



Instituto Politécnico Nacional
Escuela Superior de Cómputo



Bioinformatics

Lab Session 6. Protein Structure Prediction

González Bocio Erik Alexander 2020630163

Jorge Luis Rosas Trigueros

Fecha de realización: 4 de abril de 2022

Fecha de entrega: 25 de abril de 2022

Marco teórico:

La predicción de la estructura de las proteínas es un desafío de larga data en la biología computacional. A través de la extensión de la predicción basada en el aprendizaje profundo a orientaciones interresiduas además de las distancias, y el desarrollo de una optimización restringida por Rosetta, mostramos que se pueden generar modelos más precisos. Los resultados sobre un conjunto de 18 proteínas diseñadas de novo sugieren que el método propuesto debería ser directamente aplicable a los desafíos actuales en el diseño de proteínas de novo.

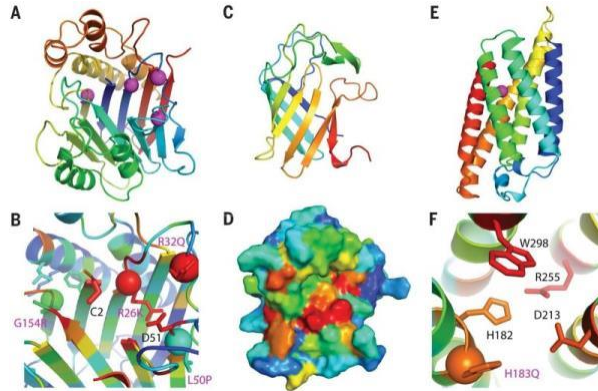


Fig. 1. Ejemplos de predicción de la estructura de proteínas.

Las proteínas son esenciales para la vida, y la comprensión de su estructura puede facilitar una comprensión mecanicista de su función. A través de un enorme esfuerzo experimental, se han determinado las estructuras de alrededor de 100.000 proteínas únicas, pero esto representa una pequeña fracción de los miles de millones de secuencias de proteínas conocidas.

Predecir la estructura tridimensional que adoptará una proteína basándose únicamente en su secuencia de aminoácidos, el componente de predicción de la estructura del "problema del plegamiento de proteínas"—ha sido un importante problema de investigación abierta durante más de 50 años.

Por su parte, SWISS-MODEL es un servidor de homología-modelado de estructura de proteínas totalmente automatizado, accesible a través del . Expasy web server, o desde el programa DeepView (Swiss Pdb-Viewer).



Fig. 2. Logotipo de SWISS-MODEL

El propósito de este servidor es hacer que el modelado de proteínas sea accesible para todos los investigadores de ciencias de la vida en todo el mundo.

RaptorX es un software y servidor web para la predicción de la estructura y función de proteínas que es gratuito para uso no comercial. RaptorX es uno de los métodos más populares para la predicción de la estructura de proteínas. Al igual que otras técnicas de reconocimiento remoto de homología / enhebrado de proteínas, RaptorX es capaz de generar regularmente modelos de proteínas confiables cuando el AMPLIAMENTE UTILIZADO PSI-BLAST no puede. Sin embargo, RaptorX también es significativamente diferente de esos métodos basados en perfiles (por ejemplo, HHPred y Phyre2) en que RaptorX sobresale en el modelado de secuencias de proteínas sin un gran número de homólogos de secuencias mediante la explotación de la información de estructura. RaptorX Server ha sido diseñado para garantizar una interfaz fácil de usar para usuarios inexpertos en métodos de predicción de la estructura de proteínas.



Fig. 3. Logotipo de RaptorX

RaptorX es desarrollado por el grupo Xu, sobresaliendo en la predicción terciaria y de contacto para secuencias de proteínas sin homólogos cercanos en el Banco de Datos de Proteínas (PDB). RaptorX predice estructuras secundarias y terciarias de proteínas, mapa de contacto y distancia, accesibilidad con disolventes, regiones desordenadas, anotación funcional y sitios de unión.

Material y Equipo:

- Computadora
- Red Wifi
- Plataforma de SWISS-MODEL
- Plataforma de RaptorX
- Plataforma del UnitPro

Desarrollo de la práctica:

Para empezar con el uso de las predicciones de estructuras de proteínas vamos a tomar una proteína del Uniprot, para esto se ingresó la siguiente búsqueda avanzada en el cuadro de búsqueda.

