

Techniques for Theoretical Prediction of Immunogenic Peptides

Subjects: Cell Biology

Contributor: Robert Friedman

Small peptides are an important component of the vertebrate immune system. They are important molecules for distinguishing proteins that originate in the host from proteins derived from a pathogenic organism, such as a virus or bacterium. Consequently, these peptides are central for the vertebrate host response to intracellular and extracellular pathogens. Computational models for prediction of these peptides have been based on a narrow sample of data with an emphasis on the position and chemical properties of the amino acids. In past literature, this approach has resulted in higher predictability than models that rely on the geometrical arrangement of atoms. However, protein structure data from experiment and theory are a source for building models at scale, and, therefore, knowledge on the role of small peptides and their immunogenicity in the vertebrate immune system. The following sections introduce procedures that contribute to theoretical prediction of peptides and their role in immunogenicity. Lastly, deep learning is discussed as it applies to immunogenetics and the acceleration of knowledge by a capability for modeling the complexity of natural phenomena.

Keywords: immunogenetics ; immunogenic peptides ; pathogenic organism ; vertebrate host ; computational models ; protein structure ; T cell receptor ; adaptive immunity ; deep learning

The adaptive immune system of vertebrates is a system that includes cells and molecules whose role is to distinguish self from the outside world (non-self). Therefore, a vertebrate host has the potential for detecting and clearing pathogenic organisms from its organ systems. A major component of adaptive immunity involves a linear chain of amino acids: the small peptides ^[1]. The small peptide is of interest since the host immune system relies on it as a marker for a determination on whether a protein originates from itself or instead of a foreign source, such as a virus or bacterium ^[2]. This system can also identify its own cells as foreign if they are genetically altered by a process that causes production of unfamiliar molecules ^{[3][4]}.

These peptides of interest are formed by cleavage of proteins in cells of the host, and they form the basis for the cellular processes of immune surveillance, and identification of pathogens and cells that operate outside their normal genetic programming ^{[3][5][6][7]}. When adaptive immunity falsely identifies a peptide derived from a protein that is essential to the individual as not originating from that individual, a phenomenon referred to as autoimmunity occurs ^{[8][9][10][11]}. A generalized example of autoimmunity is where a subset of T cells ^{[12][13]}, a name that references their development in the thymus ^{[14][15]}, falsely detects small peptides as presented on the surface of cells as originating from non-self and subsequently signals the immune system to eliminate these cells in the host ^{[16][17][18]}.

The mechanism for peptide detection is reliant on molecular binding between the peptide and a major histocompatibility complex (MHC) receptor that is expressed in the majority of cells of a vertebrate ^{[19][20][21]}. Nearly all cells of the canonical vertebrate express MHC Class 1 cell surface receptors that are capable of presenting peptides of an intracellular origin, while a subset of cell types of the immune system express MHC Class 2 cell surface receptors for presentation of peptides of an intercellular origin.

Furthermore, the mechanism described above is refined by training the T cells to perform as specialists so that they disfavor any attack on normal cells, while favoring the proliferation of the T cells that have developed to attack non-self ^{[14][21]}. This is not a deterministic process, however. The dictates of probability are present in biological systems, including: (1) the generation of genetic diversity across the various MHC receptor types, (2) the cleavage process for generation of small peptides from a protein, (3) the timeliness of the immune response to molecular evidence of a pathogen, (4) the binding strength of peptides to an MHC receptor, and (5) the requisite sample of peptides for detection of a pathogen. This system is in contrast to a human designed system (engineered) in which the structure and function originate from an artificial design and a low tolerance for the prior mentioned variability.

The aggregate of past collections of immunological peptide data is not representative of the total space of these peptides ^{[3][21]}. For example, only a small proportion of MHC molecules have been studied for their association with small peptides.

This sampling problem is related to the allelic distribution of the MHC molecules. While there are about a dozen genetic loci in clusters that code for a MHC protein receptor, the number of alleles among these loci is very high as compared to the other genetic loci in the typical vertebrate genome. In the human population, the expected number of alleles for the MHC loci is estimated in the thousands [3][22]. Correspondingly, these loci are active genetic sites of evolutionary change and generation of diversity, and—unlike the other regions of the genome—this genetic diversity has been undiminished at the genetic level by the putative bottleneck that reduced our effective population size to mere thousands of individuals [23]. Likewise, the study of immunological peptides has generally been restricted to that of the human population and animal models that serve as a proxy in the study of biomedicine and livestock [21].

Moreover, there is a preference for MHC class type as a result of model feasibility. The MHC Class 1 receptor is generally favored over that of Class 2 in modeling the MHC-peptide association, partly because in MHC Class 1, some of the amino acids of the peptide are confined in pockets of the MHC receptor [1][3][24][25]. This has led to predictive models of MHC-peptide (pMHC) binding that parameterize the position and chemical type of the amino acids of these peptides [21][26]. These models have exceeded the predictiveness as compared to models based solely on geometrical data of the atomic arrangements [3]. However, the geometrical features are expected to contribute to insight on pMHC binding and models for predicting an adaptive immune response.

