CIENCIA

launion.com.mx

@uniondemorelos

Doblando collares con pura intuición

GABRIEL DEL RÍO GUERRA

Gabriel del Río Guerra es Ingeniero Biotecnólogo por el ITSON, Maestro y Doctor en biotecnología por la UNAM, certificado como programador profesional por U.C. Berkeley, realizó 2 estancias posdoctorales en EEUU y dirigió el primer grupo para estudiar el envejecimiento mediante enfoques computacionales en el Instituto Buck. Es investigador titular C del departamento de bioquímica y biología estructural del Instituto de Fisiología Celular de la UNAM y nivel 3 del SNI. Estudia la relación estructura-función de sistemas biológicos usando enfoques de aprendizaje de máquina y biología molecular. Fundador de la empresa DProtein Inc. la cual desarrolla una plataforma para la comercialización de proteínas optimizadas para la nutrición humana, tecnología galardonada con el primer lugar del premio PROFOPI 2022. Ha recibido el reconocimiento como *Experienced Researcher* por la fundación alemana Alexander von Humboldt

Esta publicación fue revisada por el comité editorial de la Academia de Ciencias de Morelos.

Implicaciones de la herramienta de aprendizaje de máquina AlfaFold2 para la ciencia y la sociedad

Identificar la estructura tridimensional (3D) de las proteínas es fundamental para el estudio de los seres vivos y el desarrollo de nuevos medicamentos, entre otras aplicaciones biotecnológicas y para la salud. Para evaluar los avances de la identificación y predicción de la estructura de proteínas, en 1994 se organizó el concurso mundial de predicción de la estructura 3D de las proteínas (CASP por sus siglas en inglés de Evaluación Crítica de la Predicción de la Estructura de las Proteínas) y en 2020 los organizadores de este evento anunciaron que el problema planteado por este concurso había sido resuelto por la empresa inglesa DeepMind [1]. Esto lo declararon después de observar que el programa de aprendizaje profundo desarrollado por esta empresa, AlfaFold2, lograba eficiencias similares a las observadas por métodos experimentales. Incluso las predicciones hechas por AlfaFold2 sirvieron a los experimentalistas para resolver sus datos experimentales.
La relevancia de haber resuelto este problema afecta a múltiples áreas del quehacer humano, de ahí lo importante que es explicar este tema y así prepararnos para los cambios que están presentándose. Pero primero, veamos en que consiste este problema.

El problema del doblado de los collares

Una proteína es como un collar; de forma similar a los collares, las proteínas están compuestas de aminoácidos que serían las cuentas del collar, y cada cuenta o aminoácido está enlazado a través de un conector que tiene cierta flexibilidad como se muestra en la Figura 1.

FIGURA 1. REPRESENTACIÓN de los 20 aminoácidos con los que se construyen las proteínas. Cada cuenta en esta foto representa un aminoácido diferente; al igual que los aminoácidos, las cuentas tienen la capacidad de enlazarse entre sí.

Si consideramos que hay 20 diferentes aminoácidos y que las proteínas pueden construirse combinando: a) los mismos aminoácidos, b) diferentes aminoácidos y/o c) combinaciones diversas de estos, como se muestra en la Figura 2, uno puede imaginar que existen un número gigantesco de posibles proteínas diferentes, una por cada combinación diferentes de aminoácidos. Exactamente hay 20 elevado a la n proteínas diferentes (20^n), en donde n representa el número de aminoácidos que se usaron para construir la proteína. Por ejemplo, el número de proteínas diferentes combinando 100 aminoácidos (por ejemplo, 5 de cada uno de los 20 aminoácidos) es de $20^{100}=1x10^{130}$, es decir, ¡un 1 seguido de 130 ceros! Pero las proteínas pueden estar construidas con diferente número de aminoácidos, por ejemplo, 100, 101, 102,103 y así hasta 15,000 aminoácidos, que corresponde con la proteína más grande conocida (la Titina, proteína de músculo), por lo que el número posible de estas proteínas es una sumatoria de números gigantes, es decir, existe un número infinito de posibles proteínas.

FIGURA 2. LAS proteínas se construyen a partir de la unión de varios aminoácidos. Pueden incluirse en una proteína aminoácidos iguales, diferentes o repeticiones diferentes.

Las células y los seres vivos tienen en las proteínas un infinito número de moléculas para realizar las tareas que éstas requieren. Desde transformar químicamente su entorno, comunicarse entre ellos, reproducirse, entre muchas otras tareas. Para realizar esas tareas, las proteínas deben doblarse (plegarse) para adquirir formas diferentes que les permiten participar en estas tareas celulares. Pero ¿de cuántas formas distintas uno puede doblar un collar? Pareciera que también existe un número gigante de posibles formas: en forma de espiral, en forma alargada, en forma circular, en formas que combinen las anteriores por partes, etcétera. En la Figura 3 se muestran algunos ejemplos de estas formas. Sin embargo, de 300,000 estructuras diferentes de proteínas que conocemos, estas adoptan un número limitado de formas, menos de 3,000 formas diferentes [2]. Esto indica que diferentes proteínas adquieren la misma forma. A estas formas que adquieren las proteínas se le conoce como "plegados de las proteínas". El término plegado se usa para referirse al proceso de plegamiento (doblado de un collar) que condujo a esa forma o estructura 3D.
Conocemos la secuencia de millones de proteínas, cuyos plegados aun no conocemos. Por lo que el objetivo del CASP consiste en identificar qué proteína corresponde con cuál plegado. Ese es el problema que AlfaFold2 resolvió. AlfaFold2 no resolvió el problema del plegamiento, solo resolvió el problema de asignar una proteína a alguno de los menos de 3,000 plegados conocidos.

