

Modelado por Homología

Lucía Chemes y Juliana Glavina

Ejercicio 1.

Luego de dos años y numerosos intentos fallidos, usted logró determinar por resonancia magnética nuclear una región de la proteína misteriosa y depositar la estructura en la base de datos de proteínas PDB (1F7W).

Años después ocurre una pandemia de una enfermedad respiratoria causada por *Actinobacillus pleuropneumoniae* que está causando un rápido aumento en la mortalidad de la población porcina, trayendo terribles consecuencias en la actividad económica mundial. Una vez que se logró aislar la cepa responsable, se cree que una proteína que comparte casi el 25 % de identidad con la proteína misteriosa es un posible blanco para el diseño de una droga. Sin embargo, se desconoce la estructura de la misma. Como usted es el único experto en esa proteína en el mundo, la Asociación del Centro Médico Epidemiológico (ACME) se pone en contacto con usted en busca de una solución. Para solucionar el problema, Ud. decide primero intentar un modelado por homología de la nueva proteína.

1. Ingrese la secuencia de la proteína misteriosa patogénica en HHPred
<https://toolkit.tuebingen.mpg.de/tools/hhpred>

```
>Pathogenic Mistery Protein
```

```
MELHILFFILAGLLIAVLISFSLWSARREKSRIFSNTFSTRPPSTPINNIVSDVPPSLNPQS  
YAQTTGQHGETEADNPVQIQQEVESLREIKINLPGQDSAAYQSKVEETPIYSGQPVLVPVQP  
QYQTQVQYQTQPHIEPAFTQAPQSPIAEATSVLEQSVEELERQAAQGDVDIYSDASVRVEL  
AKNSMQADSVAEQKPVAENNMLTLYVVAPEGQQFRGDYVVQSLEALGFQYGEYQIFHRHQHM  
GNSASPVIFSVANMMQPGIFDLTKIEHFSTVGLVLFMHLPSEGNDVVNFKLLLKTENLAQA  
LGGFVLNEHREIFDENSQRQSYLARVS
```

2. Seleccione el hit que le parezca más conveniente y luego seleccione en la parte superior *Model using selection*. ¿Qué se muestra en la nueva ventana?
3. Haga click en *Forward to Modeller* y luego en *Submit*. (De ser necesario ingrese la siguiente key: MODELIRANJE y luego haga click en *Submit*)
¿Qué aparece en la nueva ventana?
4. Descargue el archivo PDB (*Download PDB File*)
5. La herramienta **Verify3D** permite determinar la compatibilidad de un modelo 3D de una proteína con su secuencia aminoacídica en base a cuál es el

ambiente en el cual se encuentra cada residuo y la compatibilidad con la estructura secundaria en la que se encuentra.

Vaya a la web de **Verify3D** (<https://servicesn.mbi.ucla.edu/Verify3D/>), suba el archivo PDB obtenido en el paso anterior y corra el programa.

El gráfico reporta la calidad del modelo por posición y en él se observan tres regiones: Las posiciones con score menor a cero están mal modeladas, las posiciones con score entre cero y 0.2 están pobremente modeladas, y las posiciones con score mayor a 0.2 están modeladas con buena calidad. Verify 3D asigna como aceptado a un modelo con >80% posiciones en el área “bien modelada”.

Observe el resultado obtenido. ¿Cuál es el score global? ¿Cuál es el score de los residuos?

6. La herramienta **Procheck** permite analizar la calidad de la geometría de los residuos en una estructura proteica dada en comparación a parámetros estereoquímicos derivados de estructuras tridimensionales de alta resolución ya conocidas.
En la parte superior elija **Procheck**. en *Choose File*, cargue el archivo PDB y haga click en *Run Procheck*.
 - a. Investigue los plots de Ramachandran. ¿Qué residuos no están en el área esperada?
 - b. Investigue las propiedades de los residuos. ¿Qué residuos se alejan de los ángulos diédricos esperados?
 - c. Investigue los gráficos de las longitudes de enlace en la cadena principal (*Main-chain bond lengths*) y los ángulos de unión de la cadena principal (*Main-chain bond angles*). ¿Existen aminoácidos que se alejen significativamente de los resultados esperados?
7. En base a los resultados obtenidos por Verify3D y ProCheck responda: ¿Es bueno el modelo? ¿Por qué?
8. Abra chimera y busque el modelo que determinó usted años atrás (Por si no recuerdan: *File* → *Fetch by ID* → 1F7W)
9. Luego, cargue en la misma ventana de Chimera la estructura de la proteína misteriosa patogénica (*File* → *Open*).
10. Para tener una noción de cuán similar es la estructura de dos proteínas, podemos realizar un **Alineamiento Estructural**, que consiste en superponer las estructuras de ambas proteínas en el espacio intentando alinear sus cadenas aminoacídicas. Alinear estructuras en chimera es muy fácil, sólo requiere un comando.