Recently, artificial neural networks and related machine language approaches have led to advances in knowledge of protein structure and the potential for modeling the association between proteins and other molecules [21][26][27][28][29]. These methods are capable of highly predictive models that incorporate disparate kinds of data, such as in the use of both geometrical and chemical features in estimating the binding affinity for an MHC receptor with a peptide [21]. Moreover, they are highly efficient in the case where modeling is dependent on a very large number of parameters, as commonly observed in the interactions of biological molecules. Consequently, the deep learning approaches have shown success in the prediction of protein structure across a broad sampling of the clades of living organisms [30]. These approaches are complemented through the analysis of metrics, preferably with a level of interpretability, that are capable of estimating the geometrical similarity among proteins [31][32][33].

As a whole, the study of immunogenetics relies on collecting data and building models as expected in the pursuit of knowledge [21][34]. Deep learning is applicable to these goals, for which the data collection is extensive and there is a theoretical basis for the system of interest. Ideally, this kind of scientific practice is expected to lead to a meaningful synthesis that is unmired by a collector's fondness for naming schemes and ungrounded collations of terms and studies [35]. The latter perspective resembles the practice of creating images of science, akin to an art form, that sometimes occur in the disciplines of natural science while not achieving the aim of extending knowledge through the purposeful modeling of natural phenomena [36].

References

1. Wieczorek, M.; Abualrous, E.T.; Sticht, J.; Álvaro-Benito, M.; Stolzenberg, S.; Noé, F.; Freund, C. Major Histocompatibility Complex (MHC) Class I and MHC Class II Proteins: Conformational Plasticity in Antigen Presentation. *Front. Immunol.* 2017, 8, 292.
2. Dhatchinamoorthy, K.; Colbert, J.D.; Rock, K.L. Cancer Immune Evasion through Loss of MHC Class I Antigen Presentation. *Front. Immunol.* 2021, 12, 636568.
3. Peters, B.; Nielsen, M.; Sette, A. T Cell Epitope Predictions. *Annu. Rev. Immunol.* 2020, 38, 123–145.
4. Engelhard, V.H. Structure of peptides associated with MHC class I molecules. *Curr. Opin. Immunol.* 1994, 6, 13–23.
5. Davis, M.M.; Bjorkman, P.J. T-cell antigen receptor genes and T-cell recognition. *Nature* 1988, 335, 744.
6. Serwold, T.; Gonzalez, F.; Kim, J.; Jacob, R.; Shastri, N. ERAAP customizes peptides for MHC class I molecules in the endoplasmic reticulum. *Nature* 2002, 419, 480–483.
7. Clevers, H. The T Cell Receptor/Cd3 Complex: A Dynamic Protein Ensemble. *Annu. Rev. Immunol.* 1988, 6, 629–662.
8. Theofilopoulos, A.N.; Kono, D.H.; Baccala, R. The multiple pathways to autoimmunity. *Nat. Immunol.* 2017, 18, 716–724.
9. Uemura, Y.; Senju, S.; Maenaka, K.; Iwai, L.K.; Fujii, S.; Tabata, H.; Tsukamoto, H.; Hirata, S.; Chen, Y.Z.; Nishimura, Y.; et al. Systematic Analysis of the Combinatorial Nature of Epitopes Recognized by TCR Leads to Identification of Mimicry Epitopes for Glutamic Acid Decarboxylase 65-Specific TCRs. *J. Immunol.* 2003, 170, 947–960.

