Briefings in Bioinformatics, 2024, **25(3)**, bbae247
https://doi.org/10.1093/bib/bbae247
Problem Solving Protocol

Integrating Reinforcement Learning and Monte Carlo Tree Search for enhanced neoantigen vaccine design

Yicheng Lin[1,2], Jiakang Ma[1,2], Haozhe Yuan[1,2], Ziqiang Chen[1,2], Xingyu Xu[1,2], Mengping Jiang[1,2], Jialiang Zhu[1,2], Weida Meng[1,2], Wenqing Qiu[3], Yun Liu[1,2,*]

[1]MOE Key Laboratory of Metabolism and Molecular Medicine, Department of Biochemistry and Molecular Biology, School of Basic Medical Sciences and Shanghai Xuhui Central Hospital, Fudan University, 131 DongAn Road, Shanghai, 200032, China
[2]State Key Laboratory of Medical Neurobiology and MOE Frontiers Center for Brain Science, Institutes of Brain Science, Fudan University, 131 DongAn Road, Shanghai, 200032, China
[3]Shanghai Xuhui Central Hospital, 366 North Longchuan Road, Shanghai, 200231, China

*Corresponding author. E-mail: yliu39@fudan.edu.cn

Abstract

Recent advances in cancer immunotherapy have highlighted the potential of neoantigen-based vaccines. However, the design of such vaccines is hindered by the possibility of weak binding affinity between the peptides and the patient's specific human leukocyte antigen (HLA) alleles, which may not elicit a robust adaptive immune response. Triggering cross-immunity by utilizing peptide mutations that have enhanced binding affinity to target HLA molecules, while preserving their homology with the original one, can be a promising avenue for neoantigen vaccine design. In this study, we introduced UltraMutate, a novel algorithm that combines Reinforcement Learning and Monte Carlo Tree Search, which identifies peptide mutations that not only exhibit enhanced binding affinities to target HLA molecules but also retains a high degree of homology with the original neoantigen. UltraMutate outperformed existing state-of-the-art methods in identifying affinity-enhancing mutations in an independent test set consisting of 3660 peptide–HLA pairs. UltraMutate further showed its applicability in the design of peptide vaccines for Human Papillomavirus and Human Cytomegalovirus, demonstrating its potential as a promising tool in the advancement of personalized immunotherapy.

Keywords: human leukocyte antigen; major histocompatibility complex; Reinforcement Learning; Monte Carlo Tree Search; neoantigen vaccine design

Introduction

The major histocompatibility complex (MHC) molecules, commonly referred to as human leukocyte antigen (HLA), present processed peptides on the cell surface for T cell recognition [1]. MHC class I molecules, primarily encoded by *HLA-A*, *B* and *C* genes, are expressed on the surface of almost all nucleated cells. These molecules mainly bind intracellular peptides that are 8–10 amino acids in length and present them to CD8+ T cells [2–4]. For eliciting a robust adaptive immune response, antigenic peptides must be presented at the surface of cells as peptide–HLA (pHLA) complexes for T cell recognition [5].

Tumor cells often harbor somatic mutations or virally introduced genetic sequences that are not present in healthy cells. A fraction of these aberrant sequences produce non-self-peptides termed neoantigens [6]. These can be presented by MHC class I molecules, leading toward CD8+ T cells-mediated cytotoxic tumor destruction. As neoantigens are unique to

Yicheng Lin is a PhD student at School of Basic Medical Sciences, Fudan University, China. He works on both the theory of deep learning, reinforcement learning and their applications in various biological questions.
Jiakang Ma is a PhD student at School of Basic Medical Sciences, Fudan University, China. His work focuses on the mechanisms of the MHC pathway during immunotherapy of cancers.
Haozhe Yuan is a master student at School of Basic Medical Sciences, Fudan University, China. He focuses on the applications of deep learning and machine learning on large-scale genomic data.
Ziqiang Chen is a PhD student at School of Basic Medical Sciences, Fudan University, China. He works on the applications of deep learning on medical images.
Xingyu Xu is a PhD student at School of Basic Medical Sciences, Fudan University, China. His primary focus is the bioinformatic analyses of large-scale genomic data.
Mengping Jiang is a read Resolution master student at School of Basic Medical Sciences, Fudan University, China. Her work focuses on the epigenetic regulation of MHC molecules.
Jialiang Zhu is a master student at School of Basic Medical Sciences, Fudan University, China. His work focuses on the epigenetic regulation of MHC molecules.
Weida Meng is a postdoctoral fellow at School of Basic Medical Sciences, Fudan University, China. He specializes in the fields of molecular mechanism of epigenetic regulation.
Wenqing Qiu is an associate professor at Shanghai Xuhui Central Hospital, Fudan University, China. Her major research interests include genetic and epigenetic regulation of MHC pathway during tumor immune escape.
Yun Liu is a professor at School of Basic Medical Sciences, Fudan University, China. More information can be found at the website of his group: http://Liulab.fudan.edu.cn.

Received: March 4, 2024. **Revised:** April 26, 2024. **Accepted:** May 7, 2024

tumor cells [7], they are promising candidates for therapeutic cancer vaccines [8]. Recent clinical trials have underscored the feasibility and safety of personalized neoantigen vaccines in various cancers [9–11]. Yet, the high polymorphism of MHC molecules and the variability in binding affinity of neoantigens to different MHC alleles complicate the process of peptide vaccine design [12].

In the past few years, numerous *in silico* screening pipelines have been developed to identify the optimal peptide for peptide vaccine design. Among various criteria evaluated, most existing approaches involve searching for peptides with high binding affinity to specific HLA alleles using *in silico* binding affinity prediction tools [13–19]. A similar strategy has also been applied in the context of tumor neoantigen design [9, 10, 20–23]. However, the process of individually searching through numerous peptides using these tools is inefficient. Moreover, these methods predominantly rank neoantigen candidates based on their binding affinity to specific HLA alleles of patients, potentially overlooking candidates with lesser binding affinity but higher prevalence. Only a minor fraction of top-ranked antigens have proven their immunogenicity in experimental setups [24–26]. Triggering cross-immunity by utilizing peptide mutations that have enhanced binding affinity for target HLA molecules, while preserving their homology with the original antigen, can be a promising avenue for neoantigen vaccine design.

For an 8-mer antigen peptide, with 20 potential amino acids for each position, it can, in theory, yield 20^8 candidate mutations, posing a significant challenge in identifying potentially effective mutations. Hence, developing an efficient approach to identifying affinity-boosting peptide mutations is of paramount importance. Recent advancements in this field include the automatically optimized mutated peptides (AOMP) program [27], published in 2022, which was based on TransPHLA [27], a transformer-based model for predicting pHLA binding affinity. However, AOMP utilizes rank-based heuristic approach, which interprets the attention matrices from TransPHLA [27] as an indicator of the significance of pHLA interactions. This can be misleading [28, 29] and may thus result in its imperfect performance.

Reinforcement Learning (RL) trains agents to optimize long-term rewards through their interactions with intricate environments, and it has found applications in various fields, including robotics [30], game playing [31], finance [32], mathematics [33] and beyond. Monte Carlo Tree Search (MCTS) [34, 35] is a heuristic search algorithm used for decision-making. Their combination has propelled several achievements, including in game playing like AlphaGo [36], AlphaGo Zero [37] and Muzero [38] and in protein engineering [39], highlighting their effectiveness in navigating complex scenarios. In this study, leveraging the synergies of RL and MCTS, we developed an algorithm that identifies peptide mutations with enhanced binding affinity to target HLA molecules while retaining high homology. Unlike available approaches such as AOMP, which utilizes the attention matrices for rank-based heuristic methods, we postulate that the inherent pattern within the attention matrices holds vital information for mutation selection. By integrating supervised training based on AOMP's mutation strategies with unsupervised RL and MCTS, we evaluated mutation results in an independent test set comprising 3660 pHLA pairs. By applying UltraMutate to two candidate peptides derived from human papillomavirus (HPV) and human cytomegalovirus (HCMV), we demonstrated the versatile applicability of our methodology in peptide vaccine design.

