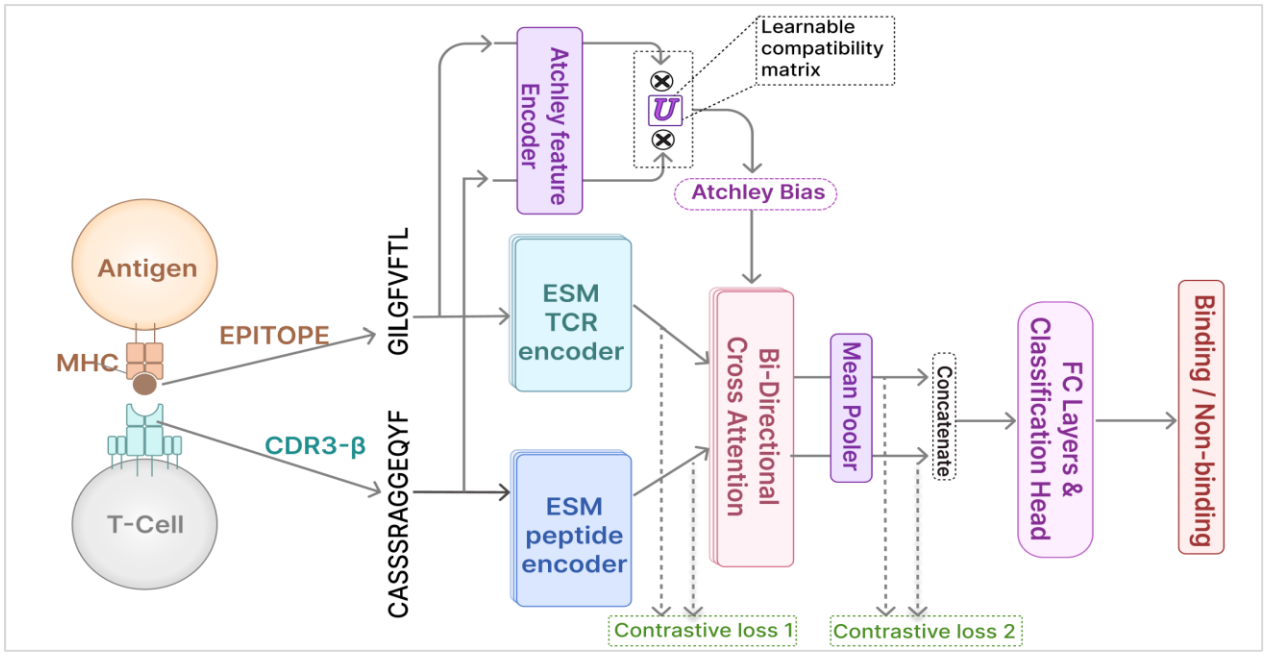

Figure 1: Shows the architectural overview of TCR-ECHO *illustrating the processing pipeline from TCR peptide interaction to binding prediction. The input sequences are encoded bi-modally by ESM and also Atchley factors. The specialized cross attention uses the Atchley factor as a bias to guide its interactions. Two levels of contrastive losses further aid the model in the final binding prediction.*

2.3 Hierarchical Contrastive Learning Framework

A key innovation in our approach is the use of contrastive learning at two distinct levels of representation. This hierarchical strategy as shown in Hierarchical Contrastive Learning [46] progressively refines the model's understanding from sequence-level patterns to interaction-level compatibility.

2.3.1 Residue-Level Contrastive Learning

Following the initial encoding by ESM-2, but before any interaction modeling through cross-attention, we apply a contrastive loss directly on the encoder outputs. At this stage, the TCR and peptide representations exist as sequence-level embeddings of dimension [L, H], where L represents the sequence length (for peptides, typically 9-15 residues; for TCRs, typically 20-30 residues for CDR3 regions), and H is the hidden dimension. While these embeddings have captured rich evolutionary and structural information from pre-training, they have not yet been adapted for the specific task of binding prediction.

The key insight motivating this residue-level loss is that it provides a complementary view of sequence similarity distinct from global pooled representations. Rather than comparing averaged embeddings, we compute similarity through fine-grained residue-to-residue comparisons. The idea being to pre-structure the embedding space such that TCRs binding to the same peptide are pulled together while those binding to different peptides are pushed apart. This creates peptide-specific clusters in the TCR embedding space, effectively teaching the encoders to recognize sequence patterns associated with specificity for peptides. The residue-level contrastive loss is computed as:

$$\mathcal{L}enc = -\frac{1}{B}\sum i = 1^{B} \log \frac{\exp(s_{ii}/\tau)}{\sum_{j=1}^{B} \exp(s_{ij}/\tau)}$$

where s_{ij} represents the averaged similarity between all residue pairs from **peptide *i*** and **TCR *j*,** B is the batch size and τ is the temperature hyperparameter. This formulation ensures that the entire sequence contributes to the learning signal, not just individual positions.

Inspired by SimCLR [47], our negative sampling strategy is carefully designed to create meaningful contrasts. For each positive TCR-peptide pair, we construct negatives by pairing the peptide with TCRs that bind to different peptides, but only when those peptides have a minimum Levenshtein distance from each other. This distance constraint ensures that negative pairs represent genuinely different binding specificities rather than minor sequence variants, leading to more robust learning. This is shown as Contrastive loss 1 in **Figure 1**.

2.3.2 Interaction-Level Representations and Contrastive Learning

After cross-attention, we obtain interaction-aware representations for both TCR and peptide sequences. These are pooled using mean pooling across sequence positions, providing a robust summary that captures information from all residues rather than relying on a single position.

At this stage, we apply a second loss based on InfoNCE [48], shown as Contrastive loss 2 in **Figure 1**, that operates on the interaction-enriched representations:

$$\mathcal{L}int = -\frac{1}{B}\sum i = 1^{B} \log \frac{\exp(\langle \mathbf{z}_i^{pep}, \mathbf{z}^{tcr} i\rangle/\tau)}{\sum j=1^{B} \exp(\langle \mathbf{z}_i^{pep}, \mathbf{z}_j^{tcr}\rangle/\tau)}$$

This loss encourages binding pairs to have similar interaction patterns after accounting for their molecular complementarity. Unlike the residue-level loss that operates on raw sequence encodings, this loss leverages the rich interaction information captured by cross-attention, which is particularly effective for learning binding compatibility.

