

## Proteínas *de novo*: del diseño a la funcionalización

*De novo proteins: from design to functionalization*

Oscar de Jesús Rodríguez Meza,<sup>1</sup> Miguel Antonio Costas Basín<sup>1†</sup> y Daniel Alejandro Fernández Velasco<sup>2</sup>

### Resumen

El diseño de proteínas *de novo* es un campo innovador con aplicaciones relevantes en medicina y biotecnología. Consiste en crear proteínas “desde cero”, con secuencias de aminoácidos completamente distintas a las presentes en la naturaleza. Este trabajo presenta la evolución del campo, desde sus inicios con el diseño minimalista y racional en los años ochenta, hasta el uso de modelos de lenguaje basados en aprendizaje profundo. El objetivo principal es ofrecer a estudiantes de licenciatura y posgrado un panorama general sobre los avances en el diseño de proteínas *de novo* y su vínculo con el plegamiento y la funcionalidad proteica.

**Palabras clave:** diseño de proteínas *de novo*, plegamiento de proteínas, biotecnología avanzada, aprendizaje profundo en biología, proteínas funcionales artificiales.

### Abstract

*De novo* protein design is an innovative field with relevant applications in medicine and biotechnology. It involves creating proteins “from scratch,” with amino acid sequences entirely different from those found in nature. This work presents the evolution of the field, from its beginnings with minimalist and rational design in the 1980s to the current use of language models based on deep learning. The main objective is to provide undergraduate and graduate students with an overview of the significant advances in *de novo* protein design and its connection to protein folding and functional protein development.

**Keywords :** *de novo* protein design, protein folding, advanced biotechnology, deep learning in biology, artificial functional proteins.

### CÓMO CITAR:

Rodríguez Meza, O. de J., Costas Basín, M. A., y Fernández Velasco, D. A. (2025, julio-septiembre). Proteínas *de novo*: del diseño a la funcionalización. *Educación Química*, 36(3). <https://doi.org/10.22201/fq.18708404e.2025.3.90539>

<sup>1</sup> Facultad de Química, Universidad Nacional Autónoma de México, México.

<sup>2</sup> Facultad de Medicina, Universidad Nacional Autónoma de México, México.

## Introducción

Las proteínas son uno de los grupos de biomoléculas más importantes en los seres vivos, ya que desempeñan diversas funciones vitales, como la catálisis enzimática, la señalización celular y el soporte estructural, entre otras. Los humanos producimos más de 10 000 proteínas diferentes, y se ha estimado que su concentración puede llegar hasta 20 millones de moléculas por célula (Beck et al., 2011).

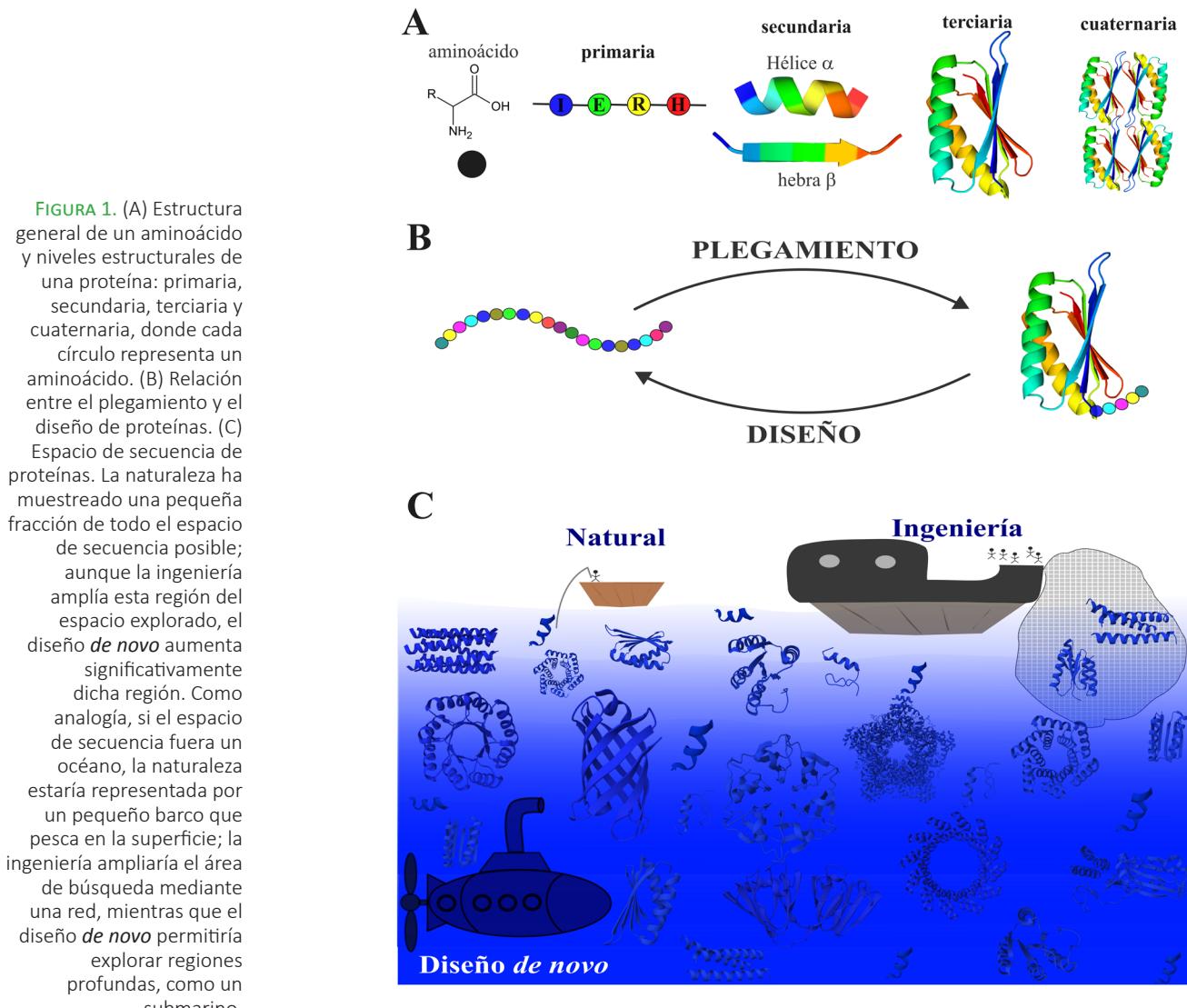
La gran mayoría de las proteínas conocidas se componen de 20 aminoácidos diferentes. Cada uno está formado por un átomo de carbono central unido a cuatro sustituyentes (Figura 1A): tres de ellos son comunes a todos los aminoácidos —un grupo amino, un grupo carboxilo y un átomo de hidrógeno—, mientras que el cuarto sustituyente, denominado grupo “R”, varía entre 20 opciones distintas que difieren en tamaño y propiedades químicas.

Los aminoácidos funcionan como bloques de construcción para formar un polímero, el cual puede adoptar ciertas conformaciones estabilizadas por interacciones entre los residuos que lo componen. Para que una proteína lleve a cabo su función, debe adoptar una conformación característica determinada por la secuencia de aminoácidos (denominada *secuencia primaria*); a esta disposición espacial de los átomos se le llama estado o conformación nativo/a.

El estado nativo es complejo, ya que involucra interacciones entre miles de átomos. Por ello, la estructura de las proteínas suele representarse en diferentes niveles. Se conoce como *estructura secundaria* a las conformaciones locales que adoptan los residuos adyacentes, entre las que destacan las hélices alfa ( $\alpha$ ) y las hebras beta ( $\beta$ ). Al conjunto de todos los elementos estructurales locales que conforman la forma global de la cadena de aminoácidos se le denomina *estructura terciaria*, la cual puede incluir elementos segregados ( $\alpha + \beta$ ) o combinados ( $\alpha/\beta$ ). El estado funcional de muchas proteínas requiere de la presencia de más de una cadena polipeptídica, lo que da lugar a la *estructura cuaternaria*.

Por medio de diferentes técnicas experimentales —difracción de rayos X, resonancia magnética nuclear y crio-microscopía electrónica— se ha podido determinar la estructura tridimensional de las proteínas en estado nativo, lo que permite describir, a nivel atómico, la organización espacial de sus residuos. Actualmente, existen más de 235 000 estructuras diferentes depositadas en el PDB (*Protein Data Bank*; Berman et al., 2000; Burley et al., 2025), y se han predicho más de 200 millones de *estructuras* (Varadi et al., 2024). Se estima que en la naturaleza existen entre 1000 y 10 000 tipos distintos de plegamiento (Kolodny et al., 2013).