NOT database:(type:pdb) length:[100 TO *] AND reviewed:no AND organism:"Homo sapiens (Human) [9606]"

Donde no hay archivo pdb asociado, ósea que no hay estructura conocida para estas proteínas, con una longitud de 100 en adelante, que no estén revisadas y sean de Homo Sapiens, ósea del humano.

Por lo que se eligió la proteína Amyloid-beta A4 (A0A140VJC8), se eligió esta ya que luego de buscar se encontró que funciona como un receptor de la superficie celular y realiza funciones fisiológicas en la superficie de las neuronas relevantes para el crecimiento de neuritas, la adhesión neuronal y la axonogénesis, por lo que resulto interesante trabajar con esta proteína.

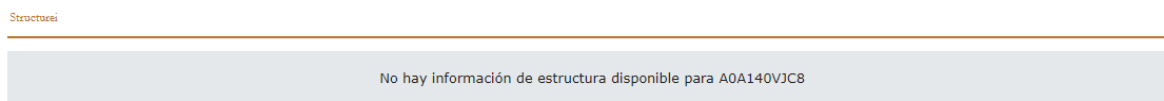


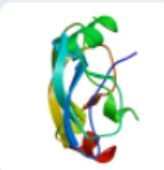
Fig. 4. No hay estructura de esta proteína

Luego de haber hecho la elección de esta proteína pasamos a hacer el segundo punto de la práctica el cuál es visitar la página de SWISS-MODEL y con ayuda de la secuencia construir el modelo por Homología.

Por lo tanto, ingresamos la secuencia y hacemos el Build Model


Fig. 5. Formulario del Build Model de la proteína seleccionada

Al hacer el Build Model, tenemos el siguiente resultado:



Model 03 ▾

Structure Assessment

 ☐

Oligo-State

Ligands

GMQE

QMEANDisCo Global

Homo-dimer

None

0.03

0.70 ± 0.08

(matching prediction)

QMEANDisCo Local

QMEAN Z-Scores ▾

Template

Seq Identity

Coverage

Description

1aap.1.A

100.00%


ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN

X-RAY CRYSTAL STRUCTURE OF THE PROTEASE INHIBITOR

DOMAIN OF ALZHEIMER'S AMYLOID BETA-PROTEIN


PRECURSOR

Model-Template Alignment ▾



Model 02 ▾

Structure Assessment

 ☐

Oligo-State ⓘ

Ligands ⓘ

GMQE ⓘ

QMEANDisCo Global ⓘ

Homo-dimer

1 x BU4 ⓘ ▾

0.19

0.71 ± 0.05

(matching prediction)

QMEANDisCo Local

QMEAN Z-Scores ▾

Template

Seq Identity

Coverage

Description

3ktm.1.A

100.00%

Amyloid beta A4 protein

Structure of the Heparin-induced E1-Dimer of the Amyloid Precursor

Protein (APP)

Model-Template Alignment ▾



Fig. 6. Resultados del Build Model de la proteína seleccionada

Obteniendo el siguiente resultado con el primer modelado

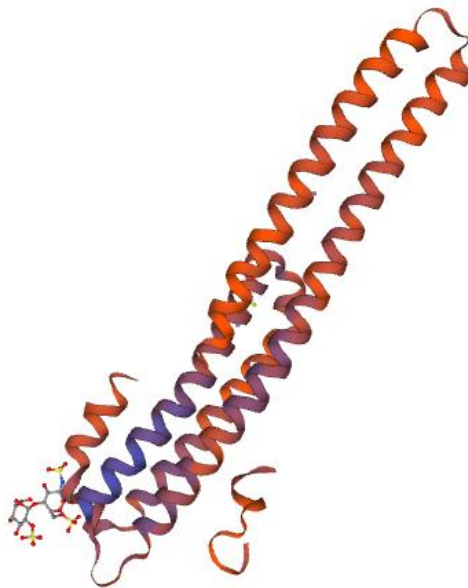


Fig. 7. Resultado de la predicción, en este caso la página nos mostró que era el más cercano de acuerdo con los templates

Al hacer uso del Structure Assessment tenemos la siguiente estructura de Ramachandran

