

SWISS-MODEL es un generador de modelos estructurales de proteínas cuyas estructuras aún no están depositadas en el PDB

Rodrigo Arreola-Barroso¹ , Verónica Quintero-Hernández² , Jesús Muñoz-Rojas³ , Paul Gaytán¹ , América Rivera-Urbalejo^{3,4} , María Rosete-Enríquez⁵ , Lucía Martínez-Martínez⁶ , Victor Rivelino Juárez-González^{7*} 

¹Departamento de Ingeniería Celular y Biocatálisis, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Av. Universidad 2001, Col. Chamilpa, Cuernavaca, Morelos 62210 México. ²CONACYT-Laboratorio de Ecología Molecular Microbiana (LEMM), Centro de Investigaciones en Ciencias Microbiológicas (CICM), Instituto de Ciencias (IC), Benemérita Universidad Autónoma de Puebla (BUAP), Puebla, México. ³Grupo Ecology and Survival of Microorganisms, LEMM, CICM, IC, BUAP, Puebla, México. Edificio IC-11, Ciudad Universitaria, San Manuel, Puebla, México. C. P. 72570. ⁴Facultad de Estomatología, BUAP, Puebla, México. ⁵Laboratorio de Macromoléculas, Facultad de Ciencias Biológicas, BUAP, Puebla, México. ⁶Laboratorio de Biología Molecular, Centro de Investigación, Facultad de Medicina UNAM-UABJO, Universidad Autónoma “Benito Juárez” de Oaxaca, Oaxaca, México. ⁷Departamento de Medicina Molecular y Bioprocesos, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Av. Universidad 2001, Col. Chamilpa, Cuernavaca, Morelos, México. CP: 62210.

*Email autor para correspondencia: rivelino.juarez@ibt.unam.mx

Recibido: 19 abril 2022. **Aceptado:** 12 septiembre 2022

RESUMEN

En los últimos 15 años, los científicos han mejorado la habilidad para generar modelos estructurales de las proteínas, cuya estructura tridimensional (3D) se desconoce, gracias al crecimiento del número de estructuras depositadas en la base de datos *Protein Data Bank* (PDB). En la actualidad, uno de los métodos más usados y más rápidos para la generación de modelos estructurales es el servidor bioinformático SWISS-MODEL, creado para el modelado por homología de estructuras 3D, que comparten hasta 30% de identidad en su secuencia de aminoácidos con otras proteínas de estructura conocida. La calidad de los modelos resultantes se evalúa con varios parámetros bioquímicos (por ejemplo: QMEAN, RAMACHANDRAN plot). El modelo puede mejorarse al incluir el SWISS-MODEL en una línea de trabajo, seguido del servidor CHARMM-GUI y el programa GROMACS. Mientras el servidor CHARMM-GUI aplica al modelo producido, parámetros de un campo de fuerza para crear un sistema proteína-agua, bajo condiciones relevantes biológicamente, apto para simulación, el programa GROMACS minimiza la energía del modelo hasta alcanzar una estructura energéticamente estable, más cercana a como se encuentra en solución o en el sistema biológico. Los modelos generados por esta línea de trabajo pueden ser analizados a detalle por los biólogos estructurales en programas para visualización, como PyMOL, para obtener un mayor entendimiento del fenómeno biológico bajo estudio.

Palabras clave: CHARMM-GUI; GROMACS; PyMOL; SWISS-MODEL.

ABSTRACT

In the last 15 years, scientists have improved their ability to generate structural models for proteins, whose three-dimensional (3D) structure is unknown, thanks to the growing number of structures stored in the Protein Data Bank (PDB) database. Nowadays, one of the most widely used and fastest methods for model generation is SWISS-MODEL, a structural bioinformatics web server created for homology modeling of 3D protein structures, sharing down to 30% identity of their amino acid sequences. The quality of the resulting models is assessed with different biochemical parameters (QMEAN, RAMACHANDRAN plot). The model can be refined by including SWISS-MODEL in a pipeline, followed by the CHARMM-GUI server and the GROMACS programs. While the CHARMM-GUI server applies force field parameters to generate a protein-water system, under biologically relevant conditions, suitable for simulation, the GROMACS program minimizes the model energy to energetically stable structure, closer to that found in solution or in a biological system. The models generated by such pipelines can be analyzed in detail by structural biologists in visualization programs, such as PyMOL, to better understand the biological phenomenon under study.

Keywords: CHARMM-GUI; GROMACS; PyMOL; SWISS-MODEL.

INTRODUCCIÓN

Estudiar los procesos estructurales, moleculares y bioquímicos mediante los cuales los seres vivos y los virus logran que sus proteínas trabajen de manera coordinada en el espacio y tiempo con otras moléculas con las que interactúan (ácidos nucleicos, carbohidratos, lípidos, etc), es fundamental para entender las funciones que permiten la vida en los organismos, microorganismos y la propagación de los virus.

La relevancia de la estructura proteica para el entendimiento de los fenómenos biológicos a nivel molecular se ve reflejado en el número de estructuras depositadas en el banco de datos de proteínas (*Protein Data Bank*), abreviado como PDB. Esta base de datos es el almacén de información estructural a nivel atómico, donde

se encuentran depositadas todas las estructuras tridimensionales de las macromoléculas biológicas determinadas por la comunidad científica, la cual incluye a proteínas, ácidos nucleicos, carbohidratos y lípidos. Asimismo, la especie con mayor número de estructuras depositadas en el PDB hasta hoy es la humana o (*Homo sapiens*) (https://www.rcsb.org/stats/explore/scientific_name_of_source_organism), cuyo conocimiento ha permitido entender las enfermedades humanas a nivel atómico. Entre los grupos menos representados en este depósito de estructuras se encuentran las bacterias y los virus; en ambos casos, el conocimiento de las estructuras proteicas tiene relevancia en áreas: como la salud, la agricultura, la producción de vacunas, etc.



