

The dawn of biophysical representations in computational immunology

Eric Wilson^{1,†}, Akshansh Kaushik^{*2,3}, Soumya Dutta^{*2,3}, and Abhishek Singharoy^{2,3,†}

¹Department of Immunology and Immunotherapy, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

²School of Molecular Sciences, Arizona State University

^{*}Equal contribution, ordered by first name

[†]Corresponding authors

Abstract

Computational immunology has been the breeding ground of some of the best bioinformatics work of the day. By melding diverse data types, these approaches have been successful in associating genotypes with phenotypes. However, the representations (or spaces) in which these associations are mapped have primarily been constructed from some omics-oriented sequence data typically derived from high-throughput experiments. In this perspective, we highlight the importance of biophysical representations for performing the genotype-phenotype map. We contend that using biophysical representations reduces the dimensionality of a search problem, dramatically expedites the algorithm, and more importantly, offers physical interpretability to the classes of clustered sequences across different layers of complexity - molecular, cellular or macro-level. Such biophysical interpretations offer a firm basis for the future of bioengineering and cell-based therapies.

1 Introduction

The core responsibility of our immune system is to protect the body from pathogens and cancers. The need to target and activate the immune system reproducibly has been underscored by the recent pandemic and the rise of anticancer therapies that rely on immunological mechanisms. This has generated a focused enthusiasm for gaining a detailed description of the immune system. However, the human immune system is incredibly complex, and often regarded as one of the most challenging topics in biology. The sheer size of sequence and population diversity in proteins associated with the immune system presents a formidable obstacle to mapping their network of interactions within a tractable space. For instance, T-cell recognition of antigens is driven by HLA genes encoding for Major Histocompatibility Complexes (or MHCs), are among the most polymorphic germline genes in the human genome that contain tens of thousands of variants across populations[1]. Moreover, the somatic hypermutations involved in the function of T-cell and B-cell receptors make them the most polymorphic human proteins in known existence, with theoretical estimates of T-cell Receptor (or TCR) diversity reaching over 10^{61} potential sequences. A more conservative estimate places TCR diversity in the range of 10^7 receptors[2], which still offers an incredibly vast range of human variations. The desire to account for this diversity and predict its associated non-linear relationships has motivated the genesis of the field of computational immunology to develop methods to analyze and predict immune outcomes based on this data (Figure-

This peer-reviewed article has been accepted for publication but not yet copyedited or typeset, and so may be subject to change during the production process. The article is considered published and may be cited using its DOI.

10.1017/qrd.2025.7

This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.

1). Computational immunology has transformed our understanding of the immune system by enabling the integration of massive amounts of biochemical and biological data. Simple mathematical models to study disease transmission can be traced to the early 20th century[3, 4]. By leveraging population data, it clarified the relationship between the size of mosquito populations and malaria incidence which led to improved malaria control. The power of computational immunology expanded significantly in the information age with the advent of high-throughput sequencing, proteomics, and the growing availability of experimental and clinical data further empowered by advances in computational technology. These advancements have enabled computational techniques to tackle more complex immunological questions. Consequently, computational immunology has now been applied to a broad spectrum of applications in-

cluding vaccine design[5], predicting population-level mortality rates[6], and forecasting the outcomes of immune checkpoint blockade therapies[7].

Due to the availability and ease of collection of protein and amino acid sequence information, most computational immunology approaches primarily rely on sequence data for their predictions[8, 9, 10]. However, recent advances in machine learning and protein modeling have caused an explosion in the synergistic incorporation of biophysical information and modeling into existing computational immunology approaches[11, 12]. Such integrations have already shown to have a profound improvement on the accuracy of models, but also grant the ability to gain novel insights into previous intractable mechanisms, advancing computational immunology models into the next era. In the following perspective, we will explore immune-related models ranging from atomistic environments to the macro-level systems, demonstrating how biophysics can be used to enhance predictive accuracy and improve our overall understanding of immune responses.

