

Modelado de proteínas

De la secuencia a la función

Andrea Monserrat Negrete Paz
Universidad Michoacana de San Nicolás de Hidalgo.
anegrete.pz@gmail.com

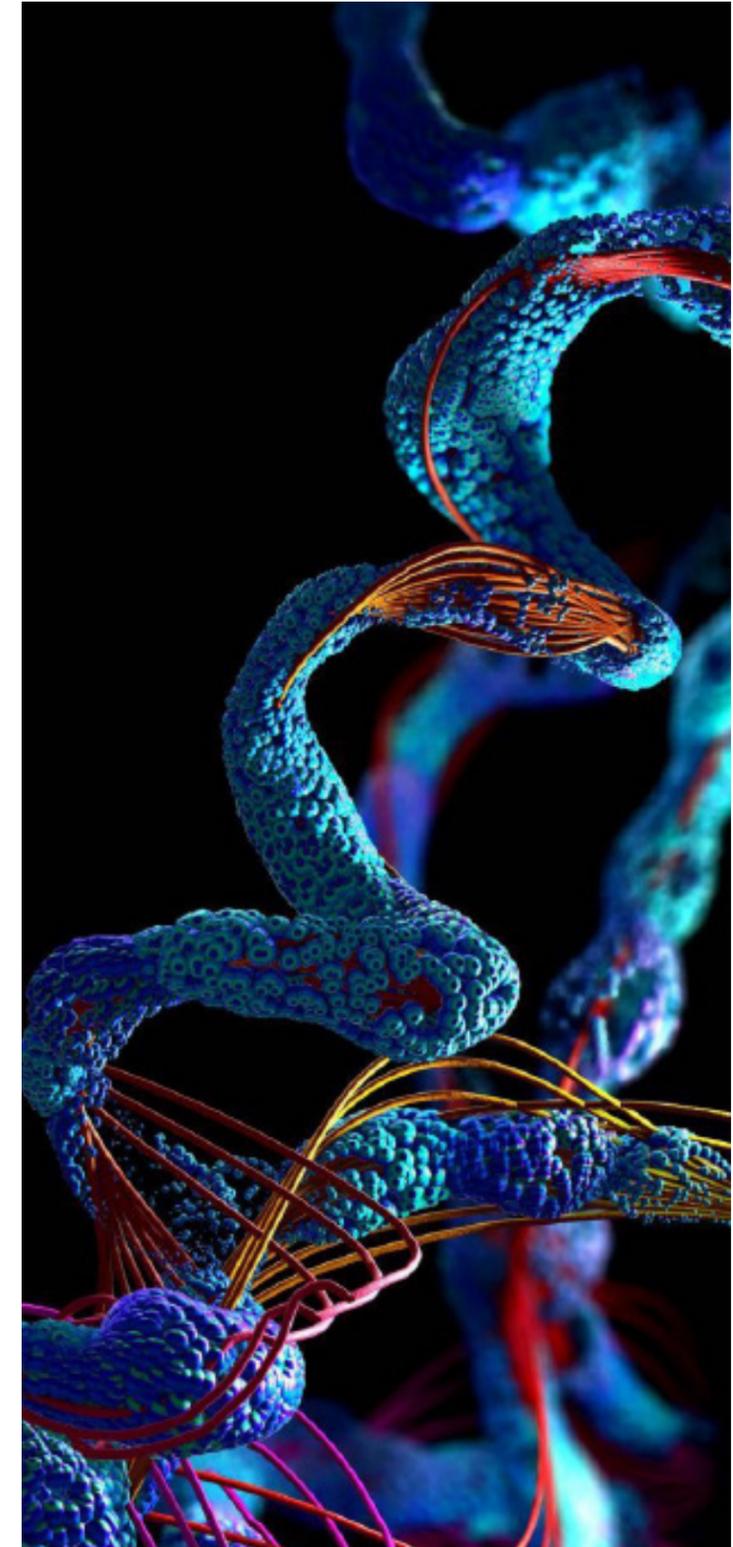
¿Cómo
funciona?