Sin embargo, las proteínas presentes en la naturaleza representan una pequeña fracción de todas las secuencias posibles. Esto nos lleva a preguntarnos si es posible diseñar computacionalmente secuencias que adopten alguno de estos tipos de plegamiento, o incluso uno completamente nuevo. Este es el objetivo del *diseño de proteínas de novo* (Woolfson et al., 2015). Gracias a esta tecnología, es posible generar proteínas con estructuras y funciones completamente novedosas, las cuales pueden emplearse para desarrollar nuevos fármacos, biocatalizadores industriales o herramientas para la degradación de contaminantes.



### El problema del plegamiento y el diseño de proteínas: dos caras de la misma moneda

El proceso mediante el cual una proteína sin estructura definida adopta el estado nativo se conoce como *plegamiento*. En 1973, Christian Anfinsen descubrió que este proceso es espontáneo *in vitro*, y propuso la “hipótesis termodinámica del plegamiento”, la cual establece que el estado nativo es el de menor energía (Anfinsen, 1973).

A lo largo de los años, se ha intentado predecir la estructura de mínima energía — es decir, el estado nativo — a partir de la secuencia de aminoácidos. Esta no es una tarea sencilla. Por ejemplo, si consideramos que cada residuo puede adoptar tres estados posibles (aunque en realidad son muchos más), una proteína típica de 100 residuos podría adoptar  $3^{100} = 5 \times 10^{47}$  conformaciones distintas. Si explorar cada conformación tomara un femtosegundo, revisar todas las posibles llevaría  $10^{27}$  años. Sin embargo, las proteínas

logran plegarse en segundos. Este análisis matemático es conocido como la *paradoja de Levinthal*. El hecho de que las proteínas alcancen su estado nativo tan rápidamente sugiere la existencia de conformaciones preferenciales o *rutas de plegamiento* (Levinthal, 1969).

El llamado “problema del plegamiento de proteínas” consiste en predecir la estructura de una proteína a partir de su secuencia de aminoácidos (Figura 1B). Este problema ha sido abordado desde diversas perspectivas, siendo la primera —y más lógica— la aproximación fisicoquímica, basada en la comprensión de las interacciones que estabilizan la estructura, así como en reglas que rigen el comportamiento individual y colectivo de los aminoácidos (Dill, 1990; Dill y Chan, 1997). Este enfoque explica por qué las proteínas se pliegan, pero no ha sido capaz de predecir su estructura nativa con precisión.

Por otro lado, es posible modificar la estructura y función de las proteínas alterando su secuencia mediante técnicas como la mutagénesis puntual, la evolución dirigida, la ingeniería o el rediseño de proteínas naturales. No obstante, estas estrategias parten de una secuencia conocida, lo que limita los cambios radicales en sus propiedades.

En este contexto, conviene considerar el número de proteínas diferentes presentes en los organismos existentes: alrededor de  $10^{12}$ . Aunque este número es considerable, resulta pequeño comparado con el número de secuencias posibles para una proteína de 100 aminoácidos, que asciende a  $20^{100}$ , es decir,  $1.3 \times 10^{130}$ , un número mucho mayor que el de átomos estimados en el universo ( $10^{82}$ ). Por tanto, la evolución natural solo ha explorado una fracción diminuta del espacio de secuencia. Si imaginamos dicho espacio como un océano, la evolución natural ha explorado solo la superficie; la ingeniería de proteínas amplía esta zona, mientras que el diseño *de novo* sondea regiones más profundas (Figura 1C).

El diseño de proteínas también se conoce como el “problema inverso del plegamiento” (Figura 1B), ya que, en lugar de predecir la estructura que adoptaría una secuencia, se busca una secuencia que adopte una estructura deseada (Yue y Dill, 1992). Por ello, los avances en predicción estructural tienen un impacto directo en el diseño, y viceversa. Un ejemplo claro ocurrió a finales de 2020, cuando se reportó un hito en la ciencia de proteínas: el desarrollo del algoritmo *AlphaFold2*, capaz de predecir con alta precisión la estructura de proteínas a partir de su secuencia (Jumper et al., 2021). Este programa permite predecir la estructura de todas las secuencias depositadas en bases de datos, así como de secuencias diseñadas, ayudando a evaluar si adoptarán o no la estructura esperada. *AlphaFold2* y otros programas basados en inteligencia artificial (IA) han impulsado enormemente el diseño de proteínas *de novo*.

El Premio Nobel de Química 2024 reconoció que ambos problemas son dos caras de la misma moneda al galardonar a David Baker por el desarrollo de algoritmos para el diseño de proteínas, y a John Jumper y Demis Hassabis por el desarrollo de *AlphaFold2* para la predicción estructural (Rodríguez-Sotres y Aguayo-Ortíz, 2025).

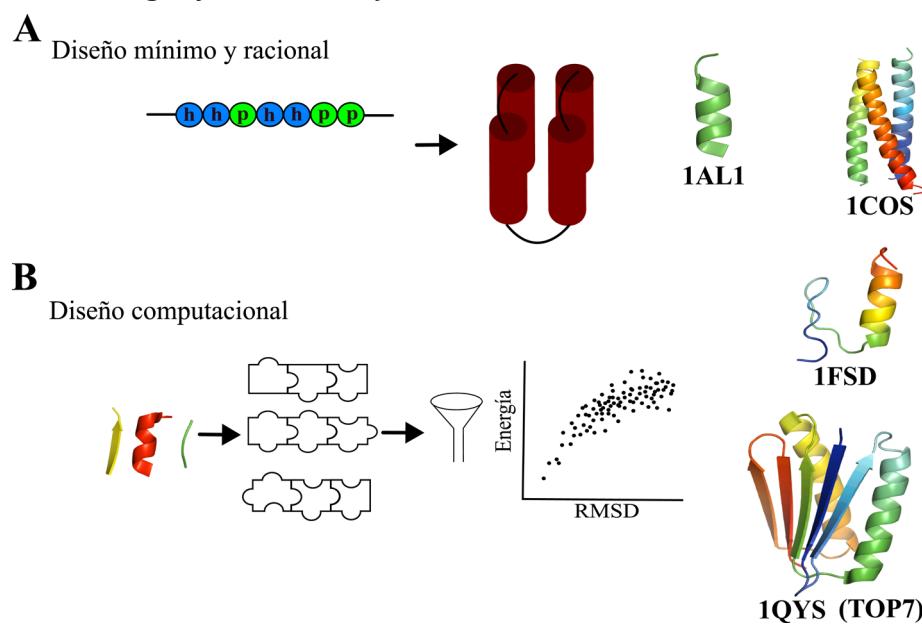
Al igual que un submarino permite explorar las profundidades del océano, el diseño de proteínas *de novo* ofrece la herramienta necesaria para ampliar la exploración del espacio de secuencia (Figura 1C). La historia de esta disciplina puede dividirse de diversas formas (Korendovych y DeGrado, 2020; Woolfson, 2021); por simplicidad, en este texto se divide en dos grandes “eras”: la primera abarca el diseño *de novo* racional y minimalista, desde los años ochenta hasta principios de los 2000; la segunda corresponde al diseño computacional, que va desde la aproximación basada en fragmentos hasta el uso de modelos de lenguaje, en desarrollo hasta la actualidad.

## Diseño *de novo* minimalista y racional de proteínas

En consonancia con la frase de Richard Feynman: “Lo que no puedo crear, no lo entiendo”, a inicios de los años ochenta surgió la idea de generar proteínas artificiales para profundizar en el conocimiento sobre la relación secuencia-estructura. En particular, se sabía que la geometría del estado nativo está estabilizada por el empaquetamiento de residuos hidrofóbicos, excluidos del solvente en el interior de la proteína, y asistida por puentes de hidrógeno formados entre grupos polares.

Por otro lado, se encontró que las cadenas laterales de los aminoácidos adoptan conformaciones particulares llamadas *rotámeros*, en función de la estructura secundaria. Con esta información, se establecieron patrones de residuos hidrofóbicos (h) y polares (p), utilizando una colección limitada de aminoácidos (Leu como h, y Lys y Glu como p), para diseñar secuencias hp que dieran lugar a pequeños péptidos con estructura secundaria, capaces de asociarse gracias al efecto hidrofóbico (Figura 1A). Este diseño minimalista fue exitoso en la obtención de estructuras helicoidales monoméricas y oligoméricas (Figura 2A) (DeGrado et al., 1989). Al incorporar información bioquímica y evolutiva a este diseño, fue posible construir *de novo* una proteína con capacidad de unir iones Zn<sup>2+</sup> (Handel y DeGrado, 1990; Regan y Clarke, 1990).

