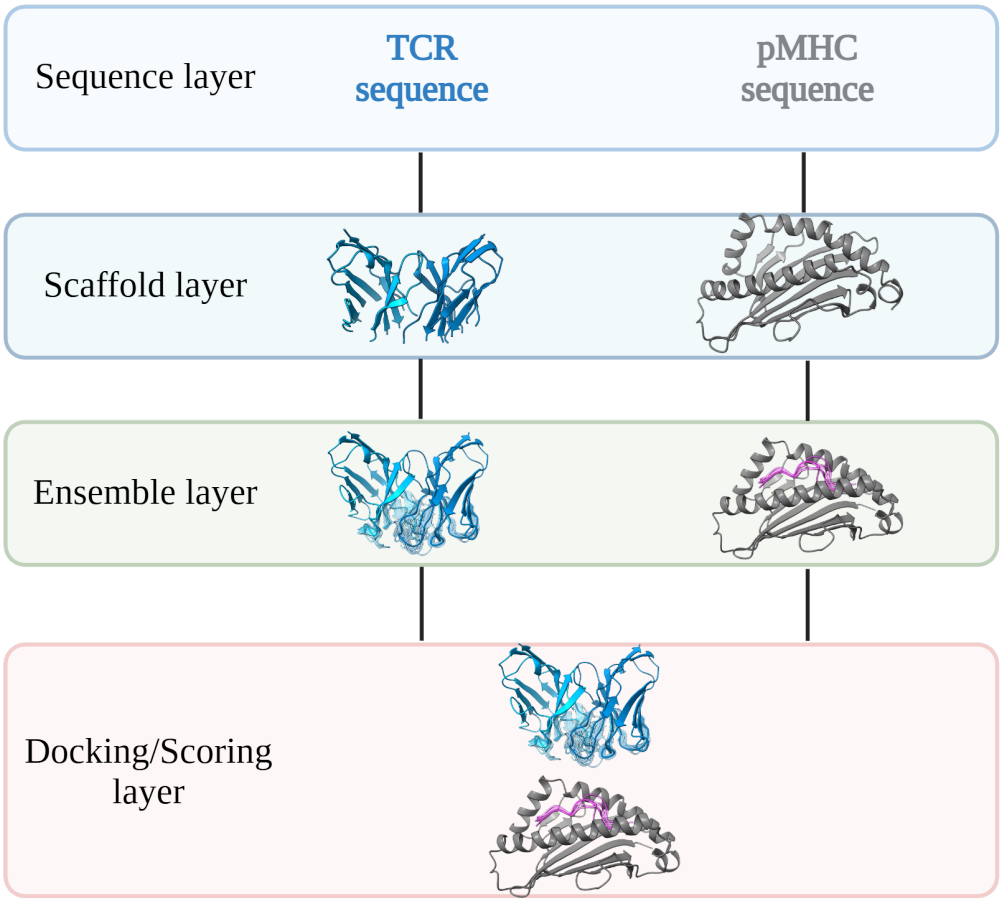
Graphical Abstract

STEGG: Structural TCR-pMHC Ensemble Generator and Gallery

Jared K. Slone, Alexander W. Bock, Mauricio Menegatti Rigo, Chengxuan Zou, Alexandre Reuben, Lydia E. Kavraki



Highlights

STEGG: Structural TCR-pMHC Ensemble Generator and Gallery

Jared K. Slone, Alexander W. Bock, Mauricio Menegatti Rigo, Chengxuan Zou, Alexandre Reuben, Lydia E. Kavraki

- Recent advances in protein structure prediction have improved modeling accuracy, but efforts to capture flexibility and diverse conformations remain lacking.
- Accurately modeling interactions between T cell receptors (TCRs) and peptide-MHCs (pMHCs) has potential implications across immunology, including in the design of immunotherapy treatments for cancer.
- Our software, STEGG, generates hundreds of low-energy, structurally diverse 3D conformations of TCR-pMHC class-I complexes from sequence input.
- Physics-based analysis of the resulting distribution of conformations can be used to accurately predict TCR-pMHC interactions.
- A public database of TCR-pMHC structural ensembles is provided to support future research.

STEGG: Structural TCR-pMHC Ensemble Generator and Gallery

Jared K. Slone[a], Alexander W. Bock[a], Mauricio Menegatti Rigo[b], Chengxuan Zou[a], Alexandre Reuben[c], Lydia E. Kavraki[a,d,1]

[a] *Computer Science, Rice University, Houston, 77005, TX, USA*
[b] *Department of Molecular Biology and Biotechnology, Federal University of Rio Grande do Sul, Porto Alegre, 90010, Rio Grande do Sul, Brazil*
[c] *Thoracic/Head & Neck Medical Oncology, MD Anderson Cancer Center, Houston, 77030, TX, USA*
[d] *The Ken Kennedy Institute, Rice University, Houston, 77005, TX, USA*

Abstract

Understanding how T cell receptors (TCRs) recognize peptides presented by class-I major histocompatibility complexes (pMHCs) is a central challenge in immunology, with implications for infectious disease research, autoimmunity, and the development of personalized cancer immunotherapies. While recent advances in 3D protein structure prediction have enabled more accurate modeling of protein complexes, current methods fail to capture the flexible nature of TCR-pMHC interactions. Due to the high flexibility in the CDR loops of the TCR and the potential flexibility of the peptide, these interactions are dynamic and can adopt a range of energetically favorable binding modes that are not well represented by a single static structure.

