

Biological databases useful for epitope mapping and immune response simulation

Bases de datos biológicas de utilidad para el mapeo de epítomos y la simulación de respuestas inmunes

Adrián Alejandro Vitón-Castillo¹  , Carlos Alfredo Miló Valdés¹  , Lidia Cecilia Pérez Acevedo²  

¹Universidad de Ciencias Médicas de Pinar del Río. Hospital Pediátrico Provincial Docente “Pepe Portilla”, Departamento de Inmunología. Pinar del Río, Cuba.

²Centro de Inmunología Molecular. La Habana, Cuba.

Cite as

Vitón-Castillo AA, Miló Valdés CA, Pérez Acevedo LC. Biological databases useful for epitope mapping and immune response simulation. SAP Biomedical & Chemical Engineering Innovation. 2025; 4:300. <https://doi.org/10.56294/evk2025300>

Publication History

Submitted: 11-11-2024

Revised: 22-02-2025

Accepted: 18-06-2025

Published: 19-06-2025

Corresponding Author

Adrián Alejandro Vitón-Castillo 

ABSTRACT

The application of bioinformatics has facilitated the development of subareas of science and subspecialties, such as immunoinformatics. Immunoinformatics is extremely useful in vaccine development, the prediction of potential epitopes, and the simulation of immune responses. However, this requires the availability of curated, high-quality data that has been experimentally tested and validated. This study describes databases that are useful in epitope mapping and immune response simulation. The Immune Epitope Database and Analysis Resource, Universal Protein Resource, and Protein Data Bank are three such databases that are relevant for the use of immunoinformatics tools in epitope mapping and immune response simulation. These databases, together with others, whether general or specialized, form part of the biological data ecosystem on which most immunoinformatics tools for epitope prediction and immune response simulation are built and operate. All have their advantages and limitations, so the selection of one or the other should be based on an analysis of the needs and objectives of the researcher and their projects.

Keywords: Immunoinformatics; Vaccines; Epitopes; Antigens; Computational Biology; Vaccine Development.

RESUMEN

La aplicación de la bioinformática ha facilitado el desarrollo de subáreas de la ciencia y subespecialidades, como es el caso de la inmunoinformática. La inmunoinformática es de gran utilidad en el desarrollo de vacunas, la predicción de epítomos potenciales y la simulación de la respuesta inmune, sin embargo para ello, requiere de la disponibilidad de datos curados y de calidad, experimentalmente probados y validados. El presente estudio describe bases de datos de utilidad en el mapeo de epítomos y la simulación de respuestas inmunes. Immune Epitope Database and Analysis Resource, Universal Protein Resource y Protein Data Bank constituyen tres de estas bases de datos con relevancia para el uso de herramientas de inmunoinformática en el mapeo de epítomos y la simulación de respuestas inmunes. Estas bases de datos, en conjunto con otras, ya sean generales o especializadas, forman parte del ecosistema de datos biológicos sobre el cual se construyen y operan gran parte de las herramientas de inmunoinformática en la predicción de epítomos y simulación de la respuesta inmune. Todas presentan sus ventajas y limitantes, de ahí que la selección de una u otra debe realizarse en base del análisis de necesidades y objetivos del investigador y sus proyectos.

Palabras clave: Inmunoinformática; Vacunas; Epítomos; Antígenos; Biología Computacional; Desarrollo de Vacunas.

INTRODUCTION

Advances in Information and Communication Technologies have not only promoted the development of Computer Science and Information Technology but have also had a direct impact on all areas of life. The biomedical field is no exception, as it is rapidly evolving in many medical specialties.

The introduction of bioinformatics in Health Sciences, and more recently, Artificial Intelligence, has led to the development of other fields, such as immunoinformatics.