Methods

We proposed a novel method called UltraMutate, which consists of supervised learning of the SL policy network, reinforcement learning of the RL policy network and MCTS to identify promising affinity-enhancing peptide mutations for a given non-binding pHLA pair.

Supervised learning using the AOMP dataset

To identify affinity-enhancing peptide mutations for a given pHLA pair, we could theoretically utilize RL algorithms from scratch, without relying on prior expert knowledge. However, this would require extensive exploration within the entire mutation action space. Thus, our initial step involved using a supervised learning (SL) policy network to learn the mutation strategy of AOMP [27]. We employed the AOMP algorithm to mutate peptides that did not bind to their corresponding HLA molecules within TransPHLA's pHLA dataset [27]. This yielded a dataset of approximately 1.8 million pHLA pairs with around 2.7 million AOMP mutations (the AOMP dataset). The AOMP dataset was subsequently divided into training, validation and test sets in an 8:1:1 ratio.

Specifically, three distinct attention matrices for a given pHLA pair were extracted through TransPHLA's forward pass (Fig. 1A): (1) the pHLA attention score matrix, obtained from the forward pass of TransPHLA with the specific pHLA pair as input; (2) the accumulative attention score matrix, representing the HLA molecule's aggregate scores for all peptides matching the length of the target peptide; and (3) the contribution ratio matrix, representing the relative contribution of each amino acid site compared to all possible peptide amino acid sites. Instead of ranking segments in these attention matrices as was done in AOMP, we use neural networks to extract useful information from these attention matrices, along with the embedded matrix of sequences for peptide and HLA (Fig. 1A).

The SL policy network parameterized by σ (denoted as p_σ) consists of four heads, each processing one of four types of observed contents, collectively referred to as state s. A softmax layer at the end provides a conditional probability distribution for all potential mutation actions, which is a 300-dimensional vector designed to represent the mutation probabilities for 20 amino acids across a 15-mer peptide.

We employed stochastic gradient ascent to maximize the likelihood of the mutation actions a that AOMP selected in state s.

$$\Delta\sigma \propto \frac{\partial \log p_\sigma\ (a|s)}{\partial\sigma}$$

Reinforcement learning of the policy network

The second step in the process is aimed to improve the policy network by policy gradient RL. This process yields the RL policy network, which has the same architecture and initial parameters as the SL policy network. In each iteration, the RL policy network selects mutation actions for randomly sampled pHLA pairs. Each mutation action is assessed using a reward function. The design of reward function takes two factors into account: (1) the extent of change in binding probability brought by a mutation (obtained through TransPHLA [27]) and (2) the degree to which the mutation alters the peptide's homology (Fig. 1). Based on these, the reward function (r) was formulated as:

$$r = -r_f + \ln\ A + S$$
$$S = \begin{cases} 10 & \textbf{if } A \geq 0.5 \\ 0 & \textbf{if } A < 0.5 \end{cases}$$

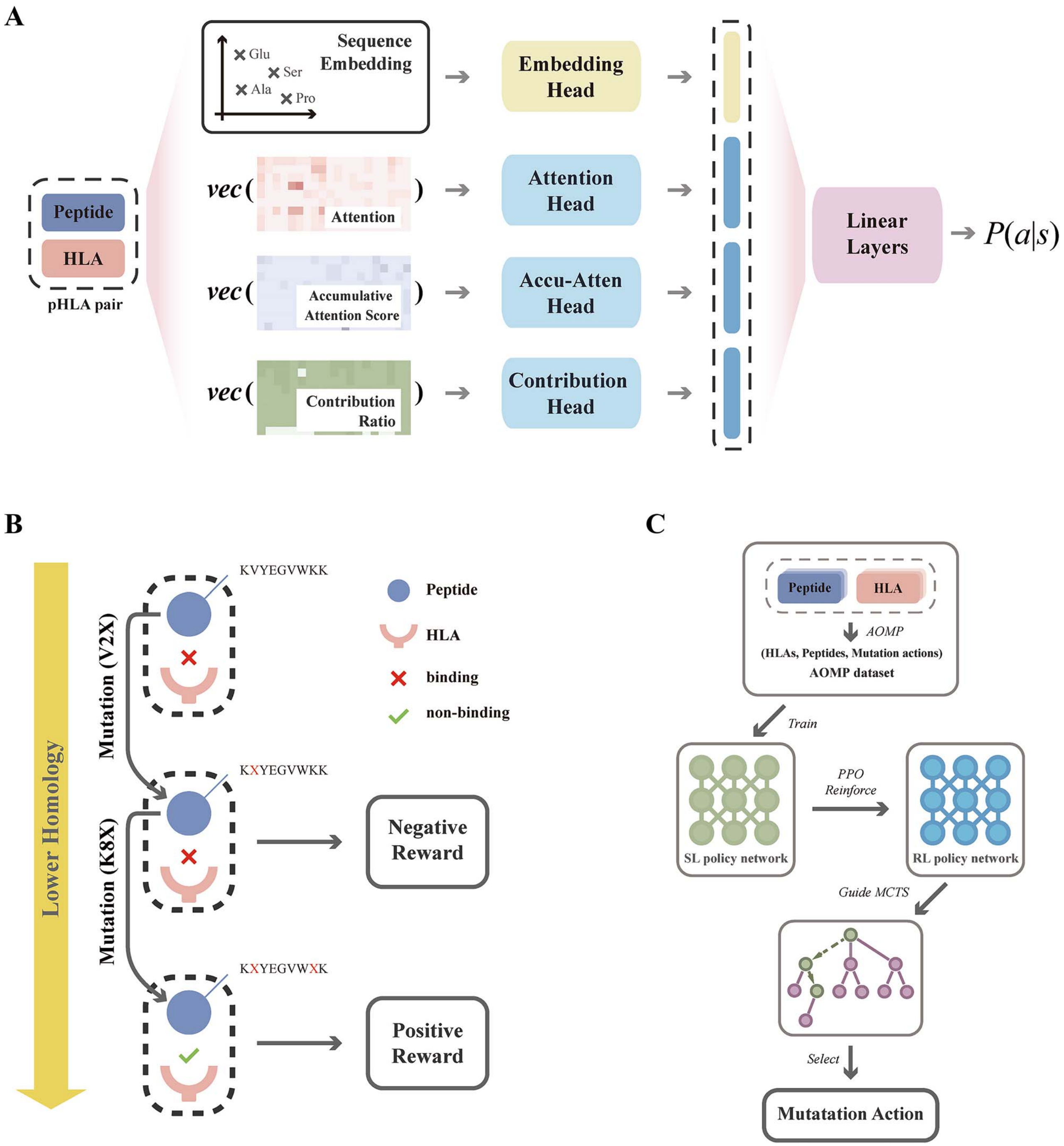


Figure 1. Overview of UltraMutate's architecture and workflow. (**A**) For a given pHLA pair, the SL policy network takes embedded sequences and flattened vectors from three distinct attention matrices as input. It uses a linear layer to combine features from four input heads to generate a conditional probability distribution $P(s|a)$ for all possible mutation actions. (**B**) The interactive environment for the agent, where rewards are given based on the binding probability of the mutated peptide with the HLA molecule and the homology between the mutated and original peptides. (**C**) UltraMutate's overall workflow includes: (1) supervised learning with the AOMP dataset; (2) fine-tuning the policy network using RL with PPO; and (3) integrating with an MCTS algorithm to determine the final mutation action.