We introduce **STEGG: Structural TCR-pMHC Ensemble Generator and Gallery**, a computational framework for generating structural ensembles of TCR-pMHC complexes from amino acid sequences. STEGG generates a range of distinct low-energy 3D conformations for each TCR-pMHC pair, capturing the diversity and flexibility of TCRs, pMHCs, and their joint interactions. Unlike conventional molecular dynamics simulations, which are computationally expensive and scale poorly, STEGG utilizes a domain specific sampling algorithm, enabling efficient 3D modeling.

We show that STEGG not only recovers conformations with low RMSD to experimentally solved structures but also finds a diverse range of biologically relevant binding modes and flexible poses. To support further research, we provide a publicly available database of structural ensembles for TCR-pMHC

pairs.

By going beyond static structure generation, STEGG presents a new method for studying the molecular drivers of TCR-pMHC interactions. The STEGG web server is freely available at `https://stegg.kavrakilab.rice.edu/`.

Keywords: TCR-pMHC, 3D protein modeling, conformational sampling, molecular interactions

1. Introduction

The interaction of T-cell receptors (TCRs) with a peptide bound to Major Histocompatibility Complex (peptide-MHC or pMHC) is a crucial part of the adaptive immune response [1]. TCRs are proteins located on the surface of T cells that recognize and bind to specific pMHCs, which display peptides derived from intracellular and extracellular proteins. If a given TCR is reactive against the displayed peptide, this binding between TCRs and pMHCs will trigger cytokine release, invoking an adaptive immune response against the cell presenting the peptide [1]. The ability to predict and manipulate these interactions has driven advances in the development of novel vaccines and targeted immunotherapies for cancer [2]. Structural insight into TCR–pMHC recognition could further guide therapeutic design, yet predicting their three-dimensional structures from sequence remains a challenge [2, 3, 4].

On the TCR side, several computational tools have been specifically developed for TCR modeling [5, 6, 7, 8, 9, 10, 11], and others have relied on general-purpose protein modeling software [12, 13, 14, 15, 16]. Although large-scale foundation models can generate highly accurate TCR structures, specialized methods tailored to TCRs achieve faster runtimes, more efficient sampling of relevant conformations, and, in many cases, greater accuracy [17]. While much of the sequence and structure of TCRs is highly conserved, the hyper-variable regions, especially CDR3 loops, remain difficult to model due to their flexibility and distinct conformational states (i.e., TCR-free vs. TCR-bound) [6, 18, 17, 19]. This conformational adaptability complicates structure prediction and continues to challenge even the most advanced deep learning models [20, 21].

On the pMHC side, accurate modeling of the three-dimensional pMHC complex is essential to understand antigen presentation and T cell recogni-

tion. The orientation and geometry of a peptide within the MHC binding groove can influence biological phenomena such as molecular mimicry and T cell cross-reactivity, which are highly relevant for the design of safe and effective immunotherapies [2, 22]. Several computational approaches have been developed to model pMHC complexes [23, 24, 25, 26, 27, 28]. These methods differ in their reliance on templates, sampling strategies, and treatment of peptide flexibility. Both the MHC and, in particular, the peptide exhibit substantial conformational diversity. To address this challenge, some modeling pipelines emphasize generating ensembles of plausible peptide conformations rather than a single static structure [23, 24, 26]. Such ensemble-based strategies better capture the range of geometries a peptide can adopt, increasing the chances of identifying functionally relevant conformations.

Given the difficulties associated with modeling the TCR and pMHC individually, it is not surprising that it remains difficult to predict the 3D structure of TCR–pMHC complexes. The interface depends not only on the conformation of each molecule, but also on its binding geometry [29]. Experiments have also shown that both peptides and CDR loops can undergo conformational changes upon binding, further complicating modeling efforts [19]. Several computational strategies have been developed to address this problem. Some methods begin with separately modeled or experimentally determined TCR and pMHC structures and then attempt to predict their binding orientation [30, 31]. Other approaches, including recent deep learning-based methods, attempt to model the TCR and the pMHC within the same pipeline [7, 11, 10]. Despite these innovations, major limitations persist. Current tools struggle to correctly identify the docking orientation of TCRs on pMHCs and to correctly model the geometry of the peptide and CDR loops within the bound structure. Consequently, accurate and efficient modeling of the geometry/orientation of TCR–pMHC complexes remains an open problem in computational immunology.

In this work, we use an ensemble-based approach for pMHC modeling [26] and we employ a hybrid strategy for TCR modeling. A deep learning–based method is used to model the conserved regions of the TCR. For the highly flexible CDR loops, we integrated an optimization-based modeling approach designed to better sample and refine conformations. This combination allows us to generate diverse 3D models of TCRs at relatively low computational cost. Additionally, we created a fine-tuned scoring function from Rosetta [32] that ranks TCR-pMHC conformations better than the existing scoring functions. Our method, STEGG, is capable of producing 3D conformations

that are close to experimental structures in the PDB. STEGG also produces a measurably diverse ensemble of conformations and binding modes, which can elucidate patterns in TCR-pMHC interactions.

