



## IA que descifra el lenguaje de las proteínas

*Xavier Soberón y Osbaldo Resendis hablaron de AlphaFold, el programa de IA que predice la estructura de las proteínas.*

**Elena Sánchez Ascencio**

24 de marzo de 2025

Investigadores del Centro de Ciencias de la Complejidad (C3) de la UNAM emplean una tecnología de inteligencia artificial, conocida como AlphaFold, para predecir la estructura espacial de las proteínas a partir de su secuencia primaria con el fin de estudiar enfermedades complejas como el cáncer, la diabetes, e incluso, la microbiota intestinal.

Las proteínas son moléculas lineales en cadena que se pliegan espontáneamente para formar una estructura tridimensional (3D) y cumplen funciones esenciales en los organismos como regular procesos dentro las células, reparar tejidos y proteger de enfermedades.

El proceso para producir proteínas funciona cuando la célula copia la información genética y la traduce para generar una cadena larga de compuestos llamados aminoácidos. Sin embargo, conocer la estructura de las proteínas para entender los procesos biológicos en los que están involucradas no ha sido un trabajo sencillo para la ciencia.

Por ello, en el evento Lunch Complejo que organiza el C3 de la UNAM, Francisco Xavier Soberón Mainero, coordinador general del C3 e investigador del Instituto de Biotecnología (IBT) de la UNAM, y Osbaldo Resendis Antonio, coordinador académico del C3 e investigador del Instituto Nacional de Medicina Genómica (INMEGEN), hablaron sobre AlphaFold, un programa de inteligencia artificial que predice la estructura 3D de las proteínas a partir de su secuencia de aminoácidos.

AlphaFold es un sistema desarrollado por [DeepMind](#), una subsidiaria de Google, que se basa en el deep learning y en mecanismos de atención para identificar patrones entre la secuencia de aminoácidos y la estructura 3D de proteínas. Mediante este entrenamiento, los desarrolladores crearon un modelo computacional que, con bases de datos de proteínas conocidas, aprende y predice estructuras 3D.

Notablemente, los modelos de AlphaFold son capaces de competir, en tiempo y recursos, con los métodos clásicos de alta precisión como la cartografía por rayos X o espectroscopía de resonancia magnética nuclear (RMN).

“En los ochentas, teníamos un puñado de secuencias de diferentes bichos, hoy en día tenemos un millón de secuencias de diferentes bichos”, dijo Soberón. Esto quiere decir que en la actualidad se tiene una gran cantidad de información genética sobre diferentes seres vivos.



Al trabajar con inteligencia artificial, en conjunto con las bases de datos actuales de secuencias proteicas y genéticas, **AlphaFold es una herramienta crucial para entender enfermedades, diseñar medicamentos más efectivos, estudiar interacciones biológicas entre especies hasta analizar el comportamiento de las proteínas del sistema inmune ante una infección.**

Por ejemplo, Christian Poitras y [colaboradores](#) de la Universidad de Montreal, con un método llamado AlphaFold Pairs, modelaron la interacción entre proteínas de un virus que infecta células humanas y proteínas del sistema inmune humano (anticuerpos).

La interpretación de este tipo de interacciones entre proteínas puede ayudar a implementar alternativas de fármacos para combatir enfermedades y comprender más a fondo los procesos inmunitarios.

Según los investigadores del C3, AlphaFold ha sido un avance disruptivo en el campo de la biofísica y la bioingeniería, ya que además de predecir estructuras proteicas 3D también se pueden hacer comparaciones entre diferentes especies de seres vivos e interpretar relaciones evolutivas para entender las funciones de las proteínas a lo largo del tiempo.

Por ejemplo, en un [estudio](#) de la revista Biomolecules del 2023 investigadores de la Academia



Osbaldo Rensendis habló del avance que significa AlphaFold para el estudio de las proteínas..  
Foto: Elena Sánchez Ascencio/ C3-UNAM.

