TRAP: A Contrastive Learning-Enhanced Framework for Robust TCR-pMHC Binding Prediction with Improved Generalizability

Jingxuan Ge[1,2], Jike Wang[1,2], Qing Ye[1], Liqiang Pan[1], Yu Kang[1], Chao Shen[1], Yafeng Deng[2], Chang-Yu Hsieh[1,*], Tingjun Hou[1,*]

[1]College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, Zhejiang, China

[2]CarbonSilicon AI Technology Company, Ltd., Hangzhou 310018, Zhejiang, China

***Corresponding authors**

Tingjun Hou

E-mail: tingjunhou@zju.edu.cn

Chang-Yu Hsieh

E-mail: kimhsieh@zju.edu.cn

Supplementary Table 1. Performance of different prediction methods in scenario 1.

Model	AUC	AUPR
TRAP	0.9155±0.0108	0.8366±0.0074
NetTCR-2.0	0.8507±0.0021	0.6363±0.0028
epiTCR (epitope)	0.8805±0.0004	0.6796±0.0006
epiTCR (pMHC)	0.8896±0.0001	0.6834±0.0006
TEIM	0.8621±0.0013	0.6598±0.0053

Supplementary Table 2. Performance of different prediction methods in scenario 2.

Model	AUC	AUPR
TRAP	0.7507±0.0440	0.3454±0.0410
NetTCR-2.0	0.6522±0.0147	0.1663±0.0173
epiTCR (epitope)	0.6354±0.0038	0.2938±0.0057
epiTCR (pMHC)	0.6775±0.0049	0.3391±0.0035
TEIM	0.6241±0.0084	0.2470±0.0125

Supplementary Table 3. Epitope-level AUC and AUPR scores in scenario 1.

Model	AUC	AUPR
TRAP	0.7988±0.2591	0.6404±0.3429
NetTCR-2.0	0.6984±0.2601	0.4306±0.3549
epiTCR (epitope)	0.7482±0.2777	0.5864±0.3514
epiTCR (pMHC)	0.7757±0.2590	0.6083±0.3532
TEIM	0.7400±0.2698	0.5424±0.3556

Supplementary Table 4. Epitope-level AUC and AUPR scores in scenario 2.

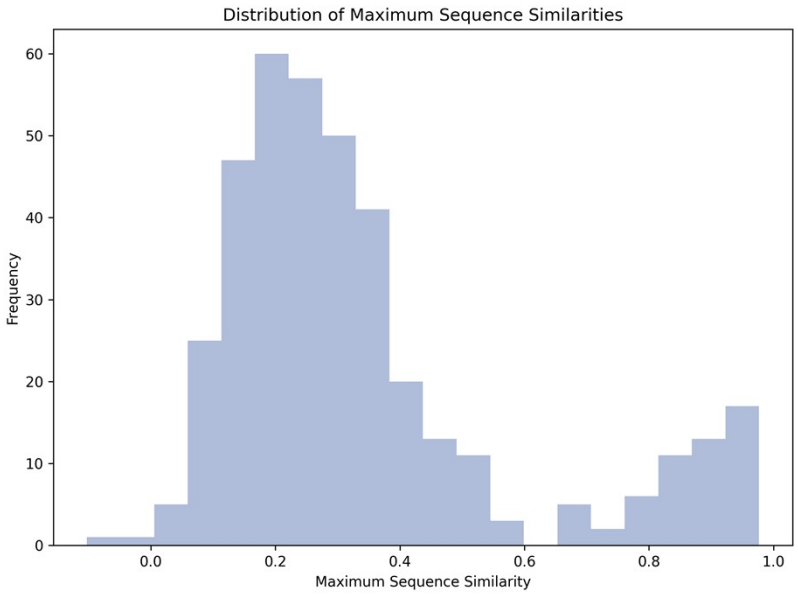
Model	AUC	AUPR
TRAP	0.7474±0.2903	0.4761±0.4036
NetTCR-2.0	0.6469±0.2580	0.2743±0.3165
epiTCR (epitope)	0.6174±0.3025	0.3311±0.3834
epiTCR (pMHC)	0.6725±0.2918	0.3835±0.3995
TEIM	0.6198±0.3014	0.3140±0.3585

Supplementary Table 5. False positive rates for epitopes in different negative sample strategies.

Epitope	Model	Unified	Randomly
KLGGALQAK	TRAP	68.7%	94.9%
	epiTCR (epitope)	95.3%	100.0%
	epiTCR (pMHC)	96.9%	100.0%
YVLDHLIVV	TRAP	13.2%	43.2%
	epiTCR (epitope)	32.1%	100.0%
	epiTCR (pMHC)	35.2%	100.0%
GLCTLVAML	TRAP	12.8%	24.3%
	epiTCR (epitope)	8.1%	99.9%
	epiTCR (pMHC)	45.7%	100.0%

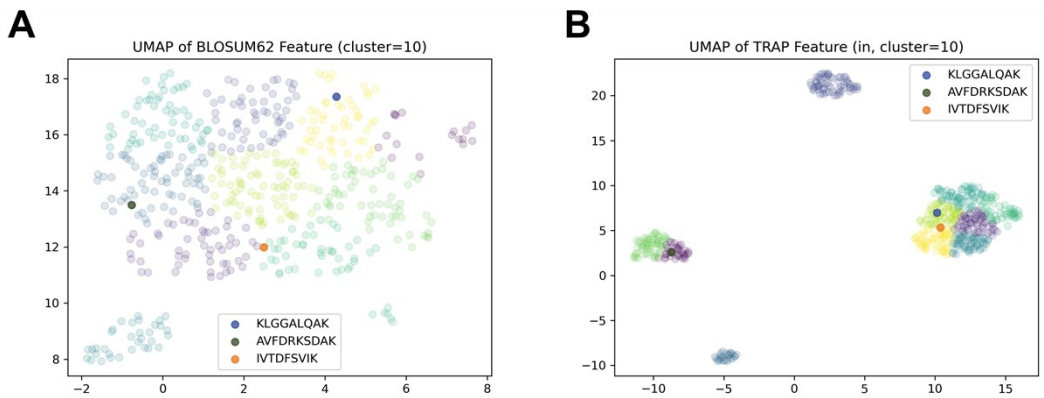
Supplementary Table 6. MM/GBSA calculation results of binding free energy.

Complex	ΔG (kcal/mol)
Crystal Structure	-51.9544±5.1208
AITR	-51.0081±6.2229
AIRQ	-78.0301±6.5745



Supplementary Figure 1. In scenario 2, the similarity distribution between the epitope in the test set and the epitope with the highest similarity in the training set.

We utilized the pairwise2 module from Biopython to conduct sequence alignment and compute the similarity between antigen peptides in the test and training sets for scenario 2. The gap opening penalty was set to -10, the gap extension penalty to -0.5, and the self-alignment score of the original sequence was used as the benchmark for similarity calculation (a few scores are below 0 due to large differences, resulting in negative penalties computed by the alignment algorithm). For each epitope in the test set, we iterated over all epitopes in the training set to compute similarity and selected the one with the highest similarity for further analysis.



Supplementary Figure 2. A, B show UMAP dimensionality reduction after K-Means clustering of BLOSUM62 and input of TRAP feature for pMHCs, respectively.