2 A perspective on biophysical models

Computational immunology has been dominated by bioinformatics, primarily due to a push from recent findings in genomic and proteomic technologies that compose around 31 different databases today[13]. Historically, it allows the study of complex protein-protein interactions across a diversity of sequences[14]. Recently, deep learning approaches offer rapid access to molecular structures from sequences[15], which has extended the realm of bioinformatics to structure-guided models of immune interactions[16]. However, the physical formulation of intermolecular interactions is statistical, which entails an ensemble description of conformations that remains obscure in the bioinformatics approaches. These ensembles capture transition in the order-disorder transition of the molecules, flexibility, and thermal effects, as well as solvation and microenvironmental impacts on structure. Attempts to overcome such limitations of traditional computational immunology open the doors for employing biophysical tools to take MHC, TCRs, and antibody predictions beyond the sequence-only or sequence-structure paradigm[17, 18, 19]. Notwithstanding the computationally expensive biophysical simulations, it generates unique representations and metrics that connect collective molecular properties with phenotypic and even population outcomes. We break down the biophysical advances in the realm of atomistic, molecular, whole-cell, and macro-level modeling, and highlight how biophysical entities of Table 1 are acting or can be leveraged as novel representations for learning in computational immunology, as complements to the traditional sequence or structural methods.

2.1 Atomistic description

We start with biophysical descriptors in computational biology arising from detailed interactions of antibodies, MHCs, and TCRs.

2.1.1 Free energy description of antibodies

Since the first antibody structure was deposited in 1976, the number of antibody structures in the PDB has grown, and it now represents approximately 2.1% of the total entries[20]. Many computational tools now use only the antibody data, as opposed to general protein data, due to the increased performance[21, 22]. To this end, the Structural Antibody Database or SAbDab collects, curates, and presents an ensemble of antibody structures from the PDB[23]. Such databases allow for the prediction of the affinity of antibody-antigen interfaces by combining the biophysics of protein-protein interactions with deep learning approaches[24]. In fact, a significant improvement in the ranking and prediction of affinity predictions is observed by combining all-atom free energy methods like Free energy perturbation or FEP+ with focused machine learning approaches like QuanSA[25]. Using such a combination of biophysics and informatics, the affinity of CR3022 antibody is optimized to the spike protein of the SARS-CoV-2 Omicron strain, achieving a high success rate with up to a 17-fold affinity increase[26]. Going beyond simple geometric 3D coordinate representations of ligands [25], a novel metric of multiple-ligand alignment is employed using so-called pocket fields to learn affinities. Unlike the learning of real geometries that are quite high dimensional, the learning of smoother functions like the 3D fields (with known map to the SMILE or peptide sequences) offers learning across a broad diversity of molecular identity and conformation, without overfitting the loss function. In conclusion, the application of free energy augmented antibody design underscore the growing power of biophysical modeling to not only understand but also engineer biological systems for specific therapeutic outcomes.

2.1.2 Structural modeling of MHC (Major Histocompatibility Complex)

In 1968, Snell examined the concept of transplantation and came across the term histocompatibility polymorphism[27, 28]. MHC proteins play a crucial role in immune mechanisms due to their involvement in activating T cells and B cells[29, 30]. Structural modeling of these complexes offers insights into the mechanism of the several pathways relevant to immunogenicity[31]. The MHC protein is one of the most polymorphic proteins in humans[1], but despite the high polymorphism, the structure of the MHC binding groove is highly conserved[11]. Researchers found that the second and last residues are key anchors for peptide binding to the MHC class-I binding groove[29], a discovery made through X-ray diffraction studies[32]. Since countless peptides can bind to MHC, many generated by frameshift events, and lack evolutionary context for multi-sequence alignments, crystallizing all polymorphic complexes is unfeasible. A biophysical approach is thus needed to model MHC-peptide complexes for further study.

