

IA en la predicción de proteínas 3D mediante AlphaFold

AI-based Protein 3D Prediction using AlphaFold

Alejandro Hernández-Soto¹, Alexander Barquero²

Hernández-Soto, A; Barquero, A. Ia en la predicción de proteínas 3D mediante AlphaFold. *Tecnología en Marcha*. Vol. 39 N° especial sobre Inteligencia Artificial. Febrero, 2026. Pág. 203-221.

 <https://doi.org/10.18845/tm.v39i5.8492>



1 Centro de Investigación en Biotecnología (CIB), Instituto Tecnológico de Costa Rica. Costa Rica.

 alhernandez@itcr.ac.cr

 <https://orcid.org/0000-0001-9435-5117>

2 Escuela de Ciencias de la Computación e Informática, Universidad de Costa Rica. Costa Rica.

 alexander.barqueroelizando@ucr.ac.cr

 <https://orcid.org/0000-0001-9397-627X>

Palabras clave

Plegamiento tridimensional de proteínas; estructura proteica; alineamiento múltiple; biotecnología; ciencias de la computación.

Resumen

La predicción tridimensional de las proteínas con una alta aproximación a la conformación real es posible conceptualmente mediante modelos matemáticos según el dogma Anfinsen. Sin embargo, lograrlo es inviable debido a las múltiples conformaciones que añaden complejidad por la paradoja de Levinthal. Una manera de resolver este acertijo es usando modelos de aprendizaje automático a partir de estructuras que ya han sido dilucidadas experimentalmente. El programa AlphaFold permite hacer predicciones *de novo* a partir de algoritmos de aprendizaje automático. En el presente trabajo se explica en qué consiste esta herramienta, las métricas que se utilizan para interpretar los resultados, alternativas y un ejemplo práctico para aprender a ejecutar una predicción y concluye con los principios de ética y uso de la misma.

Keywords

Three-dimensional protein folding; protein structure; multiple alignment; biotechnology; computer science.

Abstract

Three-dimensional protein prediction with a high approximation to the actual conformation is conceptually possible using mathematical models according to the Anfinsen dogma. Still, it is impractical due to the multiple conformations that add complexity due to the Levinthal paradox. One way to solve this puzzle is by using machine learning models based on structures that have already been elucidated using artificial intelligence. The AlphaFold program allows *de novo* predictions by using machine learning algorithms. This paper explains the tool's components and the metrics used to interpret the results. It provides optional ways to access the program, along with a practical example to learn how to execute a prediction. It concludes with the principles of use and ethics of the tool.

Introducción

En 1972 Christian Anfinsen obtuvo el Premio Nobel de Química al proponer que la estructura tridimensional de una proteína en condiciones ambientales fisiológicas puede estar principalmente determinada por su secuencia de aminoácidos bajo condiciones fisiológicas controladas, lo que se conoce como el Dogma Anfinsen [1]. Sin embargo, las múltiples conformaciones que puede tomar una estructura proteica hacen que el tiempo para una correcta predicción sea exponencial y, por tanto, poco posible de calcular desde cero a partir de reglas matemáticas, lo que se conoce como paradoja de Cyrus Levinthal [2]. A pesar de ello, es posible aproximarse al plegamiento de una proteína sin entender todo el proceso, sino analizando el comportamiento y las estructuras existentes, y usándolos como base para una nueva predicción. Es ahí donde los modelos de aprendizaje automático (en inglés: machine learning), que permiten identificar relaciones complejas de datos, facilitan un acercamiento a la estructura 3D de las proteínas. Este es el caso de AlphaFold, una herramienta cuya primera versión tuvo su debut en 2020 en la competencia de predicción de proteínas CASP (*Critical Assessment of Techniques for Protein Structure Prediction*), conocida por ser bianual y existir desde 1994. Específicamente en el CASP14 (sitio web: <https://predictioncenter.org/casp14/>

index.cgi), AlphaFold, en su segunda versión, obtuvo puntajes superiores al 90/100 en un número considerable de objetivos, demostrando un desempeño notable en esa edición. Para efectos teóricos, nótese que el término para esta aproximación en inglés es “*global distance from ground truth* (GDT_TS)”.

AlphaFold es una aplicación de predicción del plegamiento tridimensional de las proteínas con inteligencia artificial mediada por modelos de aprendizaje automático. La estructura de red neuronal de AlphaFold fue descrita en 2021 [3], y fue merecedora del Premio Nobel de Química 2024 atribuido a David Baker, Demis Hassabis y John Jumper (Comunicado del Premio Nobel 2024 sitio web: <https://www.nobelprize.org/prizes/chemistry/2024/press-release/>). El sistema puede predecir estructuras para proteínas sin modelos experimentales, aunque su exactitud depende de la información evolutiva disponible, y aunque no es un modelo basado en homología, sí alinea a secciones de proteínas similares, que tienden a evolucionar en el tiempo con la misma estructura.

La herramienta permite que cualquier usuario pueda pedirle a la plataforma predicciones previas, o hacer una nueva predicción. Esto incluye interacción con proteínas, ligandos y ácidos nucleicos con muy alta precisión [4]. Esta herramienta de inteligencia artificial para biotecnología ha tenido un impacto reconocido en la comunidad científica e industrial. La herramienta ha generado un impacto muy amplio, reflejado en millones de consultas, usos y referencias en diversos entornos científicos e industriales, un alto número de citas bibliográficas y una adopción muy veloz (DeepMind impact stories sitio web: <https://deepmind.google/science/alphafold/impact-stories/>). En investigación estructural de proteínas, la calidad de un modelo se evalúa comúnmente mediante la desviación cuadrática media (RMSD, por sus siglas en inglés Root mean square deviation), que cuantifica la diferencia entre el modelo predicho y la estructura experimental tras la superposición atómica. En muchos contextos estructurales, un RMSD cercano a 0.6 Å se interpreta como una coincidencia de precisión casi atómica. Valores alrededor de 1 Å suelen considerarse indicativos de una predicción precisa en múltiples aplicaciones. En contraste, un RMSD superior a 2 Å generalmente indica discrepancias notables, aunque su interpretación depende del tipo de proteína y de la resolución esperada. AlphaFold puede lograr predicciones cercanas a 1 Å. Es importante recordar que las predicciones de AlphaFold siguen siendo modelos computacionales y pueden contener regiones de baja confianza (*low-confidence regions*), especialmente en segmentos flexibles o pobremente conservados.

¿Cómo funciona AlphaFold?

AlphaFold funciona mediante alineamientos múltiples (MSA, por sus siglas en inglés Multiple Sequence Alignment) a partir de secuencias similares de distintos organismos (evolución) y mediante una representación de pares de aminoácidos (estructura). Con el MSA, se colocan múltiples secuencias de proteínas similares una al lado de la otra, y se genera consultando secuencias de aminoácidos de proteínas en comparación con una secuencia de entrada. El objetivo del alineamiento es asignar un valor a las similitudes y a las diferencias evolutivas. Existen aminoácidos que coevolucionan, esto es, que si cambia o muta uno, cambiará el aminoácido con el que interactúa o está en estrecho contacto. Esta coevolución o covariación preserva la estructura de la proteína. Por el contrario, aquellos fragmentos que evolucionan de manera independiente, posiblemente no tienen contacto directo o interacción con otros [5], [6], [7]. Para construir este mapa de relacionamiento, se hacen alineamientos múltiples con cientos o miles de secuencias lo que permite una predicción “confiable”, lo que es clave para una buena predicción.

Un MSA de alta calidad es esencial para que AlphaFold produzca una predicción precisa de la estructura de las proteínas. Una MSA diversa y profunda, con la mayor cantidad posible de secuencias en el alineamiento ayudará a AlphaFold a identificar señales coevolutivas y utilizarlas para descubrir la estructura 3D de la proteína. Los MSA poco profundos suelen estar entre las principales causas de predicciones inexactas y la razón más común para las predicciones AlphaFold fallidas. Ahora bien, esto se puede optimizar para casos donde la predicción no es la mejor (ver punto 1.3).

En el caso de la representación de cada par (2) de aminoácidos (pairwise representation), se les asigna una representación independiente que le permite al programa tomar la información del MSA y compararla para poder asignar las “relaciones coevolutivas”. Con esto se logra identificar la posición y distancia entre los residuos de aminoácidos.