Particularmente, en el caso de los virus resultaría interesante poder conocer la estructura y la función de la mayoría de sus proteínas; sólo por citar un caso, se desconocen aún los mecanismos moleculares por los cuales el virus del Zika y sus proteínas causan defectos congénitos en los fetos de las mujeres embarazadas infectadas [1] (Figura 1). Una estrategia para conocer estos mecanismos, sería

generar modelos tridimensionales de proteínas del virus del Zika a partir de estructuras de otros virus o microorganismos, previamente depositadas en el PDB, y así obtener información estructural y funcional para develar como se llevan a cabo estos defectos congénitos en los fetos.

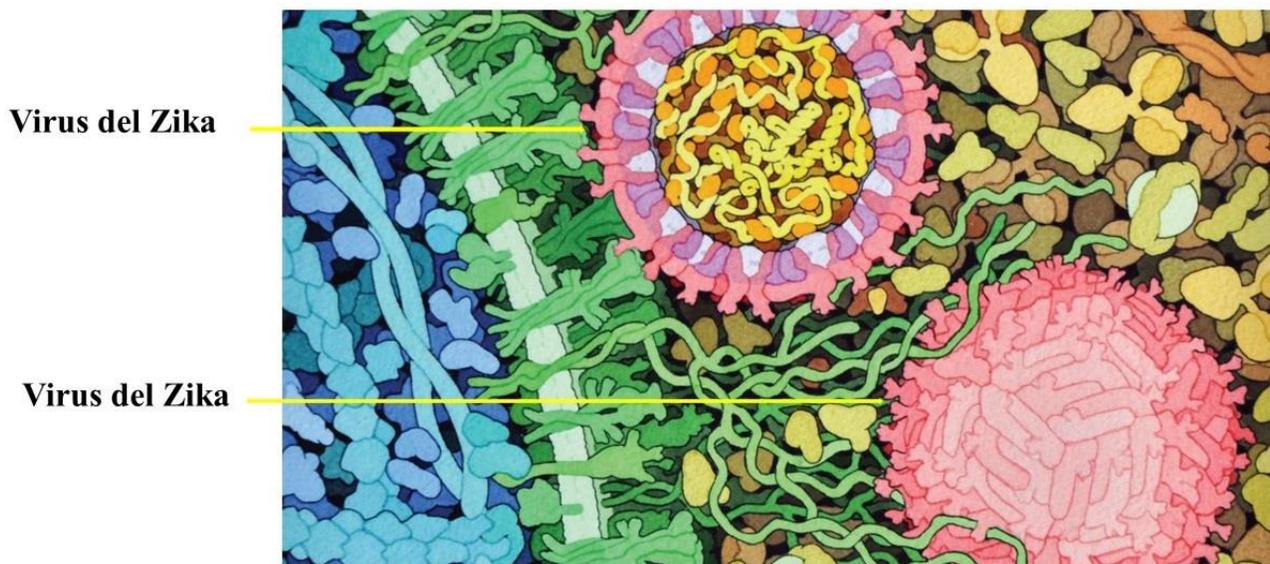


Figura 1. Virus del Zika. Pintura basada en evidencia experimental sobre el grado de empaquetamiento que tienen las proteínas del virus del Zika infectando células humanas. Imagen generada por David S. Goodsell. RCSB Protein Data Bank; doi: 10.2210/rcsb_pdb/goodsell-gallery-015.

La importancia de distintas técnicas para poder determinar la estructura tridimensional de las macromoléculas

Desde que, en 1949, Linus Pauling usó la estructura de la hemoglobina, la proteína que transporta el oxígeno en la sangre, para explicar la causa de la anemia falciforme, quedó clara la utilidad que tendría el conocimiento estructural de las proteínas para comprender su función biológica [2]. La estructura tridimensional (3D)

de las proteínas devela información muy valiosa para entender su función; por ejemplo, podemos deducir los residuos de aminoácidos importantes para su interacción con otras moléculas (incluidas otras proteínas), su regulación y estabilidad. En general, el conocimiento de la estructura es importante para una comprensión integral de los sistemas biológicos. Gracias al progreso constante de las metodologías para determinar las estructuras

3D de las macromoléculas (proteínas, ácidos nucleicos, carbohidratos, etc.), hoy conocemos, con detalle atómico, la estructura de 191,869 proteínas depositadas en el PDB [3,4]. Esta riqueza de información ha sido generada por técnicas de frontera como: la difracción de rayos X sobre cristales de proteínas, que aprovecha que cuando un haz de rayos X pasa por un monocristal de proteína crea un patrón único que depende del arreglo de sus átomos [5,6]; la resonancia magnética nuclear (RMN), que utiliza la información originada de las transiciones de espín nuclear, que ocurren en los núcleos de algunos átomos cuando se someten a campos magnéticos y cuya frecuencia de transición depende del ambiente químico en el que se encuentra el átomo [7,8] y la crio-microscopía electrónica o cryo-em, que construye fotografías de las macromoléculas a detalle atómico a partir del análisis conjunto de múltiples imágenes producidas por un haz de electrones interactuando con una disolución de la macromolécula congelada a temperaturas criogénicas. [9,10].

