

PREMIO NOBEL DE QUÍMICA 2024 A LOS AVANCES EN LA COMPUTARIZACIÓN DEL DISEÑO Y PREDICCIÓN DE ESTRUCTURAS DE LAS PROTEÍNAS

PABLO J. AZURMENDI¹, ISABEL A. LÜTHY²

¹Servicio de Nefrología Experimental y Bioquímica Molecular, Instituto de Investigaciones Médicas Alfredo Lanari, Facultad de Medicina, UBA, IDIM UBA-CONICET, ²Instituto de Biología y Medicina Experimental - CONICET, Buenos Aires, Argentina

E-mail: azurmendi.pablo@lanari.uba.ar, i.luthy@ibyme.org.ar

La Real Academia Sueca de las Ciencias ha decidido otorgar el Premio Nobel en Química 2024 a David Baker, Demis Hassabis y John Jumper, por el diseño computacional de proteínas y la predicción de su estructura¹.

Las primeras estructuras tridimensionales (3D) de proteínas fueron determinadas por cristalografía de rayos X hace unos 65 años. La estructura tridimensional les permite ejercer su acción biológica, función y propiedades bioquímicas. En 1972, Christian Anfinsen recibió el mismo Premio Nobel por encontrar que la estructura 3D está codificada en la secuencia de aminoácidos de la cadena polipeptídica, iniciando una larga búsqueda científica para predecir dicha estructura a partir de las secuencias. Debe notarse que, si bien las secuencias de ADN en bases de datos públicas se acercan actualmente a tres mil millones, el número de secuencias proteicas identificadas en organismos ronda los 200 millones, el banco de datos de proteínas *Protein Data Bank* solo contiene alrededor de 200 000 estructuras proteicas².

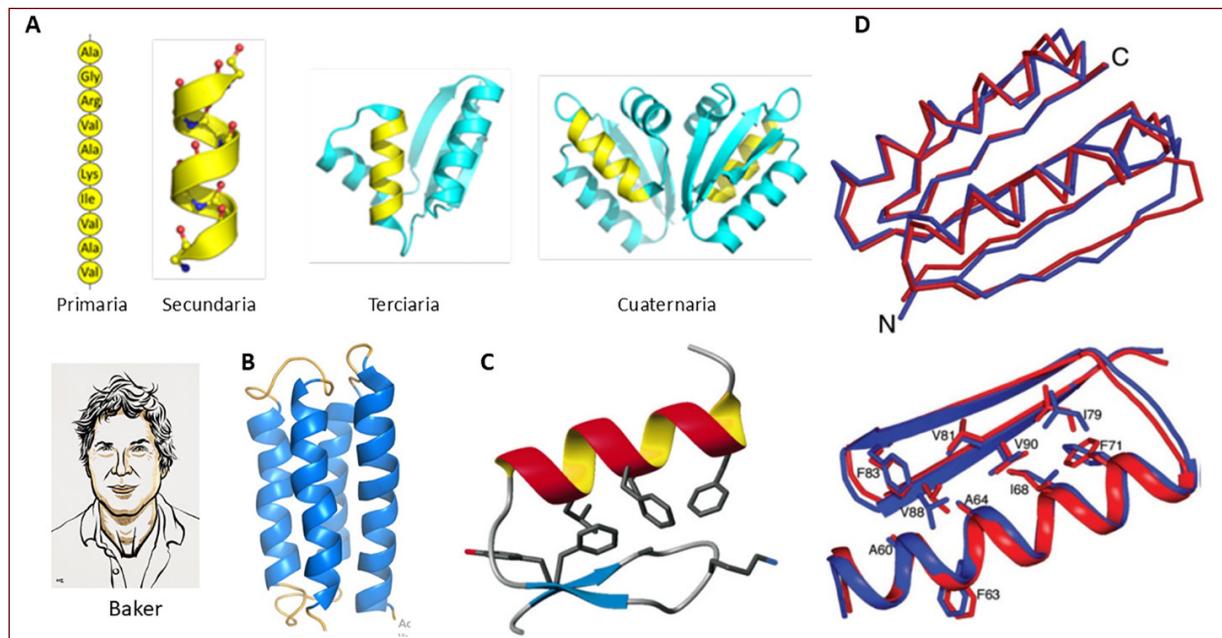
Es interesante repasar la jerarquía de la estructura de las proteínas: La estructura primaria es la secuencia de aminoácidos que está determinada por la secuencia correspondiente de tripletes de bases de ADN. La secundaria es formación de patrones geométricos regulares de α -hélices y láminas- β . Por último, las estructuras terciarias y cuaternarias son la forma detallada