**FIGURA 2.** Estrategias de diseño de proteínas *de novo*. (A) Diseño mínimo y racional. Patrones de residuos hidrofóbicos (h) y polares (p) útiles para el diseño de  $\alpha$ -hélices. Se muestran las estructuras depositadas en el PDB con los códigos 1AL1 y 1COS. (B) Diseño computacional. A partir de esta estrategia fue posible obtener proteínas funcionales (1FSD), así como emplear pequeños fragmentos de estructura secundaria que generan un gran número de modelos, los cuales se evalúan mediante el software Rosetta con base en su energía y su similitud estructural, a través de un valor denominado RMSD. Los diseños con mayor probabilidad de éxito muestran un gráfico Energía vs. RMSD con forma de embudo (Leaver-Fay, 2011). Este enfoque permitió desarrollar una gran cantidad de proteínas *de novo*, incluida TOP7, una proteína con un plegamiento jamás observado en la naturaleza.



## Diseño computacional de proteínas *de novo*

El desarrollo de algoritmos computacionales que consideran el reempaquetamiento del núcleo hidrofóbico y la minimización de energía impulsó una nueva era en el diseño de proteínas *de novo*, gracias a los trabajos de Handel (Desjarlais y Handel, 1995b) y Mayo (Dahiyat et al., 1997; Street y Mayo, 1999). El diseño computacional de proteínas es un problema de optimización: dada la estructura del esqueleto a diseñar, se buscan las secuencias de aminoácidos que adopten la conformación de menor energía. El muestreo continúa hasta que ya no es posible encontrar una secuencia con menor energía.

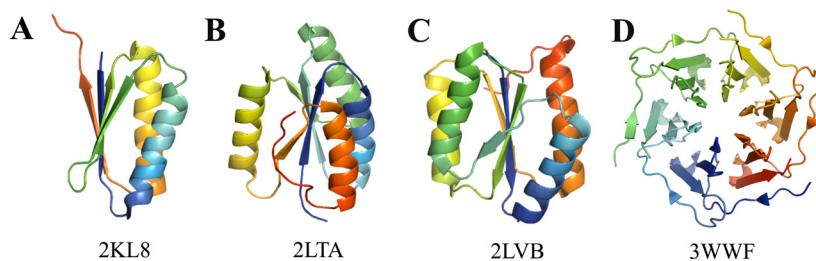
Con esta metodología se diseñaron diversas proteínas, siendo representativa la proteína tipo dedo de zinc “1FSD”, formada por 28 residuos (Dahiyat et al., 1997). Esta

proteína fue el primer rediseño computacional completo del *backbone* de una estructura existente en la naturaleza (Figura 2B).

Como complemento a esta aproximación, se desarrolló Rosetta, una suite de programas de diseño computacional que incluye la estrategia de ensamble de fragmentos. Para ello, se utiliza una librería de pequeños fragmentos de 3 y 9 residuos, extraídos del PDB, con longitudes y orientaciones relativas bien definidas. Tanto la selección del esqueleto como la de la secuencia se evalúan mediante criterios energéticos y estructurales (Figura 2B).

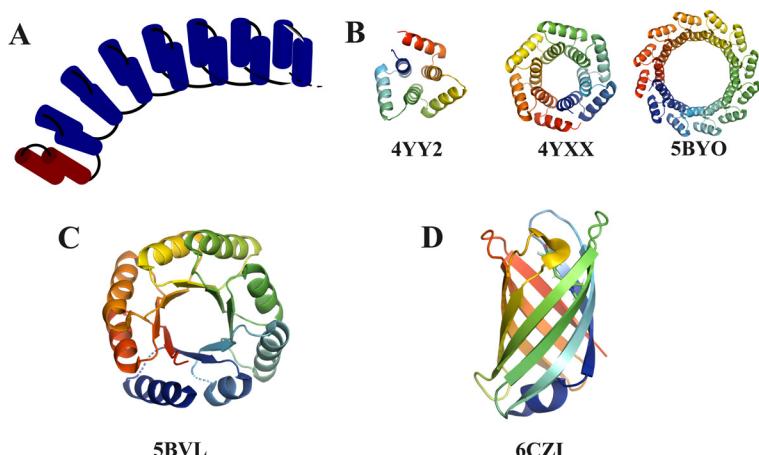
Mediante este enfoque, en 2003 se logró un hito en el diseño al validar experimentalmente una proteína con un plegamiento jamás visto en la naturaleza: TOP7 (1QYS; Kuhlman et al., 2003). Gracias a esta metodología, implementada principalmente en la suite Rosetta, se ha incrementado exponencialmente el número de proteínas diseñadas *de novo* (Huang et al., 2016). En conjunto, este trabajo del Dr. David Baker lo hizo merecedor del Premio Nobel de Química 2024.

TOP7 incluyó el diseño de hebras beta. La incorporación de este tipo de estructuras permitió el desarrollo de diversas proteínas  $\alpha/\beta$ , como los plegamientos tipo ferredoxina, bucle-P, Rossmann (Figura 3) (Lin et al., 2015; Koga et al., 2012; Marcos et al., 2017), y la serie de proteínas tipo pizza (Voet et al., 2014).



**FIGURA 3.** Proteínas *de novo* con plegamiento tipo  $\alpha\beta$ . Plegamiento tipo (A) ferredoxina, (B) bucle-P, (C) Rossmann y (D) tipo pizza.

Otra variante del diseño de proteínas *de novo* consiste en generar un pequeño segmento con estructura secundaria definida y repetirlo numerosas veces, ya sea de manera lineal o para formar una estructura de conformación cerrada cíclica (Figura 4). En este segundo caso, es posible diseñar estructuras donde la primera y la última unidad estructural interactúan. Por ejemplo, en los toroides  $\alpha$ -helicoidales (Doyle et al., 2015), el barril TIM y el barril  $\beta$  (Huang et al., 2016b; Dou et al., 2018).



**FIGURA 4.** Diseño de proteínas a través de unidades repetidas. (A) La repetición de unidades ideales puede generar proteínas de gran tamaño. (B) Toroides  $\alpha$ -hélice *de novo*. (C) Barril TIM *de novo*. (D) Barril  $\beta$  *de novo*.

La topología tipo barril TIM fue durante muchos años un objetivo de diseño, ya que es la topología más común en las enzimas. Su estructura consiste en ocho hebras  $\beta$  paralelas en el interior del barril, rodeadas por ocho hélices  $\alpha$ , con un motivo  $\beta$ -asa- $\alpha$  que se repite ocho veces. El primer diseño exitoso de un barril TIM *de novo* se logró en 2016 (Huang et al., 2016). Posteriormente, esta proteína sirvió como plataforma para rediseños subsecuentes que incrementaron su estabilidad (Romero-Romero et al., 2021), añadieron extensiones de estructura secundaria y terciaria (Beck et al., 2024; Caldwell et al., 2020), incorporaron nuevas características (Kordes et al., 2022) y modificaron su simetría interna (Koch et al., 2024). No obstante, hasta la fecha no ha sido posible generar un barril TIM *de novo* con actividad catalítica.

En cuanto al barril  $\beta$ , después de varias rondas de optimización, se logró diseñar uno con capacidad de unir un compuesto fluorescente (Dou et al., 2018).

### Proteínas funcionales diseñadas *de novo*

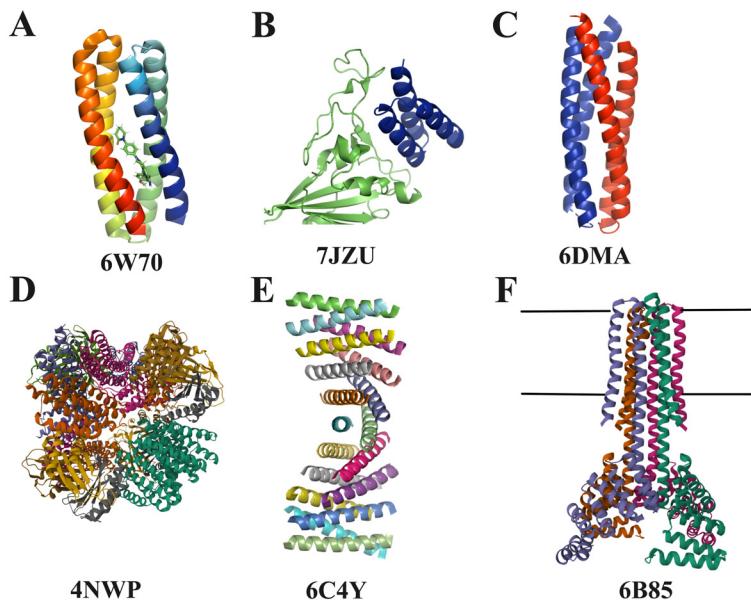
En sus orígenes, el diseño funcional de proteínas se realizaba mediante la inserción de motivos estructurales funcionales nativos; es decir, el arreglo tridimensional de grupos funcionales en el sitio activo de una proteína natural se incrustaba en un andamiaje diseñado *de novo*, como en el caso de la proteína tipo dedo de zinc “1FSD” (Dahiyat et al., 1997). Sin embargo, con el paso del tiempo, y gracias a nuevas estrategias y metodologías, ha sido posible generar proteínas funcionales completamente *de novo* (Pan y Kortemme, 2021).