A partir de toda esta información, el programa puede empezar a interpretar los datos. Para hacerlo, AlphaFold utiliza una red neuronal, que es un sistema de aprendizaje automático inspirado en el cerebro humano, capaz de aprender a reconocer patrones complejos a partir de datos. En este caso, la red neuronal de AlphaFold se llama Evoformer [3]. La innovación clave de esta arquitectura es el uso del mecanismo de atención (*Attention Mechanism*). A diferencia de los métodos computacionales clásicos que analizan secuencias de forma lineal, este mecanismo permite a la red analizar simultáneamente en partes distantes de la proteína. De esta manera, la IA puede ‘entender’ que un aminoácido en la posición 10 interactúa con uno en la posición 300, identificando nuevos pares al procesar el contexto global de la estructura y no solo de sus vecinos inmediatos. Evoformer toma luego la información del MSA y la representación de pares, para refinar e ir ajustando la estructura. Posteriormente, un módulo estructural toma lo hecho y lo compara con la secuencia original para crear una primera propuesta de estructura tridimensional, que termina de modelar al colocar los aminoácidos y ajustando sus posiciones. El proceso se repite de manera iterativa o se “recicla” 3 veces para construir una predicción [3] (Fig 1).

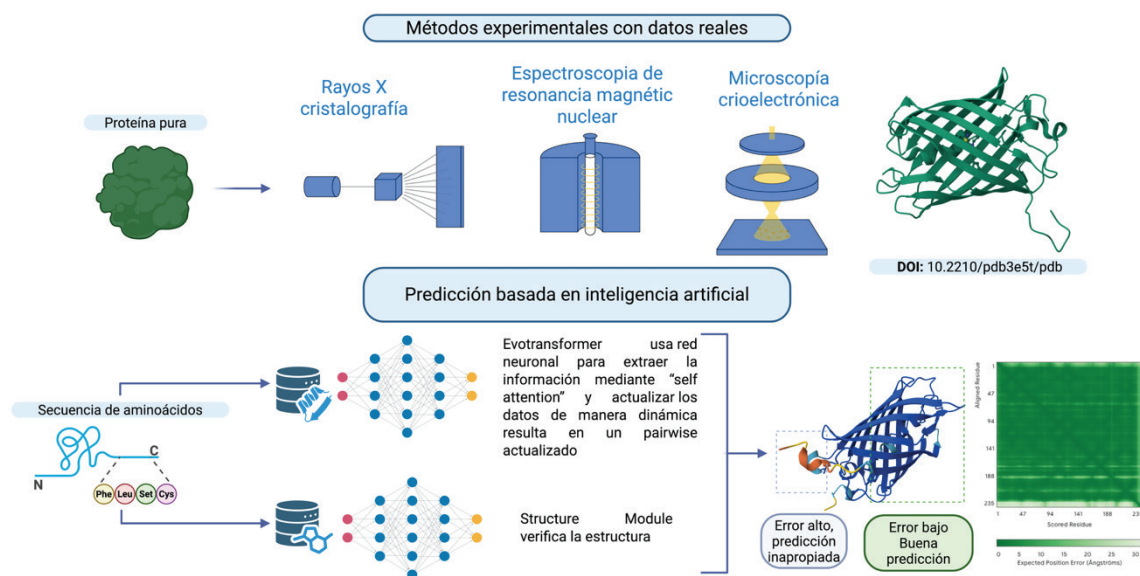


Figura 1. Ilustración comparativa de los métodos tradicionales para obtener la estructura tridimensional de las proteínas y lo que ocurre con inteligencia artificial, donde se reconstruye la estructura tridimensional mediante comparaciones con bases de datos de estructuras publicadas. Nótese que en el ejemplo se muestra la proteína rojo fluorescente cuyo identificador en la Protein Data Bank es 3e5t Doi:10.2210/pdb3e5t/pdb, Uniprot Q8ISF8, mientras que en la predicción es un resultado obtenido en el servidor de

AlphaFold disponible con el mismo identificador en el sitio web: <https://alphafold.ebi.ac.uk/entry/Q8ISF8>

La versión más reciente de AlphaFold, incorpora además modelos de difusión (*Diffusion Models*). Esta aplicación de la IA generativa no construye la proteína un bloque a la vez, sino que comienza con ruido aleatorio y refina progresivamente la posición de los átomos hasta eliminar el ruido y revelar la estructura precisa. Esto permite modelar con mayor exactitud no solo proteínas, sino sus interacciones [4].

Métricas de AlphaFold

pLDDT = Va de 0 - 100.

El “predicted Local Distance Difference Test” es una métrica que determina la confianza de la estructura local, por encima del 80% es bueno, por encima de 90 es la mayor confianza. Aún por encima de 70 corresponde a una predicción correcta de la proteína, pero con posibles diferencias en la ubicación de los residuos (Fig. 2).

El pLDDT determina la concordancia de la predicción versus una estructura experimental y se basa en la prueba de diferencia de distancia local C α (IDDT-C α), que es una puntuación evalúa la exactitud de las distancias locales [8]. Este valor puede variar a lo largo de una proteína por lo que existirán partes de la predicción con alta confianza y otras que no [9].

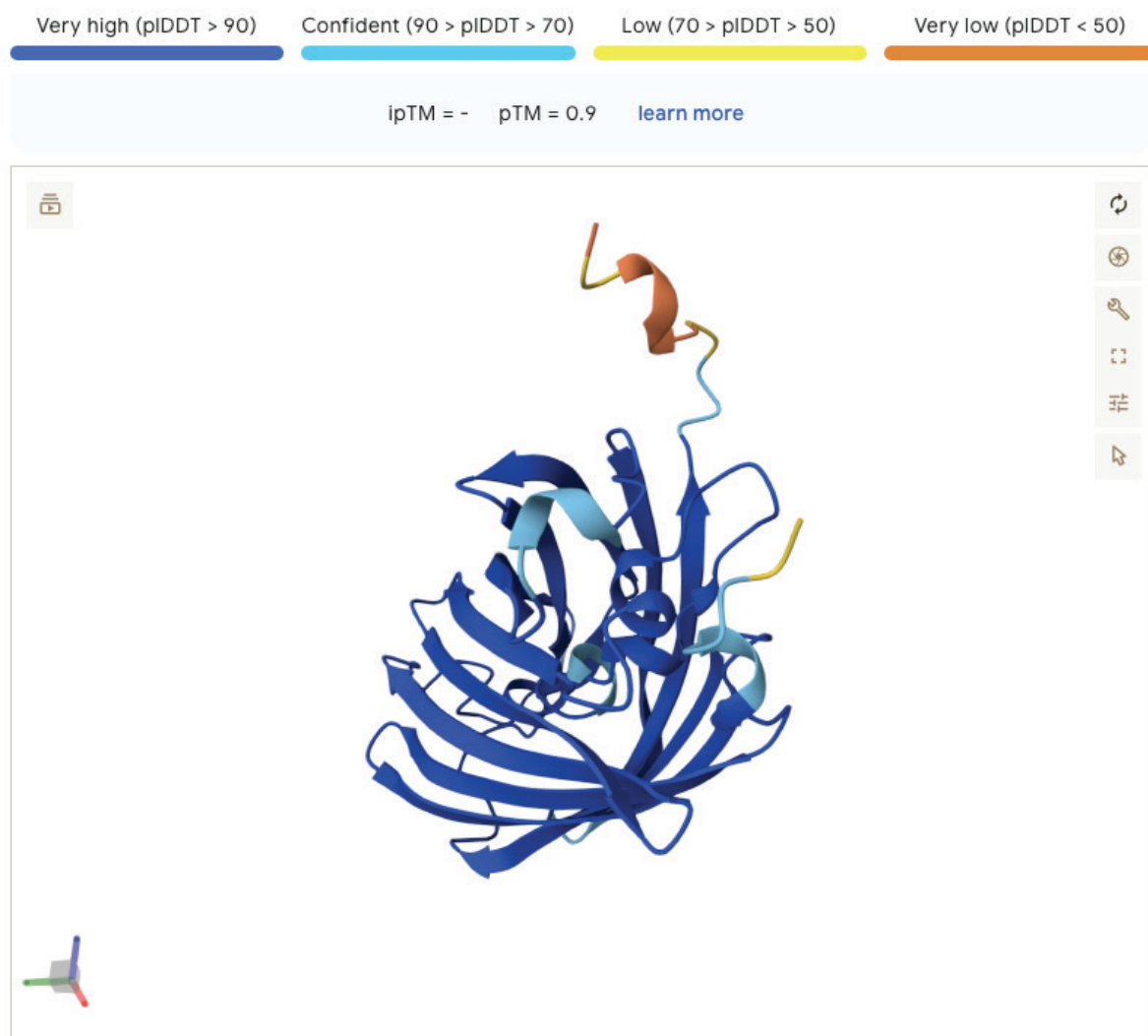
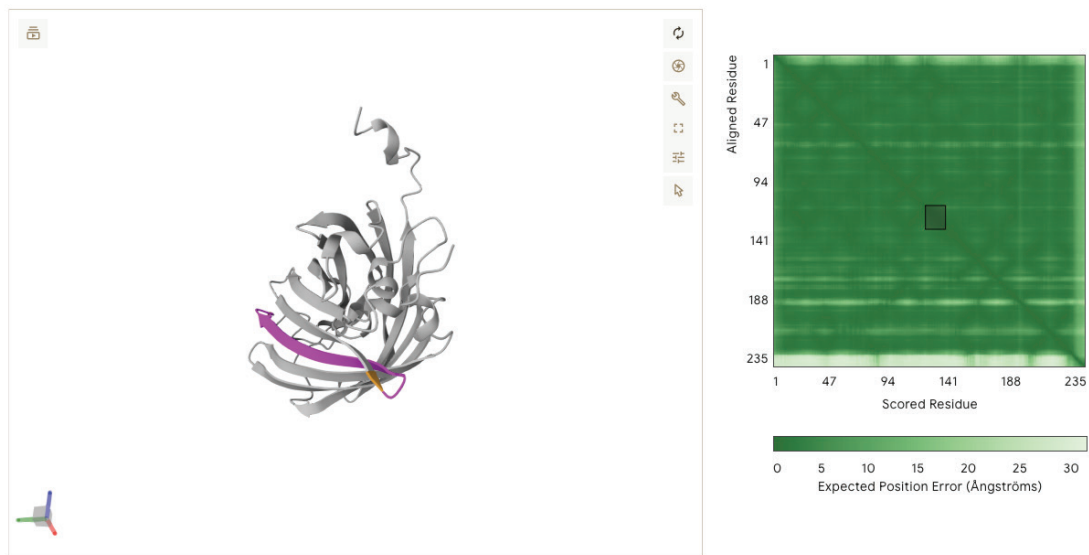


