MODELIZACIÓN POR HOMOLOGÍA

CEFIRE 2011

Antes de empezar

El objetivo de este tutorial es proporcionar a los alumnos un esquema básico en la modelización por homología y la visualización de las estructuras obtenidas. Es importante destacar que el proceso descrito a continuación y las herramientas empleadas son sólo un ejemplo de entre la gran variedad de metodologías y herramientas que existen en la web.

Introducción

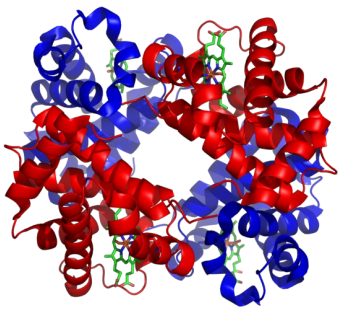
A la hora de estudiar una proteína, lo que más nos interesa es su función. Actualmente, sabemos que **hay una gran relación entre la estructura que adopta la proteína y la función que realiza**. Tener la estructura tridimensional (3D) una proteína nos puede ayudar a averiguar qué aminoácidos contribuyen a su estabilidad, por qué una región está más conservada o menos, qué aminoácidos forman la superficie de la proteína, cuáles son los directamente involucrados en la función e incluso los que interaccionan con otra molécula.

El principio que se esconde detrás de la modelización por homología es la suposición de que **la estructura está más conservada que la secuencia**. Los dominios funcionales de las proteínas suelen estar estructuralmente conservados y es, precisamente, esto lo que nos permite poder modelar una secuencia basándonos en las características estructurales de otras proteínas conocidas, a las que llamaremos *templates*.

Los pasos para la creación del modelo son:

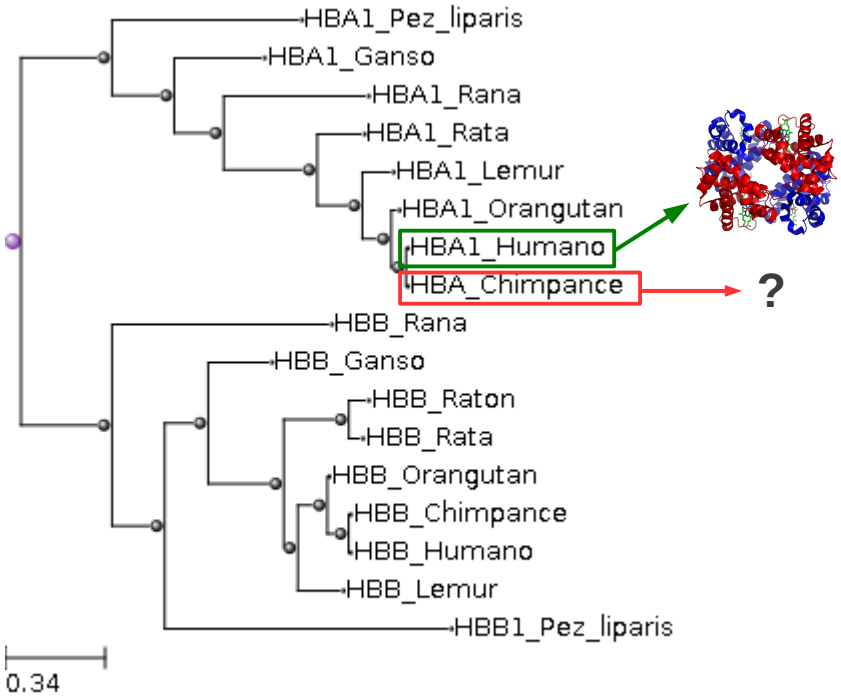
1. Búsqueda de los templates
2. Alineamiento de nuestra secuencia con los templates
3. Obtención del modelo
4. Evaluación del modelo

La Hemoglobina β



La hemoglobina (Hb) es la proteína encargada de transportar el oxígeno a los tejidos. Está formada por cuatro cadenas: α_1, α_2 (en rojo) y 2 cadenas β (en azul). A cada una de estas cadenas se une un grupo *hemo*, caracterizado por tener un átomo de hierro capaz de captar una molécula de oxigeno.

Al estudiar evolutivamente el gen de la Hb, observamos que por un proceso de duplicación génica el gen de la Hb dio lugar a la Hemoglobina α (HBA) y a la Hemoglobina β (HBB). Actualmente, conocemos la estructura tridimensional de la HBA1 para humano pero no para chimpancé (*Pan troglodites*).



