T-cell receptor specificity landscape revealed through de novo peptide design

Gian Marco Visani,[1,*] Michael N. Pun,[2] Anastasia A. Minervina,[3]
Philip Bradley,[4] Paul Thomas,[3] and Armita Nourmohammad[1,2,4,5,*]

[1] *Paul G. Allen School of Computer Science and Engineering, University of Washington, 85 E Stevens Way NE, Seattle, WA 98195, USA*
[2] *Department of Physics, University of Washington, 3910 15th Avenue Northeast, Seattle, WA 98195, USA*
[3] *St. Jude Children's Research Hospital, Memphis, TN, USA*
[4] *Fred Hutchinson Cancer Center, 1241 Eastlake Ave E, Seattle, WA 98102, USA*
[5] *Department of Applied Mathematics, University of Washington, 4182 W Stevens Way NE, Seattle, WA 98105, USA*

T-cells play a key role in adaptive immunity by mounting specific responses against diverse pathogens. An effective binding between T-cell receptors (TCRs) and pathogen-derived peptides presented on Major Histocompatibility Complexes (MHCs) mediate an immune response. However, predicting these interactions remains challenging due to limited functional data on T-cell reactivities. Here, we introduce a computational approach to predict TCR interactions with peptides presented on MHC class I alleles, and to design novel immunogenic peptides for specified TCR-MHC complexes. Our method leverages HERMES, a structure-based, physics-guided machine learning model trained on the protein universe to predict amino acid preferences based on local structural environments. Despite no direct training on TCR-pMHC data, the implicit physical reasoning in HERMES enables us to make accurate predictions of both TCR-pMHC binding affinities and T-cell activities across diverse viral epitopes and cancer neoantigens, achieving up to 72% correlation with experimental data. Leveraging our TCR recognition model, we develop a computational protocol for de novo design of immunogenic peptides. Through experimental validation in three TCR-MHC systems targeting viral and cancer peptides, we demonstrate that our designs—with up to five substitutions from the native sequence—activate T-cells at success rates of up to 50%. Lastly, we use our generative framework to quantify the diversity of the peptide recognition landscape for various TCR-MHC complexes, offering key insights into T-cell specificity in both humans and mice. Our approach provides a platform for immunogenic peptide and neoantigen design, opening new computational paths for T-cell vaccine development against viruses and cancer.

I. INTRODUCTION

T-cells play a key role in the adaptive immune system, contributing to immune surveillance and response to pathogens. T-cells recognize pathogen-derived epitopes in the form of short protein fragments (peptides) displayed on specific molecules known as Major Histocompatibility Complexes (MHCs). The binding between a T-cell receptor (TCR) and a peptide-MHC (pMHC) is essential to mediate a protective T-cell response to an infection. To defend the host against a multitude of pathogens, a highly diverse TCR repertoire is generated through a somatic VDJ recombination process, and selected to target infected cells with remarkable sensitivity and specificity.

Predicting the TCR-pMHC binding specificity is an important step in characterizing the efficacy of the immune system in countering different pathogens. Such understanding would aid with disease diagnoses and development of diagnostic tests for early detection of autoimmune diseases [1]. A predictive model for TCR-pMHC specificity can be used to design engineered TCRs that recognize cancer antigens [2], enhancing the effectiveness of adoptive cell transfer and Chimeric Antigen Receptor (CAR) T-cell therapies [3, 4], as well as the design of soluble TCRs and bispecific T-cell engager therapeutics [5]. Moreover, it can enable the design of optimized antigens for vaccines to elicit robust T-cell responses against specific pathogens or tumors [6, 7].

One bottleneck in deciphering the TCR-pMHC code is the lack of well-curated datasets. High-throughput immune repertoire sequencing, combined with comparative analyses of hosts facing similar immune challenges, and the development of functional assays (e.g., pMHC-tetramer staining [8] or MIRA [9]), have enabled the identification of diverse TCR pools with potential reactivity to specific antigens [10], collected in repositories like Immune Epitope Database (IEDB) [11] and VDJdb [12]. However, the current databases are unbalanced, with only a handful of dominant peptides (e.g. derived from SARS-CoV-2, CMV, EBV, Influenza, etc.) and MHC alleles present, each linked to hundreds of TCRs. As a result, statistical inference and machine learning methods trained on such data to predict TCR-pMHC reactivities lack generalizability beyond their training data [13–15]. On the structural side, only a few hundred TCR-pMHC crystal structures have been experimentally generated to date. This is in part because TCR-pMHC interactions

* Correspondence should be addressed to: Gian Marco Visani: gvisan01@cs.washington.edu, and Armita Nourmohammad: armita@uw.com.

are typically of low affinity and are difficult to stabilize to the level needed for CryoEM or X-ray crystallography [16]. Importantly, this scarcity hinders computational prediction methods like AlphaFold in modeling reliable TCR-pMHC structures [17, 18].

Machine learning models, such as large language models, pre-trained on vast unannotated protein datasets can generate meaningful data representations [19, 20]. Fine-tuning these models on smaller, labeled datasets for specific tasks (e.g., predicting mutation impacts on stability or function) has proven highly effective [21–24]. Pre-trained sequence-based protein language models have been employed to predict reactivities of antibodies to antigens [23, 24], or TCRs to pMHC complexes [15, 25]. However, lack of generalizability due to sparse and skewed training data has remained a challenge in these models.

Protein-protein interactions in general, and immune-pathogen interaction in particular, rely on the complementarity in the 3D structures and amino acid compositions of the interacting protein pairs. Even though the interacting amino acids can be far apart in a protein sequence, they are close to each other in structure, resulting in local interactions that can determine immune-pathogen recognition. Machine learning models trained on protein structures instead of sequences can learn local structural representations for amino acid statistics, which tend to be more generalizable across different protein families and beyond their training sets [26, 27, 30–33].

Here, we present a structure-based approach to predict interaction affinities between TCRs and peptides presented on MHC class I, and to design reactive peptides for specific TCR-MHC-I complexes (TCR-MHC for short). Due to the limited availability of structural data for TCRs, we use HERMES– a physics-guided, 3D-equivariant machine learning model trained on the entire protein universe to predict amino acid preferences based on their local structural environments [26, 27]. We demonstrate that HERMES reliably predicts interaction affinities for diverse TCR-pMHC complexes, and the activity of T-cells in response to peptides presented by MHC class I molecules. Moreover, we show that HERMES can be used to design novel peptides binding specific TCR-MHC complexes, achieving up to 50% experimental validation accuracy across diverse TCR-MHC systems. Our peptide design platform can facilitate the computational design of neoantigen libraries for cancer vaccines. By leveraging our design algorithms, we characterize the specificity of a diverse range of TCRs, providing a quantitative measure for the diversity of the peptide-MHC antigens that a TCR can recognize in humans and mice.

II. MODEL

T-cell response is mediated by the interactions between TCRs and pMHC complexes. To model the TCR-pMHC interaction, we use HERMES [26], a 3D rotationally equivariant structure-based neural network model, trained on the protein universe; see Fig. 1A. HERMES predicts the amino acid propensity from its surrounding 3D protein environment, a measure that we have previously shown to serve as a reliable proxy for assessing the effect of mutations on protein stability or binding [26, 27].

For TCR-pMHC complexes, we seek to determine how changes in a peptide sequence impact the binding affinity of a TCR-pMHC complex, and ultimately, the T-cell response to the antigen. To do so, we characterize a score, which we term peptide energy, for a peptide σ, given its surrounding TCR and MHC structural environment,

$$E_{\text{peptide}}(\sigma; \text{TCR, MHC}) = \sum_{i=1}^{\ell} E(\sigma_i; \text{TCR, MHC}, \sigma_{/i}) \tag{1}$$

We assume that each peptide residue contributes linearly to the total energy by the amount $E(\sigma_i; \text{TCR, MHC}, \sigma_{/i})$, where σ_i is the amino acid at position i of the peptide. A residue's energy contribution E is evaluated by HERMES (logits of the network), taking in as input the atomic composition of the surrounding TCR, MHC, and the rest of the peptide $\sigma_{/i}$ (excluding the ith residue) within a 10 Å of σ_i's α-Carbon; see Fig. 1 and SI for details.

To compute peptide energy, we input to HERMES an experimentally or computationally determined structure of a specified TCR-MHC complex bound to at least one peptide variant. Experimental data on the impact of peptide substitutions on the TCR-pMHC structure is limited, and computational models often fail to capture subtle conformational changes in the structure due to amino acid substitutions in a peptide. Given these constraints, we adopt two approaches to estimate peptide energies across diverse peptide variants for a given TCR-MHC complex:

- HERMES-*fixed*: In the simplest approach, we choose a TCR-pMHC structure with the same TCR and MHC as our query and a peptide sequence closest to the peptide of interest. This structure is then used as our template in HERMES to compute the peptide energy as described in eq. 1 (Fig. 1B). This method does not alter the underlying structure and assumes that peptide amino acid substitutions do not significantly change the conformation of the TCR-pMHC complex.

- HERMES-*relaxed*: Since amino acid substitutions can locally reshape a protein complex, we introduce a more involved protocol to account for these structural changes. We begin with the closest available structure for the TCR-pMHC complex. We then

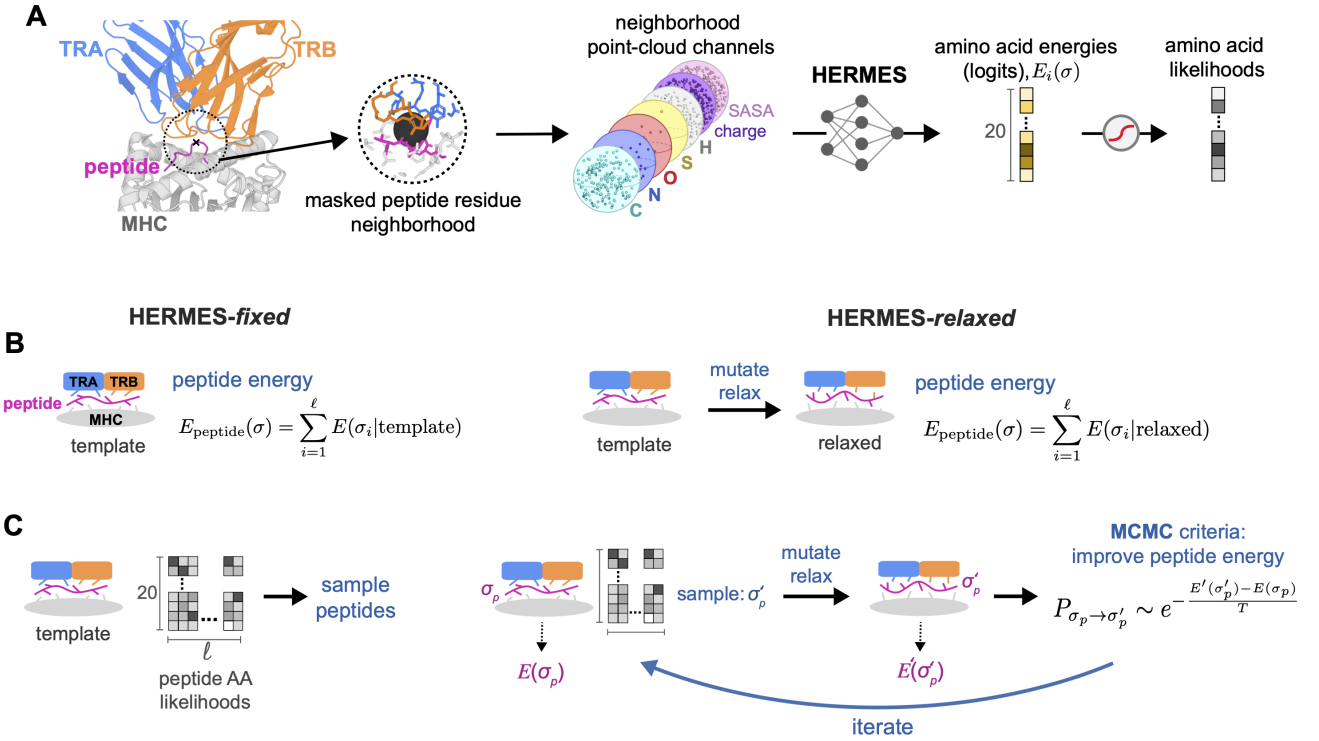

Figure 1. **HERMES model for predicting and designing TCR-pMHC interactions. (A)** HERMES is a structure-based, rotationally equivariant neural network pre-trained on the protein universe to predict amino acid preferences at a residue from its 3D structural environment [26, 27]. It takes as input all atoms within 10Å of a focal residue (with that residue's atoms masked), featurizing them by atom type, charge, and SASA. HERMES then outputs energies $E_i(\sigma)$ for the 20 possible amino acids σ at residue i, which serve as logits in a softmax to estimate amino acid likelihoods. **(B)** To predict TCR-pMHC affinity, we use two protocols: HERMES-*fixed* (left) and HERMES-*relaxed* (right). In HERMES-*fixed* we use a fixed the structural template closest to the TCR-pMHC of interest as input to HERMES, and evaluate the peptide energy as the sum of HERMES-computed per-residue energies for the peptide. In HERMES-*relaxed* mutations are introduced to the template's peptide to match the peptide of interest, followed by local repacking in PyRosetta [28]. Peptide energy is then evaluated similar to HERMES-*fixed*, but using the relaxed structure. **(C)** We also use two protocols for peptide design. Similar to (B) we fix the closest structural template in HERMES-*fixed*, and use HERMES a likelihood matrix for possible amino acids at each peptide position, which we sample to generate new peptide sequences. In HERMES-*relaxed* (right), we begin with a random peptide σ, and locally pack it with PyRosetta, and also generate an amino acid likelihood matrix (similar to HERMES-*fixed*) for this packed structure. We sample a new sequence σ' from this likelihood matrix, and pack it with PyRosetta. We evaluate peptide energies $E_{\text{peptide}}(\sigma)$, and $E_{\text{peptide}}(\sigma')$ for both peptides, using their respective PyRosetta-relaxed structures, and we use these energies as the criterion to sample improved sequences with MCMC (SI). This procedure is then iterated until convergence. Peptides designed by both approaches are then filtered by TCRdock [17] or Alpha-Fold 3 [29] to assure the stability of the complex.

mutate the original peptide to the desired state and apply a relaxation procedure using PyRosetta to properly pack the substituted amino acid [28] (SI). During relaxation, side-chain and backbone atoms of the peptide are allowed to move, while only the side-chain atoms of the MHC and TCR chains within a 10 Å radius of the peptide are flexible. We then calculate the peptide energy with HERMES (eq.1), using the relaxed structure as input (Fig. 1B); since PyRosetta's relaxation procedure is stochastic, we average the peptide energy across 100 realizations of these relaxations.

III. RESULTS

Predicting binding affinity of TCR-pMHC complexes

Binding between TCRs and pMHC complexes is necessary to mediate a T-cell response. The binding affinity of a natural TCR-pMHC complex is relatively low, with a dissociation constant K_d in the range of $1 - 100\mu$M [16, 38, 39], in contrast to nanomolar affinities for antibody-antigen complexes. Engineered TCRs can achieve much higher affinities in a nanomolar [40] to picomolar [41, 42] range.

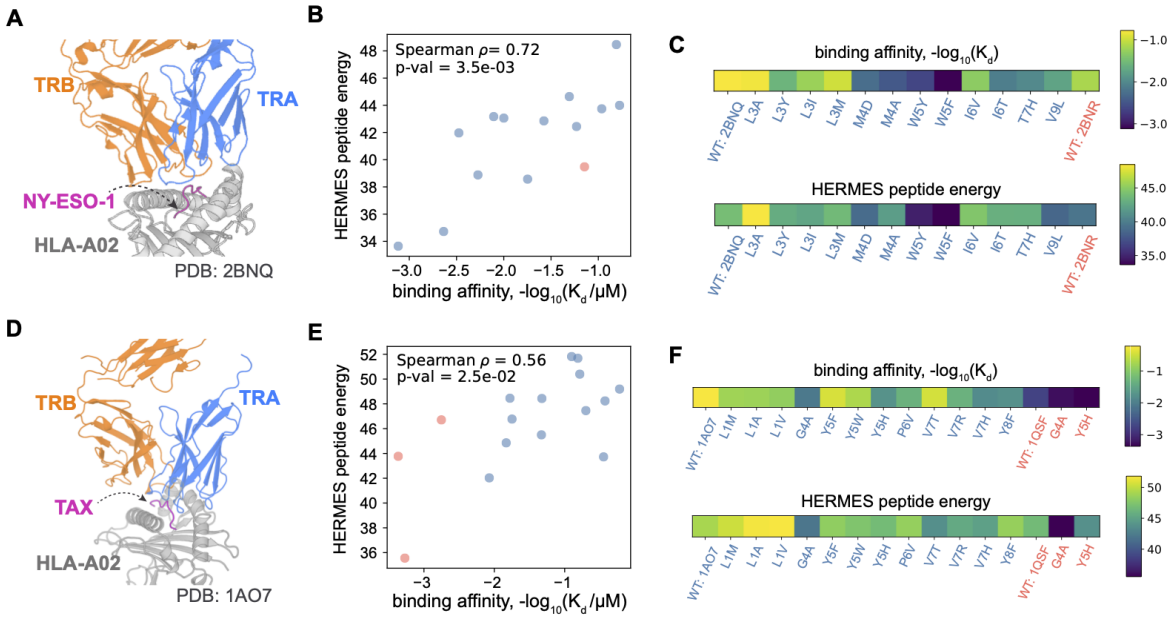

Figure 2. **Predicting binding affinity of TCR-pMHC complexes. (A)** The structure of the NY-ESO-1 peptide (purple), in complex with HLA-A*02:01 and a purified specific 1G4 TCR is shown (PDB ID: 2BNQ [34]). This structure together with that of a peptide variant with C9V mutation (PDB ID: 2BNR [34]) are used as templates to predict the binding affinity of the TCR-MHC complex to other peptide variants. **(B)** The binding affinities ($-\log_{10}(K_d/\mu\text{M})$) TCR-MHC complexes with 14 distinct peptides, measured using surface plasmon resonance, are shown against the predicted HERMES-*relaxed* peptide energies, using the closest structural template for each variant; 2BNQ is used for blue points and 2BNR for the orange point. The Spearman ρ between predictions and measurements is reported in the panel. **(C)** The heatmaps show the experimental affinity values (top) and predicted peptide energies (bottom) for all the peptide variants, with the identity of the mutations color coded based on the closest structural template. **(D-F)** Similar to (A-C) but for the Tax viral peptide in complex with HLA-A*02:01 and a specific human TCR. Two structural templates are used [35] for evaluating peptide energies with PDB IDs: 1AO7 (shown in D, and used for blue points), and 1QSF (used for orange points).

Surface plasmon resonance (SPR) spectroscopy provides accurate measurements of the dissociation constant (K_d) for specific TCR-pMHC systems. In these experiments, one protein–either the TCR or the MHC–is immobilized on a conducting plate, and the other is introduced in solution to bind to it. This binding alters the local refractive index near the plate, affecting the resonance signal. We used a published dataset of SPR-measured affinities for two TCR-MHC complexes binding to an ensemble of peptides, for which crystal structures of the complexes with at least one peptide are available [43, 44]; see Table S1 for details.

In our first example, we examined TCR 1G4 binding to NY-ESO-1 peptide variants, which is a cancer-testis antigen commonly expressed in many cancers and is targeted by immunotherapies [45]. We used two structural templates, one with the wild-type peptide SLLMWITQC (NY-ESO-1), and the other with a more immunogenic peptide, in which the Cysteine at position 9 is substituted by Valine (C9V) [34, 46] (Fig. 2A). For variants differing by a single amino acid, our predictions with HERMES-*relaxed* correlated well (72% Spearman correlation) with experimentally measured affinities from refs. [43, 44](Fig. 2B, C).

Next, we tested the A6 TCR binding to the HTLV-1 Tax peptide LLFGYPVYV. We employed two structural templates: one with the wild-type, the other with the mutant peptide with substitution Y8A [35] (Fig. 2D and Table S1). Again, the HERMES-*relaxed* peptide energies correlated well (56% Spearman ρ) with the affinities, measured in ref. [44] (Fig. 2E, F).