El aporte de estas técnicas se debe al avance tecnológico que ha permitido mejorar la resolución y disminuir el ruido con equipos más potentes y sensores más sensibles, también al uso de nuevas metodologías para analizar la información generada a partir del experimento. Además, el uso combinado de las tres técnicas descritas anteriormente ha contribuido para conocer las estructuras de las macromoléculas difíciles de determinar.

Métodos de modelado de las proteínas

El gran número de estructuras generadas, a partir de datos experimentales, y el gran avance del poder de cálculo de las computadoras en los últimos 15 años han hecho posible el uso de computadoras para la predicción estructural de las proteínas cuya configuración tridimensional no está depositada en el PDB. En la actualidad, los métodos computacionales han logrado predecir las estructuras 3D e interacciones entre proteínas, estos métodos se pueden separar en 3 categorías: el primero es el modelado basado en plantillas (*Template Based Modeling*) o templados presentes en el PDB, el segundo es acoplamiento proteína-proteína (*Protein-Protein Docking*), que predice la estructura del complejo proteico a partir de las estructuras de las proteínas individuales y, finalmente, el tercero es el modelado híbrido (*Hybrid/Integrative Modeling*), en donde los datos experimentales se fusionan con los cálculos usando potenciales de interacción entre residuos [11]. Todos estos métodos contribuyen a entender el funcionamiento de las macromoléculas y sus interacciones con otras macromoléculas y su entorno acuoso. Aquí, particularmente, hablaremos del modelado basado en plantillas o templados del que somos expertos.

Modelado de proteínas mediante SWISS-MODEL (Web-Based Integrated Service Dedicated to Protein Structure Homology Modelling)

Para poder generar un modelo estructural, los

programas usados (SWISS-MODEL [12], I-TASSER [13], etc.); toman la secuencia primaria de una proteína con estructura desconocida (proteína blanco) e identifican una o varias proteínas con un origen evolutivo común (proteínas homólogas) con estructura conocida en el PDB, utilizando el mayor porcentaje de identidad con la proteína blanco (proteínas templado). Después, se ajustan los aminoácidos de la proteína blanco para que coincidan con la estructura de la proteína templado. Terminado este proceso, se evalúa la concordancia entre la estructura del templado y la proteína blanco, para ello, se usan varios parámetros que permiten calificar el porcentaje de la secuencia blanco que puede ser modelada usando la estructura y la calidad de dicho modelado.

Al aumentar la accesibilidad a bases de datos como el PDB, y el número de estructuras que almacenan, los investigadores de la biología estructural de macromoléculas, pueden identificar mejores plantillas para las proteínas para las que solo se conoce su secuencia de aminoácidos. Esta condición ha vuelto el modelado estructural basado en plantillas o plantillas una importante fuente para generar modelos estructurales con detalle atómico. Estas metodologías usan el gran poder de cálculo computacional disponible hoy en día para identificar el templado y ajustar a esta estructura depositada en el PDB, una secuencia primaria de aminoácidos, con la que comparte un porcentaje de identidad en secuencia arriba de 30%. La calidad del modelo resultante depende del porcentaje de identidad entre las

secuencias de aminoácidos comparadas y el algoritmo empleado.

El programa predictor de estructuras de acceso libre SWISS-MODEL (liga de acceso <https://swissmodel.expasy.org/>; [12,14]) (Figura 2), pionero en el campo del modelado automatizado desde hace más de 25 años, cuenta con una biblioteca de plantillas o plantillas que se actualiza continuamente de acuerdo con las estructuras depositadas en el PDB. Además, proporciona estimaciones de calidad en varias etapas del proceso de modelado. Por ejemplo, indica el porcentaje de identidad y cobertura entre la secuencia a modelar y los diferentes plantillas y, una vez que se genera el modelo, la función de QMEAN evalúa la calidad del modelo a escala global y local [15]. La calidad aceptable del QMEAN para los modelos se encuentra entre 0 y -4.0. El programa SWISS-MODEL (Figura 3), es uno de los modeladores de proteínas más utilizados en la actualidad por científicos estructurales para la generación de modelos tridimensionales basados en homología, debido a que presenta varias ventajas sobre otros modeladores en el área, entre las que destacan: generación de modelos tridimensionales en poco tiempo (hay casos que con solo unos minutos basta), colocación de los puentes disulfuro, si presenta alguno la proteína analizada, predicción de la estructura cuaternaria en la proteína analizada y visualización de gráficos de alta calidad. En contraste, otros modeladores como I-TASSER (liga de acceso <https://zhanggroup.org/I-TASSER/>; [16]) y Phyre2 (liga de acceso <http://www.sbg.bio.ic.ac.uk/phyre2/html/page>

[.cgi?id=index](#); [17]), tienen tiempos de espera de varias horas, o incluso días, para producir modelos de calidad comparable a los generados con SWISS-MODEL. Sin embargo, SWISS-MODEL produce modelos usando plantados con al menos 30% de identidad con la proteína blanco, por lo que generalmente, la calidad del modelo generado es baja; en cuyo caso, el

servidor I-TASSER, puede ser capaz de generar modelos aceptables al usar múltiples plantados para generar un modelo quimérico, en el que diferentes segmentos de las secuencias se ajustan a plantados diferentes maximizando la identidad entre la secuencia blanco y los plantados.