Here, the term $-r_f$is a fixed negative value of -1 to encourage the model to identify the desired mutated peptide with minimal mutation actions. The variable A is the binding probability for the current pHLA pair as predicted by TransPHLA. The variable S is a reward that takes the constant value of 10 when the mutated peptide's binding probability predicted by TransPHLA meets a predefined threshold of 0.5, in line with TransPHLA's default demarcation between peptides that bind and those that don't.

We implemented Proximal Policy Optimization (PPO) [40] as the RL algorithm to train the RL policy network (denoted asp_θ), tuned from the trained SL policy network. The RL policy network's parameter, θ, is updated using stochastic gradient descent to maximize the objective L.

$$L\left(s,a,\theta_k,\theta\right) = \min\left(\frac{p_\theta\left(a|s\right)}{p_{\theta_k}\left(a|s\right)} A^{p_{\theta_k}}\left(s,a\right), g\left(\epsilon, A^{p_{\theta_k}}\left(s,a\right)\right)\right),$$

where

$$g(\epsilon, A) = \begin{cases} (1+\epsilon) A & A \geq 0 \\ (1-\epsilon) A & A \leq 0 \end{cases}$$

The advantage $A^{p_\theta}(a|s)$ is the difference between the Q-value $Q^{p_\theta}(s,a)$, the expected return of selection action a in state s and the value $V^{\varphi_k}(s)$, the predicted return of state s by the critic network parameterized by φ in the k-th iteration. ϵ is a hyperparameter that we set at 0.2, as used in the previous paper [40].

MCTS

After fine-tuning the RL policy network through 15 000 mini-batches of training, we combine it with a modified MCTS [34, 35] algorithm to select a mutation by lookahead search. Starting from the root state, which represents the original non-binding peptide–HLA pair, the search tree is expanded through successive mutation actions selected from a given state. In this search tree, each node representing state s contains edges (s,a) for all legal mutation actions a. Each edge stores a set of statistics,

$$\{P(s,a), N(s,a), R(s,a), U(s,a), u(s,a)\}$$

where $P(s,a)$ is the prior probability given by the forward pass of the RL policy network and $N(s,a)$is the total number of visits to the edge (s,a). The rollout value, $R(s,a)$, is defined by $R(s,a) = k\,\mathbf{ln}(H)+\beta$, where H is the homology score of the mutated peptide in the terminal state s_T given a sequence of actions $a_{t...T}$ sampled from the RL policy network $p_\theta$at time steps from t to T (the terminal time step). We set constants k and β at 5 and 2, respectively. The discounted rollout value for an edge is denoted as $U(s,a)$, calculated by $U(s,a) = \frac{1}{N(s,a)}R(s,a)$. The term $u(s,a)$ is defined as the following using a variant of the PUCT algorithm [41]:

$$u(s,a) = c_{puct}P(s,a)\frac{\sqrt{\sum N(s,\cdot)}}{1+N(s,a)}$$

At each of these time steps t, an action a_t is selected to maximize the sum of $U(s_t,a)$ and $u(s_t,a)$, formalized as:

$$a_t = \underset{a}{\mathbf{argmax}}\,(U(s_t,a)+u(s_t,a))$$

The remaining stages of the searching process, specifically expansion, evaluation and backpropagation, conform to the established MCTS method [34, 35]. Deploying this variant of MCTS on each pHLA pair, with a search depth of 10 and 60 simulation iterations, produces the findings detailed in the Results section.

The overall workflow of the mutation algorithm integrating the above policy networks and MCTS is illustrated in Fig. 1C.

Structure prediction and molecular docking

AlphaFold2 [42] was utilized to predict the three-dimensional structures of the peptide and HLA molecule based on their sequence information. Following the prediction, molecular docking was performed using the ClusPro server [43–46], which enabled us to analyze the bonding interactions within the pHLA complex under consideration. The docking results generated by ClusPro were then visualized using PyMOL [47], allowing for a detailed analysis of the interactions between the amino acids of the peptide and HLA molecule.

Peptide competitive binding assay for HLA-A molecules

The competitive peptide binding assay was performed as previously described [48] with minor modifications. Briefly, HMy2.CIR cell lines expressing HLA-A*11:01 [49] were collected and then washed with ice-cold elution buffer (0.131-M citric acid, 0.061-M Na_2HPO_4, pH 3.3 and filtered through a 0.22-μm filter) for 1 min, which was followed by immediate neutralization with ice-cold IMDM-2 medium supplemented with 0.5% bovine serum albumin (BSA). After centrifugation, the cells were resuspended to a density of 1×10^6 cells/ml in IMDM-2 medium with 1 μg/ml β_2M. Subsequently, 25 μl of either an unlabeled competitor original or a mutated peptide (5 or 15 μM) were added to the wells of a 96-well U-bottom plate, together with 25 μl of an fluorescein isothiocyanate (FITC)-labeled reference peptide (KVFPKALINK) at 300 nM. The plate was then incubated for 24 h at 4°C in darkness. Control wells lacking competitor test peptides (the negative control) and those with an unlabeled positive control peptide (QVPLRPMTYK) were both included.

After incubation, cells were centrifuged at 600 rpm for 5 min at room temperature and then washed twice with 100 μl of ice-cold PBS containing 0.5% BSA. The cell pellets were then resuspended in 150 μl of PBS for flow cytometry analysis. The competitive inhibition percentage at 5 μM (5-μM inhibition %) or 15 μM (15-μM inhibition %) was calculated using the following formula: [1 − (sample % − background %)/(max % − background %)] × 100% [49]. The term sample % refers to the percentage of FITC-positive cells in the experimental well; the term background % is the percentage of FITC-positive cells in the no-peptide control well and the term max % is the percentage of FITC-positive cells in wells without competitor test peptides. The IC50 value indicates the concentration of unlabeled competitor peptide needed to inhibit 50% of the FITC-labeled reference peptide's binding, with an IC50 < 5 μM (5-μM inhibition >50%) indicating a high binding affinity, an IC50 between 5 and 15 μM (5-μM inhibition <50% and 15-μM inhibition >50%) indicating an intermediate binding affinity and an IC50 > 15 μM indicating low or no binding affinity (5-μM inhibition of 20–50% or 15-μM inhibition of 30–50% means a low binding affinity; 5-μM inhibition of <20% or 15-μM inhibition of <30% means no binding affinity).

Results

Evaluation of UltraMutate with other state-of-the-art methods

To assess UltraMutate's efficacy in identifying peptide mutations with enhanced binding affinity while minimizing the loss of homology with the original one, we conducted a comparative analysis with the AOMP algorithm and the RL-based approach, PepPPO [50]. We evaluated an independent dataset of 3660 negative pHLA pairs, covering all 366 different HLA–peptide length combinations (peptide lengths varying from 8 to 14) available in TransPHLA's independent test set [27]. We employed this peptide selection process to ensure a thorough examination of both methods across a spectrum of various pHLA combinations, which was originally introduced in the evaluation of AOMP [27]. UltraMutate, AOMP and PepPPO were utilized in parallel to identify potential mutations capable of increasing the peptide's binding affinity to the target HLA molecules. The binding affinity of the original and mutated peptides was assessed using TransPHLA. Mutations were deemed successful if they resulted in a binding probability surpassing TransPHLA's default threshold of 0.5 in less than

Table 1. Comparison of UltraMutate, AOMP and PepPPO using 3660 test samples

	UltraMutate	AOMP	PepPPO
Successful samples	3640 (99.45%)	3627 (99.10%)	2671 (72.98%)
Samples w. 1 mutation site	1855 (50.68%)	1664 (45.46%)	174 (4.75%)
Samples w. 2 mutation sites	1572 (42.95%)	1563 (42.70%)	771 (21.07%)
Samples w. 3 mutation sites	198 (5.41%)	336 (9.18%)	964 (26.32%)
Samples w. 4 mutation sites	15 (0.41%)	64 (1.75%)	762 (20.82%)
Failed samples	20 (0.55%)	33 (0.90%)	989 (27.02%)

four mutation actions. Among all tested pHLA pairs, UltraMutate achieved a success rate of 99.45%, surpassing AOMP's 99.10% and PepPPO's 72.98% (Table 1).

