Science Advances
AAAS

Supplementary Materials for

RACER-m leverages structural features for sparse T cell specificity prediction

Ailun Wang *et al.*

Corresponding author: Xingcheng Lin, xingcheng_lin@ncsu.edu; Jason T. George, jason.george@tamu.edu

Sci. Adv. **10**, eadl0161 (2024)
DOI: 10.1126/sciadv.adl0161

This PDF file includes:

1 Supporting Methods

1.1 Training data selection for RACER-m.

The RACER-m training set consists of TCR-p-MHC complex structures restricted to the HLA-A*02:01 allele, collected from the Protein Data Bank, which initially comprises 66 complex structures. However, it was observed that when trained on these 66 structures, RACER-m systematically underestimated the binding affinities of strong binders specific to NLV peptides and their variants. To address this issue, we incorporated three additional structures of strong binders from [43] in which 6 strong binders were reported when combining NLV variants with TCR RA14 and 3 of them were provided with p-MHC structures. By combining these p-MHC structures with TCR RA14 to form the TCR-p-MHC complex structure and adding them as supplementary training cases, we expanded the training set to a size of 69. The inclusion of these three NLV strong binders effectively resolved the systematic underestimation problem concerning the predictions of NLV-specific strong binders, while preserving the excellent predictive power for other strong binders in the ATLAS dataset.

1.2 Collection of point-mutant weak binders for 1E6.

To test the performance of RACER-m in terms of discerning strong binders from point-mutant weak binders, we collected point-mutant weak binders from a comprehensive peptide-mutagenesis study by Bulek *et al.* [21]. Through the mutational scan, Bulek *et al.* assessed the impact of point- mutations on the binding of peptide ALWGPDPAAA to the 1E6 TCR with the tumor necrosis factor (TNF). Since it was pointed out that the 1E6 TCR was tolerant to changes in peptide residues Ala1, Leu2, Ala8, Ala9 and Ala10 [21], we collected point-mutations at positions 3 to 7 with TNF equal or smaller than 25, and considered them as point-mutant weak binders for 1E6.

2 Supporting Figures

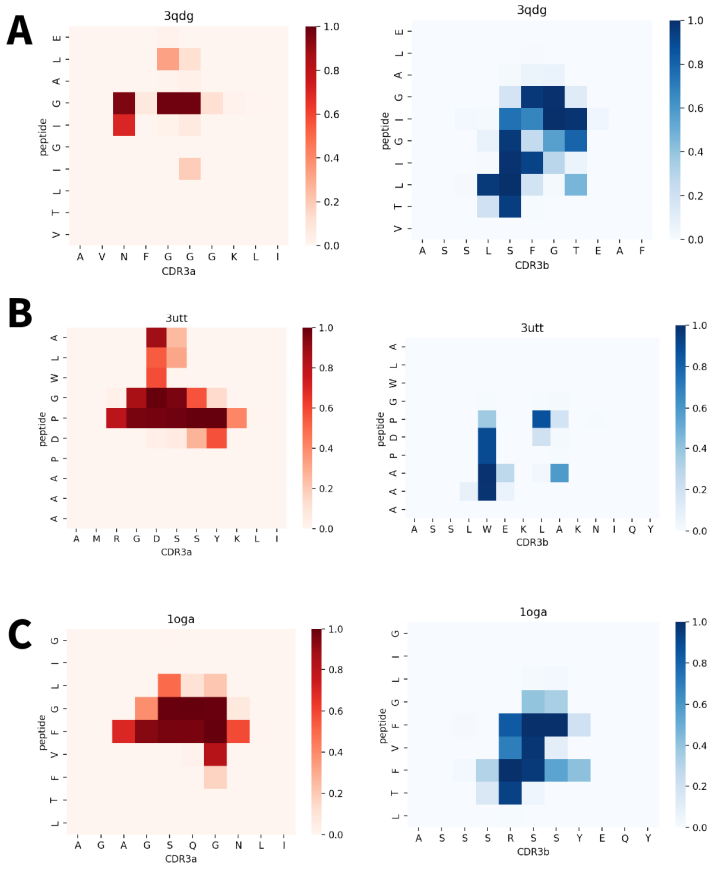

Figure S1: **Contact maps of crystal structures 3QDG, 3UTT and 1OGA.** Each contact map was calculated by measuring the proximity $W_{i,j}$ between each residues of peptide (residue i) and CDR loops (residue j) based on their mutual distance (d) using a smoothed step function: $W_{i,j} = (1 - \tanh(d - d_{max}))/2$, where $d_{max} = 8.5$Å. Only Cβ atoms were used for the mutual distance calculation (except for glycine, where the Cα atom was used).

