

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.

INTRODUCCIÓN

Los avances en el campo de las Tecnologías de la Información y las Comunicaciones han favorecido el desarrollo no solo de las Ciencias Informáticas o la Información, sino que ha tenido impacto directo en todas las áreas de la vida. El campo biomédico no se queda atrás, impactando en la rápida evolución de muchas especialidades médicas. La introducción de la bioinformática en Ciencias de la Salud, y más recientemente la Inteligencia Artificial, ha devenido en el desarrollo de otros campos, como por ejemplo la inmunoinformática.

La inmunoinformática surge debido a la aplicación de la bioinformática a la inmunología, mediante la aplicación de técnicas computacionales al estudio de las dinámicas celulares y moleculares implicadas en los procesos del sistema inmunitario.⁽¹⁾ Por ello, se considera como una rama de la bioinformática orientada al análisis *in silico*, la modelación y gestión de datos inmunológicos complejos.⁽²⁾ Para ello, se integran conocimientos matemáticos, estadísticos, bioquímicos, inmunológicos y otros, con el fin de mapear, analizar, y extrapolar, a gran escala, la estructura y función del sistema inmunitario.⁽¹⁾

Para ello, la inmunoinformática se orienta a transformar los datos inmunológicos resultantes de los procesos experimentales en problemas computacionales tratables mediante el desarrollo de modelos y algoritmos. El fin de la aplicación de estos procedimientos es traducir los mismos en información significativos desde el punto de vista inmunológico y biológico, con aplicación en los problemas de la práctica profesional.⁽³⁾

El alcance de la inmunoinformática es amplio, de ahí que abarque múltiples áreas de investigación en materia de inmunología. En esencia, la aplicación de herramientas y técnicas de inmunoinformática permite desarrollar “laboratorios virtuales”, donde se pueden realizar experimentos *in silico* para generar y probar hipótesis, diseñar candidatos vacunales, así como simular escenarios inmunológicos complejos, que por los métodos tradicionales resultarían costosos o imposibles de estudiar.⁽⁴⁾ Gracias a esto, la inmunoinformática actúa no solo como un sistema de apoyo para la investigación en inmunología, sino que puede servir de guía para una experimentación racional y minimizando el gasto de recursos.

Sin embargo, existe una importante barrera para el desarrollo de la inmunoinformática: la disponibilidad de datos biológicos de calidad. Para ello, se han hecho esfuerzos públicos y privados para el desarrollo de bases de datos biológicos y repositorios con diferentes fines, los cuales resultan de utilidad en este campo. El presente estudio describe bases de datos de utilidad en el mapeo de epítomos y la simulación de respuestas inmunes.

DESARROLLO

El proyecto Genoma Humano ha cambiado profundamente el enfoque biológico en el proceso salud – enfermedad. Su inicio en 1990 catalizó el desarrollo de la medicina ómica, mediante el mapeo del genoma humano. La información biológica resultante de este proyecto necesitó ser almacenada, y en gran medida compartida, dando inicio a las bases de datos biológicas.⁽⁵⁾ Aunque en un momento determinado solo se había mapeado completamente, o en forma de borrador de calidad algunos genomas,^(6,7) actualmente existe una amplia diversidad de genomas mapeados, así como proteínas.

De igual forma que existen bases de datos gnómicas, existen otras especializadas, por ejemplo, en almacenar los resultados del mapeo de epítomos. A continuación, se describen tres de ellas.

Immune Epitope Database and Analysis Resource (IEDB)

La IEDB (<https://www.iedb.org/>) es un recurso público, en acceso abierto, financiado por el *National Institute of Allergy and Infectious Diseases* (NIAID) de los Estados Unidos. Este, se encarga de la catalogación y diseminación de datos experimentales sobre epítomos de anticuerpos, así como ligandos del Complejo Mayor de Histocompatibilidad (MHC).⁽⁸⁾

Esta base de datos incluye cerca de 2,2 millones de epítomos diferentes. Abarca una amplia gama de especies, entre ellas, humanos, primates no humanos, y cubre contextos de enfermedades infecciosas, alergias, autoinmunidad y trasplantología. De igual forma, no solo constituye un repositorio de datos biológicos, sino que es una plataforma de herramientas de análisis que permite la predicción y estudio de epítomos.^(8,9)