Esto nos lleva a dos principales caminos en la generación de proteínas funcionales: el primero consiste en diseñar una proteína *de novo* que sirva como andamiaje y, posteriormente, rediseñarla para una función específica. Un ejemplo claro es el diseño *de novo* de un barril  $\beta$  fluorescente (Dou et al., 2018). El segundo camino considera la función deseada desde el inicio. Por ejemplo, Polizzi et al. reportaron el diseño de una proteína con cuatro hélices  $\alpha$  con capacidad de unir porfirina, mediante el diseño simultáneo del núcleo hidrofóbico y del sitio de unión (Polizzi et al., 2017). Ambos enfoques han sido exitosos, aunque, por practicidad, suele preferirse el primero.

La unión de un ligando suele ser la primera función por diseñar. Para ello, se requiere un método de muestreo y evaluación de afinidad que permita crear el sitio de unión más adecuado. Diversos diseños han tenido éxito, como el ya mencionado barril  $\beta$ , que utiliza un método llamado *RIF docking*, mediante el cual se genera un ensamble de miles de millones de posicionamientos posibles de cadenas laterales de aminoácidos que favorecen las interacciones con el ligando (Dou et al., 2018). Otro ejemplo es el método *COMBS* (*Convergent Motifs for Binding Sites*), en el cual se divide al ligando de interés en fragmentos (pequeñas subestructuras) y se buscan los residuos que puedan interactuar con cada uno para generar el sitio de unión. Este algoritmo ha sido aplicado para diseñar proteínas capaces de unir fotooxidantes (Mann et al., 2025), anticoagulantes (apixabán; Figura 5; Polizzi y DeGrado, 2020) y otros fármacos con alta afinidad (Lu et al., 2024).

En ocasiones, la función de las proteínas requiere interacción con otras proteínas. Mediante análisis paralelos masivos, ha sido posible diseñar miles de pequeñas proteínas con potencial para unirse a la hemaglutinina del virus de influenza y a la neurotoxina botulínica tipo B. Para ello, se superpusieron segmentos helicoidales del andamiaje base sobre la superficie de las hélices de las proteínas blanco, a fin de identificar los residuos clave y maximizar la afinidad (Chevalier et al., 2017). Por otra parte, utilizando el método

*RIF docking*, se diseñaron inhibidores de la proteína Spike del SARS-CoV-2 con muy alta afinidad (Cao et al., 2020), siendo esta colección una de las favoritas del Dr. David Baker (según su entrevista durante la notificación del Premio Nobel de Química 2024).

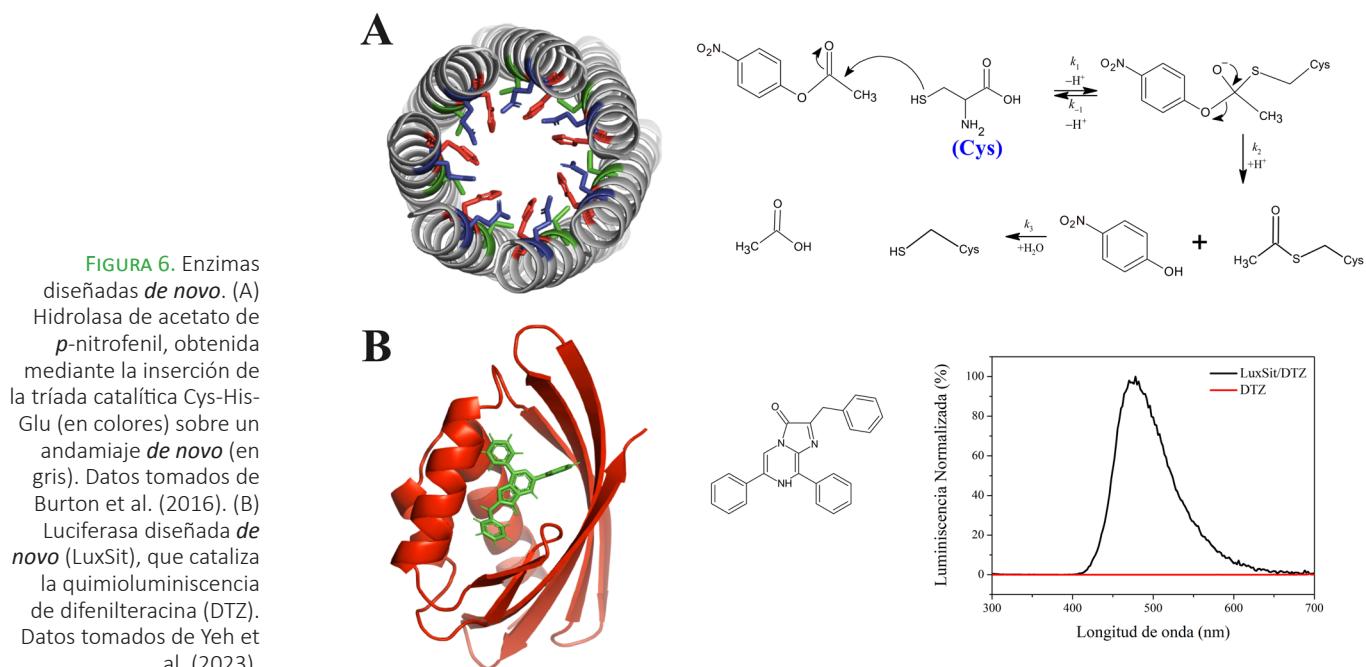


**FIGURA 5.** Proteínas funcionales diseñadas *de novo*. (A) Proteína de unión a apixabán. (B) Proteína *de novo* (en azul) con afinidad por la proteína Spike del virus SARS-CoV-2. (C) Heterodímero de proteínas *de novo*. (D) Nanocompartimento proteico *de novo*. (E) Estructura tipo amiloide a partir de proteínas *de novo*. (F) Proteína transmembrana multipaso *de novo*.

Asimismo, el diseño *de novo* de estructuras supramoleculares ha sido posible gracias al diseño de superficies proteína-proteína, lo que ha permitido generar agregados formados por proteínas idénticas (homooligómeros) o diferentes (heterooligómeros) (Fallas et al., 2017; Chen et al., 2019), nanocompartimentos proteicos (King et al., 2014) y estructuras tipo amiloide (Zhang et al., 2018; Figura 5). Otras hazañas incluyen el diseño de proteínas de membrana (Lu et al., 2018) y de proteínas tipo interruptor, que cambian de conformación en respuesta a un estímulo externo, como el pH (Boyken et al., 2019) o la presencia de iones metálicos (Wei et al., 2020; Figura 5).

Una de las funciones primordiales de las proteínas, la actividad catalítica, ha representado un gran reto en el diseño *de novo* (Dawson et al., 2019). Aunque los avances en esta área han sido limitados, son relevantes. Por ejemplo, el grupo de Woolfson reportó el diseño exitoso de una enzima *de novo*, mediante la inserción de una tríada catalítica Cys-His-Glu sobre un andamiaje *de novo*, lo que resultó en una proteína con actividad hidrolasa (Burton et al., 2016; Figura 6A). Otros ejemplos exitosos son las metaloenzimas artificiales (Lovelock et al., 2022) y el uso de aminoácidos no canónicos para la catálisis (Burke et al., 2019).

