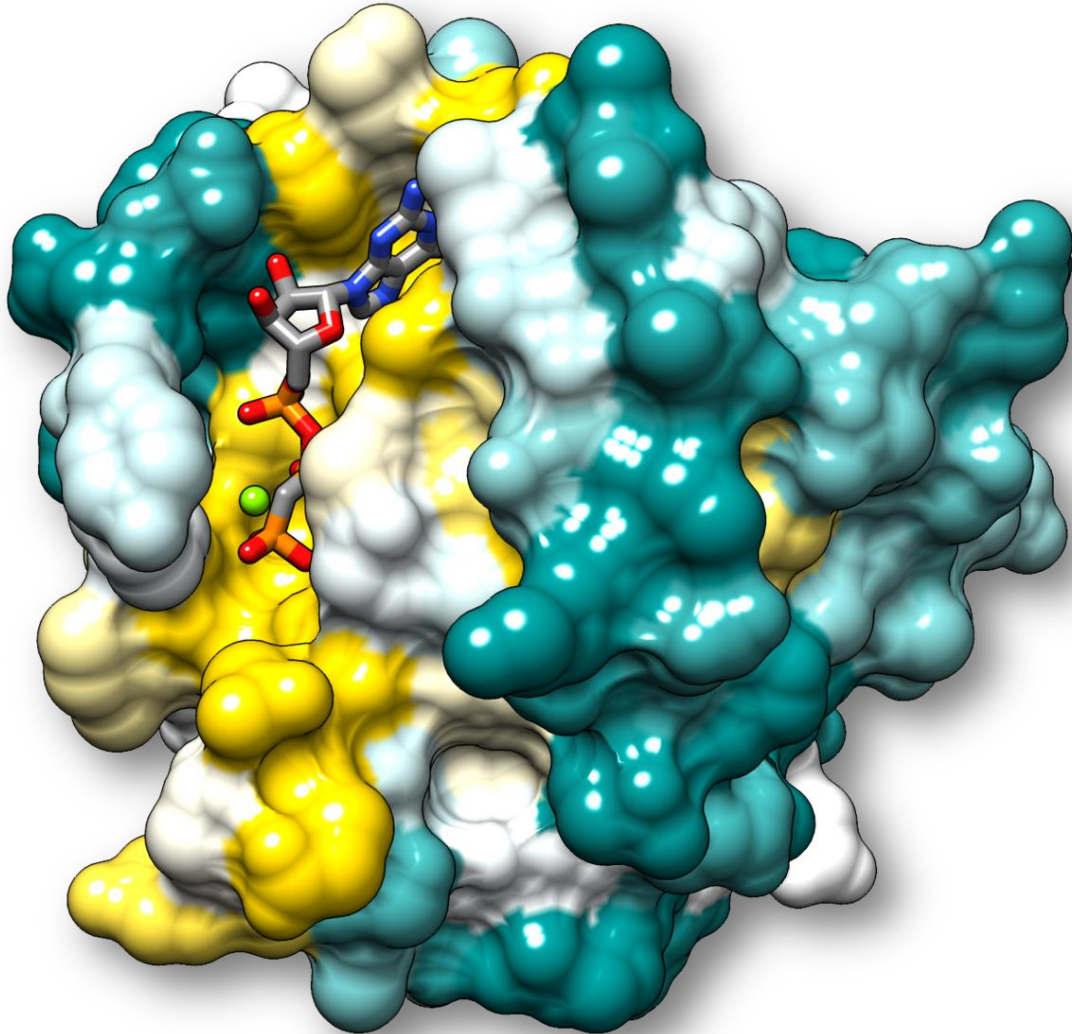


# DE LOS VIDEOJUEGOS A LA ESTRUCTURA TRIDIMENSIONAL DE LAS PROTEÍNAS



Tras más de medio siglo, un equipo de investigadores de [DeepMind](#) afirma haber resuelto el rompecabezas del plegamiento de las proteínas, esto es, los códigos internos que determinan su tridimensionalidad.

Las proteínas son secuencias de aminoácidos que se retuercen espacialmente. A la secuencia de aminoácidos se le denomina *estructura primaria*. A los plegamientos de esta secuencia se les nombra como *estructura secundaria* y *terciaria*. Cuando una proteína está formada por varias cadenas (verbigracia, la insulina) los distintos monómeros (dos en el caso de la insulina) se ovillan

especialmente entre ellos: *estructura cuaternaria*. Se conocían tiempo ha, las dos estructuras secundarias básicas (*hélice- $\alpha$*  y *hoja plegada  $\beta$* ). Sin embargo el plegamiento a larga distancia de las estructuras tridimensionales terciaria y cuaternaria apenas se había estudiado; y aún menos las interacciones en que se fundamentaban.

