

Capítulo 3

Modelado *in silico* y bioquímica computacional: Aportes al diagnóstico de enfermedades

Miriam Avendaño-Villegas¹

Carlos Romero-Díaz¹

Alexis Martínez-Barras¹

Idarh Matadamas-Ortíz²

Margarito Martínez-Cruz¹

DOI: <https://doi.org/10.61728/AE26004846>



¹ Tecnológico Nacional de México/ Instituto Tecnológico de Oaxaca/División de Estudios de Posgrado e Investigación. México.

² Tecnológico Nacional de México/ Instituto Tecnológico de Oaxaca/Departamento de Sistemas y computación. México.

Resumen

El modelado *in silico* y la bioquímica computacional se han consolidado como herramientas fundamentales para el estudio de biomoléculas y el análisis de enfermedades desde una perspectiva molecular. Estas aproximaciones permiten simular interacciones entre proteínas, ácidos nucleicos y compuestos químicos mediante métodos computacionales, facilitando la predicción de procesos biológicos sin necesidad de realizar experimentos invasivos o costosos en las etapas iniciales de la investigación. En los últimos años, el acceso a bases de datos moleculares, el desarrollo de algoritmos de predicción estructural y el aumento de la capacidad de cómputo han favorecido la aplicación de estas herramientas en el campo biomédico.

En este capítulo se presentan ejemplos recientes del uso de herramientas de bioquímica computacional desarrolladas por investigadores mexicanos, particularmente en el análisis de enfermedades endémicas del sur-sureste de México, como el dengue y otras arbovirosis. Se describen aplicaciones que incluyen minería de datos clínicos para diferenciar entre dengue con signos de alerta y dengue grave, así como modelos computacionales para la clasificación de enfermedades arbovirales mediante conjuntos de datos sintéticos derivados de la literatura científica.

Asimismo, se abordan otras aplicaciones del modelado molecular en el estudio de enfermedades hereditarias, metabólicas y neurodegenerativas, destacando cómo técnicas como el acoplamiento molecular, la dinámica molecular y el modelado estructural permiten analizar el impacto funcional de mutaciones genéticas, estudiar la interacción entre biomoléculas y evaluar posibles herramientas diagnósticas a nivel molecular.

Finalmente, se discute el papel de estas herramientas como apoyo al diagnóstico médico y a la investigación biomédica, enfatizando su relevancia en regiones con recursos limitados donde el acceso a pruebas diagnósticas especializadas puede ser restringido. En este contexto, la bioquímica computacional representa una alternativa prometedora para

fortalecer el estudio de enfermedades desde un enfoque molecular y contribuir al desarrollo de estrategias de diagnóstico temprano.

1. Introducción

El modelado *in silico* y la bioquímica computacional han transformado la manera en que se estudian las biomoléculas y los procesos biológicos asociados a la salud y la enfermedad. Tradicionalmente, el análisis bioquímico dependía principalmente de experimentos *in vitro* o *in vivo*, los cuales requieren reactivos, infraestructura especializada y largos periodos de experimentación. Sin embargo, el desarrollo de herramientas computacionales ha permitido complementar estos enfoques mediante simulaciones que reproducen el comportamiento molecular en entornos virtuales.

El término *in silico* hace referencia a experimentos realizados mediante sistemas computacionales, cuyo nombre deriva del silicio, elemento utilizado en la fabricación de microprocesadores. En este contexto, el modelado *in silico* permite simular interacciones entre biomoléculas, como proteínas, ácidos nucleicos y fármacos, utilizando información estructural disponible en bases de datos especializadas.

En términos generales, estas herramientas permiten analizar grandes volúmenes de información molecular, genética o clínica mediante métodos computacionales, facilitando el estudio de mecanismos bioquímicos complejos y la identificación de patrones asociados con distintos procesos biológicos. Estas aproximaciones pueden contribuir a la generación de hipótesis, la predicción de interacciones moleculares y el análisis de datos relacionados con enfermedades, complementando los estudios experimentales y clínicos.