Overall, HERMES predictions aligned closely with this limited experimental affinity measurements and performed as well as or better than BLOSUM-based scoring [47], ProteinMPNN [30], and TCRdock [17] (Table S2). Notably, unlike genome-averaged methods such as BLOSUM, HERMES is sensitive to the local structural environment, assigning different scores to the same mutation depending on context (Fig. S1). HERMES can predict relative peptide binding affinities for TCR-MHC complexes, provided a suitable structural template is available. Prediction quality suffers in the absence of reliable structural templates, as seen also in other methods such as TCRdock [17] (Table S2). Lastly, HERMES-*relaxed* peptide energies show better agreement with the affinity measurements than HERMES-*fixed* (Table S2). However, the small size of this dataset prevents us from drawing broader conclusions about the relative utility of

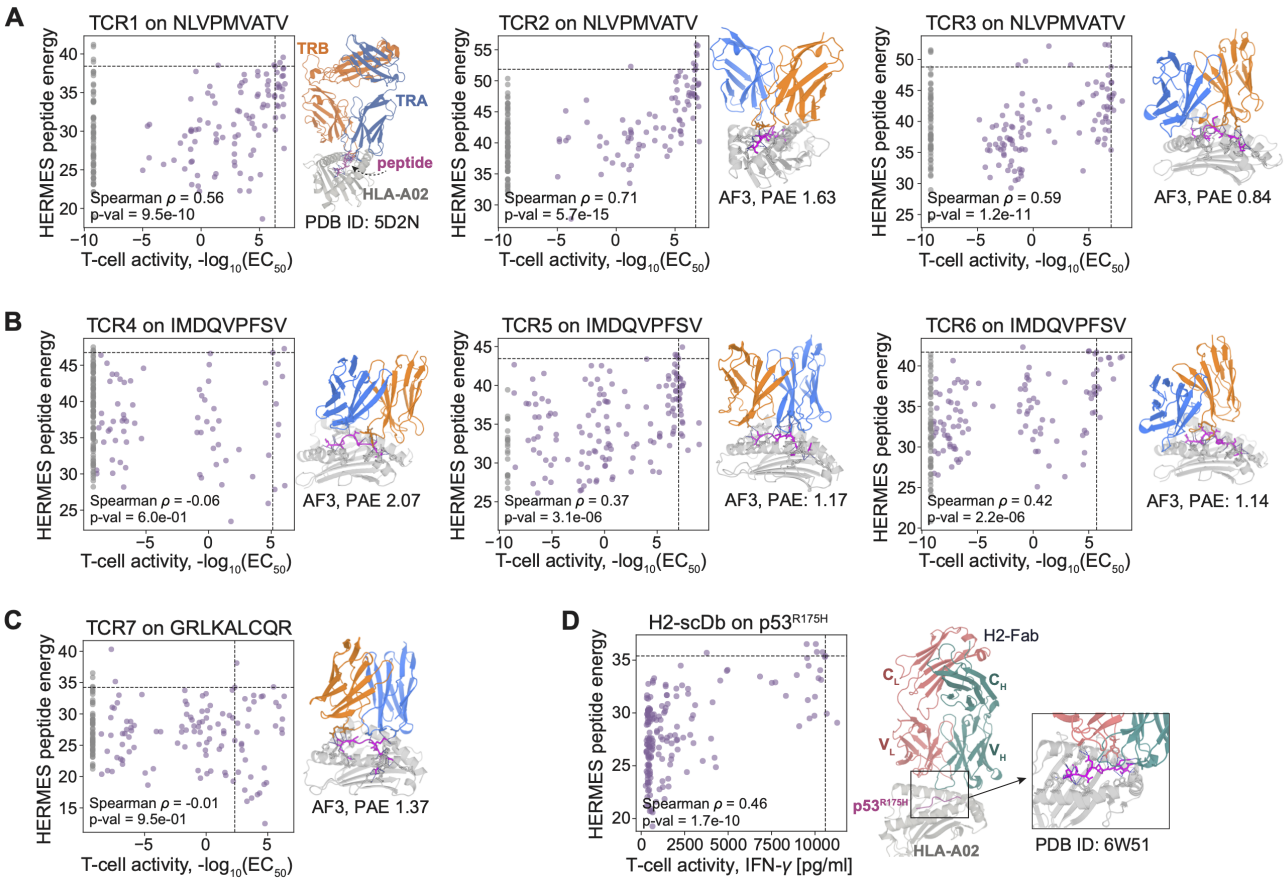

Figure 3. **Predicting T-cell activity in response to peptide variants. (A-C)** The predicted HERMES-*fixed* peptide energies versus experimentally measured T-cell activity are shown for (A) HLA-A*02:01-restricted CMV peptide variants interacting with TCR1-TCR3 (panels), (B) HLA-A*02:01-restricted melanoma gp100 peptide variants interacting with TCR4-TCR6, and (C) HLA-B*27:05-restricted pancreatic cancer peptide variants interacting with TCR7. In each case, T-cell activity in response to all single amino acid mutations from the wild-type is measured. T-cell activity is based on fitted EC_{50} values from dose-response curves of 4-1BB$^+$ CD8$^+$ T-cells (SI; data from [36]). Gray points indicate peptides with T-cell responses below detection limit (no EC_{50} obtained) and are excluded from the reported Spearman correlations. Except for TCR1 (resolved structure, PDB ID: 5D2N [37]), the other TCR-pMHC complexes were modeled with AF3, and the model's peptide PAE for each wild-type peptide is denoted. These structures are used as templates to estimate peptide energies with HERMES-*fixed*; see Tables S1, S3 for details on the templates. For all structures, TCR-α (TRA) is shown in blue, TCR-β (TRB) in orange, the peptide in magenta, and HLA in gray. **(D)** The predicted HERMES-*fixed* peptide energies are shown against experimentally measured T-cell activity for single point mutants of the HLA-A*02:01-restricted p53^{R175H} neoepitope from ref. [5]. T-cell response (IFN-γ production) is mediated through a bi-specific antibody binding to the pMHC complex. The co-crystallized structure of the bi-specific antibody with the p53^{R175H}-HLA-A*02:01 complex (PDB ID: 6W51 [5]) is used as the template for our HERMES-*fixed* predictions. For all systems (A-D), the heatmaps for the predicted and the experimentally measured mutational effects are shown in Fig. S2.

either approach for this problem.

Predicting T-cell response to peptide antigens

Aside from the affinity of TCR-pMHC interactions, other factors play important roles in determining whether and how a T-cell responds. These factors include the phenotypic state of a T-cell (e.g., naive or memory), the reactivity of its TCR to competing self-antigens presented on MHC molecules [48, 49], and the avidity of TCR-pMHC interactions, determined by the clustering of TCRs with co-receptors on the T-cell membrane and the density of pMHC molecules on antigen-presenting cells [50]. T-cell activation can be characterized using a diverse set of functional assays that measure T-cell proliferation, labeled cellular activation markers through flow cytometry, and production of cytokines such as interferon gamma (IFN-γ). As such, measuring T-cell activity is much easier than measuring the affinity of TCR-pMHC com-

plexes, which requires techniques like SPR. Although our structure-based HERMES model is more attuned to predicting the binding affinity of TCR-pMHC complexes, the ease of data production prompted us to assess the accuracy of our method in predicting T-cell activity.

As the first system, we examined T-cell responses to all single-point mutants of a native peptide, as measured in ref. [36] for different TCR-pMHC systems. Specifically, we make predictions for (i) three TCRs (TCRs 1-3) recognizing variants of a highly immunogenic HLA-A*02:01-restricted human cytomegalovirus (CMV) epitope NLVPMVATV (NLV), (ii) three TCRs (TCRs 4-6) recognizing variants of the weaker HLA-A02:01-restricted melanoma gp100 self-antigen IMDQVPFSV, and (iii) a TCR (TCR 7) recognizing variants of weakly immunogenic HLA-B*27:05-restricted pancreatic cancer neopeptide GRLKALCQR, believed to have elicited an effective immune response in long-term cancer survivors [36]. T-cell activity was quantified by determining the half-maximal effective concentration (EC_{50}) from the dose-response curves of 4-1BB$^+$ CD8$^+$ T-cells across varying peptide concentrations; see Methods on details of inferring EC_{50}.

With the exception of TCR1 (where a TCR-pMHC template with HLA-A*02:01 and one peptide variant was already available), we relied on AlphaFold3 (AF3) [29] to model the remaining TCR-pMHC complexes; see Tables S1, S3 for template information and AF3 sequence inputs for TCRs 2-7. As shown in Figs. 3A, and S2, HERMES-*fixed* peptide energy predictions (eq. 1) correlate well with experimental EC_{50}, reaching up to 71% Spearman correlation for TCR2 recognizing the CMV peptides. Correlations are lower for TCRs 4-6 targeting variants of the melanoma gp100 self-antigen, with TCR6 achieving 42% Spearman correlation as the best prediction in this class (Fig. 3B). HERMES underperforms for TCR7 responding to the pancreatic cancer neopeptide GRLKALCQR (Fig. 3C). Still, except for TCR4, which has the largest AF3 Predicted Alignment Error (PAE) of 2.07 at the TCR-pMHC interface (i.e., a lower confidence in its fold), HERMES predictions outperform those of TULIP [15], the state-of-the-art sequence-based language model trained to predict TCR-pMHC interactions (Table S4). The reduced accuracy of HERMES for cancer epitopes compared to the variants of the viral CMV epitope could reflect the impact of antagonism from self-antigens similar to tumor epitopes, hindering T-cell responses [48, 49]—though additional experiments are needed to confirm this.

Next, we examined the therapeutic T-cell responses to all single point mutants of the HLA-A*02:01-restricted p53-derived neoepitope p53^{R175H}, HMTEVVRHC, measured in ref. [5]. This T-cell therapy approach used a designed bi-specific antibody that binds both to the TCR-CD3 complex on T-cells and to the variants of the p53^{R175H}-MHC complex on tumor cells (Table S1). A strong binding to an epitope activates the engaged T-cell, and the extent of such activity was measured by the production of IFN-γ. The resulting HERMES-*fixed* predicted peptide energies (eq. 1) show a Spearman $\rho = 46\%$ with the experimentally measured IFN-γ production (Figs. 3D, S2).

Our analysis shows that, despite not being explicitly trained for this purpose, HERMES peptide energy effectively predicts T-cell responses to neoepitopes presented by different MHC-I alleles. Moreover, HERMES predictions are not limited to TCR-pMHC interactions but also extend to antibody-pMHC interactions, commonly leveraged in T-cell therapies for their higher achievable affinities. The HERMES predictions on T-cell activities perform as well or better than other computational methods including BLOSUM62 [47], ProteinMPNN [30], and TCRdock [17]; see Table S4 for a detailed benchmark of different approaches, including different HERMES models.

Design of novel peptides reactive to a TCR-MHC complex

Vaccination protocols aim to elicit robust immune responses that induce lasting immune memory in individuals against viral infections and cancer. Peptide-based cancer vaccines represent a promising therapeutic strategy to mount an effective T-cell response, offering the potential to be highly personalized to an individual's unique tumor antigens and T-cell repertoire [7, 55]. However, the clinical success of cancer vaccines has been limited by factors such as low peptide immunogenicity and insufficient uptake by antigen-presenting cells [56].

Recent approaches have employed deep mutational scanning (DMS) experiments to identify more immunogenic cancer neoantigens [36]. However, these experiments are often limited to scanning single amino acid mutations from the wild-type (native) peptide due to the combinatorial explosion when exploring multiple mutations.

Here, we present a pipeline for designing peptides with enhanced TCR-MHC reactivity by leveraging HERMES's ability in predicting T-cell response to peptides presented by MHC-I molecules. Our structure-based approach starts with an existing template structure–experimentally or computationally resolved–of a TCR-MHC in complex with at least one peptide variant, which we refer to as the wild-type peptide. We use two methods to sample peptides, with different degrees of complexity; see Fig. 1C for a schematic description of the two pipelines.

In the basic pipeline (HERMES-*fixed*), we sequentially mask the atoms of each amino acid along the peptide, one at a time. Using the HERMES model, we predict the probability of each amino acid type for a given residue, based on the local structural environment within 10 Å of the masked residue. Repeating this procedure for every amino acid along the peptide yields a position weight matrix (PWM) that represents the probabilities of differ-

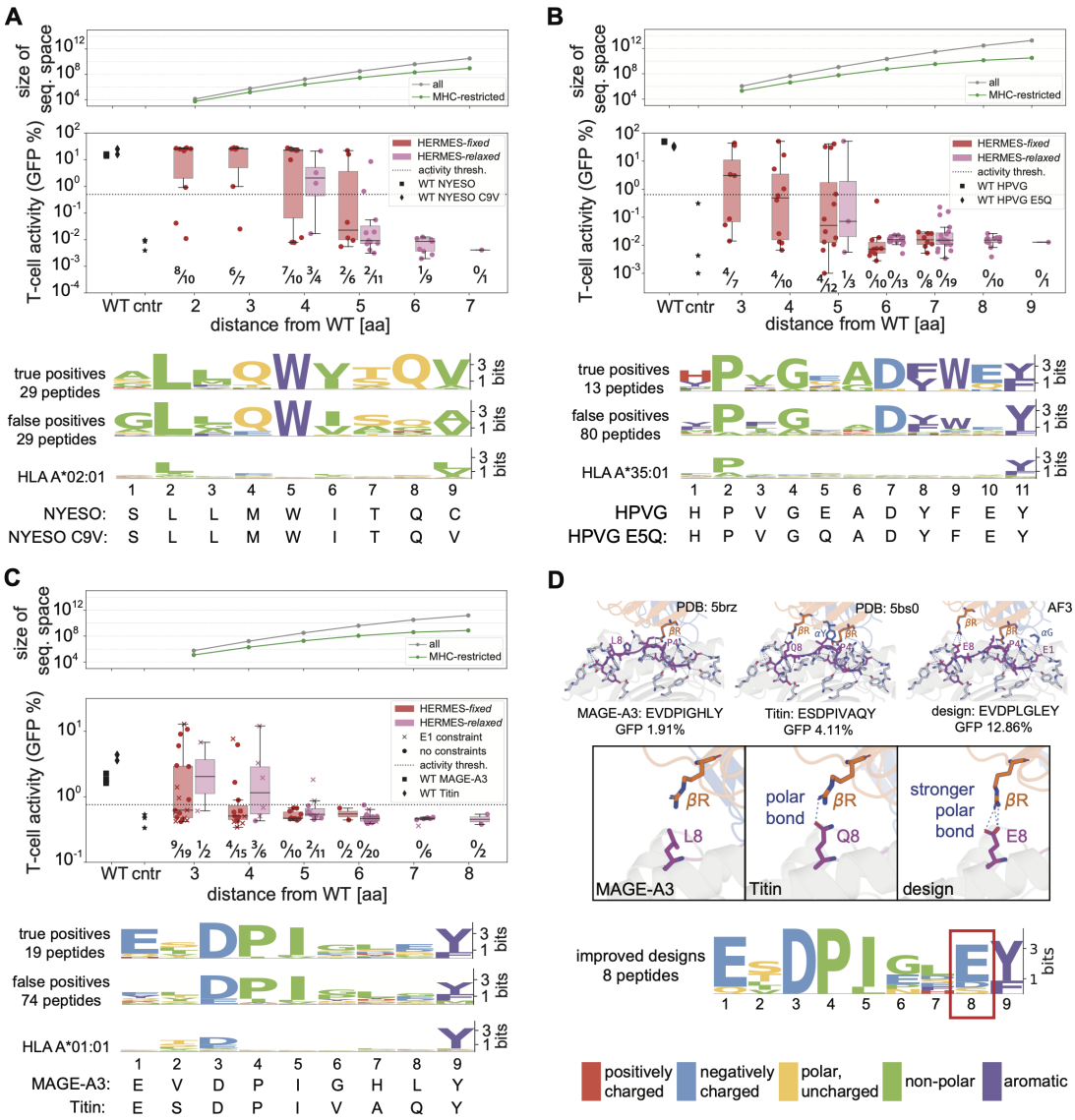

Figure 4. **De novo peptide design to elicit a T-cell response. (A-C)** Validation results are shown for designed peptides built upon the wild-type templates from (A) NY-ESO peptide against TCR 1G4 and HLA-A*02:01 (PDB IDs: 2BNR, 2BNQ [34]), (B) EBV epitopes against the TK3 TCR and the HLA-B*35:01 (PDB IDs: 3MV7 [51], 4PRP [52]), and (C) MAGE-A3 and Titin epitopes against an engineered TCR and HLA-A*01:01 (PDB IDs: 5BRZ, 5BS0 [53]). Designs were validated using Jurkat cells with endogenous NFAT-eGFP reporter expressing the paired TCR, interacting with peptide presented by the MHC in each system. The reported T-cell activities (percentage GFP-positive) are averaged over three replicates. The box plots show T-cell activity for designed peptides at different Hamming distances from the wild-type; the black line is the median, the box spans the 25-75% quantiles, and dots indicate individual data points. Red and pink boxes/markers correspond to designs with HERMES-*fixed* and HERMES-*relaxed* models, respectively; square markers denote wild-type peptides, and stars are negative controls (no peptide). The dashed line marks the activation threshold, and numbers below each bar show the relative number of successful designs out of total designs. The top plot in each panel shows the size of the peptide sequence space at different distances from the wild-types, with the gray line indicating the complete sequence space, and the green indicating the typical number of sequences likely to be presented by the MHC of each system; see SI for details. The bottom shows sequence motifs as PWMs for (i) true-positive designs that induced T-cell activation, (ii) false positives that did not, and (iii) the MHC presentation motif from the MHC Motif Atlas [54]. **(D)** 3D structures of TCR–pMHC complexes focused on the peptide are shown for MAGE-A3 (PDB ID: 5BRZ), Titin (PDB ID: 5BS0), and the highest-activity design in (C) modeled by AF3 [29]. The enhanced activity for Titin versus MAGE-A3, and for the designed peptide versus both wild-types follows from the formation of polar bonds between position 8 of the peptide (magenta) and the CDR3 loop of TCRβ (orange). The PWM highlights the prominence of E_8 among the eight designs with improved T-cell activity in (C).

ent amino acids at each peptide position. Peptides are then sampled by drawing from this PWM. It should be

noted that HERMES-*fixed* retains the structural fingerprints of the original peptide, as it does not account for local structural changes induced by amino acid substitutions; see Fig. 1C and SI for more details.

To incorporate structural flexibility and relaxation following amino acid substitutions, we introduce the design pipeline of HERMES-*relaxed*, using simulated annealing and Markov chain Monte Carlo (MCMC) sampling. We begin with the template TCR-MHC structure and a completely random peptide sequence, which is packed *in silico* within the structure using PyRosetta's packing functionality, followed by the Relax protocol [28]. During relaxation, both side-chain and backbone atoms of the peptide are allowed to move, while only the side-chain atoms of the MHC and TCR chains within 10 Å of the peptide are flexible. With this relaxed structure, we apply the HERMES-*fixed* method to generate a PWM and sample a new peptide, which is then packed and relaxed to form a new structure. We use MCMC to determine whether to accept the newly sampled peptide by comparing its HERMES peptide energy (favorability), evaluated in its relaxed conformation, to that of the original peptide (eq. 1). We then iteratively perform this procedure, incrementally reducing the MCMC temperature at each step, until the results converge; see Figures 1C, and S3, and SI for more details.

The peptides designed by both approaches are then filtered by TCRdock [17] or AlphaFold3 [29], using the Predicted Alignment Error (PAE) between TCR and pMHC interfaces, to assure folding and binding of the resulting TCR-pMHC structures. For each system, we define a specific PAE threshold close to the TCRDock PAE evaluated for the system's TCR-MHC in complex with the native (wild-type) peptide, which is known a priori to activate T-cells. As a result, the PAE thresholds vary slightly across different systems; see SI for details.

We tested the accuracy of our peptide design pipeline in three systems, with the structural templates taken from (i) the cancer-testis antigen NY-ESO-1 in complex with the TCR 1G4 and HLA-A*02:01, (ii) a peptide derived from the Epstein-Barr virus (EBV) in complex with the TK3 TCR and the HLA-B*35:01, and (iii) the immunotherapeutic target MAGE-A3 in complex with an engineered TCR and HLA-A*01:01. For brevity, we refer to these three systems as NY-ESO, EBV and MAGE, respectively; See Table S5 for information on the specific templates used in each case. It should be noted that as we progressed from NY-ESO to EBV and MAGE, we chose to explore the sequence space farther from the wild-type peptides, leading to more challenging designs.

We validated our designs using Jurkat cells expressing a specific TCR and an endogenous NFAT-eGFP reporter to indicate T-cell activation in response to different peptides. We tested the TCR-MHC specificity in all systems–NY-ESO, MAGE, EBV–under 96 different conditions, including the de novo designed peptides, positive controls (wild-type peptides from the template structures), and an unstimulated "no peptide" control. Peptides were presented on artificial APCs (aAPCs), expressing the specified HLA in each system. For NY-ESO designs, we determined the the percentage of GFP-positive cells through both flow cytometry and fluorescence microscopy. Given the consistency between the two experimental approaches (Fig. S4) we relied on fluorescence microscopy only to measure GFP levels induced by different peptides for MAGE and EBV designs; see SI for experimental details.

We used two structural templates for the NY-ESO system: one containing the original 9-amino acid peptide SLLMWITQC, and another with the more immunogenic C9V substitution [34, 46]. Our de novo designs differed from the closest wild-type peptide by 2 to 7 amino acids, with a TCRDock PAE $\leq$ 5.5. For this system only, we selected 35 negative designs (designed by HERMES, but with PAE $>$ 5.5) for experimental testing: only one of these was a false negative, which interestingly, had a desirable AF3 PAE score (as low as the wild-type peptides); see Figs. S5, S6. For the 58 positive designed peptides (PAE $\leq$ 5.5), we achieved an overall design accuracy of 50%, with GFP levels significantly higher than the negative control; see SI for details. The accuracy of predictions decreases as the sequence divergence of the de novo designs from the wild-type peptides increases (Fig. 4A, S6). HERMES-*fixed* achieved higher design accuracy of 70% within its smaller designed sequence subspace of up to 5 substitutions from the wild-types, whereas HERMES-*relaxed* explored sequences further from the wild-type peptides (4-7 substitutions) but with a reduced accuracy of 24% (Fig. 4A). The main difference between the true and false positive sequences was at position 8–from strongly preferring glutamine within the true positives, to having similar preferences among glutamine, glycine, arginine, or methionine within the false positives (Fig. 4A).

For the EBV system, we used two structural templates: one with the 11-amino acid peptide HPVGEADYFEY—commonly referred to as HPVG—derived from the viral latent nuclear protein EBNA-1, and the other with an epitope from a wide-spread viral variant, with a glutamine at position 5 [52]. Unlike NY-ESO, the HPVG peptide has a more flexible conformation with a helical turn within the MHC binding pocket. Our de novo designs differed from the closest wild-type peptide by 3 to 9 amino acids, with a TCRDock PAE $\leq$ 5.1. Across the 93 de novo designs, we achieved an overall accuracy of 14%, associated with the peptides that induced significant GFP levels in T-cells. The accuracy of predictions decreases with increasing the sequence divergence from the wild-types (Fig. 4B, S7). Notably, we achieved 40% accuracy among the designs within 3-5 amino acid distance of one of the wild-type templates, while none of the designs with larger than 5 amino acid distance were successful. Lastly, we observe a strong preference for glutamic acid at position 10 of the successful designs relative to the false positives (Fig. 4B).

For the MAGE system, we used two structural tem-

plates: one with the 9-amino acid MAGE-A3 peptide EVDPIGHLY and another with the Titin-derived self-peptide ESDPIVAQY, which is expressed in cardiomyocytes [53]. The Titin-derived peptide is known to have triggered an off-target response of the affinity-enhanced MAGE-A3-specific engineered TCR that resulted in severe cardiac toxicity and deaths of two patients [57]. Our de novo designs differed from the closest wild-type peptide by 3 to 8 amino acids, with a TCRDock PAE $\leq$ 5.2 (Fig. 4C). From the resulting 93 de novo designs, 20% activated T-cells with significant levels of GFP expression. Similar to the other systems, the accuracy of predictions decreases with increasing the sequence divergence from the wild-types (Fig. 4C). We achieved 30% accuracy among the designs with 3-5 to amino acid distance, while none of the designs with larger than 5 amino acid distance from the wild-types were successful.

Fig. 4C shows differences in amino acid composition between successful and failed designs. Alanine and glycine scanning experiments have previously suggested that MAGE-A3 peptide variants with an E-DPI---Y motif generally activate the T-cells in this system [58]. Consistent with these findings, the fraction of T-cell activating variants in our designs increases as their sequences more closely resemble this motif (Fig. S8).

In a subset of designs using the MAGE-A3 template, we fixed the glutamic acid at position 1 (E_1), which is part of the E-DPI—Y pattern. In both MAGE-A3 and Titin E_1 is within 4 Å of the MHC and forms strong polar interactions with it (Fig. 4D). Fixing E_1 improved the design accuracy to 1/3, compared to 1/9 when this amino acid was not constrained. Even without this constraint in the design protocol, E_1 appeared in 2/3 of successful designs, and in none of the unsuccessful ones (Fig. S8). Notably, among the more distant designs made with HERMES-*relaxed*, only those containing a fixed E_1 were able to activate T-cells (Fig. 4C). This suggests that constraining essential amino acids can enable broader exploration of sequence space at other positions, possibly due to epistatic interactions. However, further investigation is needed to confirm this.

