INVERSE FOLDING MODELS TRACK STRUCTURAL COMPATIBILITY IN PMHC-TCR COMPLEXES

Sergio E. Mares*
Adimab, Mountain View, CA 94043, USA
Center for Computational Biology, University of California, Berkeley, CA 94720, USA

Brian Petersen
Adimab, Mountain View, CA 94043, USA

Kyle Barlow
Adimab, Mountain View, CA 94043, USA

ABSTRACT

Predicting cross-reactivity in T-cell receptor (TCR)-based cancer therapies is a major challenge. Previous clinical studies failed to identify cross-reactive targets, ultimately resulting in fatal cardiac toxicity in two patients, and the combinatorial vastness of the promiscuous peptide-MHC (pMHC) landscape is beyond the reach of experimental sampling. While inverse folding (IF) models can sample peptide sequences from a distribution conditioned on structure, whether they recover the native peptide or map the tolerated neighborhood underlying cross-reactivity is unknown. Here we sampled peptides from four IF models across 43 pMHC-TCR crystal structures and scored them for predicted presentation and structural compatibility. We show that IF models do not recover the native peptide or rank binding strength. Instead, they sample from the sequence space compatible with the structural context they are conditioned on, offering a handle on the tolerated neighborhood that drives cross-reactivity. Together, these results establish IF model sampling distributions as an effective probe of the constraints governing peptide recognition and cross-reactivity, critical for TCR-based therapies.

1 INTRODUCTION

T-cell receptor (TCR)-based therapies have shown promising results for cancer treatment. However, TCR cross-reactivity poses a difficult challenge, as previous clinical studies have failed to identify cross-reactive targets, ultimately resulting in fatal cardiac toxicity in two patients (Linette et al., 2013; Cameron et al., 2013). Cross-reactive peptide-MHC-TCR (pMHC-TCR) interactions remain a fundamental barrier to scaling these personalized medicine approaches. Experimental methods sample a narrow slice of the pMHC landscape, and the combinatorially vast nature of the promiscuous pMHC is beyond the reach of experimental sampling. Computational approaches have shown promise in evaluating potential pMHC interactions, whereas TCR context poses a more difficult question. Previous work has shown that pMHC-TCR experimentally resolved structures provide a grounded structural prior for sequence recovery of the CDR3 interface. Inverse folding (IF) models sample sequences from a distribution conditioned on the structure, and across 38 pMHC-TCR complexes, they have been shown to recover the CDR3 interface residues, indicating that IF models capture non-trivial distributional shifts from the pMHC-TCR interface. In this work we asked whether IF models recover the native peptide of the crystal structure, whether the sequences they sample instead reflect the structural constraints of the interface and whether the presence of the TCR reshapes that distribution. We then asked what that sampled sequence space can tell us about cross-reactivity.

*Work performed during an internship at Adimab, LLC. Correspondence to sergio.mares@berkeley.edu.

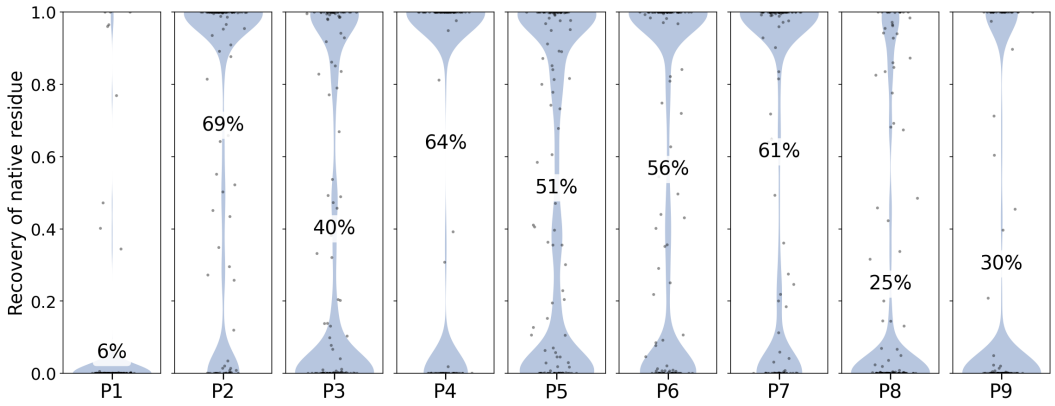

Figure 1: Per-position recovery of the native residue, 30 9-mer pMHC-TCR structures across 4 models.

2 Materials and Methods

2.1 Panel and models

We downloaded 43 human pMHC-TCR crystal structures from the Protein Data Bank (Berman et al., 2000) (Supplementary Table S1). Every structure includes a bound peptide, the MHC chains and a complete TCR $\alpha\beta$ heterodimer (Bjorkman et al., 1987) (Supplementary Figure S1). Out of these, 40 structures are HLA-A*02:01 and the remaining three are one HLA-B*51:01 (4MJI), one HLA-B*08:01 (1MI5) and one HLA-B*35 (2AK4). The average crystallographic resolution is 2.48 Å. This set has 29 distinct peptides and 22 distinct CDR3β sequences. We applied four models: ProteinMPNN (Dauparas et al., 2022), ProteinMPNN (no MHC; the same model architecture retrained with MHC/TCR-class structures excluded), ESM-IF1 (Hsu et al., 2022) and LigandMPNN (Dauparas et al., 2025) (Supplementary Table S2). For every run, we only sampled the peptide chain. Additionally, as an ablation experiment, we removed the TCR from the same structures and sampled 10,000 sequences for each condition at a temperature of T = 0.1. For pMHC binding affinity prediction we applied MHCflurry (O'Donnell et al., 2020) and ESMCBA (EvolutionaryScale Team, 2024; Mares et al., 2025), and set a threshold of 500 nM for strong binders (Sette et al., 1994).

3 Results

3.1 Models Sample a Strongly Position-Biased Sequence Space

To establish how much of the peptide sequence space these models explore, we sampled 10,000 designs per structure and model combination in the pMHC-TCR context at T=0.1. All four models sampled on average 49 unique sequences per structure, a highly restricted space (0.5%), yet these sets are largely model-specific, with ProteinMPNN variants sharing more designs with each other than any other pair of models (Supplementary Figure S2). Although exact overlap is minimal (a median of 2.7% across models for a single structure), ProteinMPNN models are within 2 mutations of each other, whereas ESM-IF1 designs sit 4-5 mutations from the nearest design of any other model (Supplementary Figure S3). Per-position Shannon entropy shows that every model imposes strong position-specific bias (Supplementary Figures S4–S8).

3.2 The two MHC anchors impose different constraints and recover differently

As shown in prior work, IF models sample a sequence-distribution conditioned on the structure and have been shown to identify functionally critical residues (Dauparas et al., 2022; Hsu et al., 2022; Ribeiro-Filho et al., 2024; Deng et al., 2025; Luo et al., 2023). We asked whether critical binding positions at the N-terminal (P2) and C-terminal (PΩ) peptide contacts shape the recovery

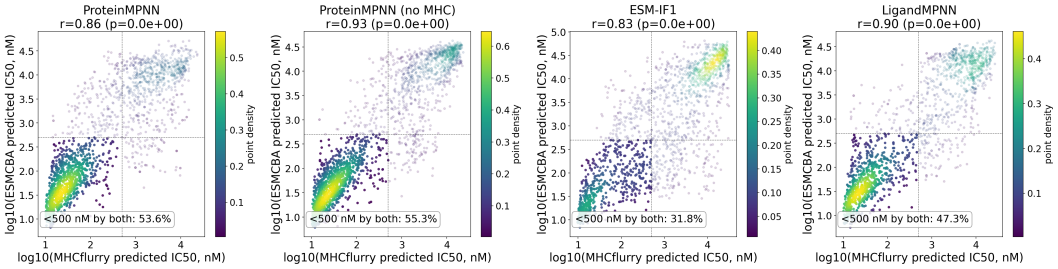

Figure 2: MHCflurry vs. ESMCBA predicted IC50 for every unique design, one panel per model.

of the native sequence across the 30 nonamer structures and 4 models. We found that recovery is biased in an unexpected way: P2, seated in pocket B, constrains a specific side chain and recovers well above the interior average (69% vs 50% interior, $p = 0.0073$), whereas PΩ, seated in pocket F, constrains the backbone but not side-chain identity and recovers significantly below it (30% vs 50%, $p = 0.010$) (Falk et al., 1991; Madden, 1995) (Figure 1). Both are structurally critical for pMHC binding, yet only P2 is constrained to a single side chain. Canonically, HLA-A*02:01 anchor residues are aliphatic non-polar (Rammensee et al., 1999). The native peptides' anchor residues are more hydrophobic than their own interior (Kyte-Doolittle scale (Kyte and Doolittle, 1982), p < 0.0001 for both), and the designs reproduce that chemistry without recovering the native residue (p < 10^{-10} for both) (Supplementary Figure S9).