En México, el uso de herramientas computacionales aplicadas a la bioquímica y la biomedicina ha experimentado un crecimiento considerable en la última década, impulsado por la disponibilidad de datos genómicos y por el desarrollo de infraestructura computacional en centros de investigación y universidades.

En particular, estas herramientas han sido aplicadas al estudio de enfermedades infecciosas y metabólicas relevantes para la salud pública en distintas regiones del país. En el contexto del sur-sureste de México,

donde enfermedades como el dengue, zika y chikungunya representan un desafío constante para los sistemas de salud, el uso de modelos computacionales ha permitido desarrollar herramientas que facilitan la interpretación de datos clínicos y moleculares asociados a estas patologías (1, 2, 3).

De esta manera, la bioquímica computacional no solo contribuye al estudio básico de las biomoléculas, sino que también permite integrar información genética, estructural y clínica para comprender mejor los mecanismos moleculares de diversas patologías y apoyar el desarrollo de herramientas diagnósticas.

El proceso general mediante el cual las herramientas computacionales permiten analizar biomoléculas y apoyar el diagnóstico de enfermedades se resume en la Figura 1.

2. Herramientas computacionales y análisis de datos en el estudio de enfermedades

2.1 Bases de datos biomoleculares y recursos computacionales

El desarrollo del modelado *in silico* ha sido posible gracias a la disponibilidad de bases de datos que almacenan información estructural, genética y funcional de biomoléculas. Estas bases de datos constituyen el punto de partida para numerosos análisis computacionales, ya que permiten obtener las secuencias o estructuras tridimensionales necesarias para realizar estudios de acoplamiento molecular, dinámica molecular o predicción estructural.

Entre las bases de datos más utilizadas se encuentran Protein Data Bank (PDB), UniProt, BRENDA, PubChem y los recursos del National Center for Biotechnology Information (NCBI). Estas plataformas proporcionan información sobre secuencias primarias, estructuras tridimensionales, funciones enzimáticas, interacciones moleculares y características fisicoquímicas de diversas biomoléculas (4, 5).

Por ejemplo, el Protein Data Bank almacena estructuras tridimensionales determinadas experimentalmente mediante cristalografía de rayos

X, resonancia magnética nuclear o microscopía electrónica (5). Por su parte, UniProt integra información detallada sobre proteínas, incluyendo su función biológica, dominios estructurales, variantes asociadas con enfermedades y localización subcelular.

Las bases de datos biomoleculares constituyen el punto de partida para la mayoría de los análisis computacionales, ya que proporcionan información estructural, funcional y genética necesaria para el modelado molecular. En la Tabla 1 se presentan algunas de las bases de datos más utilizadas en estudios de bioquímica computacional.

2.2 Herramientas de modelado molecular

Dentro del modelado *in silico* existen diversas herramientas que permiten estudiar el comportamiento de biomoléculas desde diferentes perspectivas. Entre las más utilizadas se encuentran el acoplamiento molecular, la dinámica molecular y los programas de visualización estructural.

Acoplamiento molecular (Molecular Docking)

El acoplamiento molecular es una técnica computacional utilizada para predecir la interacción entre dos moléculas, generalmente una proteína y un ligando. El objetivo principal es estimar la afinidad de unión y la orientación espacial del ligando dentro del sitio activo de la proteína.

Programas como AutoDock, AutoDock Vina y HEX (Tabla 2) permiten realizar estos cálculos mediante algoritmos que exploran múltiples configuraciones posibles entre las moléculas. Estos programas estiman la energía de interacción entre el ligando y el receptor, permitiendo identificar configuraciones potencialmente favorables desde el punto de vista termodinámico (6, 7).

Por ejemplo, AutoDock Vina es un software ampliamente utilizado debido a su precisión submicromolar y su capacidad para evaluar múltiples conformaciones del ligando dentro del sitio activo de una proteína. Este tipo de herramientas ha sido utilizado para estudiar proteínas implicadas en diversas enfermedades, incluyendo receptores asociados con procesos de proliferación celular y desarrollo de cáncer.