2.4 Physicochemically-Informed Cross-Attention Mechanism

The heart of our model is a novel cross-attention mechanism that explicitly incorporates physicochemical properties into the attention computation. This design choice stems from the fundamental principle that molecular recognition depends not just on sequence patterns but on the complementary physicochemical properties of interacting residues.

We encode each amino acid using five Atchley factors: hydrophobicity, steric bulk, codon composition, electrostatic charge, and helix/structure propensity [22]. These factors, derived from multivariate statistical analysis of amino acid properties, capture the key physicochemical dimensions relevant to protein interactions. Before computing attention, we align these Atchley encodings with the sequence lengths by truncating to match the actual amino acid positions, excluding special tokens.

Our attention mechanism computes two parallel attention patterns. The standard attention captures learned sequence relationships through the scaled dot-product:

$$\text{Attention}_{\text{sequence}} = \text{softmax}\left(\frac{Q_{\text{tcr}} K_{\text{peptide}}^T}{\sqrt{d_h}}\right)$$

while a separate physicochemical attention models compatibility based on Atchley factors through learned bilinear [52] forms:

$$\text{Score}_{\text{atchley}} = \text{softmax}(A_{\text{tcr}} U A_{\text{peptide}}^T)$$

where $A_{tcr}, $A_{peptide}$ are the Atchley factor representations and $\boldsymbol{U}$ is a learned **5 X 5** bilinear form matrix that captures interactions between physicochemical properties. These two attention mechanisms are combined through a learnable mixing parameter $\boldsymbol{\rho}$:

$$\rho = (\tanh(\lambda) + 1)/2$$

$$\text{Attention} = (1 - \rho) \cdot \text{Attention}_{\text{sequence}} + \rho \cdot \text{Score}_{\text{atchley}}$$

The combined attention weights are then applied to the peptide values to produce the final output . This design allows the model to discover the optimal balance between sequence-learned patterns and physicochemical compatibility. The

mixing parameter λ is initialized to zero, giving equal initial weighting to both attention mechanisms, which the model then adapts during training.

We apply this attention bidirectionally [40, 20]—TCR attending to peptide and peptide attending to TCR, recognizing that molecular binding is inherently a mutual interaction where both partners influence each other.

2.5 Classification and Training Objective

The final classification layer receives the concatenated TCR and peptide representations after interaction modeling. We employ a two-layer fully connected network with dropout regularization, reflecting the complexity of the binding decision while preventing overfitting. For the classification loss, we use focal loss to address potential class imbalance between binding and non-binding pairs. This loss function naturally focuses training on hard-to-classify examples, improving the model's ability to handle challenging cases near the decision boundary. The complete training objective combines all three losses with carefully tuned weights:

$$\mathcal{L}total = \mathcal{L}focal + w1 \cdot \mathcal{L}enc + w2 \cdot \mathcal{L}int$$

These weights, determined through systematic hyperparameter search, reflect the relative importance of each objective. The highest weight on interaction-level contrastive loss (2.0) emphasizes learning from interaction patterns, while the moderate weight on encoder-level loss (0.8) provides beneficial regularization without dominating the training dynamics.

2.6 Design Rationale

Our approach introduces several novel elements to TCR-peptide binding prediction. The hierarchical contrastive learning framework progressively refines representations from sequence patterns to interaction compatibility, mirroring the biological reality that binding depends on both individual sequence features and their interactions. The residue-level loss computes similarity through averaged pairwise residue interactions, while the interaction-level loss operates on globally pooled features, providing different geometric constraints on the embedding space. While both losses aim to cluster TCRs by their binding peptides, they observe this objective through different lenses—local residue compatibilities versus global sequence similarities. This creates position-aware learning signals where residues that contribute more to the overall similarity receive proportionally stronger gradient updates.

The physicochemically-informed attention mechanism explicitly models the molecular basis of recognition, going beyond pure sequence-based learning to incorporate domain knowledge about amino acid properties. The bidirectional nature of our cross-attention reflects the mutual influence between binding partners, a key aspect often overlooked in simpler architectures. By combining modern deep learning techniques with biological domain knowledge, our model achieves a balance between learning capacity and interpretability, providing not just predictions but insights into the features driving TCR-peptide recognition.

3 Results

3.1 Robust Performance on Seen & Unseen Peptides

To evaluate our model's performance, we conducted comprehensive comparisons with recent state-of-the-art methods as well as established benchmarks. All evaluated methods tackle TCR-epitope specificity prediction using solely the CDR3β region and peptide sequences presented by MHC molecules.

Recent state-of-the-art methods include PanPep, which employs disentanglement learning combined with meta-learning and neural Turing machines to achieve robust performance on both seen and unseen peptides, and VitTCR, which leverages vision transformer architecture by encoding TCR-peptide interactions as AtchleyMaps using Atchley factors, constructing five-channel contact maps between TCR and peptide sequences. **Established benchmarks** comprised pMTnet [27], which utilizes transfer learning to harness TCR sequence information and pMHC binding knowledge, DLpTCR [49], an ensemble deep learning framework integrating three models for immunogenic peptide recognition, and ERGO2 [26], an updated version incorporating LSTM or autoencoder encodings for both CDR3s and

epitopes with additional features including MHC information. We trained our model on the identical training set used by PanPep, combining their majority training and meta-training datasets. Following PanPep's experimental design, we evaluated performance on their seen peptide (majority) test set containing peptides observed during training. Detailed information on datasets and model training can be found in the Appendix.

To ensure fair comparison, we utilized PanPep's published model and retrained VitTCR using their official codebase on the same training data. Our model demonstrated superior performance on seen peptides, achieving overall AUROC and AUPRC scores exceeding 0.8 (**Figure 2A**). The statistical significance of these improvements was confirmed with p-values of 4.57e-02 and 1.87e-01 for AUROC comparisons, and 3.13e-02 and 4.36e-02 for AUPRC comparisons (**Figure 2A**).