2. Methods

Given input sequences, STEGG generates diverse ensembles of TCR-pMHC conformations by first modeling the TCR and pMHC separately, producing ensembles for each, then docking and scoring the ensemble (Figure 1A).

2.1. TCR Modeling

Our primary focus when modeling the TCR is capturing conformational diversity and relevant flexibility. Since the CDR loops are directly contacting the pMHC, accurately modeling these flexible loops is of the utmost importance. As such, our algorithm, T-RECS, focuses mostly on modeling possible conformations for these loops. T-RECS starts by modeling an initial TCR conformation with the tool tfold-TCR [11]. This tool has similar accuracy to leading modeling frameworks while being orders of magnitude faster than AlphaFold3 or TCRmodel2 [11, 17]. Next, the conserved region of the TCR is kept constant and the loop closure/modeling algorithm Randomized Coordinate Descent (RCD) [33**?**] is used to sample thousands of backbone conformations for the CDR loops. These conformations are scored with a knowledge-based scoring function [**?**] and the best conformations are retained. The algorithm FASPR [34] is then used to add side chains back to the CDR loops.

2.2. pMHC Modeling

We model the peptide–MHC (pMHC) complex using APE-Gen 2.0 [26], which generates ensembles of 3D conformations to capture peptide flexibility within the binding groove. APE-Gen 2.0 utilizes both homology modeling and conformational sampling. In the homology modeling stage, a library of experimentally resolved pMHC structures is used as structural templates. Conformational sampling for the peptide is done by treating the portion of the peptide between the anchor positions as a loop and performing several iterations of loop modeling without the MHC present. Peptide conformations are subsequently docked back into the MHC binding groove with SMINA [35]. We modified APE-Gen 2.0 by replacing MODELLER [12] with the

deep learning–based pMHC modeling framework from tfoldTCR [11]. This modification is applied only when no suitable MHC reference structure is available in the PDB, allowing users to run APE-Gen 2.0 without a MODELLER key. For all experiments, APE-Gen 2.0 was executed with an increased number of peptide conformations selected for energy minimization (200 instead of 100), while all other parameters were kept at their default values. Additional details regarding hyperparameter settings are provided in Appendix E. Users of the local Dockerized version of STEGG may adjust hyperparameters as needed for specialized applications. APE-Gen 2.0, and by extension STEGG, currently support modeling of pMHC class-I complexes only. Modeling of class-II complexes is beyond the scope of this work.

2.3. TCR-pMHC Docking

We leverage HADDOCK3 [36] to dock the ensembles of TCR and pMHC conformations. HADDOCK was found to be the most accurate docking protocol for TCR-pMHCs under realistic conditions, out of several methods considered [31]. We set the CDR3 loops and the epitope as active residues during docking. Our docking protocol consists of three sequential stages: rigid-body docking, flexible refinement, and energy minimization. Clustering and filtering are applied between each stage. The number of models retained at each step was chosen empirically to balance computational efficiency with sufficient exploration of conformational space (see Appendix E). While we are able to capture conformations that are similar to existing crystal structures with fewer samples than used here, the extra sampling paired with our clustering and filtering works towards our goal of capturing a diverse set of binding conformations for a given TCR-pMHC, not just replicating a crystallographic conformation.

2.4. Scoring for Conformation Selection

Following the docking stage, TCR-pMHC conformations are filtered and ranked according to geometric and energetic criteria. While metrics like RMSD and DockQ [37] are necessary for establishing the quality of structural modeling methods, they require access to known reference structures. To score conformations in ensembles generated by STEGG, we propose a custom scoring function that uses a set of reference-free energetic and geometric properties. We define as features: the 19 terms comprising the standard Rosetta energy function [38], pairwise distances between TCR-pMHC components, and pairwise angles between a set of vectors computed for the TCR

and MHC. Using a partial least-squares regression model, we fit these features to DockQ scores to produce a reference-free structural quality ranking for conformations within an ensemble (See Appendix A for details).

3. Results & Discussion

3.1. Metrics and Evaluation of 3D Protein Conformations

Accurately assessing the quality of 3D protein models is important to inform their use in downstream applications. The most common metric for assessing computationally generated protein models is root-mean-squared-distance (RMSD) [39]. While RMSD can provide insight into the quality of a generated structure, it can be sensitive to outliers, including those far away from the region of interest in a proposed model [39]. In the case of TCR-pMHC complexes, small changes in the docking angle can result in large changes in RMSD when the atoms in the TCR are far from the binding site are considered. Local Distance Difference Test (lDDT) [40] evaluates protein structure predictions without requiring rigid-body superposition. By calculating the conservation of local interatomic distances, lDDT scores models robustly even in the presence of large domain shifts caused by slight differences in TCR-pMHC docking angle. DockQ [37] [41] is a structural similarity measure specialized for evaluating protein-ligand and protein-protein docking models. Given a model and a reference structure, DockQ computes RMSD values between the model and reference for 1) the ligand component, which is either explicitly indicated or inferred to be the smaller of the two components, and 2) the interface region, defined as all residues in either component within a fixed distance of any residue in the other component. Importantly, DockQ also computes the proportion of inter-component residue contacts identified in the reference structure that are reproduced in the model. By employing DockQ as a quality measure, we are able to localize RMSD calculations to the site of TCR-pMHC interaction and excludes isometric differences in non-interacting TCR regions resulting from variation in docking angles, while relaxing penalties for minor geometric variation at the docking site in favor of counting discrete contacts that comprise the interaction.

