

# Química Computacional

## Taller 2

Gian Pietro Miscione y Sebastián Franco Ulloa

Módulo de Mecánica Molecular

Mayo 2, 2017

QUIM-3516

Este taller tiene como objetivo familiarizar al estudiante con *PrepWizard*, *LigPrep* y *Glide*, las herramientas distribuidas por Schrödinger para preparar un receptor, ligandos y hacer *docking*, respectivamente. Los estudiantes deberán manipular varias de las opciones ofrecidas por los programas y así entender a fondo su funcionamiento. A lo largo del taller, los estudiantes estudiarán un sistema biológico de una ADN girasa la cual está involucrada en el proceso de replicación del ADN bacteriano, por lo que su inhibición puede resultar en la muerte del microorganismo. Por ende, es un receptor llamativo para el diseño de medicamentos.

## 1 Preparación del receptor

El primer paso para un cálculo de *docking* es la preparación del receptor, el cual, al venir de la base de datos PDB, contiene varios problemas discutidos en clase con anterioridad.

- Para descargar el archivo *\*.pdb* correspondiente al receptor de interés, se debe ir a la página principal de la base de datos PDB (<http://www.rcsb.org/pdb/home/home.do>) y en el motor de búsqueda escribir “DNA gyrase”.
- Para filtrar más los resultados, en el costado izquierdo puede seleccionar en la sección de **ORGANISM**, la bacteria *Staphylococcus Aureus*.
- Buscar y seleccionar la entrada con código **2XCS**. En esta página puede encontrar toda la información pertinente del cristal.
- En el recuadro **Download Files**, seleccionar **PDB Format** para descargar el archivo *2xcs.pdb*.
- Abrir con *gedit* el archivo *2xcs.pdb* e identificar los residuos y átomos faltantes al igual que los residuos con más de una conformación (Ayuda: Vaya a la línea 1088).
- Ir a Maestro y, al igual que en el taller anterior, importar la estructura que se descargó. Aunque de la configuración y la versión de Maestro, en verde se debería estar mostrando nuestro ligando de interés.
- Jugar con la visualización de la proteína con las opciones del recuadro **Style** localizado en la parte superior derecha de Maestro. Debe identificar cuántas cadenas, moléculas, residuos, átomos y cargas tiene el sistema. También debe localizar la doble hebra de ADN y la posición del ligando.
- Seleccionar en la esquina superior izquierda la opción **Protein Preparation Wizard**. Esta es la interfaz para la preparación del receptor.
- Nótese que en el recuadro **Import structure into Workspace** se puede descargar el archivo *\*.pdb* escribiendo el código en el recuadro de texto. En este caso esto no es necesario.

- En el recuadro **Preprocess the Workspace structure** deje todas las opciones predeterminada y además seleccione las opciones “Fill in missing side chains using Prime”, “Fill in missing loops using Prime” y en “Delete waters beyond # Åfrom het groups” escriba 0.0. En este taller no se considerarán aguas por simplicidad. Haga click en **Preprocess**.
- Luego puede ver, en **View Problems**, los problemas que tuvo Maestro al preparar el receptor. Por ejemplo, átomos faltantes, átomos a los que no se pudo asignar un *atom type* y residuos con conformaciones alternas (¿hay algún residuo cerca al ligando? ¿Las conformaciones diferentes corresponden a interacciones químicas diferentes?).
- En el recuadro **Review and Modify** puede ver las cadenas, moléculas no convencionales (*het*), aguas, etc,...
- En el recuadro **Refine**, se minimizan los hidrógenos los cuales se pudieron haber puesto mal en los pasos anteriores.
- Seleccione la opción “Minimize Hydrogens of Altered Species” y luego haga click en **Optimize**.
- Podemos ignorar la sección **Remove waters**, pues ya las eliminamos todas
- En la última sección **Restrained minimization**, coloque un valor de RMSD igual a 0.30 y haga click en **Minimize**. Esta minimización va a relajar el sistema evitando que hayan átomos muy cercanos entre sí.
- Note que en cada uno de los pasos descritos anteriormente se crea una nueva entrada que se importa automáticamente a Maestro, entonces, la última entrada corresponde al receptor preparado. Para no confundirlas es útil cambiarles el nombre.
- Una buena idea para tener una noción de las interacciones presentes entre el ligando nativo y la proteína es hacer un mapa de interacciones. Maestro es capaz de hacer esto por nosotros. Seleccione la estructura preparada y seleccione en la esquina superior izquierda la opción **Ligand Interaction Diagram**.
- Para *Glide* en particular no es necesario eliminar el ligando nativo del cristal para hacer el *docking*. El mismo entiende quién es el ligando nativo y lo ignora.
- Como estamos ignorando el problema de las conformaciones alternas, Maestro, por defecto, va a usar las conformaciones ‘A’.