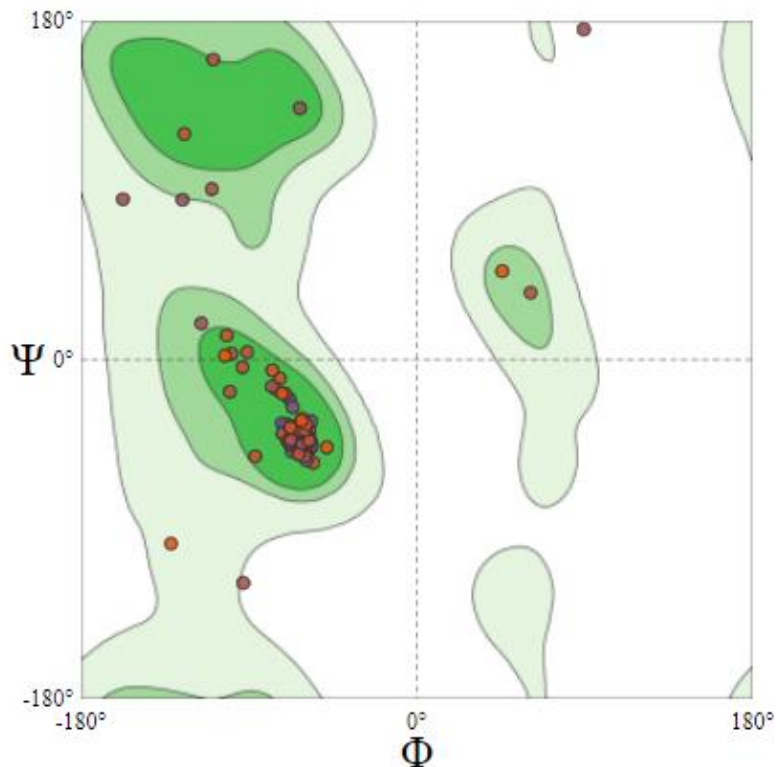


Ilustración 8. Estructura de Ramachandran del modelado más cercano según la predicción

Podemos ver que la mayoría de los ángulos diedros de la hélice están en las regiones más pobladas del diagrama de Ramachandran, diciéndonos que esa estructura puede estar en la naturaleza.

Finalmente usamos RaptorX para hacer el Threading, por lo que también colocamos la secuencia, en este caso RaptorX se lleva un par de tiempo para hacer la predicción por lo que puede ingresarse un email para poder observar los resultados.

Job Identification

Jobname:
Email:

Sequences for Prediction

Sequences:

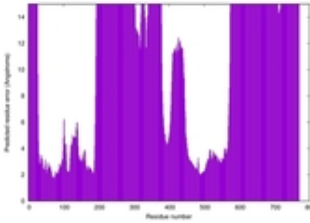

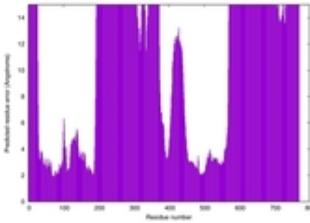

```
>seq1
ENIEVHMLNKGAEAMVFEPAIKANPGDVTFTIPVDKGHNVESIKDMIPEGA EKFKSKINENYVLTVTQPGAYLVKCT
PHYAMGMIALIAVGDSPANLDQIVSAKKPKIVQERLEKVIASAK
```

Sequence file: No se eligió ningún archivo

Ilustración 9. Formulario de RaptorX para llevar a cabo el Threading

Lamentablemente el servidor de RaptorX tuvo problemas con la cantidad de solicitudes que le llegaron, por lo que fue necesario usar otro servidor, el servidor utilizado fue el de la Universidad de Reading, luego de llevar a cabo el Threading se obtuvo el siguiente resultado:

Primero obtuvimos una serie de modelos los cuales iban de mayor a menor puntuación.

ID de modelo y enlaces DE PDF para todas las plantillas utilizadas	Confianza y valor P	Puntuación de calidad del modelo global	Gráfico de calidad de modelo local (haga clic en las imágenes para descargar gráficos)	Mapa 3D de la calidad del modelo
IntFOLD6 HHsearch multi8 TS1 5tptA 3pmrA 3k66A 3ktmA 2lp1A 5nx1D	UNCERT: 2.056E-1	0.4252		 <div>Ver modelo en 3D y descargar</div> <div>Refinar el modelo con ReFOLD3</div>
IntFOLD6 HHsearch multi6 TS1 5tptA 3ktmA 2lp1A 5nx1D	UNCERT: 2.093E-1	0.4239		 <div>Ver modelo en 3D y descargar</div> <div>Refinar el modelo con ReFOLD3</div>

