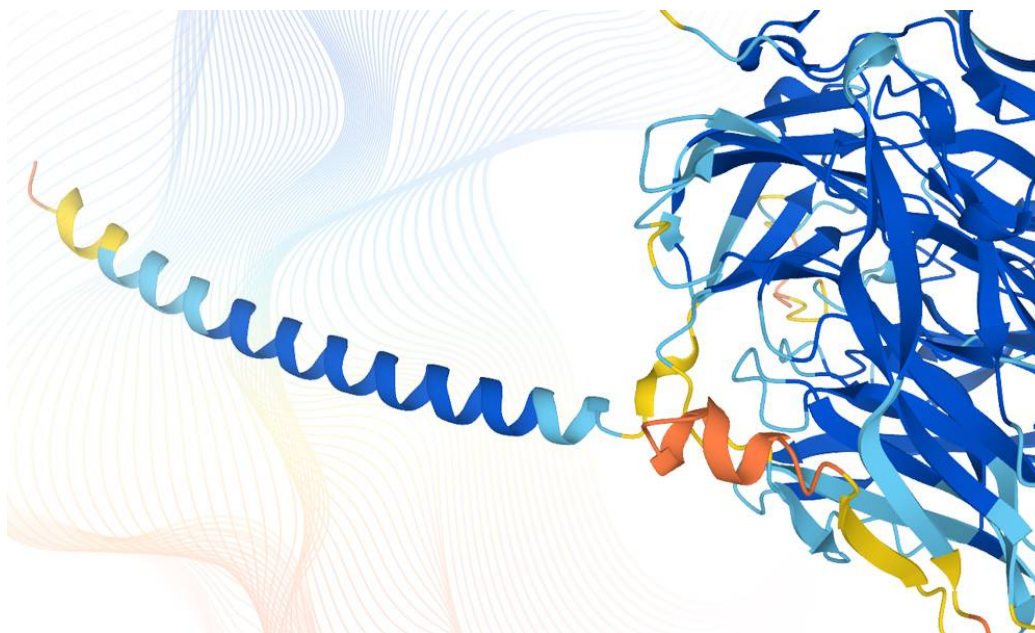


Publicada la imagen más completa de las proteínas que componen el proteoma humano

La Base de Datos de Estructura de Proteínas AlphaFold da a conocer a la comunidad científica más de 350.000 predicciones de estructuras de proteínas, incluido todo el proteoma humano, y promoverá la investigación en muchos campos para comprender el papel de estos componentes básicos de la vida.

SINC

22/7/2021 17:00 CEST



La base de datos más completa y precisa de las predicciones de estructuras del proteoma humano estará disponible de forma libre y abierta para la comunidad científica. / Karen Arnott /EMBL

DeepMind ha anunciado hoy su colaboración con el **Laboratorio Europeo de Biología Molecular** (EMBL), el principal laboratorio europeo en ciencias de la vida, para proporcionar de manera libre y abierta a la comunidad científica la **base de datos** de los modelos de predicciones de las estructuras del proteoma humano (el conjunto completo de proteínas codificadas por el genoma humano) más completa y precisa hasta la fecha.

Esto incluirá alrededor de 20.000 proteínas expresadas por el **genoma**

humano. La base de datos y el sistema de **inteligencia artificial** brindan a los biólogos estructurales nuevas y poderosas herramientas para examinar la estructura tridimensional de las proteínas, y ofrecen un tesoro de datos que podría abrir el camino a futuros avances y presagiar una nueva era para la biología basada en la inteligencia artificial.

El anuncio de hoy proporciona la imagen más completa de las proteínas que componen el proteoma humano, y la publicación de las proteínas de 20 organismos adicionales que son importantes para la investigación biológica

En diciembre de 2020, los organizadores de la evaluación comparativa Critical Assessment of Protein Structure Prediction (CASP) [reconocieron](#) AlphaFold como una solución al gran desafío de más de 50 años de [predecir la estructura de proteínas](#), lo que significó un logro asombroso en el campo.

La base de datos de estructura de proteínas **AlphaFold** (AlphaFold Protein Structure Database) se basa en esta innovación y en los descubrimientos de generaciones de científicos y científicas, desde los pioneros y las pioneras de la cristalografía y el análisis de estructura de las proteínas, hasta los miles de especialistas en predicción y biólogos y biólogas estructurales que han pasado años experimentando con proteínas desde entonces y que han compartido sus resultados de forma abierta.

La base de datos amplía drásticamente el conocimiento acumulado sobre las estructuras de proteínas, más que duplicando el número de estructuras de proteínas humanas con predicciones de alta precisión disponibles para los investigadores. Avanzar en la comprensión de estos componentes básicos de la vida, que sustentan los procesos biológicos en todos los seres vivos, permitirá a los investigadores de una gran variedad de campos acelerar su trabajo.

La semana pasada se publicó en la revista *Nature* la metodología de la última e innovadora versión de AlphaFold, el sofisticado sistema de inteligencia artificial anunciado en diciembre pasado que impulsa estas

predicciones de estructura, y su código fuente abierto. El anuncio de hoy coincide con un segundo artículo de *Nature* que proporciona la **imagen más completa de las proteínas** que componen el proteoma humano, y la publicación de las proteínas de 20 organismos adicionales que son importantes para la investigación biológica.

“Nuestro objetivo en DeepMind siempre ha sido construir **inteligencia artificial** y utilizarla como una herramienta para ayudar a acelerar el ritmo del descubrimiento científico, y mejorar así el conocimiento del mundo que nos rodea”, explica el fundador y director ejecutivo de DeepMind, **Demis Hassabis**.