Lastly, several of our designs outperformed the wild-type peptides in activating T-cells: eight surpassed Titin's GFP levels, and eleven exceeded MAGE-A3's (Fig. 4C). These improved designs strongly favor glutamic acid (E) at position 8 of the peptide, where MAGE-A3 has leucine (L), and Titin has glutamine (Q) (Fig. 4D). In Titin, the glutamine forms a polar bond with an arginine (R) in the interacting TCRβ, and it is likely responsible for increasing Titin's activity relative to MAGE-A3's. The glutamic acid substitution in our designs further strengthens this polar bond (Fig. 4D), resulting in a stronger T-cell response. For NY-ESO and EBV, the GFP levels induced by the wild-type peptides are relatively high, and many of our designs induced comparable T-cell responses (Fig. 4).

Although no direct experimental comparisons were performed, computational evaluations using TCRdock PAE and presentation score by MHC-I alleles indicate that HERMES designs outperform or match those generated by ProteinMPNN [30], the corresponding MHC position weight matrix [54, 59–61], or the BLOSUM substitution matrix [47]; see SI and Figs. S9, S10, and S11 for details. A notable exception is in the MAGE system, where ProteinMPNN designs generally have more favorable PAE scores, albeit with a lower proportion of designs passing the antigen presentation filter by the MHC molecule (Fig. S11).

In all the three systems we used the PAE values from TCRdock and AF3 as filtering criteria to select designs for experimental validation (see SI). The PAE values reported by the structure prediction algorithms have been shown to be noisy indicators of protein folding fidelity and design quality [62]. While larger experimental libraries are needed to draw definitive conclusions, it appears that PAEs from TCRdock and AF3 have some complementary predictive strengths (Figs. S5, S12), and therefore, combining their information could lead to a more robust design decisions (SI).

Overall, HERMES demonstrates a potential for designing highly immunogenic peptides up to five amino acid substitutions away from the wild-type–a task that would otherwise require exploring a vast sequence space of roughly 10^8 to 10^{10} possibilities for a 9-residue to an 11-residue peptide (Fig. 4A-C). This capability makes HERMES a powerful platform for the de novo design of immunogenic peptides and neoantigens, offering a new computational path for the development of peptide-based T-cell vaccines against viruses and cancer.

Structural basis for TCR specificity

TCRs exhibit substantial degeneracy in their recognition, with some autoimmune T-cells experimentally shown to recognize over one million distinct peptides [64]. However, the full extent of TCR degeneracy for typical receptors remains unclear, largely due to the limitations in high-throughput experimental assays for TCR recognition of many peptides presented on different MHCs.

Our computational peptide design framework helps address this limitation, at least for the TCR-MHC complexes with a known structural template. Specifically, for a given TCR-MHC pair, we can us HERMES-*fixed* model to generate a peptide position weight matrix (PWM), representing the ensemble of peptides presented by the MHC and recognized by the TCR. We use the entropy of this PWM as a proxy for TCR degeneracy.

Fig. 5A shows the entropy of these peptide distributions for 105 TCR-MHC pairs across 10 human MHC-I alleles; see Table S6 for details on these structures. The median peptide entropy across these TCR-MHC pairs is approximately 10 bits, indicating that a typical TCR-MHC pair can recognize on the order of $\simeq 10^3$ peptides. In contrast, examining the degeneracy of MHC-I presentation alone in humans–using the peptide PWMs gath-

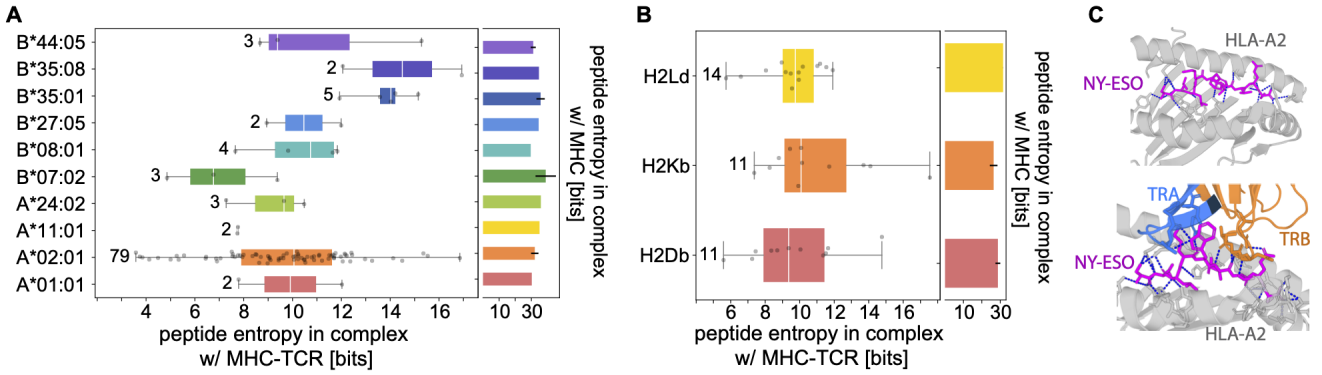

Figure 5. **Evaluating T-cell specificity from de novo designed peptides.** **(A)** T-cell specificity is measured by peptide entropy, which is computed from the HERMES-*fixed* predicted amino acid likelihood matrix, starting from a TCR-pMHC structural template. Each point on the plot corresponds to the entropy estimate from a distinct TCR-pMHC template, grouped by their MHC allele. The box plots display the distribution of peptide entropies for TCR-pMHC templates sharing the same MHC allele, with a white line marking the median and the box spanning the 25% to 75% quantiles. The number on the side of each box indicates the number of distinct structural templates we used for that MHC class; see Table S6 for details on these structures. The side plot shows peptide degeneracy for MHC presentation alone, evaluated as the entropy of the peptide position weight matrices gathered from the MHC Motif Atlas [54]. The error bars indicate the standard deviation of peptide degeneracies across motifs of different lengths, associated with a given MHC allele. **(B)** Similar to (A) but for degeneracy of peptides interacting with TCRs and MHCs in mice. **(C)** Polar bonds (blue dashed lines) between the NY-ESO peptide (magenta) and HLA-A*02:01 (gray) is shown at the top (PDB ID: 1S9W [63]). The additional polar bonds formed between the peptide and TRA (blue) and TRB (orange) of an interacting TCR are shown at the bottom (PDB ID: 2BNR [34]). These additional polar bonds limit the degeneracy of the peptides that can interact with a TCR-MHC complex compared to those presented by the same MHC.

ered from the MHC Motif Atlas [54]–yields a median entropy of 31 bits, implying that MHC-I molecules can typically recognize $\simeq 2 \times 10^9$ distinct peptides (Fig. 5A). A similar analysis on 36 TCR-MHC pairs spanning three mouse MHC-I alleles shows comparable trends, with a MHC-I degeneracy on the order of $\simeq 5 \times 10^8$ peptides (29 bits), and the TCR-MHC degeneracy of $\simeq 10^3$ peptides (10 bits); see Fig. 5B, and Table S6 for details on these structures. This pronounced reduction in entropy—observed by comparing the distribution of peptides recognized by MHC-I molecules alone to that recognized by TCR-MHC-I complexes—quantifies how TCRs constrain the accessible antigenic shape space, as illustrated by the polar interactions that they form with peptides in Fig. 5C.

Because our analyses focus on interactions involving typical TCR and MHC molecules, our entropy-based degeneracy estimates are smaller than those from previous theoretical studies, which often relied on experimental observations from autoimmune T-cells or limited synthetic peptide libraries [65]. Moreover, the HERMES-*fixed* model generates PWMs based on peptides whose backbone conformations match those of the wild-type template, likely underestimating the true degeneracy of TCR-MHC complexes. Although HERMES-*relaxed* can partially address this issue, its designs are only reliable within roughly five amino acids of the native peptides (Fig. 4), preventing a full exploration of the possible peptide poses for a given TCR-MHC complex. Nonetheless, our estimate can serve as informative lower bounds on typical MHC-I and TCR degeneracies.

IV. DISCUSSION

In this work, we introduced a structure-based machine learning model to predict TCR-pMHC interactions and to design novel reactive peptides for TCR-MHC complexes. Our approach builds on HERMES, a physically motivated equivariant neural network model that we previously developed [26, 27]. Trained on the protein universe, HERMES predicts amino acid propensities based on their atomic environments in protein structures, and has been shown to implicitly learn biophysical interatomic interactions in proteins [27], enabling it to make highly generalizable predictions. This generalizability is particularly important for TCR-pMHC complexes, where only a modest number of experimentally resolved structures are available, and computational algorithms often struggle to accurately predict new structures [17, 18].

We found that HERMES-predicted peptide energies closely tracked both TCR-pMHC binding affinities and peptide-induced T-cell activities across diverse systems, including variants from a CMV-derived viral epitope [36], the NY-ESO cancer-testis antigen [43, 44], and a melanoma self-antigen [36]. Notably, HERMES also accurately predicted the activity of p53-derived neoepitope variants in the context of a bispecific antibody-mediated T-cell therapy [5]. This result underscores the algorithm's broad applicability and generalizability:

given a reliable structural template close to the system of interest, HERMES can predict peptide-induced T-cell responses, whether through direct TCR interactions or antibody-mediated mechanisms.

Based on the HERMES predictions for TCR-pMHC reactivities, we presented a computational protocol for de novo design of immunogenic peptides that can elicit robust T-cell responses. Starting from the structural template of a TCR-MHC with a native (wild-type) peptide, we developed an MCMC search guided by HERMES peptide energies to explore the vast sequence space of roughly 10^8 (10^{10}) possibilities for 9-residue (11-residue) peptides, to identify immunogenic candidates. Experimental validation across three systems—with native peptides from NY-ESO and MAGE cancer epitopes, plus a viral epitope from EBV—revealed that our designs reliably activated T-cells, even with up to five substitutions from the native peptide used in the structural template. Notably, several MAGE-derived variants outperformed or matched the native peptide in T-cell activation, demonstrating the method's potential to enhance peptide potency. These results show the potential of HERMES for designing peptide-based cancer vaccines with neoantigens, providing an avenue for personalized immunotherapy strategies.

Accurately modeling T-cell specificity remains challenging, in part due to the limited functional data on TCR-epitope pairs, and the highly cross-reactive nature of TCR-epitope interactions. A design protocol like the one presented here, coupled with experimental validation, offers a promising path forward for future studies. By iteratively generating candidate peptide libraries, testing them experimentally, and refining the model with each iteration, it should be possible to more efficiently explore the high dimensional antigenic sequence space. This active learning strategy has a potential for revealing the "shape space" of TCR-epitope interactions, with broad applicabilities for early disease diagnosis, and for design of targeted immunotherapies. Moreover, this approach can help mitigate immunotherapy-related toxicity by identifying and excluding potential self-reactivities–beyond single point mutations in the target epitope–such as the off-target Titin reactivity observed in an engineered T-cell therapy against MAGE-A3 that led to a severe cardiac toxicity [57].

Although our findings highlight the promise of structure-based models for immune recognition, a key limitation remains the scarcity of high-quality TCR-pMHC structural templates, despite advances from tools like AlphaFold3 [29] and TCRDock [17]. On the other hand, while TCR sequences are more widely available, sequence-based models often lack generalizability because existing paired TCR-pMHC data are skewed toward just a few peptides [11, 12]. A promising direction for future work lies in developing multi-modal flexible methods that integrate structural and sequence-based models, leveraging both the depth and generalizability offered by structural information and the scalability of sequence data.

Although our study primarily examines MHC class I-restricted peptides, the broad applicability of structure-based modeling suggests that similar methods could be adapted for the more complex TCR-peptide-MHCII interactions, thereby offering new insights into CD4+ T-cell responses. Moreover, incorporating additional immunological factors—such as TCR clustering, co-receptor interactions, cross-reactivity with self-antigens, and functional immune profiling—can further enhance the accuracy of T-cell response predictions and improve the resulting de novo peptide design pipeline.

V. DATA AND CODE AVAILABILITY

Data and code for all the analyses can be found in the Github repository: https://github.com/StatPhysBio/tcr_antigen_design. Code for the original HERMES [26] model can be found in the Github repository: https://github.com/StatPhysBio/hermes.

ACKNOWLEDGMENT

We thank Mikhail V. Pogorelyy for valuable comments and help with CellCyte. This work has been supported by the National Institutes of Health MIRA awards R35 GM142795 (AN, GV, MNP) and R35 GM141457 (PB), the CAREER award from the National Science Foundation grant 2045054 (AN), the National Institutes of Health R01 award AI136514 (PT and PB), the Royalty Research Fund from the University of Washington no. A153352 (AN, MNP), the Allen School Computer Science & Engineering Research Fellowship from the Paul G. Allen School of Computer Science & Engineering at the University of Washington (GV), the Microsoft Azure award from the eScience institute at the University of Washington (AN, GV, MNP), the TIRTL Bluesky Initiative at St. Jude Children's Research Hospital (PT), and the ALSAC at St. Jude Children's Research Hospital (PT and AAM). This work is also supported, in part, through the Departments of Physics and Computer Science and Engineering, and the College of Arts and Sciences at the University of Washington (AN, GV, MNP). This work was performed in part during the 2023 summer workshop "Statistical Physics and Adaptive Immunity" at the Aspen Center for Physics, which is supported by National Science Foundation grant PHY-2210452. This work also benefited from discussions during the 2024 program "Interactions and Co-evolution between Viruses and Immune Systems" at the Kavli Institute for theoretical physics (KITP), which is supported by National Science Foundation grants PHY-2309135, and PHY-2309135, and the Gordon and Betty Moore Foundation Grant No. 2919.02.

COMPETING INTEREST STATEMENT

PT is on the Scientific Advisory Board of Immunoscape and Shennon Bio, has received research support and personal fees from Elevate Bio, and consulted for 10X Genomics, Illumina, Pfizer, Cytoagents, Sanofi, Merck, and JNJ.

[1] Attaf M, Huseby E, Sewell AK (2015) $\alpha\beta$ T cell receptors as predictors of health and disease *Cell. Mol. Immunol.* 12:391–399.

[2] Liu B, et al. (2024) Design of high specificity binders for peptide-MHC-I complexes *bioRxiv* 2024.11.28.625793.

[3] Leidner R, et al. (2022) Neoantigen T-cell receptor gene therapy in pancreatic cancer *N. Engl. J. Med.* 386:2112–2119.

[4] Mirazee JM, et al. (2024) Enhancing CAR T cell therapy through in silico modeling of hinge sequence length based on epitope location *Biophys. J.* 123:552a.

[5] Hsiue EHC, et al. (2021) Targeting a neoantigen derived from a common TP53 mutation *Science* 371:eabc8697.

[6] Blass E, Ott PA (2021) Advances in the development of personalized neoantigen-based therapeutic cancer vaccines *Nat. Rev. Clin. Oncol.* 18:215–229.

[7] Rojas LA, et al. (2023) Personalized RNA neoantigen vaccines stimulate T cells in pancreatic cancer *Nature* 618:144–150.

[8] Zhang SQ, et al. (2018) High-throughput determination of the antigen specificities of T cell receptors in single cells *Nat. Biotechnol.* 36:1156–1159.

[9] Nolan S, et al. (2025) A large-scale database of T-cell receptor beta (TCRβ) sequences and binding associations from natural and synthetic exposure to SARS-CoV-2 *Front. Immunol.* 16:1488851.

[10] Dash P, et al. (2017) Quantifiable predictive features define epitope-specific T cell receptor repertoires *Nature* 547:89–93.

[11] Vita R, et al. (2019) The Immune Epitope Database (IEDB): 2018 update *Nucleic Acids Res.* 47:D339–D343.

[12] Shugay M, et al. (2018) VDJdb: a curated database of T-cell receptor sequences with known antigen specificity *Nucleic Acids Res.* 46:D419–D427.

[13] Isacchini G, Walczak AM, Mora T, Nourmohammad A (2021) Deep generative selection models of T and B cell receptor repertoires with soNNia *Proc. Natl. Acad. Sci. U. S. A.* 118:e2023141118.

[14] Weber A, Born J, Rodriguez Martínez M (2021) TITAN: T-cell receptor specificity prediction with bimodal attention networks *Bioinformatics* 37:i237–i244.

[15] Meynard-Piganeau B, Feinauer C, Weigt M, Walczak AM, Mora T (2024) TULIP: A transformer-based unsupervised language model for interacting peptides and T cell receptors that generalizes to unseen epitopes *Proc. Natl. Acad. Sci. U. S. A.* 121:e2316401121.

[16] Piepenbrink KH, Gloor BE, Armstrong KM, Baker BM (2009) Methods for quantifying T cell receptor binding affinities and thermodynamics *Methods Enzymol.* 466:359–381.

[17] Bradley P (2023) Structure-based prediction of T cell receptor:peptide-MHC interactions *Elife* 12:e82813.

[18] Yin R, et al. (2023) TCRmodel2: high-resolution modeling of T cell receptor recognition using deep learning *Nucleic Acids Res.* 51:569–576.

[19] Rives A, et al. (2021) Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences *Proc. Natl. Acad. Sci. U. S. A.* 118:e2016239118.

[20] Lin Z, et al. (2023) Evolutionary-scale prediction of atomic-level protein structure with a language model *Science* 379:1123–1130.

[21] Madani A, et al. (2023) Large language models generate functional protein sequences across diverse families *Nat. Biotechnol.* 41:1099–1106.

[22] Schmirler R, Heinzinger M, Rost B (2024) Fine-tuning protein language models boosts predictions across diverse tasks *Nat. Commun.* 15:7407.

[23] Wang Y, et al. (2024) An explainable language model for antibody specificity prediction using curated influenza hemagglutinin antibodies *Immunity* 57:2453–2465.e7.

[24] Hie BL, et al. (2024) Efficient evolution of human antibodies from general protein language models *Nat. Biotechnol.* 42:275–283.

[25] Bravi B, et al. (2023) A transfer-learning approach to predict antigen immunogenicity and T-cell receptor specificity *Elife* 12:e85126.

[26] Visani GM, et al. (2024) HERMES: Holographic Equivariant neuRal network model for Mutational Effect and Stability prediction *bioRxiv* 2024.07.09.602403.

[27] Pun MN, et al. (2024) Learning the shape of protein microenvironments with a holographic convolutional neural network *Proc. Natl. Acad. Sci. U. S. A.* 121:e2300838121.

[28] Chaudhury S, Lyskov S, Gray JJ (2010) PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta *Bioinformatics* 26:689–691.

[29] Abramson J, et al. (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3 *Nature* 630:493–500.

[30] Dauparas J, et al. (2022) Robust deep learning-based protein sequence design using ProteinMPNN *Science* 378:49–56.

[31] Blaabjerg LM, et al. (2023) Rapid protein stability prediction using deep learning representations *Elife* 12:e82593.

[32] Michalewicz K, Barahona M, Bravi B (2024) ANTIPASTI: Interpretable prediction of antibody binding affinity exploiting normal modes and deep learning *Structure* 32:2422–2434.e5.

[33] Diaz DJ, et al. (2024) Stability Oracle: a structure-based graph-transformer framework for identifying stabilizing mutations *Nat. Commun.* 15:6170.

[34] Chen JL, et al. (2005) Structural and kinetic basis for heightened immunogenicity of T cell vaccines *J. Exp. Med.* 201:1243–1255.

[35] Garboczi DN, et al. (1996) Structure of the complex between human T-cell receptor, viral peptide and HLA-A2 *Nature* 384:134–141.

[36] Łuksza M, et al. (2022) Neoantigen quality predicts immunoediting in survivors of pancreatic cancer *Nature* 606:389–395.

[37] Yang X, et al. (2015) Structural basis for clonal diversity of the public T cell response to a dominant human Cytomegalovirus Epitope *J. Biol. Chem.* 290:29106–29119.

[38] Davis MM, et al. (1998) Ligand recognition by alpha beta T cell receptors *Annu. Rev. Immunol.* 16:523–544.

[39] Zhong S, et al. (2013) T-cell receptor affinity and avidity defines antitumor response and autoimmunity in T-cell immunotherapy *Proc. Natl. Acad. Sci. U. S. A.* 110:6973–6978.

[40] Soto CM, et al. (2013) MHC-class I-restricted CD4 T cells: a nanomolar affinity TCR has improved anti-tumor efficacy in vivo compared to the micromolar wild-type TCR *Cancer Immunol. Immunother.* 62:359–369.

[41] Li Y, et al. (2005) Directed evolution of human T-cell receptors with picomolar affinities by phage display *Nat. Biotechnol.* 23:349–354.

[42] Dunn SM, et al. (2006) Directed evolution of human T cell receptor CDR2 residues by phage display dramatically enhances affinity for cognate peptide-MHC without increasing apparent cross-reactivity *Protein Sci.* 15:710–721.

[43] Aleksic M, et al. (2010) Dependence of T cell antigen recognition on T cell receptor-peptide MHC confinement time *Immunity* 32:163–174.

[44] Pettmann J, et al. (2021) The discriminatory power of the T cell receptor *Elife* 10:e67092.

[45] Esfandiary A, Ghafouri-Fard S (2015) New York esophageal squamous cell carcinoma-1 and cancer immunotherapy *Immunotherapy* 7:411–439.

[46] Chen JL, et al. (2000) Identification of NY-ESO-1 peptide analogues capable of improved stimulation of tumor-reactive CTL *J. Immunol.* 165:948–955.

[47] Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks *Proc. Natl. Acad. Sci. U. S. A.* 89:10915–10919.

[48] Altan-Bonnet G, Germain RN (2005) Modeling T cell antigen discrimination based on feedback control of digital ERK responses *PLoS Biol.* 3:e356.

[49] Achar SR, et al. (2022) Universal antigen encoding of T cell activation from high-dimensional cytokine dynamics *Science* 376:880–884.

[50] Stone JD, Chervin AS, Kranz DM (2009) T-cell receptor binding affinities and kinetics: impact on T-cell activity and specificity *Immunology* 126:165–176.

[51] Gras S, et al. (2010) Allelic polymorphism in the T cell receptor and its impact on immune responses *J. Exp. Med.* 207:1555–1567.

[52] Liu YC, et al. (2014) A molecular basis for the interplay between T cells, viral mutants, and human leukocyte antigen micropolymorphism *J. Biol. Chem.* 289:16688–16698.

[53] Raman MCC, et al. (2016) Direct molecular mimicry enables off-target cardiovascular toxicity by an enhanced affinity TCR designed for cancer immunotherapy *Sci. Rep.* 6:18851.

[54] Tadros DM, Eggenschwiler S, Racle J, Gfeller D (2023) The MHC Motif Atlas: a database of MHC binding specificities and ligands *Nucleic Acids Res.* 51:D428–D437.

[55] Vormehr M, Türeci, Ö, Sahin U (2019) Harnessing tumor mutations for truly individualized cancer vaccines *Annu. Rev. Med.* 70:395–407.

[56] Song K, Pun SH (2024) Design and evaluation of synthetic delivery formulations for peptide-based cancer vaccines *BME Front.* 5:0038.

[57] Linette GP, et al. (2013) Cardiovascular toxicity and titin cross-reactivity of affinity-enhanced T cells in myeloma and melanoma *Blood* 122:863–871.

[58] Cameron BJ, et al. (2013) Identification of a Titin-derived HLA-A1-presented peptide as a cross-reactive target for engineered MAGE A3-directed T cells *Sci. Transl. Med.* 5:197ra103.