3.2. Benchmark on Experimental Structures from STCRDab

While STEGG's objective is to produce diverse ensembles of conformations, being able to accurately replicate experimentally determined TCR-pMHC crystal structures is still important. We benchmarked structures pro-

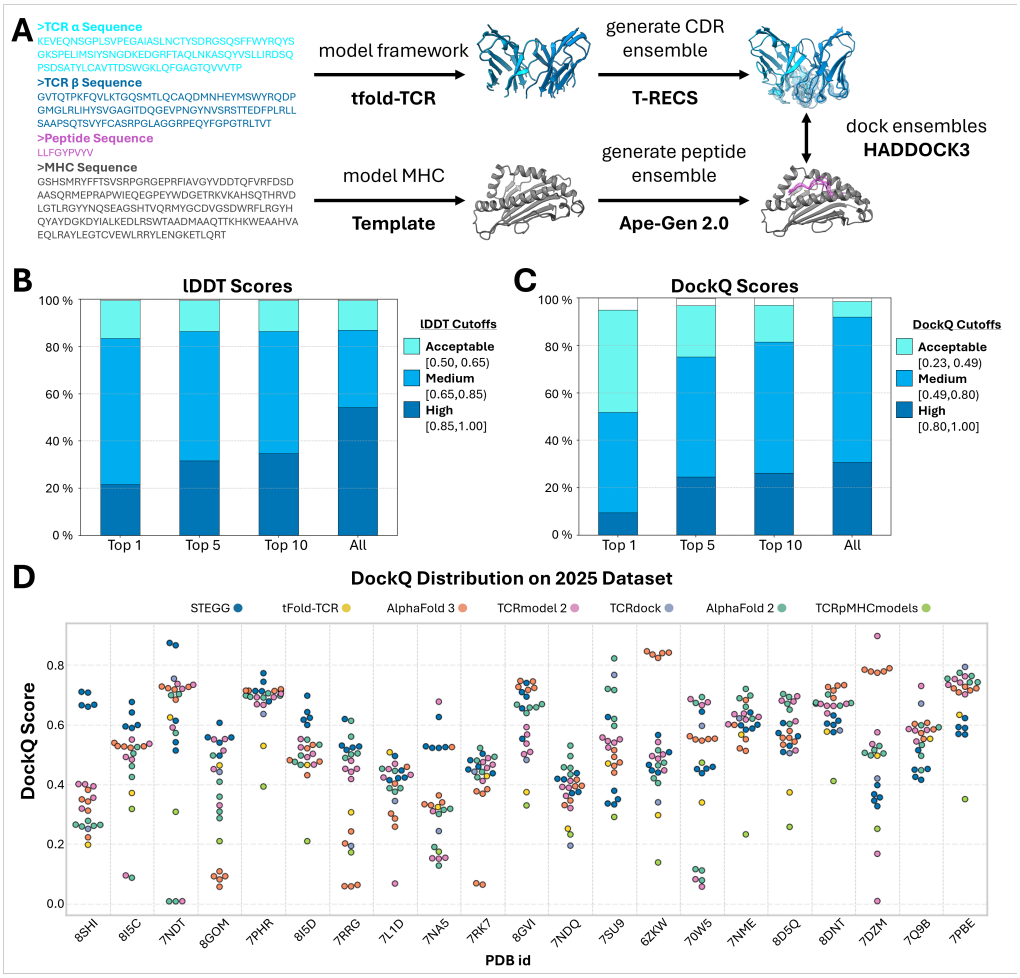


Figure 1: **A**: Visualization of our modeling pipeline. STEGG takes amino acid sequences as input then utilizes tfold-TCR, Ape-Gen 2.0, T-RECS, and HADDOCK 3 to generate 3D conformational ensembles. **B**: lDDT Scores for complexes in STCRDab. **C**: DockQ Scores for complexes in STCRDab. **D**: DockQ distributions for STEGG (blue dots) and other models on a set of 21 structures recently added to the PDB. For readability, only the top 5 conformations produced by STEGG are included here.

duced by STEGG against 199 experimentally solved structures collected from STCRDab [42] (collection date July 2025). The quality of STEGG's models relative to these reference structures is detailed in Figure 1 parts B and C.

To assess STEGG's accuracy in reproducing experimentally derived conformations, we report the success rates for lDDT and DockQ scores: the

percentage of STCRDab complexes for which STEGG produced at least one acceptable, medium, or high-quality model within the top 1, 5, 10, and *ALL* ranked conformations. We adopted the standard CAPRI quality categories (i.e., Acceptable, Medium, High). As shown in Figure 1 parts B and C, STEGG produced quality structures for nearly every complex in the benchmark set. To avoid data-leakage and ensure fair rankings, a separate instance of our custom scoring function weights was set with the exclusion of each PDB and used to rank conformations for that PDB in a leave-one-out fashion.