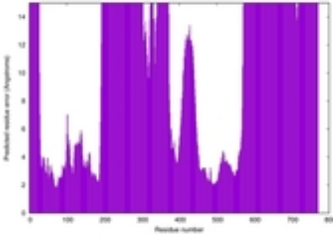

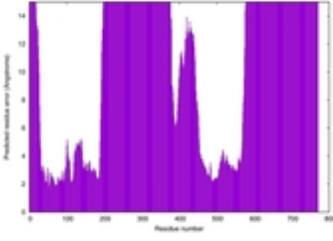

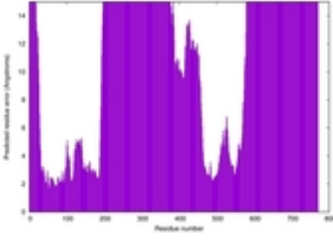

<p>IntFOLD6 IntFOLDTS140 multi2 TS1</p> <p>5tptA 3ktmA 4ntwB 2lp1A</p>	<p>UNCERT: 2.128E-1</p>	<p>0.4227</p>		 <p>Ver modelo en 3D y descargar</p> <p>Refinar el modelo con ReFOLD3</p>
<p>IntFOLD6 dPPAS multi8 TS1</p> <p>5tptA 3k66A 3pmrA 3ktmE</p>	<p>UNCERT: 2.153E-1</p>	<p>0.4218</p>		 <p>Ver modelo en 3D y descargar</p> <p>Refinar el modelo con ReFOLD3</p>
<p>IntFOLD6 dPPAS2 multi4 TS1</p> <p>3ktmE 3pmrA 3k66A</p>	<p>UNCERT: 2.177E-1</p>	<p>0.4210</p>		 <p>Ver modelo en 3D y descargar</p>

Fig. 10. Resultados del Threading de la proteína seleccionada

El más votado fue el que se muestra a continuación:

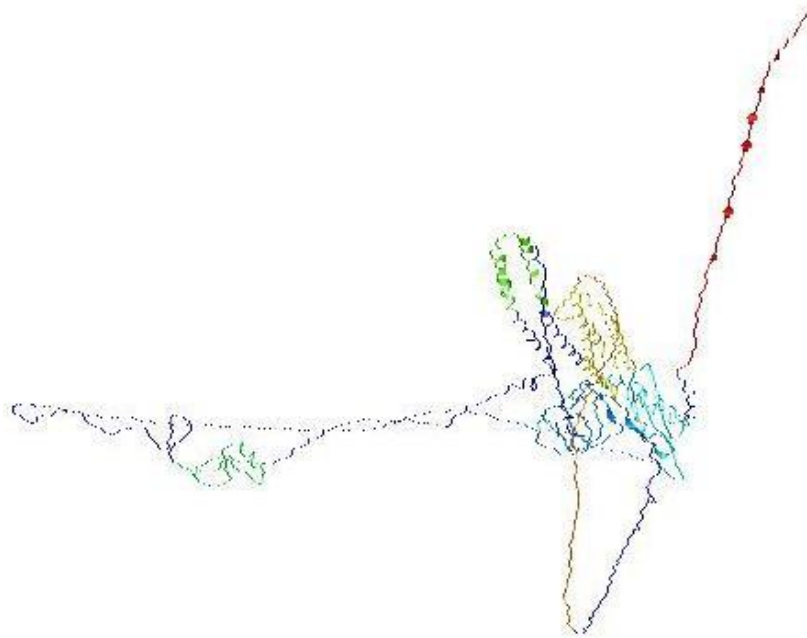


Fig. 11. Estructura más cercana a la proteína seleccionada según la página de la universidad

Este tiene una puntuación de calidad del modelo global de 0.4252, siendo este el más alto, también tenemos otros como los son:

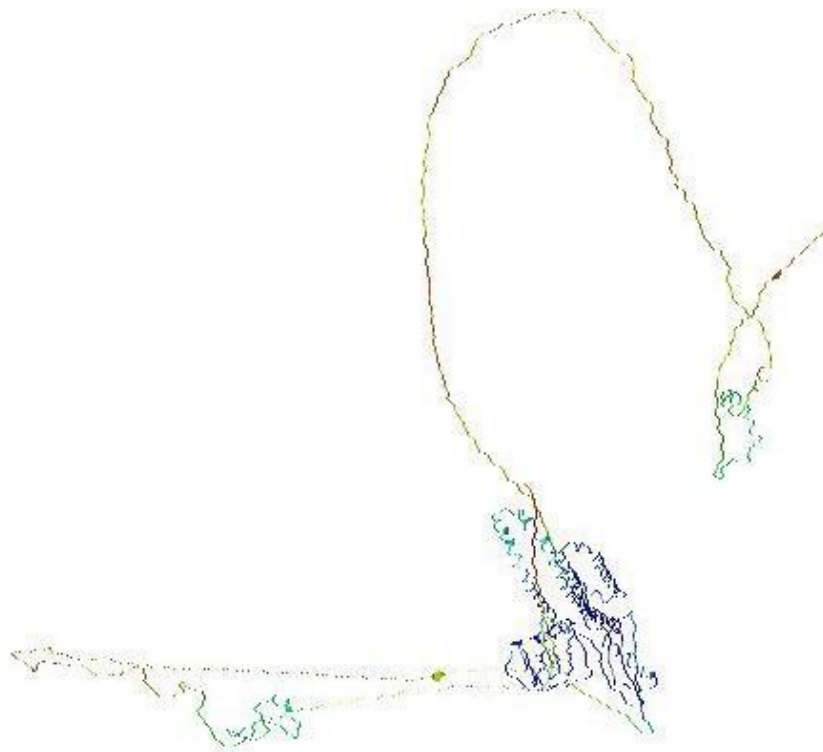


Fig. 12. Segunda estructura más cercana según la puntuación

Con una puntuación de calidad del modelo global de 0.4239

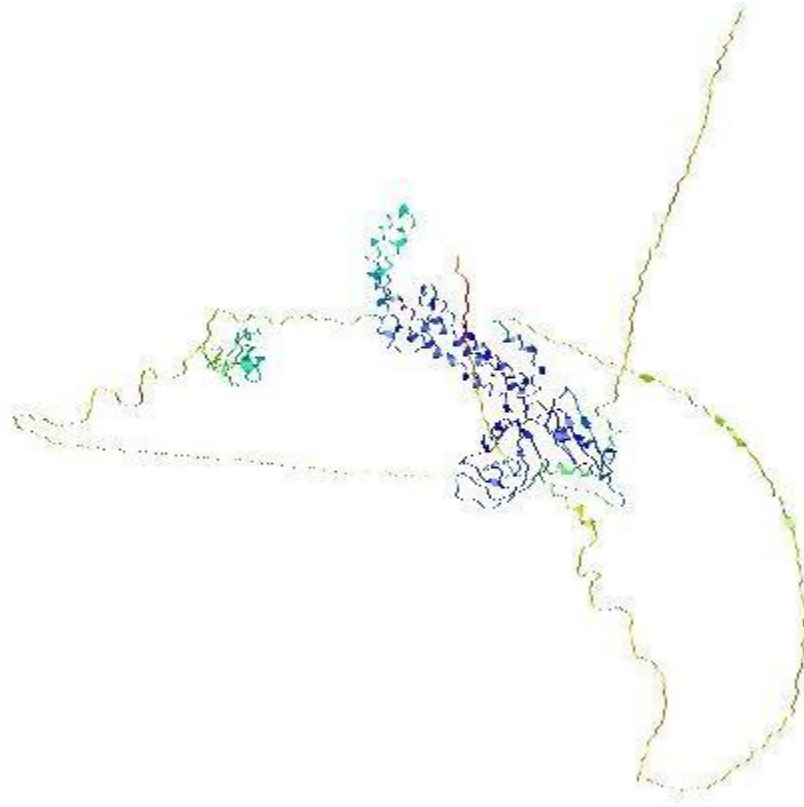


Fig. 13. Tercera estructura más cercana según la puntuación

La cual tiene una puntuación de 0.4227

Al igual que estos hay más cada vez con menos puntuación, por lo que nos quedamos con el primer modelado.

La predicción de trastornos para la proteína dada por la puntuación de probabilidad de trastorno y el número de residuo queda de la siguiente manera