[59] Lundegaard C, et al. (2008) NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11 *Nucleic Acids Res.* 36:W509–12.

[60] Lundegaard C, Lund O, Nielsen M (2008) Accurate approximation method for prediction of class I MHC affinities for peptides of length 8, 10 and 11 using prediction tools trained on 9mers *Bioinformatics* 24:1397–1398.

[61] Andreatta M, Nielsen M (2016) Gapped sequence alignment using artificial neural networks: application to the MHC class I system *Bioinformatics* 32:511–517.

[62] Roney JP, Ovchinnikov S (2022) State-of-the-Art Estimation of Protein Model Accuracy Using AlphaFold *Phys. Rev. Lett.* 129:238101.

[63] Webb AI, et al. (2004) Functional and structural characteristics of NY-ESO-1-related HLA A2-restricted epitopes and the design of a novel immunogenic analogue *J. Biol. Chem.* 279:23438–23446.

[64] Wooldridge L, et al. (2012) A single autoimmune T cell receptor recognizes more than a million different peptides *J. Biol. Chem.* 287:1168–1177.

[65] Sewell AK (2012) Why must T cells be cross-reactive? *Nat. Rev. Immunol.* 12:669–677.

Supplementary Information

T-cell receptor specificity landscape revealed through de-novo peptide design

Gian Marco Visani, Michael N. Pun, Anastasia A. Minervina, Philip Bradley, Paul Thomas, and Armita Nourmohammad

Data and code availability. Data and code for all the analyses can be found in the Github repository: https://github.com/StatPhysBio/tcr_antigen_design. Code for the original HERMES [30] model can be found in the Github repository: https://github.com/StatPhysBio/hermes.

I. HERMES NEURAL NETWORK MODEL FOR AMINO ACID PROPENSITIES

HERMES is a self-supervised structure-based machine learning model for proteins [30, 31]. HERMES has a 3D rotationally equivariant architecture and is trained to predict a residue's amino acid identity from its surrounding atomic environment within a 10 Å of the central residue's C-α in the 3D structure. The input to the neural network is a point cloud of atoms from a given structural neighborhood, in which all atoms associated with the central residue are masked. These point clouds are featurized by atom types, including computationally-added hydrogens, partial charge, and Solvent Accessible Surface Area (SASA). The point clouds are first projected onto the orthonormal Zernike Fourier basis, centered at the position of the (masked) central residue's C-α (Fig. 1A). The resulting *holographic* projections are fed to a stack of SO(3)-equivariant layers, the output of which is SO(3)-invariant, and is then passed through a multi-layer perceptron (MLP) to generate the desired predictions. The MLP's outputs are termed pseudo-energies, associated with the 20 different possible amino acids that can be placed at the masked residue, and act as logits within a softmax function to compute probabilities of each amino acid.

Each HERMES model is an ensemble of 10 individually-trained architectures. In this work, we used HERMES models trained on structural data with different levels of added noise: 0 noise (the original PDB structure), and 0.5 noise, which is obtained by adding Gaussian noise to the 3D coordinates of the original structure, with standard deviation of 0.5 Å. We only use models that were trained on the wild-type amino acid classification task [30]. Models were pre-trained on structures from 30% similarity split of ProteinNet's CASP12 set [32], resulting in around ∼10k training structures.

We refer the reader to refs. [30, 31] for further details on the architecture, training procedure, training data preparation, and the mathematical introduction to SO(3)-equivariant models in Fourier space.

II. TCR-PMHC BINDING AFFINITY AND T-CELL ACTIVITY PREDICTION

Peptide energy calculation with HERMES. To predict the affinity of TCR-pMHC complexes or the activity of TCRs induced by their interactions with a pMHC complex, we introduce the predicted energy for a peptide σ, given its surrounding TCR and MHC,

$$E_{\text{peptide}}(\sigma; \text{TCR, MHC}) = \sum_{i=1}^{\ell} E(\sigma_i; \text{TCR, MHC}, \sigma_{/i}). \tag{S1}$$

We assume that each peptide residue contributes linearly to the total energy by the amount $E(\sigma_i; \text{TCR, MHC}, \sigma_{/i})$, where σ_i is the amino acid at position i of the peptide. A residue's energy contribution E is evaluated as its associated logit value from HERMES, taking in as input the atomic composition of the surrounding TCR, MHC and the rest of the peptide $\sigma_{/i}$ (Fig. 1A).

To compute peptide energies, HERMES operates on the structural composition of the protein complex. However, data on the impact of substitutions on the TCR-pMHC structure is limited, and computational models often fail to capture subtle conformational changes in the structure due to amino acid substitutions in a peptide. To address this issue, we adopt two approaches to estimate energy for a diverse array of peptides for a given TCR-MHC complex:

- HERMES-*fixed*: In the simplest approach, we select the TCR-pMHC structure with matching TCR and MHC to our query and with the peptide sequence that most closely matches the peptide of interest and directly input it into HERMES. The peptide energy is subsequently calculated as described in eq. S1 (Fig. 1B). This method does

not modify the underlying structure, effectively assuming that any amino acid substitutions do not significantly alter the protein conformation.

- HERMES-*relaxed*: Since amino acid substitutions can locally reshape the protein complex, we introduce a more involved protocol to account for these structural changes. We begin with the available crystal structure of the TCR-pMHC complex that closely matches our query as template. We then mutate the original peptide to the desired state and relax the structure using PyRosetta's cartesian_ddg fastrelax protocol [33]. During relaxation, side chain and backbone atoms of the peptide are allowed to move, while only the side chain atoms of the MHC and TCR chains within a 10Å of the peptide are flexible. We then calculate the peptide energy with HERMES (eq.S1), using the relaxed structure of the mutated variant as input (Fig. 1B). Since PyRosetta's relaxation procedure is stochastic in nature, we average peptide energy across 100 realizations of the relaxation.

Evaluation of EC_{50} from dose response measurements against peptide libraries. Ref. [34] presents dose response measurements for fraction of 4-1BB$^+$ CD8$^+$ T-cells across varying peptide concentrations. These measurements are done for peptide libraries of single point mutants from a given wild-type peptide, across 7 different TCR-MHC systems: three TCRs recognizing variants of the highly immunogenic HLA-A02:01-restricted human cytomegalovirus (CMV) epitope NLVPMVATV (NLV); three TCRs recognizing variants of the weaker HLA-A02:01-restricted melanoma gp100 self-antigen IMDQVPFSV; and one TCR recognizing variants of the weakly immunogenic HLA-B*27:05-restricted pancreatic cancer neo-peptide GRLKALCQR [34].

The half-maximal effective concentration (EC_{50}) from the dose response curves of 4-1BB$^+$ CD8$^+$ T-cell fractions can quantify the level of T-cell activity in response to each peptide. We used our own fits to estimate the EC_{50} associated with T-cell responses. Specifically, we fitted a Hill function with background to the dose response measurements for each peptide from ref. [34],

$$f(x) = \frac{(A - A_0)x^n}{x^n + K^n} + A_0 \tag{S2}$$

where $f(x)$ is the fraction of 4-1BB$^+$ CD8$^+$ T-cells pulsed by a peptide at concentration x, A is the response amplitude, A_0 is the background activation, n is the Hill coefficient, and K is the half-maximal effective concentration EC_{50}. Similar to ref. [34], we regularized our fitted Hill functions, whereby the deviation of the amplitude A and the Hill coefficient n from 1, and the background activation A_0 from 0 are penalized. The code to fit these curves is available in our github repository.

Computational modeling of TCR-pMHC templates in ref. [34] with AlphaFold3. While there exists a crystal structure of TCR1 in complex with our the correct pMHC (PDB ID: 5D2N [35]), the same is not true for TCRs 2-7. For these, we resort to structure prediction with AlphaFold3 (AF3) [36]. Specifically, we fold the TCR-pMHC structures using the wild-type sequence only, resulting in one wild-type structure per system. We ran predictions using the AlphaFold Server, with no random seed, on the sequences listed in Table S3. Of the 5 output structures predicted by AF3, we considered the highest confidence structure based on the Predicted Alignment Error (PAE) score of the TCR-pMHC interface.

III. DE-NOVO PEPTIDE DESIGN

Peptide design algorithm. Given a TCR and an MHC, we use HERMES to design de novo peptides that can form a stable TCR-pMHC complex, and potentially, elicit a T-cell response. Our approach starts with an existing template TCR-pMHC structure. We refer to the peptide in the template as wild-type; see Table S5 for the templates we use in our analyses. Similar to evaluating the peptide energy, we use two pipelines to sample novel peptides:

- HERMES-*fixed*: This is our basic pipeline, in which we generate novel peptide sequences by conditioning on the fixed TCR-pMHC template structure. Specifically, we sequentially mask the atoms of each amino acid along the peptide and then use HERMES to compute the probability $P(\sigma_i|\mathbf{x}_i)$ of different amino acid types σ_i at the masked residue i, conditioned on the surrounding atomic neighborhood $\mathbf{x}_i$, within a 10 Å radius of the masked residue's C-α. The neighborhood $\mathbf{x}_i$ can include atoms from the TCR, the MHC, and the other amino acids in the peptide. Repeating this procedure for every residue along the peptide allows us to construct a Position Weight Matrix (PWM) that represents the probabilities of different amino acids at each peptide position, conditioned on the surrounding structural neighborhood. We generate peptides using multinomial sampling from the PWM; see the schematic in Fig. 1C.

- HERMES-*relaxed*: This pipeline incorporates simulated annealing and Markov chain Monte Carlo (MCMC) sampling to account for structural changes resulting from peptide substitutions. We start with the template TCR-MHC structure and an initial random peptide sequence $\sigma^{(0)}$, which is *in-silico* packed via PyRosetta, followed by the Relax protocol, yielding a relaxed structure $\mathbf{x}^{(0)}$. During relaxation, both side chain and backbone atoms of the peptide are allowed to move, while only the side chain atoms of the MHC and TCR within 10 Å of the peptide are flexible. Following the HERMES-*fixed* protocol, we then evaluate the PWM associated with this relaxed structure and use it to generate a new peptide sequence, $\sigma^{(1)}$. This sequence is again packed and relaxed, producing a new structure, $\mathbf{x}^{(1)}$. Using eq. S1, we evaluate the energies of the peptides in their respective (relaxed) structures as $E(\sigma^{(0)};\mathbf{x}^{(0)})$ and $E(\sigma^{(1)};\mathbf{x}^{(1)})$, and apply the Metropolis-Hastings criterion to accept the transition from $\sigma^{(0)}$ to $\sigma^{(1)}$ with probability:

$$P_{\sigma^{(0)}\to\sigma^{(1)}} = \min\left(1, e^{-\frac{E(\sigma^{(1)};\mathbf{x}^{(1)})-E(\sigma^{(0)};\mathbf{x}^{(0)})}{T}}\right) \tag{S1}$$

where T is the temperature. This process is iterated, incrementally reducing T at each step, until convergence. The peptide obtained at the final iteration represents the protocol's de novo design. Repeating this procedure many times produces a diverse set of peptides spanning various distances from the wild-type sequence; see the schematic in Fig. 1C and Algorithm S1 for more details, and the the convergence curves in Fig. S3.

Input: TCR-pMHC structure $\mathbf{x}$, maximum number of iterations N;
Output: peptide sequence $\sigma^{(N)}$;
$T_0 \leftarrow 10$;
$T_f \leftarrow 0$;
$a \leftarrow 0.95$;
$T^{(0)} \leftarrow T_0$;
$\sigma^{(0)} \leftarrow$ random sequence;
$\mathbf{x}^{(0)} \leftarrow \text{RosettaMutate}(\mathbf{x}^{(0)}, \sigma^{(0)})$;
$\mathbf{x}^{(0)} \leftarrow \text{RosettaFastRelax}(\mathbf{x}^{(0)})$;
$E^{(0)}_{\text{peptide}} \leftarrow \sum_{i\in\mathcal{R}_{\text{peptide}}} E(\sigma^{(0)}_i; \mathbf{x}^{(0)}_i)$;
for $k = 1, 2, \ldots, N$ **do**
  for $i \in \mathcal{R}_{peptide}$ **do**
    $\sigma^{(\text{temp})}_i \sim \text{Multinomial}(\text{softmax}([E(\text{aa}; \mathbf{x}^{(k-1)}_i)/T^{(k)}]_{\text{aa}\in\text{amino-acids}}))$;
  end
  $\mathbf{x}^{(\text{temp})} \leftarrow \text{RosettaMutate}(\mathbf{x}^{(k-1)}, \sigma^{(\text{temp})})$;
  $\mathbf{x}^{(\text{temp})} \leftarrow \text{RosettaFastRelax}(\mathbf{x}^{(\text{temp})})$;
  $E^{(\text{temp})}_{\text{peptide}} \leftarrow \sum_{i\in\mathcal{R}_{\text{peptide}}} E(\sigma^{(\text{temp})}_i; \mathbf{x}^{(\text{temp})}_i)$;
  $p \sim \text{Uniform}(0, 1)$;
  if $p < \exp(-(E^{(temp)}_{peptide} - E^{(k-1)}_{peptide})/T^{(k)})$ **then**
    $E^{(k)}_{\text{peptide}} \leftarrow E^{(\text{temp})}_{\text{peptide}}$;
    $\mathbf{x}^{(k)} \leftarrow \mathbf{x}^{(\text{temp})}$;
    $\sigma^{(k)} \leftarrow \sigma^{(\text{temp})}$;
  else
    $E^{(k)}_{\text{peptide}} \leftarrow E^{(k-1)}_{\text{peptide}}$;
    $x^{(k)} \leftarrow \mathbf{x}^{(k-1)}$;
    $\sigma^{(k)} \leftarrow \sigma^{(k-1)}$;
  end
  $T^{(k+1)} \leftarrow (T_0 - T_f)a^k + T_f$;
end

Algorithm S1: HERMES-*relaxed* protocol for generating TCR-MHC-specific peptide sequences. For brevity, we define $\mathbf{x}_i \equiv (\text{TCR}, \text{MHC}, \sigma_{/i})$ and drop the arguments of E_{peptide}.

Filtering metrics for peptide design with HERMES and benchmarking of different design approaches. For each system, we first selected designs predicted by NetMHCPan to be at least weak binders to their target MHCs (EL_rank ≤ 2.0) [37]. We then used TCRdock [38] to filter peptides for reactivity. TCRdock is a structure prediction algorithm built on top of AlphaFold2 [39], and refined to model TCR-pMHC interactions. Its PAE score for the TCR-pMHC interface can be used to distinguish reactive from decoy peptides [38], and therefore, can provide a metric for filtering our designed peptides; other work has similarly used the PAE of structure-prediction algorithms to filter designed protein binders [40]. Using TCRdock, we retained those peptides whose PAE scores at the TCR-pMHC interface fell below a threshold, as illustrated in Figs. S9, S10, and S11). The distribution of PAE scores is highly system specific [40], so we set each system's threshold near the TCRdock PAE value of its *wild-type* peptides in complex with the respective TCR-MHC, as the wild types are known binders. Among the filtered peptides, we selected 93 for experimental validation, ensuring that designs from both HERMES-*fixed* and HERMES-*relaxed*, as well as from models with 0.0 and 0.5Å noise levels were represented. We also enforced a minimum difference of 2-3 amino acids (depending on the system) between the designs and the wild-type peptides.

It should be noted that the AF3 PAE score can also discriminate between reactive and decoy peptides (Figs. S5, S12), making it also suitable for filtering. However, a full-access implementation of AF3 was unavailable before our experimental validations. Therefore, we used AF3 only as a secondary filter for designs in the EBV and MAGE systems, and for benchmarking against TCRdock's PAE score for the designed peptides (Figs. S5); see details below.

Design parameters for different TCR-pMHC systems. Here, we provide the details on the specifics of peptide selection for each TCR-MHC system. It should be noted that as we progressed from NY-ESO to EBV and MAGE, we chose to explore the sequence space farther from the wild-type peptides, leading to more challenging designs.

- *NY-ESO:* The wild-type peptides (NYESO: SLLMWITQC and NYESO V9C: SLLMWITQV) have TCRdock PAEs of 5.17 and 5.18, respectively. As positive designs, we selected 58 peptides with the lowest TCRdock PAE values (all below 5.35) and a Hamming distance of at least two amino acids from their respective wild-type sequences. For negative designs, we uniformly sampled 35 peptides among those designed by HERMES whose TCRdock PAE values fell between 5.5 and 7.0. We did not enforce all HERMES models to be represented equally among these designs, nor did we enforce equal representation of the template structures. However, we note that all models are represented among positive designs (Fig. S6).

- *EBV:* The wild-type peptides (HPVG: HPVGEADYFEY and HPVG E5Q: HPVGQADYFEY) have TCRdock PAEs of 4.90 and 4.82, and AF3 PAEs of 1.15 and 1.13, respectively. We selected 93 positive peptides using a TCRdock's PAE score cutoff of 5.1 and AF3 PAE score cutoff of 2.0. Specifically, we sampled uniformly below these cutoffs with the following design protocols: HERMES-*fixed* at both noise levels using the HPVG structure (3MV7) as template (24 peptide), HERMES-*fixed* at both noise levels, using the HPVG E5Q structure (4PRP) as template (23 designs); HERMES-*relaxed* at both noise levels, using the HPVG structure (3MV7) as template (24 peptides), and HERMES-*relaxed* at both noise levels using the HPVG E5Q structure (4PRP) as template (23 designs); see Fig. S7.

- *MAGE:* The wild-type peptides (MAGE-A3: EVDPIGHLY and Titin: ESDPIVAQY) have TCRdock PAEs of 5.21 and 5.03, and AF3 PAEs of 1.00 and 0.98, respectively. We selected 93 positive peptides using a TCRdock's PAE score cutoff of 5.2; the AF3 PAE scores for all of these designed peptide are comparable to that of the wild-types. For these designs, we sampled uniformly below the PAE cutoffs with the following design protocols: HERMES-*fixed* at both noise levels using MAGE-A3 structure (5BRZ) as template (15 peptides); HERMES-*fixed* at both noise levels using the MAGE-A3 structure (5BRZ) as template with E_1 fixed (15 peptides); HERMES-*relaxed* at both noise levels using the MAGE-A3 structure (5BRZ) as template (16 peptides); HERMES-*relaxed* at bothsee Fig. S7 noise levels using the MAGE-A3 structure (5BRZ) as template with E_1 fixed (15 peptides); HERMES-*fixed* at both noise levels and using the Titin structure (5BS0) as template (16 peptides); HERMES-*relaxed* at both noise levels using the Titin structure (5BS0) as template (16 peptides); see Fig. S8.

Characterizing the size of the sequence space for design. The size of the sequence space to explore grows exponentially with increasing distance from the wild-type peptide. If no restriction is imposed (i.e., all amino acids are permissible with equal probabilities), the size of the sequence space follows,

$$\Omega_0(d) = \binom{\ell}{d} 19^d \tag{S2}$$

where ℓ is the length of the peptide, and d is the distance from the wild-type sequence.

MHC presentation imposes constraints on amino acid usages in peptides, especially at the anchor residues. Let $p_i(a)$ denote the probability of amino acid a at position i of the peptide, e.g., set by the sequence-specific preferences

of the MHC allele. The probability of introducing a mutation at position i depends on the prominence of the wild-type amino acid, and is equal to $q_i = 1 - p_i(\text{wt})$. The typical number of non-wildtype amino acids realized at position i can be estimated by $n_i = 2^{S_i}$, where $S_i = -\sum_{a \neq \text{wt}} \hat{p}_i(a) \log_2 \hat{p}_i(a)$ is the entropy associated with the non-wildtype amino acid frequencies, and $\hat{p}_i(a) = \frac{p_i(a)}{1-p_i(\text{wt})}$ is the normalized probability of amino acid a, obtained by excluding the wild-type from the set. Therefore, the typical size of the sequence space at Hamming distance d from the wild-type peptide—subject to constraints imposed by the amino acid frequencies $p_i(a)$ (e.g., due to MHC restriction)—is given by,

$$\Omega(d) = \sum_{\{i_1,\ldots,i_d\}} (q_{i_1} n_{i_1}) \ldots (q_{i_d} n_{i_d}) \equiv \sum_{\{i_i,\ldots,i_d\}} 2^{\hat{S}_{i_1} + \cdots + \hat{S}_{i_d}} \tag{S3}$$

where the summation runs over all $\binom{\ell}{d}$ ways to choose d distinct sites at which mutations may occur, and $q_i n_i$ is the expected number of substitutions associated with site i—evaluated as the likelihood of mutating site i times the typical number of viable amino acid replacements at that site—and $\hat{S}_i = S_i + \log_2 q_i$ is the resulting adjusted entropy at site i.

In principle, evaluating $\Omega(d)$ in eq. S3 involves summation over a combinatorially large number of terms, which can become computationally prohibitive. In practice, however, many of the sites have the same adjusted entropy, which simplifies the evaluations of $\Omega(d)$. For example, the adjusted entropies for mutations away from the 9 amino acid NY-ESO peptide restricted by HLA-A*02:01 yield the following (rounded) values: $[4, 2, 4, 4, 4, 3, 4, 4, 2]$ in bits, indicating six sites with 4 bits of entropy (or $2^4 = 16$ typical substitutions), one site with 3 bits (8 typical substitutions), and two sites with 2 bits (4 typical substitutions). To demonstrate, in this case, the number of sequences at distance $d = 5$ from the wild-type follows,

$$\begin{aligned}\Omega_{\text{NYESO}}(d=5) &= \binom{6}{5}16^5 + \binom{6}{4}16^4 \times \left[\binom{2}{1}4^1 + 8^1\right] + \binom{6}{3}16^3 \times \left[\binom{2}{1}4^1 \times 8^1 + \binom{2}{2}4^2\right] + \binom{6}{2}16^2 \times \binom{2}{2}4^2 \times 8^1 \\ &= 29,065,216\end{aligned} \tag{S4}$$

Similar calculations can be done to characterize the size of the sequence space at varying distances from the wild-type, as shown in Fig. 4.

IV. ALTERNATIVE COMPUTATIONAL METHODS FOR SCORING PEPTIDES AND DESIGN

We compare our approach to alternative methods (listed below), based on which we characterize peptide scores akin to HERMES's peptide energy, and when possible, use them to design novel peptides. A detailed comparison between the accuracy of different methods in predicting the affinity of TCR-pMHC interactions and the peptide-induced activation of T-cells is presented in Tables S2 and S4. Comparison between the fidelity of the designs–only according to computational metrics–is shown in Figures S9, S10, and S11. All the designed peptides as well as the code to generate samples are available in our github repository.