In 93.6% of pHLA pairs, UltraMutate identified peptide mutants with no more than two mutation sites, maintaining a homology score that exceeded 0.8. This performance surpasses AOMP's performance of 88.2%. In contrast, PepPPO identified only 25.8% of the pHLAs with two or fewer mutation sites. This discrepancy can be attributed to PepPPO's training method, which focuses on optimizing peptide–HLA binding affinity while not considering the preservation of homology. Therefore, UltraMutate provides a more effective mutation strategy for conserving the homology of the original peptide. UltraMutate does take longer computational time than AOMP and PepPPO in optimizing peptides (Supplementary Table 1), primarily due to its implementation of MCTS, which entails running searches using trained neural networks. In real-world applications, it remains efficient and practical, especially when deployed in parallel.

UltraMutate's strategic advantage over attention weights–based heuristics

We delved deeper into the factors contributing to UltraMutate's better performance compared to the rank-based heuristic approach of AOMP. Unlike UltraMutate, AOMP relies on attention weights to focus on specific input segments during mutation processes. However, interpreting these attention weights as direct contributions can lead to inaccuracies or potential misinterpretations [28, 29]. Hence, a heuristic algorithm derived from such interpretations may not always yield the best results. To evaluate UltraMutate's effectiveness over such heuristic strategies, we examined three pHLA pairs as case studies.

For the pairing of the HLA allele HLA-C*07:01 with the peptide RYEDPDAPL, derived from Human gammaherpesvirus 4, the heuristic AOMP method prioritized the mutation sites of leucine at position 9 (9L), tyrosine at position 2 (2Y) and arginine at position 1 (1R) as the first mutation sites, guided by the high accumulative attention scores and contribution ratios corresponding to these peptide positions with the HLA molecule (Fig. 2). AOMP's selection on mutation sites was influenced by two main factors: the accumulative attention score from all peptides with a length of nine amino acids and the contribution ratio from the specific attention matrix for the pHLA pair under study. After the first step of mutation, AOMP identified some more mutation sites in order to reach the required binding affinity level, resulting in the final peptide that met the requirements within two mutation sites in total.

UltraMutate, however, identified the optimal mutation action, E3Y, on an alternative mutation site (3E) that did not rank highest in either accumulative attention score or contribution ratio (Fig. 2). This strategic selection by UltraMutate highlights its capability to surpass heuristic methods like AOMP in pinpointing optimal mutation sites, thus optimizing the peptide's binding affinity without the need for extra mutations. This instance is not unique but is one of many where UltraMutate consistently outperformed the attention weights–based strategy of AOMP (Supplementary Figures 1 and 2). These examples provide evidence of UltraMutate's superior ability in identifying optimal mutation actions by leveraging the inherent pattern within the attention matrices.

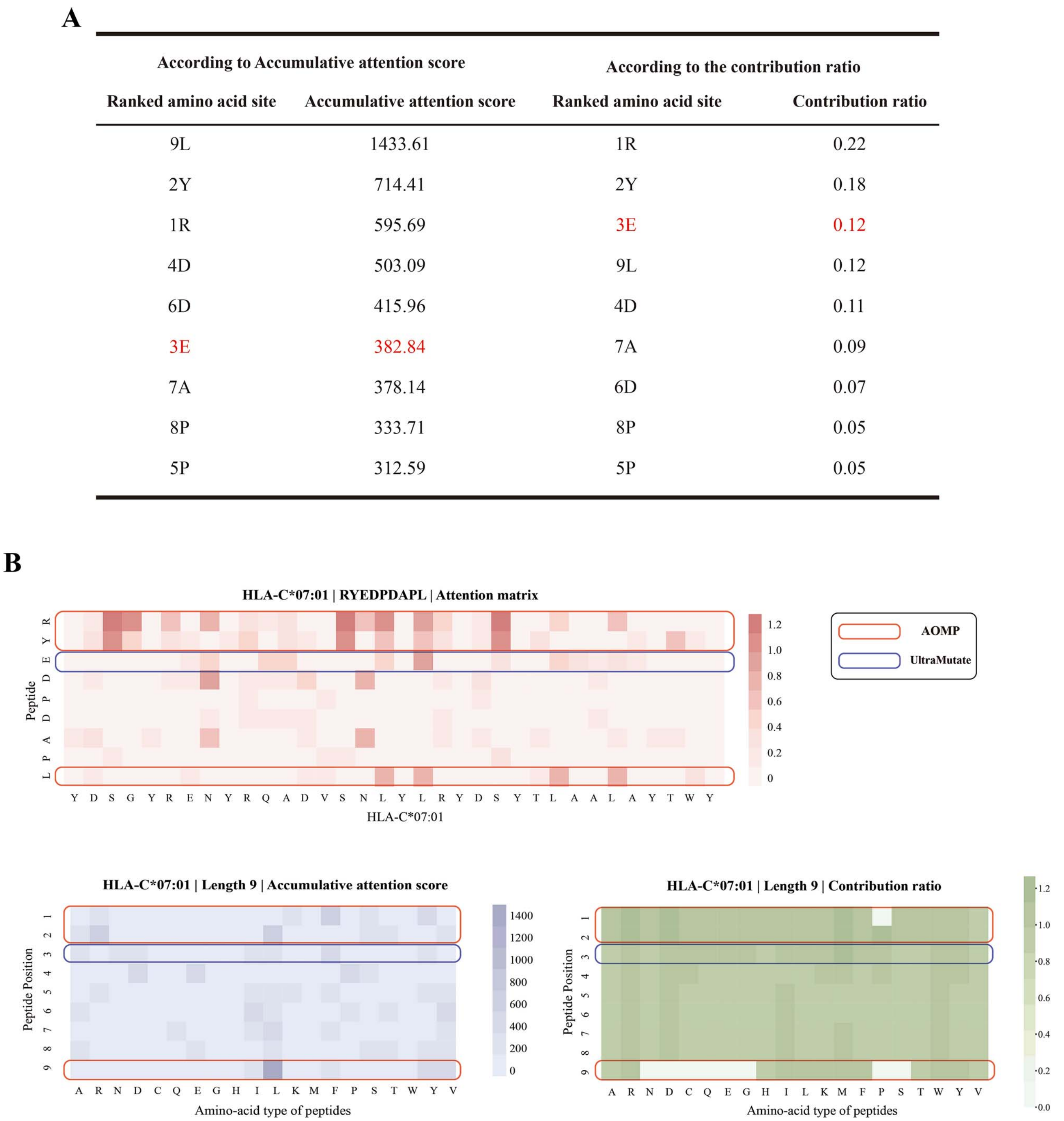
Application of UltraMutate in HPV and HCMV peptide vaccines

UltraMutate's efficacy was further evaluated in the context of potential peptide vaccine design. Human papillomavirus (HPV) is the most prevalent viral infection of the reproductive tract, with 99.7% of cervical cancers attributable to the virus [51]. Extensive clinical trials have provided compelling evidence that HPV vaccination can significantly reduce the incidence of cervical cancer [52]. In particular, HPV type 52 has a notable incidence in East Asia, accounting for 6.5% of invasive cervical cancers in China and 8.7% in Japan [53]. The high prevalence of this HPV type underscores an urgent need for its effective therapeutic vaccines.