To evaluate if the recovery of the native sequence is a function of the structure's resolution, we first pooled across the 43 structures; the correlation does not reach significance (r = -0.22, p = 0.16), and only ESM-IF1 shows a relationship individually (r = -0.37, p = 0.015; Supplementary Figure S10). We then used five structures (2P5W, 2BNR, 2F53, 2F54 and 2PYE) that are independent crystals of the identical peptide (SLLMWITQC, the NY-ESO-1 epitope (Chen et al., 2005; Sami et al., 2007)) differing primarily in resolution (1.9–2.7 Å). Declining resolution decreased recovery for ProteinMPNN (no MHC) (r = -0.97, p = 0.006) and, less clearly, ESM-IF1 (r = -0.84, p = 0.073), and increased the number of unique designs for both ProteinMPNN variants (r = 0.97, p = 0.006 and r = 0.94, p = 0.017; Supplementary Figure S11). LigandMPNN showed the opposite trend (r = +0.19 and r = -0.16, both n.s.).

3.3 Peptides sampled by IF models are predicted to be presented

We asked whether those sampled peptides would be presented, scoring every unique design from the 40 HLA-A*02:01 structures for predicted HLA-A2 binding with two unrelated methods, MHCflurry (O'Donnell et al., 2020) and ESMCBA (EvolutionaryScale Team, 2024; Mares et al., 2025), neither of which had seen this panel (Figure 2). The two methods correlated strongly on the sampled designs (r = 0.83–0.93 per model in full context, 0.81–0.93 without the TCR), and the principal component analysis (PCA) of the ESMCBA embeddings showed that the first component (74.1% of variance) ranks designs in the same order as the ESMCBA prediction (ρ = 1.000) and correlates r = 0.88 with $\log_{10}$ MHCflurry IC_{50} (Supplementary Figure S12). We found that ProteinMPNN and LigandMPNN consistently sampled higher predicted affinity than ESM-IF1. Below the 500 nM cutoff by both methods (Sette et al., 1994), 47–55% of ProteinMPNN-family and LigandMPNN designs passed, compared with 32% for ESM-IF1 (χ^2 = 263.34, p = 9×10^{-57}).

3.4 Without the TCR, IF models sample more widely but less confidently

Given that the models sample sequences compatible with the pMHC groove, we then evaluated the effect of the TCR on what these models propose. We performed an ablation experiment, resampling every structure without the TCR from the input. Every model explored substantially more diverse sequence space without the TCR, with unique design counts roughly doubling or tripling (ProteinMPNN 2282 → 6949; ProteinMPNN (no MHC) 2619 → 4652; ESM-IF1 1858 → 3525; LigandMPNN 1723 → 4561; Supplementary Figure S13). This larger space skewed toward higher predicted affinity, but not uniformly across models. ProteinMPNN (no MHC) showed the largest shift (strong-binder rate by both methods 55%→81%), and LigandMPNN a moderate one (47%→55%),

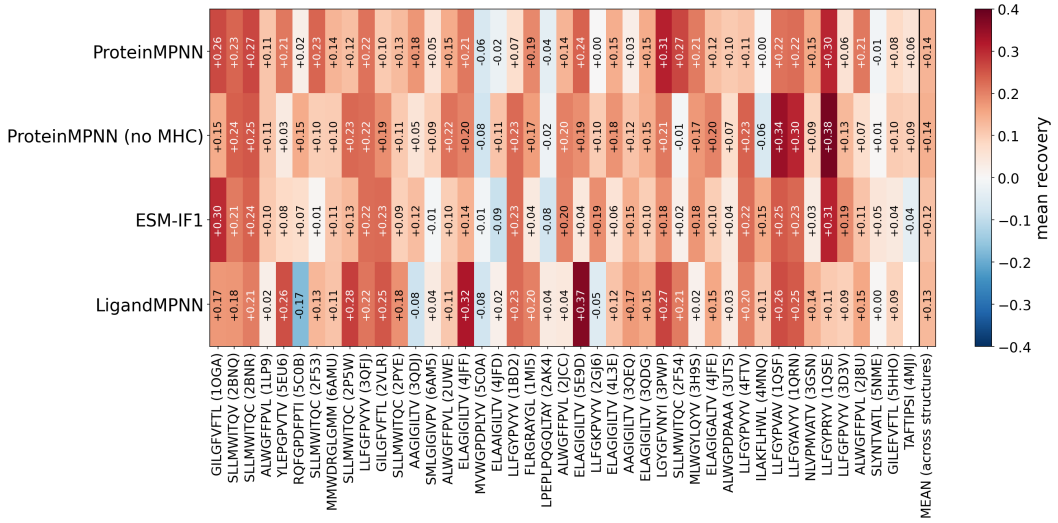

Figure 3: Recovery with TCR, per crystal and model, ordered by resolution.

while ProteinMPNN and ESM-IF1 changed less (54%→59% and 32%→38%) (Supplementary Figure S14). Recovery moved the other way, falling from 44%-48% to 29%-36%, and the size of that loss varied by crystal (Figure 3).

Leveraging the models' negative log likelihood (NLL) as a confidence metric for the sequences, we explored the distributional shifts in NLL induced by the TCR context. ProteinMPNN and ProteinMPNN (no MHC) report an average per-residue NLL for each design, where lower values indicate higher confidence; LigandMPNN reports an overall confidence score, for which higher values indicate higher confidence. Across the HLA-A*02:01 structures, removing the TCR shifted every model to a lower confidence (ProteinMPNN mean 0.988 $\rightarrow$ 1.239; ProteinMPNN (no MHC) 0.974 $\rightarrow$ 1.094; LigandMPNN 0.409 $\rightarrow$ 0.319; Mann-Whitney $p < 10^{-147}$ for all three; Supplementary Figure S15). Per-structure analysis showed the shift reaching significance in almost every structure (33 of 40 for ProteinMPNN, 33 of 39 for ProteinMPNN (no MHC) and 30 of 38 for LigandMPNN; Supplementary Figure S16). Removing the TCR moved the sampled sequences into a nearly disjoint set that both methods scored as strong binders. There was nearly no overlap (<1%) between the sequences sampled with and without the TCR, and the rate for <500 nM rose from 48% to 60%, indicating that the TCR interface shaped the sampling distribution toward the native epitope and far from the promiscuous pMHC sequence space.

3.5 Recovery Tracks Structural Compatibility, Not Binding Affinity

From the sequence recovery analysis, we found a bias at the P2 anchor, and the sampled sequences scored as strong pMHC binders in TCR context. We next asked whether sequence recovery can also serve as a proxy for functional pMHC-TCR interactions. We cross-referenced SKEMPI 2.0 (Jankauskaitė et al., 2019) and IEDB (Vita et al., 2019) against each structure's crystallized TCR clone, matched by CDR3 sequence, and identified 16 (structure, position) pairs with an experimentally measured mutation that alters TCR binding. At these positions, designs recover the native residue at 49% with the TCR and 27% without, whereas the rest were at 44–48% and 29–36%. Then, we evaluated the question with the experimental binding affinities. Restricting to single mutants yielded 23 complex/position pairs in SKEMPI across 11 human MHC class I complexes with measured $\Delta\Delta G$. At critical mutant positions (>1 kcal/mol), the recovery gain from TCR context was not significant (+0.139 vs +0.111; p = 0.19, Mann-Whitney; Supplementary Figure S17). We next used the crystallized DMF5 receptor (PDB: 6AM5 and 6AMU), which presents the same HLA-A*02:01 with two different peptides and has been pivotal in defining cross-reactive ligands in pMHC-TCR (Riley et al., 2018). The same study characterized two 10-mers and 13 peptide variants measured for KD and thermal stability (Riley et al., 2018). The first, SMLGIGIVPV, carries a central hydrophobic GIG core motif; the second, MMWDRGLGMM, carries a DRG motif with a

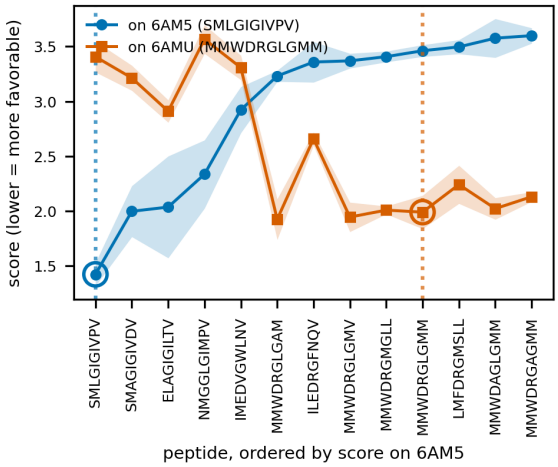

Figure 4: Thirteen DMF5 peptides scored on both backbones; rings mark each structure's own peptide.

charged residue. Models capture the well-characterized core, the GIG motif, which was recovered effectively compared to the rest of the peptide sequence ($p = 0.016$, paired Wilcoxon). We asked how the IF models score and rank each peptide on its own backbone and on the others. Across all three scoring models, each structure's own crystallized peptide ranked 2nd on average of 13 on its own backbone and 11th on average of 13 on the other (Figure 4), and the score therefore seems to read the backbone it is conditioned on. Across the nine peptides with a measured KD (four are reported as >500 μM or not detected), the score carried no information about affinity on either backbone, for any of the three models ($|\rho| \leq 0.32$, all $p > 0.4$; Supplementary Figure S18). IF models preferentially scored peptides that are compatible with a given groove and backbone, which is a quantity relevant to cross-reactivity, orthogonal to rank or binding strength. Scoring a graded peptide panel on each backbone reproduces this ordering, with measured tolerated substitutions closest to the crystal peptide and shuffled or random peptides furthest, though the sampled designs themselves fall outside the range of every measured binder, so the score orders compatibility rather than enumerating it (Supplementary Figure S19).