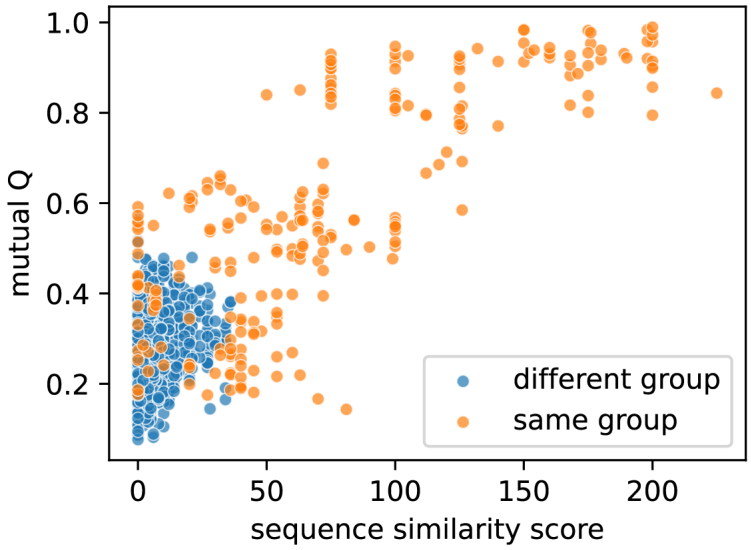

Figure S2: **Relationship between structure and sequence similarities of TCR-pMHC complexes.**