Dinámica molecular

La dinámica molecular permite simular el movimiento de átomos y moléculas a lo largo del tiempo mediante ecuaciones físicas que describen las fuerzas que actúan entre ellos.

Programas como GROMACS o AMBER permiten estudiar la estabilidad de complejos moleculares, como interacciones proteína-ligando, anticuerpo-antígeno o lectina-carbohidrato. Estas simulaciones se realizan generalmente en escalas de tiempo del orden de nanosegundos a microsegundos (8, 9).

Mediante estos análisis es posible evaluar parámetros como estabilidad estructural, cambios conformacionales y fluctuaciones atómicas, lo que permite comprender mejor el comportamiento dinámico de las biomoléculas (6).

Programas de visualización molecular

Permiten representar gráficamente estructuras biomoleculares y analizar sus características estructurales.

Entre los más utilizados se encuentran UCSF Chimera, PyMOL, VMD y Avogadro, los cuales permiten visualizar proteínas, ácidos nucleicos y complejos moleculares utilizando distintos formatos de representación, como cintas (ribbons), superficies moleculares o modelos de bolas y varillas.

Estas herramientas facilitan la interpretación de resultados obtenidos mediante simulaciones computacionales y permiten analizar interacciones moleculares, distancias entre átomos, enlaces de hidrógeno y cambios estructurales en biomoléculas.

2.3 Aplicaciones del análisis computacional de datos clínicos en enfermedades infecciosas

Los modelos computacionales también han sido utilizados para analizar enfermedades infecciosas relevantes para la salud pública en el contexto regional del sur-sureste de México. En este sentido, el análisis computacional de grandes conjuntos de datos clínicos permite identificar patrones asociados con la progresión de la enfermedad y apoyar la toma de decisiones médicas.

Un ejemplo de ello corresponde al estudio realizado en el Hospital “Aurelio Valdivieso” en Oaxaca, donde se emplearon técnicas de minería de datos (data mining) y aprendizaje automático (machine learning) para diferenciar entre dengue con signos de alerta y dengue grave (1).

En este estudio se analizaron datos clínicos de 1341 pacientes, utilizando algoritmos de clasificación para identificar patrones asociados con la progresión de la enfermedad. Los resultados mostraron precisiones de hasta 82.9 % en la clasificación entre dengue con signos de alerta y dengue grave, destacando como predictores principales la hematemesis, la ascitis y el edema (1).

De manera complementaria, otros estudios han desarrollado modelos computacionales para la clasificación de enfermedades arbovirales que presentan síntomas clínicos similares. Estas enfermedades, como dengue, zika y chikungunya, son transmitidas por mosquitos del género *Aedes* y frecuentemente presentan manifestaciones clínicas iniciales indistinguibles, lo que dificulta su diagnóstico en las primeras etapas de la infección.

En un estudio reciente se desarrolló un marco computacional para la clasificación de enfermedades arbovirales utilizando conjuntos de datos sintéticos derivados de información epidemiológica reportada en la literatura científica (2). Para ello se generó un conjunto de datos con aproximadamente 28 000 registros clínicos y 67 síntomas estandarizados, incluyendo dengue, zika, chikungunya e influenza como control negativo.

Este conjunto de datos permitió entrenar distintos algoritmos de clasificación, entre ellos perceptrón multicapa, redes neuronales, máquinas de vectores de soporte y árboles de decisión. Los resultados mostraron niveles de precisión cercanos al 92% y valores de AUC superiores a 0.98, lo que demuestra la viabilidad de estos modelos como herramientas de apoyo en el diagnóstico preliminar de enfermedades arbovirales (2).

Posteriormente, se desarrolló una base de datos sintética ampliada derivada de la literatura científica, la cual integró más de 48 000 registros clínicos estandarizados correspondientes a dengue, zika, chikungunya e influenza (3). Esta base de datos constituye un recurso útil para el entrenamiento y validación de futuros modelos computacionales orientados al diagnóstico temprano de enfermedades infecciosas, particularmente en regiones donde la disponibilidad de pruebas de laboratorio puede ser limitada.