4 Discussion

IF models continue to prove useful in protein optimization and protein engineering (Dauparas et al., 2022; 2025; Hsu et al., 2022), and have shown their capabilities in recovering variable loop regions (Ribeiro-Filho et al., 2024). However, this raises the question of whether they capture the biophysics of an interface interaction or sample sequences with a structural prior. Here we show that structural conditioning is not sufficient to recovered the native sequence of the peptide, and that removing TCR context shifts both recovery and the sampled sequence space each model is confident in. The MHC anchors reveal mechanistic differences: pocket B constrains a single side chain, while pocket F constrains chemistry rather than side-chain identity. Residue recovery succeeds in pocket B and fails in pocket F. IF models sample a sequence distribution with strong predicted pMHC interactions, but the confidence scores reveal compatibility with the groove and backbone the model was conditioned on rather than the strength of binding. We also find that the ProteinMPNN models and ESM-IF1 sample distinct regions of sequence space and predict fewer binders, which we speculate may be due to ESM-IF1's training on monomeric structures, which differs from the multi-chain structures in ProteinMPNN's training data. This work positions IF models as tools to characterize the sequence space compatible with a groove and a TCR. That compatible sequence space is precisely the tolerated neighborhood that gives rise to cross-reactivity. Their sampling distributions therefore serve as a probe of the constraints governing peptide recognition and cross-reactivity, which are critical for TCR-based therapies.

ACKNOWLEDGEMENTS

Special thanks to Drs. Arvind Sivasubramanian, Juergen Nett, Garrett Rappazzo, Max Vásquez and the Adimab team for their support, feedback and helpful discussions.

REFERENCES

Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The Protein Data Bank. *Nucleic Acids Research*, 28(1):235–242, 2000. doi: 10.1093/nar/28.1.235.

Pamela J. Bjorkman, Mark A. Saper, Boudjema Samraoui, William S. Bennett, Jack L. Strominger, and Don C. Wiley. Structure of the human class I histocompatibility antigen, HLA-A2. *Nature*, 329(6139):506–512, 1987. doi: 10.1038/329506a0.

Brian J. Cameron, Andrew B. Gerry, Joseph Dukes, Jane V. Harper, Vivekanandan Kannan, Frayne C. Bianchi, Francis Grand, Joanna E. Brewer, Minnal Gupta, Gabriela Plesa, Giovanna Bossi, Annelise Vuidepot, Alex S. Powlesland, Alison Legg, Katherine J. Adams, Alan D. Bennett, Nicholas J. Pumphrey, Daniel D. Williams, Gwendolyn Binder-Scholl, Irina Kulikovskaya, Bruce L. Levine, James L. Riley, Angel Varela-Rohena, Edward A. Stadtmauer, Aaron P. Rapoport, Gerald P. Linette, Carl H. June, Namir J. Hassan, Michael Kalos, and Bent K. Jakobsen. Identification of a titin-derived HLA-A1-presented peptide as a cross-reactive target for engineered MAGE A3-directed T cells. *Science Translational Medicine*, 5(197):197ra103, 2013. doi: 10.1126/scitranslmed.3006034.

Jean-Li Chen, Guillaume Stewart-Jones, Giovanna Bossi, Nikolai M. Lissin, Linda Wooldridge, Eric Man Lik Choi, Gerhard Held, P. Rod Dunbar, Robert M. Esnouf, Malkit Sami, Jonathan M. Boulter, Pierre Rizkallah, Christoph Renner, Andrew Sewell, P. Anton van der Merwe, Bent K. Jakobsen, Gillian Griffiths, E. Yvonne Jones, and Vincenzo Cerundolo. Structural and kinetic basis for heightened immunogenicity of T cell vaccines. *Journal of Experimental Medicine*, 201 (8):1243–1255, 2005. doi: 10.1084/jem.20042323.

Y. Nancy Chen and Anna M. Bulek. Approval of the first TCR-based cell therapy. *Molecular Therapy*, 32(10):3241–3242, 2024. doi: 10.1016/j.ymthe.2024.09.007. Afamitresgene autoleucel (TECELRA) received U.S. FDA accelerated approval on 2 August 2024.

Justas Dauparas, Ivan Anishchenko, Nathaniel Bennett, Hua Bai, Robert J. Ragotte, Lukas F. Milles, Basile I. M. Wicky, Alexis Courbet, Rob J. de Haas, Neville Bethel, Philip J. Y. Leung, Timothy F. Huddy, Sam Pellock, Doug Tischer, Frederick Chan, Bruce Koepnick, Hannah Nguyen, Alex Kang, Banumathi Sankaran, Asim K. Bera, Neil P. King, and David Baker. Robust deep learning-based protein sequence design using ProteinMPNN. *Science*, 378(6615):49–56, 2022. doi: 10.1126/science.add2187.

Justas Dauparas, Gyu Rie Lee, Robert Pecoraro, Lulu An, Ivan Anishchenko, Cameron Glasscock, and David Baker. Atomic context-conditioned protein sequence design using LigandMPNN. *Nature Methods*, 22(4):717–723, 2025. doi: 10.1038/s41592-025-02626-1.

Arthur Deng, Karsten Householder, Fang Wu, Sebastian Thrun, K. Christopher Garcia, and Brian Trippe. Predicting mutational effects on protein binding from folding energy. *arXiv preprint arXiv:2507.05502*, 2025. doi: 10.48550/arXiv.2507.05502.

EvolutionaryScale Team. ESM Cambrian: revealing the mysteries of proteins with unsupervised learning. Technical report, EvolutionaryScale, 2024. https://evolutionaryscale.ai/blog/esm-cambrian.

Kirsten Falk, Olaf Rötzschke, Stefan Stevanović, Günther Jung, and Hans-Georg Rammensee. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature*, 351(6324):290–296, 1991. doi: 10.1038/351290a0.

David S. Hong, Brian A. Van Tine, Sarina Biswas, Cheryl McAlpine, Melissa L. Johnson, Anthony J. Olszanski, Jeffrey M. Clarke, Dejka Araujo, George R. Blumenschein, Partow Kebriaei, Quan Lin, Alex J. Tipping, Joseph P. Sanderson, Ruoxi Wang, Trupti Trivedi, Stavros Rafail, Nancy Rafail, Elliot Norry, and Aaron P. Rapoport. Autologous T cell therapy for MAGE-A4$^+$ solid tumors in adults with advanced disease: a phase 1 trial. *Nature Medicine*, 29(1):104–114, 2023. doi: 10.1038/s41591-022-02128-z.

Chloe Hsu, Robert Verkuil, Jason Liu, Zeming Lin, Brian Hie, Tom Sercu, Adam Lerer, and Alexander Rives. Learning inverse folding from millions of predicted structures. In *Proceedings of the 39th International Conference on Machine Learning (ICML)*, volume 162 of *PMLR*, pages 8946–8970, 2022.

Justina Jankauskaitė, Brian Jiménez-García, Justas Dapkūnas, Juan Fernández-Recio, and Iain H. Moal. SKEMPI 2.0: an updated benchmark of changes in protein-protein binding energy, kinetics and thermodynamics upon mutation. *Bioinformatics*, 35(3):462–469, 2019. doi: 10.1093/bioinformatics/bty635.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 2021. doi: 10.1038/s41586-021-03819-2.

Jack Kyte and Russell F. Doolittle. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, 157(1):105–132, 1982. doi: 10.1016/0022-2836(82)90515-0.

Gerald P. Linette, Edward A. Stadtmauer, Marcela V. Maus, Aaron P. Rapoport, Bruce L. Levine, Lyndsey Emery, Leslie Litzky, Adam Bagg, Beatriz M. Carreno, Patrick J. Cimino, Gwendolyn K. Binder-Scholl, Dominic P. Smethurst, Andrew B. Gerry, Nicholas J. Pumphrey, Alan D. Bennett, Joanna E. Brewer, Joseph Dukes, Jane Harper, Helen K. Tayton-Martin, Bent K. Jakobsen, Namir J. Hassan, Michael Kalos, and Carl H. June. Cardiovascular toxicity and titin cross-reactivity of affinity-enhanced T cells in myeloma and melanoma. *Blood*, 122(6):863–871, 2013. doi: 10.1182/blood-2013-03-490565.