For zero-shot evaluation on unseen peptides, we additionally compared against the established methods pMTnet, DLpTCR, and ERGO2. We utilized test sets carefully curated by removing peptides present in the training data to ensure unbiased evaluation. Our model maintained exceptional performance on novel peptides, achieving AUROC and AUPRC scores approaching 0.8, substantially outperforming competing methods (**Figure 2C**). While PanPep and VitTCR achieved scores around 0.7, the older benchmark methods scored closer to 0.5, approaching random performance levels. To investigate the relationship between prediction accuracy and peptide novelty, we analyzed individual peptide performance as a function of sequence similarity to training data. Using Levenshtein distance as a metric of sequence novelty, we computed the minimum distance from each test peptide to all training peptides (**Figure 2D**). Our model exhibited consistent performance across varying degrees of sequence novelty, with only modest performance degradation for highly divergent sequences (maximum Levenshtein distances). This suggests robust generalization capabilities beyond simple sequence memorization.

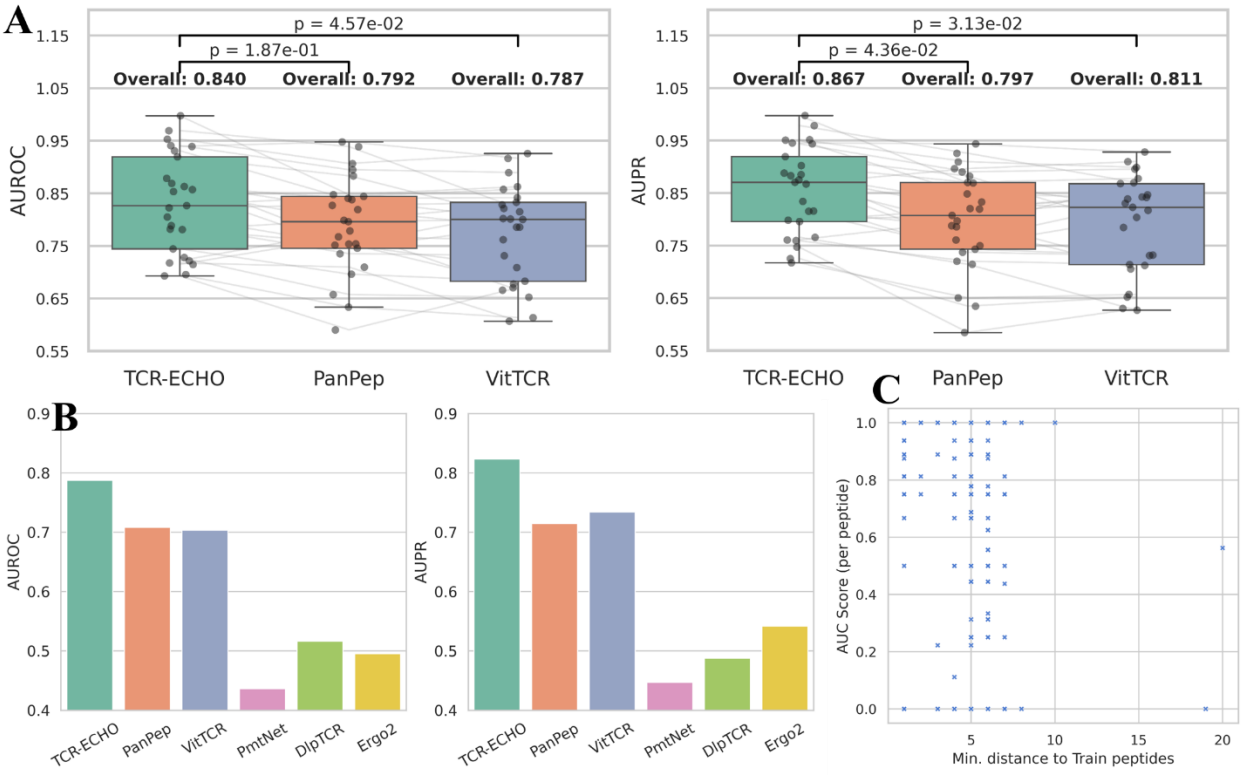

Figure 2: Performance on Panpep's seen and unseen test sets: (A) *Box plots showing AUROC and AUPRC distributions on the seen peptides across individual peptides for TCR-ECHO (ours), PanPep, and VitTCR. Gray lines connect peptides.* ***(B)*** *AUC and AUPRC scores of TCR-ECHO vs other methods on the unseen (zero shot) test setting.* ***(C)*** *Scatter plot showing individual unseen peptide AUC scores (y-axis) against Levenshtein distance to the most similar training peptide.*

3.2 Real-World COVID-19 TCR Recognition

To evaluate our model's real-world applicability in viral immune recognition, we tested its generalization capability on a completely independent COVID-19 dataset from the ImmuneCODE project [28, 29]. This dataset consists of TCR sequences from COVID-19 exposed or diagnosed subjects, along with virus-specific TCRs [30]. The dataset contains 1,100,136 samples with 541 unique epitopes and was preprocessed according to the setting described by PanPep (Methods section). While we have tried our best to follow PanPep's directions to curate the same dataset they used, it will not be identical (negative pairs will differ), which may lead to differences in the PanPep scores we report compared with those in the paper.

Our model significantly outperformed existing methods, achieving AUROC and AUPRC scores of 0.71. In comparison, other models showed considerably lower performance: PanPep (AUROC: 0.50, AUPRC: 0.51), ERGO2 (AUROC: 0.49, AUPRC: 0.49), and VitTCR (AUROC: 0.60, AUPRC: 0.62). The precision-recall curves shown in **Figure 3A** particularly highlight our model's superior ability to maintain precision across different recall levels. These results are especially noteworthy given the completely novel nature of the viral epitopes (unseen), since the model was trained on pre-pandemic human sequences. Our model's robust performance on this independent dataset demonstrates its potential as a reliable tool for identifying TCR-epitope interactions in emerging viral diseases, where rapid and accurate prediction of immune responses is crucial.