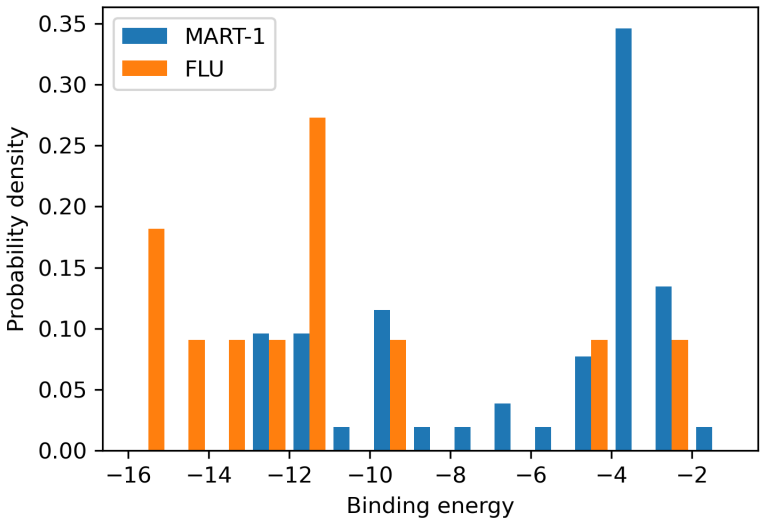

Figure S3: **Distribution of predicted binding energies of MART-1 and FLU.**

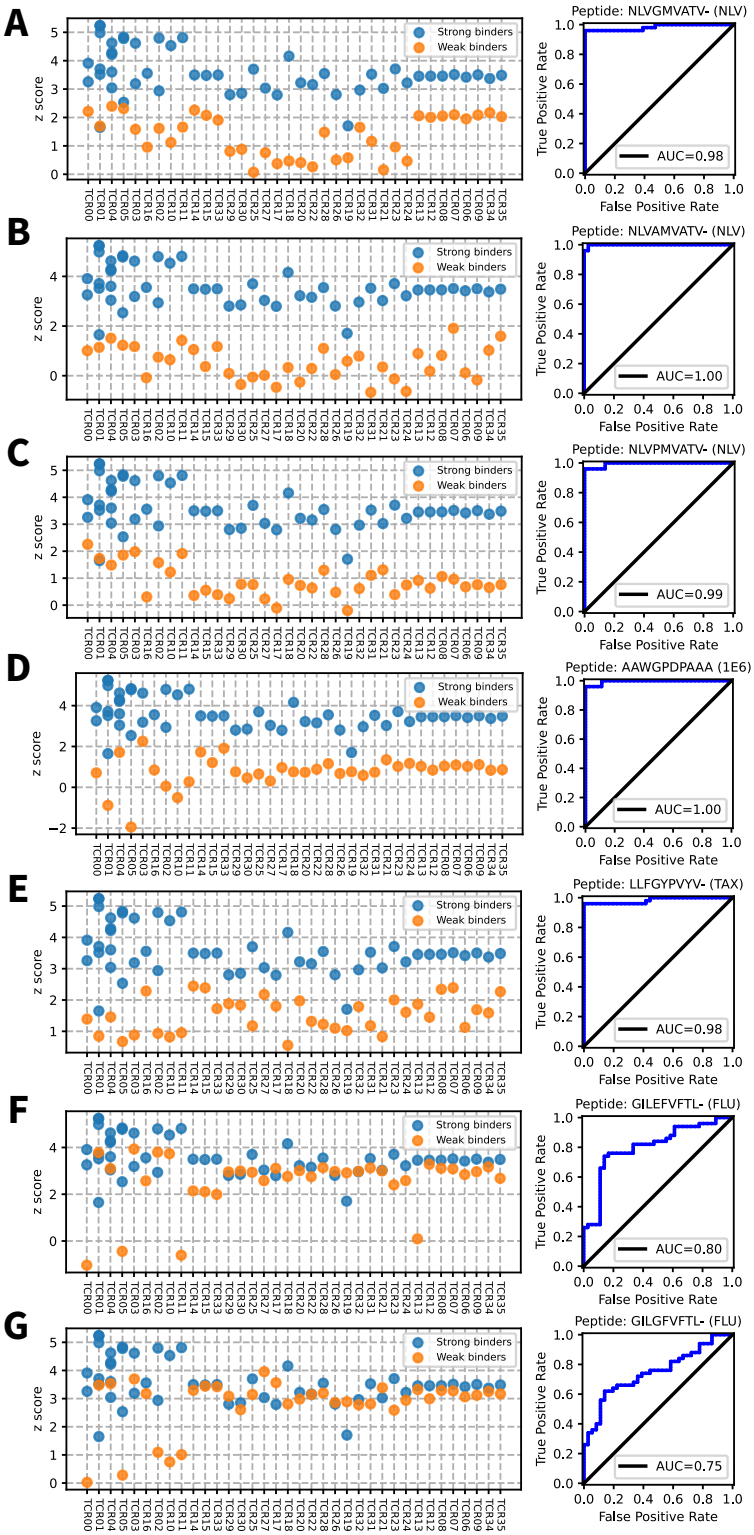

Figure S4: **Comparison of predicted z-scores between strong binders of MART-1 (blue) and weak binders (orange) generated by mismatching MART-1 TCRs with peptides from 1E6, TAX, NLV, and FLU.**

