



ICLR 2026

Quantifying Cross-Attention Interaction in Transformers for Interpreting TCR-pMHC Binding

Jiarui Li¹, Zixiang Yin¹, Haley Smith²,
Zhengming Ding¹, Samuel J Landry², Ramgopal R. Mettu^{1*}

¹ Department of Computer Science, Tulane University

² Biochemistry and Molecular Biology, Tulane University School of Medicine

* Corresponding Author

Project Page: <https://qcai.jiarui.li/>



ICLR
International Conference On
Learning Representations



BACKGROUND

TCR-pMHC Binding Prediction



Metu Landry
Laboratory



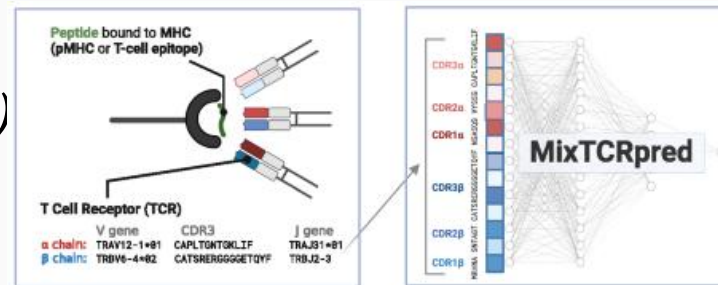
INPUTS:

1. CDR3 α Chain (e.g., $[ALGDHSGSWQLI]$)
2. CDR3 β Chain (e.g., $[ASSLRTGANSDYT]$)
3. Peptide (e.g., $[GVYATSSAVRLR]$)

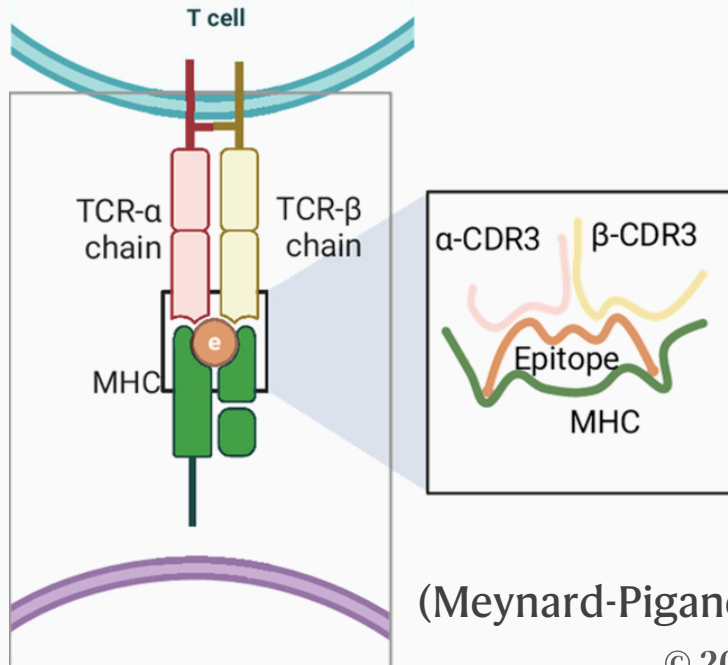
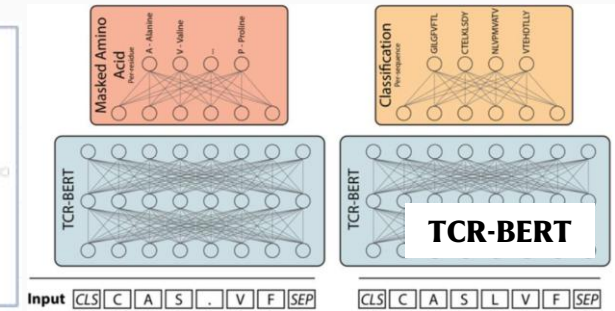
OUTPUTS:

1. Binder/non-binder (*True/False*)

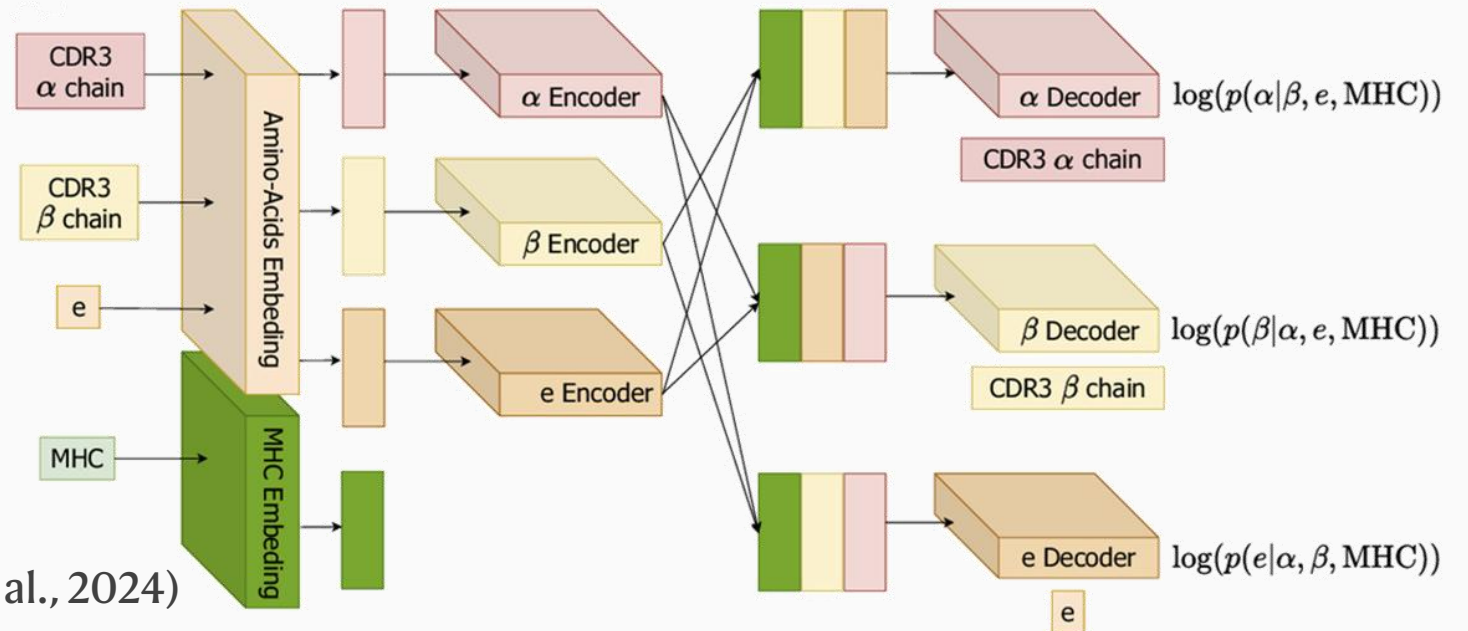
(Groce et al, 2024)



(Wu et al, 2024)



(Meynard-Piganeau et al., 2024)