3.3 Binding Scores better represent Clonal Expansion

T cells that exhibit stronger binding affinity to their cognate peptide-MHC complexes are more likely to undergo clonal expansion, a fundamental principle underlying adaptive immune responses [50]. To validate whether our predicted binding scores can effectively capture this biological relationship, we evaluated our method using the benchmark single-cell datasets previously employed by PanPep, containing CD8+ T-cell profiles from two healthy donors across 44 distinct peptide-MHC complexes. Our method demonstrated substantially improved performance compared to both PanPep and UMI-based indicators across all evaluation metrics. For binary classification of clonal expansion status, our approach achieved AUC scores of 0.703 and 0.677 for donors 1 and 2, respectively, representing improvements of 0.138 and 0.094 over PanPep. Area under the precision-recall curve (AUPRC) showed even more pronounced improvements, with our method achieving 2.7-fold and 2.4-fold improvements over PanPep for the two donors (0.255 vs 0.093 and 0.286 vs 0.121, respectively).Most importantly, our method exhibited stronger correlations with actual clonal expansion ratios, achieving Spearman correlation coefficients of 0.373 and 0.270 compared to PanPep's 0.280 and 0.234, representing relative improvements of 33% and 15%. Both methods substantially outperformed UMI-based indicators (correlations of 0.046 and 0.016), confirming that sequence-based binding predictions are superior to UMI counts for predicting clonal expansion patterns. These results demonstrate that our method provides more accurate indicators for identifying clonally expanded T cells and better captures the underlying binding affinity relationships that drive T-cell clonal dynamics.

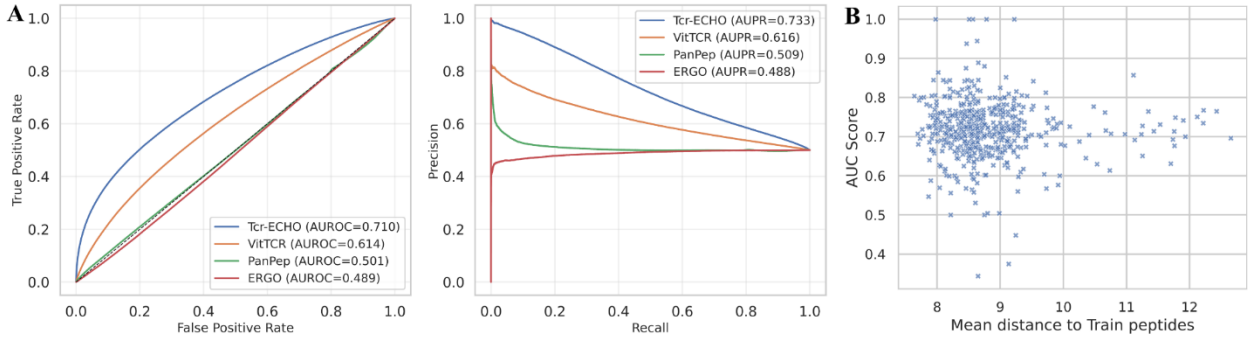

Figure 3: (A) *ROC and Precision-Recall curves comparing TCR-ECHO against VitTCR, PanPep, and ERGO2 on the independent COVID-19 dataset, showing ESM-ViT's superior performance.* ***(B)*** *Scatter plot showing individual COVID-19 peptide AUC scores (y-axis) against Mean Levenshtein distance to training peptides.*

3.4 Component contribution through Ablation

To validate our architectural design choices and understand the contribution of individual components, we conducted systematic ablation experiments using PanPep's dataset with enhanced negative sampling. We generated more robust negative samples by pairing peptides with TCRs that bind to other peptides, requiring a minimum Levenshtein distance of 3 between peptides to ensure meaningful contrasts. We evaluated five model variants against our full architecture (**Table 1**). The **no cross-attention** variant replaced our attention mechanism with simple mean pooling and concatenation, resulting in substantial performance degradation, particularly for unseen peptides (AUROC: 0.69→0.61). The **no Atchley features** variant performed standard bidirectional cross-attention without physicochemical bias, showing moderate performance loss on unseen peptides (AUROC: 0.69→0.65), but replacing them with **Blosum62 [53]** substitution matrix encodings largely recovers performance (AUROC: 0.67). Most critically, removing **contrastive learning losses** caused complete model collapse, demonstrating their fundamental importance for learning meaningful representations. The **no biased attention** variant concatenated Atchley factors to embeddings rather than using our biased attention mechanism, yielding inferior performance (unseen AUROC: 0.69→0.61). Finally, we tried **training the ESM model from scratch** with a random initialization; the large drop in scores for both settings shows the contribution of the rich evolutionary information gained from pretraining. The results reveal that while performance drops on seen peptides are relatively modest, the impact on unseen peptides is substantially larger, highlighting each component's role in generalization. Cross-attention mechanisms and physicochemically informed attention prove especially critical for handling novel peptides, suggesting they capture fundamental binding principles rather than dataset-specific patterns. The severe performance collapse without contrastive learning underscores its essential role in structuring the representation space for meaningful TCR-peptide similarity relationships.

Model variants	**Seen peptides**		**Unseen peptides**	
	AUROC	**AUPRC**	**AUROC**	**AUPRC**
Full model	**0.76**	**0.78**	**0.69**	**0.72**
No cross-attention	0.72	0.75	0.61	0.65
No Atchley features	0.72	0.75	0.65	0.69
No Atchley→Blosum62	0.73	0.76	0.67	0.71
No contrastive loss 1	0.74	0.75	0.59	0.56
No contrastive loss 2	0.68	0.65	0.52	0.55
No biased attention	0.71	0.73	0.61	0.64
No pre-trained ESM	0.58	0.57	0.56	0.58

Table 1: *Performance impact of ablating model components on seen and unseen peptide prediction.*

4 Discussion and Conclusion

Our model demonstrates robust performance across diverse evaluation scenarios, consistently outperforming both recent state-of-the-art methods and established benchmarks for TCR-peptide binding prediction. This success stems from an architecture designed to computationally model the biological reality of immune recognition interactions.

The bidirectional cross-attention mechanism captures the fundamental principle that molecular recognition involves mutual assessment between binding partners, with each residue contributing to overall interaction compatibility. This approach mirrors the physical process where TCR and peptide residues evaluate complementarity during binding. The integration of Atchley factor bias provides essential physicochemical context—hydrophobicity, polarity, and structural propensities—that governs real biochemical interactions beyond sequence patterns alone. Our hierarchical

contrastive learning framework proves particularly effective at structuring the representation space to distinguish binding from non-binding pairs.

The COVID-19 validation and T cell expansion analyses demonstrate practical utility for immunotherapy and vaccine development. The model's ability to predict immune responses to novel viral epitopes, despite training exclusively on pre-pandemic data, suggests it captures fundamental recognition principles rather than dataset-specific patterns. This capability could accelerate therapeutic development during emerging infectious disease outbreaks.