Figura 2. Captura de pantalla de una predicción en el servidor de AlphaFold obtenido a partir de la secuencia de aminoácidos de la proteína rojo fluorescente Uniprot Q8ISF8. La predicción puede ser accedida en el repositorio de AlphaFold con el mismo identificador en el sitio web: <https://alphafold.ebi.ac.uk/entry/Q8ISF8>. Nótese los valores de pLDDT en color azul cuando son mayores a 90, y naranja cuando son menores a 50.

PAE

Es el “Predicted Aligned Error” entre la estructura y la predicción. Esto es, una medida de la confianza en la posición relativa de un par de aminoácidos en la estructura, representada gráficamente para todos los pares en el gráfico PAE (Fig. 3).



Non-commercial use only, subject to [AlphaFold Server Output Terms of Use](#); no use in docking or screening tools.

Information

Type	Copies	Sequence
Protein	1	10 MVSKGEELIK 70 ILATSFMYGS 130 PSN 190 TYRSKKT VGN 20 ENMRMKVVME 80 RTFIKYPADI 140 PSN 200 IKMPGVHAYD 30 GSVNGHQFKC 90 PDDFKQSFPE 150 TKGWEPNTEM 210 HRLERIEESD 40 TGEGEGRPYE 100 GFTWERVTRY 160 MYPADGGLRG 220 NETYVVQREV 50 GVQTMRIKVI 110 EDGGVVTVTQ 170 YTDIALKVDG 230 AVAKYSNLGG 60 EGGPLPFAFD 120 DTSL 180 GGHLHCNFVT 240 GMDELYK

Figura 3. Captura de pantalla de una predicción en el servidor de AlphaFold obtenida a partir de la secuencia de aminoácidos de la proteína rojo fluorescente Uniprot Q8ISF8. La predicción puede ser accedida en el repositorio de AlphaFold con el mismo identificador en el sitio web: <https://alphafold.ebi.ac.uk/entry/Q8ISF8>. Nótese el recuadro en el gráfico de color verde que corresponde al “Expected Position Error”, indicando la ubicación de la predicción con un tono oscuro cuando el error es bajo. Al marcar el recuadro se despliega en la imagen tridimensional la ubicación de la estructura y en el recuadro de la secuencia de aminoácidos, los aminoácidos que corresponden.

El PAE se puede visualizar como un gráfico de puntos. La gráfica 2D contiene los residuos de proteínas a lo largo de los ejes. Cada uno de los recuadros tiene una tonalidad que corresponde a la predicción. La intensidad de coloración verde es un indicativo del error de distancia esperado en Ångströms (Å) para un par de residuos. Un mosaico de color verde oscuro es un indicador positivo de una predicción con un error bajo, mientras que un recuadro con un verde claro es un indicador de un alto error y, por tanto, una predicción deficiente (Fig. 3).

El gráfico PAE habitualmente presenta una línea diagonal de color verde oscuro que va desde la parte superior izquierda hasta la parte inferior derecha. Esto representa residuos alineados contra sí mismos, donde la confianza siempre es alta por definición, por lo que no es informativo y puede ignorarse. La información biológicamente relevante suele encontrarse principalmente en las regiones alejadas de la diagonal (Fig. 3).

Es importante notar que se puede seleccionar una región del gráfico. La selección resaltará la región o regiones en cuestión en la estructura 3D, para que se pueda examinar la relación entre la secuencia, el segmento correspondiente del gráfico PAE y la estructura.

AlphaFold-Multimer, se usa para interacciones

AlphaFold-Multimer es una versión especializada del software diseñada para predecir estructuras de complejos proteicos. AlphaFold-Multimer incorpora principalmente dos métricas adicionales: pTM e ipTM. La puntuación del modelado de plantilla predicha (pTM) y la puntuación del modelado de interfaz predicho (ipTM). Ambas se basan en la puntuación de modelado de plantilla (TM), que mide la precisión de la estructura global de la proteína y es menos sensible a imprecisiones localizadas.

El pTM es una medida integrada de qué tan bien AlphaFold-Multimer ha predicho la estructura general del complejo. Es la puntuación de TM prevista para una superposición entre la estructura prevista y la estructura verdadera hipotética. Una puntuación de TM superior a 0,5 significa que el pliegue general previsto para el complejo podría ser similar a la estructura real. Una puntuación de TM inferior a 0,5 significa que la estructura prevista probablemente sea incorrecta: la puntuación de pTM sigue la misma definición. La puntuación pTM debe interpretarse con cautela. Por ejemplo, una situación en la que una de las proteínas que interactúan es más grande y su estructura se predice correctamente, mientras que una proteína asociada más pequeña se predice incorrectamente. La puntuación pTM resultante del complejo podría estar dominada por la proteína más grande y mostrar una puntuación pTM superior a 0,5.

Por el contrario, el ipTM mide la precisión de las posiciones relativas previstas de las subunidades que forman el complejo proteína-proteína. Los valores superiores a 0,8 suelen asociarse con predicciones confiables, mientras que los valores inferiores a 0,6 sugieren probablemente una predicción fallida. Los valores de ipTM entre 0,6 y 0,8 son una zona gris donde las predicciones pueden ser correctas o incorrectas. Estos valores suponen un modelado con múltiples pasos de reciclaje, por lo que el proceso de predicción alcanza un grado de convergencia. Las regiones desordenadas y las regiones con una puntuación pLDDT baja pueden afectar negativamente a la puntuación ipTM incluso si la estructura del complejo se predice correctamente.

El ipTM puede resultar más útil para los usuarios que pTM. Esto se debe a que la calidad de la predicción de las posiciones relativas de las subunidades y la calidad de toda la predicción compleja son altamente interdependientes: si las posiciones relativas de las subunidades son correctas (como se refleja en una puntuación ipTM alta), los usuarios pueden esperar que todo el complejo también sea correcto.

En la práctica, su confianza general en las predicciones de multímeros debe basarse en una combinación de todas las métricas, incluidas el pTM e ipTM, así como el pLDDT y PAE.

Alternativas de uso y optimización de AlphaFold

Alineamientos múltiples con posibilidad de optimización

La calidad de la predicción depende de la calidad del alineamiento de secuencias múltiples (MSA), que se genera en función de los datos de la secuencia de proteínas. Los MSA creados con el mecanismo predeterminado suelen funcionar bastante bien, pero puede haber casos difíciles. Se puede proporcionar el propio MSA o utilizar implementaciones accesibles como ColaFold [10].