BACKGROUND

Interpretable Machine Learning



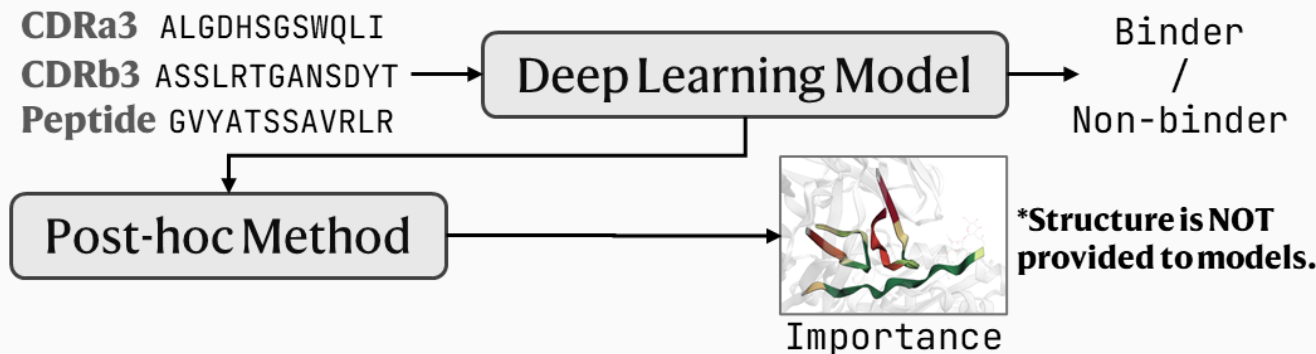
Metu Landry
Laboratory



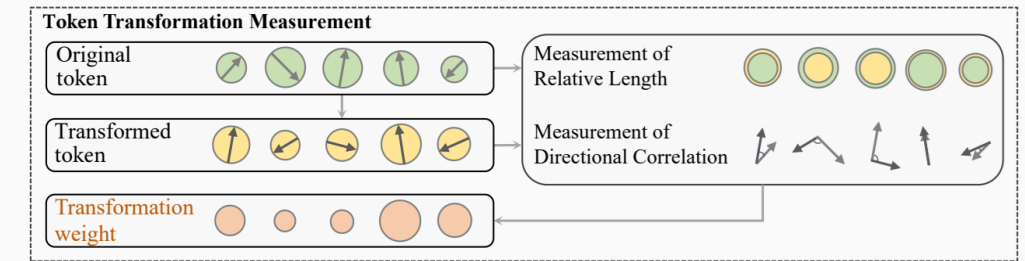
Black Box



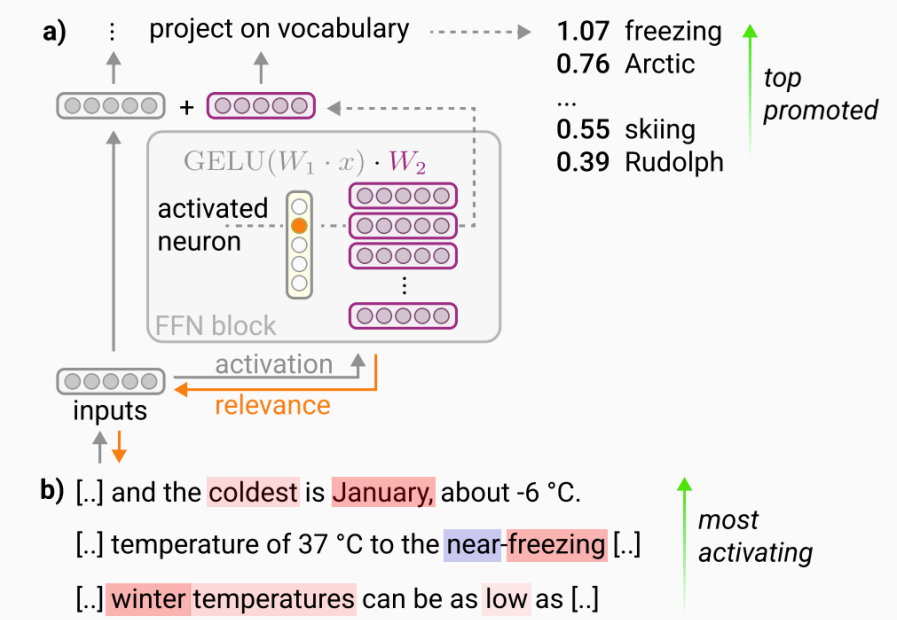
Post-hoc Interpretation



TokenTM (Wu et al., 2024)



AttnLRP (Achtibat et al., 2024)