Lo que vamos a hacer en este tutorial es crear la estructura tridimensional de la hemoglobina α (HBA1) de chimpancé basándonos en una estructura homóloga conocida, la humana.

1. Búsqueda de los templates

Para crear un modelo lo primero que debemos saber es la secuencia de la proteína que queremos a modelar y a la que desde ahora llamaremos *target*. **Uniprot** (http://www.uniprot.org/) es una de las bases de dados de proteínas más completas que existen actualmente. Vamos a dirigirnos ahí para extraer la secuencia de nuestra proteína.

Desde el cuadro de búsqueda vamos a especificarle a Uniprot que estamos interesados en la proteina "HBA1" de chimpancé. Escribe el nombre de la proteína y clica en "Advanced search" (Figura 1).

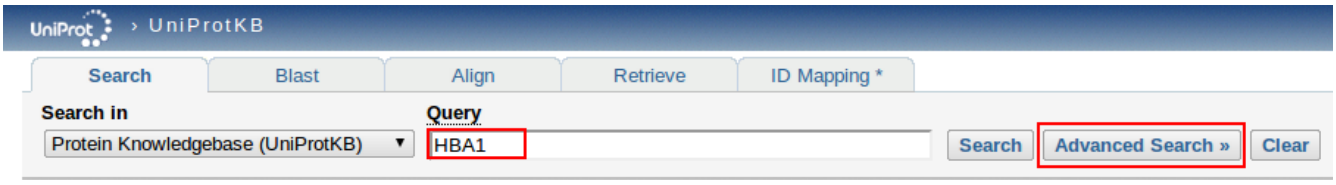


Figura 1: Uniprot

Se nos mostrará un nuevo cuadro de búsqueda. Ahí vamos a especificar que el organismo que queremos es el chimpancé. En el campo "Field" selecciona "Organism" y en "Term" escribimos "Chimpanzee". Finalmente, ejecutamos la búsqueda con "Add & search" (Figura 2).

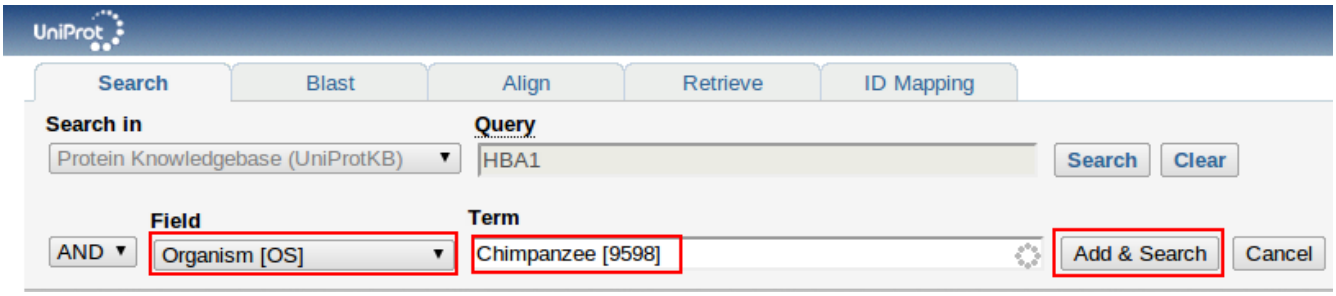


Figura 2: Búsqueda avanzada en UniProt

Verás que con esas especificaciones sólo observamos un resultado, es precisamente el que queremos. Haz click en el código de la proteína y, automáticamente, te mostrará todo lo que sabe Uniprot sobre ella.

Desde el apartado "Sequences" selecciona FASTA. Normalmente vamos a encontrar estas secuencias en formato FASTA, ya que es el formato más utilizado por los programas de análisis (Figura 3). Copia la secuencia que te muestra en un editor de texto y guárdala.

```
>sp|P21667|HBB_LEMVA Hemoglobin subunit beta OS=Lemur variegatus GN=HBB PE=1 SV=1
TFLTPEENNHVTSLWGKVNVEKVGGEALGRLLVVYPWTQRFFESFGDLSSPDAIMGNPKV
KAHGKKVLTAFSEGLHHLDNLKGTFAQLSELHCDKLHVDPQNFKLLGNVLVIVLAHHFGN
DFSPQTQAAFQKVVTGVANALAHKYH
```

Figura 3: Formato FASTA

Una vez obtenida la secuencia, vamos a buscar los *templates*. Para poder crear nuestro modelo tenemos que encontrar proteínas homólogas a nuestro *target,* cuya estructura sea conocida. Para eso vamos a utilizar **BLAST** (http://blast.ncbi.nlm.nih.gov/), una aplicación web que nos va a ayudar a encontrar secuencias homólogas a nuestro *target*.