Conventionally, there are three ways to model structures: molecular dynamics, molecular docking, and homology modeling[33]. The unifying protocol to design a model for MHC is as follows: the first part is to generate a peptide conformation using a PDB template, the second step involves docking of the peptide, and finally optimizing the overall structure. Multiple sources are available to model MHC-I complexes such as DockTope, GradDock, APE-Gen, AlphaFold2, and RoseTTAfold[34, 35, 36, 37]. Although these methods are highly accurate, some of them are highly computationally heavy or applicable only to the MHC class-I molecule due to the heterodimeric binding pocket observed in MHC class-II molecules. Recently, a state-of-the-art method, *PANDORA*, shows potential to design even MHC class-II molecules, and also offers some tunability while modeling. Its energy-based definition of loop conformations is shown to outperform most of the methods previously introduced in terms of accuracy and computational efficiency[38]. However, there still is a need for a tool that models complex structures by capturing the biophysical attributes of the peptide-MHC instead of exploiting sequence similarity and templates. Large datasets to benchmark biophysical properties across a range of MHC systems - similar to MISATO (for MD simulations of 20,000 protein-ligand systems) or 100-protein NMR spectra (for protein dynamics) - do not yet exist in this space. A very promising result is that semi-empirical quantum mechanical representations can now be embedded in these data sets to refine the associated protein structures. Once similar datasets start existing for the broad class of MHC proteins, such quantum chemistry representations can likely be extended to the peptide-MHC predictions, for example, with *PANDORA* or other tools. Ultimately, improvement to MHC modeling and subsequent extraction of generalizable biophysical properties will lead to better predictions of immunogenicity. Highlighting this point, a thorough structural study demonstrated that a non-anchor position mutation in a MHC-I peptide, presented by an ovarian cancer tumor, modified both the structural and dynamic properties of the bound complex. These changes resulted in optimal confirmations for interaction and subsequent activation of cognate T cells[39]. Such an observation would be difficult, if not impossible to determine from sequence alone and emphasizes the value of structural considerations when studying immunogenicity.

2.1.3 Catch bond description of TCRs

Catch bonds have been referred to as the interaction between various biomolecules and biomolecular surfaces, where the lifetime of the bond increases with the application of tensile force on the bond [40, 41]. The atomistic detail of catch bond formation had remained elusive for a long period of time, but the general explanation was given by a two-state model or a two-pathway model. In the two-state model, the receptor-ligand complex is theorized to exist in two distinct states, a short-lived and a long-lived state. The application of force loosens the interaction between the binding site and a regulatory site which drives the whole complex towards the long-lifetime state[40]. In the two-pathway model the receptor-ligand complex undergoes unbinding via two distinct pathways with different K_{off} values, and the application of tensile force triggers the allosteric change, that leads the unbinding to happen via the pathway with a high energy barrier, thereby the long-lifetime[42]. Such catch bonding has been observed at TCR-peptide-MHC immune synapse, and more importantly, immunogenicity has been attributed to the strength of the catch bond formation[43]. Hence, catch bonds offer a biophysical descriptor of MHC alleles for presenting peptides to the TCRs. Interestingly, unlike binding affinity, catch bonds uniquely capture the system's out-of-equilibrium properties. Therefore, it can capture the state of the immune synapse under stress, which rectifies the frozen stationary picture of complexes drawn by the affinity measures. This descriptor is computable using Steered MD simulations [44] and more recently using metadynamics methodologies [45], offering insights into how sequence changes reflect in non-equilibrium interaction changes. However, both the experimental and computational biophysical methods for tracking catch bonds are resource-intensive, so high-throughput measurements are yet missing, in turn impacting the extensive use of this information in immunology models. The advent of reinforcement learning with Jarzynski's equality and so-called 'stiff-spring approximations' [46] to formulate a space of molecular ac-

tions using steered MD simulations presents a promising step forward in rapidly modeling at least the 2-state model of the catch bonds as another biophysical descriptor in computational immunology[43]. A more rigorous consideration of catch bond formation has practical implications for enhancing T cell based cancer immunotherapies. A recent study showed low affinity TCRs can be optimized to acquire catch bonding characteristic allowing for potent activation at relative weak 3D binding affinities[47]. This has the ability to drive strong antitumor immune response with a lower risk of potentially life threatening cross-reactivity.

2.2 Molecular description

The translation from atomistic to molecular biophysical representation has become popular to allow algorithms to distinguish self vs. non-self interactomes. The biophysical representations of glycans underpinning the pathogen entry path offer some stark examples. By employing tools like variational autoencoders, the so-called 'glycan shield' of spike proteins was dissected to detect the role of specific glycan size, orientation, and chemistry [48]. A physical interpretation of the latent spaces was determined from protein-glycan contacts. Subsequently, we engineered the glycan shield based on their contact representation to reduce the infectivity of the NL63 coronavirus by nearly 50%[49]. This idea of monitoring contacts was also extrapolated to monitor inter-glycan interactions between the cell surface of the influenza virus and those of chicken and human cell surface glycolyx [50]. Again, by translating fluorescence signals into a contact matrix representation, support vector machines were successful in identifying the critical density of glycans that make the H1N1 cells in mammalian cells show a greater binding than when grown in egg cells. Finally, the protein-protein contact matrices also found application in vector design for AstraZeneca and J&J's COVID vaccines, implication platelet factor proteins in blood clotting side-effects of the vaccine candidate[51]. Altogether, contact matrices can offer a robust biophysical representation, where in molecular interactions can be classified to be self or non-self.