Esta combinación de datos experimentales curados y herramientas de predicción convierte a la IEDB en una plataforma activa que impulsa la investigación, facilitando el ciclo iterativo de predicción computacional y validación experimental, esencial para el avance en el campo.⁽¹⁰⁾

Universal Protein Resource (UniProt)

UniProt (<https://www.uniprot.org/>) es un consorcio colaborativo de bases de datos interconectados a un repositorio central. Es un recurso de libre acceso centrado en datos referentes a proteínas, sus secuencias, funciones y anotaciones.

Su repositorio o base de datos central es UniProtKB (KnowledgeBase) Esta se divide en dos sistemas, Swiss-Prot y Translated EMBL Nucleotide Sequence Data Library (TrEMBL). Swiss-Prot contiene entradas de proteínas manualmente anotadas, revisadas por expertos, lo cual garantiza información de elevada calidad e información curada, y por ende, una actualización más lenta.⁽¹¹⁾ Por su parte, TrEMBL contiene datos de proteínas anotadas de forma automatizada, resultante de la traducción de secuencias codificantes de ADN, provenientes de bases de datos de nucleótidos como GenBank. Esto permite tener a disposición de la comunidad una base de datos con una mayor cobertura y de forma rápida, sin embargo, las anotaciones pueden ser menos detalladas que en Swiss-Prot.⁽¹²⁾

Otra de las bases de datos dentro de UniProt es la UniProt Reference Clusters (UniRef); esta contiene conjuntos secuenciales de UniProtKB agrupados a diferentes niveles de secuencias, lo cual facilita la reducción de redundancias y por ende una mayor facilidad de análisis comparativos.^(13,14,15)

La tercera base de datos dentro de este ecosistema es el UniProt Archive (UniParc), un archivo exhaustivo que contiene todas

las secuencias de proteínas disponibles públicamente, rastreando todas las versiones de las secuencias. Se ha catalogado como una fuente de datos biológicos indispensable para obtener secuencias primarias de antígenos de interés, sus funciones, localización, variantes y modificaciones post-traduccionales, lo cual resulta crucial en la búsqueda y desarrollo de nuevos candidatos vacunales, así como la simulación e interpretación de predicciones de epítomos.⁽¹²⁾

Protein Data Bank (PDB)

PDB (<https://www.rcsb.org/>) es la base de datos primaria mundial de información sobre estructuras tridimensionales de ácidos nucleicos, proteínas y ensamblaje macromolecular complejo. Para su obtención se han empleado una amplia gama de técnicas, como la difracción de rayos X de cristales, resonancia magnética nuclear en disoluciones y más recientemente, la crio-microscopía electrónica.^(16,17)

Esta base de datos es esencial para el análisis de epítomos conformacionales de Linfocitos B y su predicción, pues estos dependen de la estructura 3D plegada del antígeno; también es crucial para el modelado y simulación de interacciones epítomo-anticuerpo o péptido-MHC. Poseer una estructura experimental de la PDB o un modelo 3D fiable que derive de homología, o más recientemente de predicciones basadas en Inteligencia Artificial es un factor determinante para la aplicación de muchas de las herramientas inmunoinformáticas de predicción de epítomos y simulación de interacciones en la respuesta inmune.

CONSIDERACIONES FINALES

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.

REFERENCIAS BIBLIOGRÁFICAS

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 – Iji.org s. f. <https://www.iji.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>

FINANCIACIÓN

Ninguna.

CONFLICTO DE INTERESES

Ninguno.

CONTRIBUCIÓN DE AUTORÍA

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

Redacción – borrador inicial: Adrián Alejandro Vitón Castillo, Carlos Alfredo Miló Valdés, Lidia Cecilia Pérez Acevedo.

Redacción – revisión y edición: Adrián Alejandro Vitón Castillo, Carlos Alfredo Miló Valdés, Lidia Cecilia Pérez Acevedo.