En la página web seleccionamos "protein blast" (Figura 4).

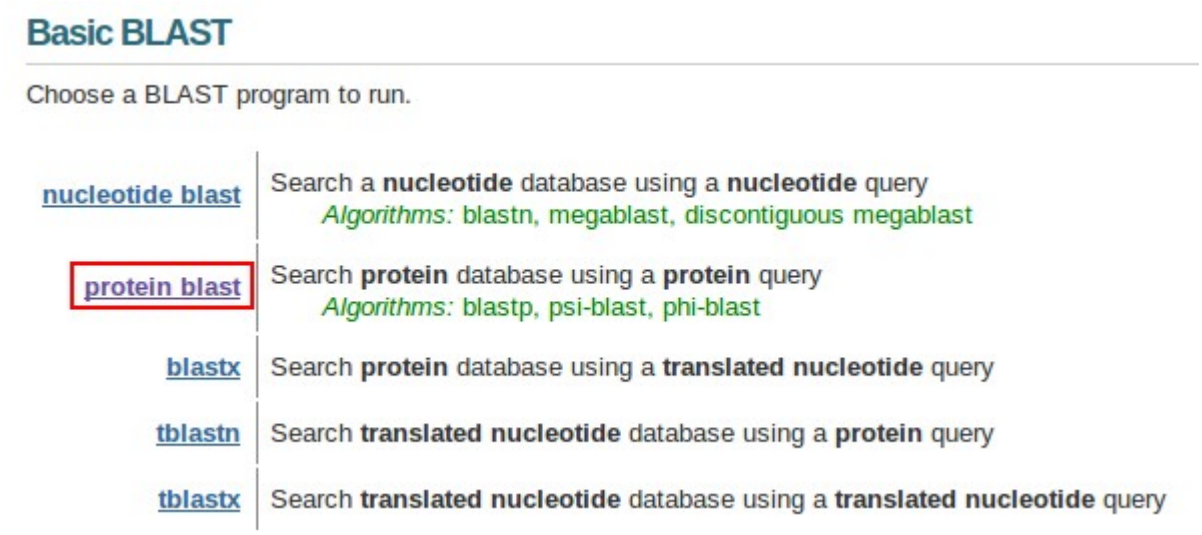
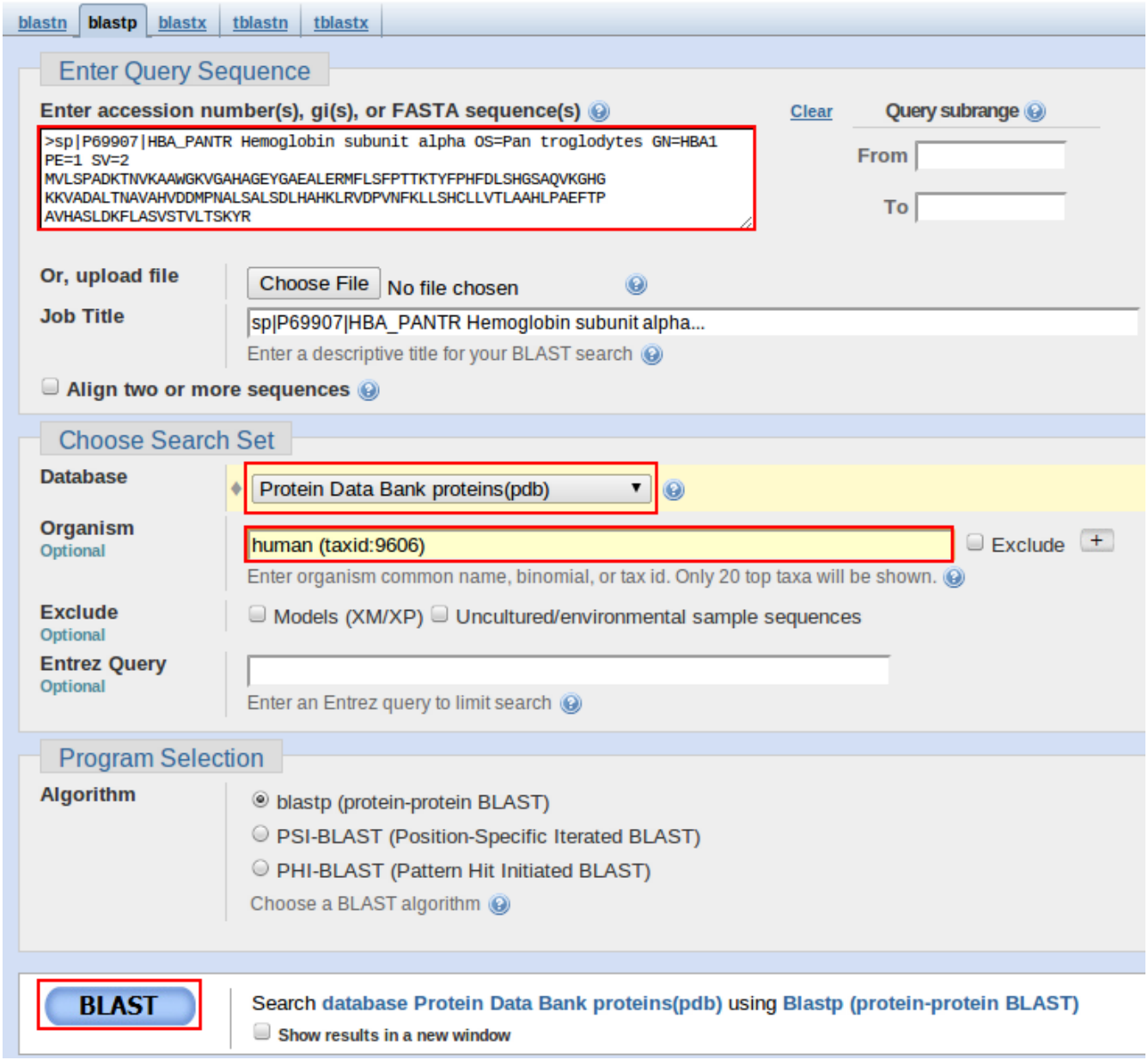