2.3 Cellular description

Whole cell models, though scarce, have found applications in computational immunology. A mechanistic, multiscale mathematical model of immunogenicity for therapeutic proteins was formulated by recapitulating key biological mechanisms, including antigen presentation, activation, proliferation, and differentiation of immune cells, secretion of antidrug antibodies, as well as *in vivo* disposition of antibodies and therapeutic proteins [52]. The multiscale model structure can be represented by the subcellular, cellular, and whole-body levels. To represent the physiology of MHC-II, a key parameter used in these models involves the number of T-epitope-MHC, *in silico* T cell epitope prediction and experimental measurements of their MHC-binding affinities, which is scaffolded within a two-compartment drug pharmacokinetics model. Using adalimumab as an example therapeutic protein, the model is able to simulate immune responses against adalimumab in individual subjects and in a population and also provides estimations of immunogenicity incidence and drug exposure reduction that can be validated experimentally [52, 53]. Most of the cell models in immunology are agent-based that use the automaton algorithm with specific mechanistic logics or rules. Interestingly these rules show remarkable similarity with classical thermodynamic and kinetic principles, such as landscapes and equations of motion [54]. Such models have found applications in CD4+ T cell responses to influenza infections, multiscale mechanistic modeling of human dendritic cells, and have potential applications in dendritic cell-based targeted cell therapies [55, 56].

2.4 Macro description

The integration of molecular immunology concepts into macro-level analyses has already demonstrated significant potential in elucidating disease associations. A notable example is the use of patient-specific MHC genotypes to predict disease risk. For instance, large-scale analyses involving 9,176 cancer patients revealed that MHC-I genotypes were predictive of the tumor mutational landscape[57]. This study found that oncogenic mutations were more likely to occur in regions not presented by the patient's MHC-I molecules, suggesting that gaps in antigen presentation contribute to tumor evolution. Similarly, patients undergoing immune checkpoint blockade therapies have shown improved responses when their MHC-I genotype allows for the presentation of a more diverse array of potential peptides[58]. More recently, biophysical approaches have been applied to link MHC-I genotypes with disease risk and progression[11]. Recently, we created a diverse protein ensemble of 5,281 MHC-I protein binding grooves, generating 211,240 structural models, which were subsequently translated into a simplified representation of electrostatic properties (5,281 averaged electrostatic maps). A subset of these maps, those with known MHC-I binding motifs, were used to train an *Inception* neural network capable of predicting MHC-I binding motifs from electrostatic maps alone. Beyond the ability to perform high throughput proteome-scale binding predictions, the predicted binding motifs were utilized to construct interaction networks that accurately

classified HIV disease progression and immune checkpoint therapy response. At the population level, applications of MHC-I genotype analysis have revealed further insights. A consensus MHC-I prediction model, *ensembleMHC*, demonstrated that populations enriched for MHC-I alleles capable of strongly binding multiple peptides from SARS-CoV-2 structural proteins exhibited lower mortality rates during the pre-vaccination phase of the COVID-19 pandemic[6]. This suggests that MHC-I diversity and peptide-binding capacity at the population level may serve as predictors of disease outcomes in emerging viral threats. These findings highlight some of the promise of MHC genotype based analysis in both disease risk assessment and therapeutic strategy development. MHC analysis can aid in predicting susceptibility to autoimmune diseases and cancer while also informing vaccine design by optimizing patient antigen selection.