Vaya a *Tools* → *Structure Comparison* → *MatchMaker*

Se abrirá una nueva ventana.

En *Reference structure* (el panel de la izquierda) puede seleccionar una de las estructuras de referencia. Esta estructura es la que se mantendrá fija. (Ej. 1F7W)

En *Structure(s) to match* (el panel de la derecha) seleccione la estructura que será superpuesta y alineada con la que se eligió como referencia. (Ej. el modelo)

Observe el resultado del alineamiento: ¿Son parecidas las estructuras? ¿En donde se observan las mayores diferencias?

11. Para ver cómo se corresponde el grado de similitud estructural con el grado de similitud en secuencia podemos realizar un alineamiento de ambas secuencias guiado por el alineamiento estructural. Para esto, vaya a:

Tools → *Structure comparison* → “*Match->Align*”

Ahora, observando la estructura y el alineamiento responda:

- A. ¿Qué son las regiones marcadas en rosa en el alineamiento?
 - B. ¿Este alineamiento, identifica regiones que no alinean estructuralmente? ¿A qué se debe?
 - C. En la parte superior de la ventana del alineamiento de secuencia vaya a *Headers* y seleccione RMSD:ca
¿Qué regiones poseen mayor RMSD? ¿A qué elementos estructurales corresponden? Para responder esto, seleccione estas regiones con el mouse en el alineamiento y visualícelas en la estructura alineada
12. Para cuantificar el alineamiento de secuencia obtenido, podemos calcular el % de identidad de secuencia. Para ello, en la ventana del alineamiento de secuencias vaya a: *Info* → *Percent identity*. Seleccione una estructura en *Compare* y la otra estructura en *with*. En *Divide by* seleccione *longer sequence length*. Presiona en Ok.

¿Qué valor de identidad de secuencia obtiene? ¿Porque cree que difiere del reportado anteriormente? ¿Las sustituciones observadas en las secuencias son conservativas?

En base a los resultados obtenidos. ¿Intentaría obtener experimentalmente la estructura de la nueva proteína, o confiaría en el modelo?

Ejercicio 2.

Usted es un famoso ecólogo que desde siempre sintió un especial interés por las ranas. Durante un viaje de campaña se encontró con unas ranas muy inusuales que poseían una fascinante coloración azul. Luego de años de investigación y muchos subsidios invertidos, su becario descubrió que esta coloración se debe a la existencia de una proteína en la linfa de las ranas que es capaz de conjugar biliverdina. Luego de aislar la proteína, obtiene su secuencia:

```
>Hypsiboas_punctatus_BP
MRVLLILGVVVLSTLAFAHHEEGHHDEDLKDDHDPFLPEDHKKALFVYQKPALNNINFA
FKMYRQLARDHPTENIVISPVSISSALALLSLGAKGHTHSQIVERLGYNTSEIPEQQIHE
SFHKQLDVVDDKDRDLEFEHGNALFTCKEHKIHQTFLDDAKKFYHSEVIPTDFKNTEEAK
NQINSYVEKSTHGKITNILDSDVDQDAMIALINFIYLRANWQHPPFDEKLTKEGDFHVDKDT
TVKVPFMRRRGIYKMAYTDDIIMVTIPYNGSVEMFLAMTKMGKLSELEQNLNRERSLKWR
EIMQYQLIDLSPKLSVSGILNLKETLSKLGIVDVFSNHADLSGITDESHLKVSKAIHKA
MMSFDEHGTEAAPATAAEADPLMLPPHFKFDYPFIFRVQDLKTKNPLLVGRIANPQK
```

Utilizando la secuencia, el becario busca en las bases de datos y descubre que su proteína es homóloga a una superfamilia de proteínas conocidas como serpinas compartiendo un 43% de identidad de secuencia con la proteína de humanos.

