

Predicción estructural de conotoxinas con AlphaFold2

Introducción

AlphaFold es un programa de inteligencia artificial desarrollado por DeepMind que realiza predicciones de la estructura de proteínas a partir de la secuencia primaria, mediante el uso de redes neuronales. La segunda versión de este programa se desarrolló en 2020 y obtuvo resultados sobresalientes, comparado a otros sistemas existentes [1].

En este tutorial se indica cómo realizar un ejercicio de predicción estructural a partir de una secuencia primaria, usando una implementación de Alphafold2 y RosettaFold en Google Colab.

Uso de AlphaFold2 en Google colab.

Colab es una herramienta de la suite de Google, que permite escribir y ejecutar código Python desde un navegador. Para hacer uso de esta herramienta es necesario y suficiente contar con una cuenta activa de correo gmail.

La implementación de AlphaFold2 en colab se encuentra en:

<https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>

Es importante que antes de ejecutar los bloques de código, hagas una lectura del texto que acompaña al código. La implementación de AlphaFold2 consiste de 5 bloques de código, los cuales deben ejecutarse secuencialmente de acuerdo al orden en el que se encuentran dispuestos. Una vez que accedas a la liga mencionada, notarás que hay un ejemplo de prueba que está precargado, en el cual aparece una secuencia “query”, un “jobname” que en este caso es “test” y algunos parámetros establecidos, como número de modelos a generar, número de ciclos, etc. Es recomendable que realices un primer ejercicio usando el ejemplo precargado, para asegurarse que todo esté funcionando correctamente. Si la ejecución del ejercicio es exitosa, al ejecutar el último bloque de código se descargará a tu equipo un archivo .zip que contiene los resultados de la ejecución.

Predicción estructural de conotoxinas.

Una vez realizada la prueba indicada en el apartado anterior, puedes proceder a realizar el ejercicio de predicción estructural con las secuencias de conotoxinas que se te han asignado. Para cada secuencia, el “jobname” a utilizar será el correspondiente ID de la secuencia. Los demás parámetros a utilizar serán los que están marcados por default, es decir, los mismos que se usaron en la prueba del apartado anterior. Para cada péptido modelado, se deben descargar los archivos de resultados.

[1] <https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology>

Lineamientos para reporte

Escribir un reporte individual que incluya al menos lo siguiente:

- Breve descripción de los procedimientos para obtener los modelos (preprocesamiento de la secuencia *query*, parámetros seleccionados para la predicción estructural y cualquier otro aspecto que consideren pertinente)
- Resultados y discusión (presentación de modelos, evaluación de la calidad (gráficas), posibles observaciones que relacionen la calidad de los resultados con el MSA y cualquier otro aspecto que consideren relevante)
- Conclusión
- Referencias

En DropBox, ir en la carpeta **Conotoxins_AF2_2023**;

<https://www.dropbox.com/scl/fo/m4x0aaxc0g92zxy8w4nl5/h?dl=0&rlkey=ntyynrses a9e0rc6xug1zh8df>

- Subir su archivo en formato PDF a la carpeta correspondiente a **Reportes**.
- Subir sus resultados en formato Excel a la carpeta correspondiente a **Archivos**
Por cada secuencia crear una carpeta con su código PDB (p.ej. *A0A0B5A8Q6*), subir los resultados .zip de AF2 (las predicciones en formato .pdb, archivos .a3m and .log, imágenes .png) y RosettaFold

Si tienen alguna duda, pueden ponerse en contacto escribiendo un correo a fabien.plisson@cinvestav.mx