Immunoinformatics arises from applying bioinformatics to immunology through the application of computational techniques to study the cellular and molecular dynamics involved in the immune system processes.⁽¹⁾ It is therefore considered a branch of bioinformatics focused on *in silico* analysis, modeling, and managing complex immunological data.⁽²⁾ To this end, mathematical, statistical, biochemical, immunological, and other knowledge is integrated to map, analyze, and extrapolate the immune system's structure and function on a large scale.⁽¹⁾

To this end, immunoinformatics aims to transform immunological data from experimental processes into computable problems by developing models and algorithms. The purpose of applying these procedures is to translate them into meaningful information from an immunological and biological point of view, with applications in professional practice.⁽³⁾

The scope of immunoinformatics is broad, covering multiple areas of immunology research. Essentially, immunoinformatics tools and techniques enable the development of 'virtual laboratories' where *in silico* experiments can be conducted to generate and test hypotheses, design vaccine candidates, and simulate complex immunological scenarios that would be costly or impossible to study using traditional methods.⁽⁴⁾ Thanks to this, immunoinformatics acts not only as a support system for immunology research but can also serve as a guide for rational experimentation and minimizing resource expenditure.

However, there is a significant barrier to the development of immunoinformatics: the availability of quality biological data. To this end, public and private efforts have been made to develop biological databases and repositories for different purposes, which are helpful in this field. This study describes databases useful for epitope mapping and immune response simulation.

DEVELOPMENT

The Human Genome Project has profoundly changed the biological approach to the health-disease process. Its inception in 1990 catalyzed the development of omic medicine by mapping the human genome. The biological information from this project needed to be stored and, to a large extent, shared, giving rise to biological databases.⁽⁵⁾ Although, at one point, only a few genomes had been completely mapped or were available in draft form,^(6,7) there is now a wide variety of mapped genomes and proteins.

Just as there are genomic databases, there are also specialized databases, for example, those that store epitope mapping results. Three of these are described below.

Immune Epitope Database and Analysis Resource (IEDB)

The IEDB (<https://www.iedb.org/>) is a public, open-access resource funded by the National Institute of Allergy and Infectious Diseases (NIAID) in the United States. It is responsible for cataloging and disseminating experimental data on antibody epitopes and Major Histocompatibility Complex (MHC) ligands.⁽⁸⁾

This database includes nearly 2,2 million different epitopes. It covers various species, including humans and non-human primates, and contexts of infectious diseases, allergies, autoimmunity, and transplantology. Similarly, it is a repository of biological data and a platform of analysis tools that enable the prediction and study of epitopes.^(8,9)

This combination of curated experimental data and prediction tools makes the IEDB an active platform that drives research by facilitating the iterative cycle of computational prediction and experimental validation, which is essential for advancement in the field.⁽¹⁰⁾

Universal Protein Resource (UniProt)

UniProt (<https://www.uniprot.org/>) is a collaborative consortium of databases interconnected to a central repository. It is a freely accessible resource focused on data relating to proteins, their sequences, functions, and annotations.

UniProtKB (KnowledgeBase) is its repository or central database. Swiss-Prot and Translated EMBL Nucleotide Sequence Data Library (TrEMBL) are divided into two systems. Swiss-Prot contains manually annotated protein entries reviewed by experts, guaranteeing high-quality, curated information and, therefore, slower updates.⁽¹¹⁾ TrEMBL automatically annotates protein data by translating DNA coding sequences from nucleotide databases such as GenBank. This gives the community a database with greater coverage and faster access; however, the annotations may be less detailed than in Swiss-Prot.⁽¹²⁾

Another database within UniProt is UniProt Reference Clusters (UniRef), which contains sequential sets of UniProtKB grouped at different sequence levels, reducing redundancies and thus facilitating comparative analysis.^(13,14,15)

The third database within this ecosystem is the UniProt Archive (UniParc), a comprehensive archive containing all publicly available protein sequences, tracking all versions of the sequences. It has been cataloged as an indispensable biological data source for obtaining primary sequences of antigens of interest, their functions, location, variants, and post-translational modifications, which is crucial in the search and development of new vaccine candidates, as well as the simulation and interpretation of epitope predictions.⁽¹²⁾

Protein Data Bank (PDB)

PDB (<https://www.rcsb.org/>) is the world's primary database of information on three-dimensional structures of nucleic acids,

proteins, and complex macromolecular assemblies. Many techniques have been used to obtain this information, such as X-ray diffraction of crystals, nuclear magnetic resonance in solutions, and, more recently, cryo-electron microscopy.^(16,17)

This database is essential for analyzing and predicting B-cell conformational epitopes, as these depend on the folded 3D structure of the antigen. It is also crucial for modeling and simulating epitope-antibody or peptide-MHC interactions. Having an experimental structure from the PDB or a reliable 3D model derived from homology, or more recently from predictions based on artificial intelligence, is a determining factor for applying many bioinformatics tools for epitope prediction and simulation of interactions in the immune response.