A peptide sequence, ELQRREVYK, originating from the E6 oncoprotein of HPV type 52, has been recognized as an HLA-A*11:01-restricted peptide through an *in vitro* peptide-binding assay [54]. However, the binding affinity predicted by NetMHCpan [55] for this pHLA pair is classified as weak, suggesting its suboptimal efficacy. Consistently, this pHLA pair does not elicit positive T-cell responses, as evidenced by its underperformance in the interferon-gamma enzyme-linked immunospot (IFN-γ ELISPOT) assay [54]. Considering the high prevalence of HLA-A*11:01 allele (21.143%) in the Chinese population [56], identifying mutations of the ELQRREVYK peptide with enhanced binding affinity to HLA-A*11:01 that elicit a robust cross-immune response could have significant clinical implications.

We employed UltraMutate to this task, which successfully identified a single mutation, substituting the first amino acid from E to A (E1A). This mutation yielded the peptide ALQRREVYK, which was subsequently predicted as a strong binder to HLA-A*11:01 by both NetMHCpan and TransPHLA (Table 2). The mutated peptide retained substantial similarity to the original peptide, with a homology score of 0.89. Molecular docking studies of the original (ELQRREVYK) and mutated (ALQRREVYK) peptides with HLA-A*11:01 were conducted using the ClusPro online server [43–46]. The statistics of contact residues and the structure of the binding complex visualized with PyMOL [47] further revealed enhanced binding affinity of the mutated peptide (Fig. 3A). In contrast, AOMP was unable to generate a mutation within a single site and required at least two mutation sites to produce a mutated peptide with a diminished homology score of 0.78.

To further confirm the binding affinity enhancement, we performed a competitive binding assay [48, 49] by introducing the



A

According to Accumulative attention score		According to the contribution ratio	
Ranked amino acid site	Accumulative attention score	Ranked amino acid site	Contribution ratio
9L	1433.61	1R	0.22
2Y	714.41	2Y	0.18
1R	595.69	3E	0.12
4D	503.09	9L	0.12
6D	415.96	4D	0.11
3E	382.84	7A	0.09
7A	378.14	6D	0.07
8P	333.71	8P	0.05
5P	312.59	5P	0.05

Figure 2. Comparative analysis of UltraMutate and AOMP for the pairing of HLA-C*07:01 with the peptide RYEDPDAPL. (**A**) The summary table showing a ranking of amino acid positions determined by AOMP based on the accumulative attention score and contribution ratio for the original peptide, where UltraMutate identified another mutation site (3E) that does not have the highest accumulative attention score or contribution ratio. (**B**) Graphical representations of amino acid positions in three attention matrices for the pHLA pair extracted through TransPHLA's forward pass. Mutation sites selected by AOMP were marked by the corresponding circles.

original or mutated peptides to HMy2.CIR cells expressing HLA-A*11:01 [49]. Flow cytometry analysis revealed a leftward shift in the fluorescence peak for the mutated peptide (Fig. 3B), indicating its efficient binding to the HLA-A*11:01 allele. In contrast, the original peptide showed no discernible shift (Fig. 3B). The binding affinities and IC50 values are summarized in Fig. 3C, with the mutated peptide displaying an intermediate binding affinity within the 5–15-μM range.

We applied UltraMutate to another candidate peptide, KEVNSQLSL from the IE1 protein of HCMV [57, 58], targeting the HLA-B*15:10 allele. The proposed single-site mutation of the first amino acid K to F (K1F) led to a substantial increase in binding as predicted by both NetMHCpan and TransPHLA (Table 2). The mutated peptide, FEVNSQLSL, exhibited an impressive homology score of 0.89. However, AOMP suggested a peptide with two mutation sites, resulting in a lower homology score of 0.78. The structural representations and statistics of contact residues of the binding complexes for this pHLA pair are also depicted in Supplementary Figure 3. Collectively, these findings underscore UltraMutate's efficacy in scenarios of peptide vaccine design.

Table 2. Evaluation of binding potential for peptides derived from HPV and HCMV

	Peptide sequence	HLA allele	NetMHCpan evaluation		TransPHLA evaluation	
			%Rank	Bind level	Binding probability	Binding result
Original peptide (HPV)	ELQRREVYK	HLA-A*11:01	1.864	Weak binder	0.003	Non-binding
Mutated peptide (HPV)	ALQRREVYK	HLA-A*11:01	0.408	Strong binder	0.995	Binding
Original peptide (HCMV)	KEVNSQLSL	HLA-B*15:10	1.002	Weak binder	0.000	Non-binding
Mutated peptide (HCMV)	FEVNSQLSL	HLA-B*15:10	0.269	Strong binder	0.993	Binding

%Rank: rank of the predicted binding score compared to a set of random natural peptides introduced by NetMHCpan. Peptides will be identified as weak binder if the %Rank is above the threshold of the strong binders (by default, 0.5%) but below the threshold of the weak binders (by default, 2%). Binding probability: the binding probability predicted by TransPHLA. Peptides will be identified as non-binding to the HLA allele if the binding probability is below the default threshold 0.5.

Discussion

In recent years, several *in silico* screening pipelines have emerged to predict the binding affinity between the pHLA pair, which greatly facilitates the design of peptide vaccines. Notable tools, including NetMHCPan and NetMHCIIPan, [55] have been employed in the peptide selection processes for developing vaccines against pathogens such as *Staphylococcus saprophyticus* [19], MERS-CoV [14] and SARS-CoV-2 [13, 16, 17]. Additionally, HLAPred was utilized in the design of a peptide vaccine against *Dermatophagoides pteronyssinus* [18]. While these existing methods employ *in silico* binding affinity prediction tools to screen through numerous candidate antigens, the process of individually screening for peptides with a high binding affinity for specific HLA alleles remains inefficient. Moreover, the ranking of candidate peptides is predominantly based on their binding affinity, which may neglect candidate pHLA pairs with a higher prevalence of HLA molecules. Exploring the potential of triggering cross-immunity by screening peptide mutations with enhanced binding affinity to target HLA molecules, while preserving homology with the original one, offers another option, which can be easily achieved with UltraMutate. Through the integration of UltraMutate into the existing pipeline of vaccine development, we envision that it can not only enhance the efficiency of vaccine candidate selection but also streamline the path toward clinical translation by prioritizing peptides with greater therapeutic potential.

