Research Highlight | Open Access

AI Agent

Fan *et al. AI Agent* **2026**, *2*, 31 **DOI:**10.20517/aiagent.2026.30

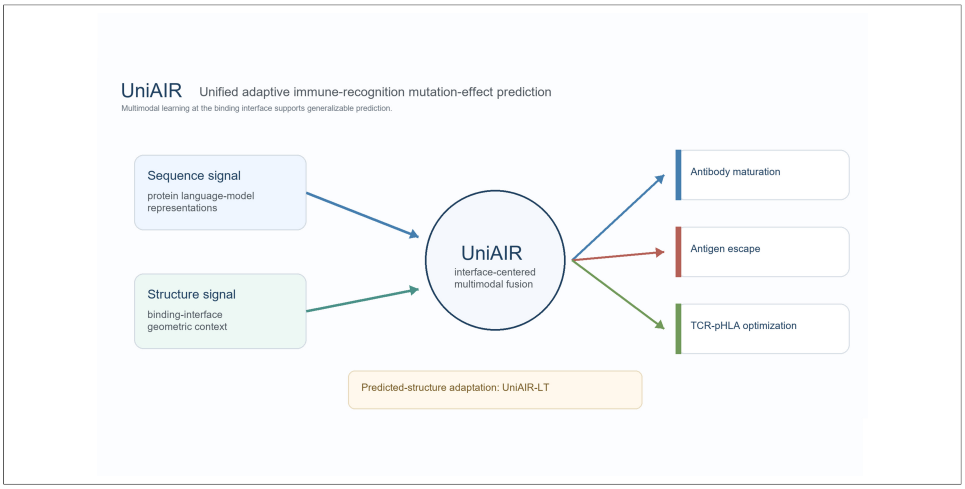
UniAIR: a unified map for mutation effects in adaptive immune recognition

Weisi Fan[1], Hamad Ullah[1], Yongqi Chen[2], Ziheng Chen[3]

Citation: Fan, W.; Ullah, H.; Chen, Y.; Chen, Z. UniAIR: a unified map for mutation effects in adaptive immune recognition. *AI Agent* **2026**, *2*, 31. https://dx.doi.org/10.20517/aiagent.2026.30

Received: 19 Jun 2026 **First Decision:** 21 Aug 2026 **Revised:** 21 Aug 2026 **Accepted:** 22 Sep 2026 **Published:** 5 Nov 2026

Academic Editor: Jianjun Liu **Copy Editor:** Tong Wang **Production Editor:** Tong Wang



WHY MUTATION-EFFECT PREDICTION MATTERS

Adaptive immune recognition depends on precise molecular contacts. Antibodies recognize antigens in humoral immunity, while T cell receptors (TCRs) recognize peptide antigens presented by major histocompatibility complex (MHC) or human leukocyte antigen (HLA) molecules in cellular immunity. Mutations at these binding interfaces can alter affinity, specificity and immune escape, directly affecting antibody engineering, viral surveillance, peptide-vaccine design and TCR-centered immunotherapy.

The challenge is the size and diversity of the mutation landscape. Exhaustive experimental testing is rarely practical, and available datasets differ in assay type, molecular system and structural completeness. UniAIR addresses this problem by treating the binding interface as a shared modeling object across antibody-antigen and TCR-pHLA/TCR-pMHC recognition, rather than building separate predictors for each narrow task.

Check for updates

[1]Department of Computer Science, Iowa State University, Ames, IA 50014, USA.
[2]Department of Statistical Science, University of California, Santa Cruz, Santa Cruz, CA 95064, USA.
[3]Department of Computer Science, Stony Brook University, New York, NY 11790, USA.

Correspondence to: Dr. Weisi Fan, Department of Computer Science, Iowa State University, Ames, IA 50014, USA. E-mail: weisifan@iastate.edu