FINAL CONSIDERATIONS

These databases, together with others, whether general or specialized, form part of the biological data ecosystem on which many of the immunoinformatics tools for epitope prediction and immune response simulation are built and operate. All have advantages and limitations, which is why selecting one or the other must be based on an analysis of the needs and objectives of the researcher and their projects.

BIBLIOGRAPHICAL REFERENCES

1. Tong JC, Ren EC. Immunoinformatics: Current trends and future directions. *Drug Discov Today* 2009;14:684-9. <https://doi.org/10.1016/j.drudis.2009.04.001>.
2. Patronov A, Doytchinova I. T-cell epitope vaccine design by immunoinformatics. *Open Biol* 2013;3:120139. <https://doi.org/10.1098/rsob.120139>.
3. Guarra F, Colombo G. Computational Methods in Immunology and Vaccinology: Design and Development of Antibodies and Immunogens. *J Chem Theory Comput* 2023;19:5315-33. <https://doi.org/10.1021/acs.jctc.3c00513>.
4. Vargas Aguilar AA, Fuentes Condori R, Vargas Aguilar AA, Fuentes Condori R. Estudios in Silico, Simulando Vida en un Entorno Virtual. *Gaceta Médica Boliviana* 2021;44:278-9. <https://doi.org/10.47993/gmb.v44i2.263>.
5. Hood L, Rowen L. The Human Genome Project: big science transforms biology and medicine. *Genome Medicine* 2013;5:79. <https://doi.org/10.1186/gm483>.
6. Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, et al. The complete genome of an individual by massively parallel DNA sequencing. *Nature* 2008;452:872-6. <https://doi.org/10.1038/nature06884>.
7. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, et al. The Diploid Genome Sequence of an Individual Human. *PLOS Biology* 2007;5:e254. <https://doi.org/10.1371/journal.pbio.0050254>.
8. Immune Epitope Database and Analysis Resource (IEDB) | NIAID: National Institute of Allergy and Infectious Diseases 2024. <https://www.niaid.nih.gov/research/immune-epitope-database>
9. Chun S, Muthu M, Gopal J, Paul D, Kim DH, Gansukh E, et al. The unequivocal preponderance of biocomputation in clinical virology. *RSC Adv* 2018;8:17334-45. <https://doi.org/10.1039/c8ra00888d>.
10. Zhang R, Sarmientor J, De Ocampo AL, Hernandez R. Fruit and vegetable self-billing system based on image recognition. *Data and Metadata*. 2024;3:397.
11. Immune Epitope Database – lji.org s. f. <https://www.lji.org/research/immune-epitope-database/> (accedido 30 de mayo de 2025).
12. UniProt. UniProt s. f. <https://www.uniprot.org> (accedido 30 de mayo de 2025).
13. EMBL-EBI. The UniProt databases | UniProt s. f. <https://www.ebi.ac.uk/training/online/courses/uniprot-quick-tour/the-uniprot-databases>
14. Chavula P, Kayusi F, Agura Kayus B. Application of Artificial Intelligence in Tree Care in Sub-Saharan Africa. *LatIA*. 2023; 1:325.
15. Lorusso T, Blaustein M, García Carrillo M. Bioinformatic analysis of the potential effects on human health of herbicides for alternative use to glyphosate and atrazine. *SCT Proceedings in Interdisciplinary Insights and Innovations*. 2024; 2:293. <https://doi.org/10.56294/piii2024293>

16. RCSB PDB: Homepage s. f. <https://www.rcsb.org/>

17. Protein Data Bank s. f. <https://www.ehu.eus/biofisica/juanma/rasmol/pdb.htm>

FUNDING

None.

CONFLICT OF INTEREST

None.

AUTHOR CONTRIBUTION

Conceptualisation: Adrián Alejandro Vitón Castillo, Carlos Alfredo Miló Valdés, Lidia Cecilia Pérez Acevedo.

Drafting – initial draft: Adrián Alejandro Vitón Castillo, Carlos Alfredo Miló Valdés, Lidia Cecilia Pérez Acevedo.

Writing – review and editing: Adrián Alejandro Vitón Castillo, Carlos Alfredo Miló Valdés, Lidia Cecilia Pérez Acevedo.