Para entender las diferencias con la proteína de humanos, estuvo muy interesado en obtener la estructura tridimensional de la proteína de rana. Sin embargo, todos los intentos de cristalización fallaron rotundamente. Su subsidio se está terminando rápidamente pero afortunadamente, un becario muy interesado en bioinformática y el modelado por homología lo salva de su desesperación.

Utilizando la herramienta HHPred el becario encontró que el mejor template era: 3NE4, Chain A, correspondiente al inhibidor de tripsina humano (Alpha-1-antitrypsin, P01009). Utilizando herramientas de modelado desarrolla un modelo 3D y le asegura que el modelo es de muy buena calidad.

Desconfiando de los resultados de su becario, Ud. decide analizar la calidad del modelo obtenido. Para esto utiliza todas las herramientas que conoce:

1. Descargue el modelo creado por su becario (Hypsiboas_punctatus_BP.pdb). Utilice las herramientas aprendidas en el punto anterior (Verify3D) e investigue los resultados obtenidos. ¿Le parece que su becario estaba en lo cierto, o equivocado?

Para explicar las diferencias obtenidas analice las estructuras como se indica en los puntos siguientes (2, 3 y 4) usando Chimera y IUPred:

2. Utilizando el modelo generado y el PDB (3NE4) utilizado como molde realice un alineamiento estructural en Chimera. ¿Qué diferencias observa? ¿Tiene relación con lo obtenido por Verify3D?
3. En Chimera observe el alineamiento de secuencia ¿En qué regiones hay mayor número de Gaps? Calcule el RMSD. ¿En qué regiones se observan las mayores diferencias? ¿A qué estructura corresponde? ¿Por qué cree que ocurre esto?
4. Ingrese la secuencia de la rana en IUPred2A (<https://iupred2a.elte.hu/plot>). ¿Qué relación encuentra con lo obtenido por Verify3D?

En base a los resultados de su análisis, responda:

- ¿Pudo explicar todas las regiones de menor calidad reportadas por Verify3D?
- Ud. Se sacó un nuevo subsidio donde tiene plata para seguir haciendo estudios estructurales de esta proteína: le daría alguna indicación a su nuevo becario, para que tenga más suerte al intentar cristalizarla?

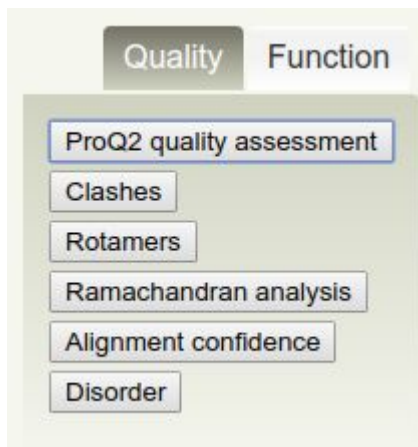
Ejercicios Adicionales

Ejercicio Adicional 1.

Después del exitoso resultado que obtuvo, ahora más relajado luego de haber salvado a la población porcina, decide rechequear los resultados obtenidos con la proteína misteriosa patogénica 1 utilizando otro programa que requiere más tiempo para modelar su proteína.

1. Ingrese la secuencia de la proteína en la web de Phyre2 (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>) Espere... Espere... Espere... (Puede llevar horas).
2. Observe los resultados obtenidos. ¿Cuál es el valor de “*Confidence*” del modelo? ¿Cuál es el coverage de la secuencia? ¿A qué se debe este valor?
3. Descargue el modelo (haciendo click sobre la representación) y trate de alinearlos en chimera con los otros modelos, ¿es muy diferente? Investigue las regiones con mayor valor de RMSD ¿A qué estructuras corresponde? ¿Qué concluye sobre las regiones más difíciles de modelar?
4. En la web de Phyre, vaya a la sección *Detailed template information*. Haga click en *Run Investigator*.

Explore en *Quality* las distintas opciones:



Al final de la página posee el alineamiento y puede ir estudiando estas características posición por posición.

¿Son muy diferentes los resultados obtenidos en comparación a los anteriores?