Shitong Luo, Yufeng Su, Zuofan Wu, Chenpeng Su, Jian Peng, and Jianzhu Ma. Rotamer density estimator is an unsupervised learner of the effect of mutations on protein-protein interaction. *bioRxiv*, 2023. doi: 10.1101/2023.02.28.530137.

Dean R. Madden. The three-dimensional structure of peptide-MHC complexes. *Annual Review of Immunology*, 13:587–622, 1995. doi: 10.1146/annurev.iy.13.040195.003103.

Sergio Mares, Ariel Espinoza, and Nilah Ioannidis. Continued domain-specific pre-training of protein language models for pMHC-I binding prediction. In *Proceedings of the 20th Machine Learning in Computational Biology Meeting*, volume 311 of *PMLR*, pages 304–325, 2025.

Don Mason. A very high level of crossreactivity is an essential feature of the T-cell receptor. *Immunology Today*, 19(9):395–404, 1998. doi: 10.1016/S0167-5699(98)01299-7.

Leland McInnes, John Healy, and James Melville. UMAP: uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018. doi: 10.48550/arXiv.1802.03426.

Timothy J. O'Donnell, Alex Rubinsteyn, and Uri Laserson. MHCflurry 2.0: improved pan-allele prediction of MHC class I-presented peptides by incorporating antigen processing. *Cell Systems*, 11(1):42–48.e7, 2020. doi: 10.1016/j.cels.2020.06.010.

Hans-Georg Rammensee, Jutta Bachmann, Niels P. N. Emmerich, Oskar A. Bachor, and Stefan Stevanović. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics*, 50 (3–4):213–219, 1999. doi: 10.1007/s002510050595.

Helder V. Ribeiro-Filho, Gabriel E. Jara, João V. S. Guerra, Melina Cheung, Nathaniel R. Felbinger, José G. C. Pereira, Brian G. Pierce, and Paulo S. Lopes-de Oliveira. Exploring the potential of structure-based deep learning approaches for T cell receptor design. *PLOS Computational Biology*, 20(10):e1012489, 2024. doi: 10.1371/journal.pcbi.1012489.

Timothy P. Riley, Lance M. Hellman, Marvin H. Gee, Juan L. Mendoza, Jesus A. Alonso, Kendra C. Foley, Michael I. Nishimura, Craig W. Vander Kooi, K. Christopher Garcia, and Brian M. Baker. T cell receptor cross-reactivity expanded by dramatic peptide-MHC adaptability. *Nature Chemical Biology*, 14(10):934–942, 2018. doi: 10.1038/s41589-018-0130-4.

Malkit Sami, Pierre J. Rizkallah, Steven Dunn, Peter Molloy, Ruth Moysey, Annelise Vuidepot, Emma Baston, Penio Todorov, Yi Li, Fang Gao, Jonathan M. Boulter, and Bent K. Jakobsen. Crystal structures of high affinity human T-cell receptors bound to peptide major histocompatibility complex reveal native diagonal binding geometry. *Protein Engineering, Design and Selection*, 20(8):397–403, 2007. doi: 10.1093/protein/gzm033.

Alessandro Sette, Antonella Vitiello, Barbara Reherman, Peggy Fowler, Ramila Nayersina, W. Martin Kast, Cornelis J. M. Melief, Carla Oseroff, Lauri Yuan, John Ruppert, John Sidney, Maria-França del Guercio, Scott Southwood, Ralph T. Kubo, Robert W. Chesnut, Howard M. Grey, and Francis V. Chisari. The relationship between class I binding affinity and immunogenicity of potential cytotoxic T lymphocyte epitopes. *The Journal of Immunology*, 153(12):5586–5592, 1994. doi: 10.4049/jimmunol.153.12.5586.

Ian Sillitoe, Nicola Bordin, Natalie Dawson, Vaishali P. Waman, Paul Ashford, Harry M. Scholes, Camilla S. M. Pang, Laurel Woodridge, Clemens Rauer, Neeladri Sen, Mahnaz Abbasian, Sean Le Cornu, Su Datt Lam, Karel Berka, Ivana Hutarova Varekova, Radka Svobodova, Jon Lees, and Christine A. Orengo. CATH: increased structural coverage of functional space. *Nucleic Acids Research*, 49(D1):D266–D273, 2021. doi: 10.1093/nar/gkaa1079.

Randi Vita, Swapnil Mahajan, James A. Overton, Sandeep Kumar Dhanda, Sheridan Martini, Jason R. Cantrell, Daniel K. Wheeler, Alessandro Sette, and Bjoern Peters. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Research*, 47(D1):D339–D343, 2019. doi: 10.1093/nar/gky1006.

Linda Wooldridge, Julia Ekeruche-Makinde, Hugo A. van den Berg, Anna Skowera, John J. Miles, Mai Ping Tan, Garry Dolton, Mathew Clement, Sian Llewellyn-Lacey, David A. Price, Mark Peakman, and Andrew K. Sewell. A single autoimmune T cell receptor recognizes more than a million different peptides. *Journal of Biological Chemistry*, 287(2):1168–1177, 2012. doi: 10.1074/jbc.M111.289488.

A Supplementary Figures

PDB	Peptide	Len.	Allele	Res. (Å)	CDR3α	CDR3β
1BD2	LLFGYPVYV	9	HLA-A*02:01	2.50	AAMEGAQKLV	ASSYPGGGFYEQY
5HHO	GILEFVFTL	9	HLA-A*02:01	2.95	AGAGSQGNLI	ASSIRSSYEQY
2VLR	GILGFVFTL	9	HLA-A*02:01	2.30	AGAGSQGNLI	ASSSRASYEQY
1OGA	GILGFVFTL	9	HLA-A*02:01	1.40	AGAGSQGNLI	ASSSRSSYEQY
3QEQ	AAGIGILTV	9	HLA-A*02:01	2.59	AGGTGNQFY	AISEVGVGQPQH
1LP9	ALWGFFPVL	9	HLA-A*02:01	2.00	ALFLASSSFSKLV	ASSDWVSYEQY
2J8U	ALWGFFPVL	9	HLA-A*02:01	2.88	ALFLASSSFSKLV	ASSDWVSYEQY
2JCC	ALWGFFPVL	9	HLA-A*02:01	2.50	ALFLASSSFSKLV	ASSDWVSYEQY
2UWE	ALWGFFPVL	9	HLA-A*02:01	2.40	ALFLASSSFSKLV	ASSDWVSYEQY
2AK4	LPEPLPQGQLTAY	13	HLA-B*35	2.50	ALSGFYNTDKLI	ASPGLAGEYEQY
3UTS	ALWGPDPAAA	10	HLA-A*02:01	2.71	AMRGDSSYKLI	ASSLWEKLAKNIQY
5C0A	MVWGPDPLYV	10	HLA-A*02:01	2.46	AMRGDSSYKLI	ASSLWEKLAKNIQY
5C0B	RQFGPDFPTI	10	HLA-A*02:01	2.03	AMRGDSSYKLI	ASSLWEKLAKNIQY
3GSN	NLVPMVATV	9	HLA-A*02:01	2.80	ARNTGNQFY	ASSPVTGGIYGYT
4MJI	TAFTIPSI	8	HLA-B*51:01	2.99	ATDDDSARQLT	ASSLTGGGELF
4MNQ	ILAKFLHWL	9	HLA-A*02:01	2.74	AVDSATALPYGYI	ASSYQGTEAF
5EU6	YLEPGPVTV	9	HLA-A*02:01	2.02	AVLSSGGSNYKLT	ASSFIGGTDTQY
4JFD	ELAAIGILTV	10	HLA-A*02:01	2.46	AVNDGGRLT	AWSETGLGMGGWQ
4JFE	ELAGIGALTV	10	HLA-A*02:01	2.70	AVNDGGRLT	AWSETGLGMGGWQ
4JFF	ELAGIGILTV	10	HLA-A*02:01	2.43	AVNDGGRLT	AWSETGLGMGGWQ
3QDG	ELAGIGILTV	10	HLA-A*02:01	2.69	AVNFGGGKLI	ASSLSFGTEAF
3QDJ	AAGIGILTV	9	HLA-A*02:01	2.30	AVNFGGGKLI	ASSLSFGTEAF
6AM5	SMLGIGIVPV	10	HLA-A*02:01	2.39	AVNFGGGKLI	ASSLSFGTEAF
6AMU	MMWDRGLGMM	10	HLA-A*02:01	2.15	AVNFGGGKLI	ASSLSFGTEAF
4L3E	ELAGIGILTV	10	HLA-A*02:01	2.56	AVNFGGGKLI	ASSWSFGTEAF
2P5W	SLLMWITQC	9	HLA-A*02:01	2.20	AVRPLLDGTYIPT	ASSYLGNTGELF
2PYE	SLLMWITQC	9	HLA-A*02:01	2.30	AVRPLLDGTYIPT	ASSYLGNTGELF
2BNQ	SLLMWITQV	9	HLA-A*02:01	1.70	AVRPTSGGSYIPT	ASSYVGNTGELF
2BNR	SLLMWITQC	9	HLA-A*02:01	1.90	AVRPTSGGSYIPT	ASSYVGNTGELF
2F53	SLLMWITQC	9	HLA-A*02:01	2.10	AVRPTSGGSYIPT	ASSYVGNTGELF
2F54	SLLMWITQC	9	HLA-A*02:01	2.70	AVRPTSGGSYIPT	ASSYVGNTGELF
5NME	SLYNTVATL	9	HLA-A*02:01	2.94	AVRTNSGYALN	ASSDTVSYEQY
5E9D	ELAGIGILTV	10	HLA-A*02:01	2.51	AVTKYSWGKLQ	ASRPGWMAGGVELY
1QRN	LLFGYAVYV	9	HLA-A*02:01	2.80	AVTTDSWGKLQ	ASRPGLAGGRPEQY
1QSE	LLFGYPRYV	9	HLA-A*02:01	2.80	AVTTDSWGKLQ	ASRPGLAGGRPEQY
1QSF	LLFGYPVAV	9	HLA-A*02:01	2.80	AVTTDSWGKLQ	ASRPGLAGGRPEQY
2GJ6	LLFGKPVYV	9	HLA-A*02:01	2.56	AVTTDSWGKLQ	ASRPGLAGGRPEQY
3D3V	LLFGFPVYV	9	HLA-A*02:01	2.80	AVTTDSWGKLQ	ASRPGLAGGRPEQY
3H9S	MLWGYLQYV	9	HLA-A*02:01	2.70	AVTTDSWGKLQ	ASRPGLAGGRPEQY
3PWP	LGYGFVNYI	9	HLA-A*02:01	2.69	AVTTDSWGKLQ	ASRPGLAGGRPEQY
3QFJ	LLFGFPVYV	9	HLA-A*02:01	2.29	AVTTDSWGKLQ	ASRPGLAGGRPEQY
4FTV	LLFGYPVYV	9	HLA-A*02:01	2.74	AVTTDSWGKLQ	ASRPGLMSAQPEQY
1MI5	FLRGRAYGL	9	HLA-B*08:01	2.50	ILPLAGGTSYGKLT	ASSLGQAYEQY

