A Unified Framework for TCR-pMHC Structural Model Assessment

Alex Ascunce-París[1], Miguel Romero-Durana[1], Alfonso Valencia[1,2], Roc Farriol-Duran[1] and Victor Guallar[1,2]

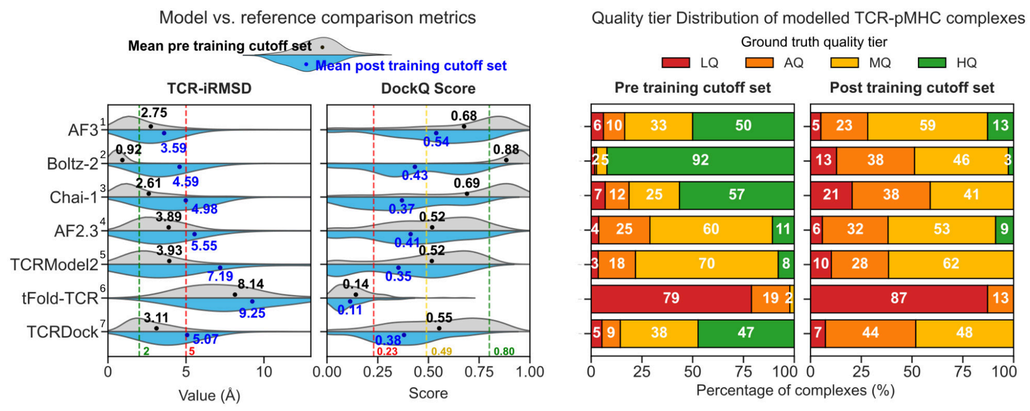
[1]Computational Biology Group, Life Sciences Department, Barcelona Supercomputing Center (Spain), [2]Catalan Institution for Research and Advanced Studies (ICREA) (Spain)

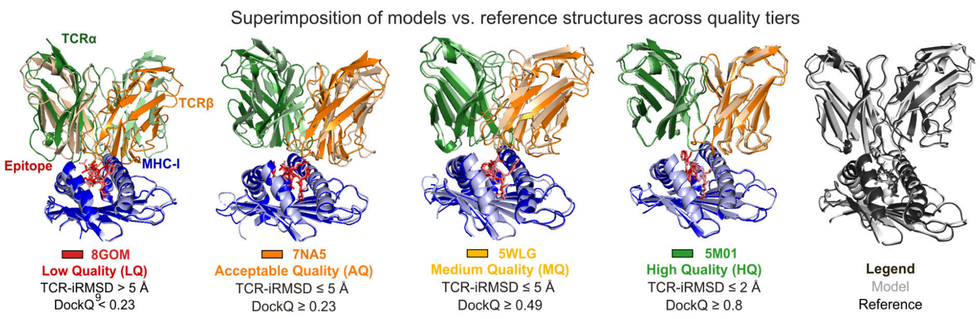
Abstract

T-cell receptor (TCR) recognition of peptide–MHC complexes (pMHCs) is central to adaptive cellular immunity. Structural characterization of these interactions is crucial for understanding TCR specificity and further developing TCR-based immunotherapies. Yet experimental structures remain scarce, with only 275 TCR-pMHC class I complexes available in the PDB[8] by January 2026. Computational modeling can help to expand this limited structural space. Downstream applications require both accurate structural models and robust methods to assess model quality. However, TCR diversity and CDR loop flexibility make accurate prediction difficult, and reliable model quality evaluation remains challenging in the absence of experimental references. To assess the current state of the art, we benchmarked seven modeling tools on TCR-pMHC class I complexes in the PDB, comparing predicted with experimental structures and splitting complexes by deposition date relative to the training cutoff. AlphaFold3 generalized best, achieving the highest overall structural quality on complexes deposited after the cutoff. We then developed an interpretable framework that uses confidence metrics extracted from the models themselves to predict the quality tier a model would receive when compared with its experimental counterpart. We applied it to 21,775 AlphaFold3 models representing 4,355 TCR-pMHC complexes from TCRvdb and IMMREP23. Higher predicted quality tiers were consistently enriched for experimentally validated, cognate TCR-pMHC interactions, including epitopes absent from the PDB and our classifier's training data. Overall, reference-free quality assessment provides a practical way to prioritize reliable TCR-pMHC models at scale.