Código fuente de AlphaFold

<https://github.com/google-deepmind/alphafold>

Alternativas

Existe un sitio colaborativo para estructuras 3D-Beacons Network, sitio web: <https://www.ebi.ac.uk/pdbe/pdbe-kb/3dbeacons/>

Rosetta Fold del Laboratorio de Baker [11], es una alternativa a AlphaFold del laboratorio de Jumper [4]. RoseTTAFold2 puede usarse en una versión web en: <https://neurosnap.ai/service/RoseTTAFold2>, y permite mejorar la predicción para moléculas más complejas para lo cual se ajusta el número de “recycles” (Fig. 4).

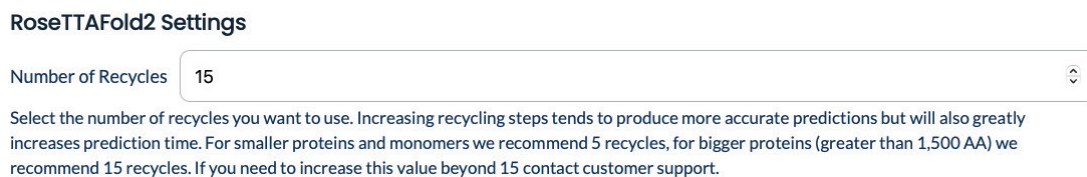


Figura 4. Captura de pantalla de una alternativa a AlphaFold llamada RosettaFold. Nótese la ventana de número de ciclos de reciclaje, este número es de 3 por defecto, por lo que un aumento puede permitir una mejor predicción.

Cómo usar la herramienta

La herramienta se puede usar de distintas maneras; a continuación se describen dos de las opciones más comunes para usuarios web.

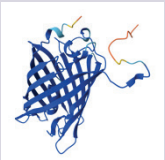
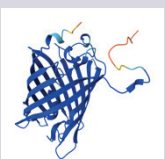
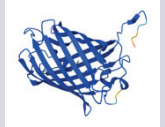
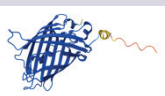
La primera es su versión de búsqueda como repositorio de predicciones “AlphaFold Protein Structure Database” hechas con anterioridad. Esta versión se encuentra en el enlace, sitio web: <https://alphafold.ebi.ac.uk>. Este repositorio es resultado de la alianza entre el Instituto de Bioinformática Europea (EBI) y Google DeepMind. Las predicciones son además reportadas en el sitio de referencia de proteínas UniProt, sitio web: <https://www.uniprot.org/>, y también en Foldseek, buscador web: <https://search.foldseek.com/search> [12]. Adicionalmente, la base de datos está disponible para su uso [13].

La segunda es su versión de servidor “AlphaFold Server”. Su más reciente versión, AlphaFold, permite someter a predicción las proteínas con o sin interacción con otras moléculas. Para ello se debe ingresar al servidor en el sitio web: <http://alphafoldserver.com/> con alguna cuenta de Google, o iniciar desde un usuario nuevo en el sitio web: <https://deepmind.google/technologies/alphafold/alphafold-server/>.

Ejemplo práctico

En el apartado anterior, se mencionó que es posible usar el repositorio. La manera más veloz de disponer de la predicción 3D es acceder a “AlphaFold Protein Structure Database” que la herramienta ya ha hecho con anterioridad en el sitio web: <https://alphafold.ebi.ac.uk>. A continuación, un ejemplo de estructuras que están disponibles en el repositorio (cuadro 1).

Cuadro 1. Ejemplos de predicciones hechas en AlphaFold y disponibles en el repositorio abierto de la herramienta.

Secuencia	AlphaFold	3D
>ATE88097.1 mRuby3 MVSKGEELIKENMRMKVMEGS VNGHQFKCTGEGEGRPYEGVQT MRIKVIEGGPLPFAFDILATSFMYG SRTFIKYPADIPDFKQSFPEG FTWERVTRYEDGGVTVTQDT SLEDGELYNNVKKVRGVNFPNSG PVMQKKTKGWEPNTEMMYPADG GLRGYTDIALKVDGGGHLHCNFVT TYRSKKTGVGNKMPGVHAVDHRL ERIEESDNETYVVQREVAVAKYSN LGGGMDELYK	Identity: 79% Positives: 207/237 (87%) Gaps: 7/237 (3%) Average pLDDT= 94.31 (Very High) Proteína Uniprot A0A182DW87 https://alphafold.ebi.ac.uk/entry/A0A182DW87	
>OL631606.1 Electra MVSKGEELIENMRMKVMEGS VNGHQFKCTGEGEGRPYEGVQT MRIKVIEGGPLPFAFDILATSLFG SKTFIKYPADIPDFKQSFPEGFT WERVTRYEDGGVTVTQDTSLEDG GLVYNNVKKVRGVNFPNSGPVMQKK TEGWEPFTEMMYPANGGLRGYTDI ALKVDG DGHLANIVTTYRSKKTVG DIKMPGVHAVDYRLERVEESDNETY VVLREVAVAKYSNLGGGMDELYK	Identity: 75% Positives: 202/237 (85%) Gaps: 7/237 (3%) Average pLDDT= 94.31 (Very High) Proteína Uniprot A0A182DW87 https://alphafold.ebi.ac.uk/entry/A0A182DW87	
>AYV61462.1 mKelly2 MELIKENMRMKLYMEGTGVGNHH FKCTAEGEGKPYEGTQTQRIKV VEGGPLPFAFDILATCFMYGSKTFIN YPQGIPTDFKQSFPEGFTWERVT TYEDGGVLTVTQDTSLQDGYLIYN VKLRGVNFPNSGPVMQKKTLG WEPTTEYLVPVDGGLEGTCDMAL RLVGGGHLKCNLKTYSKKPA KNLKMPGEHQVDRKLERIKEADNE TYEQHEVAVAR	Identity: 87% Positives: 197/218 (90%) Gaps: 0/218 (0%) Average pLDDT 96.62 (Very High) Proteína Uniprot=D0VX33 https://alphafold.ebi.ac.uk/entry/D0VX33	
>ATE88096.1 mClover3 MVSKGEELFTGVVPILVELDGDVNG HKFSVRGEGEGDATNGKLTCLKICTT GKLVPWPWTLVTTFGYGVACFSRYP DHMKQHDFFKSAMPEGYVQERT ISFKDDGTYKTRAEVKFEGDTLVN RIELKGIDFKEDGNILGHKLEYNFN SHYVYITADKQKNCIKANFKIRHN VEDGSVQLADHYQQNTPIGDGPVLL PDNHYLSHQSKLSKDPNEKRDHNV LLEFVTAAGITHGMDELYK	Identity: 96% Positives: 229/236 (97%) Gaps: 0/236 (0%) Average pLDDT= 93.69 (Very High) Proteína Uniprot A0A059PIR9 https://alphafold.ebi.ac.uk/entry/A0A059PIR9	
>QUA55773.1 mGold MVSKGEELFTGVVPILVELDGDVNG HKFSVSGEGEGDATYGLTLKFICTT GKLVPWPWTLVTSGLYGLQCFARYP DHMKQHDFFKSAMPEGYVQER TIFFKDDGNYKTRAEVKFEGDTLVN RIELKGIDFKEDGNILGHKLEYNN SHNVYITADKQKNGIKANFKIRH NIEDGGVQLADHYQQNTPIGDGPV LLPDNHYLSYQSKLSKDPNEKRDH MVLLEFVTAAGITLGMDELYK	Identity: 97% Positives: 234/239 (98%) Gaps: 0/239 (0%) Average pLDDT: 96.25 (Very High) Proteína Uniprot A0A7S7YQ80 https://alphafold.ebi.ac.uk/entry/A0A7S7YQ80	

La segunda opción es realizar la predicción a partir de la secuencia de aminoácidos de una proteína. Para ello, es necesario ingresar al servidor de AlphaFold <https://deepmind.google/technologies/alphafold/alphafold-server/> es posible ingresar proteínas fluorescentes de *Entacmaea quadricolor* tales como la ATE88097.1 mRuby3 de color rojizo, OL631606.1 Electra de color azul, AYV61462.1 mKelly2 de color rojizo, así como proteínas no relacionadas cuya estructura 3D es similar pero en su secuencia de aminoácidos es distinta por ejemplo *Aequorea victoria* como ATE88096.1 mClover3 de color verde y QUA55773.1 mGold de color amarillo (cuadro 2).

Cuadro 2. Secuencia de aminoácidos correspondientes a proteínas fluorescentes provenientes de *Entacmaea quadricolor* y *Aequorea victoria*.