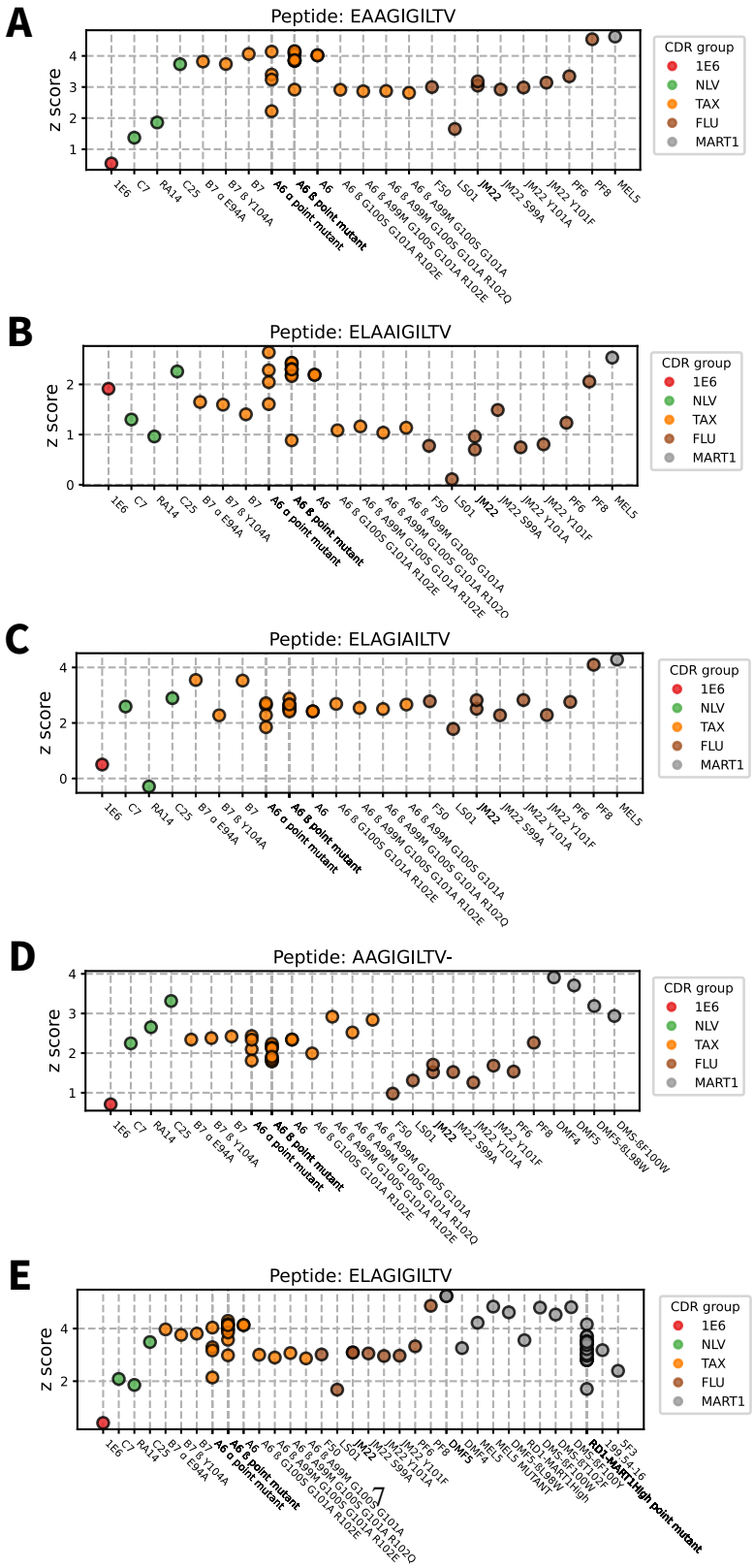

Figure S5: **Comparison of predicted z-scores between strong binders of MART-1 (grey) and weak binders (red, green, orange, and brown) generated by mismatching MART-1 peptides with TCRs specific to 1E6, TAX, NLV, and FLU.**

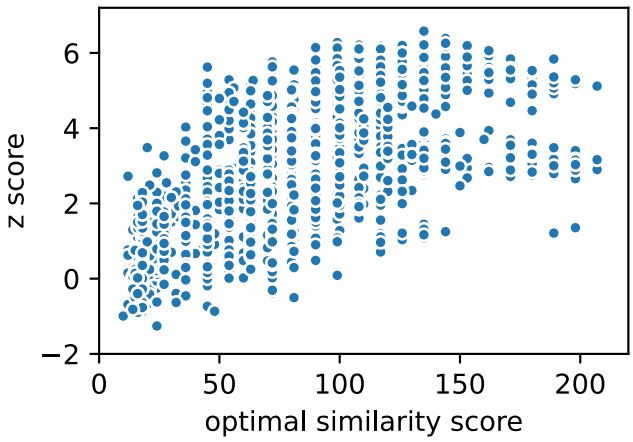

Figure S6: **Z score vs. optimal sequence similarity score for 10x genomics dataset. [12, 42]**

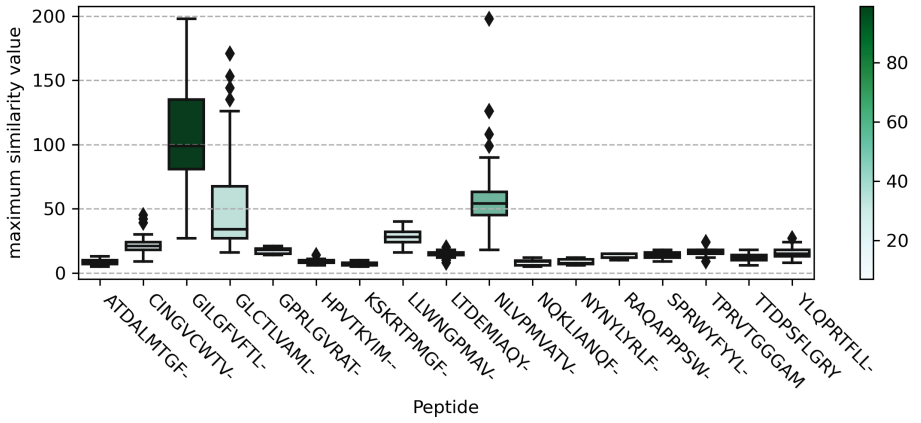

Figure S7: **Distribution of optimal sequence similarity score for strong binders in VDJdb.**