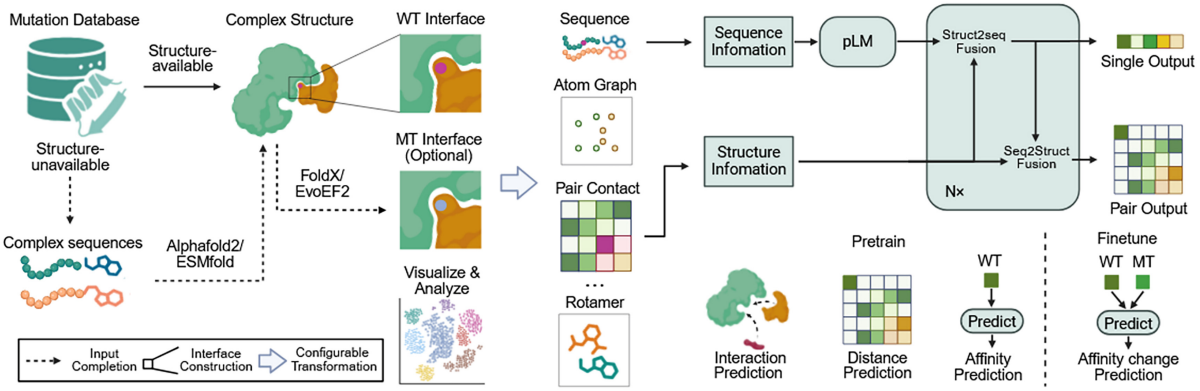


Figure 1. UniAIR architecture and workflow, adapted from the official UniAIR GitHub repository[2]. Source material is distributed under the repository's MIT license. pLM: Protein language model; WT: wild type.

WHAT UniAIR PROVIDES

Han *et al.* introduced UniAIR [Figure 1], short for Unified Adaptive Immune Recognition, as a modular multimodal framework for mutation-effect prediction across adaptive immune-recognition scenarios[1]. Published in *Nature Machine Intelligence* on 27 May 2026, the work aims to improve generalization across immune systems by combining standardized interface-centered data processing, sequence-structure fusion, multiexpert prediction and adaptation to predicted structures[1-3].

UniAIR differs from general protein language-model, protein-protein interaction and structure-based ΔΔG predictors by focusing on mutation-centered immune interfaces. Its public repository describes S-Former for sequence-structure fusion and supports comparison with baseline families including Evolutionary Scale Modeling (ESM2)-like protein language models[4], Protein-Protein Interaction (PPI) former[5], Gearbind, ΔΔG Prediction (DDGPred), UniBind and the Relative Distance Embedding (RDE) based ΔΔG baseline referred to as RDE-DDG in Figure 2[2].

HOW THE MODEL IS ORGANIZED

A key contribution of UniAIR is its standardized data layer. For mutation-effect prediction on binding affinity, the repository lists fields such as protein, interacting partners, mutation, Protein Data Bank (PDB) identifier, label, chain identifiers, valid chains, wild-type and mutant structure paths, and fold assignments[2]. For ESSM pretraining, it lists related fields including protein, partners, PDB identifier, chain information, structure path, binding free energy and split information[2].

The framework also handles heterogeneous labels with task-aware configuration rather than a single universal conversion. The repository exposes fields such as cols_label, loss_types and loss_weights, and provides separate modes for training, testing, transfer, few-shot prediction and scanning[2]. Thus, UniAIR's generalization claim is best understood as a shared interface-centered representation and workflow across tasks, not as evidence that affinity changes, escape fractions, binary labels and half-maximal inhibitory concentration (IC50) values are directly interchangeable.

For structural inputs, mutant structures can be prepared through the repository's make_input function, with EvoDesign physical energy function 2 (EvoEF2) used in preprocessing; wild-type structures can be generated with evolutionary ccale modeling fold (ESMFold)[4] or open-source protein folding framework (OpenFold) scripts[2]. When experimental structures are unavailable, the paper describes UniAIR-LT as a lightweight latent-space adaptation module designed to reduce mismatch between predicted and experimentally