En los últimos años, gracias al auge de la inteligencia artificial, en especial de los modelos de aprendizaje profundo (*deep learning*), se ha abierto una nueva era en el diseño de proteínas funcionales *de novo*. Por ejemplo, mediante la aplicación de un método llamado “alucinación”, se diseñó una proteína capaz de catalizar selectivamente la quimioluminiscencia oxidativa de la difenilteracina (DTZ), un sustrato sintético de luciferina, con una eficiencia catalítica comparable a la de las luciferasas nativas, pero con mayor especificidad por el sustrato (Figura 6B; Yeh et al., 2023). El campo de la inteligencia artificial sigue en desarrollo y ha brindado resultados emocionantes; se recomienda al lector consultar los trabajos de Khakzad et al. (2024), Chu et al. (2024) y Kortemme (2024) para profundizar en el tema.



**FIGURA 6.** Enzimas diseñadas *de novo*. (A) Hidrolasa de acetato de *p*-nitrofenil, obtenida mediante la inserción de la triada catalítica Cys-His-Glu (en colores) sobre un andamiaje *de novo* (en gris). Datos tomados de Burton et al. (2016). (B) Luciferasa diseñada *de novo* (LuxSIt), que cataliza la quimioluminiscencia de difenilteracina (DTZ). Datos tomados de Yeh et al. (2023).

## Perspectivas

En general, el diseño de proteínas *de novo* ha sido un campo fascinante y complejo que ha puesto a prueba el conocimiento sobre fisicoquímica de proteínas, a la vez que ha contribuido a su mejor comprensión. Tal ha sido su impacto que su principal exponente fue reconocido con el más reciente Premio Nobel de Química.

Aunque los avances han sido notables, aún quedan muchos retos por resolver. Se espera que el desarrollo de nuevos algoritmos basados en aprendizaje profundo permita superar desafíos actuales. Uno de los principales es el bajo porcentaje de éxito: para obtener una función o estructura específica, es común requerir múltiples “rondas” de diseño, en las que deben analizarse experimentalmente decenas o centenas de secuencias —un proceso costoso— hasta encontrar una que cumpla con los criterios originales. Esto es especialmente relevante en el diseño *de novo* de enzimas, donde los casos exitosos aún son escasos.

Además, la aplicación de estas proteínas como agentes terapéuticos representa un campo innovador de alto impacto, gracias a su potencial como agentes modulables, económicos, eficaces y biocompatibles. La capacidad de diseñar catalizadores específicos permitirá su uso en diversos sectores industriales. Aunado a las reacciones enzimáticas conocidas, se podrán diseñar nuevas enzimas capaces de llevar a cabo reacciones químicas inéditas en los seres vivos.

Es importante señalar que, como cualquier otra tecnología, el diseño de proteínas también puede usarse con fines poco éticos, como la creación de toxinas, enzimas u otras proteínas con potencial como armas biológicas. Por ello, es fundamental legislar sobre el tema y establecer mecanismos de seguridad que limiten o impidan su uso militar.

Por último, queda por comprobar si los nuevos algoritmos basados en inteligencia artificial serán lo suficientemente precisos como para permitir el diseño de proteínas exitoso en laboratorios de todo el mundo, y no solo en aquellos con grandes recursos. A pesar de

las limitaciones y de los posibles usos indebidos, el diseño de proteínas es una tecnología emergente con un enorme potencial para atender las necesidades en alimentación, salud y biorremediación que enfrentamos como especie.

## Referencias

- Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science*, 181(4096), 223–230. <https://doi.org/10.1126/science.181.4096.223>
- Beck, J., Shanmugaratnam, S., y Höcker, B. (2024). Diversifying *de novo* TIM barrels by hallucination. *Protein Science*, 33(6), e5001. <https://doi.org/10.1002/pro.5001>
- Beck, M., Schmidt, A., Malmstroem, J., Claassen, M., Ori, A., Szymborska, A., Herzog, F., Rinner, O., Ellenberg, J., y Aebersold, R. (2011). The quantitative proteome of a human cell line. *Molecular Systems Biology*, 7(1), 549. <https://doi.org/10.1038/msb.2011.82>
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., y Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Research*, 28(1), 235–242. <https://doi.org/10.1093/nar/28.1.235>
- Boyken, S. E., Benhaim, M. A., Busch, F., Jia, M., Bick, M. J., Choi, H., Klima, J. C., Chen, Z., Walkey, C., Mileant, A., Sahasrabuddhe, A., Hodge, E. A., Byron, S., Quijano-Rubio, A., Sankaran, B., King, N. P., Lippincott-Schwartz, J., Wysocki, V. H., ... y Baker, D. (2019). *De novo* design of tunable, pH-driven conformational changes. *Science*, 364(6441), 658–664. <https://doi.org/10.1126/science.aav7897>
- Burke, A. J., Lovelock, S. L., Frese, A., Crawshaw, R., Ort Mayer, M., Dunstan, M., Levy, C., y Green, A. P. (2019). Design and evolution of an enzyme with a non-canonical organocatalytic mechanism. *Nature*, 570(7760), 219–223. <https://doi.org/10.1038/s41586-019-1262-8>
- Burley, S. K., Bhatt, R., Bhikadiya, C., Bi, C., Biester, A., Biswas, P., Bittrich, S., Blaumann, S., Brown, R., Chao, H., Chithari, V. R., Craig, P. A., Crichlow, G. V., Duarte, J. M., Dutta, S., Feng, Z., Flatt, J. W., Ghosh, S., Goodsell, D. S., ... y Zardecki, C. (2025). Updated resources for exploring experimentally-determined PDB structures and Computed Structure Models at the RCSB Protein Data Bank. *Nucleic Acids Research*, 53(D1), D564–D574. <https://doi.org/10.1093/nar/gkae1091>
- Burton, A. J., Thomson, A. R., Dawson, W. M., Brady, R. L., y Woolfson, D. N. (2016). Installing hydrolytic activity into a completely *de novo* protein framework. *Nature Chemistry*, 8(9), 837–844. <https://doi.org/10.1038/nchem.2555>
- Caldwell, S. J., Haydon, I. C., Piperidou, N., Huang, P. S., Bick, M. J., Sjöström, H. S., Hilvert, D., Baker, D., y Zeymer, C. (2020). Tight and specific lanthanide binding in a *de novo* TIM barrel with a large internal cavity designed by symmetric domain fusion. *Proceedings of the National Academy of Sciences*, 117(48), 30362–30369. <https://doi.org/10.1073/pnas.2008535117>
- Cao, L., Goreshnik, I., Coventry, B., Case, J. B., Miller, L., Kozodoy, L., Chen, R. E., Carter, L., Walls, A. C., Park, Y. J., Strauch, E. M., Stewart, L., Diamond, M. S., Veesler, D., y Baker, D. (2020). De *novo* design of picomolar SARS-CoV-2 miniprotein inhibitors. *Science*, 370(6515), 426–431. <https://doi.org/10.1126/science.abd9909>

- Chen, Z., Boyken, S. E., Jia, M., Busch, F., Flores-Solis, D., Bick, M. J., Lu, P., VanAernum, Z. L., Sahasrabuddhe, A., Langan, R. A., Bermeo, S., Brunette, T. J., Mulligan, V. K., Carter, L. P., DiMaio, F., Sgourakis, N. G., Wysocki, V. H., y Baker, D. (2019). Programmable design of orthogonal protein heterodimers. *Nature*, 565(7737), 106–111. <https://doi.org/10.1038/s41586-018-0802-y>
- Chevalier, A., Silva, D. A., Rocklin, G. J., Hicks, D. R., Vergara, R., Murapa, P., Bernard, S. M., Zhang, L., Lam, K. H., Yao, G., Bahl, C. D., Miyashita, S. I., Goreshnik, I., Fuller, J. T., Koday, M. T., Jenkins, C. M., Colvin, T., Carter, L., Bohn, A., ... y Baker, D. (2017). Massively parallel de novo protein design for targeted therapeutics. *Nature*, 550(7674), 74–79. <https://doi.org/10.1038/nature23912>
- Chu, A. E., Lu, T., y Huang, P. S. (2024). Sparks of function by de novo protein design. *Nature Biotechnology*, 42(2), 203–215. <https://doi.org/10.1038/s41587-024-02133-2>
- Dahiyat, B. I., y Mayo, S. L. (1997). De novo protein design: Fully automated sequence selection. *Science*, 278(5335), 82–87. <https://doi.org/10.1126/science.278.5335.82>
- Dahiyat, B. I., Sarisky, C. A., y Mayo, S. L. (1997). De novo protein design: Towards fully automated sequence selection. *Journal of Molecular Biology*, 273(4), 789–796. <https://doi.org/10.1006/jmbi.1997.1341>
- Dawson, W. M., Rhys, G. G., y Woolfson, D. N. (2019). Towards functional de novo designed proteins. *Current Opinion in Chemical Biology*, 52, 102–111. <https://doi.org/10.1016/j.cbpa.2019.06.011>
- DeGrado, W. F., Wasserman, Z. R., y Lear, J. D. (1989). Protein design, a minimalist approach. *Science*, 243(4891), 622–628. <https://doi.org/10.1126/science.2464850>
- Desjarlais, J. R., y Handel, T. M. (1995). De novo design of the hydrophobic cores of proteins. *Protein Science*, 4(10), 2006–2018. <https://doi.org/10.1002/pro.5560041006>
- Desjarlais, J. R., y Handel, T. M. (1995b). New strategies in protein design. *Current Opinion in Biotechnology*, 6(4), 460–466. [https://doi.org/10.1016/0958-1669\(95\)80076-X](https://doi.org/10.1016/0958-1669(95)80076-X)
- Dill, K. A. (1990). Dominant forces in protein folding. *Biochemistry*, 29(31), 7133–7155. <https://doi.org/10.1021/bi00483a001>
- Dill, K. A., y Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nature Structural Biology*, 4(1), 10–19. <https://doi.org/10.1038/nsb0197-10>
- Dou, J., Vorobieva, A. A., Sheffler, W., Doyle, L. A., Park, H., Bick, M. J., Mao, B., Foight, G. W., Lee, M. Y., Gagnon, L. A., Carter, L., Sankaran, B., Ovchinnikov, S., Marcos, E., Huang, P. S., Vaughan, J. C., Stoddard, B. L., y Baker, D. (2018). De novo design of a fluorescence-activating β-barrel. *Nature*, 561(7724), 485–491. <https://doi.org/10.1038/s41586-018-0509-0>
- Doyle, L., Hallinan, J., Bolduc, J., Parmeggiani, F., Baker, D., Stoddard, B. L., y Bradley, P. (2015). Rational design of α-helical tandem repeat proteins with closed architectures. *Nature*, 528(7583), 585–588. <https://doi.org/10.1038/nature16191>