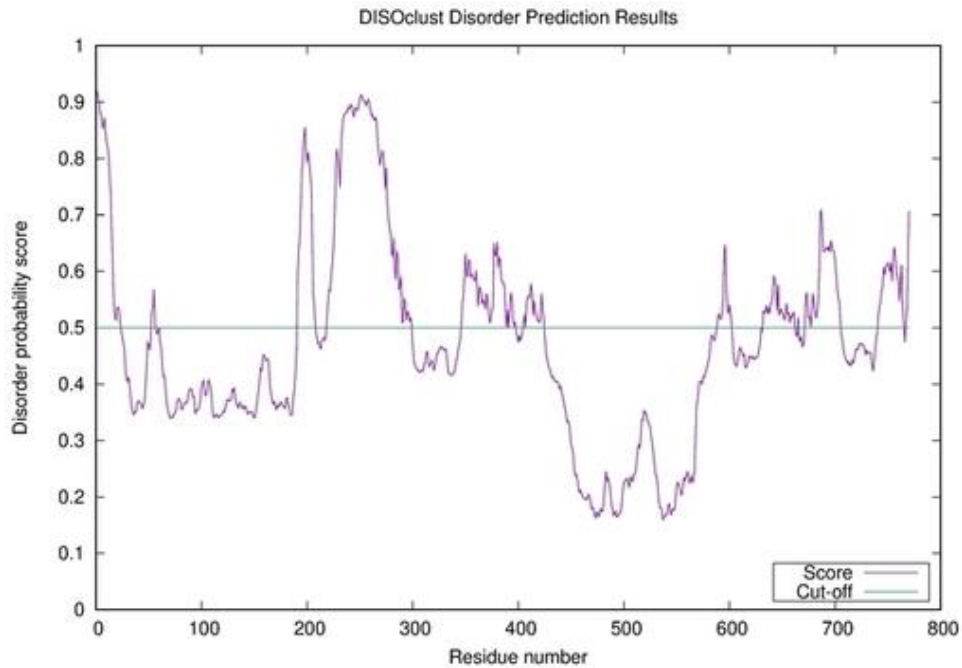


Fig. 14. Predicción de trastornos para la proteína

Mientras que la predicción de límites queda:

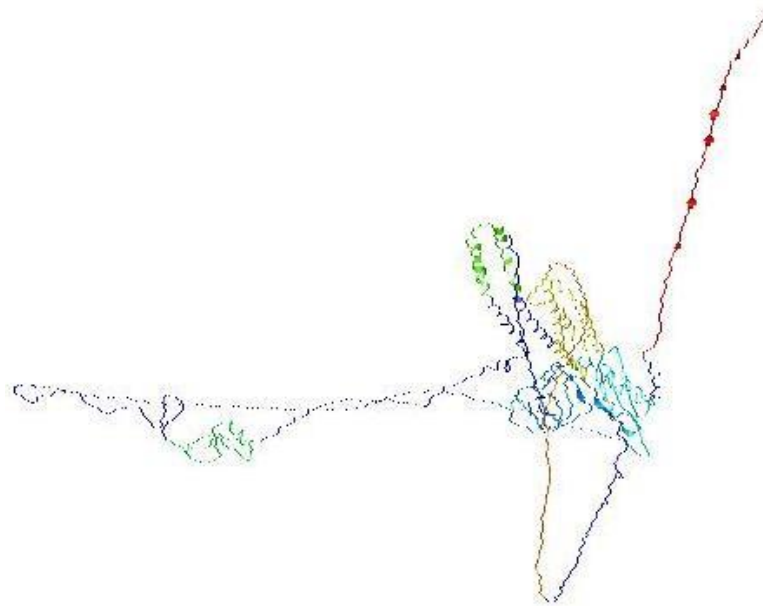


Ilustración 15. Predicción de límites para la proteína

Conclusiones y recomendaciones:

Con esta práctica vimos una introducción a la predicción de estructuras de proteínas, también lo llevamos a cabo con una proteína que no tiene estructura en el Uniprot, así, con ayuda de algunos servidores dedicados a la predicción

encontramos varios modelos que pueden ser el de la proteína elegida, en este caso hicimos dos, una por homología y otra con threading.

Me gusto esta práctica ya que primero pudimos ver como con ayuda de varios templates se puede hacer una predicción de la estructura, también con el threading había que esperar algún tiempo para poder ver los resultados.

Bibliografía:

- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., & Ronneberger, O. et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583-589. <https://doi.org/10.1038/s41586-021-03819-2>
- *RaptorX web servers for protein sequence, structure and functional study*. Raptorx.uchicago.edu. (2022). Retrieved 8 April 2022, from <http://raptorx.uchicago.edu/>.
- *SWISS-MODEL*. Swissmodel.expasy.org. (2022). Retrieved 8 April 2022, from <https://swissmodel.expasy.org/>.
- Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S., & Baker, D. (2020). Improved protein structure prediction using predicted interresidue orientations. *Proceedings Of The National Academy Of Sciences*, 117(3), 1496-1503. <https://doi.org/10.1073/pnas.1914677117>