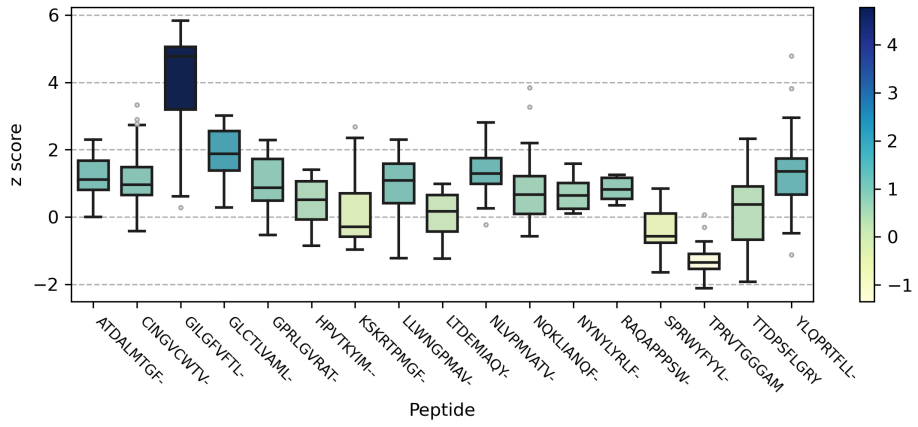

Figure S8: **Distribution of RACER-m predicted z-scores for strong binders in VDJdb.**

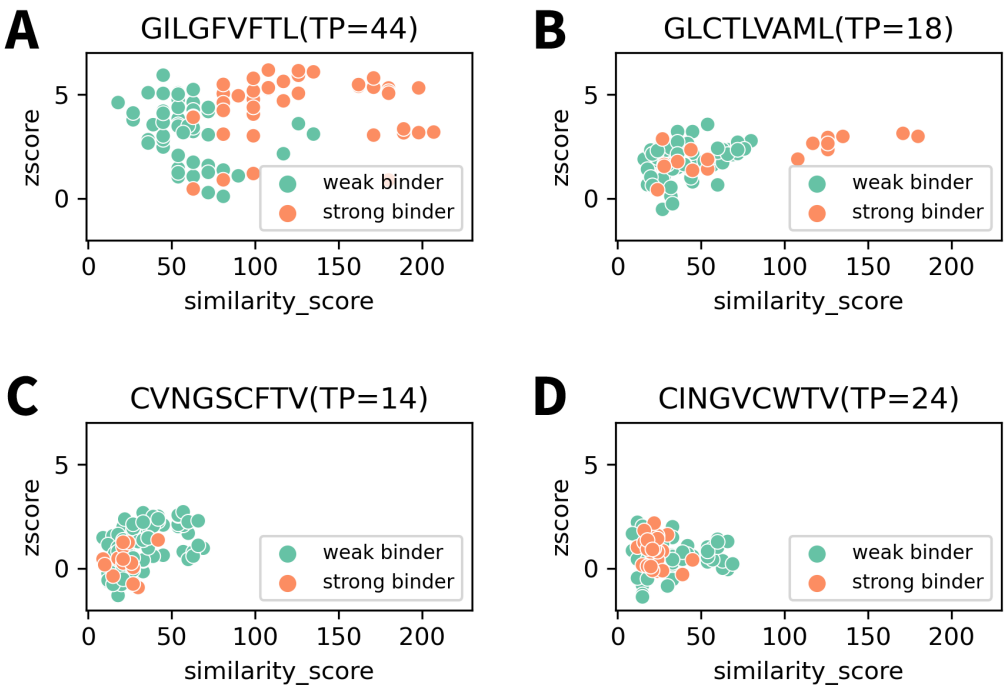

Figure S9: **Z score vs. optimal sequence similarity score for dataset from Grant *et al.* [34].**