- Fallas, J. A., Ueda, G., Sheffler, W., Nguyen, V., McNamara, D. E., Sankaran, B., Pereira, J. H., Parmeggiani, F., Brunette, T. J., Cascio, D., Yeates, T. R., Zwart, P., y Baker, D. (2017). Computational design of self-assembling cyclic protein homo-oligomers. *Nature Chemistry*, 9(4), 353–360. <https://doi.org/10.1038/nchem.2673>
- Handel, T., y DeGrado, W. F. (1990). De novo design of a Zn<sup>2+</sup>-binding protein. *Journal of the American Chemical Society*, 112(18), 6710–6711. <https://doi.org/10.1021/ja00174a039>
- Huang, P. S., Boyken, S. E., y Baker, D. (2016). The coming of age of de novo protein design. *Nature*, 537(7620), 320–327. <https://doi.org/10.1038/nature19946>
- Huang, P. S., Feldmeier, K., Parmeggiani, F., Fernandez Velasco, D. A., Höcker, B., y Baker, D. (2016b). De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nature Chemical Biology*, 12(1), 29–34. <https://doi.org/10.1038/nchembio.1966>
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... y Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>
- Khakzad, H., Igashov, I., Schneuing, A., Goverde, C., Bronstein, M., y Correia, B. (2023). A new age in protein design empowered by deep learning. *Cell Systems*, 14(11), 925–939. <https://doi.org/10.1016/j.cels.2023.10.006>
- King, N. P., Bale, J. B., Sheffler, W., McNamara, D. E., Gonen, S., Gonen, T., Yeates, T. O., y Baker, D. (2014). Accurate design of co-assembling multi-component protein nanomaterials. *Nature*, 510(7503), 103–108. <https://doi.org/10.1038/nature13404>
- Koch, J. S., Romero-Romero, S., y Höcker, B. (2024). Stepwise introduction of stabilizing mutations reveals nonlinear additive effects in de novo TIM barrels. *Protein Science*, 33(3), e4926. <https://doi.org/10.1002/pro.4926>
- Koga, N., Tatsumi-Koga, R., Liu, G., Xiao, R., Acton, T. B., Montelione, G. T., y Baker, D. (2012). Principles for designing ideal protein structures. *Nature*, 491(7423), 222–227. <https://doi.org/10.1038/nature11600>
- Kolodny, R., Pereyaslavets, L., Samson, A. O., y Levitt, M. (2013). On the universe of protein folds. *Annual Review of Biophysics*, 42(1), 559–582. <https://doi.org/10.1146/annurev-biophys-083012-130432>
- Kordes, S., Romero-Romero, S., Lutz, L., y Höcker, B. (2022). A newly introduced salt bridge cluster improves structural and biophysical properties of de novo TIM barrels. *Protein Science*, 31(2), 513–527. <https://doi.org/10.1002/pro.4249>
- Korendovych, I. V., y DeGrado, W. F. (2020). De novo protein design, a retrospective. *Quarterly Reviews of Biophysics*, 53, e3. <https://doi.org/10.1017/S0033583519000131>
- Kortemme, T. (2024). De novo protein design—From new structures to programmable functions. *Cell*, 187(3), 526–544. <https://doi.org/10.1016/j.cell.2023.12.028>

Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L., y Baker, D. (2003). Design of a novel globular protein fold with atomic-level accuracy. *Science*, 302(5649), 1364-1368. <https://doi.org/10.1126/science.1089427>

Leaver-Fay, A., Tyka, M., Lewis, S. M., Lange, O. F., Thompson, J., Jacak, R., Kaufmann, K. W., Renfrew, P. D., Smith, C. A., Sheffler, W., Davis, I. W., Cooper, S., Treuille, A., Mandell, D. J., Richter, F., Ban, Y. E. A., Fleishman, S. J., Corn, J. E., Kim, D. E., ... y Bradley, P. (2011). ROSETTA3: An object-oriented software suite for the simulation and design of macromolecules. In *Methods in Enzymology* (Vol. 487, pp. 545-574). Academic Press. <https://doi.org/10.1016/B978-0-12-381270-4.00019-6>

Levinthal, C. (1969). How to fold graciously. *Mossbauer Spectroscopy in Biological Systems*, 67, 22-24.

Lin, Y. R., Koga, N., Tatsumi-Koga, R., Liu, G., Clouser, A. F., Montelione, G. T., y Baker, D. (2015). Control over overall shape and size in de novo designed proteins. *Proceedings of the National Academy of Sciences*, 112(40), E5478-E5485. <https://doi.org/10.1073/pnas.1509508112>

Lovelock, S. L., Crawshaw, R., Basler, S., Levy, C., Baker, D., Hilvert, D., y Green, A. P. (2022). The road to fully programmable protein catalysis. *Nature*, 606(7912), 49-58. <https://doi.org/10.1038/s41586-022-04456-z>

Lu, L., Gou, X., Tan, S. K., Mann, S. I., Yang, H., Zhong, X., Gazgalis, D., Valdiviezo, J., Jo, H., Wu, Y., Diolaiti, M. E., Ashworth, A., Polizzi, N. F., y DeGrado, W. F. (2024). De novo design of drug-binding proteins with predictable binding energy and specificity. *Science*, 384(6691), 106-112. <https://doi.org/10.1126/science.adl5364>

Lu, P., Min, D., DiMaio, F., Wei, K. Y., Vahey, M. D., Boyken, S. E., Chen, Z., Fallas, J. A., Ueda, G., Sheffler, W., Mulligan, V. K., Xu, W., Bowie, J. U., y Baker, D. (2018). Accurate computational design of multipass transmembrane proteins. *Science*, 359(6379), 1042-1046. <https://doi.org/10.1126/science.aaq1739>

Mann, S. I., Lin, Z., Tan, S. K., Zhu, J., Widel, Z. X. W., Bakanas, I., Mansergh, J. P., Liu, R., Kelly, M. J. S., Wu, Y., Wells, J. A., Therien, M. J., y DeGrado, W. F. (2025). De novo design of proteins that bind naphthalenediimides, powerful photooxidants with tunable photophysical properties. *Journal of the American Chemical Society*, 147(9), 7849-7858. <https://doi.org/10.1021/jacs.4c18151>

Marcos, E., Chidyausiku, T. M., McShan, A. C., Evangelidis, T., Nerli, S., Carter, L., Nivón, L. G., Davis, A., Oberdorfer, G., Tripsianes, K., Sgourakis, N. G., y Baker, D. (2018). De novo design of a non-local  $\beta$ -sheet protein with high stability and accuracy. *Nature Structural & Molecular Biology*, 25(11), 1028-1034. <https://doi.org/10.1038/s41594-018-0141-6>