El problema del plegamiento de las proteínas despertó el interés de la comunidad científica a partir de la década de 1960 gracias a las investigaciones del [matrimonio Hauptman](#) (Herbert e Isabella) y [Jerome Karle](#) quienes desarrollaron y experimentaron con una técnica (cristalografía de rayos X) que hizo posible desentrañar la estructura tridimensional de moléculas cada vez más complejas. [Herbert Hauptman](#) y [Jerome Karle](#) fueron reconocidos con el [Premio Nobel de Química en 1985](#) por sus investigaciones sobre la estructura cristalina de grandes moléculas.

El químico norteamericano [Linus Pauling](#), doblemente galardonado con el Premio Nobel (de Química en 1954; y [de la Paz en 1962](#) por su compromiso en la lucha contra el armamento nuclear) ya había teorizado la estructura secundaria *hélice- $\alpha$*  para las proteínas globulares. En aquella época todavía se creía que las hélices- $\alpha$  se empaquetaban juntas en formas aparentemente irregulares.

Desde entonces, el problema del plegamiento proteico se ha desgajado en tres cuestiones separadas: (1) el código de plegamiento, esto es, la termodinámica de equilibrio de las fuerzas interatómicas para una determinada secuencia de aminoácidos (estructura primaria); (2) la predicción de la tridimensionalidad de la proteína a partir del determinismo inherente en la secuencia de aminoácidos; y (3) la cinética del plegamiento, que determina que la proteína adopte la conformación correcta de entre una infinitud de posibles opciones; y que esta conformación se adopte en una fracción de tiempo casi infinitamente breve, nanosegundos o picosegundos ( $10^{-9}$  y  $10^{-12}$  segundos respectivamente). La investigación se ha centrado en proteínas solubles, dejando de lado, por ahora, las proteínas fibrosas o embebidas completa o parcialmente en membranas celulares.

Cuanto más compleja es una proteína mayor es el número de interacciones moleculares, lo que nos conduce a la denominada «paradoja de Levinthal», teorizada en el año 1969: una cadena poli-peptídica desplegada tiene muchísimos *grados de libertad* (posibilidades de rotación en cada enlace interatómico) y, en consecuencia, un número casi infinito de posibles conformaciones, que Cyrus Levinthal estimó en  $3^{300}$ . Si el plegamiento de una macromolécula siguiera *solo* las leyes del azar, se tardaría un tiempo casi infinito (en sentido matemático) hasta lograr la *conformación correcta*. Sin embargo, una proteína adopta *espontáneamente* su conformación en nanosegundos ( $10^{-9}$  segundos); incluso en picosegundos ( $10^{-12}$  segundos).

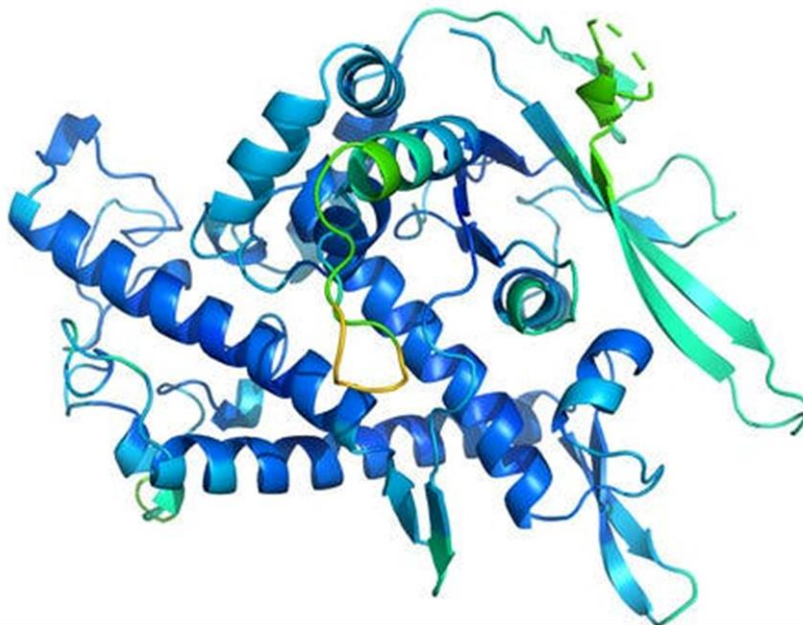
Un hito importante en la ciencia de las proteínas es la hipótesis del grupo de trabajo dirigido por Christian Anfinsen, partiendo de sus experimentos sobre la enzima ribonucleasa, que cataliza el clivaje del ARN en fragmentos de menor tamaño. Según esta teorización la estructura espacial de la proteína es determinada termodinámicamente por la secuencia de aminoácidos (*estructura primaria*) y las condiciones de la solución; y no tanto por la ruta del plegamiento cinético. Así pues, la estructura nativa [de una proteína] no depende de si la proteína se sintetiza en un ribosoma, con la ayuda de chaperonas (proteínas que *ayudan* al plegamiento de otras proteínas) o en forma aislada en un tubo de ensayo. Existen excepciones, no obstante. Tal es el caso de la insulina, la proteasa- $\alpha$ , o las serpinas (apócope de «serina-proteasas»), en las que la forma biológicamente activa queda *atrapada* cinéticamente en su conformación correcta. El modelo teórico de Anfinsen permitió el estudio del plegamiento proteico *in vitro* (en el tubo de ensayo) en lugar del entorno intracelular, mucho más complejo y costoso. Concluyó, además, que el enrollamiento de la cadena proteica es una cuestión termodinámica, no modificándose cuando se producen cambios limitados en la secuencia de aminoácidos como consecuencia de mutaciones. De este modo, el organismo está *protegido* frente a algunas (tal vez muchas) mutaciones. De otro modo la supervivencia de las especies estaría seriamente comprometida.

