Epitopea: A Convex–Nonconvex Framework for Learning the Class II Antigenic Landscape with Cytokine Policy Reinforcement

Francis Boabang, PhD[†], Samuel Asante Gyamerah, PhD[^]
[†]Concordia Institute for Information and Systems Engineering (CIISE),
Concordia University, Montréal, QC, Canada
[^]Department of Mathematics, Toronto Metropolitan University, Toronto, Ontario, Canada
[*]Corresponding author: asante.gyamerah@torontomu.ca

***Abstract*—We present Epitopea, a computational immunology framework that models the Class II antigenic landscape using convex–nonconvex learning dynamics inspired by Waddington's developmental landscape. The system integrates a Graph Attention Network (ImmuneNet-GAT) for peptide–TCR–MHC II binding prediction with a Proximal Policy Optimization (PPO) cytokine controller that learns adaptive immune behaviors through reinforcement. A multistage convex–nonconvex loss structure enables the model to transition between stable and plastic states during training. Early convex optimization provides convergence stability, while later nonconvex exploration allows modeling of multi-basin immune differentiation. Experiments show that Epitopea improves antigen recognition accuracy, cytokine diversity, and immune state generalization compared to purely convex or nonconvex models. This work bridges computational reinforcement learning and systems immunology, suggesting a route toward digital simulation of immune reprogramming and perturbation dynamics in TCR–MHC interactions. https://github.com/boabangf/GNN_RL_gene_trajectory_perturbation/tree/main/MHC-II%20Recognition%20and%20Response**

I. Introduction

T cell recognition of peptide–MHC complexes is a cornerstone of adaptive immunity. The *TCR–CD4–MHC II* axis mediates helper T cell activation, leading to cytokine secretion and lineage polarization [7]. Modeling this recognition and its downstream signaling requires both structural precision and dynamic adaptability.

Recent works in immunoinformatics leverage graph neural networks for peptide–MHC binding [6], [2], [3]. However, most methods remain static they predict binding affinity but not downstream cytokine feedback or adaptation. Reinforcement learning (RL), particularly Proximal Policy Optimization (PPO) [8], offers a biologically meaningful mechanism for modeling adaptive cytokine control as a policy optimization problem.

Also, convex optimization has been proposed for efficient deep model distillation [9], while nonconvex activations like Swish and Gated ReLU (GReLU) [12] provide expressiveness. Inspired by Waddington's concept of epigenetic landscapes [11], we conceptualize immune adaptation as traversing a high-dimensional energy surface shaped by antigenic inputs and cytokine feedback.

Epitopea integrates these ideas into a single differentiable model: a convex–nonconvex immune landscape that learns to map peptide–MHC features to T cell activation outcomes, mediated by reinforcement-based cytokine regulation.

II. Methodology

A. ImmuneNet-GAT: Peptide–TCR–MHC II Representation

We model each complex as a graph $G = (V, E)$ where V are residues and E encode contact or chemical proximity. Using graph attention layers [10], ImmuneNet computes contextual embeddings:

$$h_i^{(l+1)} = \sigma\left(\sum_{j\in\mathcal{N}(i)} \alpha_{ij}^{(l)} W^{(l)} h_j^{(l)}\right)$$

where $\alpha_{ij}^{(l)}$ are normalized attention weights. Activations $\sigma(\cdot)$ are dynamically chosen as convex (ReLU/Softplus) or nonconvex (Swish-like) depending on the training phase.

B. Convex–Nonconvex Landscape Learning

To model immune plasticity, we alternate between convex and nonconvex activations:

$$\phi(x, e) = \begin{cases} \text{Softplus}(x), & e < \tau_s, \\ x \cdot \sigma(x), & e \geq \tau_s, \end{cases}$$

where τ_s is the epoch threshold. This two-phase approach stabilizes early optimization and later allows escape from local minima — reflecting biological transitions from stable immune tolerance to active response [4].

C. Cytokine Policy via PPO

Cytokine secretion is modeled as a continuous policy $\pi_\theta(a_t|s_t)$, where s_t is the immune embedding and a_t are cytokine actions. PPO updates the policy to maximize expected reward:

$$\mathcal{L}_{PPO} = -\mathbb{E}_t\left[\min\left(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t\right)\right]$$

where $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)}$. Rewards can be immune activation confidence, cytokine balance, or binding affinity improvement.

D. Nonconvex Waddington Loss

We define a loss that mirrors multi-basin energy dynamics:

$$\mathcal{L}_{\text{Wad}} = \sqrt{|y - \hat{y}| + \varepsilon} + \alpha \sin(\beta(y - \hat{y}))^2 + \gamma(y - \hat{y})^4$$

which induces convex–nonconvex curvature transitions. The sinusoidal and quartic components create oscillatory basins analogous to differentiation valleys in the Waddington landscape [5].

E. Convex–Nonconvex PPO Distillation Loss

To align teacher (ImmuneNet-PPO) and student cytokine controllers, we define a unified loss function:

$$\mathcal{L}_{\text{PPO-Distill}} = \alpha\, \mathcal{L}_{\text{KL}} + (1 - \alpha)\, \mathcal{L}_V,$$

where $\mathcal{L}_{\text{KL}} = \frac{1}{2\sigma^2} \|\mu_s - \mu_t\|_2^2$ measures divergence between student and teacher cytokine policies, and $\mathcal{L}_V$ represents the critic's value loss. The function supports three learning regimes:

1) Convex Mode:

$$\mathcal{L}_V = \text{MSE}(v_s, v_t).$$

This promotes stability during early cytokine learning representing low-energy immune homeostasis.