- *BLOSUM62 substitution matrix* [41]: The BLOSUM62 substitution matrix characterizes the favorability of substitutions between all pairs of amino acids, and has been used widely to characterize the mutational effects in proteins. As our first (and simplest) benchmark, we used BLOSUM62 to define a peptide energy (score). Specifically, we score a peptide by summing over the BLOSUM62 scores associated with the amino acid substitutions from the wild-type peptide σ^{wt} (the one present in the template structure) to the query peptide sequence σ that we are interested in,

$$E^{(\text{B62})}_{\text{peptide}}(\sigma; \sigma^{\text{wt}}) = \sum_{i=1}^{\ell} B62[\sigma_i^{\text{wt}} \to \sigma_i] \tag{S1}$$

  Notably, the BLOSUM62 predicted peptide energy only depends on the substitutions in the peptide relative to the wild-type amino acid, and does not depend on the structural context of these substitutions (Fig. S1).

  For designs with BLOSUM62, we independently sample amino acids along the peptide based on their favorability with respect to the wild-type amino acid at each position. Specifically, at a given position i we interpret the substitution score $B62[\sigma_i^{\text{wt}} \to \sigma_i]$ from the wild-type to any of the other 19 amino acids as the energy difference

between the two states, and by applying a softmax function with temperature T, we evaluate the probability of these possible substitutions. We then use multinomial sampling to draw new amino acids from this probability distribution at each position,

$$\sigma_i \sim \text{Multinomial}\left(\text{softmax}\left[\frac{B62[\sigma_i^{\text{wt}} \to \sigma]}{T}\right]_{\sigma \in \text{amino-acids}}\right) \tag{S2}$$

For our analyses, we use temperatures $T = 1.0, 2.0, 3.0$ to modulate the variability of the designed peptides.

- *MHC position weight matrices* [42]: MHC alleles exhibit distinct sequence-specific preferences for peptide presentation, most notably in the peptide's amino acid composition at anchor positions. Computational models have characterized these preferences with position weight matrices (PWMs) that reflect the the probability of different amino acids at each position for peptides presented by a given MHC allele [42–45]. As a benchmark for our designs, we use the PWMs from the MHC Motif Atlas [42] to independently sample amino acids at each site along the peptide. The resulting peptide sequences should reflect biases imposed by MHC presentation but not interactions with TCRs.

- *ProteinMPNN* [46]: ProteinMPNN is a commonly used tool for inverse-folding (i.e., designing protein sequences consistent with the backbone of a structure). It is primarily used to sample amino acid sequences conditioned on a protein's backbone structure, and optionally with partial sequence information (i.e., amino acids specified at some of the residues). As ProteinMPNN also outputs probability distributions of amino acids at different sites of the protein, it can also be used to score peptides (i.e., evaluate peptide energy) in a manner similar to HERMES, as

$$\begin{aligned} E_{\text{peptide}}^{(\text{MPNN})}(\sigma; \text{backbone}, \sigma_{\text{TCR}}, \sigma_{\text{MHC}}) &= \log P_{\text{peptide}}^{(\text{MPNN})}(\sigma; \text{backbone}, \sigma_{\text{TCR}}, \sigma_{\text{MHC}}) \\ &= \sum_{i=1}^{\ell} \log P_{\text{peptide}}^{(\text{MPNN})}(\sigma_i; \text{backbone}, \sigma_{\text{TCR}}, \sigma_{\text{MHC}}, \sigma_{/i}) \end{aligned} \tag{S3}$$

  where $P_{\text{peptide}}^{(\text{MPNN})}(\sigma_i; \text{backbone}, \text{TCR}, \text{MHC}, \sigma_{/i})$ is the ProteinMPNN estimate for the probability of amino acid σ_i at position i of the peptide, given the backbone structure of the whole protein complex, and the amino acid sequence of the MHC σ_{MHC}, TCR σ_{TCR}, and the rest of the peptide $\sigma_{/i}$. Similar to HERMES, we considered ProteinMPNN models trained with two noise levels: 0.02 Å (very little noise) and 0.20 Å.

  For designs with ProteinMPNN, we use the algorithm's native functionalities, keeping the TCR and MHC structures fixed (from the design template) and prompting the model to design only the peptide. We use two ProteinMPNN models (trained with 0.02 Å and 0.20 Å noise), and use a temperature of 0.7 for all systems.

- *TCRdock* [38]: TCRdock is a structure prediction algorithm built on top of AlphaFold2 [39], and refined to predict the structure of TCR-pMHC complexes. The algorithm model the structure of TCR-pMHC complexes by leveraging structural templates provided by its curated database, as well as the knowledge of TCR-pMHC canonical binding poses. Crucially, TCRdock's Predicted Alignment Error (PAE) for TCR-pMHC interfaces was found to well discriminate between true peptide binders and random decoys [38]. Thus, to benchmark against TCRdock, we interpret this PAE score as peptide energy, and use it to predict TCR-pMHC binding affinities and peptide-induced T-cell activities.

  TCRdock takes as input sequences of a TCR, MHC, and peptide, and identifies TCR-pMHC structures in its database with similar amino acid sequences to the query to serve as templates to model the structure of the input TCR-pMHC complex. To examine how template-query sequence similarity influences TCRdock's predictive accuracy, we ran the algorithm under two conditions: the default mode, which uses any closest-matching templates available, and the *benchmark* mode, which restricts the maximum sequence similarity of templates with the input TCR-pMHC.

  We use TCRdock to characterize peptide energies, and to filter the sequences designed by HERMES. However, TCRdock does not provide a natural way to sample peptides for design, and because of its high computational cost, we did not include TCRdock PAE calculations in our design pipeline (e.g., for MCMC sampling).

- *TULIP* [47]: TULIP is a transformer-based model trained to generate either TCR or peptide sequences conditioned on the other binding partners in a TCR-pMHC complex. The conditional likelihood function learned by TULIP can be used to score peptides within a TCR-pMHC context. Specifically, this scoring scheme was used in ref. [47] to predict the T-cell activity measurements from ref. [34], and we benchmarked our predictions

against TULIP's for this data. In Table S4, we present the Spearman correlations between TULIP predictions and the fitted EC_{50} values as reported in ref. [34], which are the same correlations reported in the TULIP paper [47]. Additionally, we provide correlations with the fitted EC_{50} values obtained using the procedure outlined in Section II and eq. S2, consistent with correlations reported in Fig. 3.

Benchmarking HERMES predictions for T-cell affinity and activity. We compared the performance of HERMES models at predicting the binding affinity and activity of T-cells against BLOSUM62, ProteinMPNN, TCRdock, and TULIP. Since all of these models are unsupervised in nature, we do not expect predictions to be in the same units as the affinity and activity values, and thus, we use Spearman correlation between data and predictions as a metric for model performance.

For predicting TCR-pMHC binding affinities (Table S2), HERMES-*fixed* and HERMES-*relaxed* with no noise consistently outperform ProteinMPNN and TCRdock-*benchmark* (the TCRdock model without using structural templates similar to the query data). When structural templates from similar complexes are incorporated, TCRdock achieves highly favorable scores, performing best in the A6 TCR system. Notably, BLOSUM62 predictions exhibit strong correlation with affinity data and outperform other models in the 1G4 TCR system. However, given the limited data size, caution is warranted in drawing broad conclusions about the effectiveness of different methods for predicting TCR-pMHC binding affinity.

For predicting T-cell activity (Table S4), HERMES-*fixed* without noise achieves the best performance for TCR1, TCR2, and TCR3, while demonstrating competitive performance for H2-scDb, TCR5, and TCR6 compared to other models. However, all models struggle to provide reliable predictions for TCR7, with the exception of ProteinMPNN-0.02, which attains a moderate 31% correlation. For TCR4, which has a poor structural model (AF3 PAE = 2.07), all machine learning models fail to produce reliable predictions, whereas BLOSUM62 achieves a moderate 33% correlation. Lastly, the structure-based models HERMES and ProteinMPNN consistently outperform the sequence-based transformer model TULIP.

Our results do not indicate a clear preference between HERMES-*fixed* and HERMES-*relaxed*: HERMES-*fixed* predicted T-cell activity better, while HERMES-*relaxed* predicted binding affinity between TCR and pMHC better. However, it is important to note that in all systems tested here, the scored peptides are highly similar to the wild-type peptides, with a maximum Hamming distance of just one amino acid from a wild-type peptide found in one of the available structural templates. As a result, relaxation may not be essential for achieving high performance. Interestingly, noised HERMES models underperform compared to the un-noised counterparts. This is in contrast to prior observations, where noised models achieve higher performances in predicting the stability effect of mutations in proteins [30].

Benchmarking HERMES designs against alternative models. Although we did not experimentally test the designs by the different models, we used both peptide presentation by MHC predicted by NetMHCPan and TCRdock's PAE as benchmarking metrics across different design methods. We note that these are only computational metrics and should not be taken as ground truth; this is particularly true for results using TCRdock PAE as a proxy for binding, which are much more likely farther from the ground truth than the easier-to-predict MHC presentation.

Figs. S9, S10, and S11 show the fraction of peptides meeting each criterion for designs produced by various HERMES models, ProteinMPNN, BLOSUM62, and the system's MHC peptide preferences. Consistently, across the three systems, and for all models, the fraction of designs passing the PAE threshold decreases with increasing Hamming distance from the wild-types' sequences (Figs. S9, S10, and S11; panels E, and G). Also consistently, but more surprisingly, models trained with Gaussian noise applied to the structure—for both HERMES and ProteinMPNN—perform worst than their noise-less counterparts (Figs. S9, S10, and S11; panels E, and G).

HERMES models are consistently the best at designing peptides that are scored well for presentation by the respective MHC alleles (panel C of Figs. S9, S10, and S11). No single model stands out as being consistently best across the three systems at designing peptides above the TCRdock PAE threshold: HERMES outperforms ProteinMPNN in the NY-ESO system (Fig.S9), performs similarly in the EBV system (Fig.S10), and performs worse in the MAGE system (Fig. S11). BLOSUM62 overall makes the worst designs compared to HERMES and ProteinMPNN, based on both MHC presentation and TCRdock PAE scores. Designs sampled from the MHC motif, while scoring well on MHC presentation, consistently score poorly for TCRdock PAE. We expected HERMES-*relaxed* to design better binders than HERMES-*fixed* at greater sequence distances from the wild-types. However, this holds only for the NY-ESO system (Fig. S9) and not for EBV or MAGE (Figs. S10, S11). Overall, a broader benchmarking study, ideally with experimental validation, is needed to rigorously evaluate the effectiveness of different peptide design methods.

V. EXPERIMENTAL PROTOCOL TO MEASURE PEPTIDE ACTIVITIES

Artificial antigen-presenting cell lines. We obtained the full-length coding sequences of HLA-A*02:01, HLA-A*01:01, and HLA-B*35:01 from the IMGT/HLA database [48]. Gene fragments were synthesized by Genscript and were cloned into the lentiviral pLVX-EF1α-IRES-Puro (Clontech) vector. We used Lipofectamine 3000 Transfection Reagent (Invitrogen) to transduce the 293T packaging cell line (ATCC CRL-3216) with the psPAX2 packaging plasmid (Addgene plasmid #12260), the pMD2.G envelope plasmid (Addgene plasmid #12259), and the HLA-encoding lentiviral vector. The lentivirus-containing supernatant was collected after 24 and 48 hours post-transfection and was filtered through a 0.45um SFCA filter (Thermo Scientific). The lentivirus was concentrated using Lenti-X Concentrator (Takara Bio) according to the manufacturer's protocol, resuspended in DPBS (Gibco), and stored at -80C for future use. To generate monoallelic HLA cell lines for HLA-A*02:01 and HLA-B*35:01, HLA-deficient K562 cells (ATCC CCL-243) were transduced with lentivirus. Since K562 cells have a high expression of both MAGE-A3 and Titin antigens, we transduced murine EL4 cell lines (ATCC TIB-39) with HLA-A*01:01 lentivirus to produce HLA-A*01:01 presenters lacking the natural expression of the tested antigens. At 48 hours post-transduction, K562 cells were transferred to Iscove's Modified Dulbecco's Medium containing 10% FBS, 2mM L-glutamine, 100 U/mL penicillin/streptomycin, and 2 μg/mL puromycin (Gibco). EL4 cells were transferred to RPMI media containing 10% FBS, 2mM L-glutamine, 100 U/mL penicillin/streptomycin, and 2 μg/mL puromycin (Gibco). The antibiotic selection continued for one week before transferring cells to the antibiotic-free media.

TCR-expressing Jurkat 76.7 cell lines. Nucleotide sequences of TCRa and TCRb chains were generated using stitchr [49]. Both chains were modified to incorporate murine constant segments (TRAC*01 and TRBC2*01) to enhance TCR surface expression. For NYESO and EBV-specific TCRs, the TCRa and TCRb chains were connected through a P2A ribosomal skip motif, synthesized as a single bicistronic gene fragment by Genscript, and cloned into the lentiviral pLVX-EF1a-P2A-mCherry-IRES-Puro (Clontech). For MAGE-specific TCR, gene fragments for the TCRa and TCRb chains were separately synthesized by Genscript and cloned into either the lentiviral pLVX-EF1a-P2A-mCherry-IRES-Puro (Clontech) or pLVX-EF1α-IRES-G418 (modified in-house from pLVX-EF1α-IRES-Puro). The 293T packaging cell line (ATCC CRL-3216) was transfected with TCR-encoding lentiviral vectors, psPAX2 packaging plasmid (Addgene plasmid #12260), and pMD2.G envelope plasmid (Addgene plasmid #12259) using Lipofectamine 3000 Transfection Reagent (Invitrogen). The lentivirus-containing supernatant was collected after 24 and 48-hours after transfection, filtered through a 0.45μm SFCA filter (Thermo Scientific), concentrated using Lenti-X Concentrator (Takara Bio) according to the manufacturer's protocol, resuspended in DPBS (Gibco), and stored at -80C for future use. To generate Jurkat cells expressing the paired TCR, TCR-null CD8-positive Jurkat 76.7 cells with NFAT-eGFP reporter were transduced with lentivirus. At 48 hours post-transduction, cells were transferred to RPMI media containing 10% FBS, 2mM L-glutamine, 100 U/mL penicillin/streptomycin, and 1 μg/mL puromycin (Gibco). For MAGE-specific cell line generation Jurkat cells were co-transduced with lentiviruses encoding TCRa and TCRb chains and cultured with both 1 μg/mL puromycin and 500 μg/mL Genectin (Gibco). The antibiotic selection continued for one week. The abTCR expression on the surface was confirmed by mCherry expression and staining with a 1:100 dilution of anti-mouse TCRb APC-Fire750 (clone H57-597, 109246, Biolegend).

Specificity validation of TCR-expressing Jurkat. Original cognate peptides and their variants were synthesized as a crude peptide library with by Genscript, and then diluted to a 4mM stock solution in DMSO (Sigma). To examine the specificity of the TCR-expressing Jurkat cell lines, we co-cultured 100,000 Jurkat cells with equal amounts of artificial APCs expressing the corresponding HLA allele in a 96-well round-bottom tissue culture plate. We tested each specificity—MAGE, EBV, NYESO—under 96 different conditions. These included one or two positive control original peptides and an unstimulated "no peptide" control. The cells were cultured in RPMI media supplemented with 10% FBS, 2mM L-glutamine, 100 U/mL penicillin/streptomycin, and 1 μg/ml each of anti-human CD28 (BD Biosciences, 555725) and CD49d (BD Biosciences, 555501). The co-cultures were pulsed with 1μM peptide in triplicate. We added DMSO to unstimulated control wells (no peptide) at concentrations equivalent to those in the peptide-stimulated wells. After incubating the cells for 24 hours at 37°C and 5% CO2, we washed them once in DPBS (Gibco) and resuspended them in 150μl DPBS for flow cytometry analysis using a custom-configured BD Fortessa with FACSDiva software (Beckton Dickenson). We determined GFP percentages using FlowJo version 10.8.1 software (BD Biosciences). Alternatively, we measured the GFP percentage in wells of a 96-well round-bottom plate using Cellcyte X (Echo), using 4x objective in spheroid mode, acquisition in green channel (800 ms exposure 3dB gain). Image analysis was done with CellCyte Studio software with total spheroid area recipe and standard defaults (50 a.u. Contrast Sensitivity, 2 a.u Smoothing, 100 μm^2 filled hole size 100 μm Min. object size), relative GFP+ spheroid area was used as output. For NY-ESO, we compared the T-cell activity level measured using flow cytometry with fluorescence microscopy, and show consistency between the two (Fig. S4). Given the consistency between the two experimental approaches, we relied on fluorescence microscopy only to measure GFP levels induced by different

peptides for MAGE and EBV.

Selecting the GFP expression threshold for T-cell activity. Experimental measurements are done across three replicates, and the reported GFP levels in Figs. 4, S5 for each condition are averaged over the three replicates. We consider a peptide to have activated T-cells if its induced replicate-averaged GFP level met two criteria: 1) it exceeded 0.5%, and 2) it surpassed the mean plus three standard deviations of the GFP level negative controls (no-peptide). According to these criteria, the GFP activation cutoffs were 0.5% for the NY-ESO system, 0.64% for the EBV system, and 0.76% for the MAGE system.

VI. EXPERIMENTAL ACTIVITIES OF THE DESIGNED PEPTIDES

Comparing design success rates across the three TCR-pMHC systems. Among the designed peptides, 50% in the NY-ESO system (Fig. 4A), 16% in the EBV system (Fig. 4B), and 26% in the MAGE system induce significant T-cell activities. However, it should be noted that the designed NY-ESO peptides were closer in sequence to their wild-type templates (2-7 amino acid differences) than the peptides designed for the EBV (3-9 differences) or MAGE (3-8 differences) systems. We find that the activity of designs decays with increasing distance from the respective wild-type, with almost no designs beyond 5 amino-acid distance from the wild-types activating their respective T-cells (Fig. 4, S6, S7, S8). This underscores the difficulty of exploring peptide variants far from the wild-type sequences. If any positive designs do exist at these greater distances, they represent an exponentially smaller fraction of the sequence space, and when in complex with the TCR-MHC, are likely to adopt a different binding mode compared to that of the wild-type.

Minor adjustments to the design protocols, aimed at better exploring the design space, may have also contributed to the observed differences in performance across the systems. For NY-ESO, which achieved the highest design success rate, peptides with highest PAE scores were chosen for experimental validation, whereas in EBV and MAGE, peptides were sampled uniformly above the set PAE thresholds. In MAGE, fixing the amino-acid at position 1 to a Glutamic Acid (E) for the one-third of the designs led to an elevated success rate in this subset (Fig. S8D); E_1 is shared between the two wild-type peptides and this design choice allowed us to assess the importance of this amino acid for function.

The design potential of different HERMES models. In panel A of Figures S6 (NY-ESO), S7 (EBV), and S8 (MAGE), we show the fidelity of designs using the four HERMES models: HERMES-*fixed* with no noise, HERMES-*fixed* with 0.5 Å noise, HERMES-*relaxed* with no noise, and HERMES-*relaxed* with 0.5 Å noise. Although PyRosetta relaxations allows HERMES-*relaxed* to explore the sequence space more broadly, the resulting designs that are made at greater distances from the wild-types seldom succeed in activating T-cells. This pattern is also observed with using models trained with 0.5 Å Gaussian noise. Indeed, successful designs lie within five amino acid substitutions from the wild-types, and for these designs, the un-noised HERMES-*fixed* model demonstrates the highest overall fidelity. To better explore the peptide energy landscape and to design sequences further from the wild-types, it may be necessary to adopt a more specialized relaxation model that accounts for the intricacies of TCR-pMHC poses, while being consistent with the neural network's model of protein environments.

Comparing TCRdock and AF3 as design filters. Using our T-cell activity measurements as the ground truth, we sought to quantify how well the TCRdock and AF3 PAE scores can be used as predictors of T-cell activation and filtering mechanisms for peptide design. For this purpose, we focus primarily on the NY-ESO system, for which we validated a balanced mix of positive and negative designs, chosen based on their TCRdock PAE scores. Here, we found that TCRdock PAE could moderately distinguish true positive from true negative designs (AUROC = 0.82, Fig. S12A). Notably, using TCRdock PAE resulted in a very low false negative rate (0.03%) but a considerably higher false positive rate (50%); see Fig. S5A top. AF3 PAE scores performed equivalently to TCRdock PAE in differentiating positives from negatives (AUROC = 0.82, Fig. S12A), though they correlated only moderately with TCRdock's PAE values predictions (Spearman's $\rho = 0.5$, Fig. S5A bottom). Indeed, some peptides deemed positive by TCRdock were classified as negative by AF3, and vice versa (Fig. S5A bottom). In the EBV and MAGE systems, we only validated the activity induced by those peptides that we deemed as positive based on their TCRdock and AF3 PAE scores. Therefore, we would expect neither TCRdock PAE nor AF3 PAE to distinguish well between the true and false positives among these peptides. While this is the case for the MAGE system (AUROC of 0.62 and 0.63; Fig. S12C), in the EBV system the PAE scores show a moderate predictive power in distinguishing true positives from true negatives (AUROC of 0.68 and 70; Fig. S12B). The PAE scores of TCRdock and AF3 differed most markedly in the MAGE system (Spearman's $\rho = 0.09$ for MAGE), with TCRdock assigning relatively unfavorable PAE scores to the wild-type peptides, whereas AF3 ranked those same peptides near the top (Fig. S5C).

The PAE values reported by the structure prediction algorithms have been shown to be noisy indicators of protein folding fidelity and design quality [50]. While larger experimental libraries will be needed to draw definitive conclusions, it appears that PAEs from TCRdock and AF3 have some complementary predictive strengths, and therefore, combining their information could lead to more robust design decisions.

VII. T-CELL AND MHC SPECIFICITY AS THE ENTROPY OF THE ASSOCIATED PEPTIDE DISTRIBUTION

We characterize the specificity of a TCR-MHC complex and that of the MHC by the diversity of peptides that they can recognize. For TCR-MHC specifiity, we starting with a TCR-pMHC structural template, and use HERMES-*fixed* model to generates a peptide position weight matrix (PWM) that captures the ensemble of peptides recognized by the TCR and presented by the MHC. In Figure 5, we analyzed 105 human and 36 mouse MHC class I TCR-pMHC structures—curated by the TCRdock template database [38]—which included only MHC alleles represented in at least two structures. For each MHC allele in this set, we then characterized the distribution of peptides that the MHC can present (independently of the TCR) using PWMs inferred by the MHC Motif Atlas [42]. In both cases, we use the entropy of these PWMs, $S = -\sum_{\alpha,i} p_i(\alpha) \log_2 p_i(\alpha)$, as a proxy for degeneracy of the target (TCR-MHC or MHC), where $\alpha \in \{1, \ldots 20\}$ denotes the amino acid types, and $i \in \{1, \ldots, \ell\}$ are positions along the peptide.

[1] Visani GM, et al. (2024) HERMES: Holographic Equivariant neuRal network model for Mutational Effect and Stability prediction *bioRxiv* 2024.07.09.602403.

[2] Pun MN, et al. (2024) Learning the shape of protein microenvironments with a holographic convolutional neural network *Proc. Natl. Acad. Sci. U. S. A.* 121:e2300838121.

[3] AlQuraishi M (2019) ProteinNet: a standardized data set for machine learning of protein structure *BMC Bioinformatics* 20:311.