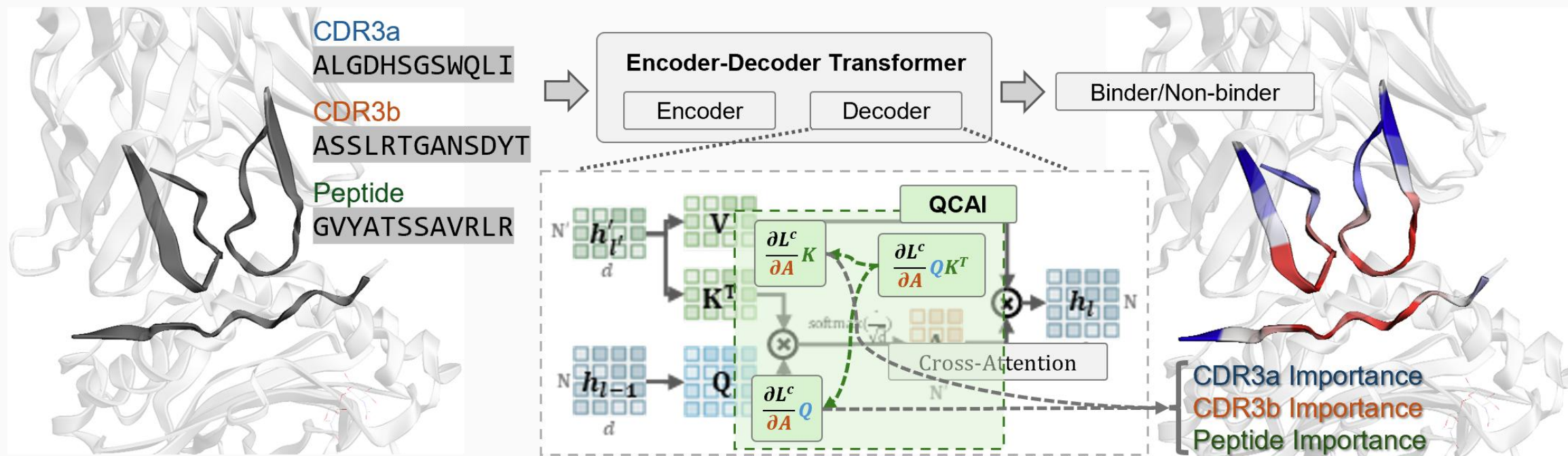


OUR APPROACH

Quantifying Cross-Attention Interaction



Metu Landry
Laboratory



OUR APPROACH

TCR-XAI Benchmark & BRHR

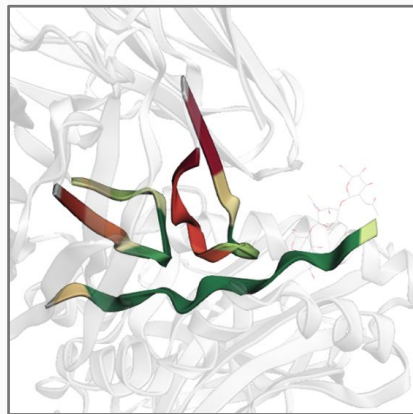


Metu Landry
Laboratory



TCR-XAI

274 High Resolution TCR-pMHC Complex Structural Data from **STCRDab** (Leem et al., 2018) and **TCR3d 2.0** (Lin et al., 2025)



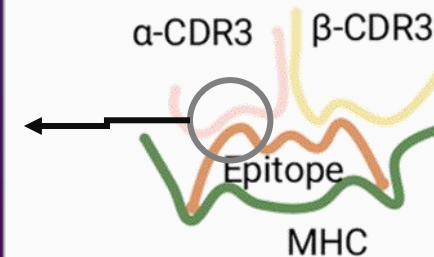
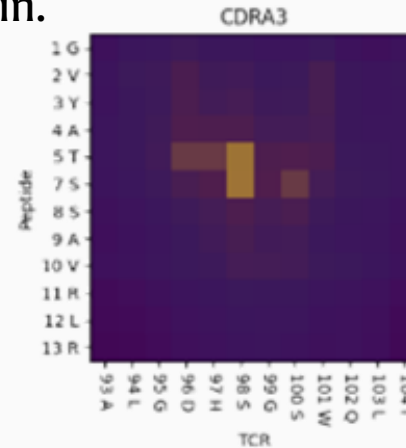
How much agreement?

BRHR

Binding Region Hit Rate

Importance vs. Binding Structure

- Residue importance correlates with binding structure: more interactions \rightarrow higher importance in the model.
- **BRHR.k**: Measures whether the top-k important residues in one chain align with the top-k residues closest to the other chain.