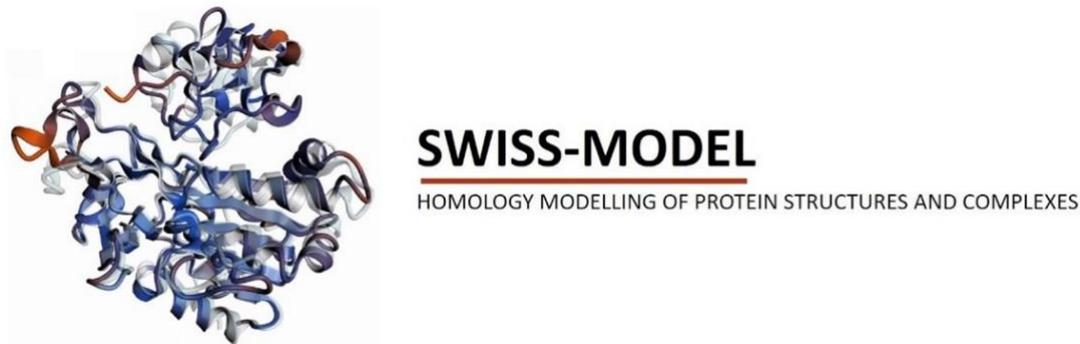


Figura 2. SWISS-MODEL es un programa generador de modelos estructurales de proteínas basados en homología con estructuras depositadas en el PDB. Imagen tomada de <https://www.youtube.com/watch?v=Ln9CYPg94Kc>.

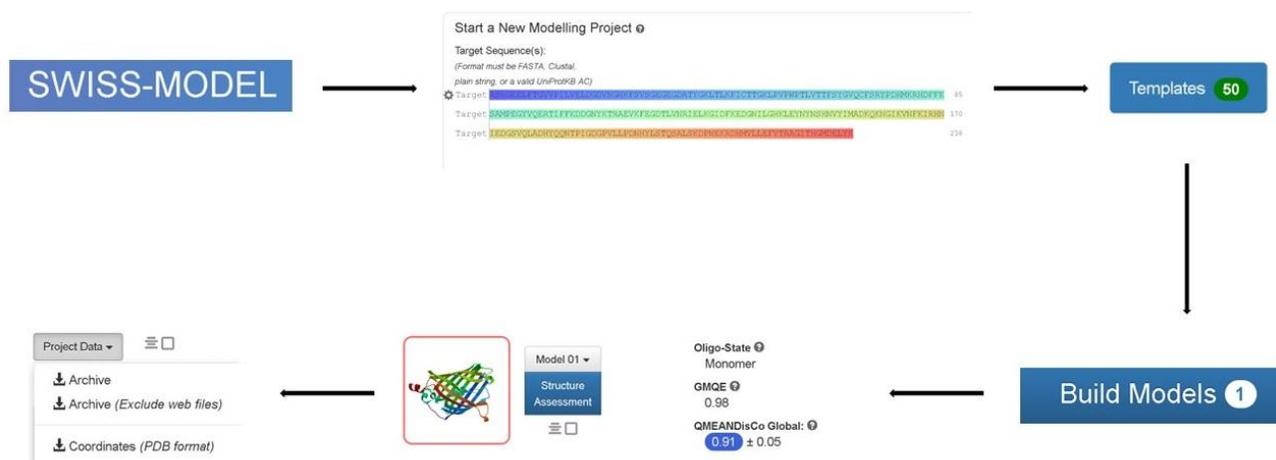


Figura 3. Diagrama de los procedimientos computacionales realizados por SWISS-MODEL para generar un modelo tridimensional basado en homología de una proteína de *E. coli*.

Al disminuir el porcentaje de identidad entre las estructuras primarias de la proteína modelada y su templado, la estructura resultante está sometida a mayor tensión, ya que al sustituir aminoácidos entre secuencias se cambia la flexibilidad y ángulos preferentes del aminoácido; es por ello que el modelo puede requerir un paso de “relajación”. Esta relajación consiste en llevar a la estructura a un mínimo energético para eliminar esta tensión y cualquier choque estérico que pudiera aparecer como resultado de diferencias en el volumen de las cadenas laterales en la proteína blanco y su templado. Dicha relajación permite obtener un modelo más cercano a la estructura nativa, que es la biológicamente funcional. Para poder llevar a cabo la “relajación”, es necesario transformar las coordenadas atómicas producidas durante el modelado en un conjunto de parámetros que definen a la proteína dentro de una simulación. Con este fin, las coordenadas del modelo generado por SWISS-MODEL son utilizadas por el programa CHARMM-GUI (*Chemistry and HARvard Molecular Mechanics*; liga de acceso <https://www.charmm-gui.org/>; [18]) (Figura 4) para asignar a todos los átomos que componen la proteína los parámetros de un campo de fuerza, que describe las interacciones entre átomos, así como los movimientos que cada átomo tiene permitido durante una simulación. Además, basado en información estructural de moléculas de referencia, adiciona los átomos de hidrógeno que no agrega SWISS-MODEL (o cualquier otro algoritmo de modelado por homología), y coloca a la proteína en una caja

de agua virtual, cuya construcción toma en cuenta la fuerza iónica, pH y iones presentes para simular las condiciones experimentales en las que se encontraría la proteína en solución [18,19]. El programa CHARMM-GUI prepara el modelo en un formato compatible con el programa de simulación molecular GROMACS (.pdb, .gro, y .psf), así como el archivo con la topología (.top), que indica la conectividad entre átomos y las reglas de interacción a seguir. Los archivos de salida que genera el programa CHARMM-GUI son tomados por el último programa de este proceso de modelado, llamado GROMACS (*GROningen Machine for Chemical Simulations*) [20] (Figura 5), el cual, mediante varios ciclos de simulación elimina los choques estéricos y lleva a la estructura a un mínimo energético de la proteína cercano a la estructura nativa que es la biológicamente funcional.