STEGG successfully generated acceptable or better models (by both lDDT and DockQ Score) for all but three complexes in the benchmark set: 5SWS, 5SWZ, and 7JWI. These three complexes are notable for exhibiting a reversed binding polarity relative to the canonical orientation observed in nearly all other solved TCR-pMHC structures. Although STEGG failed to capture this specific reversed polarity, this is a minor limitation in the context of typical immune response modeling, as previous work has established that "a canonical docking polarity is required for T cell activation" [43]. Interestingly, STEGG's docking protocol places no explicit geometric restrictions that would prohibit reversed polarity, meaning that the resulting non-reversed conformations may reflect a local energy minimum that our docking protocol found more favorable.

3.3. Comparative Benchmark on Recent Experimental Structures

Comparison with existing structure prediction tools has become increasingly difficult. Deep learning methods such as AlphaFold [16] have been trained on nearly all available protein structure data, leaving few unseen cases for fair benchmarking. In spite of these limitations, previous work has been able to identify shortcomings of deep learning based structure prediction tools in the context of TCR-pMHCs [18, 17]. We present a benchmarking of STEGG against other structure prediction models in the literature using a previously curated set of 21 experimentally solved TCR-pMHC structures, which were all deposited since the training cutoff for the most recent tool considered in our benchmark [17] No information from this set of structures (Figure 1D) was used in defining the scoring function.

In replicating the geometry at the TCR-pMHC binding interface, STEGG performs competitively. Figure 1D shows that STEGG produces TCR-pMHC conformations with better DockQ Scores than all other methods for 7 of the 21 PDBs in the 2025 benchmark dataset. Compared to other methods, STEGG performed well at modeling some complexes with extreme docking

angles. The distribution of docking angles for TCR-pMHC complexes in the PDB has a median of 43.60° and a standard deviation of 13.49° [44]. STEGG produced by far the best 3D structures, in terms of DockQ Score, for 8SHI, which has a docking angle of 21.4°, placing it in the 4^{th} percentile for docking angles among canonical solved complexes. STEGG also produced the best structure for 7RRG, which has a docking angle of 73.8°, placing it in the 98^{th} percentile for docking angles among canonical solved complexes. This performance demonstrates that STEGG's unconstrained modeling and docking approach is effective in cases of unusual binding geometries.

In replicating the exact 3D geometry observed in the crystal structures, STEGG is comparable to existing methods. We calculated global backbone RMSD for the whole TCR-pMHC complex as well as all-atom RMSDs of the most flexible parts of the TCR-pMHC complex: the six CDR loops and the peptide. The RMSD distributions for STEGG's models are comparable to those of other methods (See Appendix C for details). Importantly, STEGG produced fewer outlier structures compared to existing tools, especially for TCRs with long CDR loops. For example, for the long CDR3β loop of 7Q9B (16 amino acids), the all-atom RMSD for the STEGG model was 3.95, significantly better than the 6.02 RMSD achieved by the AlphaFold 3 model. This improved robustness is manifest through the upper quartile range for STEGG's CDR3β RMSD distribution, which is lower than that of other tools.

3.4. Use Case: Scoring for TCR-pMHC Binding Prediction

Table 1: Leave-one-out ROC–AUC performance of aggregation strategies across demonstration datasets

Dataset	**Ensemble (mean energy terms)**	**Static (Random) (median ROC–AUC)**	**Static (Best) (by Rosetta energy)**
ELAGIGILTV	0.887	0.800	0.865
FLCKMALLL	0.681	0.637	0.667
GILGFVFTL	0.732	0.675	0.666
LLWNGPMAV	0.749	0.759	0.702
TCR-1	0.775	0.687	0.731
A11Vc	0.711	0.626	0.496

To showcase the usefulness of STEGG, we conducted some experiments to predict TCR–pMHC binding specificity using conformational ensembles.

We assembled peptide-specific and TCR-specific datasets containing both binding and non-binding TCR-pMHC pairs, shown in Table 1. For ELAGIGILTV and GILGFVFTL, TCR-pMHC pairs were sampled from 10x Genomics datasets [45] using a greedy algorithm to maximize TCR diversity. For FLCMKALLL and LLWNGPMAV, binders were taken from VDJdb [46], and non-binders were created by randomly sampling TCRs binding to other antigens, again using a greedy algorithm to maximize diversity. TCR–pMHC pairs for the TCR-1 and A11Vc datasets were obtained from BATCAVE [47] and restricted to strong binders and non-binders.

For each TCR-pMHC pair, STEGG generated ensembles of 3D structures, from which we calculated Rosetta energy terms. Linear Support Vector Machines (SVMs) were then trained on these energy terms to classify binders and non-binders. We compare the performance of SVMs trained on energy terms averaged across ensembles with SVMs trained on random static conformations and the lowest-energy conformation for each ensemble. Performance was evaluated via leave-one-out cross-validation. Table 1 summarizes the results. We found that the three most influential features (i.e., those with the highest normalized coefficients in our trained SVMs) were *fa-sol* (Lazaridis-Karplus solvation energy), *fa-atr* (Lennard-Jones attraction), and *ref* (residue reference free energy). These lightweight, interpretable classifiers, based on energy terms from ensembles outperformed identical classifiers trained on static structures, highlighting the potential value of capturing diverse TCR-pMHC binding geometries. See Appendix B for additional details regarding classifier training and the composition of the six datasets.