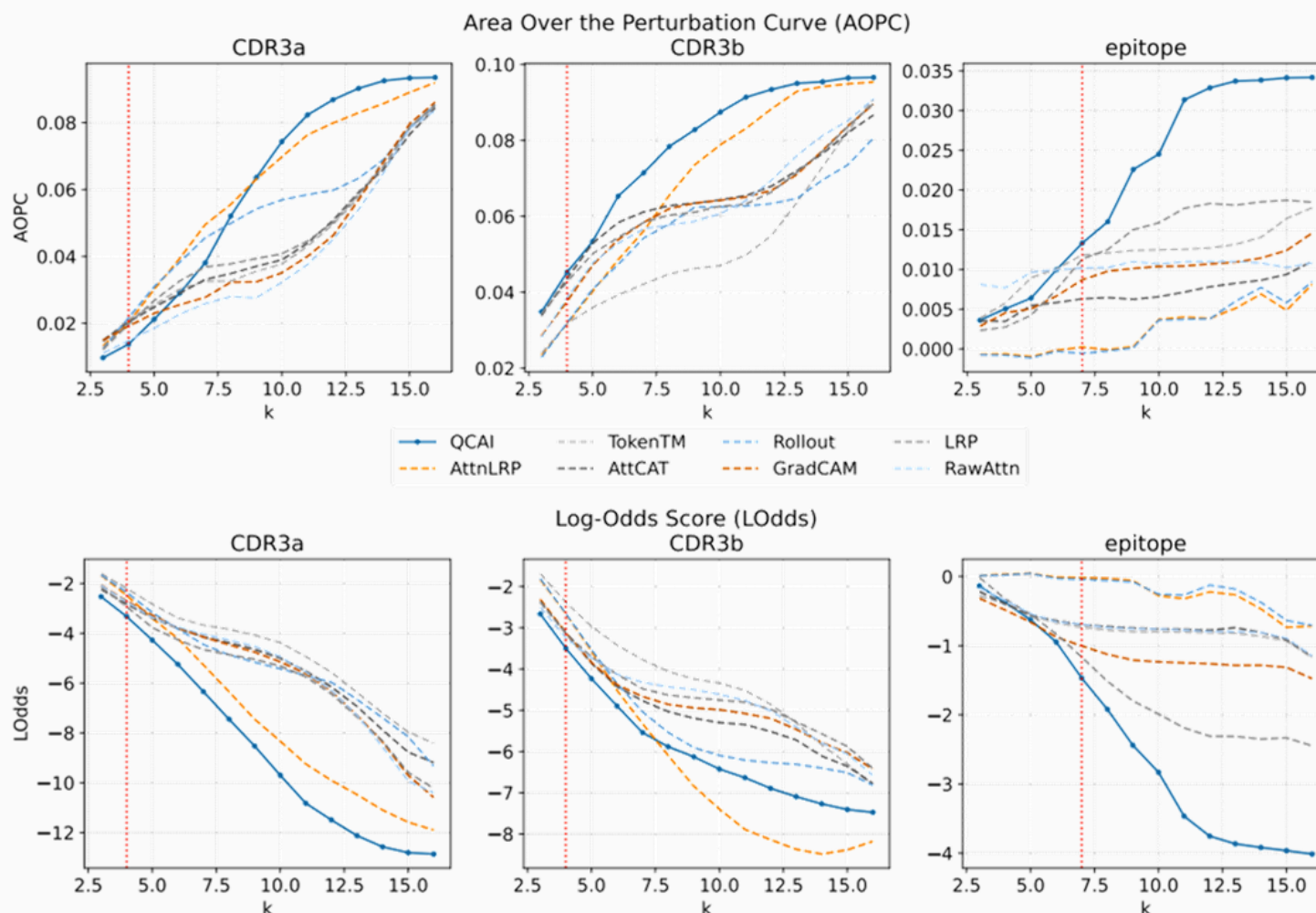
Use **distance** between two given chains to quantitatively evaluate **explanation quality**.

RESULTS

Perturbation Experiments



Metu Landry
Laboratory



LOdds

Log-Odds Score

$$\text{LOdds}(k) = \frac{1}{N} \sum_{i=1}^N \log \frac{p(\hat{y}|\tilde{\mathbf{x}}_i^k)}{p(\hat{y}|\mathbf{x}_i)}$$