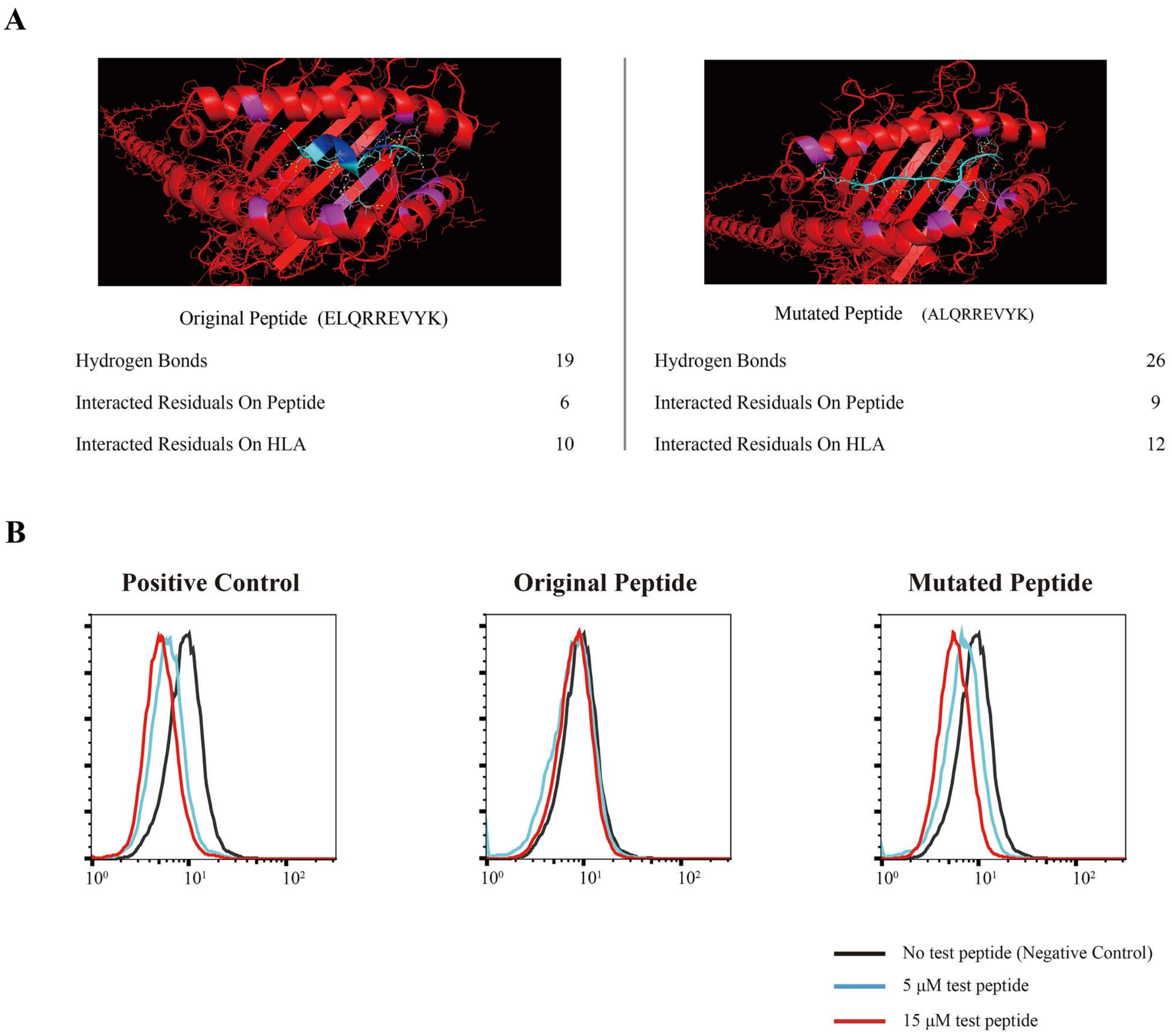
By comparing with the rank-based heuristic approach of AOMP, we revealed potential reasons leading to the better performance of UltraMutate. AOMP relies on attention weights to identify input segments for mutation actions. In the optimization of a specific pHLA pair, AOMP analyzed various attention matrices to prioritize mutation actions based on their attention weights. This approach was based on the belief that the values of attention weights represent the rankings of mutation actions. However, simply interpreting these attention weights as direct contributions can be misleading [28, 29]. Hence, a heuristic algorithm derived from such interpretations may not always achieve optimal results.

Another recently published work PepPPO shares similarities with UltraMutate in utilizing RL to identify mutation actions that enhance peptide binding affinity for a given pHLA pair. However, we've demonstrated that UltraMutate significantly outperformed PepPPO in the independent test set. The difference in performance can be attributed to several factors. First, UltraMutate leverages valuable information from TransPHLA's attention matrices. In contrast, PepPPO is trained from scratch without prior knowledge. Second, the reward function in PepPPO lacks the consideration for maintaining the homology of mutated peptides, which greatly impacts its ability in identifying peptide mutations with high homology scores. Finally, unlike UltraMutate, PepPPO does not utilize MCTS to enhance its mutation action policy, which may limit its effectiveness in exploring and exploiting the vast mutation landscape.

One important aspect to consider is the ability of UltraMutate to generalize across a wide range of HLA alleles. The training dataset of UltraMutate, derived from TransPHLA, is meticulously curated from prominent databases including IEDB [59], EPIMHC [60], MHCBN [61] and SYFPEITHI [62]. This dataset comprehensively encompasses 112 common HLA class I (40 HLA-A, 56 HLA-B and 16 HLA-C) allele types, effectively covering a substantial portion of the HLA diversity in population. In practical application, an individual typically possesses a total of six alleles across the *HLA-A*, *HLA-B* and *HLA-C* genes. Given the prevalence of these 112 *HLA* alleles within the population, it is highly probable that an individual will possess at least one of these alleles included in the dataset. Thus, we believe that UltraMutate can be generally used to identify mutated peptides capable of binding to a patient's specific HLA alleles.

The strategy of integrating RL with MCTS within the UltraMutate algorithmic framework shows significant potential, likely beyond the demonstrated scenario of peptide–HLA binding enhancement. The algorithm's adeptness at deciphering complex patterns makes it a versatile tool for tasks requiring a careful balance between biological function and molecular diversity. The efficacy of this algorithmic framework has been demonstrated in other works, notably a recent one that utilized integrated RL and MCTS for protein engineering [39]. By modulating the protein structure properties obtained from AlphaFold2, Wang *et al.* [39] illustrated the application of such strategy to the generation of proteins with desired properties. We envision that the strategic algorithmic foundation of UltraMutate will lead to advancements in a variety of fields, including drug discovery, genetic research and beyond.

UltraMutate's capabilities, while advanced, also come with certain limitations that offer avenues for future improvement. The first limitation is the absence of structural information in UltraMutate's input. It relies on sequences of peptides and HLA molecules without considering the three-dimensional configurations that are critical in determining the pHLA interactions. While prevailing binding prediction models predominantly focus on sequence information due to the inherent brevity of peptides binding to MHC, there exists potential benefit in incorporating structural data into the model. Several algorithms [63–65] have successfully incorporated structural information for protein design purposes. To address this, incorporating molecular dynamics simulations or structural modeling into UltraMutate could provide a more nuanced understanding of how mutations may affect the physical interactions between peptides and HLA molecules. Given that UltraMutate employs TransPHLA



C

HLA-A	Peptides	Peptide Sequence	5 μM Inhibition (%)	15 μM Inhibition (%)	Binding Affinity
A*11:01	Positive Control	QVPLRPMTYK	68.72	83.61	high
	Original	ELQRREVYK	24.21	19.47	no
	Mutated	ALQRREVYK	44.83	77.55	intermedium

Figure 3. Binding evaluation of the pHLA complexes for HPV peptides. (A) Visualization of the pHLA complexes for the original (ELQRREVYK) and mutated (ALQRREVYK) peptides using PyMOL, with the statistics of the bonding interactions at the bottom. (**B**) Flow cytometry results show a leftward shift in the fluorescence intensity for the mutated peptide (ALQRREVYK), indicating its enhanced binding affinity to the HLA-A*11:01 allele in contrast to the original peptide (ELQRREVYK), which does not exhibit a discernible shift. The negative control is the experiment without competitor peptides, and the positive control is the one with an unlabeled positive control peptide. (**C**) Summarized binding affinity and competitive inhibition percentages of peptides with the HLA-A*11:01 allele.

to evaluate the binding affinity of pHLA pairs, and TransPHLA performs well without structural data, the lack of this information may have a minimal impact on UltraMutate's capabilities. Nonetheless, such integrative approaches may further refine the predictions, making them more robust and applicable to real-world scenarios.