Model benchmarking

All tools drop sharply after the training cutoff, with AF3 degrading the least*

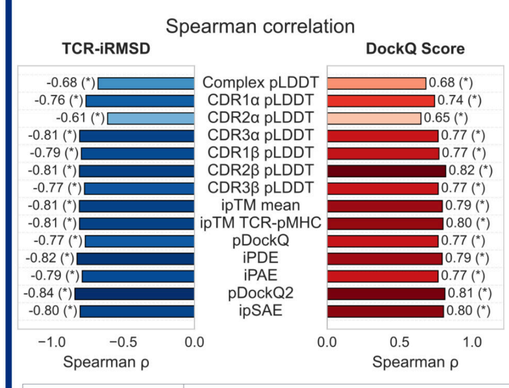


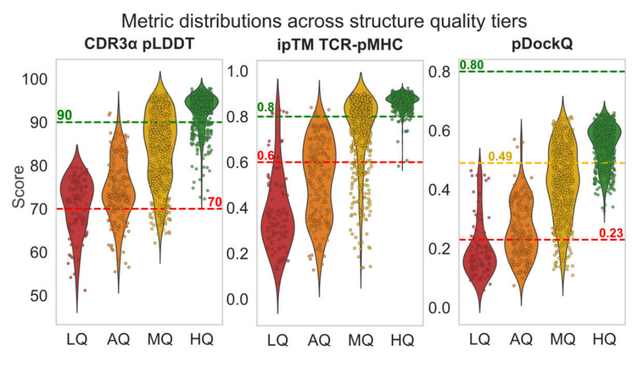


*** Complexes split by deposition date vs. training cutoff, 2023-06-01, applied uniformly across tools**

Single vs. combined metrics

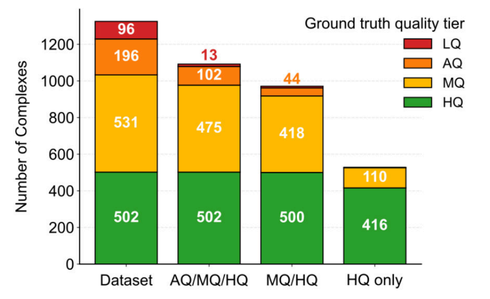
Combining metrics improves classification, at the cost of complex filters





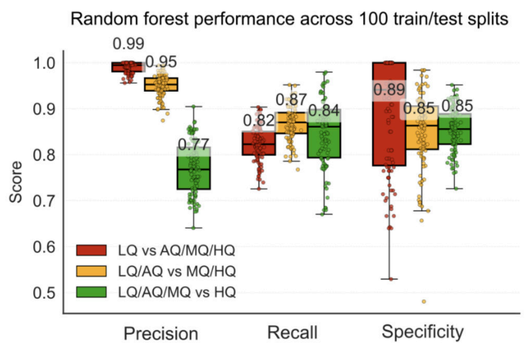
LQ vs. AQ/MQ/HQ	pLDDT CDR1α > 70 AND pLDDT CDR3α > 70 AND pLDDT CDR1β > 70 AND pLDDT CDR2β > 70 AND pLDDT CDR3β > 70 OR ipSAE > 0.60 OR ipTM (TCR-pMHC) > 0.60 AND pDockQ > 0.23 OR pDockQ2 > 0.23
LQ/AQ vs. MQ/HQ	pLDDT Complex > 70 OR pLDDT CDR1α > 70 AND pLDDT CDR2α > 70 OR pLDDT CDR3α > 70 AND pLDDT CDR1β > 70 OR pLDDT CDR2β > 70 OR pLDDT CDR3β > 70 OR ipSAE > 0.60 OR ipTM (TCR-pMHC) > 0.60 OR pDockQ2 > 0.49
LQ/AQ/MQ vs. HQ	pLDDT Complex > 90 OR pLDDT CDR1α > 90 AND pLDDT CDR2α > 90 OR pLDDT CDR3α > 90 AND pLDDT CDR2β > 90 AND pLDDT CDR3β > 90 AND ipTM > 0.80

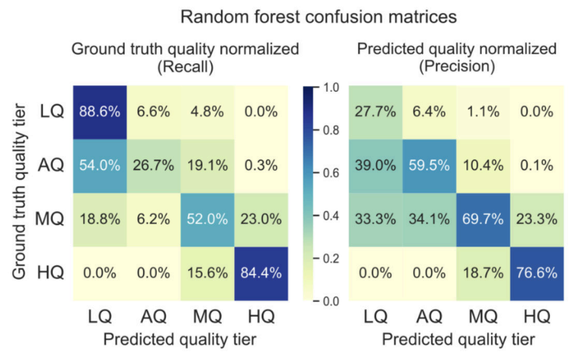
Application of balanced filters to PDB modelled structures