Proteína	Secuencia	Referencia
Entacmaea quadricolor		
Ruby3 (roja)	>ATE88097.1 mRuby3 MVSKEELIKENMRMKVMEGSVNGHQFKCTGEGEGRPY EGVQTMRIKVIIEGGPLPFAFDILATSFMYGSRFTIKYPADIP DFFKQSFPEGFTWERVTRYEDGGVVTVTQDTSLEDGELVYN VKVRGVNFPSPNGPVMQKKTKGWEPTNEMYPADGGRLRGYT DIALKVDGGGHLHCNFTTYRSKKTGNIKMPGVHAVDHRL ERIEESDNETYVQREVAVAKYSNLGGGMDELYK	[14]
Electra 1 (Azul) derivó de Ruby)	>OL631606.1 Electra MVSKEELIEENMRMKVMEGSVNGHQFKCTGEGEGR PYEGVQTMRIKVIIEGGPLPFAFDILATSLFGSKFTIKYPADIP DFFKQSFPEGFTWERVTRYEDGGVVTVTQDTSLEDGGVLYN VKVRGVNFPSPNGPVMQKKTEGWEPTNEMYPANGGLRGYT DIALKVDG DGHLANIVTTYRSKKTGDIKMPGVHAVDYRL ERVEESDNETYVVLREAVAKYSNLGGGMDELYK	[15]
mKelly2 <i>moderate acid sensitivity.</i>	>AYV61462.1 mKelly2 MELIKENMRMKLYMEGTGNGHHFKCTAEGEGK PYEGTQTQRIKVVIEGGPLPFAFDILATCFMYGSKTFINYPQGIP DFFKQSFPEGFTWERVTTYEDGGVLTVTQDTSLEDGGLIYN VKLRGVNFPSPNGPVMQKKTLGWEPTTEYLVPVDGGLEGTC MALRLVGGGHLKCNLKTYSKKPAKNLKMPEHQVDR KLERIKEADNETYVEQHEVAVAR	[16]
Aequorea victoria		
mClover (Verde) <i>high acid sensitivity.</i>	>ATE88096.1 mClover3 MVSKEELFTGVVPIVLVDGVDVNGHKFSVRGEGEGDA TNGKLTLLKFICTTGKLPVPWPTLVTTFGYGVACFSRYPDH MKQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEG DTLVNRIELKGIDFKEDGNILGHKLEYNFSHNVYITADKQKN CIKANFKIRHNVEDGSQLADHYQQNTPIGDGPVLLPDN HYLSHQSKLSKDPNEKRDMVLLEFVTAAGITHGMDELYK	[14]
mGold	>QUA55773.1 mGold MVSKEELFTGVVPIVLVDGVDVNGHKFSVSGEGEGDATYG KLTLKFICTTGKLPVPWPTLVTSGLGYGLQCFARYPDHMKQH DFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRI ELKGIDFKEDGNILGHKLEYNFSHNVYITADKQKNIGKAN FKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSYQSKL SKDPNEKRDMVLLEFVTAAGITLGMDELYK	[17]

A continuación, en el cuadro 3 y en la figura 5, se presentan los pasos a seguir para lograr realizar una predicción tridimensional de una proteína en el servidor de AlphaFold.

Cuadro 3. Pasos para hacer una predicción en AlphaFold.

Paso	Observaciones
0	Copiar alguna de las secuencias de aminoácidos de interés, por ejemplo las que hemos dispuesto en el cuadro 2.
1	Para proceder con la predicción en el Servidor de AlphaFold, se ingresa al servicio de AlphaFold https://deepmind.google/technologies/alphafold/alphafold-server/ y se selecciona “ Add entry”.
2	Añada la secuencia de aminoácidos y haga clic en continuar y hacer un preview.
3	Añada el nombre del proyecto y haga clic en confirmar el trabajo.
4	El resultado se presentará en el menú principal en el ítem de tres puntos, al final de cada resultado para mostrar los datos.

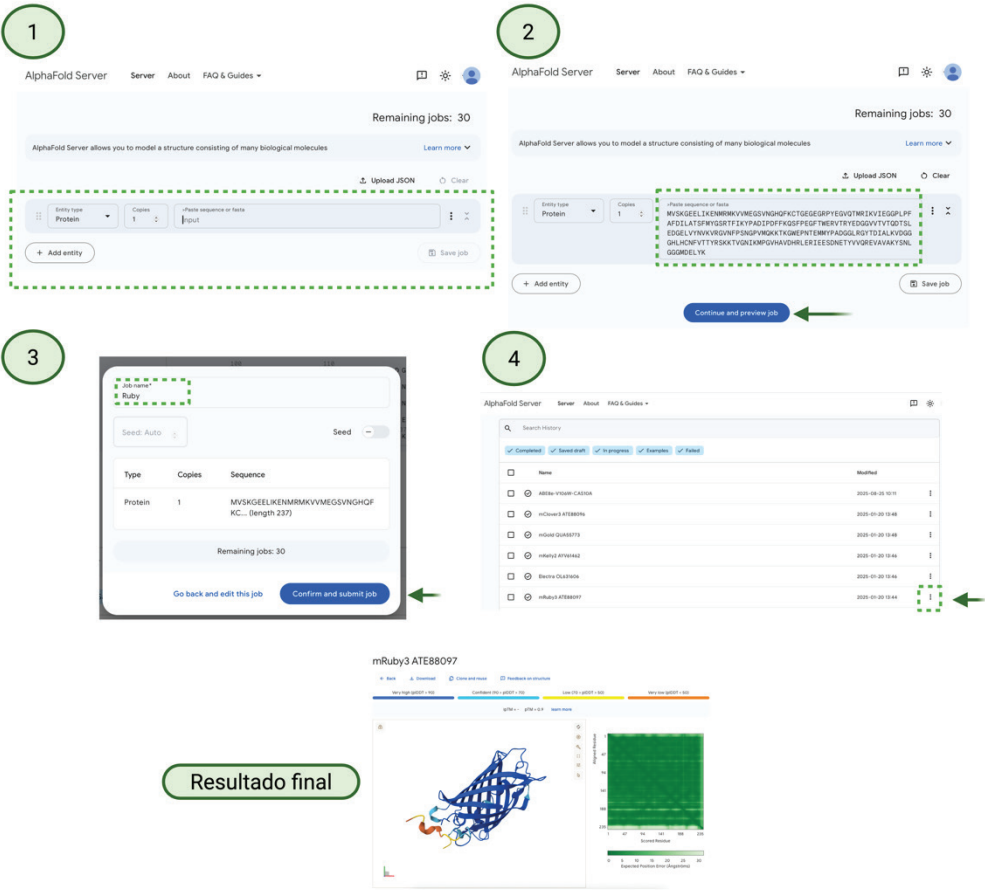
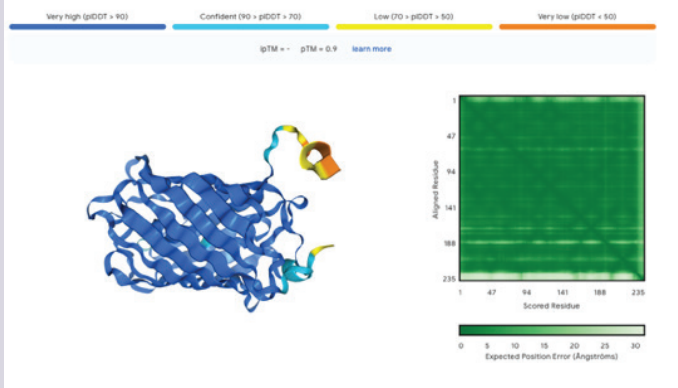
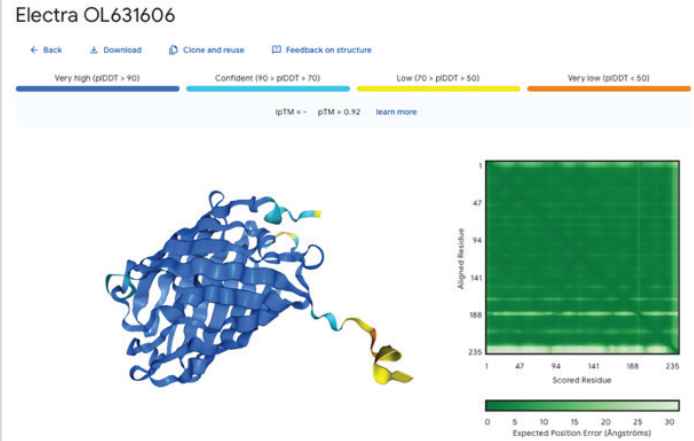
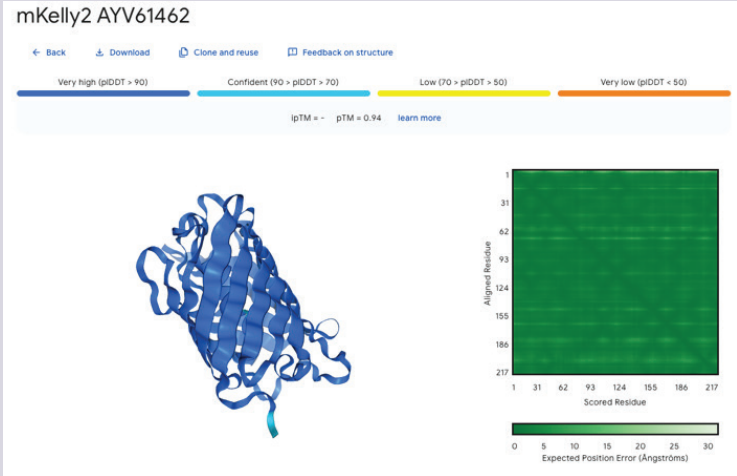
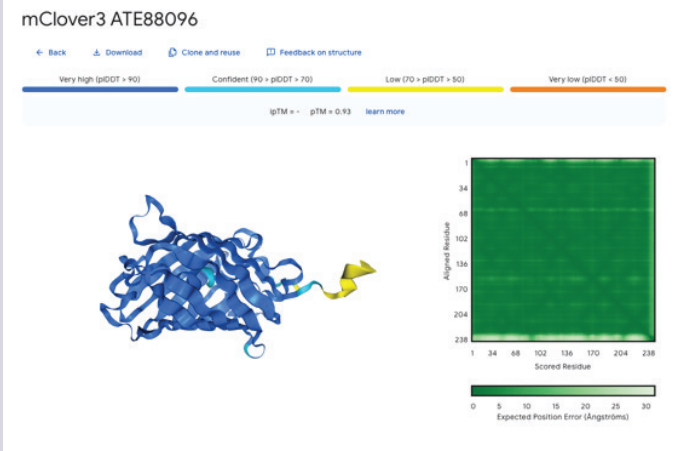
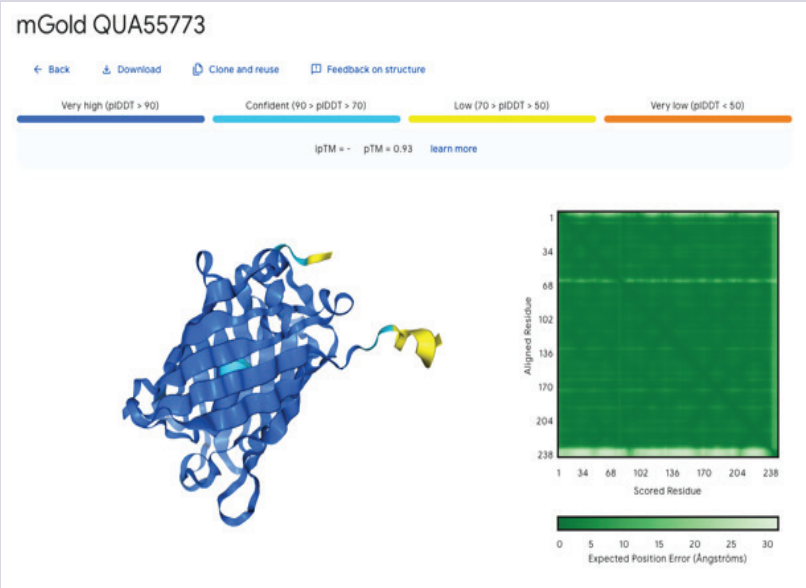