AOPC

Area Over the Perturbation Curve

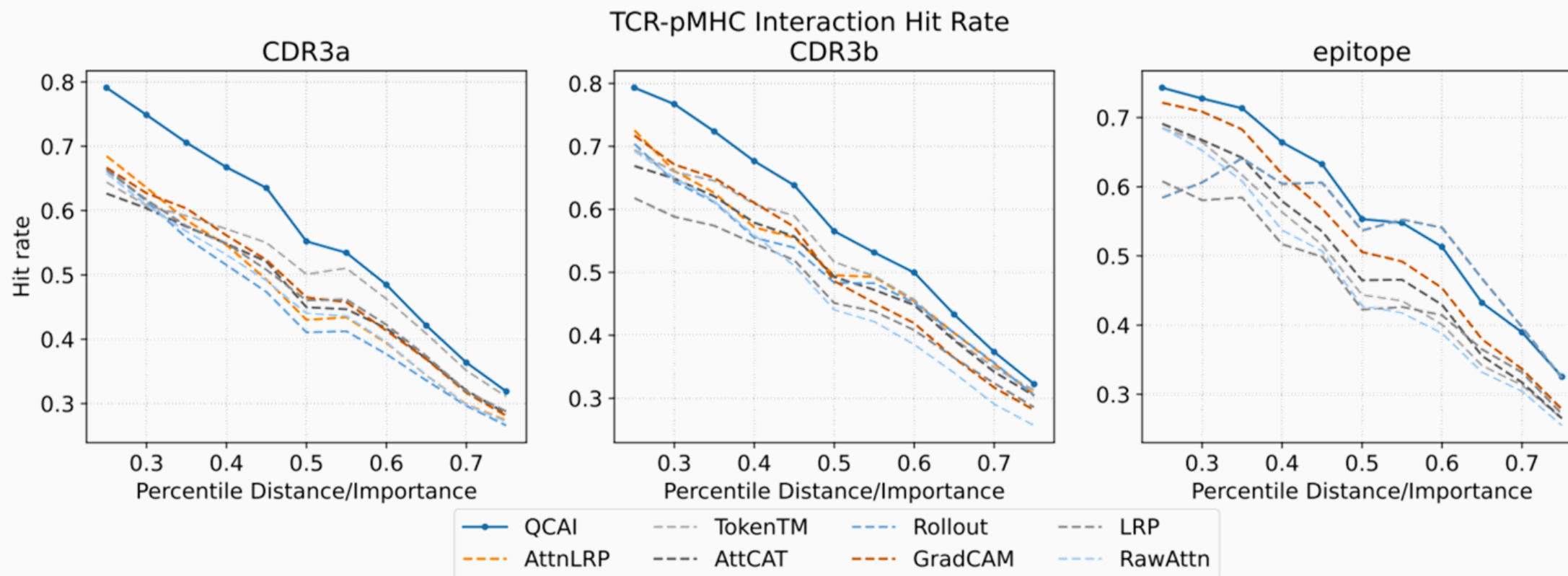
$$\text{AOPC}(k) = \frac{1}{N} \sum_{i=1}^N p(\hat{y}|\mathbf{x}_i) - p(\hat{y}|\tilde{\mathbf{x}}_i^k),$$

RESULTS

Binding Region Hit Rate



Mettu Landry
Laboratory



Binding Site Hit Rate (BRHR)

compares top-ranked residues by explanation score against top interacting residues by distance.