“Hemos utilizado AlphaFold para generar la imagen más completa y precisa del proteoma humano. Creemos que esta es la contribución más significativa que ha hecho la inteligencia artificial al avance del conocimiento científico hasta la fecha, y es un gran ejemplo de los tipos de beneficios que la inteligencia artificial puede aportar a la sociedad”, continúa.

Una ayuda a los científicos para acelerar sus descubrimientos

La capacidad de predecir computacionalmente la forma de una proteína a partir de su **secuencia de aminoácidos**, en lugar de tener que determinarla experimentalmente con técnicas minuciosas, laboriosas, y a menudo costosas, ya está ayudando a los científicos a lograr en meses lo que antes requería años de trabajo.

AlphaFold ha ayudado a acelerar la investigación de aquellos científicos y científicas que trabajan en la determinación experimental de la estructura de las proteínas

“La base de datos AlphaFold es un ejemplo perfecto del círculo virtuoso de la ciencia abierta”, explica la directora general del EMBL, **Edith Heard**.

“AlphaFold ha sido entrenado utilizando datos de recursos públicos creados por la comunidad científica, por lo que tiene sentido que sus predicciones

sean públicas. Compartir las predicciones de AlphaFold de forma abierta y gratuita permitirá a los investigadores de todo el mundo obtener nuevos conocimientos e impulsar nuevos descubrimientos. Creo que AlphaFold es una verdadera revolución para las ciencias de la vida, así como fue la genómica hace varias décadas y estoy muy orgullosa de que el EMBL haya podido ayudar a DeepMind a permitir el acceso abierto a este recurso extraordinario”, añade.

AlphaFold ya está siendo utilizado por socios como la Iniciativa de Medicamentos para Enfermedades Desatendidas ([DNDi](#), por sus siglas en inglés), que ha [avanzado en su investigación sobre curas que salvan vidas](#) para enfermedades que afectan de manera desproporcionada a las zonas más empobrecidas del mundo, o el Centro de Innovación Enzimática ([CEI](#)) que utiliza AlphaFold para ayudar a diseñar **enzimas** más rápidas para reciclar algunos de los **plásticos** más contaminantes de un solo uso.

AlphaFold ha ayudado a acelerar la investigación de aquellos científicos y científicas que trabajan en la determinación experimental de la estructura de las proteínas. Por ejemplo, un equipo de la Universidad de Colorado en Boulder utiliza las predicciones de AlphaFold para estudiar la resistencia a los antibióticos, mientras que un grupo de la Universidad de California en San Francisco las ha utilizado para [estudiar la biología del SARS-CoV-2](#).

La base de datos de estructura de proteínas AlphaFold

La [base de datos de estructura de proteínas AlphaFold](#) está basada en muchas contribuciones de la comunidad científica internacional, así como en las refinadas **innovaciones algorítmicas** de AlphaFold y en las décadas de experiencia del Instituto Europeo de Bioinformática del EMBL (EMBL-EBI) compartiendo datos biológicos mundiales. DeepMind y el EMBL-EBI están dando libre acceso a las predicciones de AlphaFold para que cualquiera pueda usar el sistema con el fin de permitir y acelerar la investigación y explorar nuevas vías de conocimiento científico.

“ *Hacer que las predicciones de AlphaFold sean accesibles a la comunidad científica internacional*

abre muchas nuevas vías de investigación, desde enfermedades desatendidas hasta nuevas enzimas para la biotecnología y mucho más

Ewan Birney, Director General Adjunto del EMBL



“Este será uno de los conjuntos de datos más importantes desde el mapa del Genoma Humano”, recalca el Director General Adjunto del EMBL y el director del EMBL-EBI, **Ewan Birney**. “Hacer que las predicciones de AlphaFold sean accesibles a la comunidad científica internacional abre muchas nuevas vías de investigación, desde enfermedades desatendidas hasta nuevas enzimas para la **biotecnología** y mucho más. Esta es una nueva y gran herramienta científica, que complementa las tecnologías existentes y nos permitirá ampliar los límites de nuestra comprensión del mundo ”.

Entre las primeras más de 350.000 estructuras publicadas en la base de datos, además del **proteoma humano**, están las proteínas de 20 organismos biológicamente significativos como *E. coli*, la **mosca de la fruta**, el **ratón**, el **pez cebra**, el **parásito de la malaria** y las **bacterias de la tuberculosis**. Se han realizado muchas investigaciones importantes sobre estos organismos, y tener estas estructuras a disposición permitirá a muchos investigadores de campos muy diferentes, desde la neurociencia hasta la medicina, acelerar su trabajo.

La base de datos y el sistema serán actualizados periódicamente a medida que se continúe invirtiendo en mejoras futuras de AlphaFold, y en los próximos meses se planea expandir enormemente la cobertura a casi todas las proteínas secuenciadas conocidas por la ciencia: más de 100 millones de estructuras que incluyen la mayoría de UniProt, la base de datos referencia.

Referencia:

Kathryn Tunyasuvunakool et al. "Highly accurate protein structure prediction for the human proteome" *Nature*

Derechos: **Creative Commons**.

TAGS

PROTEÍNA

GENOMA HUMANO

BASE DE DATOS

Creative Commons 4.0

Puedes copiar, difundir y transformar los contenidos de SINC. [Lee las condiciones de nuestra licencia](#)