en 3D de la cadena polipeptídica y la asociación de varias cadenas o subunidades polipeptídicas, respectivamente (Fig. 1A).

La predicción de la estructura proteica puede ser encarada preguntando qué secuencias aminoácidas pueden proporcionar cierto plegado. Esta es la base del diseño de la estructura proteica, un campo en que se identifica esta estructura por métodos computacionales. Este año el Premio Nobel en Química reconoce descubrimientos a este problema – predicción de la estructura a partir de la secuencia y predicción de la secuencia a partir de la estructura – con profundas implicancias. La mayoría de las estructuras de proteínas monoméricas puede ahora ser predicha con gran fidelidad, y existen enormes bases de datos de estructuras, que poseen un gran impacto en la investigación bioquímica y biológica. De la misma manera, estructuras proteicas completamente nuevas, que no están presentes en la naturaleza, pueden ser creadas ahora por diseño computacional y utilizadas para diversas aplicaciones biotecnológicas y biomédicas.

Luego de la determinación de las primeras estructuras proteicas, fue evidente que la estructura 3D poseía elementos como α -hélices y láminas- β , cuya orientación, en conjunto con las regiones de los *loops* conectores, define la topología terciaria. Podemos recordar que el patrón de α -hélices fue predicho por Linus Pauling ya en 1951². Luego de diversos intentos de estable-

Figura 1 | Representación esquemática de jerarquía estructural de las proteínas (A) y los avances impulsados por Davis Baker en su modelado. B: Estructura en haces de cuatro hélices con un interior hidrofóbico y una superficie exterior hidrófila. C: Diseño de una proteína en "dedo" de Zinc. D: Comparación de la estructura predicha para la proteína Top7 (azul) con la estructura determinada por rayos X (rojo) en el panel superior y la vista de las cadenas laterales superpuestas en el modelo diseñado y la estructura resuelta en el panel inferior



Adaptado de Ref. 2

cer la conformación, se pudo realizar un caso de diseño proteico, con una cara de cada α -hélice hidrofílica, expuesta al solvente, y las otras caras hidrofóbicas, lo que le permite un empaquetado adecuado con otras superficies igualmente hidrofóbicas que se encuentran en el interior de la estructura diseñada. Esto fue realizado por Regan y DeGrado, quienes, en 1988, construyeron una proteína de cuatro hélices que obedecía a estos principios, donde se necesitaron tres loops para conectar las cuatro α -hélices (Fig. 1B). El primer diseño exitoso de una pequeña proteína por vía computacional fue publicado por Dahiyat y Mayo en 1997, en el que eligieron un motivo de "dedo" de zinc que coordina uno o dos iones Zn^{2+} . El "dedo" de zinc es una estructura relativamente pequeña, unos 30 aminoácidos, y contiene una α -hélice y dos láminas- β , además de loops conectores cortos (motivo $\beta\beta\alpha$) (Fig. 1C).

El hallazgo en diseño computacional *de novo* de proteínas llegó en 2003, cuando David Baker y colaboradores publicaron el diseño y valida-

ción cristalográfica de una proteína α/β de 93-residuos llamada Top7³. Este remarkable logro era una proteína relativamente grande con dos α -hélices y una lámina- β hecha de cinco hebras β , y la estructura predicha coincidió con la experimental, incluyendo las posiciones de la cadena lateral. Por otro lado, los autores buscaron diseñar un patrón de plegado que no se encuentra en ninguna proteína globular del *Protein Data Bank*. Adicionalmente, el diseño de Top7 no presenta similitudes con ninguna proteína natural. Entonces, Top7 es una proteína completamente novel, en cuanto a estructura y secuencia, diseñada por computación automatizada con optimización completa de la cadena principal y las laterales (Fig. 1D).