Pan, X., y Kortemme, T. (2021). Recent advances in de novo protein design: Principles, methods, and applications. *Journal of Biological Chemistry*, 296, 100558. <https://doi.org/10.1016/j.jbc.2021.100558>

Polizzi, N. F., y DeGrado, W. F. (2020). A defined structural unit enables de novo design of small-molecule-binding proteins. *Science*, 369(6508), 1227-1233. <https://doi.org/10.1126/science.abb8330>

- Polizzi, N. F., Wu, Y., Lemmin, T., Maxwell, A. M., Zhang, S. Q., Rawson, J., Beratan, D. N., Therien, M. J., y DeGrado, W. F. (2017). De novo design of a hyperstable non-natural protein-ligand complex with sub-Å accuracy. *Nature Chemistry*, 9(12), 1157–1164. <https://doi.org/10.1038/nchem.2846>
- Regan, L., y Clarke, N. D. (1990). A tetrahedral zinc (II)-binding site introduced into a designed protein. *Biochemistry*, 29(49), 10878–10883. <https://doi.org/10.1021/bi00501a003>
- Rodriguez-Sotres, R., y Aguayo-Ortíz, R. (2025). Desentrañando los secretos de autoensamblado en las máquinas de la vida: plegamiento de proteínas y Nobel de Química 2024. *Educación Química*, 36(1), 5–13. <https://doi.org/10.22201/fq.18708404e.2025.1.89902>
- Romero-Romero, S., Costas, M., Silva Manzano, D. A., Kordes, S., Rojas-Ortega, E., Tapia, C., Guerra, Y., Shanmugaratnam, S., Rodríguez-Romero, A., Baker, D., Höcker, B., y Fernández-Velasco, D. A. (2021). The stability landscape of de novo TIM barrels explored by a modular design approach. *Journal of Molecular Biology*, 433(18), 167153. <https://doi.org/10.1016/j.jmb.2021.167153>
- Street, A. G., y Mayo, S. L. (1999). Computational protein design. *Structure*, 7(5), R105–R109. [https://doi.org/10.1016/S0969-2126\(99\)80062-8](https://doi.org/10.1016/S0969-2126(99)80062-8)
- Varadi, M., Bertoni, D., Magana, P., Paramval, U., Pidruchna, I., Radhakrishnan, M., Tsenkov, M., Nair, S., Mirdita, M., Yeo, J., Kovalevskiy, O., Tunyasuvunakool, K., Laydon, A., Žídek, A., Tomlinson, H., Hariharan, D., Abrahamson, J., Green, T., Jumper, J., ... y Velankar, S. (2024). AlphaFold Protein Structure Database in 2024: Providing structure coverage for over 214 million protein sequences. *Nucleic Acids Research*, 52(D1), D368–D375. <https://doi.org/10.1093/nar/gkad1011>
- Voet, A. R. D., Noguchi, H., Addy, C., Simoncini, D., Terada, D., Unzai, S., Park, S. Y., Zhang, K. Y. J., y Tame, J. R. H. (2014). Computational design of a self-assembling symmetrical β-propeller protein. *Proceedings of the National Academy of Sciences*, 111(42), 15102–15107. <https://doi.org/10.1073/pnas.1412768111>
- Wei, K. Y., Moschidi, D., Bick, M. J., Nerli, S., McShan, A. C., Carter, L. P., Huang, P. S., Fletcher, D. A., Sgourakis, N. G., Boyken, S. E., y Baker, D. (2020). Computational design of closely related proteins that adopt two well-defined but structurally divergent folds. *Proceedings of the National Academy of Sciences*, 117(13), 7208–7215. <https://doi.org/10.1073/pnas.1914808117>
- Woolfson, D. N. (2021). A brief history of de novo protein design: Minimal, rational, and computational. *Journal of Molecular Biology*, 433(20), 167160. <https://doi.org/10.1016/j.jmb.2021.167160>
- Woolfson, D. N., Bartlett, G. J., Burton, A. J., Heal, J. W., Niitsu, A., Thomson, A. R., y Wood, C. W. (2015). De novo protein design: How do we expand into the universe of possible protein structures?. *Current Opinion in Structural Biology*, 33, 16–26. <https://doi.org/10.1016/j.sbi.2015.05.009>

Yeh, A. H. W., Norn, C., Kipnis, Y., Tischer, D., Pellock, S. J., Evans, D., Ma, P., Lee, G. R., Zhang, J. Z., Anishchenko, I., Coventry, B., Cao, L., Dauparas, J., Halabiya, S., DeWitt, M., Carter, L., Houk, K. N., y Baker, D. (2023). De novo design of luciferases using deep learning. *Nature*, 614(7949), 774–780. <https://doi.org/10.1038/s41586-023-05696-3>

Yue, K., y Dill, K. A. (1992). Inverse protein folding problem: Designing polymer sequences. *Proceedings of the National Academy of Sciences*, 89(9), 4163–4167. <https://doi.org/10.1073/pnas.89.9.4163>

Zhang, S. Q., Huang, H., Yang, J., Kratochvil, H. T., Lolicato, M., Liu, Y., Shu, X., Liu, L., y DeGrado, W. F. (2018). Designed peptides that assemble into cross- $\alpha$  amyloid-like structures. *Nature Chemical Biology*, 14(9), 870–875. <https://doi.org/10.1038/s41589-018-0105-5>

# Espacio de secuencia de las proteínas

## Resumen

En este trabajo se describe qué es el espacio de secuencia de las proteínas y se presenta la magnitud de sus combinaciones posibles. Se proponen actividades para que el alumnado explore el diseño de proteínas *de novo* mediante la creación de secuencias aleatorias y la predicción de su estructura con herramientas de inteligencia artificial.

**Palabras clave:** Bioquímica, Premio Nobel, Química computacional, Proteínas, Inteligencia Artificial.

## Abstract

This work describes the concept of protein sequence space and presents the magnitude of its possible combinations. It proposes activities for students to explore *de novo* protein design through the creation of random sequences and the prediction of their structure using artificial intelligence tools.

**Keywords:** Biochemistry, Nobel Prize, Computational Chemistry, Proteins, Artificial Intelligence.

**E**n esta hoja didáctica se presentan algunas actividades para el alumnado y la persona docente, relacionadas con la comunicación “Proteínas de novo: del diseño a la funcionalización”, publicada en este mismo número de *Educación Química*. En especial, se busca que el estudiantado comprenda la magnitud del espacio de secuencia posible para una proteína, y que sea capaz de generar una secuencia aleatoria y predecir su estructura mediante el uso de aplicaciones basadas en inteligencia artificial (IA), en particular AlphaFold.

**¿Qué tan grande es el espacio de secuencia para una proteína?**

Para una proteína típica de 200 aminoácidos pueden existir  $20^{200}$  combinaciones posibles, lo que equivale a  $1.6 \times 10^{260}$ . Este número es mayor al estimado de átomos presentes en el universo ( $10^{82}$ ). Para comprender un poco esta cifra, podemos comenzar analizando el número de proteínas presentes en la naturaleza, que es del orden de  $10^{12}$ , es decir, un millón de millones. Aunque este número es muy grande, resulta diminuto comparado con el total de posibles secuencias ( $10^{260}$ ). Para visualizar esta diferencia, es importante recordar que se trata de exponentes, por lo que cada cero implica un orden de magnitud mayor (véase la figura 1).

**FIGURA 1.** Visualización del número de secuencias de proteínas presentes en la naturaleza (en rojo) frente al número de secuencias posibles utilizando 20 aminoácidos diferentes en una proteína de 200 aminoácidos (en negro).