Décadas de intentos diversos para resolver este problema mostraron que comparar la secuencia de proteínas (la secuencia es similar al orden en el que aparecen las cuentas del collar de izquierda a derecha; ver la figura 2 por ejemplo) solo permite resolver algunos plegados de algunas proteínas. El uso de modelos físicos que modelan las fuerzas de atracción y repulsión entre los aminoácidos en una proteína también solo pudo resolver algunos casos, el de algunas proteínas pequeñas (con pocos aminoácidos), posiblemente porque la complejidad del plegamiento de las proteínas no puede ser abordado por las computadoras actuales o porque la naturaleza del plegamiento de las proteínas no obedece a la lógica de esos modelos físicos. En ese sentido es relevante la conjetura de Johannes Kepler en 1611 sobre cómo lograr el máximo empacamiento de esferas iguales; la conjetura no considera las fuerzas de atracción y repulsión entre las esferas como en el caso de los modelos físicos que actualmente simulan el plegamiento de las proteínas. Los científicos tardaron 400 años en lograr validar que la conjetura de Kepler es válida, pero solo mediante simulaciones en computadora [3]. Existe evidencia que muestra que el empacamiento de los aminoácidos en las proteínas sigue las reglas de la conjetura de Kepler sobre el máximo empacamiento [4], lo cual sugiere que este problema podría resolverse sin considerar las fuerzas de atracción y repulsión entre los aminoácidos.

Importancia científica de predecir el doblado de los collares

¿Por qué es importante conocer la estructura 3D de las proteínas en los estudios científicos? Gracias al conocimiento de la estructura 3D de las proteínas se pueden diseñar moléculas que se unan a ciertas regiones de la proteína para evitar su actividad. La idea básica de estos diseños es encontrar moléculas que puedan "llenar" huecos en la superficie de las proteínas; la actividad de las proteínas comúnmente se localiza en huecos sobre la superficie de las proteínas, de ahí que encontrando una molécula que llene esos huecos se puede evitar que las proteínas realicen su actividad. Estos diseños basados en la estructura 3D de las proteínas han permitido a los científicos identificar de forma expedita fármacos que regulen la actividad de proteínas que son relevantes para controlar padecimientos o mejorar rendimientos en los seres vivos; por ejemplo, el mesilato de Nelfinavir (nombre comercial Viracept), fue el primer fármaco diseñado a partir de la estructura de la proteasa del virus de inmunodeficiencia humana (VIH-1) en 1997 para el tratamiento de esta enfermedad.
Se desconoce la estructura 3D para muchas de las proteínas relevantes para lo que hacemos o sentimos los seres humanos, pero ahora con los avances logrados por AlfaFold2 es posible anticipar que nuevos fármacos puedan ser identificados en un menor tiempo. Al mismo tiempo, más de la mitad de las proteínas conocidas se desconoce su actividad en una célula, y por ende se desconoce su participación en diversas conductas o actividades en los seres vivos. Existe un concurso mundial, que al igual que CASP, busca acelerar el desarrollo de predicciones confiables sobre la actividad de las proteínas (CAFA por sus siglas en inglés Evaluación Crítica de la Asignación de la Función de las Proteínas) que muestra resultados con porcentajes de confianza en sus predicciones de entre 40% y 60% [5]. Estos valores son muy similares a los del CASP antes de AlfaFold2. Por lo tanto, es muy probable que con las estructuras de AlfaFold2 se puedan mejorar las confianzas en la predicción de las actividades de las proteínas.