2.4 Aplicaciones del modelado in silico en el diagnóstico de enfermedades

El modelado estructural de proteínas representa una herramienta importante dentro del conjunto de métodos computacionales aplicados al estudio de enfermedades humanas. Cuando se identifican mutaciones mediante técnicas de secuenciación genética, no siempre es posible determinar inmediatamente si estas variaciones son responsables de una enfermedad o si corresponden a cambios sin relevancia clínica. En este contexto, el modelado in silico permite analizar cómo estas mutaciones pueden modificar la estructura tridimensional de las proteínas, afectando su estabilidad, su capacidad de interacción con otras biomoléculas o su función biológica (10, 11).

El estudio de enfermedades hereditarias constituye uno de los campos donde estas herramientas han demostrado gran utilidad. Un estudio reciente realizado en un laboratorio de diagnóstico clínico mostró que el análisis estructural computacional puede apoyar la clasificación de variantes genéticas detectadas en pacientes, permitiendo diferenciar entre mutaciones potencialmente patogénicas y cambios benignos en diversas proteínas humanas. Al integrar información genómica con modelado estructural, los investigadores lograron proponer mecanismos moleculares que explican cómo determinadas mutaciones pueden desestabilizar proteínas o alterar su funcionamiento normal. Estos resultados evidencian que la bioquímica computacional no solo tiene aplicaciones en la investigación básica, sino que también representa una herramienta útil para mejorar la interpretación clínica de la información genómica y fortalecer el diagnóstico de distintas enfermedades (10).

Un ejemplo representativo corresponde al estudio del síndrome de Lynch, una enfermedad hereditaria asociada con un mayor riesgo de desarrollar cáncer colorrectal y otros tipos de cáncer. Esta patología está relacionada con mutaciones en genes responsables de la reparación del ADN, entre ellos el gen MSH2, cuya proteína participa en la corrección de errores durante la replicación genética. Mediante cálculos estructurales y simulaciones realizadas in silico, los investigadores analizaron cómo cambios específicos en la secuencia genética pueden alterar la estabili-

dad estructural de esta proteína. Los resultados mostraron que muchas de las mutaciones asociadas con el síndrome de Lynch no eliminan directamente la función de la proteína, sino que provocan alteraciones en su plegamiento estructural, generando inestabilidad y favoreciendo su degradación celular antes de que pueda cumplir su función en los mecanismos de reparación del ADN (11).

El modelado *in silico* también ha sido utilizado para estudiar biomoléculas relacionadas con enfermedades metabólicas. Un ejemplo relevante es la diabetes mellitus, una enfermedad caracterizada por niveles elevados de glucosa en sangre y que representa uno de los principales problemas de salud pública a nivel mundial. En este contexto, diversos estudios han explorado el uso de biosensores basados en aptámeros, pequeñas secuencias de ADN o ARN capaces de reconocer moléculas específicas. Mediante simulaciones de docking y dinámica molecular, se ha analizado la interacción entre aptámeros y la molécula de insulina, evaluando parámetros como la energía de unión y la estabilidad estructural del complejo molecular. Los resultados sugieren que estos aptámeros presentan interacciones más específicas con la insulina que con otras moléculas relacionadas, lo que indica su potencial aplicación en el desarrollo de biosensores para la detección temprana de diabetes tipo 2 (12).

De forma similar, en enfermedades neurodegenerativas, el modelado molecular ha permitido explorar posibles compuestos con actividad terapéutica. Trastornos como la enfermedad de Alzheimer y la enfermedad de Parkinson se caracterizan por la degeneración progresiva de neuronas y la acumulación de proteínas mal plegadas en el sistema nervioso. Debido a la complejidad de estos procesos, la identificación de compuestos con potencial terapéutico suele requerir largos periodos de investigación experimental. En este sentido, las herramientas de bioquímica computacional permiten evaluar de manera preliminar la interacción entre compuestos naturales y proteínas asociadas con estas enfermedades.