Figura 5. Capturas de pantalla de los pasos detallados en el cuadro 3 para realizar una predicción en AlphaFold.

Los resultados que se obtienen de la predicción se presentan a continuación en un cuadro resumen (cuadro 4).

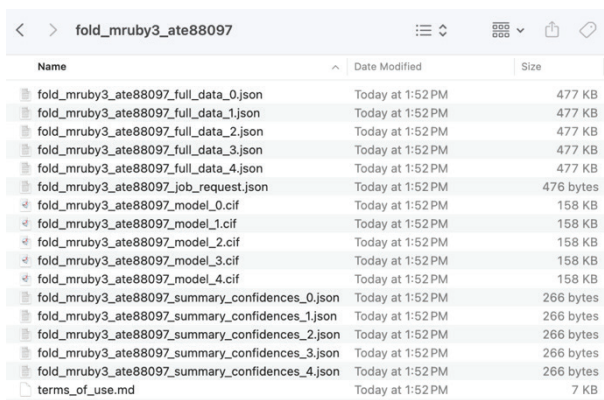
Cuadro 4. Resumen de los resultados obtenidos por AlphaFold para la predicción de estructura tridimensional de proteínas fluorescentes detalladas en el cuadro 2.

Secuencia	AlphaFold
>ATE88097.1 mRuby3 MVSKGEELIKENMRMKVVMESGVNG HQFKCTGEGEGRPYEGVQTMRIKVIEGG PLPFAFDILATSFMYGSRTFIKYPADIP DFFKQSFPEGFTWERVTRYEDGGVVT VTQDTSLEDGELVYNVKGVRGNFPSNG PVMQKKTKGWEPNTEMMYPADGGLR GYTDIALKVDGGGHLHCNFTTYRSKKT VGNKMPGVHAVDHRLERIEESDNETYV VQREVAVAKYSNLGGGMDELYK	 ipTM = -pTM = 0.9
>OL631606.1 Electra MVSKGEELIEENMRMKVVMESGVNG HQFKCTGEGEGRPYEGVQTMRIKVIEGG PLPFAFDILATSFLFGSKTFIKYPADIP DFFKQSFPEGFTWERVTRYEDGGVVT VTQDTSLEDGGLVYNVKGVRGNFPSNG PVMQKKTEGWEPFTEMMYPANGGLRGYT DIALKVDG DGHLHANIVTTYRSKKTVG DIKMPGVHAVDYRLERVEESDNETYVLR VAVAKYSNLGGGMDELYK	 ipTM = -pTM = 0.92
>AYV61462.1 mKelly2 MELIKENMRMKLYMEGTGNHHFKCTAE GEGKPYEGTQTQRIKVVEGGPLPFAFDI LATCFMYGSKTFINYPQGIPDFFKQSFPEG FTWERVTTYEDGGVLTVTQDTSLQDGyli YNVKLRGVNFPNGPVMQKKTLGWEP TEYLPVDGGLEGTCMALRLVGGGHLKC NLKTTYRSKKPAKNLKMPEGHQVDR KLERIKADNETYVEQHEVAVAR	 ipTM = -pTM = 0.94

Secuencia	AlphaFold
>ATE88096.1 mClover3 MVSKGEELFTGVVPILVELDGDVNGHKF SVRGEGEDATNGKLTCLKFICTTGKLPVP WPTLVTTFGYGVACFSRYPDHMKQHDFFK SAMPEGYVQERTISFKDDGTYKTRAEVK FEGDTLVNRIELKGIDFKEDGNILGHK LEYNFSNHYVITADKQKNCIKANFKIRHN VEDGSVQLADHYQQNTPIGDGPVLLPDN HYLSHQSKLSDPNEKRDHMLLEFVTAA GITHGMDELYK	mClover3 ATE88096 ← Back Download Clone and reuse Feedback on structure Very high (pIDDT > 90) Confident (90 > pIDDT > 70) Low (70 > pIDDT > 50) Very low (pIDDT < 50) ipTM = - pTM = 0.93 learn more  ipTM = -pTM = 0.93
>QUA55773.1 mGold MVSKGEELFTGVVPILVELDGDVNGHKF SVSGEGEGDATYGKLTCLKFICTTGKLPVP WPTLVTSLGYGQLQCFARYPDHMKQHDFFK SAMPEGYVQERTIFFKDDGNYKTRAEVK FEGDTLVNRIELKGIDFKEDGNILGHKLEY NYNSHNHYVITADKQKNGIKANFKIRHNIEDG GVQLADHYQQNTPIGDGPVLLPDNHYL SYQSKLSKDPNEKRDHMLLEFVTAAGITL GMDELYK	mGold QUA55773 ← Back Download Clone and reuse Feedback on structure Very high (pIDDT > 90) Confident (90 > pIDDT > 70) Low (70 > pIDDT > 50) Very low (pIDDT < 50) ipTM = - pTM = 0.93 learn more  ipTM = -pTM = 0.93

Nota: Para efectos prácticos se recomienda descargar el archivo de la predicción haciendo click en download.