Interpretability analyses confirm that TCR-ECHO learns biologically grounded representations. Computational alanine scanning reveals that the model correctly identifies middle CDR3β segments as dominant contributors to binding predictions, while attention weights show substantial correlation with structural burial fractions from crystal complexes, validating that the model captures physically meaningful interaction patterns despite training exclusively on sequence data (**Appendix A**).

Several constraints limit our current approach. Relying solely on CDR3β and peptide sequences may inadequately capture the full complexity of TCR-pMHC recognition, particularly given the known contributions of CDR3α chains and MHC context. The moderate performance scores (AUROC ~0.71) on COVID-19 data and occasional failure on specific novel peptides highlight these limitations. The extreme diversity of immune receptor sequences, combined with limited training data relative to the theoretical sequence space, constrains generalization potential.

Future iterations could incorporate MHC sequences and paired TCRα information to better approximate the three-dimensional binding interface. The attention patterns learned by our model may provide valuable insights for improving computational structure prediction of TCR-pMHC complexes, particularly for the highly variable CDR3β regions where experimental structures remain scarce.

To conclude, we present a novel deep learning framework that successfully integrates evolutionary sequence information with physicochemical binding principles for TCR-peptide interaction prediction. The model's strong performance across multiple validation scenarios, combined with its biological interpretability, establishes a foundation for advanced computational immunology tools. Future developments incorporating additional molecular context promise to enhance both prediction accuracy and therapeutic applications in personalized immunotherapy design.

Acknowledgements

We acknowledge support from the University of Wisconsin-Madison Badger Challenge Award to I.M.O. Additionally, T.R. acknowledges Quaid Morris, Olga Lyudovyk, and Ilyes Baali from Memorial Sloan Kettering Cancer Center for their support and mentorship in the initial stages of this work.

References

[1] Janeway, C. A., Travers, P., Walport, M., & Shlomchik, M. J. (2001). Immunobiology: The Immune System in Health and Disease. 5th edition. New York: Garland Science.

[2] Maimela, N. R., Liu, S., & Zhang, Y. (2019). Fates of CD8+ T cells in Tumor Microenvironment. Computational and Structural Biotechnology Journal, 17, 1-13. https://doi.org/10.1016/j.csbj.2018.11.004

[3] Grebenkin, I., Voevodski, K., Basharov, M., Grachev, V., & Shaytan, A. K. (2020). Binding prediction for MHC class I and II peptide complexes using deep learning. Bioinformatics, 36(22-23), 5610-5616. https://doi.org/10.1093/bioinformatics/btaa1034

[4] O'Donnell, T. J., Rubinsteyn, A., & Laserson, U. (2020). MHCflurry 2.0: Improved Pan-Allele Prediction of MHC Class I-Presented Peptides by Incorporating Antigen Processing. Cell Systems, 11(1), 42-48.e7. https://doi.org/10.1016/j.cels.2020.06.010

[5] Caron, E., Espona, L., Kowalewski, D. J., Schuster, H., Ternette, N., Alpízar, A., Schittenhelm, R. B., Ramarathinam, S. H., Lindestam Arlehamn, C. S., Chiek Koh, C., Gillet, L. C., Rabsteyn, A., Navarro, P., Kim, S., Lam, H., Sturm, T., Marcilla, M., Sette, A., Campbell, D. S., … Aebersold, R. (2015). An open-source computational and data resource to analyze digital maps of immunopeptidomes. eLife, 4, e07661. https://doi.org/10.7554/eLife.07661

[6] Townsend, A., Elliott, T., Cerundolo, V., Foster, L., Barber, B., & Tse, A. (1990). Assembly of MHC class I molecules analyzed in vitro. Cell, 62(2), 285-295. https://doi.org/10.1016/0092-8674(90)90366-m

[7] Boehm, K. M., Bhinder, B., Raja, V. J., Dephoure, N., & Elemento, O. (2019). Predicting peptide presentation by major histocompatibility complex class I: an improved machine learning approach to the immunopeptidome. BMC Bioinformatics, 20(1), 7. https://doi.org/10.1186/s12859-018-2561-z

[8] Bui, H.-H., Sidney, J., Peters, B., Sathiamurthy, M., Sinichi, A., Purton, K.-A., Mothé, B. R., Chisari, F. V., Watkins, D. I., & Sette, A. (2005). Automated generation and evaluation of specific MHC binding predictive tools: ARB matrix applications. Immunogenetics, 57(5), 304-314. https://doi.org/10.1007/s00251-005-0798-y

[9] Doytchinova, I. A., & Flower, D. R. (2001). Toward the Quantitative Prediction of T-Cell Epitopes: CoMFA and CoMSIA Studies of Peptides with Affinity for the Class I MHC Molecule HLA-A*0201. Journal of Medicinal Chemistry, 44(22), 3572-3581. https://doi.org/10.1021/jm010021j

[10] Nielsen, M., Lundegaard, C., Worning, P., Lauemøller, S. L., Lamberth, K., Buus, S., Brunak, S., & Lund, O. (2003). Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. Protein Science, 12(5), 1007-1017. https://doi.org/10.1110/ps.0239403

[11] Parker, K. C., Bednarek, M. A., & Coligan, J. E. (1994). Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. The Journal of Immunology, 152(1), 163-175.