Another limitation pertains to UltraMutate's assessment of peptide homology solely based on sequence similarity, which may not always translate to biological equivalence. Furthermore, UltraMutate focuses on HLA binding affinity without direct consideration of whether the resultant pHLA pair will activate T cells and elicit a cross-immune response. To bridge this gap, future iterations of UltraMutate could integrate sequence analysis with structural epitope mapping, potentially employing algorithms that consider the interactions between peptide–HLA pairs and T cell receptors to predict the immunogenicity of peptides. Integration of T cell receptor modeling and epitope prediction tools into the UltraMutate framework could provide a more comprehensive assessment of the peptide's ability to initiate an adaptive immune response, moving closer to a biologically relevant measure of peptide vaccines.

Conclusion

In this work, we present UltraMutate, an algorithm that leverages the power of RL and MCTS to identify peptide mutations. UltraMutate not only enhances peptide's binding affinity to target HLA molecules but also maintains a high degree of homology with the original peptide. Our results demonstrate that UltraMutate outperforms existing heuristic approaches like AOMP and RL-based approaches such as PepPPO in identifying mutations that may enhance the immunogenic potential of neoantigens. Using HPV and HCMV peptides as examples, UltraMutate demonstrated its efficacy in the development of peptide vaccines.

The integration of RL with MCTS equips UltraMutate with a robust mechanism for navigating the vast mutation landscape more effectively than other state-of-art competitors. The current version can be seamlessly integrated with the existing pipelines for peptide vaccine development and is a stepping stone toward a comprehensive peptide vaccine design tool that is designed to potentially elicit an immune response. This could represent a significant advancement in the field of neoantigen-based vaccine design and a promising step toward personalized cancer immunotherapy.

Key Points

- In the pursuit of effective neoantigen-based vaccines, we introduced UltraMutate, an innovative algorithm that leverages Reinforcement Learning (RL) and Monte Carlo Tree Search (MCTS), aiming to identify peptide mutations that not only demonstrate improved binding affinities to target HLA molecules but also maintain a significant degree of homology with the original peptide.
- UltraMutate utilized the attention information generated by TransPHLA to identify affinity-enhancing peptide mutations. Instead of the existing rank-based heuristic method like AOMP, UltraMutate uses a neural network to output the probability distribution of mutation actions and combines the policy networks with a modified MCTS to ensure its performance.
- Demonstrating superior performance on a held-out test set, UltraMutate validates its efficacy compared to existing methods such as AOMP and another RL approach, PepPPO, which lacks consideration for maintaining homology and does not involve a searching mechanism.
- UltraMutate showed its applicability in the design of peptide vaccines for Human Papillomavirus and Human Cytomegalovirus, demonstrating its potential as a promising tool in the advancement of personalized immunotherapy.

Supplementary data

Supplementary data is available at *Briefings in Bioinformatics* online.

Funding

This work was supported by funding from the National Natural Science Foundation of China (82171837, 32300484), the Science and Technology Commission of Shanghai Municipality (23JS1400400), Shanghai Municipal Science and Technology Major Project (2017SHZDZX01 and 2018SHZDZX01) and ZJLab.

Data availability

The TransPHLA's pHLA dataset was downloaded at: https://github.com/a96123155/TransPHLA-AOMP/tree/master/Dataset. All the codes used in this work are available at: https://github.com/YichengLin-lab/ultramutate.

References

1. Trowsdale J, Knight JC. Major histocompatibility complex genomics and human disease. *Annu Rev Genomics Hum Genet* 2013;**14**:301–23.
2. Bouvier M, Wiley DC. Importance of peptide amino and carboxyl termini to the stability of MHC class I molecules. *Science* 1994;**265**:398–402.
3. Zacharias M, Springer S. Conformational flexibility of the MHC class I alpha1-alpha2 domain in peptide bound and free states: a molecular dynamics simulation study. *Biophys J* 2004;**87**: 2203–14.
4. Babbitt BP, Allen PM, Matsueda G, *et al.* Binding of immunogenic peptides to Ia histocompatibility molecules. *Nature* 1985;**317**: 359–61.
5. Rossjohn J, Gras S, Miles JJ, *et al.* T cell antigen receptor recognition of antigen-presenting molecules. *Annu Rev Immunol* 2015;**33**: 169–200.
6. Smith CC, Selitsky SR, Chai S, *et al.* Alternative tumour-specific antigens. *Nat Rev Cancer* 2019;**19**:465–78.
7. Schumacher TN, Scheper W, Kvistborg P. Cancer neoantigens. *Annu Rev Immunol* 2019;**37**:173–200.
8. Lang F, Schrors B, Lower M, *et al.* Identification of neoantigens for individualized therapeutic cancer vaccines. *Nat Rev Drug Discov* 2022;**21**:261–82.
9. Ott PA, Hu Z, Keskin DB, *et al.* An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* 2017;**547**: 217–21.
10. Awad MM, Govindan R, Balogh KN, *et al.* Personalized neoantigen vaccine NEO-PV-01 with chemotherapy and anti-PD-1 as first-line treatment for non-squamous non-small cell lung cancer. *Cancer Cell* 2022;**40**:1010–1026.e11.
11. Rojas LA, Sethna Z, Soares KC, *et al.* Personalized RNA neoantigen vaccines stimulate T cells in pancreatic cancer. *Nature* 2023;**618**:144–50.
12. Lybaert L, Lefever S, Fant B, *et al.* Challenges in neoantigen-directed therapeutics. *Cancer Cell* 2023;**41**:15–40.