RESULTS

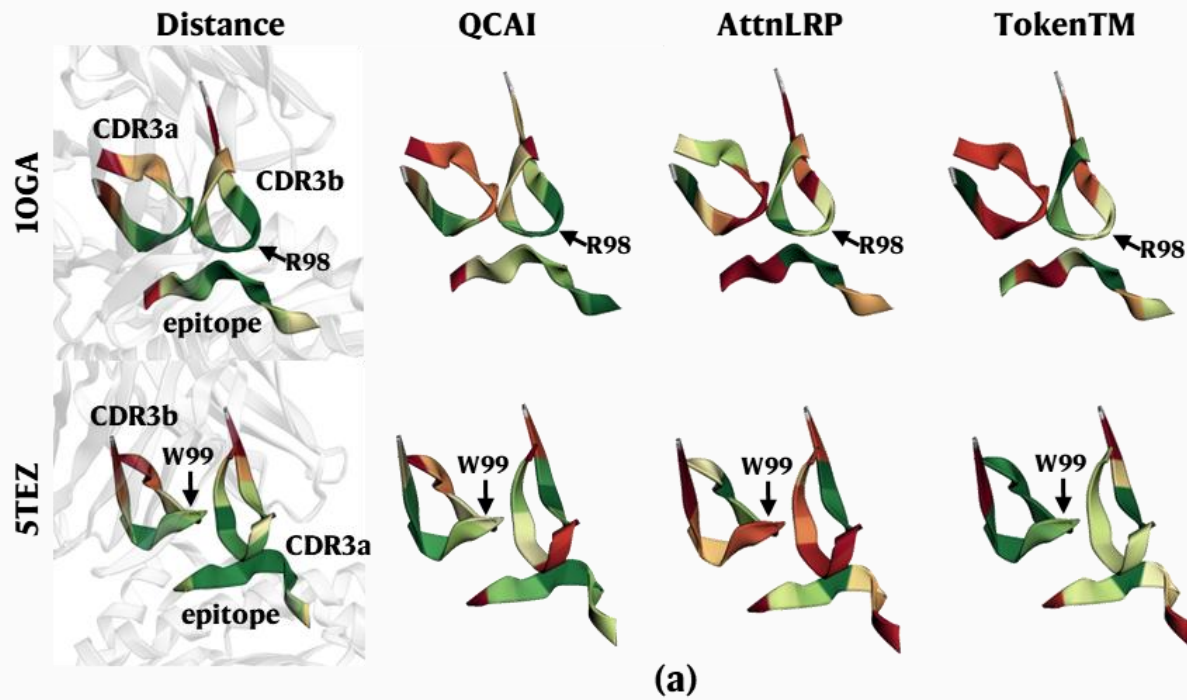
Case Studies



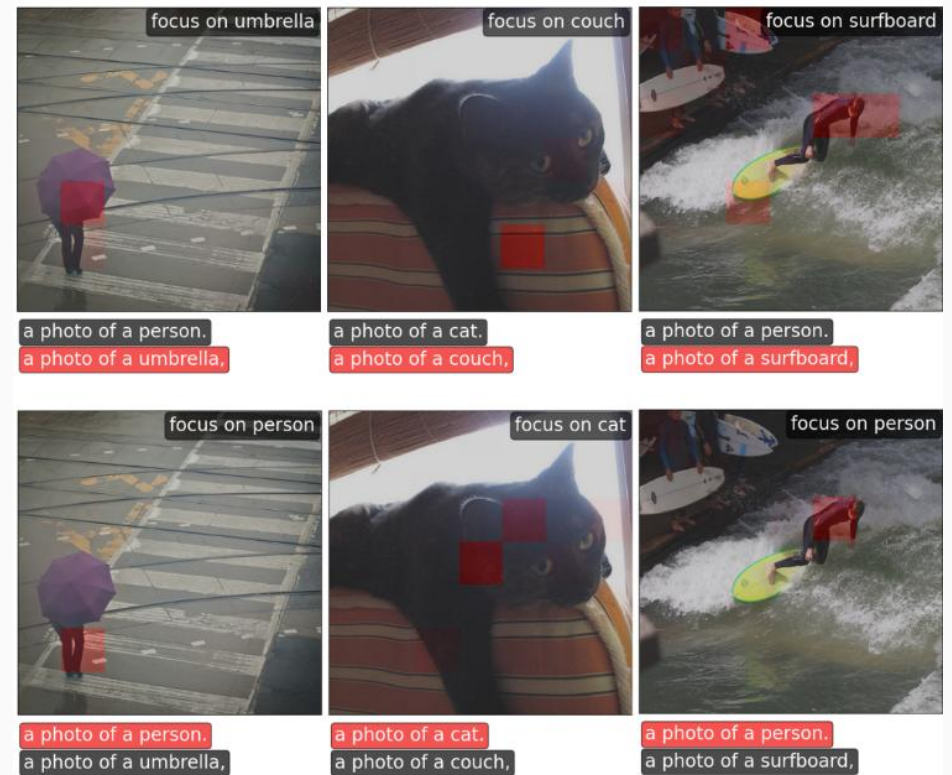
Metu Landry
Laboratory



TULIP for Influenza Matrix Protein



Vision Language Model on ImageNet



CONCLUSION

Summary & Future Study



Metu Landry
Laboratory



- **QCAI:** Quantifying Cross-Attention Interaction
- **TCR-XAI:** Quantitative Benchmark for TCR-pMHC Explanation Evaluation
- Demonstrate the role of cross-attention in TCR-pMHC Binding
- We have used QCAI for Rational Model Design (*Li et al., 2025, ACM-BCB*)
- We are working on Explain-By-Design models, and the distinction between interpretability and explainability.

Acknowledgement

- Ding Group, Landry Group, and Gragert Group;
- Harold L. and Heather E. Jurist Center of Excellence for Artificial Intelligence at Tulane University;
- National Institutes of Health (U54-CA260581);
- Tulane SOM Pilot Funding;
- Tulane Center of Excellence for Emerging and Re-emerging Infectious Diseases (CEERID) Pilot Research Program
- Lavin-Bernick Faculty Grant Proposal Research and Scholarly Activities Support.

**Project Page for Code,
Data, and Models**
(<https://qcai.jiarui.li/>)

