

## MODELADO POR HOMOLOGÍA MODELLER

### INTRODUCCIÓN

En la actualidad contamos con una gran cantidad de secuencias provenientes de los proyectos genomas. Sin embargo, las metodologías experimentales para obtener cristales que redunden en estructuras de proteínas de alta confiabilidad siguen siendo complejas. Este escenario hace que el modelado molecular constituya una herramienta muy útil para lograr dilucidar estructuras de proteínas no cristalizadas.

Existen distintas formas de obtener modelos de estructuras, las dos metodologías más utilizadas son el modelado ab-initio y el modelado comparativo. Dentro de este último grupo,

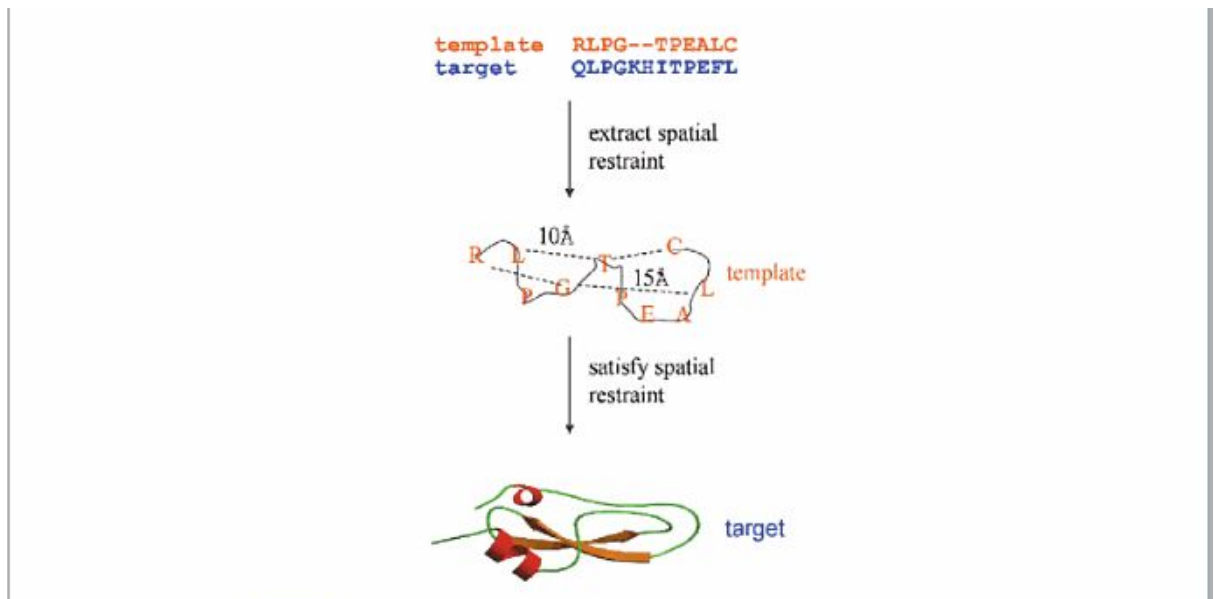
existen el modelado comparativo por homología y el modelado comparativo por enhebrado de proteínas. En este curso aprenderemos a construir modelos utilizando el método comparativo por homología.

Durante este trabajo práctico vamos a aprender a generar modelos partiendo de un único templado, lo que llamaremos 'modelado simple'. Trabajaremos sobre el sistema de albúmina de rata y como molde la salida del tp anterior.

### ACTIVIDAD

- En esta actividad vamos a trabajar con las secuencias y \*.pdb. Construiremos un alineamiento de la secuencia blanco con la del \*.pdb templado. En un primer paso vamos a convertir las secuencias en formato FASTA, a formato PIR para luego poder alinearlas, utilizando el servidor ([https://www.hiv.lanl.gov/content/sequence/FORMAT\\_CONVERSION/form.html](https://www.hiv.lanl.gov/content/sequence/FORMAT_CONVERSION/form.html)). Se puede utilizar los archivos ya dispuestos en la carpeta de ejemplo.
- Lo siguiente es construir un alineamiento entre la secuencia de interés y la secuencia del templado extraída del \*.pdb. MODELLER incluye un comando que permite alinear secuencias teniendo en cuenta la estructura secundaria presente en el \*.pdb. Pasar el \*.pdb al directorio de trabajo, junto con un alineamiento secuencial \*\_aln y construya el alineamiento con el script "alineamientotempladoseq.py". En caso de utilizar las salidas del tp 1 de modelado, puede recurrir al alineamiento secuencial entre el target y el molde construido anteriormente.  
¿Cuáles son los archivos de salida? ¿Qué información posee cada uno?
- En esta actividad construiremos modelos a partir del \*.pdb y el alineamiento generado en el paso anterior, utilizando el script "generador\_modelos\_rat.py". Este script utiliza la función automodel de MODELLER para generar 100 modelos a partir de un alineamiento, donde determinamos qué secuencia es la conocida ('known') y cuál la que vamos a modelar ('sequence'), y qué parámetros de energía queremos que MODELLER calcule para la futura selección del mejor modelo. Se obtiene un modelo 3D mediante la optimización de una función de densidad de probabilidad

molecular, o por satisfacción de restricciones espaciales.



Las distancias entre residuos tomadas del template se complementan con datos estereoquímicos, distancias de enlaces, ángulos dihedros, distancias interacciones débiles. Muchos de estos valores pueden tener valores maximos y minimos ya que se pueden derivar de varias estructuras similares al template. Todos estos datos se combinan para dar una “objective function” . El modelado consiste en optimizar esta objective function. El mejor modelo es aquél que viola el menor número de restraints (“satisfaction of the spatial restraints”) o Discrete Optimized Protein Energy (DOPE).

¿Cuáles son los archivos de salida? ¿Qué datos aportan?

- Evaluación del mejor modelo. En esta actividad seleccionaremos el mejor modelo, de entre los generados en la actividad anterior.
  1. Realizar una tabla de salida para comparar los parámetros molpdf, DOPE, GA341. Seleccionar el mejor modelo. Busque información sobre los distintos parámetros y explique brevemente de qué se trata cada uno.
  2. Para discutir: ¿Es necesario un alto número de modelos generados?
  3. También podemos usar servidores online que no permiten obtener parámetros sobre los modelos, por ejemplo:  
<https://prosa.services.came.sbg.ac.at/prosa.php>

¿Cómo interpreta los resultados? ¿Cómo organizaría los resultados para poder automatizar este análisis?