Baker y colaboradores utilizaron un programa computacional llamado Rosetta, que habían publicado anteriormente, que ensambla fragmentos estructurales cortos de estructuras proteicas con secuencia similar en *Protein Data Bank* y, simultáneamente, optimiza la secuencia y estruc-

tura con respecto a la conformación de la cadena principal. Utilizaron optimización de Monte Carlo en el cálculo con una función de energía. El programa genera muchas soluciones posibles y las ranquea en términos de energía. Baker y sus colegas continuaron desarrollando una gran variedad de estructuras proteicas utilizando dicho programa. El trabajo más reciente involucra el diseño de funciones avanzadas, lo cual plantea desafíos enormes en cuanto a la dinámica proteica, transiciones estructurales, alosterismo [un modo de regulación de la actividad de una enzima por el que la unión de una molécula en una ubicación (sitio alostérico) modifica las condiciones de unión de otra molécula, en otra ubicación (sitio catalítico) distante de la primera], efectos catalíticos, etc.

En 2008, Baker y colaboradores reportaron los primeros intentos de diseño *de novo* de una enzima, o el diseño de enzimas noveles que catalizaran reacciones que no ocurren naturalmente. Sin embargo, estas enzimas presentan niveles de catálisis menor que las naturales. No obstante, su diseño pudo ser mejorado con sucesivas rondas de evolución dirigida experimentalmente (por la cual Frances Arnold recibió el Premio Nobel en Química 2018). Baker y colaboradores entre otros muchos logros, consiguieron diseñar estructuras proteicas que unen esteroides con alta afinidad y selectividad⁴.

En cuanto a la predicción de estructuras, con los avances en la secuenciación del ADN a principios de los 90, el número de secuencias proteicas disponibles creció rápidamente. Esto permitió obtener numerosas secuencias de una familia que podían ser alineadas y comparadas. Luego de diversos métodos para aumentar la correcta predicción de la secuencia, hacia 2018 hubo un enorme incremento en la exactitud de mapas estructurales 3D y los contactos entre moléculas. Ello ocurrió con utilización de métodos de aprendizaje profundo (en inglés, *deep learning*) utilizando redes neuronales convolucionales. Consideraron que este tipo de entrenamiento tenía similitudes con el reconocimiento de imágenes. En 2018, la compañía DeepMind, fundada y dirigida por Demis Hassabis, realizó un programa basado en una red neuronal convolucional que llamaron AlphaFold (ahora cono-

cida como AlphaFold1 o AF1)⁵. El programa fue entrenado con las estructuras de *Protein Data Bank* para producir un mapa de distancias entre residuos o, mejor dicho, un mapa de probabilidad de distribuciones para las distancias entre residuos aminoacídicos, basada en alineaciones de múltiples secuencias. A partir de este mapa, se pudo construir un potencial de fuerza media y optimizarlo mediante un algoritmo de descenso de gradiente para generar estructuras. Hassabis y su equipo ya habían realizado programas de Go y ajedrez, también basados en *deep learning*.