Table S1: The forty-three-structure panel. PDB accession, native peptide, peptide length, MHC allele, crystallographic resolution and the crystallised CDR3α and CDR3β sequences for every structure in the panel. Rows are ordered by CDR3α then CDR3β, grouping structures that share a TCR clone.

Model	Params	Architecture	Training data
ProteinMPNN	1.66M	Message-passing GNN	PDB (`pdb_2021aug02`), general backbones
ProteinMPNN (no MHC)	1.66M	Message-passing GNN (identical to above)	Same corpus, MHC/TCR-class structures excluded
ESM-IF1	141.7M	GVP-GNN + transformer	CATH-derived backbones
LigandMPNN	2.62M	Message-passing GNN + ligand context	PDB, augmented with explicit ligand/heteroatom context

Table S2: **Models compared.** Parameter counts verified directly from each checkpoint's `state_dict`. ESM-IF1 is the only tool whose training set is built substantially from AlphaFold2-predicted structures (Jumper et al., 2021) rather than solved crystal structures (∼12M predicted backbones (Hsu et al., 2022), on top of its CATH-derived crystal set (Sillitoe et al., 2021)).

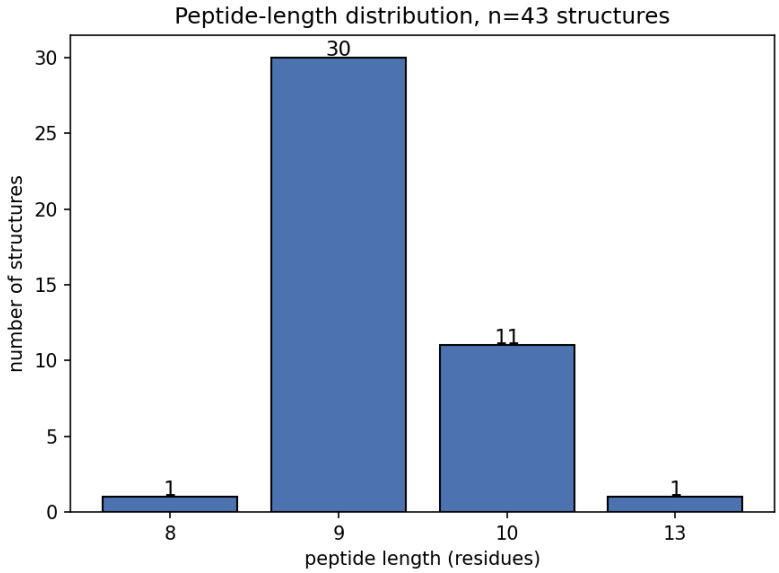

Figure S1: Peptide length distribution across all 43 structures in the panel.

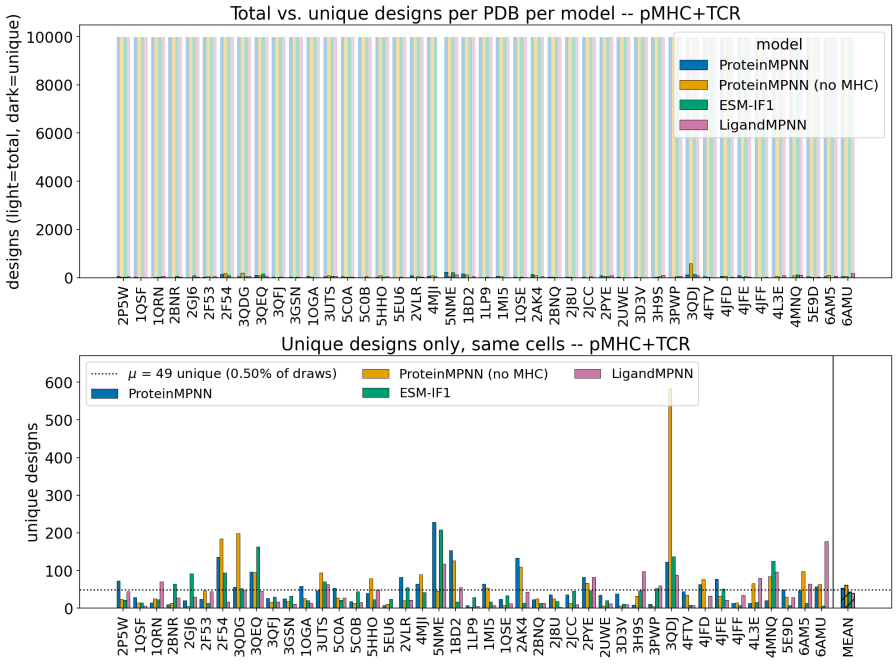

Figure S2: Total versus unique designs from 10,000 raw draws per structure and model at T=0.1.

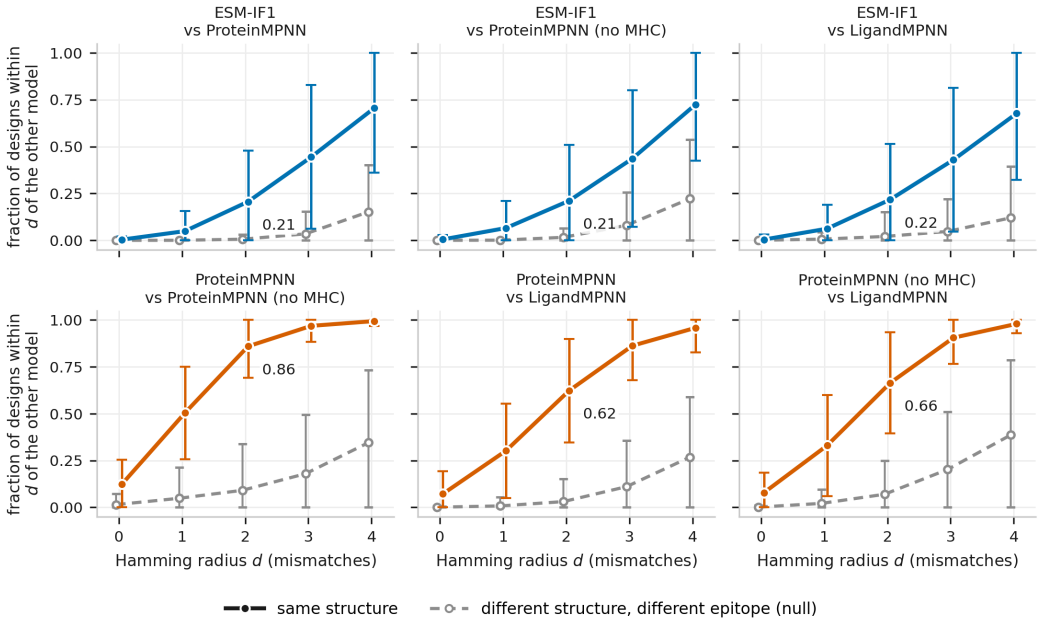

Figure S3: Cumulative fraction of one model's unique designs lying within d mismatches of the nearest design of another model, one panel per model pair against a different-structure null, labeled at $d = 2$, where the separation between the ProteinMPNN family and ESM-IF1 is widest.