4. Webserver

The STEGG web server is freely available at `https://stegg.kavrakilab.rice.edu/`. To produce a TCR-pMHC ensemble, the user just has to enter the amino sequences of the TCR, peptide, and MHC. The MHC sequences can be auto-filled by allele. Once the ensemble has been modeled, the user receives an email with a link to the results page, allowing them to visualize and download a list of ranked TCR-pMHC conformations. Visualizations are done with the MolStar 3D viewer. Users will also have the option to add their STEGG ensembles to our public database. For highly sensitive jobs or for more customized versions of our tool, please use our Docker Container, available on DockerHub or our GitHub page: `https://github.com/KavrakiLab/STEGGosaurus-docker`.

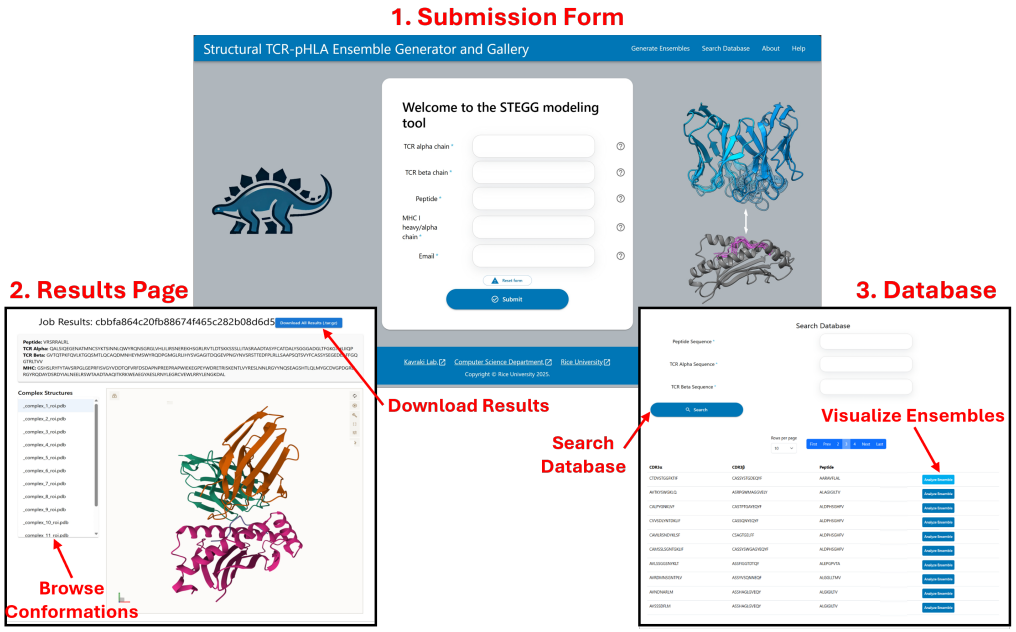


Figure 2: Structural TCR-pHLA Ensemble Generator and Gallery: Submission Form (1), Results Page (2), Database (3).

5. Database

In addition to a web server for generating ensembles of TCR-pMHC conformations, we also provide a database containing TCR-pMHC ensembles. The sequences used for generating this database were taken from the VDJdb [46] and from user input. We will continue to update and expand this database as more data becomes available. The database is accessible through the web server.

6. Funding and Acknowledgments

The authors would like to acknowledge the conversations and brainstorming sessions with members of the Kavraki Lab and the Reuben Lab that helped shape up this work. ChimeraX [48] was used to visualize and analyze protein structures.

Work by J.K.S. on this project was supported by a training fellowship from the Gulf Coast Consortia on the Training Program in Biomedical Informatics, National Library of Medicine 5T15LM007093. Work by A.W.B.

was supported by a fellowship from the NSF CSGrad4US Fellowship Program GL8EKK2U5YE9. Work by M.M.R was supported by a Computational Cancer Biology Training Program fellowship (CPRIT Grant No. RP170593) and FAPERGS (Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul). Work by L.E.K. was supported in part by National Institutes of Health NIH U01CA258512. This work was supported in part by the NOTS cluster operated by Rice University's Center for Research Computing (CRC).

7. Declarations

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] S. A. Rosenberg, N. P. Restifo, J. C. Yang, R. A. Morgan, M. E. Dudley, Adoptive cell transfer: a clinical path to effective cancer immunotherapy, Nature Reviews Cancer 8 (4) (2008) 299–308.

[2] D. Hudson, R. A. Fernandes, M. Basham, G. Ogg, H. Koohy, Can we predict T cell specificity with digital biology and machine learning?, Nature Reviews Immunology 23 (8) (2023) 511–521. doi:10.1038/s41577-023-00835-3.

[3] M. Nielsen, A. Eugster, M. F. Jensen, M. Goel, A. Tiffeau-Mayer, A. Pelissier, et al., Lessons learned from the immrep23 tcr-epitope prediction challenge, ImmunoInformatics 16 (2024) 100045. doi:10.1016/j.immuno.2024.100045.

[4] S. N. Deleuran, M. Nielsen, Nettcr-struc, a structure driven approach for prediction of tcr-pmhc interactions, Frontiers in Immunology (2025). doi:10.3389/fimmu.2025.1616328.

[5] M. S. Klausen, M. V. Anderson, M. C. Jespersen, M. Nielsen, P. Marcatili, Lyra, a webserver for lymphocyte receptor structural modeling, Nucleic Acids Research 43 (W1) (2015) W349–W355.