Finalmente, los modelos generados, son visualizados en distintos programas entre los que destacan; CHIMERA (liga de acceso <https://www.cgl.ucsf.edu/chimera/>; [21–23]), VMD (liga de acceso <https://www.ks.uiuc.edu/Research/vmd/>; [24]), PyMOL (liga de acceso <https://www.pymol.org/view.html>; [25]), entre otros. La comparación estructural entre modelos (generados mediante experimentos y cálculos computacionales) permite deducir información funcional después de realizar la superposición de modelos, para ver las diferencias estructurales en las regiones de interés. Dicha superposición puede lograrse de distintas maneras, por ejemplo, si la superposi-

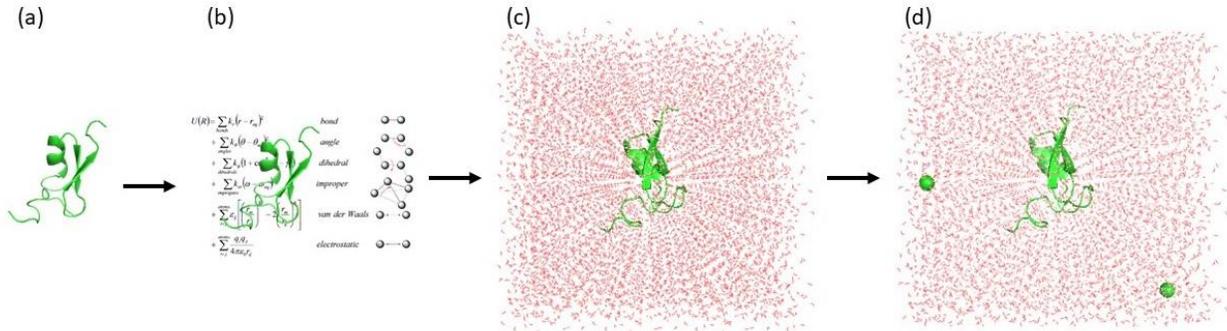


Figura 4. El servidor CHARMM-GUI prepara un sistema compatible con programas para simulaciones moleculares. (a) Lee las coordenadas del archivo pdb; (b) transforma las coordenadas en un conjunto de parámetros de un campo de fuerzas; (c) construye la caja de agua; (d) ajusta el sistema a un pH y adiciona iones.

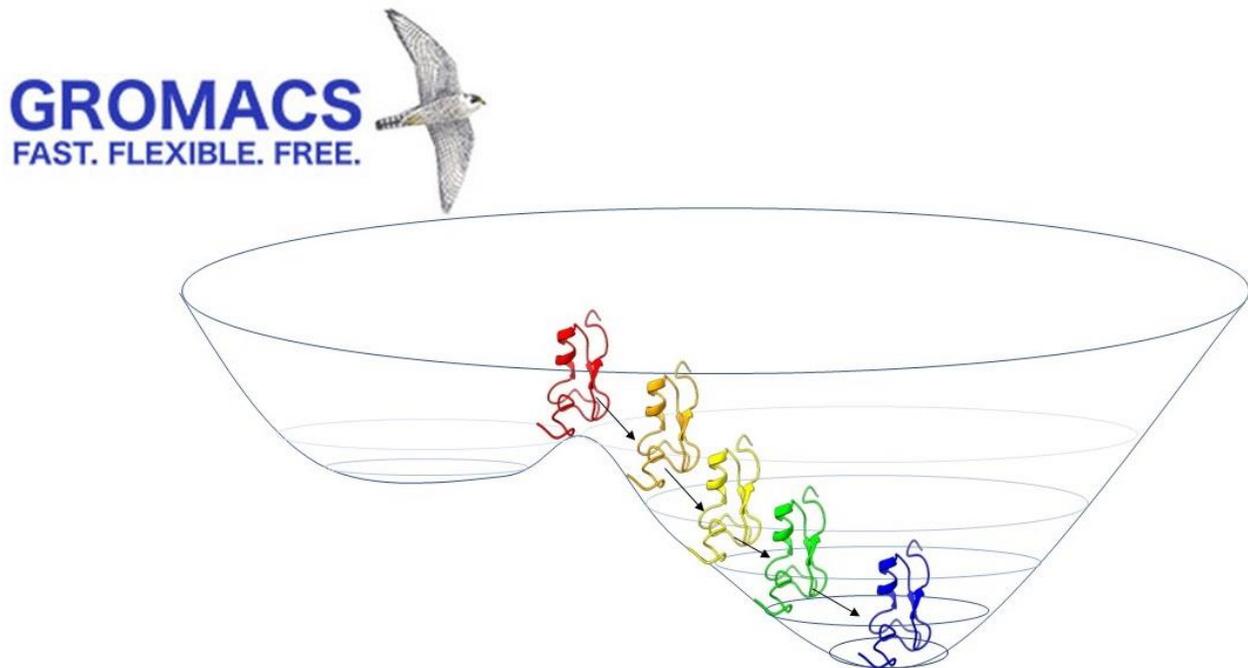


Figura 5. Minimización energética usando GROMACS. Las diferencias entre la secuencia de aminoácidos de la proteína blanco y la proteína molde usadas para generar el modelo pueden forzar a algunos aminoácidos a adoptar conformaciones que generen tensión en la estructura. GROMACS simula cambios conformacionales que "relajan" la estructura, reducen su energía y permiten al modelo adquirir una conformación más cercana a la biológicamente relevante.



ción del modelo y la estructura se realizan con PyMOL se utiliza la función de cealign, la cual indica los valores de la raíz de la desviación cuadrática media o RMSD (*Root-Mean-Square Deviation*), cuyos valores son muy cercanos a cero cuando la estructura y el modelo son muy similares. Además, la comparación entre modelos permite trasladar información bioquímica (como residuos importantes para la catálisis o cambios conformacionales) de un modelo a otro, así como explicar diferencias funcionales entre proteínas homólogas.

