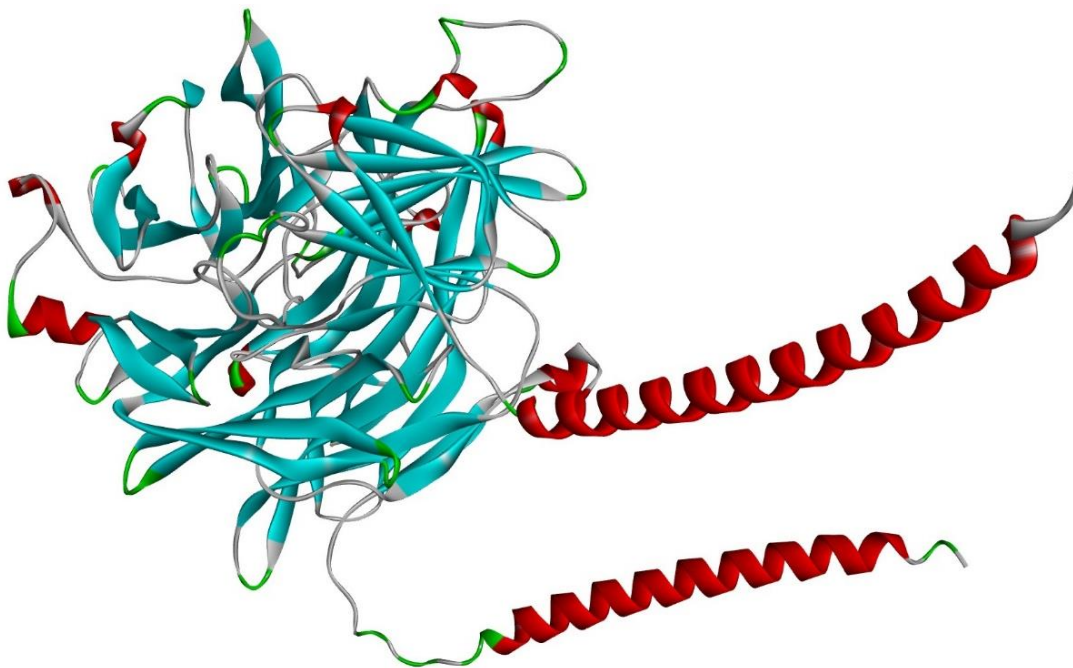




Madrid, miércoles 28 de julio de 2021

La suma de matemáticas y bioinformática desentraña las piezas básicas de los seres humanos

- Las científicas del CSIC Nuria Campillo y Carmen Fernández subrayan que la sinergia de disciplinas ha sido clave para el hito científico logrado por el sistema informático AlphaFold



Homólogo de la proteína inmunomoduladora de células T del parásito *Plasmodium falciparum*, que podría ser la responsable de protegerle del ataque del sistema inmune del hospedador. / EMBL-EBI/ AlphaFold

Hace más de 50 años, científicos de la talla de los premiados con el Nobel, **Kendrew**, **Perutz**, o **Hodgkin**, estaban metidos de lleno en el estudio de las proteínas, el milagro de la vida, mediante cristalografía de rayos X. El plegamiento, no solo de las proteínas, sino de las macromoléculas en general (enzimas y receptores), junto con poder discernir su estructura tridimensional, ha sido uno de los desafíos científicos de las últimas décadas. Diferentes disciplinas, como la ya mencionada cristalografía, la criomicroscopía electrónica, la resonancia magnética nuclear (RMN), la bioinformática y la modelización

in silico, han conseguido increíbles logros en este sentido. Sin embargo, el trabajo de resolver la estructura de estas macromoléculas sigue siendo complicado y costoso.

Y este es un desafío importante, porque las macromoléculas responsables de la vida, como las proteínas, son elementos clave para el buen funcionamiento de nuestro organismo, estando involucradas en prácticamente todos los procesos biológicos que tienen lugar en nuestras células. Poder conocer la estructura de las proteínas posibilita, de una forma más sencilla, determinar su función y, por lo tanto, poder entender y solucionar problemas cuando alguna de ellas no funciona como debería.