Figura 4: BLASTp

A continuación, pegamos la secuencia que hemos conseguido anteriormente en el cuadro indicado (Figura 3). Alternativamente, podemos cargar el archivo con la secuencia desde el botón “choose file”. En “Database” selecciona “Protein Data Bank proteins (pdb)”, en organism especificamos “human”, ya que sabemos que existen estructuras conocidas para humano de la HBA1, finalmente, cliclamos en “BLAST”.

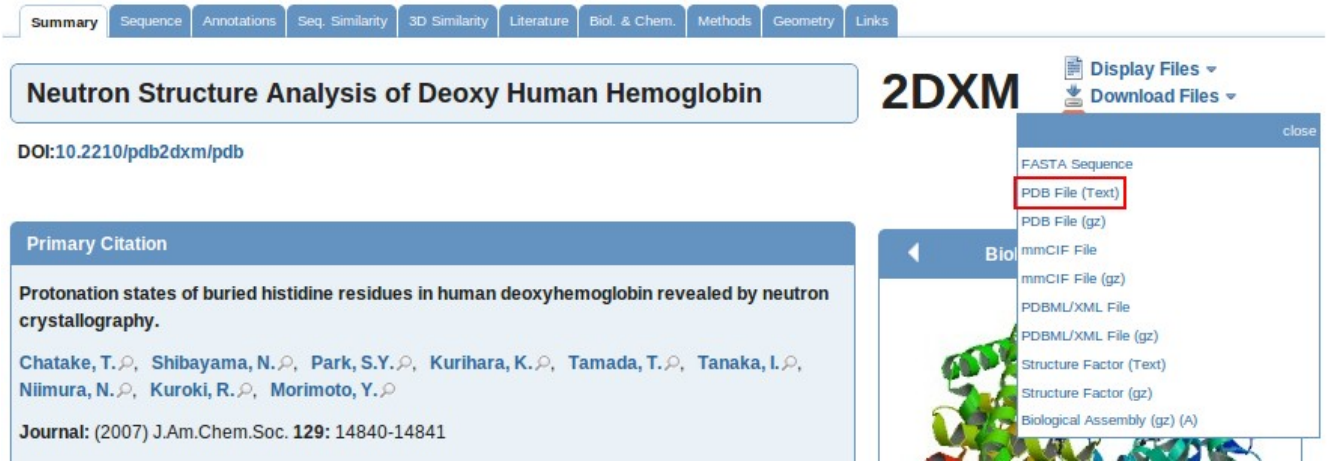


Lo que va a hacer el programa mientras esperamos es alinear nuestra secuencia contra todas las secuencias de humano e identificar aquellas que sean homólogas y que además tengan una estructura conocida y almacenada en PDB.