La importancia del modelado de proteínas

Los predictores basados en homología, como SWISS-MODEL, generan modelos estructuralmente muy similares a los plegamientos de las estructuras depositadas en el PDB de las cuales fueron formados; estos permiten analizar la función de la proteína en presencia o ausencia de sus sustratos o inhibidores. Estos análisis generan nuevo conocimiento en el PDB acerca de si estas proteínas se encuentran solas o formando complejos. Por ejemplo: las betalactamasas (enzimas capaces de inactivar antibióticos de la familia de los betalactámicos) [26,27], la triosa fosfato isomerasa (enzima glucolítica que cataliza la interconversión entre la dihidroxiacetona fosfato y el gliceraldehído 3-fosfato) [28,29], las fosfatasas alcalinas (enzimas que remueven grupos fosfato de moléculas blanco en presencia de iones como Mg^{2+} y Zn^{2+}) [30,31], la Taq polimerasa (ADN polimerasa que utiliza Mg^{2+} como cofactor para

su funcionamiento) [32,33], los anticuerpos (glucoproteínas para la defensa del huésped) [34,35], etc.

Para que un modelo tenga validez en la investigación, es fundamental que existan o se realicen varios experimentos utilizando distintas técnicas experimentales como la dispersión dinámica de luz o DLS, la cual mide el diámetro de moléculas en disolución (valor que debe coincidir con el modelo), permitiendo discernir entre monómeros y oligómeros de una proteína con alto grado de pureza en solución; la difracción de rayos X de bajo ángulo o SAXS, que genera una envoltura para moléculas en solución (cuya forma debe coincidir con la del modelo por homología); el dicroísmo circular, el cual genera un espectro del que se puede deducir el porcentaje de cada tipo de estructura secundaria (hélices α , láminas β y lazos conectores) presente en una proteína (porcentaje que debe concordar con el modelo); la determinación del intercambio protón-deuterio (H-D) usando espectroscopia de masas, para lo cual la proteína se diluye en un buffer en agua deuterada, deteniendo el intercambio H-D reduciendo el pH a 2.5, el porcentaje de intercambio H-D de cada aminoácido determinado por espectroscopia de masas depende del acceso que tenga el solvente al aminoácido y del tiempo transcurrido. Al disminuir el porcentaje de identidad entre el templado y la proteína blanco, se vuelve más importante que el modelo generado esté apoyado por estos datos experimentales para poder validarlo.

El impacto del modelado de proteínas mediante SWISS-MODEL en distintas áreas de investigación combinado con otros servidores

Existe una gran cantidad de trabajos en donde el programa SWISS-MODEL unido a otros programas computacionales ha contribuido de manera importante para fortalecer y apoyar la generación de nuevo conocimiento en diferentes áreas de investigación, aquí solo mostraremos algunos ejemplos de ello.

Se ha realizado modelado por homología de la proteína nifA para tener una mejor comprensión de su estructura y su influencia en la fijación biológica de nitrógeno, utilizando varios servidores; entre ellos, SWISS-MODEL fue importante para generar un modelo más aceptable que permita entender mejor el funcionamiento de la proteína [36]. Por medio de SWISS-MODEL y CHARMM-GUI nosotros fuimos capaces de generar modelos estructurales de la proteína escorpina de los alacranes *Hoffmannihadrurus gertschi* y *Vaejovis mexicanus*, basados en evidencias experimentales [37], y propusimos apoyados en estos modelos, mutaciones en una de las regiones de su estructura secundaria (α -hélices), para evaluar los efectos en las constantes de inhibición en el desarrollo de los oocinetos, que son uno de los estadios del parásito de la malaria [37]. La malaria es una de las enfermedades más importantes del sector salud a nivel mundial, con 241 millones de casos en el año 2021.

En la actualidad, para poder contender con el

SARS-CoV-2, causante de la COVID-19, es fundamental el desarrollo de vacunas y el diseño de fármacos que contribuyan con una alta capacidad de neutralización del virus. Para el diseño de fármacos se ha recurrido a distintas estrategias de simulación computacional como Hawk Dock [38], HDOCK [39], PockDrug [40] y SWISS-MODEL [12]. Finalmente, para el modelado por homología y las energías de unión de la molécula, se emplean herramientas computacionales que analizan la conformación, la afinidad de unión y la capacidad farmacológica del receptor unido a su ligando [41].

CONCLUSIÓN

En las últimas décadas, las técnicas para develar la estructura tridimensional de las proteínas como la difracción de los rayos X en cristales de proteína, la resonancia magnética nuclear, etc., han incrementado significativamente el número de estructuras depositadas en el PDB. Este aumento en la disponibilidad de estructuras 3D ha permitido al modelado por homología posicionarse como una herramienta para generar modelos que aporten un mayor conocimiento en diferentes áreas de investigación, entre las que destacan la biología estructural, la medicina, la biotecnología, la farmacología etc.