El plegamiento de las proteínas: ¿determinismo o resultado de múltiples interacciones?

Hasta mediados de la década de 1980, estudiábamos que el *código del plegamiento proteico* era resultado de muchas interacciones *de corto alcance*, tales como *enlaces de hidrógeno*, iónicos, covalentes y enlaces de van der Waals, junto a interacciones con el solvente (generalmente agua).

Conceptualmente, cada nivel de estructura contenía la información necesaria para adoptar la conformación inmediatamente superior. De esta manera, la estructura primaria (secuencia de aminoácidos) determinaba la estructura secundaria; ésta la estructura terciaria; y, en las proteínas poliméricas, la estructura terciaria [de cada monómero] condicionaba la forma como las cadenas poli-peptídicas separadas interaccionaban hasta formar la estructura cuaternaria.

Sin embargo, los modelos estadísticos surgidos a finales de la década de 1980 establecieron que existe un *código de plegamiento no secuencial* y que la forma como se ovilla una proteína se debe tanto a interacciones locales como a otras de más largo alcance.



Las proteínas nativas son solo 5 a 10Kcal/mol más estables que sus estados desnaturalizados (desenrollados). Es termodinámicamente poco probable que la fuerza motriz de tipo energético sea determinante para guiar el plegamiento proteico. Además, la mayoría de las proteínas concentran sus aminoácidos con

carga eléctrica neta en regiones dieléctricas situadas en las superficies expuestas, donde tienen lugar las interacciones moleculares con moléculas agonistas o antagonistas (fisiológicas o farmacológicas). Por otra parte, la estabilidad de las proteínas suele ser independiente del pH de la solución en que se hallan suspendidas, y las mutaciones que se trasladan en cambios de la distribución eléctrica suelen tener mínimos efectos sobre la estructura 3D y la estabilidad.

Las estructuras tridimensionales de las proteínas son una forma de empaquetar gran cantidad de átomos en un espacio reducido.

La biología computacional dirigida a descifrar la tridimensionalidad de una proteína se inició con la *met-enkefalina*, un péptido endógeno opiáceo con solo 5 aminoácidos. Durante la década de 1990 se produjeron notables progresos consiguiendo enhebrar tridimensionalmente estructuras cada vez más complejas generándose bases de datos cada vez más sofisticadas.

En el año 1994, John Moult desarrolló [CASP](#), acrónimo de Critical Assessment of Techniques for Protein Structure Prediction, casi un concurso entre la comunidad científica para predecir estructuras 3D de proteínas. Uno de los *softwares* desarrollados bajo el paraguas del CASP (en sus distintas ediciones) ha sido el programa [AlphaFold](#).

La ubicuidad de las proteínas es impresionante. Son los ingenieros de la fábrica celular, pero también la propia fábrica. No son las únicas moléculas, pero, sin duda, sí las más importantes. No en vano el código genético *solo* codifica proteínas, entre ellas las que regulan su propio funcionamiento.

El programa informático AlphaFold desarrollado por DeepMind posibilita la predicción del plegamiento proteico en pocas horas, un proceso que hace tan solo unas décadas requería muchos años de arduo y meticuloso trabajo.

DeepMind es propiedad de la misma empresa matriz de Google. A partir de la secuencia de aminoácidos de la proteína (su *estructura primaria*) AlphaFold puede predecir de manera rápida y confiable su estructura tridimensional (*secundaria, terciaria* e incluso *cuaternaria*).

Se espera que este avance contribuya a la mejor comprensión de muchas enfermedades y al desarrollo de nuevos medicamentos.

El grado de precisión conseguido por AlphaFold compite favorablemente con los complejos procesos físico-químicos necesarios para llevar a cabo la tarea, a lo largo de años y años de exhaustiva investigación.