[12] Zhang, H., Lund, O., & Nielsen, M. (2006). The PickPocket method for predicting binding specificities for receptors based on receptor pocket similarities: application to MHC-peptide binding. Bioinformatics, 25(10), 1293-1299. https://doi.org/10.1093/bioinformatics/btp137

[13] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2020). High Accuracy Protein Structure Prediction Using Deep Learning. 14th Critical Assessment of Protein Structure Prediction (CASP14). https://doi.org/10.1101/2020.12.09.418855

[14] Lensink, M. F., Brysbaert, G., Mauri, T., Nadzirin, N., Velankar, S., Chaleil, R. A. G., Clarence, T., Bates, P. A., Kong, R., Liu, B., Xu, R., Zhang, L., Shi, H., Chang, S., Eisenstein, M., Karczynska, A., Czaplewski, C., Lubecka, E. A., Lipska, A. G., … Wodak, S. J. (2021). Prediction of protein assemblies, the next frontier: The CASP14-CAPRI experiment. Proteins: Structure, Function, and Bioinformatics, 89(12), 1800-1823. https://doi.org/10.1002/prot.26222

[15] Elnaggar, A., Heinzinger, M., Dallago, C., Rehawi, G., Wang, Y., Jones, L., Gibbs, T., Feher, T., Angerer, C., Steinegger, M., Bhowmik, D., & Rost, B. (2020). ProtTrans: Towards Cracking the Language of Life's Code Through Self-Supervised Deep Learning and High Performance Computing. ArXiv:2007.06225 [Cs, q-Bio]. http://arxiv.org/abs/2007.06225

[16] Rao, R., Bhattacharya, N., Thomas, N., Duan, Y., Chen, X., Canny, J., Abbeel, P., & Song, Y. S. (2019). Evaluating Protein Transfer Learning with TAPE. Advances in Neural Information Processing Systems, 32. https://proceedings.neurips.cc/paper/2019/hash/37f65c068b7723cd7809ee2d31d7861c-Abstract.html

[17] Rives, A., Goyal, S., Meier, J., Guo, D., Ott, M., Zitnick, C. L., Ma, J., & Fergus, R. (2019). Biological Structure and Function Emerge from Scaling Unsupervised Learning to 250 Million Protein Sequences. BioRxiv, 622803. https://doi.org/10.1101/622803

18. Rao, R., Meier, J., Sercu, T., Ovchinnikov, S., & Rives, A. (2020). Transformer protein language models are unsupervised structure learners. BioRxiv, 2020.12.15.422761. https://doi.org/10.1101/2020.12.15.422761

[19] Rao, R., Liu, J., Verkuil, R., Meier, J., Canny, J. F., Abbeel, P., Sercu, T., & Rives, A. (2021). MSA Transformer. BioRxiv, 2021.02.12.430858. https://doi.org/10.1101/2021.02.12.430858

[20] Gasser, J., Hamoudi, M. M., & Ong, C. S. (2021). ImmunoBERT - A Pretrained Language Model to Improve MHC-I Peptide-Protein Binding Prediction (pp. 82–88).

[21] Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y. and dos Santos Costa, A., 2023. Evolutionary-scale prediction of atomic-level protein structure with a language model. Science, 379(6637), pp.1123-1130.

[22] Atchley, W. R., Zhao, J., Fernandes, A. D., & Drüke, T. (2005). Solving the protein sequence metric problem. Proceedings of the National Academy of Sciences, 102(18), 6395-6400.

[23] Jiang, M., Yu, Z., & Lan, X. (2024). VitTCR: A deep learning method for peptide recognition prediction. iScience, 27(5), 109770.

[24] Moris, P., De Pauw, J., Postovskaya, A., Gielis, S., De Neuter, N., Bittremieux, W., Ogunjimi, B., Laukens, K., & Meysman, P. (2021). Current challenges for unseen-epitope TCR interaction prediction and a new perspective derived from image classification. Briefings in Bioinformatics, 22(4), bbaa318.

[25] Gao, Y., Gao, Y., Fan, Y., Zhu, C., Wei, Z., Zhou, C., Chuai, G., Chen, Q., Zhang, H. and Liu, Q., 2023. Pan-peptide meta learning for T-cell receptor–antigen binding recognition. Nature Machine Intelligence, 5(3), pp.236-249.

[26] Springer, I., Tickotsky, N. and Louzoun, Y., 2021. Contribution of T cell receptor alpha and beta CDR3, MHC typing, V and J genes to peptide binding prediction. Frontiers in immunology, 12, p.664514.

[27] Lu, T., Zhang, Z., Zhu, J., Wang, Y., Jiang, P., Xiao, X., Bernatchez, C., Heymach, J.V., Gibbons, D.L., Wang, J. and Xu, L., 2021. Deep learning-based prediction of the T cell receptor–antigen binding specificity. Nature machine intelligence, 3(10), pp.864-875.

[28] May, D.H., Rubin, B.E., Dalai, S.C., Patel, K., Shafiani, S., Elyanow, R., Noakes, M.T., Snyder, T.M. and Robins, H.S., 2021. Immunosequencing and epitope mapping reveal substantial preservation of the T cell immune response to Omicron generated by SARS-CoV-2 vaccines. MedRxiv, pp.2021-12.

[29] Snyder, T.M., Gittelman, R.M., Klinger, M., May, D.H., Osborne, E.J., Taniguchi, R., Zahid, H.J., Kaplan, I.M., Dines, J.N., Noakes, M.N. and Pandya, R., 2020. Magnitude and dynamics of the T-cell response to SARS-CoV-2 infection at both individual and population levels. MedRxiv, 2020.07.31.20165647.

[30] Nolan, S., Vignali, M., Klinger, M., Dines, J.N., Kaplan, I.M., Svejnoha, E., Craft, T., Boland, K., Pesesky, M., Gittelman, R.M. and Snyder, T.M., 2020. A large-scale database of T-cell receptor beta (TCRβ) sequences and binding associations from natural and synthetic exposure to SARS-CoV-2. Research square.

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

[32] Ansel, J., Yang, E. et. Al. (2024). PyTorch 2: Faster Machine Learning Through Dynamic Python Bytecode Transformation and Graph Compilation [Conference paper]. 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2 (ASPLOS '24). https://doi.org/10.1145/3620665.3640366

[33] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M. and Davison, J., 2019. Huggingface's transformers: State-of-the-art natural language processing. arXiv preprint arXiv:1910.03771.

[34] Glanville, J., Huang, H., Nau, A., Wang, G.B., Peng, H., Zhou, Z., Grigorieff, N., Chen, Z. and DeKosky, B.J., 2017. Identifying specificity groups in the T cell receptor repertoire. Nature, 547(7661), pp.94-98.

[35] Dash, P., Fiore-Gartland, A.J., Hertz, T., Wang, G.C., Sharma, S., Souquette, A., Crawford, J.C., Clemens, E.B., Nguyen, T.H., Kedzierska, K. and La Gruta, N.L., 2017. Quantifiable predictive features define epitope-specific T cell receptor repertoires. Nature, 547(7661), pp.89-93.

[36] Davis, M.M. and Bjorkman, P.J., 1988. T-cell antigen receptor genes and T-cell recognition. Nature, 334(6181), pp.395-402.