El archivo debe contener los archivos en formato JSON y CIF, y el resumen en formato JSON (Fig. 6). Integralmente, estos archivos permitirán un análisis posterior de la predicción y disponer de los resultados para verificarlos en otras herramientas de visualización tridimensional.



Name	Date Modified	Size
fold_mrubby3_ate88097_full_data_0.json	Today at 1:52 PM	477 KB
fold_mrubby3_ate88097_full_data_1.json	Today at 1:52 PM	477 KB
fold_mrubby3_ate88097_full_data_2.json	Today at 1:52 PM	477 KB
fold_mrubby3_ate88097_full_data_3.json	Today at 1:52 PM	477 KB
fold_mrubby3_ate88097_full_data_4.json	Today at 1:52 PM	477 KB
fold_mrubby3_ate88097_job_request.json	Today at 1:52 PM	476 bytes
fold_mrubby3_ate88097_model_0.cif	Today at 1:52 PM	158 KB
fold_mrubby3_ate88097_model_1.cif	Today at 1:52 PM	158 KB
fold_mrubby3_ate88097_model_2.cif	Today at 1:52 PM	158 KB
fold_mrubby3_ate88097_model_3.cif	Today at 1:52 PM	158 KB
fold_mrubby3_ate88097_model_4.cif	Today at 1:52 PM	158 KB
fold_mrubby3_ate88097_summary_confidences_0.json	Today at 1:52 PM	266 bytes
fold_mrubby3_ate88097_summary_confidences_1.json	Today at 1:52 PM	266 bytes
fold_mrubby3_ate88097_summary_confidences_2.json	Today at 1:52 PM	266 bytes
fold_mrubby3_ate88097_summary_confidences_3.json	Today at 1:52 PM	266 bytes
fold_mrubby3_ate88097_summary_confidences_4.json	Today at 1:52 PM	266 bytes
terms_of_use.md	Today at 1:52 PM	7 KB

Figura 6. Captura de pantalla de los archivos descargables de AlphaFold y que son útiles para un posterior análisis de los resultados.

Nótese que en todas las anteriores la predicción corresponde a una proteína similar pero no idéntica. Es por lo anterior, que lo mejor es hacer una predicción específica de las proteínas en estudio.

Análisis de la predicción

Los datos más importantes del modelo de predicción sencilla para una única proteína son dos: el pLDDT y el PAE. Estos datos se deben interpretar en conjunto, y permiten determinar si la predicción es confiable. El pLDDT relacionado a la confianza del modelo y el PAE relaciona al error de la predicción. El resultado de las predicciones están en dos formatos: PDB y mmCIF. Estos incluyen los datos de coordenadas atómicas y el puntaje de pLDDT, y pueden observarse en distintas herramientas como por ejemplo el software PyMOL: <https://www.pymol.org/> o en iCn3D visor web: <https://www.ncbi.nlm.nih.gov/Structure/icn3d/>. AlphaFold también suministra el PAE: una indicación de su confianza y las posiciones relativas de los dominios. El PAE se almacena en un archivo independiente como JSON y se puede visualizar como un gráfico PAE. El modelado también suministra dos puntuaciones adicionales: pTM, que evalúa la precisión de la estructura general del complejo proteico, mientras que ipTM mide la precisión de las predicciones de la interfaz entre las subunidades del complejo proteína-proteína. Las puntuaciones de pTM e ipTM se encuentran en los archivos de registro y se guardan en el archivo JSON.

En la versión web del icn3d: <https://www.ncbi.nlm.nih.gov/Structure/icn3d/> se puede abrir el modelo 0 para mirarlo según se muestra en la Fig. 7. Para su mejor visualización, el sistema permite colorear el modelo a conveniencia. El usuario también puede dar clic en “analysis” para obtener un análisis de la proteína.

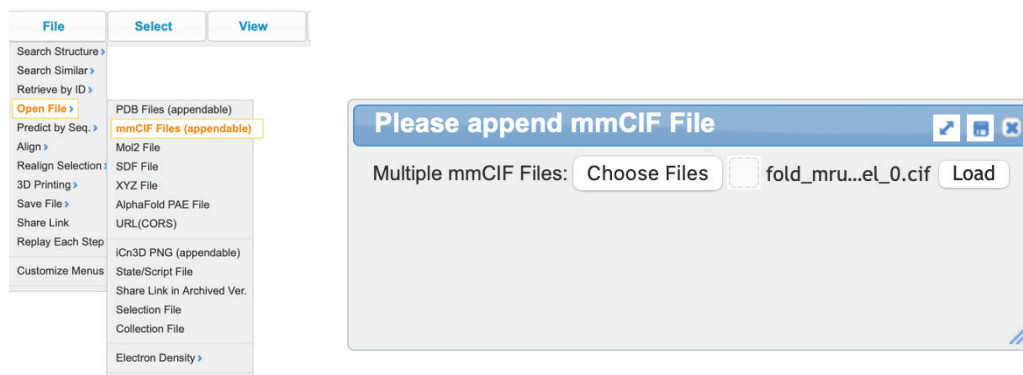


Figura 7. Captura de pantalla del programa icn3d útil para la lectura de los archivos descargados de AlphaFold en formato .cif

Verificación del modelo, pLDDT, PAE

En el sitio del visor de PAE sitio web: <https://subtiwiki.uni-goettingen.de/v4/paeViewerDemo> vamos a revisar el error de la predicción, para lo cual es necesario subir el archivo CIF y el JSON (full data) del modelo 0 según se muestra en la Fig. 8.

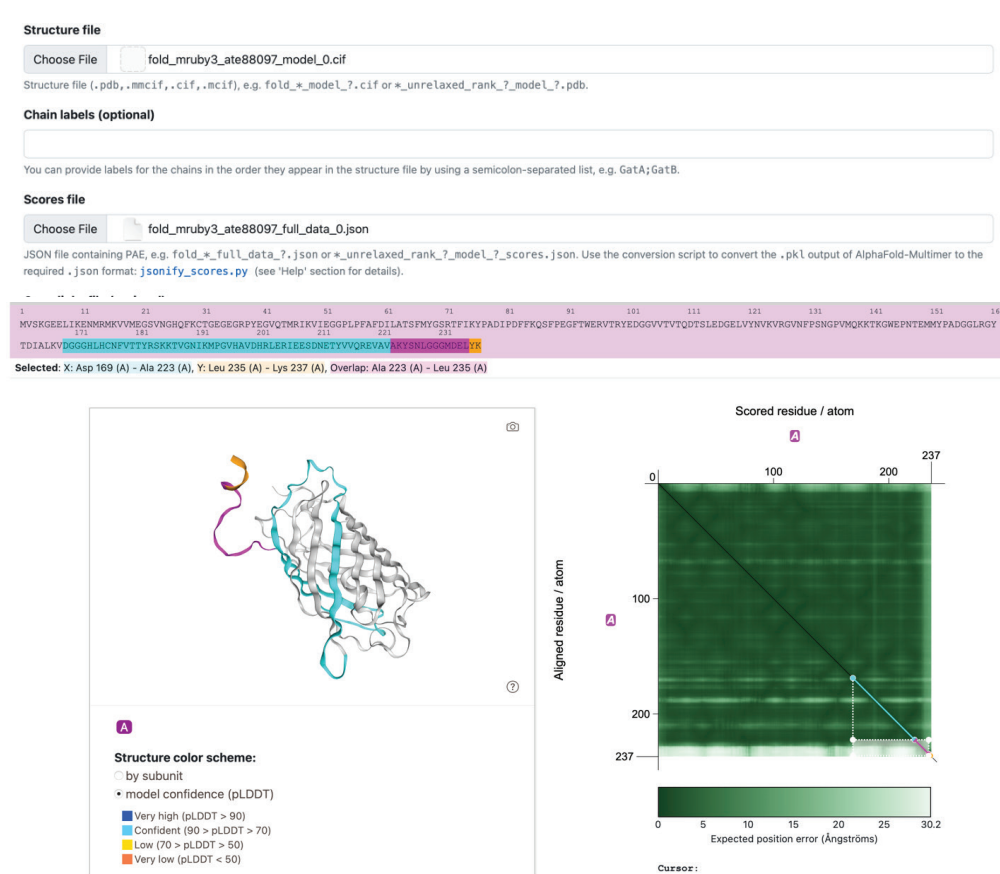


Figura 8. Captura de pantalla del visor web PAE útil para la identificar la calidad de la predicción de los archivos descargados de AlphaFold en formato CIF. Note que al marcar las zonas en blanco, el programa identificará los aminoácidos con más error de predicción.