[4] Chaudhury S, Lyskov S, Gray JJ (2010) PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta *Bioinformatics* 26:689–691.

[5] Łuksza M, et al. (2022) Neoantigen quality predicts immunoediting in survivors of pancreatic cancer *Nature* 606:389–395.

[6] Yang X, et al. (2015) Structural basis for clonal diversity of the public T cell response to a dominant human Cytomegalovirus Epitope *J. Biol. Chem.* 290:29106–29119.

[7] Abramson J, et al. (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3 *Nature* 630:493–500.

[8] Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M (2020) 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:W449–W454.

[9] Bradley P (2023) Structure-based prediction of T cell receptor:peptide-MHC interactions *Elife* 12:e82813.

[10] Jumper J, et al. (2021) Highly accurate protein structure prediction with AlphaFold *Nature* 596:583–589.

[11] Liu B, et al. (2024) Design of high specificity binders for peptide-MHC-I complexes *bioRxiv* 2024.11.28.625793.

[12] Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks *Proc. Natl. Acad. Sci. U. S. A.* 89:10915–10919.

[13] Tadros DM, Eggenschwiler S, Racle J, Gfeller D (2023) The MHC Motif Atlas: a database of MHC binding specificities and ligands *Nucleic Acids Res.* 51:D428–D437.

[14] Lundegaard C, et al. (2008) NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11 *Nucleic Acids Res.* 36:W509–12.

[15] Lundegaard C, Lund O, Nielsen M (2008) Accurate approximation method for prediction of class I MHC affinities for peptides of length 8, 10 and 11 using prediction tools trained on 9mers *Bioinformatics* 24:1397–1398.

[16] Andreatta M, Nielsen M (2016) Gapped sequence alignment using artificial neural networks: application to the MHC class I system *Bioinformatics* 32:511–517.

[17] Dauparas J, et al. (2022) Robust deep learning-based protein sequence design using ProteinMPNN *Science* 378:49–56.

[18] Meynard-Piganeau B, Feinauer C, Weigt M, Walczak AM, Mora T (2024) TULIP: A transformer-based unsupervised language model for interacting peptides and T cell receptors that generalizes to unseen epitopes *Proc. Natl. Acad. Sci. U. S. A.* 121:e2316401121.

[19] Robinson J, et al. (2003) IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex *Nucleic Acids Res.* 31:311–314.

[20] Heather JM, et al. (2022) Stitchr: stitching coding TCR nucleotide sequences from V/J/CDR3 information *Nucleic Acids Res.* 50:e68.

[21] Roney JP, Ovchinnikov S (2022) State-of-the-Art Estimation of Protein Model Accuracy Using AlphaFold *Phys. Rev. Lett.* 129:238101.

[22] Chen JL, et al. (2005) Structural and kinetic basis for heightened immunogenicity of T cell vaccines *Journal of Experimental Medicine* 201:1243–1255.

[23] Garboczi DN, et al. (1996) Structure of the complex between human T-cell receptor, viral peptide and HLA-A2 *Nature* 384:134–141.
[24] Ding YH, Baker BM, Garboczi DN, Biddison WE, Wiley DC (1999) Four A6-TCR/Peptide/HLA-A2 Structures that Generate Very Different T Cell Signals Are Nearly Identical *Immunity* 11:45–56.
[25] Hsiue EHC, et al. (2021) Targeting a neoantigen derived from a common TP53 mutation *Science* 371:eabc8697.
[26] Gras S, et al. (2010) Allelic polymorphism in the T cell receptor and its impact on immune responses *Journal of Experimental Medicine* 207:1555–1567.
[27] Liu YC, et al. (2014) A molecular basis for the interplay between T cells, viral mutants, and human leukocyte antigen micropolymorphism *J. Biol. Chem.* 289:16688–16698.
[28] Raman MCC, et al. (2016) Direct molecular mimicry enables off-target cardiovascular toxicity by an enhanced affinity TCR designed for cancer immunotherapy *Sci. Rep.* 6:18851.
[29] Cameron BJ, et al. (2013) Identification of a Titin-derived HLA-A1-presented peptide as a cross-reactive target for engineered MAGE A3-directed T cells *Sci. Transl. Med.* 5:197ra103.
[30] Visani GM, et al. (2024) HERMES: Holographic Equivariant neuRal network model for Mutational Effect and Stability prediction *bioRxiv* 2024.07.09.602403.
[31] Pun MN, et al. (2024) Learning the shape of protein microenvironments with a holographic convolutional neural network *Proc. Natl. Acad. Sci. U. S. A.* 121:e2300838121.
[32] AlQuraishi M (2019) ProteinNet: a standardized data set for machine learning of protein structure *BMC Bioinformatics* 20:311.
[33] Chaudhury S, Lyskov S, Gray JJ (2010) PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta *Bioinformatics* 26:689–691.
[34] Łuksza M, et al. (2022) Neoantigen quality predicts immunoediting in survivors of pancreatic cancer *Nature* 606:389–395.
[35] Yang X, et al. (2015) Structural basis for clonal diversity of the public T cell response to a dominant human Cytomegalovirus Epitope *J. Biol. Chem.* 290:29106–29119.
[36] Abramson J, et al. (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3 *Nature* 630:493–500.
[37] Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M (2020) 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:W449–W454.
[38] Bradley P (2023) Structure-based prediction of T cell receptor:peptide-MHC interactions *Elife* 12:e82813.
[39] Jumper J, et al. (2021) Highly accurate protein structure prediction with AlphaFold *Nature* 596:583–589.
[40] Liu B, et al. (2024) Design of high specificity binders for peptide-MHC-I complexes *bioRxiv* 2024.11.28.625793.
[41] Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks *Proc. Natl. Acad. Sci. U. S. A.* 89:10915–10919.
[42] Tadros DM, Eggenschwiler S, Racle J, Gfeller D (2023) The MHC Motif Atlas: a database of MHC binding specificities and ligands *Nucleic Acids Res.* 51:D428–D437.
[43] Lundegaard C, et al. (2008) NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11 *Nucleic Acids Res.* 36:W509–12.
[44] Lundegaard C, Lund O, Nielsen M (2008) Accurate approximation method for prediction of class I MHC affinities for peptides of length 8, 10 and 11 using prediction tools trained on 9mers *Bioinformatics* 24:1397–1398.
[45] Andreatta M, Nielsen M (2016) Gapped sequence alignment using artificial neural networks: application to the MHC class I system *Bioinformatics* 32:511–517.
[46] Dauparas J, et al. (2022) Robust deep learning-based protein sequence design using ProteinMPNN *Science* 378:49–56.
[47] Meynard-Piganeau B, Feinauer C, Weigt M, Walczak AM, Mora T (2024) TULIP: A transformer-based unsupervised language model for interacting peptides and T cell receptors that generalizes to unseen epitopes *Proc. Natl. Acad. Sci. U. S. A.* 121:e2316401121.
[48] Robinson J, et al. (2003) IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex *Nucleic Acids Res.* 31:311–314.
[49] Heather JM, et al. (2022) Stitchr: stitching coding TCR nucleotide sequences from V/J/CDR3 information *Nucleic Acids Res.* 50:e68.
[50] Roney JP, Ovchinnikov S (2022) State-of-the-Art Estimation of Protein Model Accuracy Using AlphaFold *Phys. Rev. Lett.* 129:238101.
[51] Chen JL, et al. (2005) Structural and kinetic basis for heightened immunogenicity of T cell vaccines *Journal of Experimental Medicine* 201:1243–1255.
[52] Garboczi DN, et al. (1996) Structure of the complex between human T-cell receptor, viral peptide and HLA-A2 *Nature* 384:134–141.
[53] Ding YH, Baker BM, Garboczi DN, Biddison WE, Wiley DC (1999) Four A6-TCR/Peptide/HLA-A2 Structures that Generate Very Different T Cell Signals Are Nearly Identical *Immunity* 11:45–56.
[54] Hsiue EHC, et al. (2021) Targeting a neoantigen derived from a common TP53 mutation *Science* 371:eabc8697.
[55] Gras S, et al. (2010) Allelic polymorphism in the T cell receptor and its impact on immune responses *Journal of Experimental Medicine* 207:1555–1567.
[56] Liu YC, et al. (2014) A molecular basis for the interplay between T cells, viral mutants, and human leukocyte antigen micropolymorphism *J. Biol. Chem.* 289:16688–16698.
[57] Raman MCC, et al. (2016) Direct molecular mimicry enables off-target cardiovascular toxicity by an enhanced affinity TCR designed for cancer immunotherapy *Sci. Rep.* 6:18851.

[58] Cameron BJ, et al. (2013) Identification of a Titin-derived HLA-A1-presented peptide as a cross-reactive target for engineered MAGE A3-directed T cells *Sci. Transl. Med.* 5:197ra103.

TCR / Ab	(PDBid, WT peptide)	MHC allele
1G4 TCR	(2BNR [51], SLLMWITQC); (2BNQ [51], SLLMWITQV)	HLA-A*02:01
A6 TCR	(1AO7 [52], LLFGYPVYV); (1QSF [53], LLFGYPVAV)	HLA-A*02:01
TCR1	(5D2N [35], NLVPMVATV)	HLA-A*02:01
TCR2	(AF3 model, NLVPMVATV)	HLA-A*02:01
TCR3	(AF3 model, NLVPMVATV)	HLA-A*02:01
TCR4	(AF3 model, IMDQVPFSV)	HLA-A*02:01
TCR5	(AF3 model, IMDQVPFSV)	HLA-A*02:01
TCR6	(AF3 model, IMDQVPFSV)	HLA-A*02:01
TCR7	(AF3 model, GRLKALCQR)	HLA-B*27:05
H2-scDb (Fab)	(6W51 [54], HMTEVVRHC)	HLA-A*02:01

Table S1. **Structure templates used for the binding affinity and T-cell activity analyses.** The first two rows contain information related to the analyses on binding affinity data, the rest are related to T-cell activity data. Every structure identified by a 4-letter PDBid was downloaded from the Protein Data Bank. The raw PDB file for 5D2N [35] has two copies of the TCR-pMHC complex, so we deleted one before running any analysis. Similarly, the raw PDB file for 6W51 [54] has the Antibody and pMHC coordinates of two different copies within the crystal, so we expanded the crystal using PyMOL's `symexp` command and kept a single bound Antibody-pMHC complex. The sequence inputs used to fold TCR-pMHC complexes with AlphaFold3 [36] can be found in Table S3.

Model	1G4 TCR	A6 TCR
BLOSUM62	0.83 (0.00)	0.56 (0.02)
TCRdock	0.62 (0.02)	0.88 (0.00)
TCRdock benchmark	0.25 (0.39)	-0.14 (0.62)
ProteinMPNN 0.02	-0.01 (0.97)	0.49 (0.05)
ProteinMPNN 0.20	0.12 (0.69)	0.29 (0.27)
HERMES-fixed 0.00	0.29 (0.31)	0.63 (0.01)
HERMES-fixed 0.50	0.19 (0.51)	-0.28 (0.29)
HERMES-relaxed 0.00	0.72 (0.00)	0.56 (0.03)
HERMES-relaxed 0.50	0.65 (0.01)	0.06 (0.83)

Table S2. **Benchmarking of models for predicting TCR-pMHC binding affinities.** The table lists the Spearman $\rho\,(\mathrm{p-value})$ of models' predictions against experimentally measured binding affinities for (i) 1G4 TCR in complex with HLA-A*02:01 binding to the variants of the NY-ESO-1 peptide, and (ii) the A6 TCR in complex with HLA-A*02:01 binding to the variants of the Tax peptide.

TCR	MHC seq	peptide seq	variable alpha seq	variable beta seq
TCR2	GSHSMRYFFTSVSRPGRGEPR FIAVGYVDDTQFVRFDSDAAS QRMEPRAPWIEQEGPEYWDG ETRKVKAHSQTHRVDLGTLRG YYNQSEAGSHTVQRMYGCDV GSDWRFLRGYHQYAYDGKDYI ALKEDLRSWTAADMAAQTTK HKWEAAHVAEQLRAYLEGTCV EWLRRYLENG	NLVPMVATV	DAKTTQPNSMESNEEEPVHL PCNHSTISGTDYIHWYRQLPS QGPEYVIHGLTSNVNNRMAS LAIAEDRKSSTLILHRATLRDA AVYYCILSNNNDMRFGAGTR LTVKP	SQTIHQWPATLVQPVGSPLSL ECTVEGTSNPNLYWYRQAA GRGLQLLFYSVGIGQISSEVP QNLSASRPQDRQFILSSKKLL LSDSGFYLCAWTISDLAKNIQ YFGAGTRLSVL
TCR3	GSHSMRYFFTSVSRPGRGEPR FIAVGYVDDTQFVRFDSDAAS QRMEPRAPWIEQEGPEYWDG ETRKVKAHSQTHRVDLGTLRG YYNQSEAGSHTVQRMYGCDV GSDWRFLRGYHQYAYDGKDYI ALKEDLRSWTAADMAAQTTK HKWEAAHVAEQLRAYLEGTCV EWLRRYLENG	NLVPMVATV	DQQVKQNSPSLSVQEGRISIL NCDYTNSMFDYFLWYKKYP AEGPTFLISISSIKDKNEDGR FTVFLNKSAKHLSLHIVPSQP GDSAVYFCAASGGPSYDKVI FGPGTSLSVIP	DAGVIQSPRHEVTEMGQEVT LRCKPISGHDYLFWYRQTMM RGLELLIYFNNNVPIDDSGMP EDRFSAKMPNASFSTLKIQPS EPRDSAVYFCASSPGNILGGY TFGSGTRLTVV
TCR4	GSHSMRYFFTSVSRPGRGEPR FIAVGYVDDTQFVRFDSDAAS QRMEPRAPWIEQEGPEYWDG ETRKVKAHSQTHRVDLGTLRG YYNQSEAGSHTVQRMYGCDV GSDWRFLRGYHQYAYDGKDYI ALKEDLRSWTAADMAAQTTK HKWEAAHVAEQLRAYLEGTCV EWLRRYLENG	IMDQVPFSV	GIQVEQSPPDLILQEGANSTL RCNFSDSVNNLQWFHQNPW GQLINLFYIPSGTKQNGRLSA TTVATERYSLLYISSSQTTDSG VYFCAPRRNYGQNFVFGPGT RLSVLP	NAGVMQNPRHLVRRRGQEA RLRCSPMKGHSHVYWYRQL PEEGLKFMVYLQKENIIDESG MPKERFSAEFPKEGPSILRIQ QVVRGDSAAYFCASSPGQGG YEQYFGPGTRLTVT
TCR5	GSHSMRYFFTSVSRPGRGEPR FIAVGYVDDTQFVRFDSDAAS QRMEPRAPWIEQEGPEYWDG ETRKVKAHSQTHRVDLGTLRG YYNQSEAGSHTVQRMYGCDV GSDWRFLRGYHQYAYDGKDYI ALKEDLRSWTAADMAAQTTK HKWEAAHVAEQLRAYLEGTCV EWLRRYLENG	IMDQVPFSV	QKEVEQNSGPLSVPEGAIASL NCTYSDRGSQSFFWYRQYSG KSPELIMFIYSNGDKEDGRFT AQLNKASQYVSLLIRDSQPSD SATYLCAVRIGSQGNLIFGKG TKLSVKP	EAQVTQNPRYLITVTGKKLTV TCSQNMNHEYMSWYRQDPGL GLRQIYYSMNVEVTDKGDVPE GYKVSRKEKRNFPLILESPSPN QTSLYFCASSLLQGTEAFFGQG TRLTVV
TCR6	GSHSMRYFFTSVSRPGRGEPR FIAVGYVDDTQFVRFDSDAAS QRMEPRAPWIEQEGPEYWDG ETRKVKAHSQTHRVDLGTLRG YYNQSEAGSHTVQRMYGCDV GSDWRFLRGYHQYAYDGKDYI ALKEDLRSWTAADMAAQTTK HKWEAAHVAEQLRAYLEGTCV EWLRRYLENG	IMDQVPFSV	AQKITQTQPGMFVQEKEAVT LDCTYDTSDPSYGLFWYKQP SSGEMIFLIYQGSYDQQNATE GRYSLNFQKARKSANLVISAS QLGDSAMYFCAMRESSNSNS GYALNFGKGTSLLVTP	IAGITQAPTSQILAAGRRMTLR CTQDMRHNAMYWYRQDLGL GLRLIHYSNTAGTTGKGEVPD GYSVSRANTDDFPLTLASAVPS QTSVYFCASSPRGGTADTQYF GPGTRLTVL
TCR7	GSHSMRYFHTSVSRPGRGEPR FITVGYVDDTLFVRFDSDAAS PREEPRAPWIEQEGPEYWDR ETQICKAKAQTDREDLRTLLR YYNQSEAGSHTLQNMYGCDV GPDGRLLRGYHQDAYDGKDY IALNEDLSSWTAADTAAQITQR KWEAARVAEQLRAYLEGECVE WLRRYLENG	GRLKALCQR	GIQVEQSPPDLILQEGANSTL RCNFSDSVNNLQWFHQNPW GQLINLFYIPSGTKQNGRLSA TTVATERYSLLYISSSQTTDSG VYFCAVDVGGSYIPTFGRGTS LIVHP	DAGVTQSPTHLIKTRGQQVTL RCSPKSGHDTVSWYQQALGQ GPQFIFQYYEEEERQRGNFPD RFSGHQFPNYSSELNVNALLL GDSALYLCASSLARQGSNEQYF GPGTRLTVT

Table S3. **Sequence information used to model TCR1-7 with AlphaFold3.** The table lists the amino acid sequences of the MHC, peptide, and TCR alpha and beta chains that were used as input to AlphaFold3 [36] for modeling TCR2-7, for which no experimentally determined structure is currently available.

Model	H2-scDb (Ab)	TCR1	TCR2	TCR3	TCR4	TCR5	TCR6	TCR7
BLOSUM62	-0.05 (0.52)	-0.07 (0.47)	0.14 (0.20)	0.38 (0.00)	0.33 (0.01)	0.42 (0.00)	0.55 (0.00)	-0.01 (0.89)
TCRdock	N/A	0.02 (0.83)	0.13 (0.24)	0.21 (0.03)	0.18 (0.15)	-0.02 (0.83)	0.19 (0.04)	0.06 (0.50)
TCRdock benchmark	N/A	-0.03 (0.77)	0.44 (0.00)	0.11 (0.27)	-0.22 (0.08)	0.28 (0.00)	0.14 (0.14)	0.06 (0.50)
TULIP	N/A	0.12 (0.24)	0.18 (0.09)	-0.12 (0.20)	0.04 (0.73)	0.24 (0.00)	0.29 (0.00)	N/A
TULIP* (from [47])	N/A	*0.47 (0.00)	*0.44 (0.00)	*0.14 (0.07)	*0.23 (0.00)	*0.20 (0.01)	*0.09 (0.25)	N/A
ProteinMPNN 0.02	0.53 (0.00)	0.46 (0.00)	0.53 (0.00)	0.49 (0.00)	-0.23 (0.06)	0.45 (0.00)	0.42 (0.00)	0.31 (0.00)
ProteinMPNN 0.20	0.03 (0.65)	0.27 (0.01)	0.46 (0.00)	0.40 (0.00)	-0.23 (0.06)	0.26 (0.00)	0.49 (0.00)	0.02 (0.79)
HERMES-fixed 0.00	0.46 (0.00)	0.56 (0.00)	0.71 (0.00)	0.59 (0.00)	-0.06 (0.60)	0.37 (0.00)	0.42 (0.00)	-0.01 (0.95)
HERMES-fixed 0.50	0.26 (0.00)	0.42 (0.00)	0.61 (0.00)	0.54 (0.00)	0.03 (0.83)	0.26 (0.00)	0.31 (0.00)	0.02 (0.82)
HERMES-relaxed 0.00	0.29 (0.00)	0.31 (0.00)	0.57 (0.00)	0.55 (0.00)	-0.20 (0.11)	0.37 (0.00)	0.42 (0.00)	0.14 (0.15)
HERMES-relaxed 0.50	0.16 (0.04)	0.37 (0.00)	0.50 (0.00)	0.45 (0.00)	-0.16 (0.20)	0.23 (0.00)	0.06 (0.49)	0.08 (0.37)

Table S4. **Benchmarking of models for predicting peptide-induced T-cell activities.** The table lists the Spearman $\rho\,(\mathrm{p}-\mathrm{value})$ of models' predictions against experimentally measured T-cell activities. The correlations for TCR1-7 are reported based on the EC_{50} values obtained using the procedure outlined in Section II and Eq. S2, consistent with correlations reported in Fig. 3. In addition, we present the Spearman correlations between TULIP predictions and the fitted EC_{50} values as reported in ref. [34], which are the same correlations reported in the TULIP paper [47] (TULIP* row). The TULIP* correlations also include datapoints for which T-cell responses were below the detection limit, and which we exclude from our analysis (gray points in Fig. 3).

System	**TCR**	**(PDBid, WT peptide)**	**MHC allele**
NY-ESO	1G4 TCR	(2BNR [51], SLLMWITQC); (2BNQ [51], SLLMWITQV)	HLA-A*02:01
EBV	TK3 TCR	(3MV7* [55], HPVGEADYFEY); (4PRP* [56], HPVGQADYFEY)	HLA-B*35:01
MAGE	eng. TCR	(5BRZ* [57], EVDPIGHLY); (5BS0* [57], ESDPIVAQY)	HLA-A*01:01

Table S5. **Structure templates used for peptide design.** For the PDB entries marked with an asterisk (*), we used the curated structures from the TCRdock database [38] instead of the PDB entries.