13. Rakib A, Sami SA, Mimi NJ, *et al.* Immunoinformatics-guided design of an epitope-based vaccine against severe acute respiratory syndrome coronavirus 2 spike glycoprotein. *Comput Biol Med* 2020;**124**:103967.
14. Mahmud S, Rafi MO, Paul GK, *et al.* Designing a multi-epitope vaccine candidate to combat MERS-CoV by employing an immunoinformatics approach. *Sci Rep* 2021;**11**:15431.
15. Sami SA, Marma KKS, Mahmud S, *et al.* Designing of a multi-epitope vaccine against the structural proteins of Marburg virus exploiting the Immunoinformatics approach. *ACS Omega* 2021;**6**: 32043–71.
16. Rakib A, Sami SA, Islam MA, *et al.* Epitope-based Immunoinformatics approach on Nucleocapsid protein of severe acute respiratory syndrome-Coronavirus-2. *Molecules* 2020; **25**:5088.
17. Obaidullah AJ, Alanazi MM, Alsaif NA, *et al.* Immunoinformatics-guided design of a multi-epitope vaccine based on the structural proteins of severe acute respiratory syndrome coronavirus 2. *RSC Adv* 2021;**11**:18103–21.
18. Adji A, Niode NJ, Memah VV, *et al.* Designing an epitope vaccine against Dermatophagoides pteronyssinus: an in silico study. *Acta Trop* 2021;**222**:106028.
19. Yousaf M, Ullah A, Sarosh N, *et al.* Design of Multi-Epitope Vaccine for staphylococcus saprophyticus, : Pan-genome and reverse vaccinology approach. *Vaccines (Basel)* 2022;**10**:10.
20. Chen H, Li Z, Qiu L, *et al.* Personalized neoantigen vaccine combined with PD-1 blockade increases CD8(+) tissue-resident memory T-cell infiltration in preclinical hepatocellular carcinoma models. *J Immunother Cancer* 2022;**10**:10.
21. Peng M, Mo Y, Wang Y, *et al.* Neoantigen vaccine: an emerging tumor immunotherapy. *Mol Cancer* 2019;**18**:128.
22. Gubin MM, Zhang X, Schuster H, *et al.* Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature* 2014;**515**:577–81.
23. Sahin U, Derhovanessian E, Miller M, *et al.* Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature* 2017;**547**:222–6.
24. Wells DK, van Buuren MM, Dang KK, *et al.* Key parameters of tumor epitope immunogenicity revealed through a consortium approach improve Neoantigen prediction. *Cell* 2020;**183**:818–834 e813.
25. Yadav M, Jhunjhunwala S, Phung QT, *et al.* Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature* 2014;**515**:572–6.
26. Tran E, Ahmadzadeh M, Lu YC, *et al.* Immunogenicity of somatic mutations in human gastrointestinal cancers. *Science* 2015;**350**: 1387–90.
27. Chu Y, Zhang Y, Wang Q, *et al.* A transformer-based model to predict peptide–HLA class I binding and optimize mutated peptides for vaccine design. *Nat Mach Intell* 2022;**4**:300–11.
28. Jain S, Wallace BC. Attention is not explanation. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 1 (Long and Short Papers). Minneapolis, USA: Association for Computational Linguistics, 2019, pp. 3543–56.
29. Serrano S, Smith NA. Is attention interpretable? arXiv preprint, arXiv:1906.03731, 2019.
30. Zhang T, Mo H. Reinforcement learning for robot research: a comprehensive review and open issues. *Int J Adv Robot Syst* 2021;**18**:172988142110073.
31. Souchleris K, Sidiropoulos GK, Papakostas GA. Reinforcement learning in game industry&mdash;review, prospects and challenges. *Appl Sci* 2023;**13**:2443.
32. Malibari N, Katib I, Mehmood R. Systematic review on reinforcement learning in the field of Fintech. arXiv preprint, arXiv:2305.07466, 2023.
33. Fawzi A, Balog M, Huang A, *et al.* Discovering faster matrix multiplication algorithms with reinforcement learning. *Nature* 2022;**610**:47–53.
34. Coulom R. Efficient selectivity and backup operators in Monte-Carlo Tree Search, In: van den Herik HJ, Ciancarini P, Donkers HHLM (eds) *Computers and Games. CG 2006. Lecture Notes in Computer Science*, vol **4630**. Springer, Berlin, Heidelberg, 2006: 72–83.
35. Kocsis L, Szepesvári C. Bandit Based Monte-Carlo Planning. In: Fürnkranz J, Scheffer T, Spiliopoulou M (eds) *Machine Learning: ECML 2006.Lecture Notes in Computer Science*, vol **4212**. Springer, Berlin, Heidelberg 2006, 282–93.
36. Silver D, Huang A, Maddison CJ, *et al.* Mastering the game of Go with deep neural networks and tree search. *Nature* 2016;**529**: 484–9.
37. Silver D, Schrittwieser J, Simonyan K, *et al.* Mastering the game of Go without human knowledge. *Nature* 2017;**550**: 354–9.
38. Schrittwieser J, Antonoglou I, Hubert T, *et al.* Mastering Atari, Go, chess and shogi by planning with a learned model. *Nature* 2020;**588**:604–9.
39. Wang Y, Tang H, Huang L, *et al.* Self-play reinforcement learning guides protein engineering. *Nat Mach Intell* 2023;**5**:845–60.
40. Schulman J, Wolski F, Dhariwal P, *et al.* Proximal policy optimization algorithms. arXiv preprint, arXiv:1707.06347, 2017.
41. Rosin CD. Multi-armed bandits with episode context. *Ann Math Artif Intell* 2011;**61**:203–30.
42. Jumper J, Evans R, Pritzel A, *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;**596**: 583–9.
43. Kozakov D, Hall DR, Xia B, *et al.* The ClusPro web server for protein–protein docking. *Nat Protoc* 2017;**12**:255–78.
44. Kozakov D, Beglov D, Bohnuud T, *et al.* How good is automated protein docking? *Proteins* 2013;**81**:2159–66.
45. Vajda S, Yueh C, Beglov D, *et al.* New additions to the ClusPro server motivated by CAPRI. *Proteins* 2017;**85**:435–44.
46. Desta IT, Porter KA, Xia B, *et al.* Performance and its limits in rigid body protein-protein docking. *Structure* 2020;**28**:1071–1081 e1073.
47. Schrodinger LLC. The PyMOL Molecular Graphics System, Version 1.8. 2015.
48. Kessler JH, Benckhuijsen WE, Mutis T, *et al.* Competition-based cellular peptide binding assay for HLA class I. *Curr Protoc Immunol* 2004;**Chapter 18**:Unit 18.12.
49. Ding Y, Zhou Z, Li X, *et al.* Screening and identification of HBV epitopes restricted by multiple prevalent HLA-A allotypes. *Front Immunol* 2022;**13**:847105.
50. Chen Z, Zhang B, Guo H, *et al.* Binding peptide generation for MHC Class I proteins with deep reinforcement learning. *Bioinformatics* 2023;**39**:39.
51. Walboomers JM, Jacobs MV, Manos MM, *et al.* Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol* 1999;**189**:12–9.
52. Rahangdale L, Mungo C, O'Connor S, *et al.* Human papillomavirus vaccination and cervical cancer risk. *BMJ* 2022;**379**:e070115.
53. HPV INFORMATION CENTRE. https://hpvcentre.net/ (4 December.
54. Chan PK, Liu SJ, Cheung JL, *et al.* T-cell response to human papillomavirus type 52 L1, E6, and E7 peptides in women with

transient infection, cervical intraepithelial neoplasia, and invasive cancer. *J Med Virol* 2011;**83**:1023–30.

55. Reynisson B, Alvarez B, Paul S, *et al*. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res* 2020;**48**: W449–54.
56. Zhou X-Y, Zhu F-M, Li J-P, *et al*. High-resolution analyses of human leukocyte antigens allele and haplotype frequencies based on 169,995 volunteers from the China bone marrow donor registry program. *PLoS One* 2015;**10**:e0139485.
57. Gibson L, Piccinini G, Lilleri D, *et al*. Human cytomegalovirus proteins pp65 and immediate early protein 1 are common targets for CD8+ T cell responses in children with congenital or postnatal human cytomegalovirus infection. *J Immunol* 2004;**172**: 2256–64.
58. Bunde T, Kirchner A, Hoffmeister B, *et al*. Protection from cytomegalovirus after transplantation is correlated with immediate early 1-specific CD8 T cells. *J Exp Med* 2005;**201**:1031–6.
59. Dhanda SK, Mahajan S, Paul S, *et al*. IEDB-AR: immune epitope database-analysis resource in 2019. *Nucleic Acids Res* 2019;**47**:W502–6.
60. Reche PA, Zhang H, Glutting JP, Reinherz EL. EPIMHC: a curated database of MHC-binding peptides for customized computational vaccinology. *Bioinformatics* 2005;**21**:2140–1.
61. Lata S, Bhasin M, Raghava GP. MHCBN 4.0: a database of MHC/TAP binding peptides and T-cell epitopes. *BMC Res Notes* 2009;**2**:61.
62. Rammensee H, Bachmann J, Emmerich NP, *et al*. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* 1999;**50**:213–9.
63. Anishchenko I, Pellock SJ, Chidyausiku TM, *et al*. De novo protein design by deep network hallucination. *Nature* 2021;**600**:547–52.
64. Yang KK, Wu Z, Arnold FH. Machine-learning-guided directed evolution for protein engineering. *Nat Methods* 2019;**16**:687–94.
65. Brian H, Salvatore C, Zeming L, *et al*. A high-level programming language for generative protein design. *bioRxiv* preprint, bioRxiv:2022.12.21.521526, 2022.