La calidad de los modelos generados mediante SWISS-MODEL y la rapidez con la que lo hace, en comparación con otros servidores, lo convierte en una herramienta útil para la biología estructural. Este servidor, junto con

programas computacionales como CHARMM-GUI y GROMACS, forman una línea de producción de libre acceso, rápida y versátil para obtener modelos de estructuras de proteínas que todavía no han sido depositadas en el PDB. Este conocimiento estructural mejora nuestro entendimiento acerca de la función de las proteínas y su interacción con otras moléculas, haciendo posible predecir residuos importantes para la función y entender el efecto de mutaciones reportadas en la literatura. La comprensión de fenómenos bioquímicos, con detalle atómico, puede también llevar a la identificación de blancos terapéuticos y diseño de fármacos para el tratamiento de enfermedades o trastornos genéticos.

CONFLICTO DE INTERESES

Los autores declaran la ausencia de conflicto de interés.

AGRADECIMIENTOS

Los autores agradecen al Mtro. Juan Manuel Hurtado Ramírez, M.E.M. David Santiago Castañeda Carreón, L.I. Servando Aguirre Cruz e Ing. Roberto P. Rodríguez Bahena, por su asistencia técnica en computación y al M.C. Jorge Arturo Yañez Ponce de León, por su asistencia técnica.

REFERENCIAS

[1]. Plourde AR, Bloch EM. A literature review of zika virus. *Emerg Infect Dis* 2016;22:1185–

92. <https://doi.org/10.3201/eid2207.151990>.

[2]. Gormley M. The first “molecular disease”: a story of Linus Pauling, the intellectual patron. *Endeavour* 2007;31:71–7. <https://doi.org/10.1016/j.endeavour.2007.05.009>.

[3]. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, *et al.* The protein data bank. *Acta Crystallogr Sect D Biol Crystallogr* 2002;58:899–907. <https://doi.org/10.1107/S0907444902003451>.

[4]. Burley SK. Impact of structural biologists and the Protein Data Bank on small-molecule drug discovery and development. *J Biol Chem* 2021;296:100559. <https://doi.org/10.1016/j.jbc.2021.100559>.

[5]. Blundell TL. The first resolution revolution in protein structure analysis: X-ray diffraction of polypeptide conformations and globular protein folds in 1950s and 1960s. *Prog Biophys Mol Biol* 2021;167:32–40. <https://doi.org/10.1016/j.pbiomolbio.2021.09.002>.

[6]. Scarborough NM, Godaliyadda GMDP, Ye DH, Kissick DJ, Zhang S, Newman JA, *et al.* Dynamic X-ray diffraction sampling for protein crystal positioning. *J Synchrotron Radiat* 2017;24:188–95. <https://doi.org/10.1107/S160057751601612X>.

[7]. Ishima R, Torchia DA. Protein dynamics from NMR. *Nat Struct Biol* 2000;7:740–3. <https://doi.org/10.1038/78963>.

[8]. Letertre MPM, Giraudeau P, de Tullio P. Nuclear Magnetic Resonance Spectroscopy in

- Clinical Metabolomics and Personalized Medicine: Current Challenges and Perspectives. *Front Mol Biosci* 2021;8:1–25. <https://doi.org/10.3389/fmolb.2021.698337>.
- [9]. Murata K, Wolf M. Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. *Biochim Biophys Acta - Gen Subj* 2018;1862:324–34. <https://doi.org/10.1016/j.bbagen.2017.07.020>.
- [10]. Carroni M, Saibil HR. Cryo electron microscopy to determine the structure of macromolecular complexes. *Methods* 2016;95:78–85. <https://doi.org/10.1016/j.ymeth.2015.11.023>.
- [11]. Soni N, Madhusudhan MS. Computational modeling of protein assemblies. *Curr Opin Struct Biol* 2017;44:179–89. <https://doi.org/10.1016/j.sbi.2017.04.006>.
- [12]. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, *et al.* SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res* 2018;46:W296–303. <https://doi.org/10.1093/nar/gky427>.
- [13]. Zheng W, Zhang C, Bell EW, Zhang Y. I-TASSER gateway: A protein structure and function prediction server powered by XSEDE. *Futur Gener Comput Syst* 2019;99:73–85. <https://doi.org/10.1016/j.future.2019.04.011>.
- [14]. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, *et al.* SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* 2014;42:252–8. <https://doi.org/10.1093/nar/gku340>.
- [15]. Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* 2011;27:343–50. <https://doi.org/10.1093/bioinformatics/btq662>.
- [16]. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER suite: Protein structure and function prediction. *Nat Methods* 2014;12:7–8. <https://doi.org/10.1038/nmeth.3213>.
- [17]. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 2015;10:845–58. <https://doi.org/10.1038/nprot.2015.053>.
- [18]. Kim S, Lee J, Jo S, Brooks CL, Lee HS, Im W. CHARMM-GUI ligand reader and modeler for CHARMM force field generation of small molecules. *J Comput Chem* 2017;38:1879–86. <https://doi.org/10.1002/jcc.24829>.
- [19]. Jo S, Cheng X, Lee J, Kim S, Park SJ, Patel DS, *et al.* CHARMM-GUI 10 years for biomolecular modeling and simulation. *J Comput Chem* 2017;38:1114–24. <https://doi.org/10.1002/jcc.24660>.
- [20]. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: Fast, flexible, and free. *J Comput Chem* 2005;26:1701–18. <https://doi.org/10.1002/jcc.20291>.
- [21]. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J Comput*