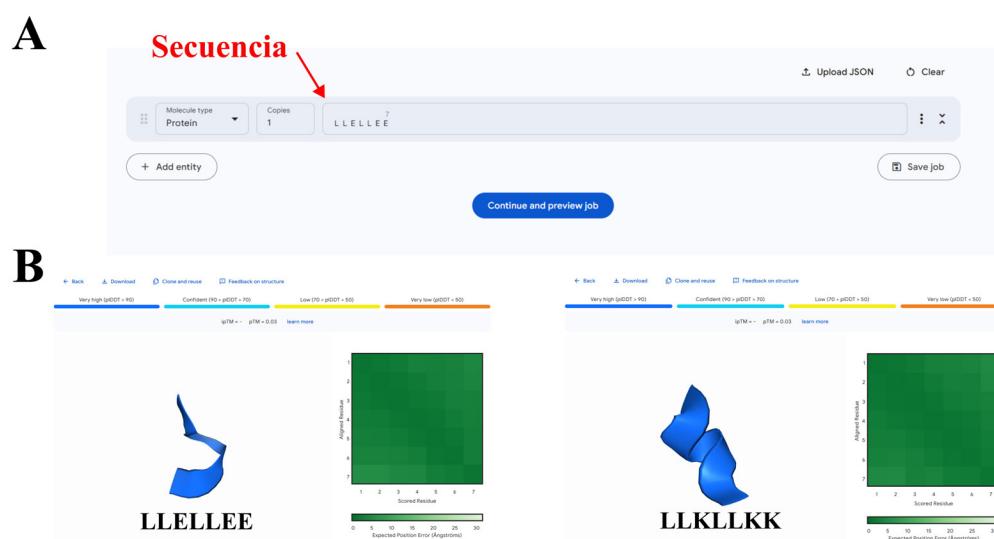
Intentemos comprender estas magnitudes en términos de distancia. Si consideramos que cada secuencia puede ocupar un picómetro ( $10^{-12}$  m), todas las secuencias naturales representarían un metro. Utilizando la misma unidad de secuencia por picómetro, y

considerando que un año luz equivale aproximadamente a  $10^{16}$  metros y que el diámetro del universo observable se estima en unos 93 mil millones de años luz ( $\sim 10^{11}$  años luz), el número de secuencias posibles abarcaría más de seis universos. En cada uno habría  $\sim 10^{39}$  secuencias. ¿Se te ocurre otra comparación entre ambas magnitudes?

### ¿Es posible generar una proteína bien estructurada con una secuencia aleatoria?

Ahora que comprendemos la magnitud del espacio de secuencia, podemos comenzar a explorarlo con secuencias pequeñas, tal como lo hizo DeGrado en los años ochenta (Korenovych y DeGrado, 2020). Antes de continuar, vale la pena recordar que las cadenas laterales de los aminoácidos son muy variadas, y pueden contener grupos hidrofóbicos (insolubles en agua) o polares (solubles en agua); entre estos últimos, algunos no tienen carga y otros presentan carga positiva o negativa.

Para diseñar una secuencia simple, utilizaremos patrones hp (hidrofóbico/polar). Por ejemplo, la leucina (L) es hidrofóbica, mientras que la lisina (K) y el ácido glutámico (E) son polares. Así, nuestras combinaciones iniciales serán: LLKLLKKy LLELLEE. Ingresaremos ambas al servidor de [AlphaFold3](#) (AF3) (Abramson et al., 2024) para predecir su estructura. Como se observa en la figura 2, en ambos casos obtenemos una pequeña hélice  $\alpha$ , lo cual era esperado.



**FIGURA 2.** Uso del servidor AlphaFold3. (A) Lugar para ingresar la secuencia de aminoácidos cuya estructura se desea predecir. (B) Resultados obtenidos para los péptidos mencionados en el texto.

Para incrementar el tamaño de la secuencia y hacerlo de forma aleatoria, podemos usar un asistente de IA. En este caso, utilizaremos Gemini (Google, 2024) y ChatGPT (OpenAI, 2024). Para obtener una secuencia congruente, se puede usar un *prompt* como el siguiente:

“Quiero que me generes una secuencia de 100 aminoácidos completamente aleatoria en código de una letra. Solo con las letras A, V, L, I, P, M, F, W, G, S, T, C, N, Q, Y, D, E, K, R y H. Sin espacio entre cada letra.”

Cada persona puede modificar la extensión de la secuencia a su criterio, aunque se sugiere aumentarla o disminuirla gradualmente para comprender mejor los resultados. En la figura 3 se muestran dos secuencias de 100 aminoácidos generadas por los asistentes de IA mencionados, así como las estructuras predichas por AF3. Estas se colorean según el valor del *pIDDT*, un parámetro que indica el grado de confiabilidad de la predicción. En el ejemplo, la predicción estructural de la secuencia generada por ChatGPT es mayoritariamente naranja, lo que indica baja confiabilidad, por lo que probablemente no adopte una estructura definida en solución. En cambio, la predicción de la secuencia generada con Gemini es mayormente azul, lo cual indica alta confiabilidad.

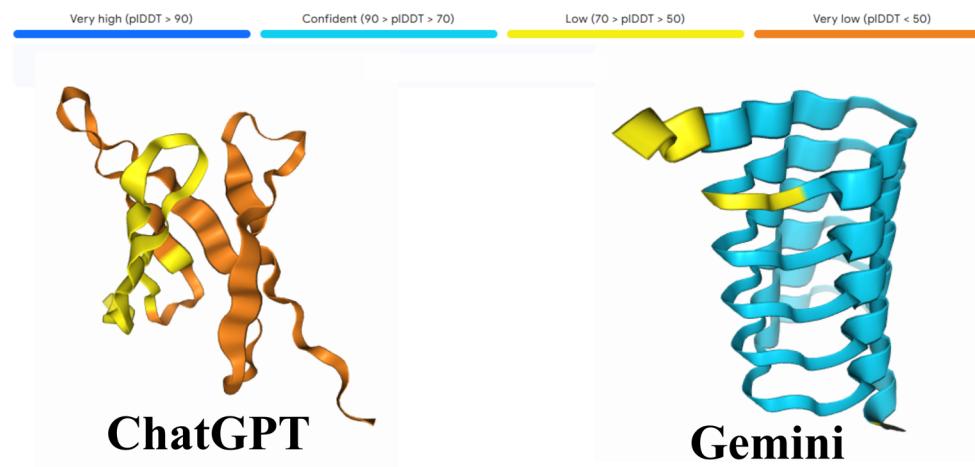
Gemini:

VPLIMFWGSCTNQYDEKRHPLIMFWGSCTNQYDEKRHPLIMFWGSCTNQYDEKRHPLIMFWGSCTNQYDEKRHPLIMFWGSCTNQYDEKRHPLIMFWGSCTNQYDEKRHPLIMFWGSCTNQYDEKRH

ChatGPT:

RLNFIATMKVPYQGHDWSICMLETKVAYFNSIWDGHTPQLRKFMCVYHSAIQDWWTGFLVPEKNSMIRTY QCAVWLPHKD

**FIGURA 3.** En la parte superior se muestran las secuencias aleatorias generadas por Gemini (Google, 2024) y ChatGPT (OpenAI, 2024), con sus estructuras predichas por AlphaFold3 (Abramson et al., 2024). En la parte media se indica la escala de confiabilidad de la predicción: el azul señala regiones que probablemente adoptarán la estructura mostrada, mientras que el naranja corresponde a regiones con baja confianza en la predicción.



Aunque la probabilidad de obtener una proteína bien plegada de manera aleatoria es baja, es interesante subrayar el potencial de las nuevas herramientas de inteligencia artificial. De hecho, AF3 se utiliza para predecir si una secuencia artificial puede adoptar la estructura deseada. No obstante, aún no existe un software o servidor que permita diseñar proteínas con unos simples clics, ya que las herramientas en desarrollo requieren conocimientos informáticos avanzados.

Para finalizar, el alumnado y la persona docente pueden comparar las estructuras predichas a partir de las secuencias generadas, y seleccionar aquellas con los mejores valores de confiabilidad. Además, dada la infinidad de combinaciones posibles, se puede explorar el espacio de secuencia de otras maneras. Por ejemplo: ¿qué pasa si repetimos 10, 15 o 20 veces el péptido LLKLLKK?, ¿qué sucede si agregamos 3 o 4 residuos de alanina o glicina entre estas repeticiones?, ¿y si añadimos una prolina entre cada repetición?, ¿si cambiamos L y K por otros residuos hidrofóbicos y polares obtendremos algo similar?, ¿existirá algún patrón de aminoácidos que genere una estructura tipo  $\beta$ ?

Como se puede ver, hay muchas preguntas interesantes por plantear. Cada estudiante es libre de elegir, cuestionar, generar y aplicar las ideas que surjan.

## Referencias

- Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D., Hung, C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., ... y Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 1–3. <https://doi.org/10.1038/s41586-024-07487-w>
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., y Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Research*, 28(1), 235–242. <https://doi.org/10.1093/nar/28.1.235>
- Google. (2024). *Gemini* (versión del 7 de noviembre). <https://gemini.google.com/app>
- Korendovych, I. V., y DeGrado, W. F. (2020). De novo protein design, a retrospective. *Quarterly Reviews of Biophysics*, 53, e3. <https://doi.org/10.1017/S0033583519000131>
- OpenAI. (2024). *ChatGPT* (versión del 7 de noviembre). <https://chatgpt.com/>