3 Outlook: Future inspired by the past of functional representations

Most of the biophysics, including the powerful integrative models we know, is predicated upon the sequence → structure → function → phenotype paradigm. With the maturation of machine learning techniques and the availability of data at various scales, researchers (particularly bioinformaticians) have been trying to bridge gaps between the different tiers of this process, starting from the age-old genotype-type modeling to CASP and AlphaFold's sequence structure up to recent attempts to go from sequence to ensemble. However, physical causality is often missing in the traditional bioinformatics models, thus far sidelining the role of AI-driven advances only to predictions of the forward direction. So, it is high time that we introduce physical ideas to conceive generative models that backmap phenotypes down to an ensemble of structures and sequences. Model representations play a central role in this mapping process. The traditional sequence of 3D coordinate structural representations require an enormous amount of training data, and prone to overfitting, nonetheless offer the most extensive models. In contrast, the thermodynamic or kinetic representations, using ideas of entropy or committor functions are quite generalizable across application domains but lack the physical interpretability[59]. Loosely, they draw analogies to the plane wave basis set representations that find application in several areas of quantum mechanics [60]. However, akin to how quantum mechanics was represented in the molecular systems using the Gaussian-like basis set representations, we posit that biophysical representations offer a segue for representing the deep learning models in the molecular space. To this end, we highlight a number of representations that are either being used or hold the potential for multiscale applications in computational immunology. Similar to how the Gaussian orbitals offer physical interpretation to highly resolved electronic structures e.g. using the molecular orbital theory, the biophysical functions (pocket fields, QM/MM charge density, binding affinity, catch bonding, contact matrices, molecular electrostatics) offer interpretability in terms of deep rooted physical theories associated (thermodynamic integration, electronic structure theory, equilibrium and non-equilibrium statistical theories, linear response theories, polymer folding, continuum mechanics) that can be projected on structure and function. Essentially they offer a physical basis to the loss functions and the latent spaces, that gets to learning both the data as well as the context. So, we propose a sustained intellectual effort in this direction.

4 Acknowledgement

A. S. was supported by a CAREER award from the NSF (MCB-1942763) and an RO1 grant from the NIH (GM095583). A.S. also acknowledges start-up grants from Arizona State University School of Molecular Sciences and Bidesign Institute's Center for Applied Structural Discovery. A.S. acknowledge funding from the Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Sciences, of the U.S. Department of Energy through grants DESC0010575. A.S. acknowledges grant DE-SC0022956 for their support, also from the Department of Energy. This material is based on work supported by the National Defense Education Program (NDEP) for Science, Technology, Engineering, and Mathematics (STEM) Education.

5 Declarations of interest:

The authors declare no conflict of interest.

References