MHC allele	(PDBid, WT peptide)
HLA-A*01:01	(5BS0, ESDPIVAQY); (5BRZ, EVDPIGHLY)
HLA-A*02:01	(4FTV, LLFGYPVYV); (5D2L, NLVPMVATV); (2F53, SLLMWITQC); (5TEZ, GILGFVFTL); (6VQO, HMTEVVRHC); (3HG1, ELAGIGILTV); (2P5E, SLLMWITQC); (7N6E, YLQPRTFLL); (6RP9, SLLMWITQV); (5C08, RQWGPDPAAV); (3H9S, MLWGYLQYV); (5C0B, RQFGPDFPTI); (6EQA, AAGIGILTV); (5EUO, GILGFVFTL); (3D39, LLFGPVYV); (6VRN, HMTEVVRHC); (5YXU, KLVALGINAV); (5EU6, YLEPGPVTV); (1BD2, LLFGYPVYV); (1OGA, GILGFVFTL); (2VLK, GILGFVFTL); (5MEN, ILAKFLHWL); (3D3V, LLFGPVYV); (1AO7, LLFGYPVYV); (5ISZ, GILGFVFTL); (5JHD, GILGFVFTL); (6DKP, ELAGIGILTV); (1QSE, LLFGYPRYV); (3QDM, ELAGIGILTV); (5JZI, KLVALGINAV); (2P5W, SLLMWITQC); (5C0C, RQFGPDWIVA); (7N1E, RLQSLQTYV); (4QOK, EAAGIGILTV); (5E6I, GILGFVFTL); (2VLR, GILGFVFTL); (6RSY, RMFPNAPYL); (3QEQ, AAGIGILTV); (3O4L, GLCTLVAML); (5C0A, MVWGPDPLYV); (5C07, YQFGPDFPIA); (5D2N, NLVPMVATV); (2F54, SLLMWITQC); (2VLJ, GILGFVFTL); (5NHT, ELAGIGILTV); (6TRO, GVYDGREHTV); (7N1F, YLQPRTFLL); (3GSN, NLVPMVATV); (6D78, AAGIGILTV); (3QDG, ELAGIGILTV); (5NQK, ELAGIGILTV); (5HHO, GILEFVFTL); (2PYE, SLLMWITQC); (5HHM, GILGLVFTL); (2GJ6, LLFGKPVYV); (5NMF, SLYNTIATL); (3UTS, ALWGPDPAAA); (5E9D, ELAGIGILTV); (6RPA, SLLMWITQV); (1QSF, LLFGYPVAV); (4EUP, ALGIGILTV); (5YXN, KLVALGINAV); (2BNQ, SLLMWITQV); (6RPB, SLLMWITQV); (5HYJ, AQWGPDPAAA); (3QFJ, LLFGFPVYV); (3UTT, ALWGPDPAAA); (3QDJ, AAGIGILTV); (4L3E, ELAGIGILTV); (5C09, YLGGPDFPTI); (6VRM, HMTEVVRHC); (1QRN, LLFGYAVYV); (2BNR, SLLMWITQC); (3PWP, LGYGFVNYI); (6AMU, MMWDRGLGMM); (6AM5, SMLGIGIVPV); (4MNQ, ILAKFLHWL); (5NME, SLYNTVATL); (5NMG, SLFNTIAVL)
HLA-A*11:01	(5WKH, GTSGSPIINR); (5WKF, GTSGSPIVNR)
HLA-A*24:02	(3VXR, RYPLTFGWCF); (3VXM, RFPLTFGWCF); (3VXS, RYPLTLGWCF)
HLA-B*07:02	(6VMX, RPPIFIRRL); (6AVF, APRGPHGGAASGL); (6AVG, APRGPHGGAASGL)
HLA-B*08:01	(1MI5, FLRGRAYGL); (3FFC, FLRGRAYGL); (4QRP, HSKKKCDEL); (3SJV, FLRGRAYGL)
HLA-B*27:05	(4G8G, KRWIILGLNK); (4G9F, KRWIIMGLNK)
HLA-B*35:01	(6BJ2, IPLTEEAEL); (4PRP, HPVGQADYFEY); (3MV7, HPVGEADYFEY); (3MV9, HPVGEADYFEY); (3MV8, HPVGEADYFEY)
HLA-B*35:08	(4PRH, HPVGDADYFEY); (4PRI, HPVGEADYFEY)
HLA-B*44:05	(3DXA, EENLLDFVRF); (3KPR, EEYLKAWTF); (3KPS, EEYLQAFTY)
H2Db	(3PQY, SSLENFRAYV); (5WLG, SQLLNAKYL); (5TJE, KAVYNFATM); (7JWI, ASNENMETM); (5SWS, ASNENMETM); (5TIL, KAPYNFATM); (7JWJ, ASNENMETM); (6G9Q, KAPYDYAPI); (5M02, KAPFNFATM); (5M00, KAVANFATM); (5M01, KAPANFATM)
H2Kb	(3PQY, SSLENFRAYV); (5WLG, SQLLNAKYL); (5TJE, KAVYNFATM); (7JWI, ASNENMETM); (5SWS, ASNENMETM); (5TIL, KAPYNFATM); (7JWJ, ASNENMETM); (6G9Q, KAPYDYAPI); (5M02, KAPFNFATM); (5M00, KAVANFATM); (5M01, KAPANFATM)
H2Ld	(2OI9, QLSPFPFDL); (4NHU, GGGAPWNPAMMI); (4MVB, QPAEGGFQL); (4MXQ, SPAPRPLDL); (2E7L, QLSPFPFDL); (4N5E, VPYMAEFGM); (6L9L, SPSYAYHQF); (3E3Q, QLSPFPFDL); (3E2H, QLSPFPFDL); (3TPU, FLSPFWFDI); (3TFK, QLSDVPMDL); (4MS8, SPAEAGFFL); (4N0C, MPAGRPWDL); (3TJH, SPLDSLWWI)

Table S6. **Structure templates to compute TCR-MHC specificities.** The table list the structures used to compute TCR-MHC specificities in Fig. 5 for different MHC class I alleles in humans and mice. Structures were collected from the TCRdock curated database [38].

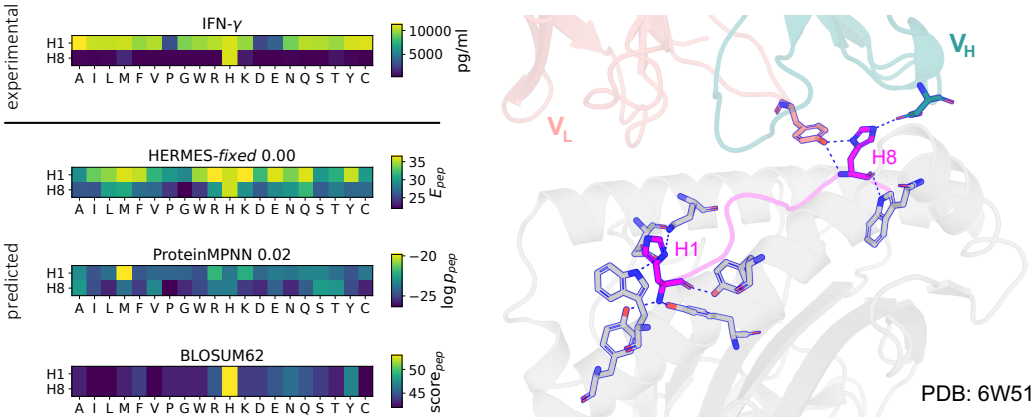

Figure S1. **Context-dependent mutational effects on T-cell activity.** The heatmaps show both experimental measurements and model predictions of T-cell activities of amino acid variants at positions 1 and 8 of the p53^{R}157H peptide, which we study in complex with the designed bi-specific antibody H2-scDB and HLA-A*02:01 [54]. The wild-type peptide has a histidine at both sites. A standard substitution matrix (such as BLOSUM62) would assign identical scores to mutations at these two sites, because they share the same wild-type residue. However, peptide binding and the subsequent T-cell activity depends on the structural context of each substitution. Here, H1 forms polar contacts only with the MHC (from the sides), while H8 is more tightly constrained by polar contacts with the TCR (right panel). HERMES-*fixed* correctly leverages structural context to predict T-cell activity as a result of substitutions away from histidines at both positions. Although ProteinMPNN also incorporates structural information, its predictions are not as accurate in this instance.

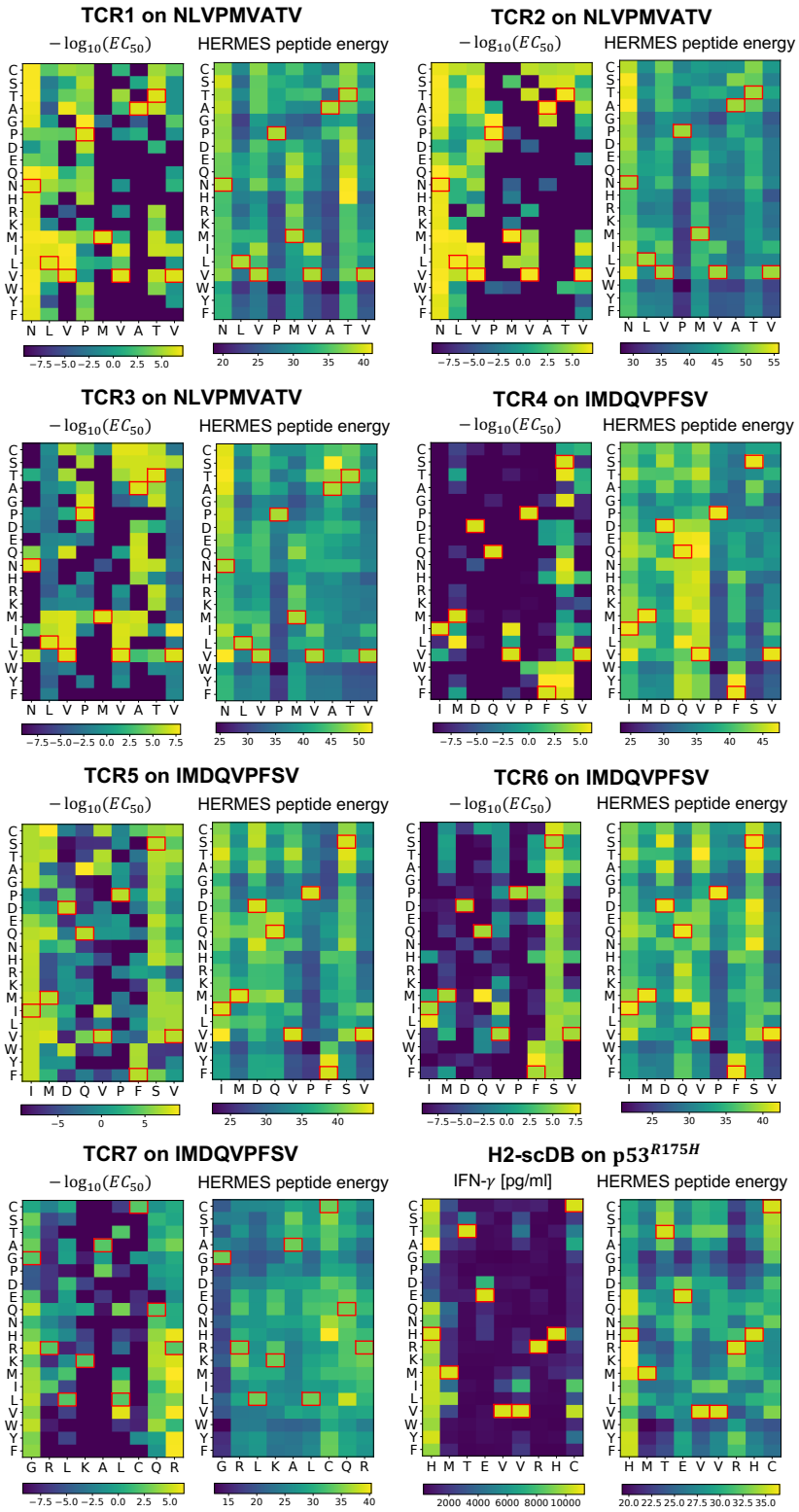

Figure S2. **Mutational effect on T-cell activities**. The heatmaps show the experimental measurements and the predictions from the HERMES-*fixed* model for the effect of all single point mutations on peptides interacting with TCR1-7 [34], and the designed bi-specific antibody H2-scDB [54] (similar to the data presented in scatter plots in Fig. 3). In the heatmaps, each column corresponds to a position in the peptide and is labeled with its wild-type amino acid (indicated by a red square on the heatmaps), and the rows show the amino acid substitutions.

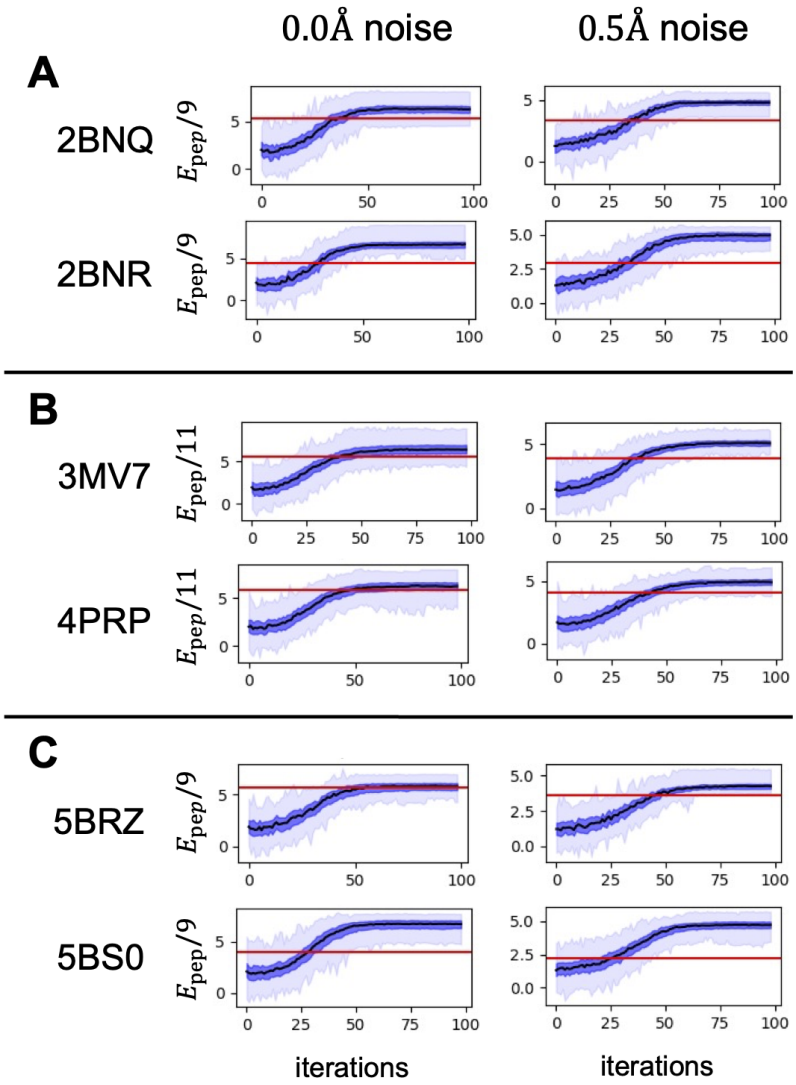

Figure S3. **Peptide energy improvement through HERMES-*relaxed* design protocol.** Peptide energy (eq. S1) divided by the length of the designed peptide is shown as a function of HERMES-*relaxed* MCMC iterations for designs made in **(A)** NY-ESO, **(B)** EBV, and **(C)** MAGE systems. In all panels, the black line represents the median across all designs (100 for NY-ESO; 200 for EBV and MAGE), with the dark blue shading indicating the 25%–75% quantile, and the light blue shading indicating the 5%–95% quantile across designs. The red line indicates the energy of the wild-type peptide in its respective structure. Designs are subdivided by the input structural template (rows) and by the noise level used in the HERMES-*relaxed* protocol (columns). The HERMES-*relaxed* protocol consistently produces peptides with more favorable (higher) peptide energies than that of the wild-type, indicating that the protocol can successfully optimize the peptide energy landscapes; models trained with noise are more successful in finding peptides with higher energy values relative to the wild-types'.

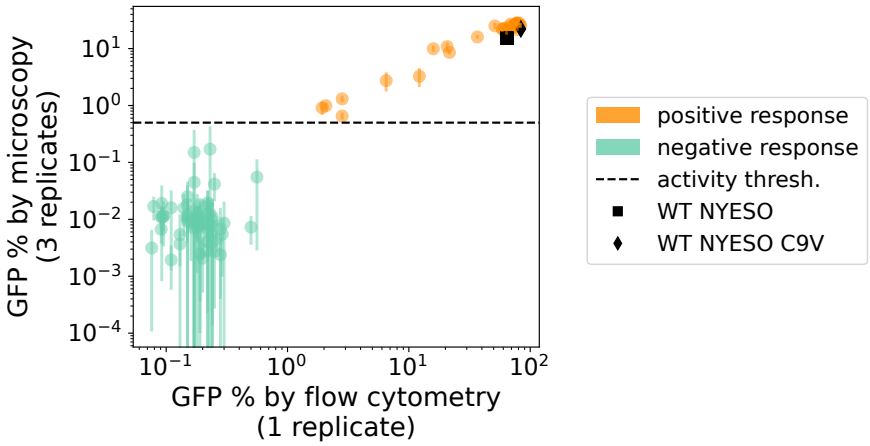

Figure S4. **Measurements of GFP levels by flow cytometry and fluorescence microscopy.** For the NY-ESO system, we performed measurements of GFP levels from peptide-induced T-cell activities, using both flow cytometry and fluorescence microscopy. Fluorescence microscopy measurements were performed in three independent replicates; markers denote the mean GFP level, and error bars show the standard deviation across the three replicates. Flow cytometry was done only on a single replicate. The two measurement methods demonstrate good agreement, with a clear separation between designs that induced significant T-cell activation (orange) and those that did not (green). The black markers indicate the activities induced by the wild-type peptides.

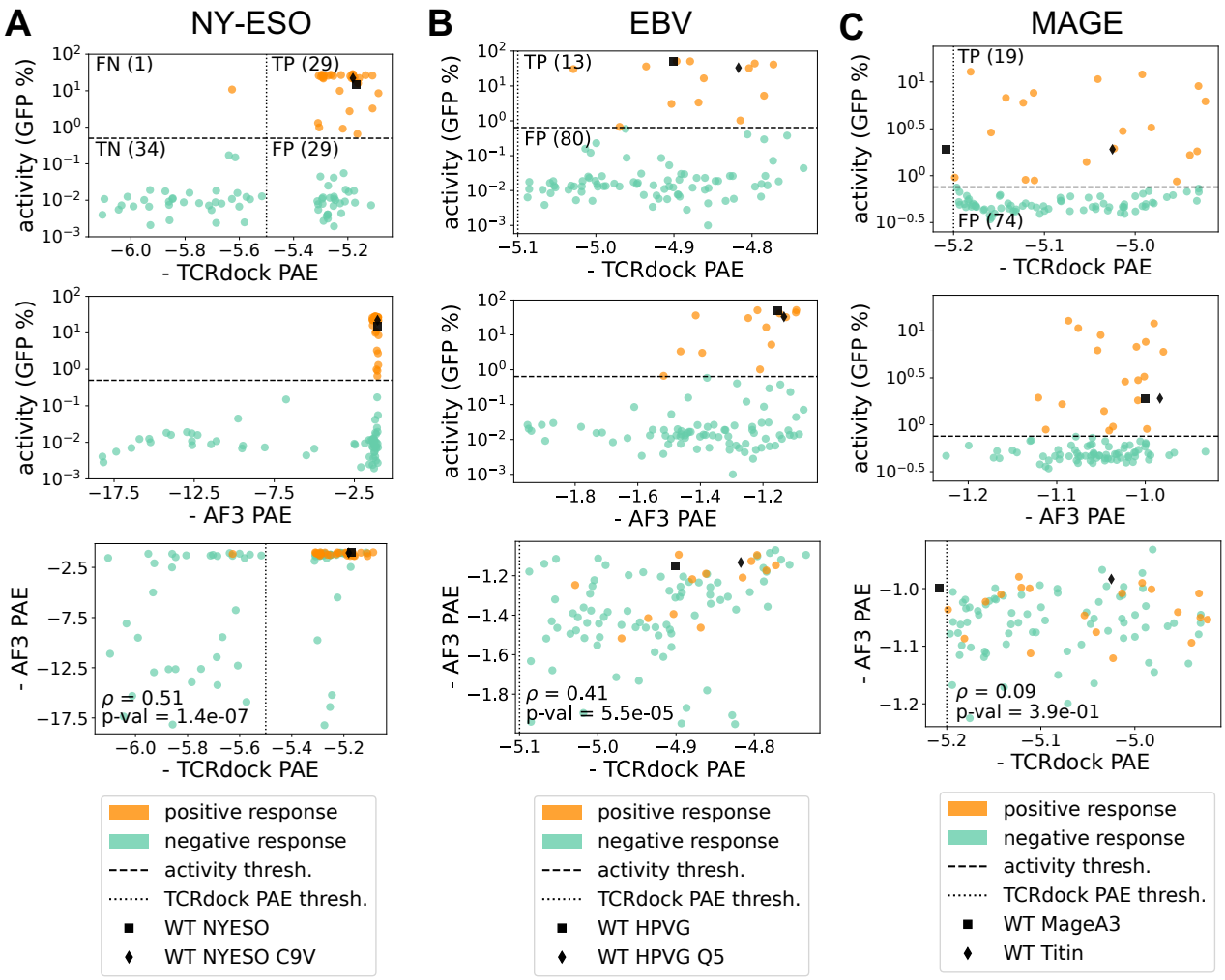

Figure S5. **TCRdock and AlphaFold3 PAEs for HERMES-designed peptides**. **(A)** For the NY-ESO system, the induced T-cell activity by the designed peptides is shown against the computed TCRdock negative-PAE (top), and the AlphaFold3 negative-PAE (center); larger negative-PAE's indicate better designs. The bottom panel compares the AlphaFold3 and TCRdock negative-PAE values for each designed peptide. The weak correlation between these two PAE metrics suggests that they can serve as complementary filters during design. Peptides that induced significant T-cell responses are shown in orange, and those below this threshold are shown in green. The vertical dashed line indicate the filtering threshold for TCRdock PAE, and the points on the right side of this line are our negative designs. **(B,C)** Similar to (A) but for the (B) EBV, and (C) MAGE systems. No negative designs (i.e., designs below the respective TCRdock negative-PAE thresholds) were experimentally validated for these systems.

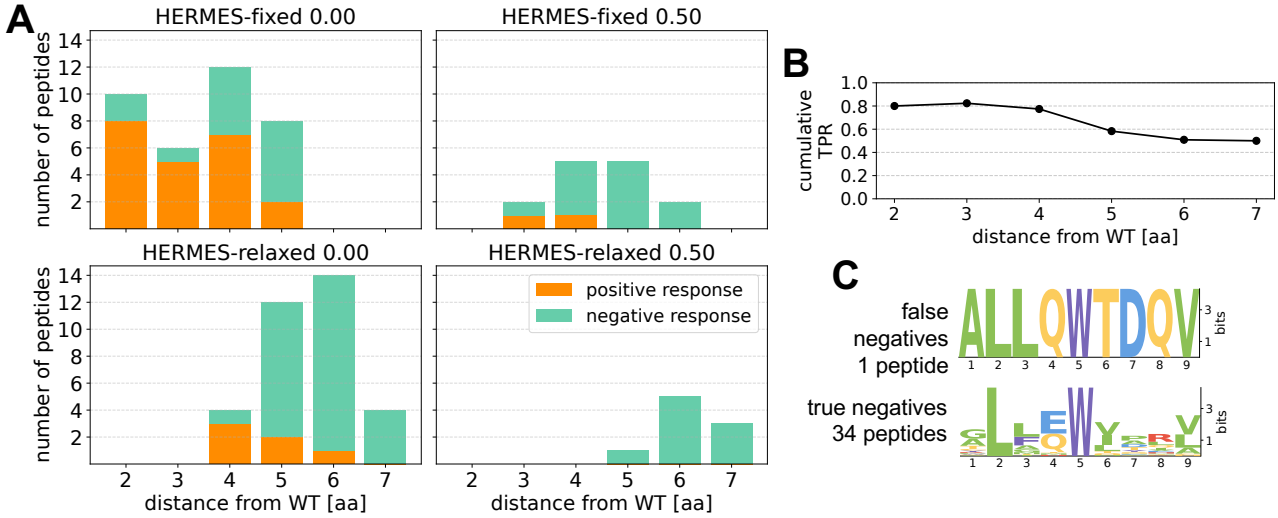

Figure S6. **Experimental T-cell activation by HERMES-designed peptides in the NY-ESO system**. **(A)** The number of peptides designed by different HERMES models (panels) that induced GFP expression beyond the T-cell activation threshold (positives; shown in orange) versus those that did not (negatives; shown in green) is shown. The designed peptides are grouped according to their minimum Hamming distance from any of the wild-type sequences used as design templates (Table S5). **(B)** The cumulative True Positive Rate (TPR) for designed peptides is shown as a function of their Hamming distance from any of the wild-types. **(C)** Logo plots show the sequence composition of the False Negative and the True Negative designs.