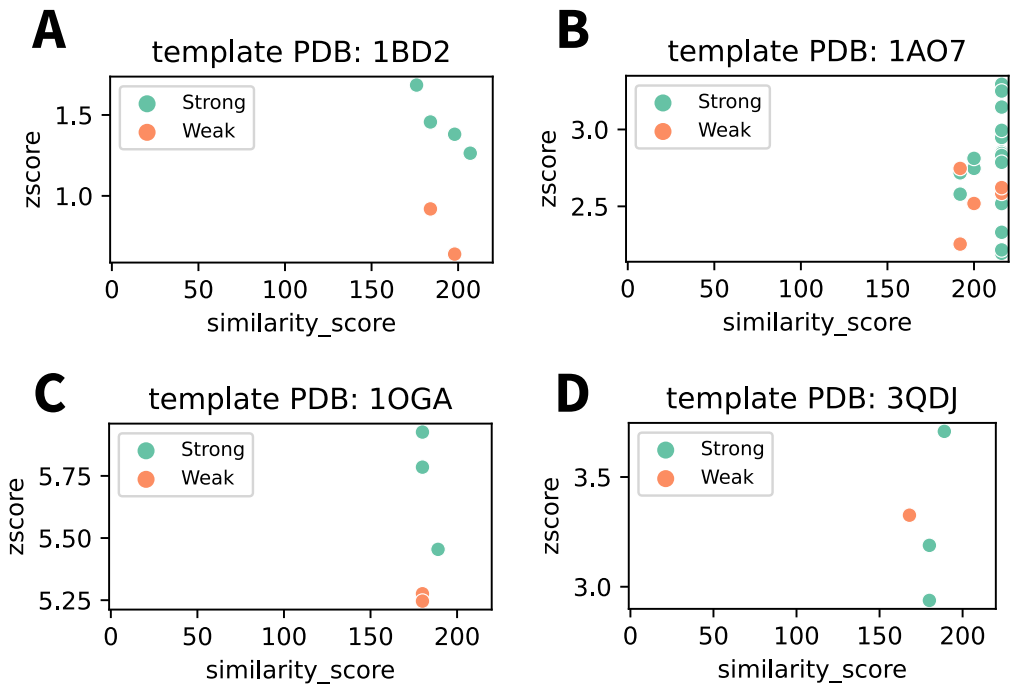

Figure S10: **Z score vs. optimal sequence similarity score for point-mutant weak binders in comparison with strong binders from ATLAS dataset [19].**

REFERENCES AND NOTES

1. L. Klein, B. Kyewski, P. M. Allen, K. A. Hogquist, Positive and negative selection of the T cell repertoire: What thymocytes see (and don't see). *Nat. Rev. Immunol.* **14**, 377–391 (2014).

2. A. Grakoui, S. K. Bromley, C. Sumen, M. M. Davis, A. S. Shaw, P. M. Allen, M. L. Dustin, The immunological synapse: A molecular machine controlling T cell activation. *Science* **285**, 221–227 (1999).

3. S. Ilyas, J. C. Yang, Landscape of tumor antigens in T cell immunotherapy. *J. Immunol.* **195**, 5117–5122 (2015).

4. A. Košmrlj, A. K. Jha, E. S. Huseby, M. Kardar, A. K. Chakraborty, How the thymus designs antigen-specific and self-tolerant T cell receptor sequences. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 16671–16676 (2008).

5. A. K. Chakraborty, A. Košmrlj, Statistical mechanical concepts in immunology. *Annu. Rev. Phys. Chem.* **61**, 283–303 (2010).

6. J. T. George, D. A. Kessler, H. Levine, Effects of thymic selection on T cell recognition of foreign and tumor antigenic peptides. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E7875–E7881 (2017).

7. K. N. Chau, J. T. George, J. N. Onuchic, X. Lin, H. Levine, Contact map dependence of a T-cell receptor binding repertoire. *Phys. Rev. E* **106**, 014406 (2022).

8. M. E. Birnbaum, J. L. Mendoza, D. K. Sethi, S. Dong, J. Glanville, J. Dobbins, E. Özkan, M. M. Davis, K. W. Wucherpfennig, K. Christopher Garcia, Deconstructing the peptide-MHC specificity of T cell recognition. *Cell* **157**, 1073–1087 (2014).