- [1] D. J. Barker, G. Maccari, X. Georgiou, M. A. Cooper, P. Flicek, J. Robinson, and S. G. E. Marsh, "The ipd-imgt/hla database," *Nucleic acids research*, vol. 51, no. D1, p. D1053–D1060, 2023.
- [2] T. Mora and A. M. Walczak, "Quantifying lymphocyte receptor diversity," in *Systems Immunology*, pp. 183–198, CRC Press, 2018.
- [3] F. Brauer, "Mathematical epidemiology: Past, present, and future," *Infectious Disease Modelling*, vol. 2, no. 2, pp. 113–127, 2017.
- [4] R. Ross, *The prevention of malaria*. John Murray, 1911.
- [5] L. He and J. Zhu, "Computational tools for epitope vaccine design and evaluation," *Current opinion in virology*, vol. 11, pp. 103–112, 2015.
- [6] E. A. Wilson, G. Hirneise, A. Singharoy, and K. S. Anderson, "Total predicted MHC-I epitope load is inversely associated with population mortality from SARS-CoV-2," *Cell Reports Medicine*, vol. 2, no. 3, 2021.
- [7] D. Chowell, L. G. T. Morris, C. M. Grigg, J. K. Weber, R. M. Samstein, V. Makarov, F. Kuo, S. M. Kendall, D. Requena, N. Riaz, and others, "Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy," *Science*, vol. 359, no. 6375, pp. 582–587, 2018.
- [8] M. C. Jespersen, B. Peters, M. Nielsen, and P. Marcatili, "BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes," *Nucleic acids research*, vol. 45, no. W1, p. W24–W29, 2017.
- [9] H. R. Ansari and G. P. S. Raghava, "Identification of conformational B-cell Epitopes in an antigen from its primary sequence," *Immunome research*, vol. 6, pp. 1–9, 2010.
- [10] B. Peters, M. Nielsen, and A. Sette, "T cell epitope predictions," *Annual Review of Immunology*, vol. 38, no. 1, pp. 123–145, 2020.
- [11] E. Wilson, J. K. Cava, D. Chowell, R. Raja, K. K. Mangalaparathi, A. Pandey, M. Curtis, K. S. Anderson, and A. Singharoy, "The electrostatic landscape of MHC-peptide binding revealed using inception networks," *Cell Systems*, vol. 15, no. 4, pp. 362–373, 2024.
- [12] P. Haste Andersen, M. Nielsen, and O. L. E. Lund, "Prediction of residues in discontinuous B-cell epitopes using protein 3D structures," *Protein Science*, vol. 15, no. 11, pp. 2558–2567, 2006.
- [13] D. J. Rigden and X. M. Fernández, "The 2024 Nucleic Acids Research database issue and the online molecular biology database collection," *Nucleic Acids Research*, vol. 52, pp. D1–D9, 10 2023.
- [14] N. Petrovsky and V. Brusic, "Computational immunology: The coming of age," *Immunology and cell biology*, vol. 80, no. 3, pp. 248–254, 2002.
- [15] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Židek, A. Potapenko, and others, "Highly accurate protein structure prediction with AlphaFold," *nature*, vol. 596, no. 7873, pp. 583–589, 2021.
- [16] P. Bradley, "Structure-based prediction of T cell receptor: peptide-MHC interactions," *Elife*, vol. 12, p. e82813, 2023.
- [17] R. Raha, Y. Ding, Q. Liu, and F.-X. Wu, "Unseen Epitope-TCR Interaction Prediction based on Amino Acid Physicochemical Properties," in *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 3122–3129, 2022.
- [18] L. Deng, C. Ly, S. Abdollahi, Y. Zhao, I. Prinz, and S. Bonn, "Performance comparison of TCR-pMHC prediction tools reveals a strong data dependency," *Frontiers in Immunology*, vol. 14, p. 1128326, 2023.
- [19] O. N. A. Demerdash and J. C. Smith, "TCR-H: explainable machine learning prediction of T-cell receptor epitope binding on unseen datasets," *Frontiers in Immunology*, vol. 15, p. 1426173, 2024.
- [20] S. Ferdous and A. C. R. Martin, "AbDb: antibody structure database—a database of PDB-derived antibody structures," *Database*, vol. 2018, p. bay040, 2018.