- Chem 2004;25:1605–12.
<https://doi.org/10.1002/jcc.20084>.
- [22]. Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* 2018;27:14–25.
<https://doi.org/10.1002/pro.3235>.
- [23]. Chen JE, Huang CC, Ferrin TE. RRDistMaps: A UCSF Chimera tool for viewing and comparing protein distance maps. *Bioinformatics* 2015;31:1484–6.
<https://doi.org/10.1093/bioinformatics/btu841>.
- [24]. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *J Mol Graph* 1996;14:33–8.
[https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
- [25]. Yuan S, Chan HCS, Hu Z. Using PyMOL as a platform for computational drug design. *Wiley Interdiscip Rev Comput Mol Sci* 2017;7.
<https://doi.org/10.1002/wcms.1298>.
- [26]. Hanes MS, Jude KM, Berger JM, Bonomo RA, Handel TM. Structural and biochemical characterization of the interaction between KPC-2 β -lactamase and β -lactamase inhibitor protein. *Biochemistry* 2009;48:9185–93.
<https://doi.org/10.1021/bi9007963>.
- [27]. Golemi D, Maveyraud L, Vakulenko S, Samama JP, Mobashery S. Critical involvement of a carbamylated lysine in catalytic function of class D β -lactamases. *Proc Natl Acad Sci U S A* 2001;98:14280–5.
<https://doi.org/10.1073/pnas.241442898>.
- [28]. Alahuhta M, Salin M, Casteleijn MG, Kemmer C, El-Sayed I, Augustyns K, *et al.* Structure-based protein engineering efforts with a monomeric TIM variant: The importance of a single point mutation for generating an active site with suitable binding properties. *Protein Eng Des Sel* 2008;21:257–66.
<https://doi.org/10.1093/protein/gzn002>.
- [29]. Parthasarathy S, Eazhisai K, Balaram H, Balaram P, Murthy MRN. Structure of Plasmodium falciparum Triose-phosphate Isomerase-2-Phosphoglycerate Complex at 1.1-Å Resolution. *J Biol Chem* 2003;278:52461–70.
<https://doi.org/10.1074/jbc.M308525200>.
- [30]. Murphy JE, Tibbitts TT, Kantrowitz ER. Mutations at positions 153 and 328 in *Escherichia coli* alkaline phosphatase provide insight towards the structure and function of mammalian and yeast alkaline phosphatases. *J Mol Biol* 1995;253:604–17.
<https://doi.org/10.1006/jmbi.1995.0576>.
- [31]. Holtz KM, Kantrowitz ER, Stec B, Myers JK, Antonelli SM, Widlanski TS. Alternate modes of binding in two crystal structures of alkaline phosphatase-inhibitor complexes. *Protein Sci* 2000;9:907–15.
<https://doi.org/10.1110/ps.9.5.907>.
- [32]. Murali R, Sharkey DJ, Daiss JL, Krishna Murthy HM. Crystal structure of Taq DNA polymerase in complex with an inhibitory Fab: The Fab is directed against an intermediate in the helix-coil dynamics of the enzyme. *Proc Natl Acad Sci U S A* 1998;95:12562–7.
<https://doi.org/10.1073/pnas.95.21.12562>.
- [33]. Kropp HM, Dürr SL, Peter C, Diederichs



K, Marx A. Snapshots of a modified nucleotide moving through the confines of a DNA polymerase. *Proc Natl Acad Sci U S A* 2018;115:9992–7.

<https://doi.org/10.1073/pnas.1811518115>.

[34]. Pinto D, Park YJ, Beltramello M, Walls AC, Tortorici MA, Bianchi S, *et al.* Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature* 2020;583:290–5.

<https://doi.org/10.1038/s41586-020-2349-y>.

[35]. Gohain N, Tolbert WD, Acharya P, Yu L, Liu T, Zhao P, *et al.* Cocystal Structures of Antibody N60-i3 and Antibody JR4 in Complex with gp120 Define More Cluster A Epitopes Involved in Effective Antibody-Dependent Effector Function against HIV-1. *J Virol* 2015;89:8840–54.

<https://doi.org/10.1128/jvi.01232-15>.

[36]. Satyanarayana SDV, Krishna MSR, Pavan Kumar P, Jeeredy S. *In silico* structural homology modeling of nif A protein of rhizobial strains in selective legume plants. *J Genet Eng Biotechnol* 2018;16:731–7.

<https://doi.org/10.1016/j.jgeb.2018.06.006>.

[37]. Vargas-Jaimes L, Rodriguez MC, Argotte-Ramos R, Juárez-González VR, Pastor N, Cesa-Luna C, *et al.* Recombinant C-Terminal Domains from Scorpine-like Peptides

Inhibit the Plasmodium berghei Ookinete Development In Vitro. *Int J Pept Res Ther* 2021;27:817–29.

<https://doi.org/10.1007/s10989-020-10130-7>.

[38]. Weng G, Wang E, Wang Z, Liu H, Zhu F, Li D, *et al.* HawkDock: a web server to predict and analyze the protein-protein complex based on computational docking and MM/GBSA. *Nucleic Acids Res* 2019;47:W322–30.

<https://doi.org/10.1093/nar/gkz397>.

[39]. Yan Y, Zhang D, Zhou P, Li B, Huang SY. HDock: A web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res* 2017;45:W365–73.

<https://doi.org/10.1093/nar/gkx407>.

[40]. Borrel A, Regad L, Xhaard H, Petitjean M, Camproux AC. PockDrug: A model for predicting pocket druggability that overcomes pocket estimation uncertainties. *J Chem Inf Model* 2015;55:882–95.

<https://doi.org/10.1021/ci5006004>.

[41]. Siddiqi MA, Rao DS, Suvarna G, Chennamachetty VK, Verma MK, Rao MVR. *In-Silico* Drug Designing of Spike Receptor with Its ACE2 Receptor and Nsp10/Nsp16 MTase Complex Against SARS-CoV-2. *Int J Pept Res Ther* 2021;27:1633–40.

<https://doi.org/10.1007/s10989-021-10196-x>.