El dogma central de la biología molecular es una teoría que establece que la información genética fluye solo en una dirección, del ADN al ARN, a la proteína [1]. Esto quiere decir que la secuencia de nucleótidos en el ADN será la base para la construcción de una proteína. Las proteínas son macromoléculas que cumplen con un rol biológico específico. Se forman de cientos o miles de unidades más pequeñas denominados aminoácidos, que se encuentran unidos mediante enlaces químicos conocidos como enlaces peptídicos formando grandes cadenas llamadas cadenas polipeptídicas. Estas macromoléculas se pliegan en formas tridimensionales estables, o conformaciones, que están determinadas por su secuencia de aminoácidos y que además se asocian a su función [2].

Pero, ¿Cómo obtenemos la estructura de una proteína a partir de su secuencia? Experimentalmente, el proceso de obtención de la estructura 3D de una proteína es un proceso complicado, debido a esto se han desarrollado técnicas orientadas a la predicción estructural “*in silico*” (por computadora). Estas técnicas se basan en el uso de la secuencia de aminoácidos como punto de partida para obtener estructuras tridimensionales bastante precisas que pueden ser utilizadas en numerosas investigaciones en el ámbito de la biología y la salud.

¿Cómo podemos inferir la estructura de una proteína?

De manera experimental, en la actualidad solo podemos hacerlo mediante dos métodos: la cristalografía de rayos X (CRX) y la espectroscopía de resonancia magnética nuclear (RMN). La finalidad de estas técnicas es crear un modelo estructural a escala atómica usando diferentes tipos de información. En la CRX se obtienen patrones de difracción que se producen debido a la dispersión de fotones de alta energía cuando el