[6] W. K. Wong, C. Marks, J. Leem, A. P. Lewis, J. Shi, C. M. Deane, Tcrbuilder: multi-state t-cell receptor structure prediction, Bioinformatics 36 (11) (2020) 3580–3581.

[7] K. K. Jensen, V. Rantos, E. C. Jappe, T. H. Olsen, M. C. Jespersen, V. Jurtz, L. E. Jessen, E. Lanzarotti, S. Mahajan, B. Peters, et al., TCRPMHCMODELS: Structural modelling of TCR-PMHC Class I complexes, Scientific Reports 9 (1) (2019). doi:10.1038/s41598-019-50932-4.

[8] S. Li, J. Wilamowski, S. Teraguchi, F. J. van Eerden, J. Rozewicki, A. Davila, Z. Xu, K. Katoh, D. M. Standley, Structural modeling of lymphocyte receptors and their antigens, in: in vitro differentiation of t-cells: methods and protocols, Springer, 2019, pp. 207–229.

[9] B. Abanades, W. K. Wong, F. Boyles, G. Georges, A. Bujotzek, C. M. Deane, Immunebuilder: Deep-learning models for predicting the structures of immune proteins, Communications Biology 6 (1) (2023) 575.

[10] R. Yin, H. V. Ribeiro-Filho, V. Lin, R. Gowthaman, M. Cheung, B.-G. Pierce, TCRMODEL2: High-resolution modeling of T-cell receptor recognition using Deep Learning, Nucleic Acids Research 51 (W1) (2023). doi:10.1093/nar/gkad356.

[11] F. Wu, Y. Zhao, Y. Xiao, C. Qin, F. Wang, Z. Wu, L.-K. Huang, X. Liu, J. Song, B. He, et al., Fast and accurate modeling of tcr-peptide-mhc complexes using tfold-tcr, bioRxiv (2025) 2025–01.

[12] B. Webb, A. Sali, Protein structure modeling with modeller, in: Functional genomics: Methods and protocols, Springer, 2017, pp. 39–54.

[13] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, et al., Highly accurate protein structure prediction with AlphaFold, Nature 596 (7873) (2021) 583–589.

[14] R. Krishna, J. Wang, W. Ahern, P. Sturmfels, P. Venkatesh, I. Kalvet, G. R. Lee, F. S. Morey-Burrows, I. Anishchenko, I. R. Humphreys, et al., Generalized biomolecular modeling and design with rosettafold all-atom, Science 384 (6693) (2024) eadl2528.

[15] C. Discovery, J. Boitreaud, J. Dent, M. McPartlon, J. Meier, V. Reis, A. Rogozhnikov, K. Wu, Chai-1: Decoding the molecular interactions of life, bioRxiv (2024). doi:10.1101/2024.10.10.615955.

[16] J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, et al., Accurate structure prediction of biomolecular interactions with alphafold 3, Nature 630 (8016) (2024) 493–500.

[17] Y. Shi, J. M. Parks, J. C. Smith, Comparative analysis of tcr and tcr-pmhc complex structure prediction tools, Journal of chemical information and modeling 65 (13) (2025) 7156–7173.

[18] H. N. Le, M. V. de Freitas, D. A. Antunes, Strengths and limitations of web servers for the modeling of TCRpMHC complexes, Computational and Structural Biotechnology Journal 23 (2024) 2938–2948. doi:https://doi.org/10.1016/j.csbj.2024.06.028.

[19] B. McMaster, C. J. Thorpe, J. Rossjohn, C. M. Deane, H. Koohy, Quantifying conformational changes in the tcr: pmhc-i binding interface, Frontiers in Immunology 15 (2024) 1491656.

[20] A. O. Stevens, Y. He, Benchmarking the accuracy of alphafold 2 in loop structure prediction, Biomolecules 12 (7) (2022) 985.

[21] T. Saldaño, N. Escobedo, J. Marchetti, D. J. Zea, J. Mac Donagh, A. J. Velez Rueda, E. Gonik, A. García Melani, J. Novomisky Nechcoff, M. N. Salas, et al., Impact of protein conformational diversity on alphafold predictions, Bioinformatics 38 (10) (2022) 2742–2748.

[22] A. F. Fonseca, D. A. Antunes, CrossDome: an interactive R package to predict cross-reactivity risk using immunopeptidomics databases, Frontiers in Immunology 14 (2023) 1142573.

[23] J. R. Abella, D. A. Antunes, C. Clementi, L. E. Kavraki, Ape-gen: A fast method for generating ensembles of bound peptide-mhc conformations, Molecules 24 (5) (2019) 881.

[24] D. F. Marzella, F. M. Parizi, D. v. Tilborg, N. Renaud, D. Sybrandi, R. Buzatu, D. T. Rademaker, P. A. 't Hoen, L. C. Xue, Pandora: a fast, anchor-restrained modelling protocol for peptide: Mhc complexes, Frontiers in Immunology 13 (2022) 878762.

[25] V. Mikhaylov, C. A. Brambley, G. L. Keller, A. G. Arbuiso, L. I. Weiss, B. M. Baker, A. J. Levine, Accurate modeling of peptide-mhc structures with alphafold, Structure 32 (2) (2024) 228–241.

[26] R. Fasoulis, M. M. Rigo, G. Lizée, D. A. Antunes, L. E. Kavraki, Apegen2. 0: Expanding rapid class i peptide–major histocompatibility complex modeling to post-translational modifications and noncanonical peptide geometries, Journal of Chemical Information and Modeling 64 (5) (2024) 1730–1750.