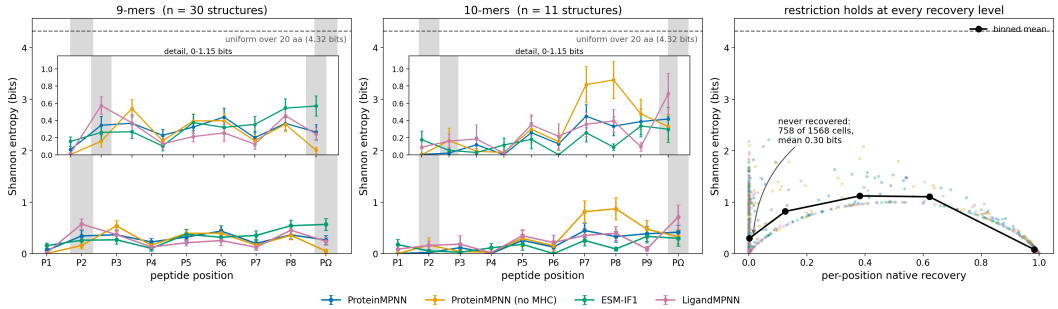

Figure S4: Per-position Shannon entropy of each model's design distribution pooled over the 43 structures (left) and against per-position native recovery (right), with anchor positions shaded and error bars showing the standard error across structures.

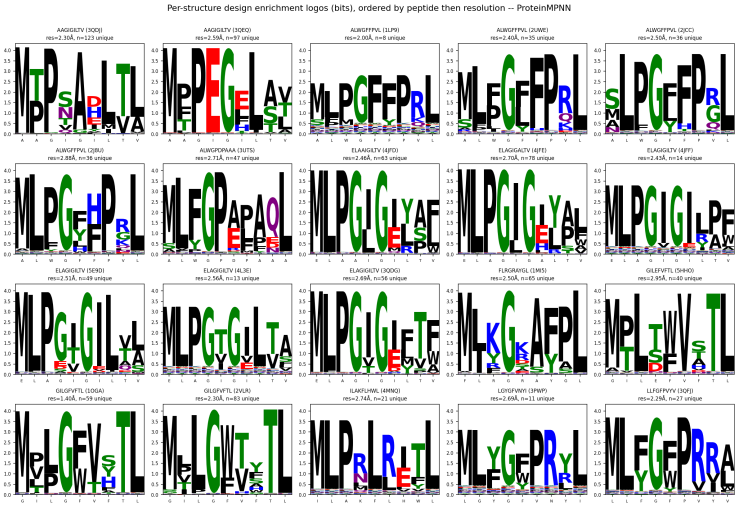

Figure S5: Per-structure design logos for ProteinMPNN, ordered by native peptide then resolution.

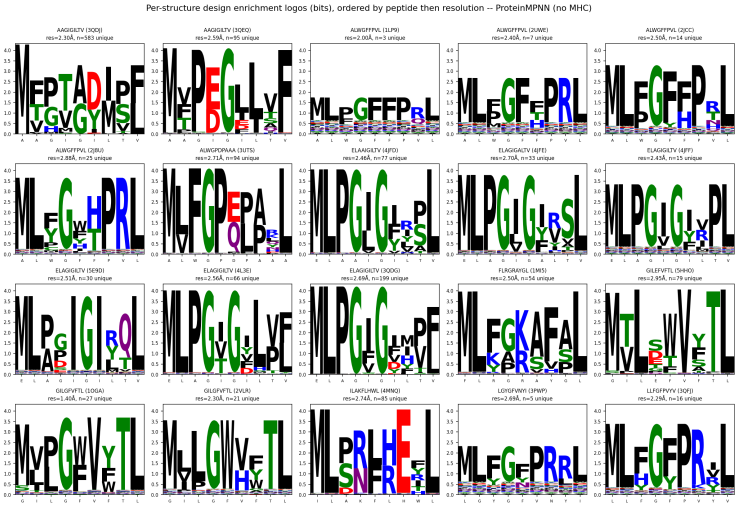

Figure S6: Per-structure design logos for ProteinMPNN (no MHC).

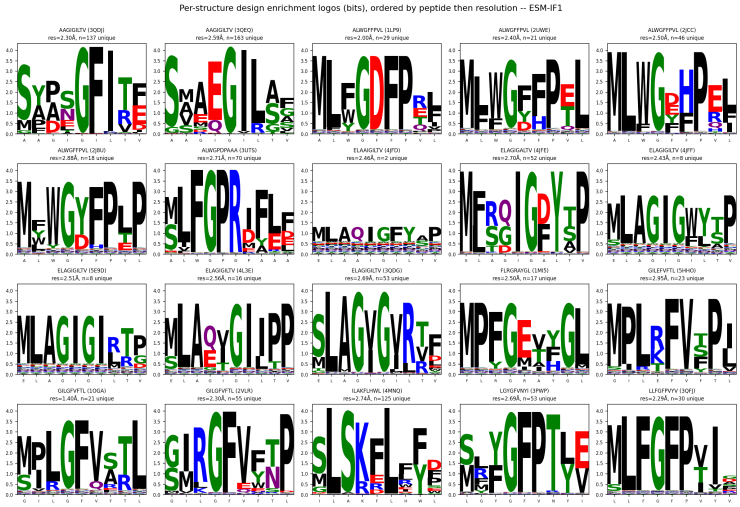

Figure S7: Per-structure design logos for ESM-IF1.

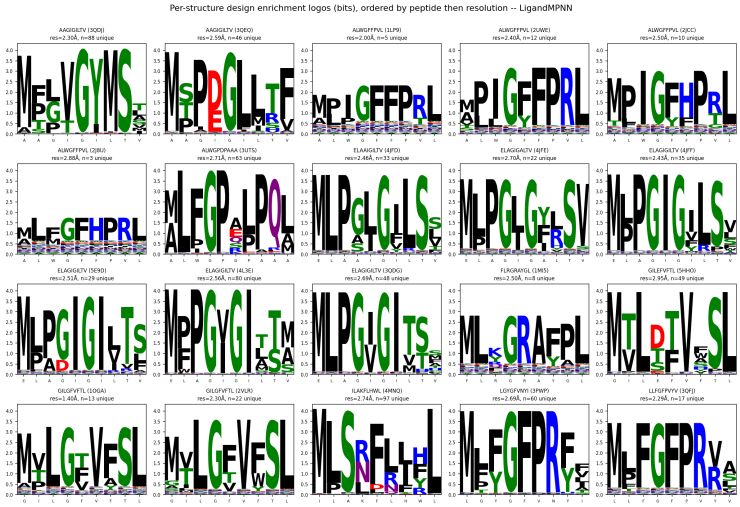

Figure S8: Per-structure design logos for LigandMPNN.

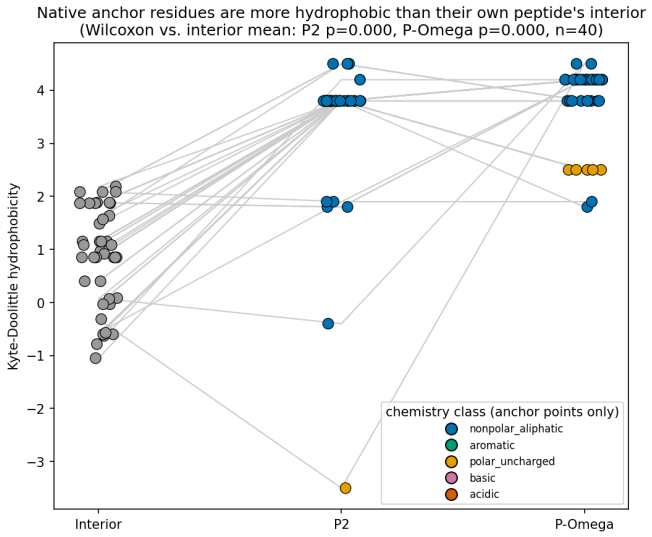

Figure S9: Kyte-Doolittle hydrophobicity of the native P2 and PΩ residues against each peptide's own interior mean, across the 40 HLA-A*02:01 structures.