Importancia social de predecir el doblado de los collares

Más allá del impacto en la ciencia, una de las lecciones más importantes que se derivan de haber resuelto este problema es que con el aprendizaje de máquina se pueden resolver problemas que aún no entendemos. Este puede ser la lección más importante para nuestra sociedad que tiende a tomar decisiones usando modelos que son entendibles, razonables, a pesar de la frecuencia con la que fallan. Considere por ejemplo su habilidad de predecir la duración de lluvias con base en el color de las nubes, su habilidad de anticipar una conducta por parte de una persona que viste de una forma particular. Si sus errores son menos que sus aciertos, el aspecto que ha usado para predecir esos acontecimientos (por ejemplo, color de las nubes) podría ser relevante para anticipar ese fenómeno. De lo contrario, es conveniente que adapte el o los aspectos que está usando para predecir. Hasta hace poco, los seres vivos éramos los únicos organismos capaces de predecir sin necesidad de entender. Considere la habilidad que tienen algunos humanos de andar en una bicicleta; para ello, los humanos aprenden a mantener el equilibrio, al tiempo que pedalean sin entender cabalmente cómo funciona la comunicación del sistema nervioso central y los músculos del cuerpo. AlfaFold2 hoy nos muestra que con el aprendizaje de máquina se puede reproducir esta habilidad e incluso mejorarla, ya que ninguna intuición o razonamiento humano fue capaz de resolver este problema en las últimas décadas. ¿Cómo llegamos a creer que la razón era la mejor forma para realizar predicciones? Esto fue un proceso de siglos que en nuestra cultura occidental, abarcaría desde la época de Tomas de Aquino (siglo XIII) hasta la del inglés Thomas H. Huxley (siglo XIX), en los que la idea de usar la razón antes que la intuición fue consolidándose [6]. En palabras de Huxley para definir el agnosticismo nos dijo que "en cuestiones intelectuales, sigue a tu razón hasta donde te lleve, sin ninguna otra consideración". Hoy AlfaFold2 nos muestra el poder de la predicción "sin razón". Es importante destacar que quienes desarrollaron AlfaFold2 no lo hicieron sin razonamientos algunos; es el problema que resuelve AlfaFold2 para el que no tenemos razonamiento alguno, aún.

Esta habilidad de predecir sin entender nos ha llevado en el pasado a desarrollar fórmulas que nos ayudan a mejorar lo que hacemos. Tome por ejemplo el caso de la primera ley de la termodinámica que establece que "La energía total de un sistema aislado ni se crea ni se destruye, permanece constante" o formalmente se escribe:

$$\Delta U = Q + W$$

en donde Q representa la cantidad de calor desprendido por un sistema, W el trabajo realizado por el sistema y ΔU el cambio en la energía del sistema. Nicolas Leonard Sadi Carnot enunció esta ley cuando el uso de las máquinas de vapor tenía importancia económica e industrial en la Inglaterra del siglo XIX; es decir, ya se usaban ampliamente las máquinas teniendo una intuición sobre su funcionamiento, pero sin tener un razonamiento completo de éstas. Carnot estaba interesado en entender los límites del rendimiento de estas máquinas y sus razonamientos ayudaron a mejorar el diseño de las máquinas hasta el día de hoy [7].
De la misma manera, AlfaFold2 ayudará a desarrollar razones que nos ayuden a entender el plegado y eventualmente el plegamiento de las proteínas.
Finalmente, es importante destacar que para lograr desarrollar AlfaFold2 se generaron un gran número de conocimientos previos, desde los años 1950s, que analizados en el contexto de aquella época no parecían tener ninguna utilidad para la sociedad humana. De hecho, AlfaFold2 no ha resuelto ningún problema de salud de la población inglesa, por ejemplo, que es de donde se derivó este programa de cómputo que sirve para doblar collares (proteínas) con pura intuición. Algunos de los ejemplos enunciados en este texto son testimonio de que el avance de nuestras sociedades se ha dado gracias a poder resolver problemas aparentemente inútiles, como doblar un collar con pura intuición. Es entendible así que aquella sociedad que no invierta en resolver problemas fundamentales, que son los que aparentemente no sirven para resolver problemas actuales prácticos, está destinada a retrasarse en su desarrollo.

FIGURA 3. DIFERENTES proteínas, construidas a partir de diferentes aminoácidos, pueden doblarse/plegarse de diferentes formas. Algunas proteínas hechas con los mismos aminoácidos pueden doblarse/plegarse con la misma forma.

Bibliografía

https://aipocrates.org/2022/09/18/alphafold2-y-el-avance-exponencial-en-la-comprension-de-los-secretos-de-la-biologia/
https://www.quimica.es/enciclopedia/Alineamiento_estructural.html
https://es.wikipedia.org/wiki/Conjetura_de_Kepler
https://www.mdpi.com/1422-0067/22/19/10359
https://biofunctionprediction.org/cafa/
https://es.wikipedia.org/wiki/Agnosticismo
https://concepto.de/leyes-de-la-termodinamica/

Esta columna se prepara y edita semana con semana, en conjunto con investigadores morelenses convencidos del valor del conocimiento científico para el desarrollo social y económico de Morelos. Desde la Academia de Ciencias de Morelos externamos nuestra preocupación por el vacío que genera la extinción de la Secretaría de Innovación, Ciencia y Tecnología dentro del ecosistema de innovación estatal que se debilita sin la participación del Gobierno del Estado.

ACADEMIA DE CIENCIAS DE MORELOS, A.C.

ESTA PUBLICACIÓN FUE REVISADA POR EL COMITÉ EDITORIAL DE LA ACADEMIA DE CIENCIAS DE MORELOS

Para actividades recientes de la academia y artículos anteriores puede consultar: **www.acmor.org.mx**
¿Comentarios y sugerencias?, ¿Preguntas sobre temas científicos? CONTÁCTANOS: editorial@acmor.org.mx