

EVALUAR MODELOS 3D

1) Elige una secuencia S de la superfamilia que elegiste para la tarea 3.

La secuencia elegida fue d1csb.1_Hsapiens.fasta que corresponde a una pro-cathepsin B de Humano.

2) Usando HHpred (<http://goo.gl/JQelUx>) selecciona al menos una estructura molde o template que puedas usar para modelar S, asegurándote que tiene menos del 90% de identidad si fuera posible.

El alineamiento elegido fue una “Cathepsin B-like peptidase (C01 family)” cuya identidad es del 58%. Se seleccionó esta porque es la que tiene la mayor identidad por debajo del 90%.

3) De acuerdo con el ejemplo de <http://goo.gl/ZZBEI9> y la documentación de MODELLER construye dos modelos M1 y M2 de S y comprueba su estima de calidad con DOPE.

Modelo 1:

DOPE score: -23843.14453

Modelo 2:

DOPE score: -23982.61133

El DOPE score es una medida que se relaciona con el α RMS (Root Mean Square).¹ Se ha propuesto una correlación en la que un valor negativo del DOPE score minimiza el valor de la distancia entre los carbonos alfa de las proteínas alineadas. En este caso, ambas tienen valores muy buenos, ya que sus números son muy negativos. Aún así, ambos valores son muy cercanos, por lo que podemos decir que los modelos son en gran medida muy similares.

4) Evalúa la calidad de los modelos M obtenidos comparándolos con la estructura conocida, que descargaste de SCOP en la tarea 3. Para ello puedes usar MAMMOTH. En tu informe por favor indica el alineamiento obtenido, el RMSD y al menos una imagen de su superposición para brevemente comentar las diferencias que observas entre cada modelo y la estructura experimental.

Alineamiento entre la secuencia de los modelos y la secuencia problema:

Este alineamiento se hizo por estructura en T-coffee (Espresso).

```
>Modelsequence _P_ 3ai8A
LPASFDAREQWPQCPTIKEIRDQGSCGSCWAFGAVEAISDRICHTNVSVEVSAEDLTCGSMCGDGCNGYPAEAWNFWTR
KGLVSGLYESHVGCPRYSIPCEHVNGSRPCTGEGDTPKCSKICEPGYSPTYKQDKHYGYNYSVSNSEKDIMAIEYKNGP
VEGAFSVYSDFLYKSGVYQHVTGEMGHAIRILGWGVENGTPYWLANSWNTDWDNGFKILRGQDHCIESEVAGIPR--
```

```
>g1csb.1 d.3.1.1 (A.;B.) (Pro)cathepsin B {Human (Homo sapiens) [TaxId: 9606]} _P_ 3k9mB
LPASFDAREQWPQCPTIKEIRDQGSCGSCWAFGAVEAISDRICHTNVSVEVSAEDLTCGSMCGDGCNGYPAEAWNFWTR
KGLVSGLYESHVGCPRYSIPCEHVNGSRPCTGE GDTPKCSKICEPGYSPTYKQDKHYGYNYSVSNSEKDIMAIEYKNGP
VEGAFSVYSDFLYKSGVYQHVTGEMGHAIRILGWGVENGTPYWLANSWNTDWDNGFKILRGQDHCIESEVAGIPRTD
```

Los valores RMSD que se obtienen de los .pdb de los modelos se ve a continuación. El RMSD (**root-mean-square deviation**) nos indica un promedio de las distancias entre los carbonos alfa de ambas proteínas en su conformación espacial sobrepuesta. Es decir que una superposición perfecta nos daría un RMSD de 0 Angstrom. En este caso vemos valores de alrededor de 13. Esto se puede deber a que, como se ve en la imagen 3D existen zonas que se alejan mucho unas de otras, haciendo que este valor vaya creciendo.

Hay que tener en cuenta que estos modelos parten de secuencias con valores bajos de identidad (en este caso 58%) y por eso vemos un valor relativamente alto del RMSD.

RMSD:

modelo 1-----

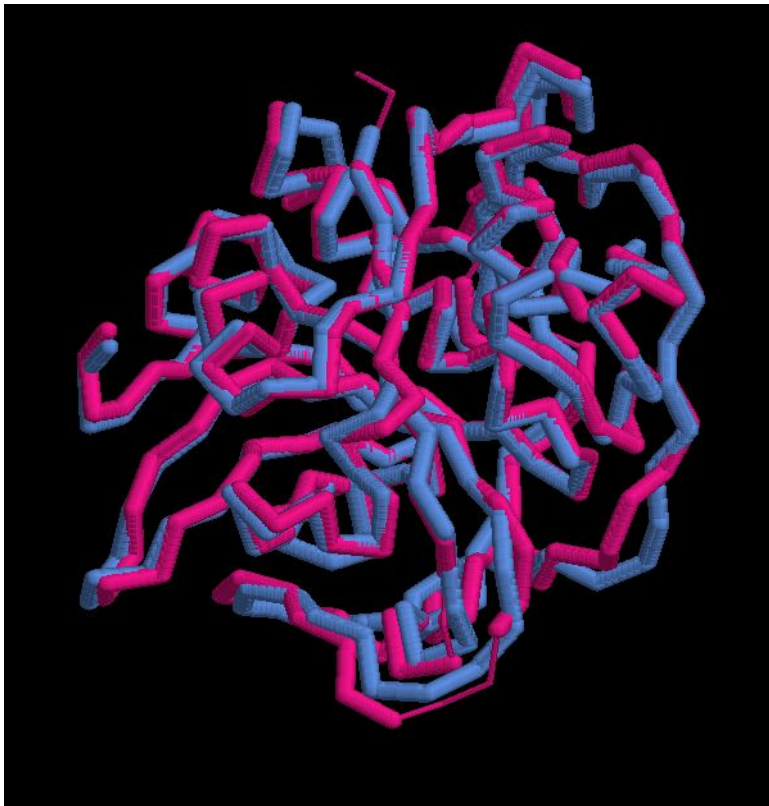
```
# Total residuos: pdb1 = 238; pdb2 = 252
# Total residuos alineados = 238
# Coordenadas originales = original.pdb
# superposicion optima- archivo PDB = align_fit.pdb
# RMSD = 13.64 Angstrom
```

modelo 2-----

```
# Total residuos: pdb1 = 238; pdb2 = 252
# Total residuos alineados = 238
# Coordenadas originales = original.pdb
# superposicion optima- archivo PDB = align_fit.pdb
# RMSD = 13.63 Angstrom
```

En la imagen se observan los gaps que se ven desde el alineamiento.

Predecir el plegamiento de una proteína no es tarea fácil pero uno se puede ayudar de algunas estructuras conocidas. Para ello se buscan secuencias similares a la secuencia en cuestión. Una vez obtenido un modelo con el cual se buscará crear un plegamiento para nuestra secuencia, es necesario ver que lo que estamos implementando sea significativo. En otras palabras, crear un modelo 3D debe de considerar no sólo la secuencia, sino propiedades que de ésta emergen para aproximar mejor su plegamiento. Es por eso que se analizan los valores del DOPE score y los RMSD. En conjunto, estos parámetros nos indican qué tan confiable es nuestra predicción. Pero también, es mejor considerar varios modelos para poder analizar sus diferentes propiedades. En este trabajo se usaron 2 modelos, pero un análisis más profundo requeriría más modelos para discernir.



Referencias:

1.- Shen, M., & Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. *Protein Science: A Publication of the Protein Society*, 15(11), 2507–2524.
<http://doi.org/10.1110/ps.062416606>