9. P. Dash, A. J. Fiore-Gartland, T. Hertz, G. C. Wang, S. Sharma, A. Souquette, J. C. Crawford, E Bridie Clemens, T. H. Nguyen, K. Kedzierska, Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature* **547**, 89–93 (2017).

10. S. C. Boutet, D. Walter, M. J. Stubbington, K. A. Pfeiffer, J. Y. Lee, S. E. B. Taylor, L. Montesclaros, J. K. Lau, D. P. Riordan, A. M. Barrio, L. Brix, K. Jacobsen, B. Yeung, X.

Zhao, T. S. Mikkelsen, Scalable and comprehensive characterization of antigen-specific CD8 T cells using multi-omics single cell analysis. *J. Immunol.* **202**, 131.4 (2019).

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

12. W. Zhang, P. G. Hawkins, J. He, N. T. Gupta, J. Liu, G. Choonoo, S. W. Jeong, C. R. Chen, A. Dhanik, M. Dillon, A framework for highly multiplexed dextramer mapping and prediction of T cell receptor sequences to antigen specificity. *Sci. Adv.* **7**, eabf5835 (2021).

13. J.-W. Sidhom, H. B. Larman, D. M. Pardoll, A. S. Baras, DeepTCR is a deep learning framework for revealing sequence concepts within T-cell repertoires. *Nat. Commun.* **12**, 1605 (2021).

14. X. Lin, J. T. George, N. P. Schafer, K. N. Chau, M. E. Birnbaum, C. Clementi, J. N. Onuchic, H. Levine. Rapid assessment of T-cell receptor specificity of the immune repertoire. *Nat. Comput. Sci.* **1**, 362–373 (2021).

15. I. Springer, N. Tickotsky, Y. Louzoun, Contribution of T cell receptor alpha and beta CDR3, MHC typing, V and J genes to peptide binding prediction. *Front. Immunol.* **12**, 664514 (2021).

16. B. G. Pierce, L. M. Hellman, M. Hossain, N. K. Singh, C. W. Vander Kooi, Z. Weng, B. M. Baker, Computational design of the affinity and specificity of a therapeutic T cell receptor. *PLoS Comput. Biol.* **10**, e1003478 (2014).

17. Chen Yanover, Philip Bradley, Large-scale characterization of peptide-MHC binding landscapes with structural simulations. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 6981–6986 (2011).

18. P. Bradley, Structure-based prediction of T cell receptor: Peptide-MHC interactions. *eLife* **12**, e82813 (2023).

19. T. Borrman, J. Cimons, M. Cosiano, M. Purcaro, B. G. Pierce, B. M. Baker, Z. Weng, ATLAS: A database linking binding affinities with structures for wild-type and mutant TCR-PMHC complexes. *Proteins* **85**, 908–916 (2017).

20. C. Szeto, C. A. Lobos, A. T. Nguyen, S. Gras, TCR recognition of peptide–MHC-I: Rule makers and breakers. *Int. J. Mol. Sci.* **22**, 68 (2020).

21. A. M. Bulek, D. K. Cole, A. Skowera, G. Dolton, S. Gras, F. Madura, A. Fuller, J. J. Miles, E. Gostick, D. A. Price, J. W. Drijfhout, R. R. Knight, G. C. Huang, Nikolai Lissin, P. E. Molloy, L. Wooldridge, B. K. Jakobsen, J. Rossjohn, M. Peakman, P. J. Rizkallah, A. K. Sewell, Structural basis for the killing of human beta cells by $CD8^+$ T cells in type 1 diabetes. *Nat. Immunol.* **13**, 283–289 (2012).

22. R. Gowthaman, B. G. Pierce, TCR3d: The T cell receptor structural repertoire database. *Bioinformatics* **35**, 5323–5325 (2019).

23. H. M. Berman, The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).

24. R, Vita, S. Mahajan, J.A. Overton, S. K. Dhanda, S. Martini, J. R. Cantrell, D. K. Wheeler, A. Sette, B. Peters, The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res.* **47**, D339–D343 (2018).

25. M. Chen, X. Lin, W. Lu, J. N. Onuchic, P. G. Wolynes, Protein folding and structure prediction from the ground up II: AAWSEM for α/β Proteins. *J. Phys. Chem. B* **121**, 3473–3482 (2017).