Rusa de Ciencias utilizaron AlphaFold para analizar dos proteínas virales conservadas en un grupo de virus llamado bacteriofagos (virus que infectan bacterias). Este estudio descubrió una relación evolutiva entre las especies lo que permite establecer parentescos filogenéticos.

## La estructura

Por su parte, Osbaldo Resendis habló de cómo la secuencia de aminoácidos dicta la estructura de las proteínas y a su vez resaltó que entender cómo una secuencia de aminoácidos se pliega en una estructura específica es clave para comprender su función biológica incluso para cuando hay alteraciones en el organismo.

“Ahora ya es posible definir interacciones entre proteínas y otros componentes por ejemplo metabolitos o incluso DNA”, dijo Resendis.

Por ejemplo, cuando las proteínas se pliegan de manera anormal a lo que deberían se pueden desarrollar enfermedades como [Alzheimer](#) o el Parkinson. Cada enfermedad está asociada a una proteína o grupo de proteínas mal plegadas, lo que provoca características patológicas y síntomas específicos.



Xavier Soberón el evento lunch complejo sobre AlphaFold del C3.  
Foto: Elena Sánchez Ascencio/ C3-UNAM.

Además, los investigadores del C3 mencionaron que puede haber excepciones en el plegamiento donde no todas las proteínas se pliegan de manera espontánea. También hay algunas que se pliegan de manera asistida, o sea que unas proteínas ayudan a otras o incluso se quedan en estados intermediarios de plegamiento.

Soberón mencionó que hasta hace cinco años los métodos utilizados para predecir estructuras proteicas trataban de comparar la secuencia de la proteína desconocida con las estructuras de proteínas ya conocidas “con éxito atractivo pero limitado”

Con la amplia gama de posibilidades que existen para la estructura y función de las proteínas, un programa como AlphaFold puede ayudar a que el trabajo sea mucho más eficiente y preciso.



**“AlphaFold, como sistema de inteligencia artificial, combina métodos evolutivos y espaciales de proteínas; es un diálogo entre la parte evolutiva y la parte espacial”, mencionó Resendis.**

Dicho de otra manera, al ingresar la secuencia de aminoácidos de la cual se quiere predecir la estructura, AlphaFold integra dos niveles de información: tanto la base de datos genéticos (evolutivos) como la de estructuras proteicas (espaciales).

En la actualidad AlphaFold ha contribuido con un 65-70% del conocimiento que se tiene sobre la estructura de proteínas humanas con alta confianza. “Hoy tenemos 200 millones de estructuras predichas en las bases de datos, lo que nos ha permitido hacer muchas cosas”, dijo Soberón.

Por ejemplo, Soberón y Resendis han colaborado en estudios de biología de sistemas, en los que se comparan datos metabólicos bajo distintas condiciones. En un [estudio](#) sobre cáncer, los investigadores identificaron patrones metabólicos que mantienen las concentraciones de tejidos cancerosos, permitiendo así su comparación con tejidos sanos y resaltar posibles modificaciones para los fenotipos malignos.

Los académicos del C3 concluyeron que el siguiente paso además de predecir y diseñar proteínas, es considerar el salto a aspectos que van más allá de la predicción de estructuras de proteínas, lo que implica aprender de los éxitos y los fracasos de AlphaFold para aplicarlo a otros problemas en la biología, como el caso de la ingeniería genética con la edición de genes.

### Ligas de interés

- AlphaFold: perspectivas sobre IA, la predicción 3D, y el diseño de proteínas:  
<https://www.youtube.com/watch?v=AnNurVNwfZ0&t=1296s>
- Usando predicciones de AlphaFold para investigación viral:  
<https://www.mdpi.com/1467-3045/45/4/240>
- Mapeo de alta precisión de interacciones físicas directas proteína-proteína entre virus y humanos utilizando el nuevo sistema computacional AlphaFold-pairs:  
<https://www.biorxiv.org/content/10.1101/2023.08.29.555151v1.abstract>