- [21] J. V. Ponomarenko and P. E. Bourne, "Antibody-protein interactions: benchmark datasets and prediction tools evaluation," *BMC structural biology*, vol. 7, pp. 1–19, 2007.
- [22] J. Młokosiewicz, P. Deszyński, W. Wilman, I. Jaszczyszyn, R. Ganesan, A. Kovaltsuk, J. Leem, J. D. Galson, and K. Krawczyk, "AbDiver: a tool to explore the natural antibody landscape to aid therapeutic design," *Bioinformatics*, vol. 38, no. 9, pp. 2628–2630, 2022.
- [23] C. Schneider, M. I. J. Raybould, and C. M. Deane, "SAbDab in the age of biotherapeutics: updates including SAbDab-nano, the nanobody structure tracker," *Nucleic acids research*, vol. 50, no. D1, p. D1368–D1372, 2022.
- [24] A. M. Hummer, C. Schneider, L. Chinery, and C. M. Deane, "Investigating the volume and diversity of data needed for generalizable antibody-antigen G prediction," *bioRxiv*, pp. 2023–2025, 2023.
- [25] A. E. Cleves and A. N. Jain, "Quantitative surface field analysis: learning causal models to predict ligand binding affinity and pose," *Journal of computer-aided molecular design*, vol. 32, pp. 731–757, 2018.
- [26] H. Cai, Z. Zhang, M. Wang, B. Zhong, Q. Li, Y. Zhong, Y. Wu, T. Ying, and J. Tang, "Pretrainable geometric graph neural network for antibody affinity maturation," *Nature communications*, vol. 15, no. 1, p. 7785, 2024.
- [27] F. Garrido, "The Discovery of the Major Histocompatibility Complex (MHC): The H-2 in Mice and the HLA in Man," in *The Major Histocompatibility Complex (MHC/HLA) in Medicine: A Personal Recollection*, pp. 1–13, Springer, 2024.
- [28] P. Hull, "Notes on dr snell's observations concerning the h-2 locus polymorphism.," 1970.
- [29] C. A. Janeway Jr, P. Travers, M. Walport, and M. J. Shlomchik, "The major histocompatibility complex and its functions," in *Immunobiology: The Immune System in Health and Disease. 5th edition*, Garland Science, 2001.
- [30] M. Wieczorek, E. T. Abualrous, J. Sticht, M. Álvaro-Benito, S. Stolzenberg, F. Noé, and C. Freund, "Major histocompatibility complex (mhc) class i and mhc class ii proteins: conformational plasticity in antigen presentation," *Frontiers in immunology*, vol. 8, p. 292, 2017.
- [31] G. L. J. Keller, L. I. Weiss, and B. M. Baker, "Physicochemical heuristics for identifying high fidelity, near-native structural models of Peptide/MHC complexes," *Frontiers in Immunology*, vol. 13, p. 887759, 2022.
- [32] C. Zhang, A. Anderson, and C. DeLisi, "Structural principles that govern the peptide-binding motifs of class I MHC molecules," Edited by I. Wilson, *Journal of Molecular Biology*, vol. 281, no. 5, pp. 929–947, 1998.
- [33] L. M. F. Bertoline, A. N. Lima, J. E. Krieger, and S. K. Teixeira, "Before and after AlphaFold2: An overview of protein structure prediction," *Frontiers in bioinformatics*, vol. 3, p. 1120370, 2023.
- [34] M. Menegatti Rigo, D. Amaral Antunes, M. de Freitas, M. de Almeida Mendes, L. Meira, M. Siniaglia, and G. Fioravanti Vieira, "DockTope: a Web-based tool for automated pMHC-I modelling," *Scientific reports*, vol. 5, no. 1, p. 18413, 2015.
- [35] H.-H. Kyeong, Y. Choi, and H.-S. Kim, "GradDock: rapid simulation and tailored ranking functions for peptide-MHC Class I docking," *Bioinformatics*, vol. 34, no. 3, pp. 469–476, 2018.
- [36] J. R. Abella, D. A. Antunes, C. Clementi, and L. E. Kavraki, "APE-gen: A fast method for generating ensembles of bound peptide-MHC conformations," *Molecules*, vol. 24, no. 5, p. 881, 2019.
- [37] P. Bryant, G. Pozzati, and A. Elofsson, "Improved prediction of protein-protein interactions using AlphaFold2," *Nature communications*, vol. 13, no. 1, p. 1265, 2022.
- [38] F. M. Parizi, D. F. Marzella, G. Ramakrishnan, P. A. C. 't Hoen, M. H. Karimi-Jafari, and L. C. Xue, "PANDORA v2. 0: Benchmarking peptide-MHC II models and software improvements," *Frontiers in Immunology*, vol. 14, p. 1285899, 2023.
- [39] J. R. Devlin, J. A. Alonso, C. M. Ayres, G. L. Keller, S. Bobisse, C. W. Vander Kooi, G. Coukos, D. Gfeller, A. Harari, and B. M. Baker, "Structural dissimilarity from self drives neoepitope escape from immune tolerance," *Nature chemical biology*, vol. 16, no. 11, pp. 1269–1276, 2020.
- [40] S. Hertig and V. Vogel, "Catch bonds," *Current Biology*, vol. 22, pp. R823–R825, 10 2012.