Un estudio reciente analizó mediante técnicas de docking y dinámica molecular el compuesto natural karanjin, evaluando su interacción con proteínas relacionadas con procesos neurodegenerativos. Las simulaciones mostraron interacciones favorables con varios blancos moleculares implicados en enfermedades como Alzheimer y Parkinson, lo que sugiere

su potencial como molécula candidata para el desarrollo de futuros fármacos. Estos resultados demuestran cómo el modelado computacional puede acelerar las primeras etapas del descubrimiento de fármacos al permitir evaluar múltiples compuestos de manera rápida antes de su validación experimental (13).

En conjunto, estos ejemplos muestran que el modelado *in silico* no sustituye el diagnóstico clínico ni las pruebas experimentales, pero proporciona información adicional sobre los mecanismos moleculares de las enfermedades, contribuyendo al desarrollo de herramientas diagnósticas y terapéuticas más precisas.

3. Conclusiones

El modelado *in silico* y la bioquímica computacional se han consolidado como herramientas fundamentales para el estudio de biomoléculas y el análisis de enfermedades desde una perspectiva molecular. Estas metodologías permiten simular interacciones entre proteínas, ácidos nucleicos y compuestos químicos, facilitando la comprensión de procesos biológicos complejos y reduciendo la necesidad de experimentación inicial en laboratorio.

Los ejemplos presentados en este capítulo muestran cómo las herramientas computacionales pueden contribuir al estudio y diagnóstico de diversas enfermedades. En el caso de enfermedades infecciosas como el dengue y otras arbovirosis, el análisis computacional de datos clínicos permite identificar patrones asociados con la progresión de la enfermedad y apoyar la toma de decisiones médicas, particularmente en regiones donde la disponibilidad de pruebas diagnósticas puede ser limitada. De manera complementaria, el modelado estructural permite analizar el impacto funcional de mutaciones genéticas y estudiar interacciones moleculares relevantes para enfermedades hereditarias, metabólicas y neurodegenerativas.

En conjunto, estas aproximaciones evidencian el potencial de la bioquímica computacional como una herramienta de apoyo para la investigación biomédica y el diagnóstico molecular. Si bien estas metodologías no sustituyen las pruebas clínicas ni los estudios experimentales, aportan información valiosa que permite comprender mejor los mecanismos moleculares asociados con diversas patologías.

En el contexto de la salud pública en regiones como el sur-sureste de México, donde enfermedades infecciosas como el dengue representan un problema importante, el desarrollo y aplicación de herramientas computacionales puede contribuir al fortalecimiento de las estrategias de diagnóstico temprano y al diseño de nuevas aproximaciones para el estudio de enfermedades. A medida que continúe creciendo la disponibilidad de datos genómicos, estructurales y clínicos, es probable que el modelado *in silico* desempeñe un papel cada vez más relevante en la investigación biomédica y en el desarrollo de estrategias diagnósticas más precisas.

4. Referencias

1. Meixueiro VHO, Hernández-Huerta MT, Pérez-Campos Mayoral LC, Cruz-Parada E, Ortiz JOL, Valdez NG, et al. Using data mining to differentiate dengue with warning signs from severe dengue: a predictive model from Oaxaca, Mexico. *Am J Trop Med Hyg.* 2025;113(3):508-15.
2. Cruz-Parada E, Vivar-Estudillo G, Pérez-Campos Mayoral L, Hernández-Huerta MT, Pérez-Santiago AD, Romero-Díaz C, et al. Proof-of-concept machine learning framework for arboviral disease classification using literature-derived synthetic data: methodological development preceding clinical validation. *Healthc (Basel).* 2026;14(2):247.
3. Cruz-Parada E, Vivar-Estudillo G, Pérez-Campos Mayoral L, Hernández-Huerta MT, Pérez-Santiago AD, Romero-Díaz C, et al. Synthetic and encoded database of dengue, Zika, chikungunya, and influenza derived from the literature. *Data (Basel).* 2026;11(2):33.
4. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 2021;49(D1):D10-7.
5. Burley SK, Berman HM, Bhikadiya C, Bi C, Chen L, Di Costanzo L, et al. Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Res.* 2019;47(D1):D520-8.
6. Vidal-Limon A, Aguilar-Toalá JE, Liceaga AM. Integration of molecular docking analysis and molecular dynamics simulations for