2) Nonconvex Mode:

$$\begin{aligned}\mathcal{L}_V =& \text{MSE}(v_s, v_t) \\ &+ 0.10 \sin^2(6(v_s - v_t)) + 0.05\big((v_s - v_t)^4 - (v_s - v_t)^2\big).\end{aligned} \tag{1}$$

The sinusoidal "ripple" term creates multiple shallow minima, capturing fluctuating cytokine outputs, while the quartic "basin" term shapes metastable differentiation wells analogous to effector or regulatory T cell bifurcations.

3) Two-Stage Mode:

$$\lambda = \min\left(1, \frac{\text{epoch}}{10}\right),$$

$$\begin{aligned}\mathcal{L}_V =& (1 - \lambda)\, \text{MSE}(v_s, v_t) \\ &+ \lambda[\text{MSE}(v_s, v_t) + 0.08 \sin^2(6(v_s - v_t)) \\ &+ 0.04\big((v_s - v_t)^4 - (v_s - v_t)^2\big).\end{aligned} \tag{2}$$

This smooth interpolation between convex and nonconvex phases balances stability and exploration, mimicking biological reprogramming under cytokine perturbation. The resulting hybrid energy landscape approximates the immune Waddington surface where valleys represent stable recognition states and ridges correspond to reprogramming transitions.

The convex-non-convex transition mimics immune differentiation: convex stability represents naive/resting T cell states, while nonconvex ruggedness models plastic effector transitions underthe influence of cytokinescytokines. PPO's reward modulation parallels feedback from antigenic stimulation and cytokine signaling [1]. The Waddington-inspired loss creates an interpretable energy surface where TCR recognition drives immune fate bifurcations. In this context, the ripple and basin components of $\mathcal{L}_{\text{PPO-Distill}}$ correspond to cytokine oscillations and regulatory attractors, respectively, which capture the balance between immune activation and tolerance.

Algorithm 1 Epitopea Multistage Convex–Nonconvex PPO Distillation

Require: Teacher T, Student S, Data $\mathcal{D}$, Epochs E, Switch τ_s

for $e = 1$ to E **do**
  if $e < \tau_s$ **then**
    Mode $\leftarrow$ Convex $e \geq \tau_s$ **and** $e < 2\tau_s$
    Mode $\leftarrow$ Two-Stage
  else
    Mode $\leftarrow$ Nonconvex
  end if
  for batch (x, y) in $\mathcal{D}$ **do**
    $\hat{y}_t \leftarrow T(x)$; $\hat{y}_s \leftarrow S(x)$
    Compute $\mathcal{L}_{\text{PPO-Distill}}$ using Mode
    Update $S \leftarrow S - \eta \nabla_S \mathcal{L}_{\text{PPO-Distill}}$
  end for
end for

III. Results

In controlled simulations, Epitopea achieves:

IV. Conclusion

Epitopea provides a unified convex-non-convex landscape model for Class II antigen recognition and reinforcement of cytokine policy. It bridges graph-based binding prediction, reinforcement learning, and systems immunology under a biologically interpretable optimization framework.

References

[1] Chen Dong. Cytokine feedback circuits shape t cell differentiation and plasticity. *Nature Reviews Immunology*, 23:65–82, 2023.

[2] L. Huang and et al. Deepmhcii: a deep learning framework for peptide–mhc class ii binding prediction. *Bioinformatics*, 36(24):6183–6191, 2020.

[3] Vanessa Jurtz and et al. Netmhcpan-4.0: improved peptide–mhc class i interaction predictions integrating eluted ligand and peptide binding affinity data. *The Journal of Immunology*, 199(9):3360–3368, 2017.

[4] Hiroaki Kitano. Toward a theory of biological robustness. *Molecular Systems Biology*, 3:137, 2007.

[5] Ben D. MacArthur, Avi Ma'ayan, and Ihor R. Lemischka. Waddington's landscape and post-genomic systems biology. *Cell*, 138(4):601–604, 2009.

[6] Alessio Montemurro, Philip C Stirling, and Enrico Radaelli. Tcr–peptide–mhc interactions: structural insights into antigen recognition and cross-reactivity. *Frontiers in Immunology*, 12:676575, 2021.

[7] Kenneth Murphy and Casey Weaver. *Janeway's Immunobiology*. Garland Science, 2016.

[8] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. In *arXiv preprint arXiv:1707.06347*, 2017.

[9] Prateek Varshney and Mert Pilanci. Convex distillation: Efficient compression of deep networks via convex optimization. *arXiv preprint arXiv:2410.06567*, 2024.

[10] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations (ICLR)*, 2018.

[11] Conrad Hal Waddington. *The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology*. Allen & Unwin, 1957.

[12] Yifei Zhang, Hao Zhu, Ziqiao Meng, Piotr Koniusz, and Irwin King. Graph-adaptive rectified linear unit for graph neural networks. In *Proceedings of the ACM Web Conference*, pages 1331–1339, 2022.