Para mayor aprendizaje se recomienda el curso oficial de AlphaFold.³

Ética y principios de uso de la herramienta AlphaFold

La predicción 3D de proteínas permite el desarrollo de proteínas de novo con fines de bienestar de la humanidad [18]. Por ejemplo, este tipo de herramientas permitió el diseño de proteínas de novo para neutralizar el veneno de serpientes, o el tratamiento de enfermedades virales [19], [20]. Sin embargo, la existencia de proteínas nuevas y cuyo comportamiento podríamos no entender en su totalidad, representa un desafío desde el punto de vista de seguridad y ética profesional. En particular, existe la posibilidad que estos avances permitan desarrollar tecnologías con fines bélicos o no relacionados con el bienestar de la humanidad.

En ese sentido, AlphaFold aclara en sus estatutos que los principios que rigen el uso de esta herramienta son consecuentes con los principios establecidos por la OCDE, en el sitio web: <https://oecd.ai/en/ai-principles>. Estos principios son (1) crecimiento inclusivo, desarrollo sostenible y bienestar; (2) protección de derechos humanos y valores democráticos, incluyendo equidad y privacidad; (3) transparencia y explicación del sistema; (4) solidez, seguridad y protección; y (5) rendición de cuentas.

Los principios específicos que gobiernan el uso de la herramienta AlphaFold están detallados en el siguiente enlace: <https://deepmind.google/discover/blog/how-our-principles-helped-define-alphafolds-release/>. Los principios tipifican el compromiso de favorecer el (1) beneficio social, (2) excelencia e integridad científica; (3) seguridad y ética, (4) rendición de cuentas ante las personas; (5) compartir el conocimiento de forma responsable y (6) diversidad, equidad e inclusión.

El sitio de AlphaFold aclara que no promueve: (1) tecnologías que causen o puedan causar daños generales. Cuando exista un riesgo significativo de daño, procederemos solo si consideramos que los beneficios superan con creces los riesgos e incorporaremos las restricciones de seguridad adecuadas; (2) armas u otras tecnologías cuyo propósito principal o implementación sea causar o facilitar directamente lesiones a las personas; (3) tecnologías que recopilen o utilicen información para vigilancia que violen las normas internacionalmente aceptadas; o (4) tecnologías cuyo propósito contravenga los principios ampliamente aceptados del derecho internacional y los derechos humanos.

Conclusiones

AlphaFold es una herramienta útil para la predicción tridimensional de estructuras proteicas, su uso es gratuito y requiere de entender las métricas de calidad y precisión que se detallan en el presente documento.

El resultado de este tipo de predicciones es posible debido al trabajo histórico de predicción mediante rayos X, cristalografía, espectrometría de resonancia magnética nuclear y microscopía crioelectrónica. Sin embargo, existen proteínas que no se han podido analizar debido a que se encuentran en espacios intermembrana o en condiciones que hacen imposible su purificación. Esto implica que existe un vacío de información estructural en ciertos casos, lo que puede limitar la precisión de las predicciones.

3 AlphaFold, A practical guide. <https://www.ebi.ac.uk/training/online/courses/alphafold/>. DOI: 10.6019/TOL.AlphaFold-w.2024.00001.1

AlphaFold también ofrece módulos para analizar interacciones más complejas con otras moléculas, aspecto que solo se menciona brevemente en este documento, pero que se recomienda explorar dado su potencial para desarrollar nuevas soluciones basadas en proteínas.

Referencias

- [1] C. B. Anfinsen, "Principles that govern the folding of protein chains," *Science*, vol. 181, no. 4096, pp. 223–230, 1973, doi: 10.1126/science.181.4096.223.
- [2] R. Zwanzig, A. Szabo, and B. Bagchi, "Levinthal's paradox," *Proc. Natl. Acad. Sci. USA*, vol. 89, no. 1, pp. 20–22, 1992, doi: 10.1073/pnas.89.1.20.
- [3] J. Jumper et al., "Highly accurate protein structure prediction with AlphaFold," *Nature*, vol. 596, no. 7873, pp. 583–589, 2021, doi: 10.1038/s41586-021-03819-2.
- [4] J. Abramson et al., "Accurate structure prediction of biomolecular interactions with AlphaFold 3," *Nature*, vol. 630, no. 8016, pp. 493–500, 2024, doi: 10.1038/s41586-024-07487-w.
- [5] U. Göbel, C. Sander, R. Schneider, and A. Valencia, "Correlated mutations and residue contacts in proteins," *Proteins*, vol. 18, no. 4, pp. 309–317, 1994, doi: 10.1002/prot.340180402.
- [6] S. A. Benner and D. Gerloff, "Patterns of divergence in homologous proteins as indicators of secondary and tertiary structure," *Adv. Enzyme Regul.*, vol. 31, pp. 121–181, 1991, doi: 10.1016/0065-2571(91)90012-B.
- [7] W. R. Taylor and K. Hatrick, "Compensating changes in protein multiple sequence alignments," *Protein Eng. Des. Sel.*, vol. 7, no. 3, pp. 341–348, 1994, doi: 10.1093/protein/7.3.341.
- [8] V. Mariani, M. Biasini, A. Barbato, and T. Schwede, "IDDT: A local superposition-free score for comparing protein structures and models," *Bioinformatics*, vol. 29, no. 21, pp. 2722–2728, 2013, doi: 10.1093/bioinformatics/btt473.
- [9] H.-B. Guo et al., "AlphaFold2 models indicate that protein sequence determines both structure and dynamics," *Sci. Rep.*, vol. 12, p. 10696, 2022, doi: 10.1038/s41598-022-14382-9.
- [10] G. Kim et al., "Easy and accurate protein structure prediction using ColabFold," 2023. [Online]. Available: <https://colabfold.mmseqs.com>
- [11] R. Krishna et al., "Generalized biomolecular modeling and design with RoseTTAFold All-Atom," *Science*, vol. 384, no. 6693, p. eadl2528, 2024, doi: 10.1126/science.adl2528.
- [12] M. van Kempen et al., "Fast and accurate protein structure search with Foldseek," *Nat. Biotechnol.*, vol. 42, no. 2, pp. 243–246, 2024, doi: 10.1038/s41587-023-01773-0.
- [13] M. Varadi et al., "AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences," *Nucleic Acids Res.*, vol. 52, no. D1, pp. D368–D375, 2024, doi: 10.1093/nar/gkad1011.
- [14] B. T. Bajar et al., "Improving brightness and photostability of green and red fluorescent proteins for live cell imaging," *Sci. Rep.*, vol. 6, p. 20889, 2016, doi: 10.1038/srep20889.
- [15] S. Papadaki et al., "Dual-expression system for blue fluorescent protein optimization," *Sci. Rep.*, vol. 12, p. 13214, 2022, doi: 10.1038/s41598-022-13214-0.
- [16] T. M. Wannier et al., "Monomerization of far-red fluorescent proteins," *Proc. Natl. Acad. Sci. USA*, vol. 115, no. 48, pp. E11294–E11301, 2018, doi: 10.1073/pnas.1807449115.
- [17] J. Lee et al., "Versatile phenotype-activated cell sorting," *Sci. Adv.*, vol. 6, no. 43, p. eabb7438, 2020, doi: 10.1126/sciadv.abb7438.
- [18] J. L. Watson et al., "De novo design of protein structure and function with RFdiffusion," *Nature*, vol. 620, no. 7976, pp. 1089–1100, 2023, doi: 10.1038/s41586-023-06415-8.
- [19] S. Vázquez Torres et al., "De novo designed proteins neutralize lethal snake venom toxins," *Nature*, vol. 639, no. 8053, pp. 225–231, 2025, doi: 10.1038/s41586-024-08393-x.
- [20] S. Tang et al., "De novo gene synthesis by an antiviral reverse transcriptase," *Science*, vol. 386, no. 6717, pp. 1–41, 2024, doi: 10.1126/science.adq0876.

Declaración sobre uso de Inteligencia Artificial (IA)

Los autores aquí firmantes declaramos que no se utilizó ninguna herramienta de IA para la conceptualización, traducción o redacción de este artículo.