- [41] B. T. Marshall, M. Long, J. W. Piper, T. Yago, R. P. McEver, and C. Zhu, "Direct observation of catch bonds involving cell-adhesion molecules," *Nature*, vol. 423, pp. 190–193, 5 2003.
- [42] E. V. Sokurenko, V. Vogel, and W. E. Thomas, "Catch bond mechanism of force-enhanced adhesion: counter-intuitive, elusive but . . . widespread?," *Cell host & microbe*, vol. 4, p. 314, 10 2008.
- [43] H.-K. Choi, P. Cong, C. Ge, A. Natarajan, B. Liu, Y. Zhang, K. Li, M. N. Rushdi, W. Chen, J. Lou, and others, "Catch bond models may explain how force amplifies TCR signaling and antigen discrimination," *Nature communications*, vol. 14, no. 1, p. 2616, 2023.
- [44] C. Schoeler, K. H. Malinowska, R. C. Bernardi, L. F. Milles, M. A. Jobst, E. Durner, W. Ott, D. B. Fried, E. A. Bayer, K. Schulten, and others, "Ultrastable cellulosome-adhesion complex tightens under load," *Nature communications*, vol. 5, no. 1, p. 5635, 2014.
- [45] W. J. Peña Coa and G. M. Hocky, "Assessing models of force-dependent unbinding rates via infrequent metadynamics," *The Journal of Chemical Physics*, vol. 156, no. 12, 2022.
- [46] S. Park and K. Schulten, "Calculating potentials of mean force from steered molecular dynamics simulations," *The Journal of chemical physics*, vol. 120, no. 13, pp. 5946–5961, 2004.
- [47] X. Zhao, E. M. Kolawole, W. Chan, Y. Feng, X. Yang, M. H. Gee, K. M. Jude, L. V. Sibener, P. M. Fordyce, R. N. Germain, *et al.*, "Tuning t cell receptor sensitivity through catch bond engineering," *Science*, vol. 376, no. 6589, p. eabl5282, 2022.
- [48] L. Casalino, A. C. Dommer, Z. Gaieb, E. P. Barros, T. Sztain, S.-H. Ahn, A. Trifan, A. Brace, A. T. Bogetti, A. Clyde, and others, "AI-driven multiscale simulations illuminate mechanisms of SARS-CoV-2 spike dynamics," *The International Journal of High Performance Computing Applications*, vol. 35, no. 5, pp. 432–451, 2021.
- [49] D. Chmielewski, E. A. Wilson, G. Pintilie, P. Zhao, M. Chen, M. F. Schmid, G. Simmons, L. Wells, J. Jin, A. Singharoy, and others, "Structural insights into the modulation of coronavirus spike tilting and infectivity by hinge glycans," *Nature communications*, vol. 14, no. 1, p. 7175, 2023.
- [50] T. M. Lucas, C. Gupta, M. O. Altman, E. Sanchez, M. R. Naticchia, P. Gagneux, A. Singharoy, and K. Godula, "Mucin-mimetic glycan arrays integrating machine learning for analyzing receptor pattern recognition by influenza A viruses," *Chem*, vol. 7, no. 12, pp. 3393–3411, 2021.
- [51] A. T. Baker, R. J. Boyd, D. Sarkar, A. Teijeira-Crespo, C. K. Chan, E. Bates, K. Waraich, J. Vant, E. Wilson, C. D. Truong, and others, "ChAdOx1 interacts with CAR and PF4 with implications for thrombosis with thrombocytopenia syndrome," *Science Advances*, vol. 7, no. 49, p. eabl8213, 2021.
- [52] X. Chen, T. Hickling, and P. Vicini, "A mechanistic, multiscale mathematical model of immunogenicity for therapeutic proteins: part 2 – model applications," *CPT: pharmacometrics & systems pharmacology*, vol. 3, no. 9, pp. 1–10, 2014.
- [53] A. Handel, N. L. La Gruta, and P. G. Thomas, "Simulation modelling for immunologists," *Nature Reviews Immunology*, vol. 20, no. 3, pp. 186–195, 2020.
- [54] L. Koopmans and H. Youk, "Predictive landscapes hidden beneath biological cellular automata," *Journal of biological physics*, vol. 47, no. 4, pp. 355–369, 2021.
- [55] K. Y. Wertheim, B. L. Puniya, A. La Fleur, A. R. Shah, M. Barberis, and T. Helikar, "A multi-approach and multi-scale platform to model cd4+ t cells responding to infections," *PLoS computational biology*, vol. 17, no. 8, p. e1009209, 2021.
- [56] S. S. Aghamiri, B. L. Puniya, R. Amin, and T. Helikar, "A multiscale mechanistic model of human dendritic cells for in-silico investigation of immune responses and novel therapeutics discovery," *Frontiers in Immunology*, vol. 14, p. 1112985, 2023.
- [57] R. Marty, S. Kaabinejadian, D. Rossell, M. J. Slifker, J. van de Haar, H. B. Engin, N. de Prisco, T. Ideker, W. H. Hildebrand, J. Font-Burgada, and others, "MHC-I genotype restricts the oncogenic mutational landscape," *Cell*, vol. 171, no. 6, pp. 1272–1283, 2017.
- [58] D. Chowell, C. Krishna, F. Pierini, V. Makarov, N. A. Rizvi, F. Kuo, L. G. T. Morris, N. Riaz, T. L. Lenz, and T. A. Chan, "Evolutionary divergence of HLA class I genotype impacts efficacy of cancer immunotherapy," *Nature medicine*, vol. 25, no. 11, pp. 1715–1720, 2019.

- [59] S. Mehdi, Z. Smith, L. Herron, Z. Zou, and P. Tiwary, "Enhanced sampling with machine learning," *Annual Review of Physical Chemistry*, vol. 75, 2024.
- [60] B. Nagy and F. Jensen, "Basis sets in quantum chemistry," *Reviews in computational chemistry*, vol. 30, pp. 93–149, 2017.

Figure 1

multiscale computational immunology models