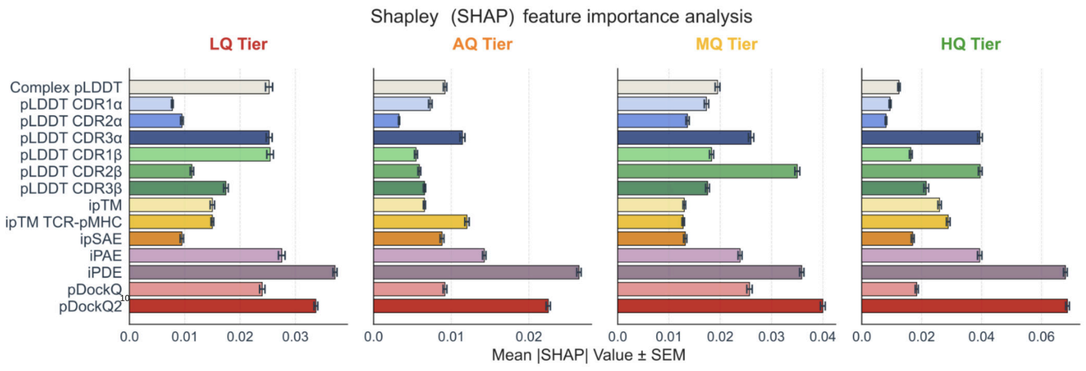


Reference-free classifier based on model confidence metrics

Reproduces quality tiers defined by comparison with PDB structures

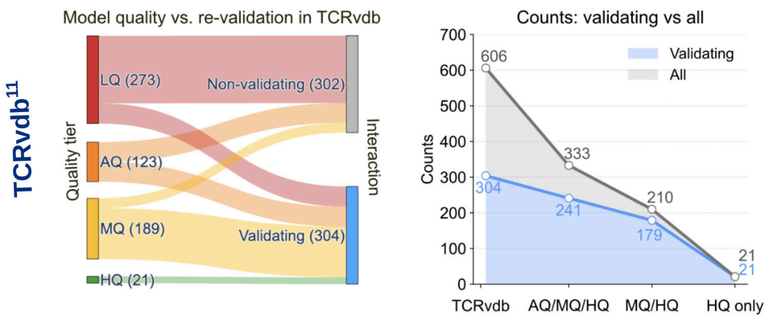


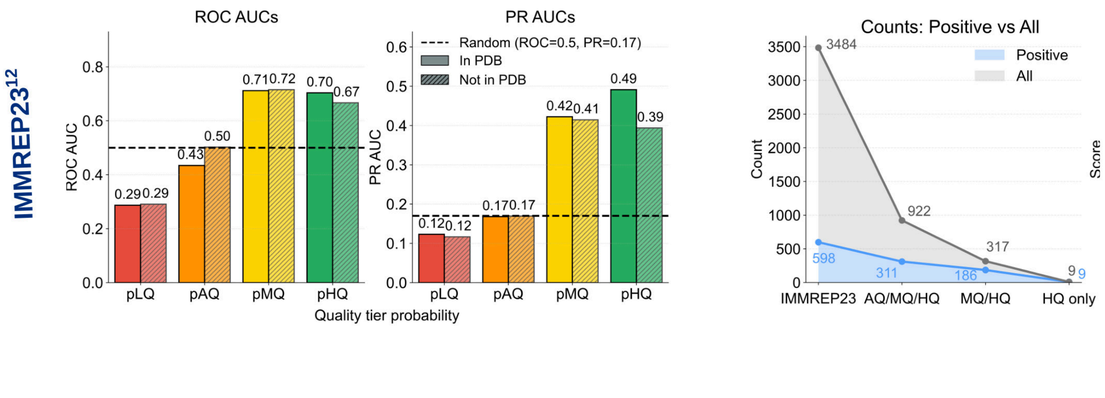




Prioritizing cognate TCR-pMHC interactions

Higher tiers are enriched for cognate interactions, including epitopes absent from the PDB and unseen by our model





Conclusions

- AlphaFold3 generalizes best among the seven tested TCR-pMHC modeling tools, achieving the highest overall structural quality on complexes deposited after the training set cutoff.
- No single confidence metric separates adjacent quality tiers. Combining metrics helps, but creates complex, threshold-dependent filters prone to overfitting.
- A random forest using reference-free confidence metrics reproduces the quality tiers. Errors occur only between adjacent tiers, limiting contamination by low-quality structures.
- Higher predicted tiers are enriched for cognate TCR-pMHC interactions in TCRvdb and IMMREP23, including epitopes absent from the PDB and unseen by our model.
- Quality-based filtering reduces large model libraries to smaller, higher-confidence candidate sets for downstream analysis or experimental screening.

References

1. Abramson et al. Nature (2024)
2. Passaro et al. bioRxiv (2025)
3. Team CD et al. bioRxiv (2024)
4. Evans et al. bioRxiv (2021)
5. Yin et al. Nucleic Acids Res (2023)
6. Wu et al. bioRxiv (2025)
7. Bradley P. eLife (2023)
8. Berman et al. Nucleic Acids Res (2000)
9. Mirabello & Wallner. Bioinformatics (2024)
10. Zhu et al. Bioinformatics (2023)
11. Messemaker et al. bioRxiv (2025)
12. Nielsen et al. ImmunoInformatics (2024)

Scan for the full paper

Miguel Romero is supported by a fellowship within the 'Generación D' initiative, Red.es, Ministerio para la Transformación Digital y de la Función Pública, for talent attraction (C005/24-ED CV1). Funded by the European Union NextGenerationEU funds, through PRTR.