- studying food proteins and bioactive peptides. *J Agric Food Chem.* 2022;70(4):934-43.
7. Paggi JM, Pandit A, Dror RO. The art and science of molecular docking. *Annu Rev Biochem.* 2024;93:389-410.
 8. Śledź P, Caflisch A. Protein structure-based drug design: from docking to molecular dynamics. *Curr Opin Struct Biol.* 2018;48:93-102.
 9. Naqvi AAT, Mohammad T, Hasan GM, Hassan MI. Advancements in docking and molecular dynamics simulations towards ligand-receptor interactions and structure-function relationships. *Curr Top Med Chem.* 2018;18(20):1755-68.
 10. McEntagart M, Smith MJ, Evans DG, Woodward ER, Raymond FL, Lalloo F, et al. Assessing the clinical utility of protein structural analysis in genomic variant classification: experiences from a diagnostic laboratory. *Genome Med.* 2022;14:83.
 11. Nielsen SV, Stein A, Dinitzen AB, Papaleo E, Tatham MH, Poulsen EG, et al. Predicting the impact of Lynch syndrome-causing missense mutations from structural calculations. *PLoS Genet.* 2017;13(4):e1006739.
 12. Mardian Y, Kurniawan F, Haryanto A, Pratiwi D, Suryani A, Wibowo FR, et al. In silico study of aptamer–insulin interaction for biosensor development in diabetes detection. *Indones J Chem.* 2024;24(3).
 13. Ogunyemi OM, Afolabi SO, Akindele AJ, Ojo OA, Ojo AB, Adewumi OB, et al. In silico investigation of karanjin as a potential therapeutic agent against neurodegenerative diseases. *Molecules.* 2022;27(9):2834.

Tabla 1. Bases de datos biomoleculares utilizadas en bioquímica computacional.

Base de datos	Información principal y uso	Sitio web
PDB	Estructuras tridimensionales de proteínas y complejos biomoleculares utilizadas en estudios estructurales y docking.	rcsb.org
UniProt	Secuencias y funciones de proteínas, dominios estructurales y variantes asociadas a enfermedades.	uniprot.org
BRENDA	Información funcional de enzimas, actividad catalítica y parámetros bioquímicos.	brenda-enzymes.org
PubChem	Estructuras y propiedades de compuestos químicos utilizados como ligandos en estudios moleculares.	pubchem.ncbi.nlm.nih.gov
NCBI	Bases de datos genómicas y herramientas para análisis de secuencias y genes.	ncbi.nlm.nih.gov

Tabla 2. Programas de acoplamiento molecular utilizados en bioquímica computacional.

Programa	Aplicación principal	Sitio web
AutoDock	Predicción de interacciones proteína-ligando mediante estimación de energía de unión	autodock.scripps.edu
AutoDock Vina	Versión optimizada de AutoDock con mayor velocidad y precisión en estudios de docking	vina.scripps.edu
HEX	Acoplamiento proteína-proteína y proteína-ligando mediante algoritmos basados en transformadas rápidas	hex.loria.fr

Figura 1. Esquema general de las principales aproximaciones computacionales aplicadas al estudio de enfermedades. La integración de datos biomoleculares, clínicos y epidemiológicos mediante modelado in silico, minería de datos y aprendizaje automático permite estudiar mecanismos moleculares, identificar patrones asociados con enfermedades y contribuir al diagnóstico y al desarrollo de estrategias terapéuticas..

