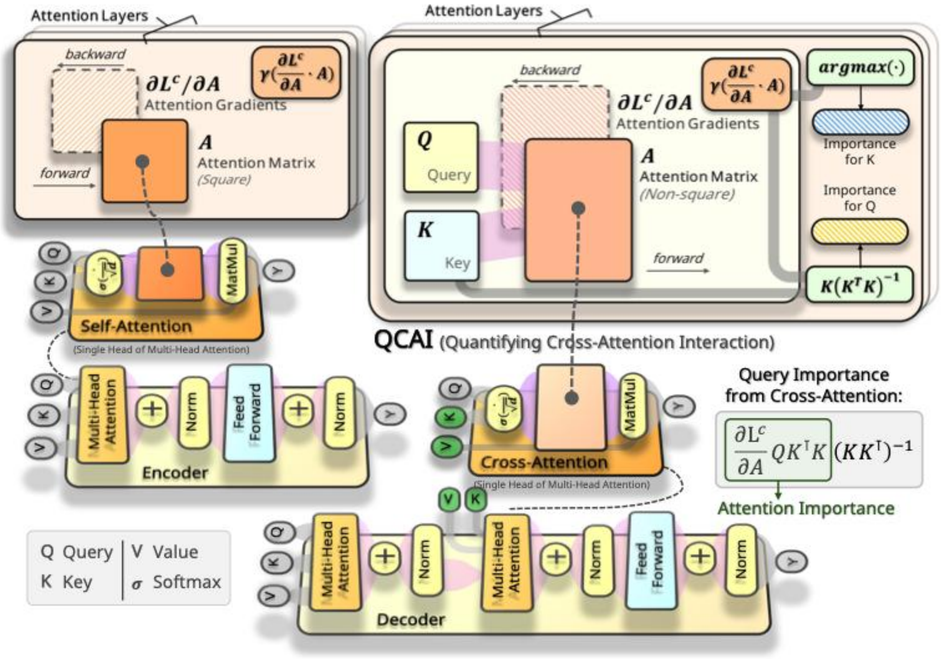
# Quantifying Cross-Attention Interaction in Transformers for Interpreting TCR-pMHC Binding

Jiarui Li [1], Zixiang Yin [1], Haley Smith [2], Zhengming Ding [1], Samuel J Landry [2], Ramgopal R. Mettu [1]

[1] Department of Computer Science, Tulane University

[2] Biochemistry and Molecular Biology, Tulane University School of Medicine

Post-hoc interpretability methods work well for encoder-only transformers with self-attention, but do not capture the contributions of cross-attention. Unlike self-attention (Q, K, V from the same input), cross-attention uses separate sources for Q and (K, V). This breaks the direct link between attention weights and query/key token importance. **QCAI** provides a way to handle this asymmetry so that cross-attention can be interpreted.

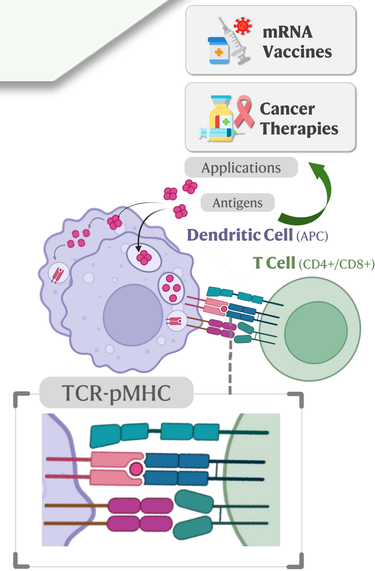



## TCR-pMHC PREDICTION

T cells play a central role in the adaptive immune system by recognizing antigens presented by Major Histocompatibility Complex (pMHC) molecules via T Cell Receptors (TCRs). Modeling T cell receptor-pMHC binding is central to understanding immune mechanisms and developing therapies.

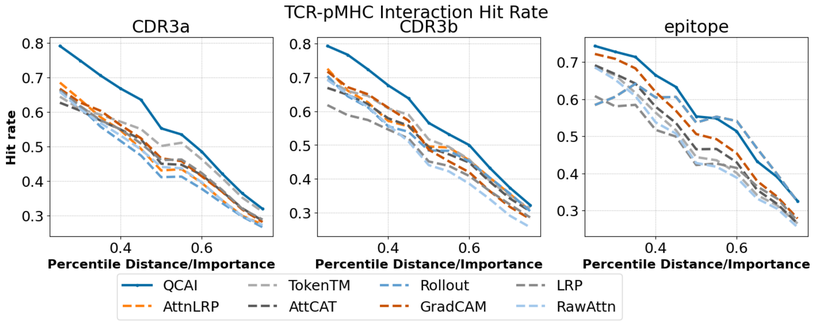
**CDR3 α Chain** (e.g., [*ALGDHSGSWQLI*])
**CDR3 β Chain** (e.g., [*ASSLRTGANSDYT*])
**Peptide** (e.g., [*GVYATSSAVRLR*])

**Binder / Non-Binder** (i.e., [*TRUE/FALSE*])



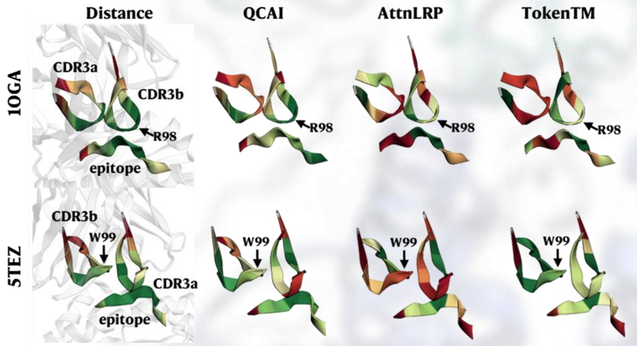

## TCR-XAI BENCHMARK

We have compiled a ground-truth TCR-pMHC dataset of **274** high resolution X-ray structures from **STCRDab** and **TCR3D 2.0.** We demonstrate the interpretability enabled by QCAI of TULIP predictions on this dataset using a quantitative metric we call **Binding Region Hit Rate** (**BRHR**).




## CASE STUDY

Using QCAI to study interactions between a TCR and peptide-MHC for an influenza matrix protein, we recapitulate the importance of two distinct residues for the same TCR in two different binding conformations.




## APPLICATION TO VISION LANGUAGE MODELS

QCAI is general and can be used in scenarios such as in a vision-language model using CLIP, to capture the contributions of image/text fusion.