La página de resultados muestra una representación gráfica del alineamiento. Una vez en la página de resultados, vamos escoger el *template* que queremos utilizar. No existe una

regla establecida acerca del número de estructuras que debemos utilizar en el modelado. Utilizar varias estructuras puede ser ventajoso, pero utilizar demasiadas puede causar problemas.

Para este tutorial, hemos escogido **2DXM** como *template* para nuestro modelo. Vamos a dirigirnos a PDB (http://www.pdb.org), vamos a buscar esta estructura en la base de datos y a descargarnos el archivo que nos permitirá verla en 3D. La figura 4 muestra cómo realizar la descarga de 2DXM desde PDB, selecciona "Download files" y escoge "PDB file".



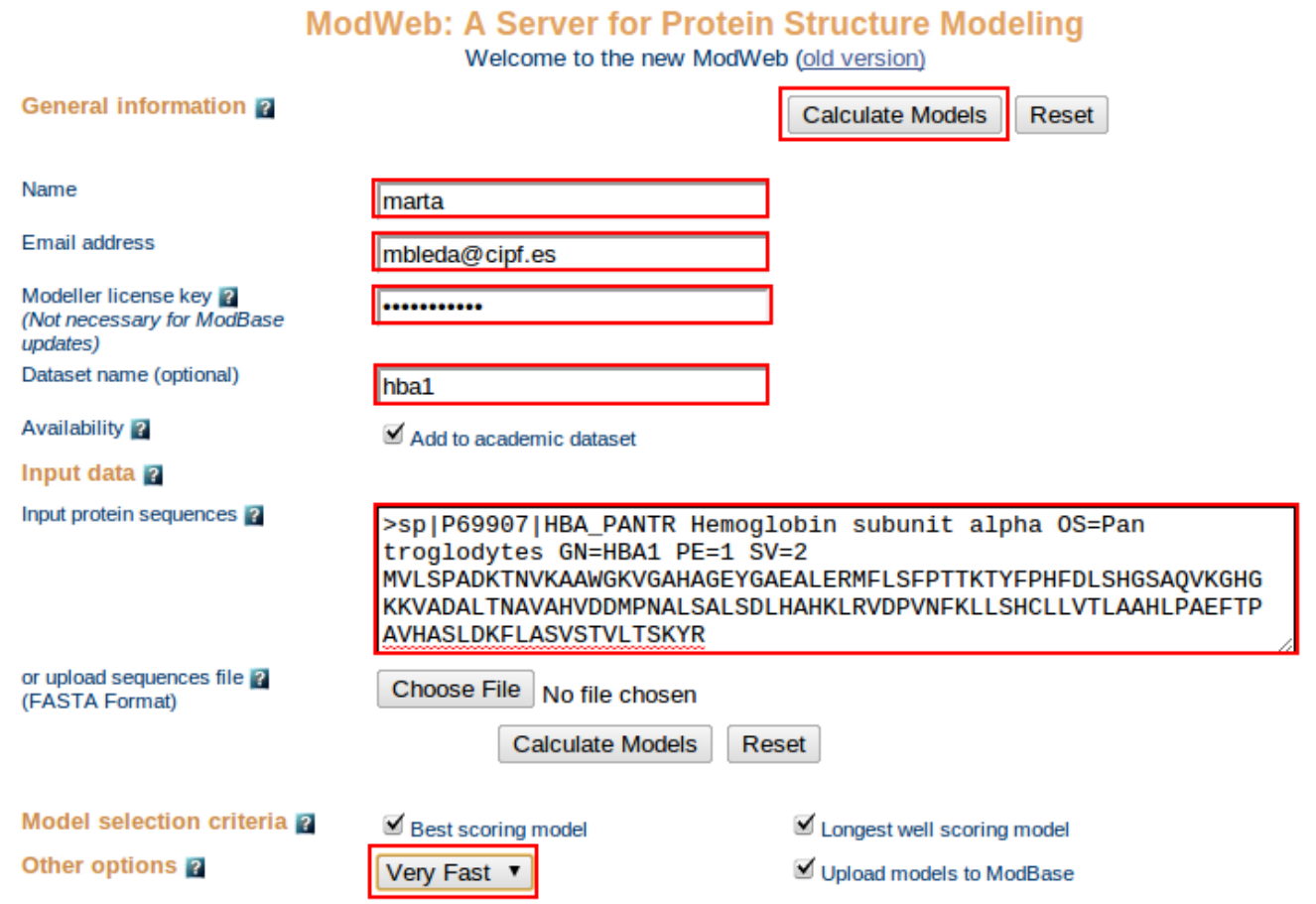
2. Obtención del modelo con ModWeb

Llegados a este punto, con nuestro *template* alineado a la secuencia de interés, teóricamente deberíamos proceder a la construcción del modelo. Lamentablemente, los programas que existen para ellos son demasiado sofisticados para verlos en un pequeño ejemplo como este. Afortunadamente, los creadores de esas complicadas herramientas han creado herramientas web que nos van a facilitar extremadamente este trabajo.

Este es el caso de ModWeb (http://salilab.org/modweb). Esta herramienta web nos permitirá crear el modelo con sólo indicarle la secuencia de nuestra proteína problema. Así que vamos a acceder a su web y a introducir estos datos (figura 5).

Vamos a indicarle algunos detalles de contacto, indica tu nombre y correo electrónico. No te asustes, no te van a enviar publicidad a tu cuenta de correo, ModWeb únicamente utilizará tu correo para avisarte cuando haya acabado de contruir tu modelo.