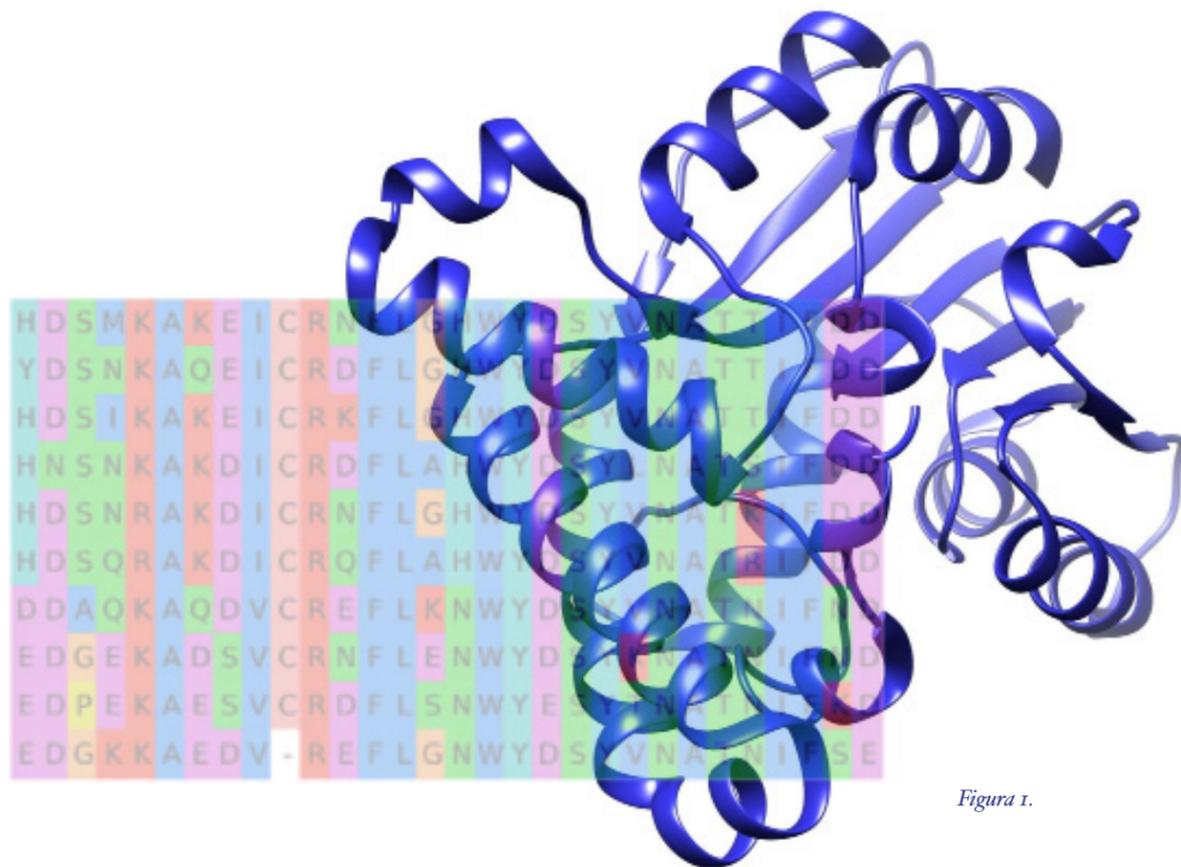


Figura 1.

X incide sobre un cristal de la proteína purificada, lo podemos ver como si hiciéramos una radiografía a la proteína. En la RMN se obtienen datos de restricciones de distancias y ángulos entre núcleos atómicos. Los átomos de la proteína purificada son alterados con ayuda de un campo magnético (como el de los imanes) y ondas de radiofrecuencia como las que usamos en una radio convencional. Como esta perturbación es sonora logramos producir una respuesta igualmente sonora, si se manipula esta respuesta mediante algoritmos matemáticos es posible averiguar la estructura de la proteína. Pero, ¿Qué pasa si no podemos emplear cristalografía de rayos X o RMN debido a que no contamos con una proteína purificada y necesitamos conocer su estructura? Recurrimos a los modelados “*in silico*”. Existen dos tipos de modelado por computadora: modelado por homología (MH) y modelado *de novo* (MN). El MH se basa en el principio de que, si dos proteínas comparten similitud de secuencia lo suficientemente alta, es probable que tengan estructuras 3D muy similares, se trata de una técnica predictiva que permite obtener una aproximación

de la conformación basada en estructuras conocidas de proteínas relacionadas (estructuras molde) que casi siempre son homólogas a nuestra proteína de estudio, es decir, poseen el mismo ancestro común. El MH consta de varios pasos: el primer paso es la selección de un molde mediante la comparación de la secuencia de aminoácidos en bases de datos. Una vez que se identifica el mejor candidato como molde se realiza un alineamiento -un alineamiento es una forma de organizar secuencias de proteínas (o ADN) para identificar regiones de similitud (figura 1)-. Acto seguido, se construye un modelo de la proteína de interés usando la información de las coordenadas atómicas espaciales del molde y finalmente se evalúa la calidad del modelo construido. Ahora bien, ¿qué pasa si no logramos identificar una estructura molde en las bases de datos? Utilizaremos MN. Una de las herramientas más novedosas de esta técnica de predicción es *AlphaFold* [3] que utiliza inteligencia artificial para construir el modelo. De manera similar al MH se parte de la secuencia de aminoácidos de la proteína de interés, se consultan bases de datos y se realiza



un alineamiento, después se identifican proteínas que puedan tener una estructura similar (molde) y se construye una representación inicial de la estructura. Esta es, en esencia, un modelo de qué aminoácidos es probable que estén en contacto entre sí. Después, *AlphaFold* toma el alineamiento, los moldes, y la representación inicial y los pasa a través de un “transformador”, cuya función es identificar qué información es más útil para la construcción del modelo, intercambiando información iterativamente. Esta información se lleva al denominado módulo de estructura que construye el modelo de nuestra proteína para finalmente obtener una lista de coordenadas que representan la posición de cada átomo de la proteína, incluidas las cadenas laterales.

Sin duda alguna, la construcción de modelados computacionales de proteínas de

estructura desconocida es una herramienta muy útil en el ámbito de la Biología Molecular. Entre sus aplicaciones se incluyen desde el análisis del efecto de mutaciones hasta la predicción de epítopes, predicción de ligandos y actividades catalíticas que pueden servir como base para la realización de estudios experimentales complejos. [100](#)

Referencias

- [1] Morange, M. (2009). The Central Dogma of molecular biology. *Resonance*, 14(3), 236-247.
- [2] Whisstock, J. C., & Lesk, A. M. (2003). Prediction of protein function from protein sequence and structure. *Quarterly reviews of biophysics*, 36(3), 307-340.
- [3] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O. & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583-589.