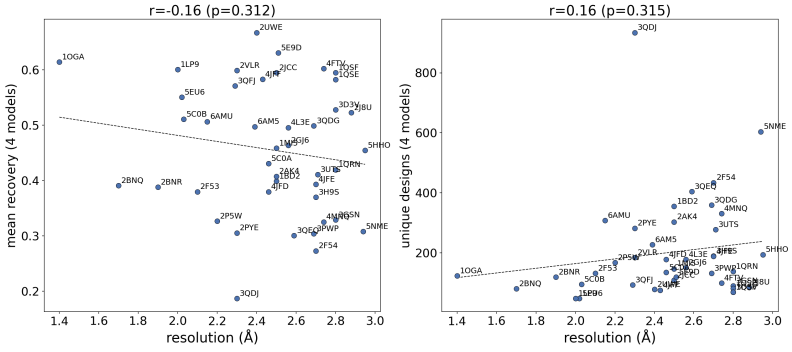

Figure S10: Resolution and unique-design count against mean recovery pooled across the panel ($r = -0.22$, $p = 0.162$ and $r = 0.15$, $p = 0.323$, $n = 43$).

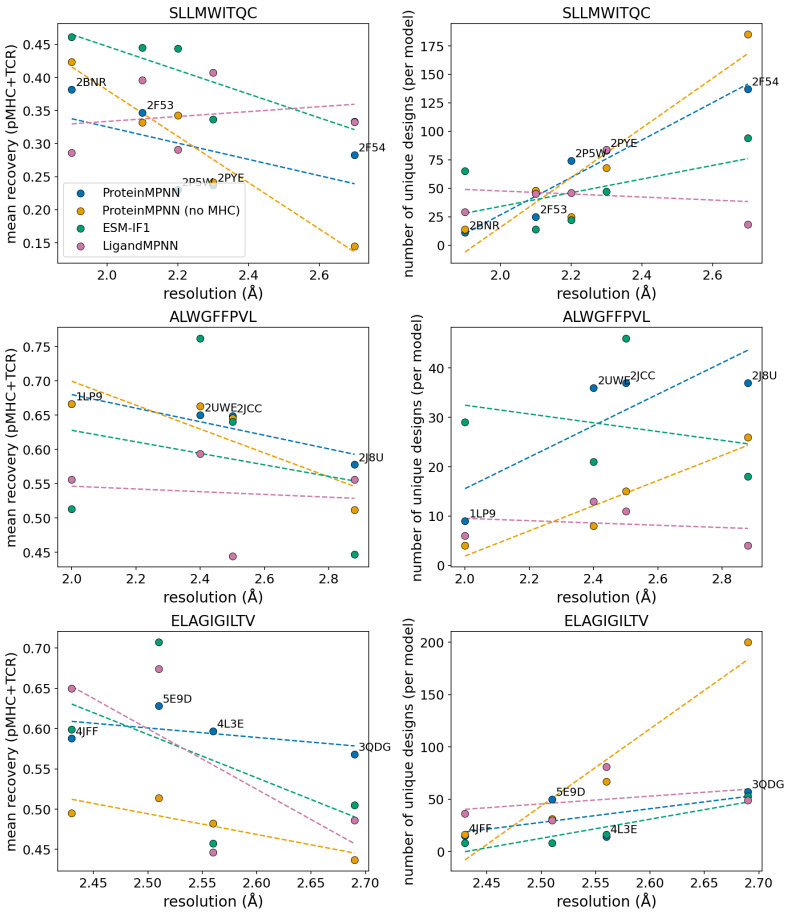

Figure S11: Resolution dependence in five same-peptide replicate structures (2P5W, 2BNR, 2F53, 2F54 and 2PYE), one panel per model.

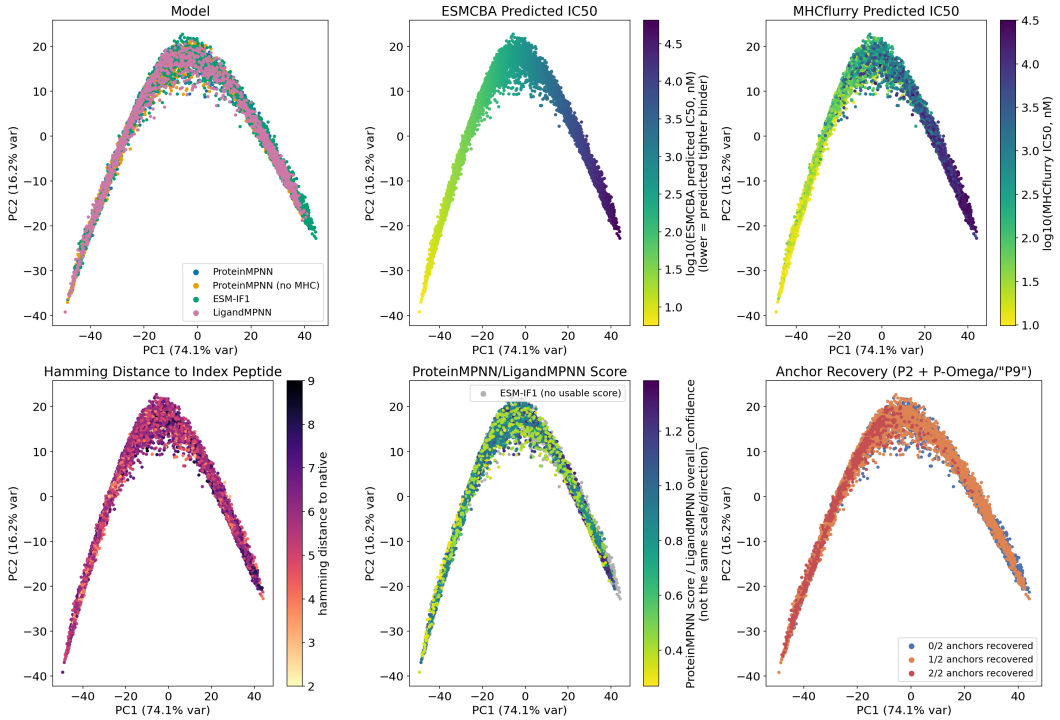

Figure S12: PCA of the ESMCBA embeddings of every unique design, recolored six ways, with the share of variance each component explains marked on its axis.

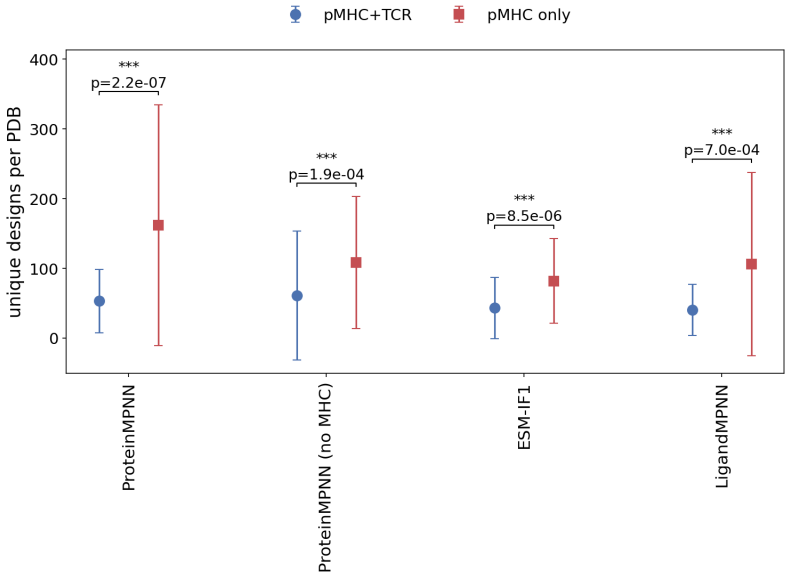

Figure S13: Unique designs per model, mean ± SD across structures, with and without the TCR.

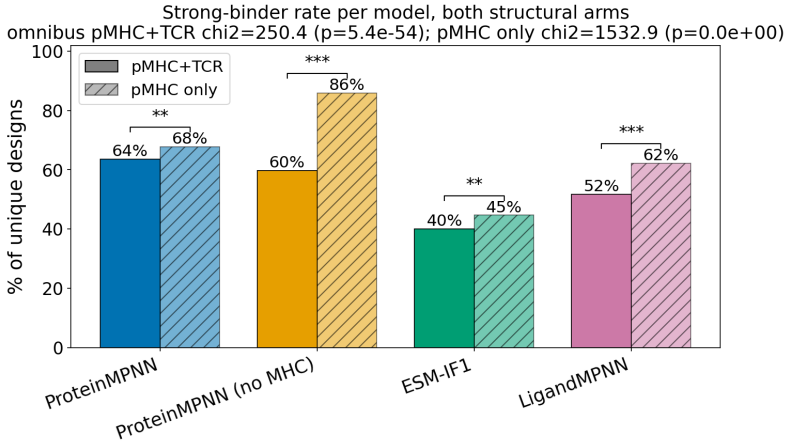

Figure S14: Strong-binder rate (MHCflurry IC50 < 500 nM) per model with the TCR present (solid bars) and removed (hatched), with Bonferroni-corrected Fisher tests of each model's shift.