ModWeb también nos va a pedir una contraseña "Modeller license key". En este campo debemos escribir "MODELIRANJE". Opcionalmente, podéis dar un nombre a vuestro trabajo en "Dataset name". A contuación, copiamos y pegamos la secuencia de la proteina de chimpancé y seleccionamos "Very fast" en "Other options". Lanzamos el trabajo con "Calculate models" y a esperar.



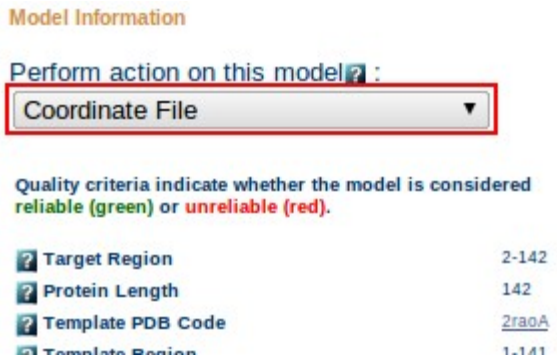
Seleccionando la opción “Very fast”, la obtención del resultado puede tardar al rededor de unos 30-45 minutos.

Podéis encontrar el resultado aquí: http://modbase.compbio.ucsf.edu/modbase-cgi/model_details.cgi?queryfile=1315557630_3892&searchmode=default&displaymode=moddetail&seq_id=6077c452d1dc6151040b2b179e2294c7MVLSSKYR

Target Region	2-142
Protein Length	142
Template PDB Code	1c7cA
Template Region	143-283
Sequence Identity	100.00%
E-Value	0
GA341	1.00
MPQS	2.24516
z-DOPE	-1.6
TSVMod Method	MTALL
TSVMod RMSD	0.491
TSVMod NO35	1
Dataset	MW-hba1_cefire
ModPipe Version	SVN:r1340:1348M
Model Date	2011-09-08

Cuando haya acabado, ves a la página de resultados, echa un vistazo a lo que ModWeb ha hecho. ¿Crees que es un buen alineamiento?

Desde la página de resultados, descárgate el modelo en formato “pdb”, para ello, en “Perform action on this model” selecciona “Coordinate File”. Te redirigirá a una nueva página. Selecciónalo todo y guárdalo en un archivo de texto.



3. Visualización y superposición del template y el modelo con PyMOL

Una vez descargado el modelo generado por ModWeb, vamos a visualizar, por fin, qué aspecto tiene nuestra secuencia en 3D. Para ello, vamos a utilizar PyMOL, un visualizador molecular.

Desde el menú superior de Ubuntu, selecciona: “Applications” > “Science” > “PyMOL Molecular Graphics System”. Verás que se abre el visualizador. Vamos a visualizar, en primer lugar, nuestro el modelo que hemos descargado para humano, para eso, desde PyMOL seleccionamos “File” > “Open...” e indicamos el archivo “2DXM.pdb”.

Vamos a jugar con ella...