[37] Turner, S.J., Doherty, P.C., McCluskey, J. and Rossjohn, J., 2006. Structural determinants of T-cell receptor bias in immunity. Nature Reviews Immunology, 6(12), pp.883-894.

[38] Tickotsky, N., Sagiv, T., Prilusky, J., Shifrut, E. and Friedman, N., 2017. McPAS-TCR: a manually curated catalogue of pathology-associated T cell receptor sequences. Bioinformatics, 33(18), pp.2924-2929.

[39] Shugay, M., Bagaev, D.V., Zvyagin, I.V., Vroomans, R.M., Crawford, J.C., Dolton, G., Komech, E.A., Sycheva, A.L., Koneva, A.E., Egorov, E.S. and Eliseev, A.V., 2018. VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. Nucleic Acids Research, 46(D1), pp.D419-D427.

[40] Devlin, J., Chang, M.W., Lee, K. and Toutanova, K., 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. North American Chapter of the Association for Computational Linguistics, pp.4171-4186.

[41] Arevalo, J., Solorio, T., Montes-y-Gómez, M. and González, F.A., 2017. Gated multimodal units for information fusion. arXiv preprint arXiv:1702.01992.

[42] Dean, J., Emerson, R.O., Vignali, M., Sherwood, A.M., Rieder, M.J., Carlson, C.S. and Robins, H.S., 2015. Annotation of pseudogenic gene segments by massively parallel sequencing of rearranged lymphocyte receptor loci. Genome medicine, 7, pp.1-8.

[43] Zhang, Yumeng, et al. "Epitope-anchored contrastive transfer learning for paired CD8+ T cell receptor–antigen recognition." *Nature Machine Intelligence* (2024): 1-15.

[44] Zhang, J., Wang, R. and Wei, L., 2024. MucLiPred: multi-level contrastive learning for predicting nucleic acid binding residues of proteins. *Journal of Chemical Information and Modeling*, *64*(3), pp.1050-1065.

[45] Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L. and Chen, W., 2022. Lora: Low-rank adaptation of large language models. *ICLR*, *1*(2), p.3.

[46] Wang, J., Li, W., Hou, C., Tang, X., Qiao, Y., Fang, R., Li, P., Gao, P. and Xie, G., 2022, October. Hcl: Improving graph representation with hierarchical contrastive learning. In *International Semantic Web Conference* (pp. 108-124). Cham: Springer International Publishing.

[47] Chen, T., Kornblith, S., Norouzi, M. and Hinton, G., 2020, November. A simple framework for contrastive learning of visual representations. In *International conference on machine learning* (pp. 1597-1607). PmLR.

[48] Oord, A.V.D., Li, Y. and Vinyals, O., 2018. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*.

[49] Xu, Z., Luo, M., Lin, W., Xue, G., Wang, P., Jin, X., Xu, C., Zhou, W., Cai, Y., Yang, W. and Nie, H., 2021. DLpTCR: an ensemble deep learning framework for predicting immunogenic peptide recognized by T cell receptor. *Briefings in Bioinformatics*, *22*(6), p.bbab335.

[50] Huang, H. et al. Select sequencing of clonally expanded CD8 + T cells reveals limits to clonal expansion. *Proc. Natl Acad. Sci. USA* 116, 8995–9001 (2019).

[51] Im, C., Zhao, R., Boyd, S.D. and Kundaje, A., 2025. Sequence-Based TCR-Peptide Representations Using Cross-Epitope Contrastive Fine-Tuning of Protein Language Models. In *International Conference on Research in Computational Molecular Biology* (pp. 34-48). Springer, Cham.

[52] Kim, J.H., Jun, J. and Zhang, B.T., 2018. Bilinear attention networks. Advances in neural information processing systems, 31.

[53] Henikoff, S. and Henikoff, J.G., 1992. Amino acid substitution matrices from protein blocks. Proceedings of the National Academy of Sciences, 89(22):10915-10919.

APPENDIX

A Additional Results

To investigate the mechanistic basis of TCR-ECHO's binding predictions, we performed two complementary interpretability analyses: positional contribution analysis via alanine scanning mutagenesis and validation of attention weights against structural data.

Alanine Scanning Reveals Position-Specific Contributions

We applied computational alanine scanning to PanPep's zero-shot test dataset, systematically mutating each CDR3β residue to alanine and measuring the resulting change in predicted binding score. Following PanPep's methodology, we divided each CDR3β sequence into five equal-length segments (N-terminal to C-terminal) and computed the mean drop in prediction score per segment.

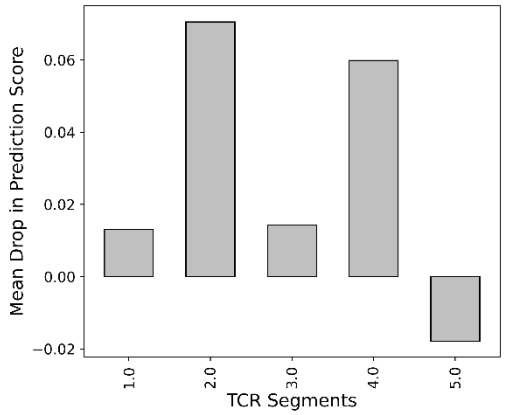

Figure 4: *Mean drop in prediction score when each CDR3β residue is mutated to alanine, aggregated across five equal-length segments*

Figure 4 shows that segments 2 and 4 contribute most significantly to binding predictions, with mean score drops of approximately 0.07 and 0.06 respectively. Notably, segment 5 (C-terminal) showed minimal contribution with near-zero or slightly negative mean drops, suggesting TCR-ECHO learns that terminal residues play a limited role in peptide recognition. The middle segments' elevated importance aligns with experimental observations that central CDR3 residues dominate antigen specificity.

Attention Weights Capture 3D Structural Features

To validate whether TCR-ECHO's attention mechanism captures biologically meaningful interactions, we analyzed the model's cross-attention weights against experimentally determined structural data. Using PanPep's curated dataset of TCR-peptide crystal complexes from IEDB, each annotated with per-residue buried fractions from 3D structures, we computed per-residue attention weights (normalized sum across peptide positions) and compared them with structural burial. The buried fraction quantifies the proportion of each residue's surface area that is hidden within the protein interior or at binding interfaces, with higher values indicating residues more deeply embedded in the structure and typically more critical for binding stability.