La inteligencia artificial ha llegado por la puerta grande a este campo, como en otras muchas áreas del conocimiento, para ayudar a resolver, por ejemplo, la predicción de la estructura 3D de una manera más rápida. Nos va a permitir poder predecir la forma y, como hemos dicho antes, el funcionamiento que tienen las proteínas (un conocimiento indispensable para el desarrollo de tratamientos eficaces para cualquier enfermedad) o las enzimas (llaves maestras en muchos procesos tanto biológicos como industriales), entre otras macromoléculas.

Recientemente se ha conocido el ganador del [concurso científico CASP 14 \(del inglés *Critical Assessment of Protein Structure Prediction*\)](#). A este concurso internacional concurren cientos de grupos de investigación para presentar el desarrollo de sus *softwares* o *aproximaciones* para predecir la estructura de proteínas a partir de sus secuencias de aminoácidos. Podríamos decir que es el escaparate de los desarrollos teóricos más prometedores para la elucidación de la estructura 3D de las proteínas. El ganador, sin ningún tipo de duda, ha sido el algoritmo AlphaFold, desarrollado por la empresa DeepMind (perteneciente a Google).

La mayoría de las predicciones de AlphaFold para la estructura 3D de las proteínas resueltas son muy similares a las que se obtienen a partir de las técnicas experimentales. AlphaFold consiguió una puntuación de casi 90 sobre 100 en un test que mide la semejanza estructural entre la determinada experimentalmente y la predicha por el algoritmo (por encima de 90 se considera como un nivel de predicción total). Nunca antes se había alcanzado una puntuación por encima de 60 en este test.

Los impresionantes resultados de AlphaFold han sido posibles gracias a la sinergia de diferentes disciplinas: matemáticas, bioinformática, biología... El sistema informático parte de la premisa de que la estructura plegada tridimensional de una macromolécula (enzimas, proteínas, receptores) se asemeja a un *grafo espacial*, donde los aminoácidos son los nodos y las aristas del grafo conectan dichos aminoácidos.

Partiendo de esta idea y utilizando el conocimiento derivado del alineamiento de cientos de miles de secuencias de proteínas de estructura 3D conocidas, AlphaFold utiliza la metodología del aprendizaje profundo (*deep learning*) para *aprender* las características físicas, geométricas y evolutivas de esos datos de entrada. Los desarrolladores han hecho un buen trabajo *enseñando* a AlphaFold a extraer información de lo ya conocido para después aplicar dicho conocimiento ante una nueva situación: determinar la estructura de proteínas para las que solo conocemos sus secuencias de aminoácidos, y hacerlo con una precisión y rapidez que no están al alcance de los métodos tradicionales.

Las aplicaciones no se han hecho esperar. [DeepMind y el Laboratorio Europeo de Biología Molecular \(EMBL\)](#), publicaron la semana pasada más de 350.000 nuevas estructuras, incluyendo, no solo unas 20.000 proteínas humanas, sino también proteínas de otros 20 organismos, como las de las bacterias *E. Coli*, la causante de la tuberculosis, entre otras, o las de los ratones que se utilizan para los experimentos en los laboratorios. Y ya hay organizaciones, como la Iniciativa Medicamentos para Enfermedades Olvidadas (DNDi) que están rastreando esta nueva base de datos de proteínas en busca de nuevos tratamientos.

Evidentemente todavía queda mucho camino por andar y numerosos desafíos a los que enfrentarnos, pero AlphaFold ya ha abierto las puertas a abordar un gran reto: el de ahondar en el secreto de la vida.

Nuria Campillo y Carmen Fernández son investigadoras del CSIC en el Centro de Investigaciones Biológicas Margarita Salas (CIB-CSIC)