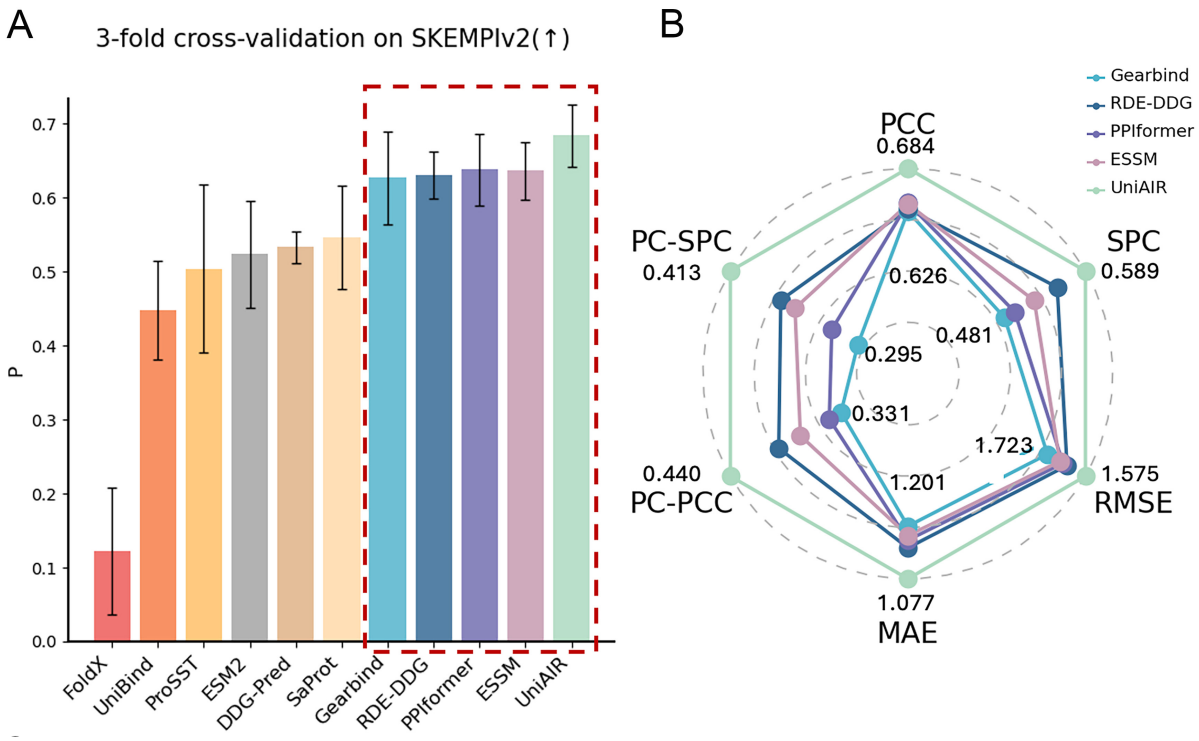


Figure 2. Performance comparison from the UniAIR repository, adapted from the official UniAIR GitHub repository[2]. (A) Three-fold cross-validation on SKEMPIv2; P denotes Pearson correlation coefficient; (B) Radar plot comparing Gearbind, RDE-DDG, PPIformer, ESSM and UniAIR across six performance metrics (PCC, SPC, RMSE, MAE, PC-PCC and PC-SPC). Each colored line connects the circular markers for one model across the six metric axes. RDE-DDG denotes the Relative Distance Embedding-based ΔΔG/binding-affinity baseline used in the figure. Source material is distributed under the repository's MIT license. PCC: Pearson correlation coefficient; SPC: Spearman correlation coefficient; RMSE: root mean square error; MAE: mean absolute error; PC-PCC: per-complex Pearson correlation coefficient; PC-SPC: per-complex Spearman correlation coefficient.

determined structures[1]. A systematic analysis of structural-confidence effects remains an important direction for future validation.

EVIDENCE, BENCHMARKS AND APPLICATIONS

The article reports that UniAIR was evaluated through large-scale benchmarking and independent tests across extracellular and intracellular immune-recognition tasks, including antibody maturation, antigen escape, TCR-pHLA optimization and cases without experimental structures[1]. The abstract states that UniAIR achieved state-of-the-art performance and robust predictions with minimal task-specific tuning[1]. To remain concise and verifiable, this highlight summarizes the benchmark scope rather than reproducing numerical metrics from inaccessible tables.

The repository documents the scale of the resources used around the framework. It lists pretraining data from SAbDab, STCRDab and PPB-Affinity; evaluation data including structural kinetic and energetic database of mutant protein interactions, version 2 (SKEMPIv2)[6] and T-cell receptor energy-based prediction (TCRen) benchmark No. 1; and downstream or independent tests involving human epidermal growth factor receptor 2 (HER2), T-cell receptor-peptide-major histocompatibility complex (TCR-pMHC), altered TCR ligand affinities and structures (ATLAS), LASSA antigen-escape datasets, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) IC50 datasets, unsupervised T-cell receptor-peptide-human leukocyte antigen (TCR-pHLA) structures and a Kirsten rat sarcoma viral oncogene homolog (KRAS) case[2].