Per-complex analysis revealed substantial variability in correlation strength. The top-performing complexes showed strong agreement between attention and structure: 3KPR (ρ=0.808), 5TEZ (ρ=0.643), 4QRR (ρ=0.613). This heterogeneity likely reflects differences in binding modes, structural resolution, and the extent to which different TCR-peptide pairs rely on buried core residues versus surface-exposed contacts.

Together, these analyses demonstrate that TCR-ECHO learns interpretable, biologically grounded representations: the alanine scanning identifies position-specific contributions matching known CDR3 biology, while attention weights correlate with structural features despite training exclusively on sequence data.

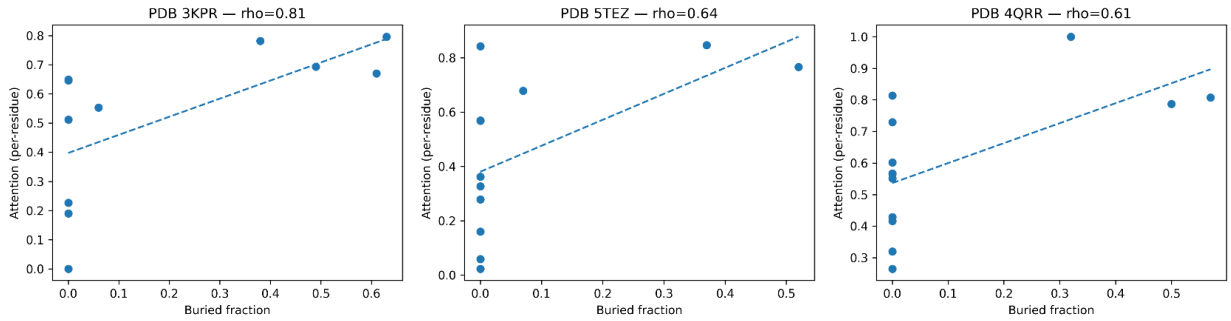

Figure 5: *Per-residue attention versus buried fraction for three top-performing complexes. Each point is one CDR3β residue; dashed lines show linear fits.*

B Datasets

For the TCR-epitope binding prediction task, we used datasets provided by PanPep [25]. The positive pairs of the training set were acquired by combining PanPep's meta dataset and majority dataset, and negative samples were created by matching epitopes from the positive samples with TCRs chosen at random from a control TCR set. This set has TCR sequences collected from 587 healthy volunteers [42, 25]. The training set had 61000 total pairs with 208 unique epitopes. The zero-shot test set used for evaluating TCR-ECHO''s performance on entirely new epitopes was the same as the one used by PanPep. This dataset contained 1714 TCR-epitope pairs, with 491 unique epitopes, which were not present in the training set.

We used the dataset provided by VitTCR to compare performance with their model. TCR-ECHOs were trained on their 5 training folds and tested on their independent dataset. Any common pairs between each fold's training data and the independent sets were removed for testing the corresponding fold.

Additionally, we curated a COVID-19 dataset following the instructions provided by PanPep. We collected the COVID-19 cohort from the ImmuneCODE project [28, 29], which initially contained 154,320 records associated with 269 different peptide groups. The TCRs binding to each peptide were balanced using the control TCR set. The final COVID-19 dataset consisted of 1,100,136 peptide-TCR pairs associated with 541 different peptides. This dataset was used to assess TCR-ECHO''s performance on a real-world, independent dataset in a zero-shot setting.

C Training details

TCR-ECHO was implemented in PyTorch, utilizing separate ESM2 encoders for TCR CDR3β and peptide sequences. The pretrained ESM2 models were obtained from the official ESM repository and Hugging Face Transformers library. A classification head with 512 hidden units connects the cross-attention output to the final prediction layer. We employed Low-Rank Adaptation (LoRA) for efficient fine-tuning of the ESM2 encoders. LoRA parameters were configured with rank r=32, scaling parameter α=32, and dropout=0.0. Fine-tuning targeted the top 5 transformer layers of each encoder to preserve lower-level evolutionary representations while adapting higher-level features for binding prediction.

TCR CDR3β sequences were truncated or padded to a maximum length of 25 residues, while peptide sequences were limited to 15 residues. During training, 10% of sequence tokens were randomly masked to encourage robust representation learning while preserving sufficient sequence information for both contrastive learning and binding prediction.

Models were trained for 200 epochs with a batch size of 32 and learning rate of 5e-5. We applied a weight decay of 5e-5 and dropout rate of 0.05 for regularization. A validation set comprising 15% of the training data was held aside

for model selection. The best model checkpoint was saved based on validation AUC score to prevent overfitting. The loss function combined focal loss (γ=2.0) for the classification objective with our hierarchical contrastive learning framework. Optimal training hyperparameters were identified using Optuna, exploring learning rates, batch sizes, contrastive loss weights, temperature values, and LoRA configurations. The reported results use the best-performing hyperparameter combination.

The hierarchical contrastive framework employed two loss components: encoder-level contrastive loss (λ_enc=0.7) and interaction-level contrastive loss (λ_int=2.0). Temperature scaling was set to τ=0.05 for both contrastive objectives. These parameters balance the contribution of sequence-level and interaction-level representations during training.

Computational Resources: All experiments were conducted on dual NVIDIA GeForce RTX 2080 Ti GPUs (11GB VRAM each). Peak GPU memory usage reached approximately 4GB during training. Training time ranged from 3-6 hours depending on dataset size. Evaluation used identical batch sizes (32) and logged performance metrics including AUROC, accuracy, and F1-score via Weights & Biases.

D Data and Code Availability

The datasets used in this study are available from the following sources: The PanPep datasets for TCR-epitope binding prediction (Majority, zero-shot and control TCR sets) can be found at https://github.com/bm2-lab/PanPep. Instructions to download VitTCR's datasets can be found at https://github.com/Jiang-Mengnan/VitTCR/tree/main. The independent Covid-19 dataset was curated from https://clients.adaptivebiotech.com/pub/covid-2020. The code for TCR-ECHO, along with the trained models and instructions for running the experiments and ablation datasets is available in the GitHub repository: https://github.com/Ong-Research/TCR-ECHO .