10. Borrmann, T.; Pierce, B.G.; Vreven, T.; Baker, B.M.; Weng, Z. High-throughput modeling and scoring of TCR-pMHC complexes to predict cross-reactive peptides. *Bioinformatics* 2020, 36, 5377–5385.
 11. Prinz, J.C. Immunogenic self-peptides—The great unknowns in autoimmunity: Identifying T-cell epitopes driving the autoimmune response in autoimmune diseases. *Front. Immunol.* 2023, 13, 1097871.
 12. Yanagi, Y.; Yoshikai, Y.; Leggett, K.; Clark, S.P.; Aleksander, I.; Mak, T.W. A human T cell-specific cDNA clone encodes a protein having extensive homology to immunoglobulin chains. *Nature* 1984, 308, 145–149.
 13. Hedrick, S.M.; Cohen, D.I.; Nielsen, E.A.; Davis, M.M. Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. *Nature* 1984, 308, 149–153.
 14. Yang, Q.; Bell, J.J.; Bhandoola, A. T-cell lineage determination. *Immunol. Rev.* 2010, 238, 12–22.
 15. Nikolich-Zugich, J.; Slifka, M.K.; Messaoudi, I. The many important facets of T-cell repertoire diversity. *Nat. Rev. Immunol.* 2004, 4, 123–132.
 16. Ashby, K.M.; Hogquist, K.A. A guide to thymic selection of T cells. *Nat. Rev. Immunol.* 2023, 23, 697.
 17. George, J.T.; Kessler, D.A.; Levine, H. Effects of thymic selection on T cell recognition of foreign and tumor antigenic peptides. *Proc. Natl. Acad. Sci. USA* 2017, 114, E7875–E7881.
 18. Smith, D.A.; Germolec, D.R. Introduction to Immunology and Autoimmunity. *Environ. Health Perspect.* 1999, 107, 661.
 19. Klein, J.; Figueroa, F. Evolution of the major histocompatibility complex. *Crit. Rev. Immunol.* 1986, 6, 295–386.
 20. Germain, R.N. MHC-dependent antigen processing and peptide presentation: Providing ligands for T lymphocyte activation. *Cell* 1994, 76, 287–299.
 21. Nielsen, M.; Andreatta, M.; Peters, B.; Buus, S. Immunoinformatics: Predicting Peptide–MHC Binding. *Annu. Rev. Biomed. Data Sci.* 2020, 3, 191–215.
 22. Radwan, J.; Babik, W.; Kaufman, J.; Lenz, T.L.; Winternitz, J. Advances in the Evolutionary Understanding of MHC Polymorphism. *Trends Genet.* 2020, 36, 298–311.
 23. Jorde, L.B. Genetic variation and human evolution. *Am. Soc. Hum. Genet.* 2003, 7, 28–33.
 24. Bjorkman, P.J.; Saper, M.A.; Samraoui, B.; Bennett, W.S.; Strominger, J.L.; Wiley, D.C. Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* 1987, 329, 506–512.
 25. Antunes, D.A.; Devaurs, D.; Moll, M.; Lizée, G.; Kavraki, L.E. General Prediction of Peptide-MHC Binding Modes Using Incremental Docking: A Proof of Concept. *Sci. Rep.* 2018, 8, 4327.
 26. Mei, S.; Li, F.; Leier, A.; Marquez-Lago, T.T.; Giam, K.; Croft, N.P.; Akutsu, T.; Smith, A.I.; Li, J.; Rossjohn, J.; et al. A comprehensive review and performance evaluation of bioinformatics tools for HLA class I peptide-binding prediction. *Brief. Bioinform.* 2020, 21, 1119–1135.
 27. Sohail, M.S.; Ahmed, S.F.; Quadeer, A.A.; McKay, M.R. In silico T cell epitope identification for SARS-CoV-2: Progress and perspectives. *Adv. Drug Deliv. Rev.* 2021, 171, 29–47.
 28. Raoufi, E.; Hemmati, M.; Eftekhari, S.; Khaksaran, K.; Mahmodi, Z.; Farajollahi, M.M.; Mohsenzadegan, M. Epitope Prediction by Novel Immunoinformatics Approach: A State-of-the-art Review. *Int. J. Pept. Res. Ther.* 2019, 26, 1155–1163.
 29. Bradley, P. Structure-based prediction of T cell receptor:peptide-MHC interactions. *eLife* 2023, 12, e82813.
 30. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021, 596, 583–589.
 31. Zhang, Y.; Skolnick, J. Scoring function for automated assessment of protein structure template quality. *Proteins Struct. Funct. Bioinform.* 2004, 57, 702–710.
 32. Zemla, A. LGA: A method for finding 3D similarities in protein structures. *Nucleic Acids Res.* 2003, 31, 3370–3374.
 33. Leman, J.K.; Szczerbiak, P.; Renfrew, P.D.; Gligorijevic, V.; Berenberg, D.; Vatanen, T.; Taylor, B.C.; Chandler, C.; Janssen, S.; Pataki, A.; et al. Sequence-structure-function relationships in the microbial protein universe. *Nat. Commun.* 2023, 14, 2351.
 34. Vita, R.; Overton, J.A.; Greenbaum, J.A.; Ponomarenko, J.; Clark, J.D.; Cantrell, J.R.; Wheeler, D.K.; Gabbard, J.L.; Hix, D.; Sette, A.; et al. The immune epitope database (IEDB) 3.0. *Nucleic Acids Res.* 2014, 43, D405–D412.
 35. Johnson, K. Natural history as stamp collecting: A brief history. *Arch. Nat. Hist.* 2007, 34, 244–258.
 36. Frede, M. Plato's Sophist on False Statements. In *The Cambridge Companion to Plato*; Kraut, R., Ed.; Cambridge University Press: Cambridge, UK, 1992; pp. 397–424.
-