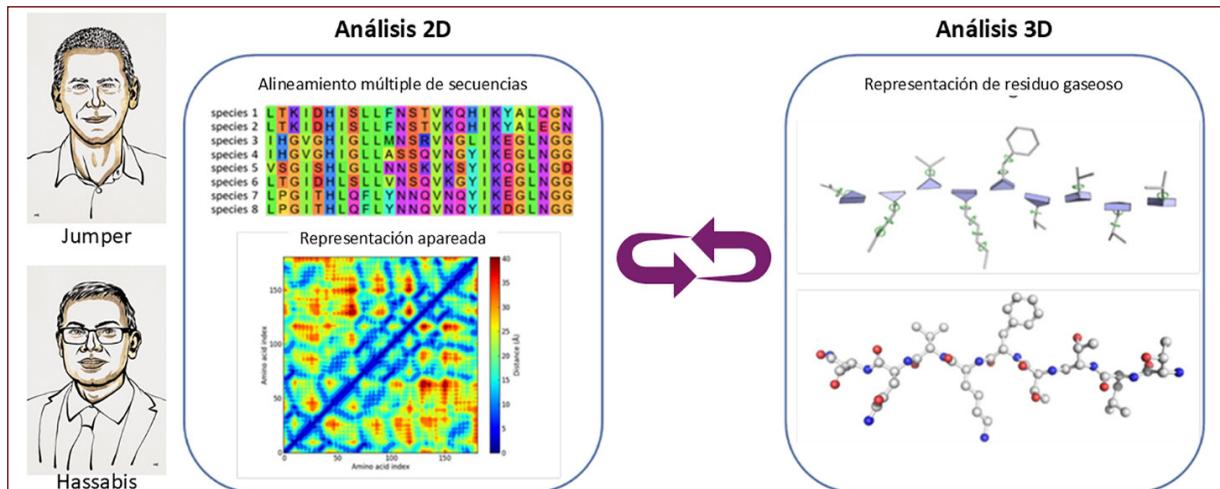
El descubrimiento verdaderamente fundamental llegó con AlphaFold2 (AF2). El grupo de trabajo de AF2 liderado por John Jumper y Hassabis finalmente logró resolver el problema de la predicción de la estructura proteica para proteínas monoméricas con una precisión de alrededor de 1 Å. Rediseñaron la red neuronal, prestando especial atención en los mecanismos de aprendizaje. Encararon el programa de forma tal que el resultado de salida sea la coordinación (Fig. 2). El hecho que AF2 sea de código abierto fue decisivo en su impacto, ya que pudo ser extensamente evaluado y validado. Una arquitectura similar a la de AF2 también fue rápidamente adoptada por Baker y colegas en el programa RoseTTAFold.

El grupo AlphaFold2 inmediatamente creó grandes bases de datos de estructuras proteicas predichas, primero del proteoma humano, y luego de la mayoría de las secuencias (más de 200 millones) disponibles en la base de datos UniProt (*Universal Protein Resource*).

Resumiendo, los logros de David Baker, Demis Hassabis y John Jumper en el campo del diseño computacional de proteínas y de su estructura son profundos, abren una nueva era en la investigación bioquímica y biológica, y permiten una enorme cantidad de nuevas posibilidades.

Se puede concluir que este Premio Nobel de Química involucra importantes hallazgos realizados recientemente en cuanto a los problemas de diseño estructural y predicción de ella mediante el uso de herramientas innovadoras de inteligencia artificial. Esto es una novedad en estos premios ya que, en general, esperan un largo tiempo antes de galardonar un descubrimiento.

Figura 2 | Representación esquemática de los dos módulos principales de AlphaFold2, programa de predicción de estructura espacial proteica desarrollado por Demis Hassabis y John Jumper. La secuencia de interés se combina con información de las bases de datos de secuencias y estructuras, que sirven como entrada para el análisis bidimensional con el módulo Evoformer. El siguiente módulo analiza la estructura tridimensional que produce un modelo 3D de la proteína correspondiente. Este proceso se itera numerosas veces de manera de ajustar ambos análisis y generar un modelo más justado y exacto.



Adaptado de Ref. 2

Bibliografía

1. The Nobel Assembly at Karolinska Institutet. The Nobel Prize in chemistry 2024. 2024. En: <https://www.nobelprize.org/prizes/chemistry/2024/press-release/>; consultado octubre 2024.
2. The Nobel Committee for Chemistry. Scientific background to the Nobel Prize in chemistry 2024. 2024: En: <https://www.nobelprize.org/uploads/2024/10/advanced-chemistryprize2024.pdf>; consultado octubre 2024.
3. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. Design of a novel globular protein fold with atomic-level accuracy. *Science* 2003; 302: 1364-8.
4. Tinberg CE, Khare SD, Dou J, et al. Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* 2013; 501: 212-6.
5. Senior AW, Evans R, Jumper J, et al. Improved protein structure prediction using potentials from deep learning. *Nature* 2020; 577: 706-10.