[27] D. Frühbuß, C. Baakman, S. Teusink, E. Bekkers, S. Jegelka, L. C. Xue, Fast and accurate peptide–mhc structure prediction via an equivariant diffusion model, bioRxiv (2025) 2025–04.

[28] C. Baakman, G. Crocioni, C. Geng, D. T. Rademaker, D. Frühbuß, Y. J. Aarts, L. C. Xue, Swiftmhc: A high-speed attention network for mhc-bound peptide identification and 3d modeling, bioRxiv (2025) 2025–01.

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

[30] F. M. Parizi, Y. J. Aarts, S. Eerden, G. Ramakrishnan, L. C. Xue, Swifttcr: Efficient computational docking protocol of tcrpmhc-i complexes using restricted rotation matrices, bioRxiv (2024) 2024–05.

[31] T. Peacock, B. Chain, Information-driven docking for tcr-pmhc complex prediction, Frontiers in immunology 12 (2021) 686127.

[32] S. Chaudhury, S. Lyskov, J. J. Gray, Pyrosetta: a script-based interface for implementing molecular modeling algorithms using rosetta, Bioinformatics 26 (5) (2010) 689–691.

[33] P. Chys, P. Chacón, Random coordinate descent with spinor-matrices and geometric filters for efficient loop closure, Journal of chemical theory and computation 9 (3) (2013) 1821–1829.

[34] X. Huang, R. Pearce, Y. Zhang, Faspr: an open-source tool for fast and accurate protein side-chain packing, Bioinformatics 36 (12) (2020) 3758–3765.

[35] D. R. Koes, M. P. Baumgartner, C. J. Camacho, Lessons learned in empirical scoring with smina from the csar 2011 benchmarking exercise, Journal of chemical information and modeling 53 (8) (2013) 1893–1904.

[36] M. Giulini, V. Reys, J. M. Teixeira, B. Jiménez-García, R. V. Honorato, A. Kravchenko, X. Xu, R. Versini, A. Engel, S. Verhoeven, et al., Haddock3: A modular and versatile platform for integrative modeling of biomolecular complexes, Journal of Chemical Information and Modeling (2025).

[37] S. Basu, B. Wallner, Dockq: A quality measure for protein-protein docking models, PLOS ONE 11 (8) (2016) 1–9. doi:10.1371/journal.pone.0161879.

[38] R. F. Alford, A. Leaver-Fay, J. R. Jeliazkov, M. J. O'Meara, F. P. DiMaio, H. Park, M. V. Shapovalov, P. D. Renfrew, V. K. Mulligan, K. Kappel, J. W. Labonte, M. S. Pacella, R. Bonneau, P. Bradley, R. L. Dunbrack, R. Das, D. Baker, B. Kuhlman, T. Kortemme, J. J. Gray, The rosetta all-atom energy function for macromolecular modeling and design, Journal of Chemical Theory and Computation 13 (6) (2017) 3031–3048. doi:10.1021/acs.jctc.7b00125.

[39] I. Kufareva, R. Abagyan, Methods of protein structure comparison, in: Homology modeling: Methods and protocols, Springer, 2012, pp. 231–257.

[40] V. Mariani, M. Biasini, A. Barbato, T. Schwede, lddt: a local superposition-free score for comparing protein structures and models using distance difference tests, Bioinformatics 29 (21) (2013) 2722–2728.

[41] C. Mirabello, B. Wallner, Dockq v2: improved automatic quality measure for protein multimers, nucleic acids, and small molecules, Bioinformatics 40 (10) (2024) btae586. doi:10.1093/bioinformatics/btae586.

[42] J. Leem, S. H. P. de Oliveira, K. Krawczyk, C. M. Deane, STCRDab: the structural T-cell receptor database, Nucleic acids research 46 (D1) (2018) D406–D412.

[43] P. Zareie, C. Szeto, C. Farenc, S. D. Gunasinghe, E. M. Kolawole, A. Nguyen, C. Blyth, X. Y. Sng, J. Li, C. M. Jones, et al., Canonical

t cell receptor docking on peptide–mhc is essential for t cell signaling, Science 372 (6546) (2021) eabe9124.

[44] V. Lin, M. Cheung, R. Gowthaman, M. Eisenberg, B. M. Baker, B. G. Pierce, Tcr3d 2.0: expanding the t cell receptor structure database with new structures, tools and interactions, Nucleic Acids Research 53 (D1) (2025) D604–D608.

[45] 10x Genomics, A new way of exploring immunity–linking highly multiplexed antigen recognition to immune repertoire and phenotype, Tech. rep (2019).

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

[47] A. Banerjee, D. J. Pattinson, C. L. Wincek, P. Bunk, A. Axhemi, S. R. Chapin, S. Navlakha, H. V. Meyer, Comprehensive epitope mutational scan database enables accurate t cell receptor cross-reactivity prediction, bioRxiv (2025) 2024–01.

[48] E. C. Meng, T. D. Goddard, E. F. Pettersen, G. S. Couch, Z. J. Pearson, J. H. Morris, T. E. Ferrin, Ucsf chimerax: Tools for structure building and analysis, Protein Science 32 (11) (2023) e4792.