26. S. S. Cho, Y. Levy, P. G. Wolynes, *P* versus *Q*: Structural reaction coordinates capture protein folding on smooth landscapes. *Proc. Natl. Acad. Sci.* **103**, 586–591 (2006).

27. R. Gowthaman, B. G. Pierce, Modeling and viewing T cell receptors using TCRmodel and TCR3d. *Methods Mol. Biol.* **2120**, 197–212 (2020).

28. M. G. Rudolph, R. L. Stanfield, I. A. Wilson, How TCRs bind MHCs, peptides, and coreceptors. *Annu. Rev. Immunol.* **24**, 419–466 (2006).

29. B. G. Pierce, Z. Weng, A flexible docking approach for prediction of T cell receptor–peptide–MHC complexes. *Protein Sci.* **22**, 35–46 (2013).

30. S. Gras, X. Saulquin, J.-B. Reiser, E. Debeaupuis, K. Echasserieau, A. Kissenpfennig, F.Legoux, A. Chouquet, M. L. Gorrec, P. Machillot, B. Neveu, N. Thielens, B. Malissen, M. Bonneville, D. Housset, Structural bases for the affinity-driven selection of a public TCR against a dominant human cytomegalovirus epitope. *J. Immunol.* **183**, 430–437 (2009).

31. S. N. Smith, Y. Wang, J. L. Baylon, N. K. Singh, B. M. Baker, E. Tajkhorshid, D. M. Kranz, Changing the peptide specificity of a human T-cell receptor by directed evolution. *Nat. Commun.* **5**, 5223 (2014).

32. 10x Genomics, "A new way of exploring immunity–linking highly multiplexed antigen recognition to immune repertoire and phenotype" (Tech. Rep., 10x Genomics, 2019).

33. P. Meysman, J. Barton, B. Bravi, L. Cohen-Lavi, V. Karnaukhov, E. Lilleskov, A. Montemurro, M. Nielsen, T. Mora, P. Pereira, A. Postovskaya, M. R. Martínez, J. Fernandez-de-Cossio-Diaz, A. Vujkovic, A. M. Walczak, A. Weber, R. Yin, A. Eugster, V. Sharma, Benchmarking solutions to the t-cell receptor epitope prediction problem: Immrep22 workshop report. *ImmunoInformatics* **9**, 100024 (2023).

34. E. J. Grant, T. M. Josephs, S. A. Valkenburg, L. Wooldridge, M. Hellard, J. Rossjohn, M. Bharadwaj, K. Kedzierska, S. Gras, Lack of heterologous cross-reactivity toward HLA-A*02:01 restricted viral epitopes is underpinned by distinct αβT cell receptor signatures. *J. Biol. Chem.* **291**, 24335–24351 (2016).

35. Z. S. Ghoreyshi, J. T. George, Quantitative approaches for decoding the specificity of the human T cell repertoire. *Front. Immunol.* **14**, (2023).

36. B. Meynard-Piganeau, C. Feinauer, M. Weigt, A. M. Walczak, T. Mora, TULIP–A transformer based unsupervised language model for interacting peptides and T-cell receptors that generalizes to unseen epitopes. bioRxiv 549669 [Preprint]. 2023. https://doi.org/10.1101/2023.07.19.549669.

37. B. P. Kwee, M. Messemaker, E. Marcus, G. Oliveira, W. Scheper, C. Wu, J. Teuwen, T. Schumacher, STAPLER: Efficient learning of TCR-peptide specificity prediction from full-length TCR-peptide data. bioRxiv 538237 [Preprint]. 2023. https://doi.org/10.1101/2023.04.25.538237.

38. N. L La Gruta, S. Gras, S. R. Daley, P. G. Thomas, Jamie Rossjohn, Understanding the drivers of MHC restriction of T cell receptors. *Nat. Rev. Immunol.* **18**, 467–478 (2018).

39. A. Davtyan, N. P. Schafer, W. Zheng, C. Clementi, Peter G Wolynes, G. A. Papoian, AWSEM-MD: Protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing. *J. Phys. Chem. B* **116**, 8494–8503 (2012).

40. W. Zheng, N. P. Schafer, A. Davtyan, G. A. Papoian, P. G. Wolynes, Predictive energy landscapes for protein–protein association. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 19244–19249 (2012).

41. B. Webb, A. Sali, Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinformatics* **54**, 5–6 (2016).

42. 10x Genomics. Tech. rep 2019.