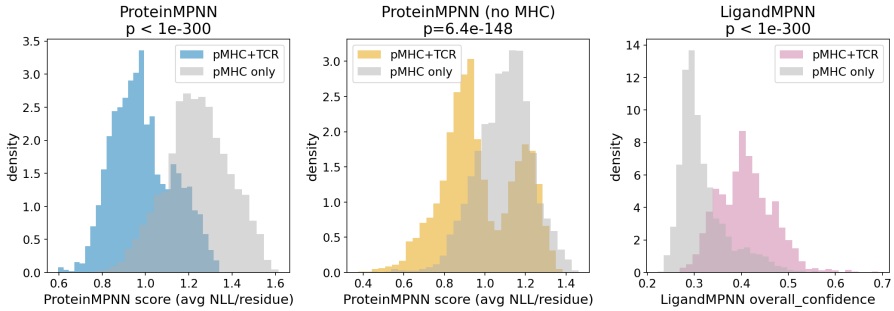

Figure S15: Each model's own confidence score for its designs, pooled across the 40 HLA-A*02:01 structures, with and without the TCR.

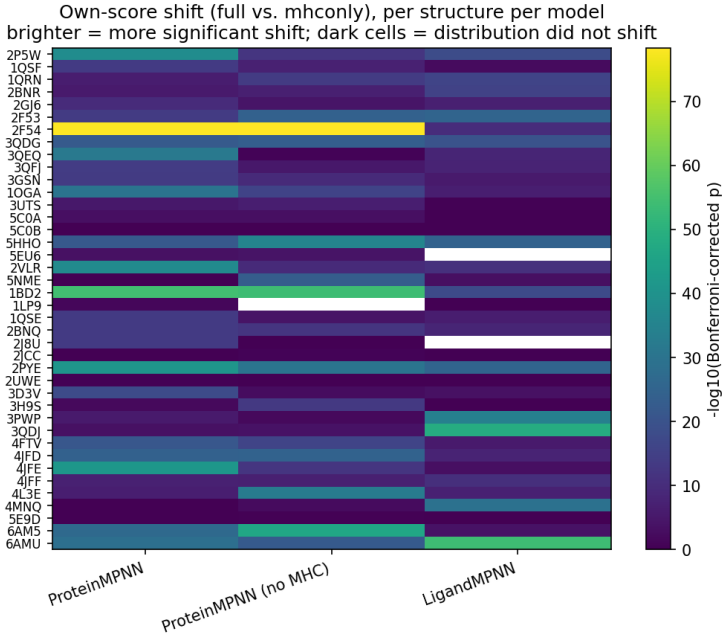

Figure S16: Bonferroni-corrected significance of the own-score shift between arms, per structure and model.

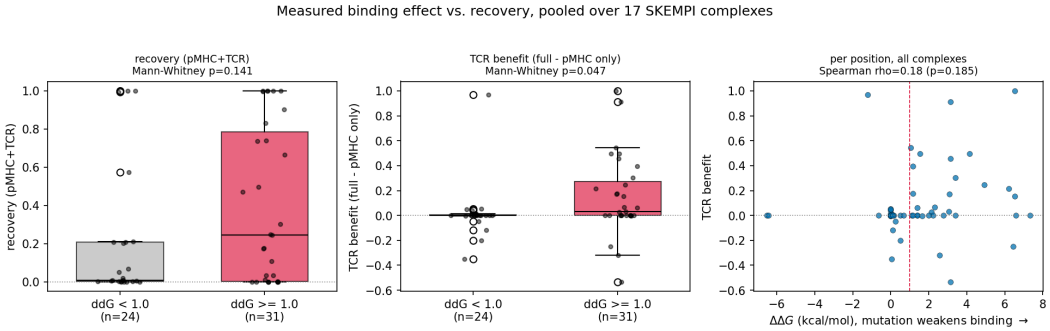

Figure S17: Measured binding effect against recovery, pooled over every SKEMPI complex with designs.

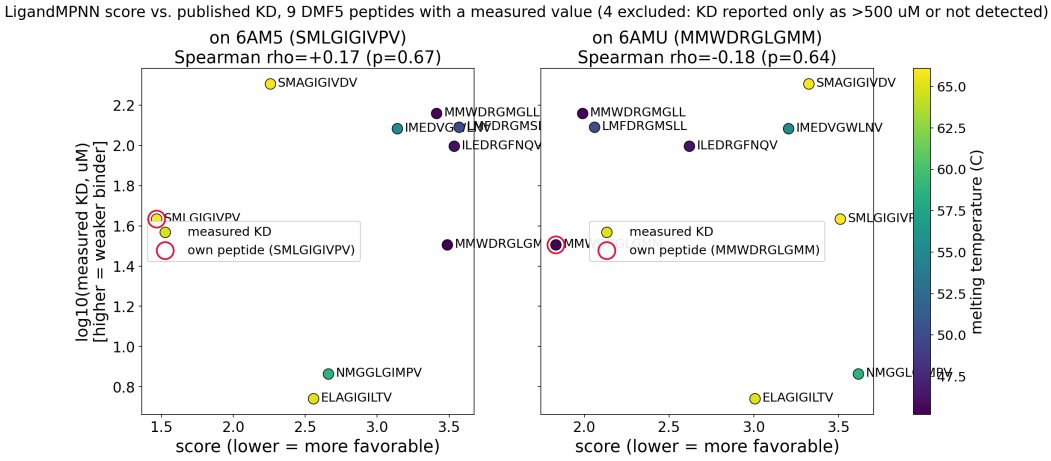

Figure S18: LigandMPNN score against published K_D for the nine DMF5 peptides with a measured value, scored on the 6AM5 (left) and 6AMU (right) backbones.

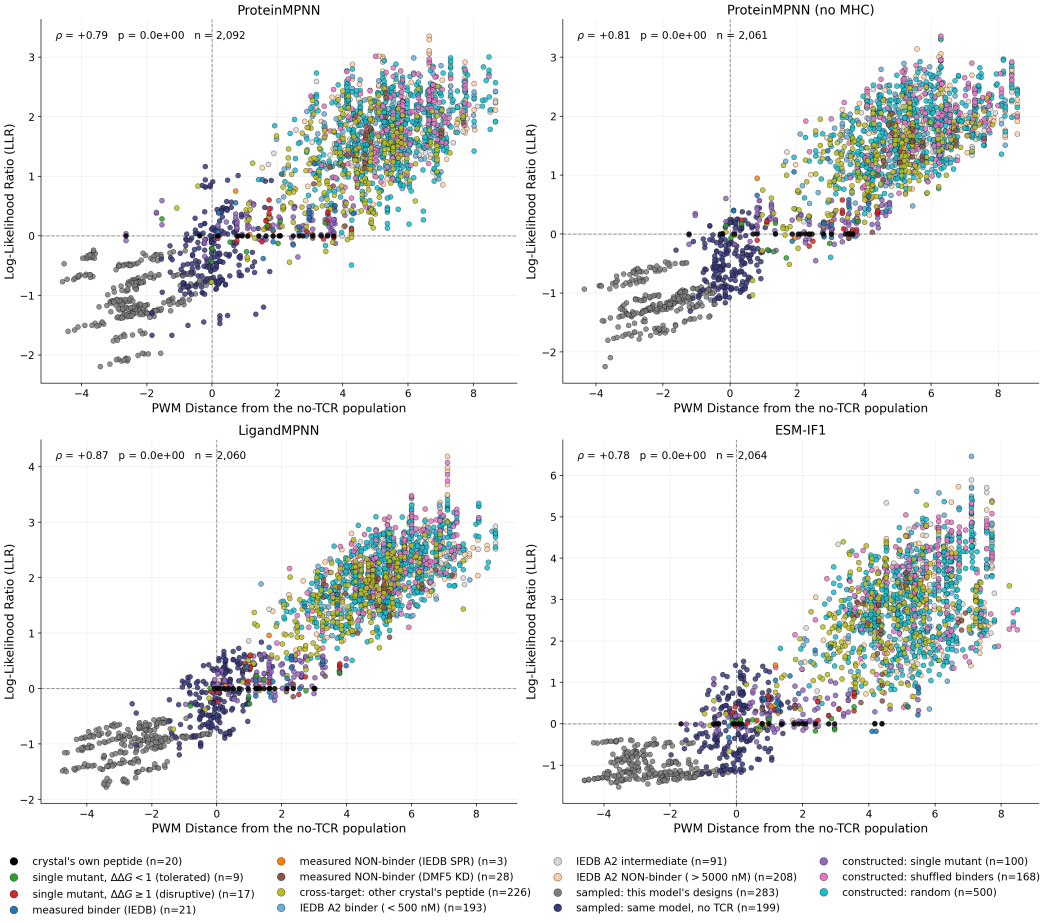

Figure S19: Profile distance against model score for a graded peptide panel on 20 backbones, both axes in nats per residue relative to each crystal's own peptide and each class, showing that the two measures agree (ρ = +0.94 to +0.96 for the MPNN models) and that the sampled designs occupy a region containing no measured binder.