The study also connects rapid artificial intelligence (AI) screening with higher-precision free-energy perturbation (FEP). The article states that UniAIR performed multiround peptide optimization of a TCR-pHLA complex under sparse feedback[1]. This pairing is useful because a fast model can prioritize broad mutation spaces, while FEP can be applied selectively to promising candidates.

OPEN RESOURCES AND REPRODUCIBILITY

The study is accompanied by public resources. *Nature Machine Intelligence* states that datasets are available through Hugging Face and that source code and model weights are available through GitHub and Zenodo[1]. The same page lists Zenodo record 10.5281/zenodo.19471285 as the initial release of UniAIR v1.0.0[1,3]. The GitHub repository is public, MIT licensed and identifies itself as the official UniAIR repository[2].

The repository provides practical entry points for cross-validation on SKEMPIv2, binary TCR-pMHC binding prediction, training on customized datasets through YAML Ain't Markup Language (YAML) files, zero-shot mutation scanning, few-shot escape prediction, UniAIR-LT training and ESSM pretraining[2]. These resources make the framework easier to inspect, reproduce and extend on new immune complexes.

WHAT IT MAY ENABLE

UniAIR's immediate value is prioritization. In antibody engineering, it can help rank mutations for experimental validation. In viral surveillance, it can help flag antigen mutations with potential escape effects when labeled data are limited. In TCR-centered immunotherapy, it can help explore peptide-HLA or TCR-pHLA mutation spaces where specificity and affinity must be balanced. These are enabling uses, not clinical claims: the paper establishes computational prediction and optimization workflows, not approved therapeutic decisions.

More broadly, UniAIR suggests that adaptive immune recognition can be modeled through a shared interface-centered representation across antibody-antigen and TCR-related systems. The framework does not replace immunological experiments, structural biology or molecular simulation. Instead, it provides a computational layer for triaging large mutation landscapes and selecting variants for deeper study.

UniAIR therefore represents a meaningful step toward a more generalizable computational framework for modeling mutation landscapes in adaptive immunity, while leaving important questions about prospective validation, assay heterogeneity and predicted-structure confidence for future work.

DECLARATIONS

Authors' contributions

Conceptualization, manuscript writing, review and editing: Fan, W.
Manuscript writing, review and editing: Ullah, H.; Chen, Y.; Chen, Z.

Availability of data and materials

Not applicable.

AI and AI-assisted tools statement

During the preparation of this manuscript, the AI tool ChatGPT (version 5.5 Pro, released 2026-04-23) was used solely for language editing. The tool did not influence the study design, data collection, analysis, interpretation, or the scientific content of the work. All authors take full responsibility for the accuracy, integrity, and final content of the manuscript.

Financial support and sponsorship

None.

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2026.

REFERENCES

1. Han, R.; Zhang, Y.; Liu, X.; et al. Generalizable mutation-effect prediction across adaptive immune recognition via unified multimodal framework. *Nat. Mach. Intell.* **2026**, *8*, 913-29. DOI
2. Han, R. The official repository of UniAIR: Generalizable mutation-effect prediction across adaptive immune recognition via unified multimodal framework. GitHub repository. https://github.com/hanrthu/UniAIR. (accessed 2026-09-29).
3. Han, R. hanrthu/UniAIR: Initial Release of UniAIR v1.0.0. Zenodo, 2026. DOI
4. Lin, Z.; Akin, H.; Rao, R.; et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science.* **2023**, *379*, 1123-30. DOI PubMed
5. Bushuiev, A.; Bushuiev, R.; Kouba, P.; et al. Learning to design protein-protein interactions with enhanced generalization. *International Conference on Learning Representations, 2024.* https://github.com/anton-bushuiev/PPIformer. (accessed 2026-09-29).
6. Jankauskaite, J.; Jimenez-Garcia, B.; Dapkunas, J.; Fernandez-Recio, J.; Moal, I. H. SKEMPI 2.0: an updated benchmark of changes in protein-protein binding energy, kinetics and thermodynamics upon mutation. *Bioinformatics* **2019**, *35*, 462-9. DOI PubMed PMC

Disclaimer/Publisher's Note: All statements, opinions, and data contained in this publication are solely those of the individual author(s) and contributor(s) and do not necessarily reflect those of OAE and/or the editor(s). OAE and/or the editor(s) disclaim any responsibility for harm to persons or property resulting from the use of any ideas, methods, instructions, or products mentioned in the content.



© The Author(s) 2026. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.