Además del progreso en investigación básica, el programa AlphaFold podría permitir el desarrollo de nuevos medicamentos, sobre todo en enfermedades vinculadas al plegamiento anormal de las proteínas, tales como la [enfermedad de alzhéimer](#) y la [fibrosis quística](#) (muscoviscidosis).

Existe, no obstante, un problema legal: qué condiciones establecerá DeepMind para compartir sus hallazgos. En cualquier caso, DeepMind es uno de los actores de un cambio radical que afecta al mundo académico, la industria y la comunidad médica, tal vez el más importante de la última década, la denominada *red neuronal*.

Una *red neuronal* es un sistema matemático modelado libremente. El programa informático (*red neuronal*) aprende habilidades analizando ingentes cantidades de datos. Por ejemplo, identificando patrones de miles de fotografías de gatos, el sistema *aprende* a reconocer un gato. En esta tecnología se fundamentan los traductores que todos llevamos en nuestros teléfonos móviles, los sistemas de reconocimiento de voz o de imágenes. Sorpresivamente, esta tecnología se aplica ahora para predecir la estructura tridimensional de las proteínas.

Partiendo del conocimiento de la tridimensionalidad de las proteínas cabe prever que se podrá anticipar cómo determinadas moléculas (verbigracia, medicamentos) se engarzarán a su estructura, estudiando (informáticamente) sus efectos agonistas o antagonistas.

Se prevé que este sistema pueda llegar a predecir la conformación 3D de una proteína con una resolución atómica.

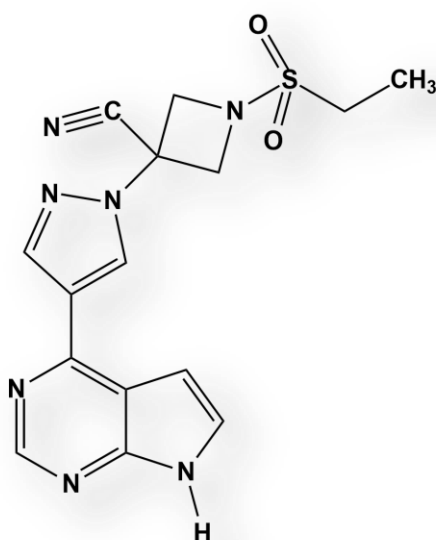
[Andrei Lupas](#), a la sazón director del Departamento de Evolución de Proteínas del Instituto Max Planck de Biología del Desarrollo, en Alemania, es uno de los



que han trabajado con el programa AlphaFold. Durante una década investigó sobre una proteína de un microorganismo parecido a las bacterias denominado arquetas (antiguamente llamadas arqueo-bacterias, filogenéticamente anteriores a las bacterias).

La proteína estudiada de las arquetas se halla imbricada, aunque no completamente, en la membrana con dominios intracelular y extracelular. Esta circunstancia dificultó la determinación de su estructura tridimensional hasta el punto de no conseguirlo tras una década de trabajo. Con el programa AlphaFold la resolución 3D esta proteína se logró en apenas treinta minutos.

Intelectualmente es muy atractiva la idea de aplicar esta tecnología al estudio de nuevos virus, como el causante de la [pandemia covid-19](#).



BARICITINIB

3-(4-(7H-pirrol[2,3-d]pirimidin-4-il)-1H-pirazol-1-il)-1-(etilsulfonil)acetidin-3-carbonitrilo

Durante la actual [pandemia de covid-19](#), una forma más simple de inteligencia artificial ha resultado útil en algunos casos. Un sistema construido por otra empresa londinense, [BenevolentAI](#), ayudó a identificar un medicamento existente, Baricitinib, que podría ser eficaz en pacientes muy graves. Se ha [completado un ensayo clínico](#) (noviembre 2020) realizado en Prati, Toscana,

Italia, pero no se han publicado los resultados (diciembre 2020). Otro estudio del uso de Baricitinib en enfermos de covid-19 no se ha iniciado todavía cuando se redacta este texto.

Baricitinib es un inhibidor de la liberación masiva de citoquinas mediada por las enzimas [tirosina-quinasa](#) JAK (JAK-1 y JAK-2). Hasta ahora se estudiaba el fármaco para prevenir el rechazo tras los injertos de órganos.

AlphaFold se podría aplicar incluso al diseño de nuevas vacunas, contra el SARS-CoV-2 y, ¡quién sabe!, tal vez contra futuros virus con potencial pandémico.

El problema del desarrollo de potenciales medicamentos usando esta técnica (y otras similares) es que la duración de los ensayos clínicos continuará siendo la etapa limitante de todo el proceso.

Esta empresa (DeepMind) se involucró en el pasado en el diseño de juegos de ordenador. Ahora ha reconducido su investigación al ámbito científico.

Zaragoza, 10 de diciembre de 2020

Dr. José Manuel López Tricas  
Farmacéutico especialista Farmacia Hospitalaria  
Farmacia Las Fuentes  
Zaragoza