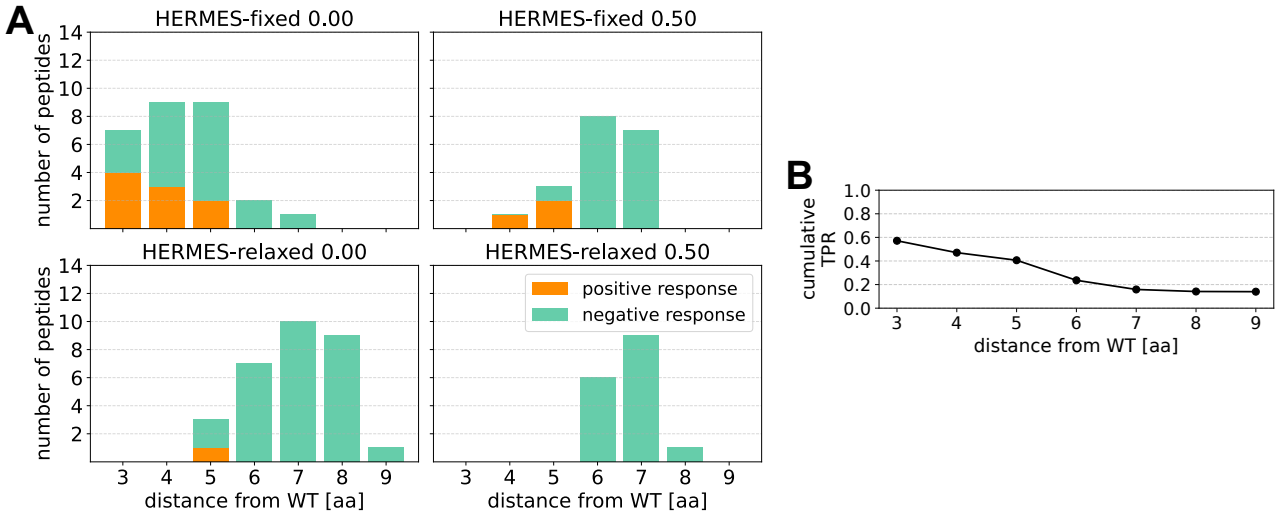

Figure S7. **Experimental T-cell activation by HERMES-designed peptides in the EBV system. (A-B)** Similar to (A-B) of Fig. S6 but for the EBV system.

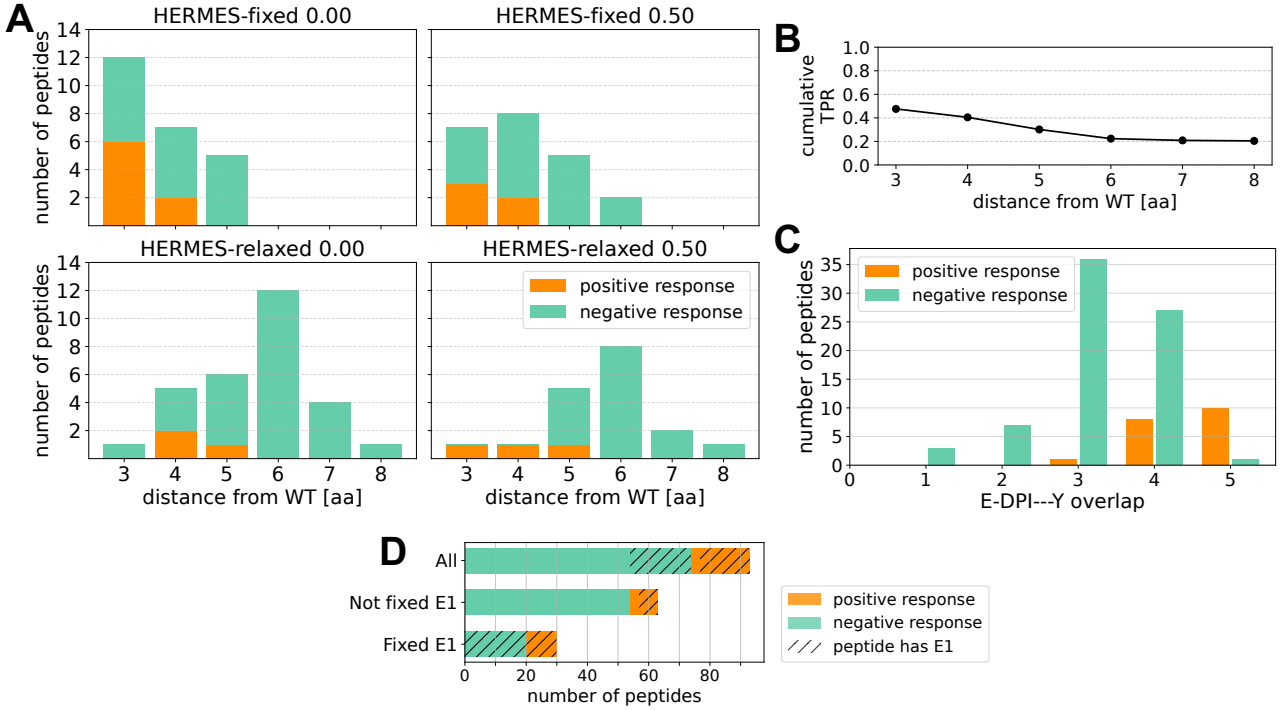

Figure S8. **Experimental T-cell activation by HERMES-designed peptides in the MAGE system. (A-B)** Similar to (A-B) of Fig. S6 but for the MAGE system. **(C)** The number of designs that induced GFP expression beyond the T-cell activation threshold (positive response; in orange) versus those that did not (negative response; in green) is shown as a function of the designs' overlap with the E-DPI---Y motif; this motif has been shown to enhance T-cell activation in this system [58]. The overlap is the number of amino acids that a design shares with the E-DPI---Y motif (maximum of 5). The fraction of positive designs increases with greater similarity to the motif. **(D)** Shown are the numbers of True Positives (orange) and False Positives (green) among designs generated under two protocols: one that enforced glutamic acid at position one (E_1)—labeled "fixed E_1"—-and one that did not enforce E_1—labeled "not fixed E_1". For each protocol, the dashed segments indicate the number of designs that contain E_1. The "All" bar presents the aggregate data from both protocols.

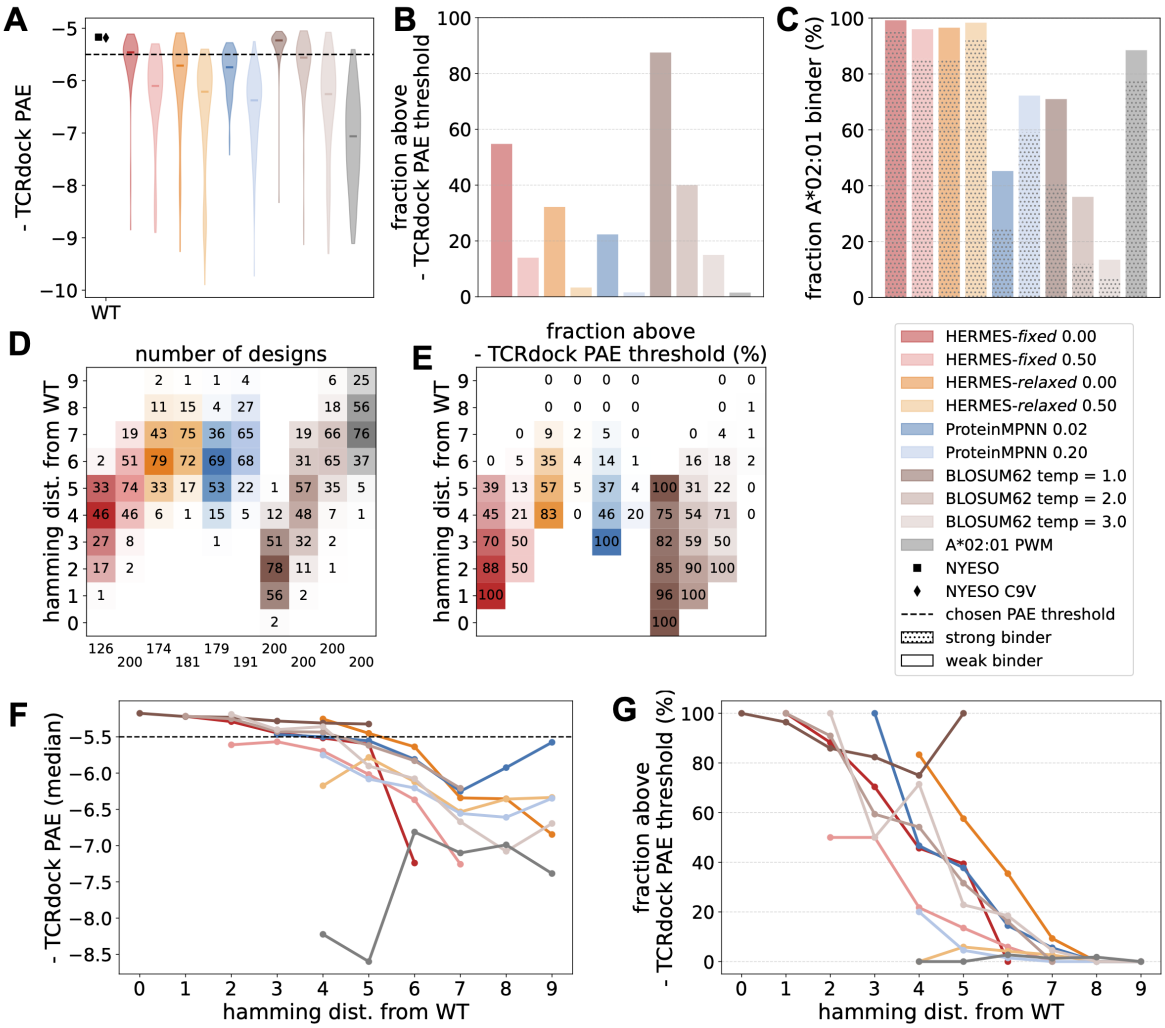

Figure S9. **Benchmarking of models for peptide design in the NY-ESO system.** **(A)** The distribution of the TCRdock negative-PAE's for peptides designed by different computational protocols (colors) are shown; in each case the score distribution for 100 independently designed peptides is shown. The horizontal dashed line indicates the PAE threshold chosen to classify positive designs. **(B)** The fractions of designed peptides that pass the TCRdock PAE threshold is shown for each method. **(C)** The fraction of designs that are predicted by NetMHCPan [37] to be presented by the template's MHC allele are shown; dotted regions correspond to strong binders (EL_rank ≤ 0.5), while the bar indicates all weak and strong binders (EL_Rank ≤ 2.0). HERMES designs are most consistently predicted to be presented by the MHC. **(D** The number of designs, and **(E)** the fraction of designs that pass the TCRdock PAE threshold are shown for each model (colors) at different Hamming distances from the wild-types; The numbers at the bottom of each column in (D) indicate the total number of unique designs. **(F)** The median of TCRdock negative-PAE across designs, and **(G)** the fraction of designs that pass the TCRdock PAE are shown for each model (colors) as a function of the Hamming distance from the wild-types

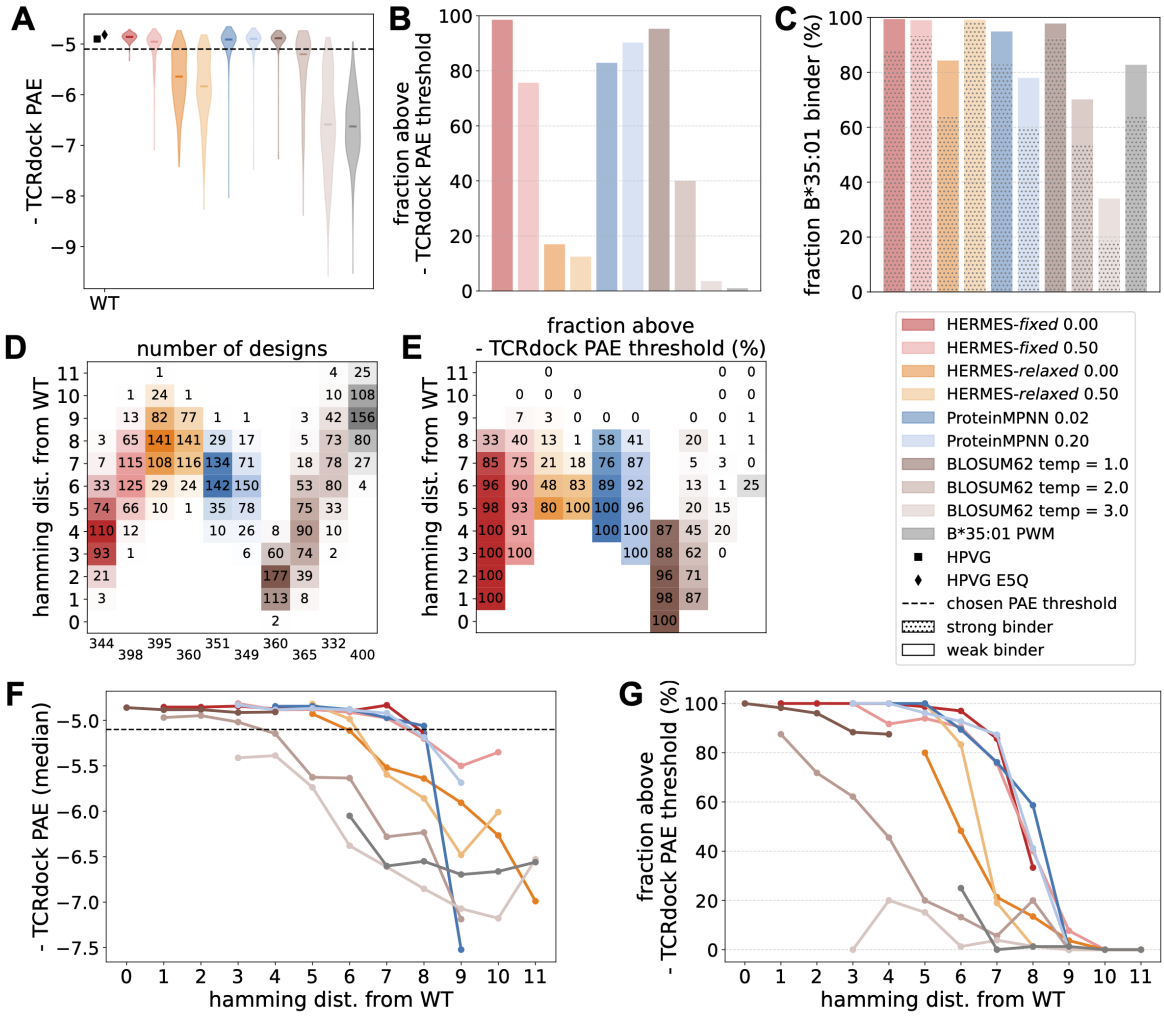

Figure S10. **Benchmarking of models for peptide design in the EBV system.** Similar to Fig. S9 but for the EBV system.

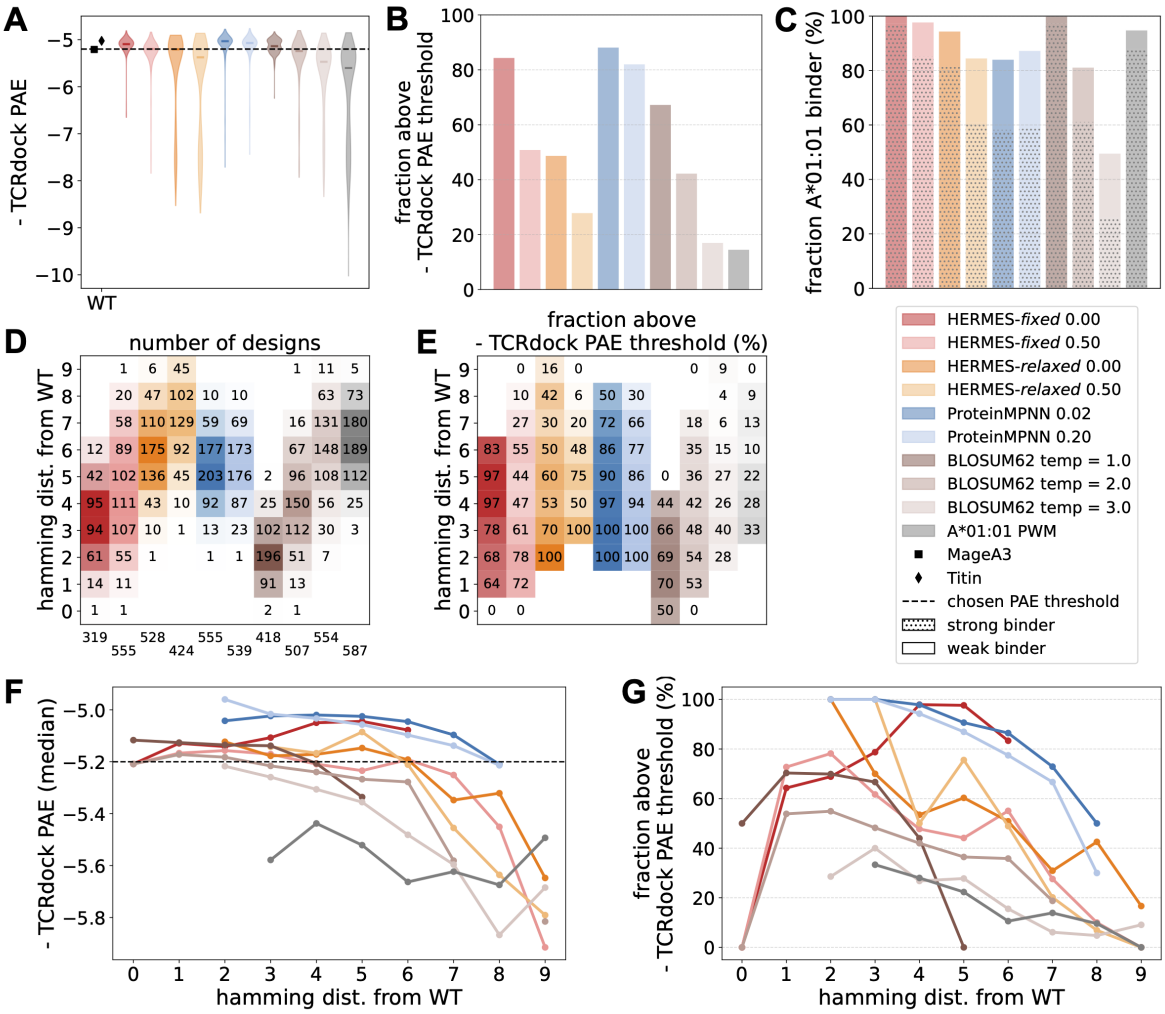

Figure S11. **Benchmarking of models for peptide design in the MAGE system.** Similar to Fig. S9 but for the MAGE system.

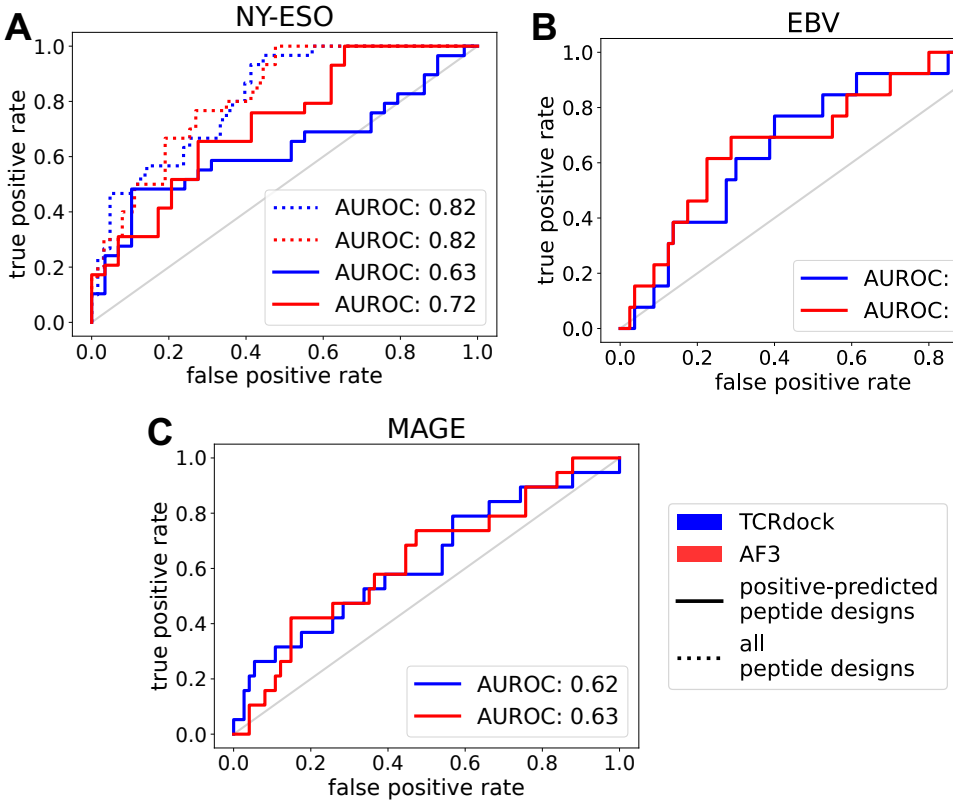

Figure S12. **Discrimination ability of TCRdock and AlphaFolf3 PAEs as design filters.** The ROC curves show the true positive versus false positive rates based on the TCRdock PAE (blue) and the AF3 PAE (red) for designed peptides made in the **(A)** NY-ESO, **(B)** EBV, and **(C)** MAGE systems; the experimentally measured peptide-induced T-cell activity (GFP%) is used to indicate the ground truth. For the NY-ESO system only we show ROC curves computed on both peptides that were predicted to activate T-cells (positives) based on their TCRdock PAE scores (full lines), and the mixture of positive and negative peptides (dotted lines). In the other systems, we only experimentally validated the peptides that we deemed to be activating (positive) based on their TCRdock PAE scores (full lines).