## 2 Generación de la cuadrícula

Procederemos con la generación de la cuadrícula la cual se hace con la herramienta *Glide* (la misma que hará el *docking*. Es decir, un único archivo que contendrá toda la información del receptor necesaria para hacer el *docking*. Esto incluye geometría, conectividad, potenciales, restricciones, etc.

- Seleccione e incluya la estructura del receptor preparado.
- Busque en *tasks* la opción “Receptor Grid Generation” y seleccione la herramienta que hace mención a *Glide*.
- Para excluir el ligando nativo del docking se le debe decir a *Glide*. Para esto, en el recuadro **receptor**, seleccione “Pick to identify the ligand” y seleccione un átomo del ligando nativo.
- Seleccione la opción “Use input partial charges” para que use las cargas de los *atom types* que se asignaron durante la preparación del receptor. Si no se hiciera esto, *Glide* volvería a asignar *atom types*.

- En la parte inferior del recuadro de **Advanced settings** se puede permitir que los átomos del receptor formen puente H con hidrógenos aromáticos, halógenos como receptores y halógenos como aceptores. En este caso ninguna de las opciones es particularmente importante.
- En el recuadro **Site** se deben especificar los parámetros de la caja interna y externa. Ya que tenemos un ligando cocrystalizado la mejor opción es centrar las cajas en el “Centroid of Workspace ligand”. En el recuadro de **Advanced setting** se pueden personalizar las dimensiones de las cajas pero mantendremos las predeterminadas.
- En el recuadro de **Constraints** vamos a darle la posibilidad a *Glide* de fijar algunas restricciones (esto no lo obligará en el paso del *docking*). Vaya a la pestaña de “H-bond/Metal” y luego seleccione los 2 átomos de oxígeno (del grupo carboxilato) presentes en los 2 residuos Asp83 (Ayuda: están interactuando con el ligando nativo).
- En este ejemplo no vamos a darle libre rotación a ninguno de los dihedros que nos ofrece *Glide* ni excluirémos ningún volumen. No obstante esto se puede hacer en las pestañas de **Rotatable Groups** y **Excluded Volumes**. OJO: No olvide que imponer restricciones al *docking* puede resultar en el sesgo de los resultados.
- Haga click en **Run**. Este cálculo debe tomar aproximadamente 3 minutos y resulta en un único archivo \*.zip.

### 3 Preparación de un ligando

Hasta este punto terminamos de lidiar con el receptor, pero aun falta preparar el ligando, lo cual se hará con la herramienta *LigPrep* de Schrödinger. En este caso haremos el *docking* del ligando nativo y compararemos la pose predicha con la original del cristal.

- Seleccione e incluya en el espacio de trabajo la estructura “2xcs” NO preparada. Haga un duplicado de la estructura haciendo *Click derecho* → *Duplicate* → *In place*.
- Ahora seleccione solo el ligando haciendo click en la letra **L** presente en la parte superior derecha de Maestro en el recuadro “Quick Select”. Luego seleccione la opción “Invert” y suprima para borrar todo excepto el ligando.
- Busque en *Tasks* la herramienta *LigPrep*. Esto debería abrir un recuadro con una única pestaña.
- Lo primero que se debe especificar es el (o los) ligando a preparar. El ligando a preparar puede estar en un archivo que no se haya importado a Maestro, o puede estar presente en el proyecto actual. Por simplicidad, haremos lo segundo.
- Asegurese de que el ligando esté incluido en el área de trabajo y en la opción “Use structures from:” selecciones “Workspace (# included entries).”
- Al hacer *docking* de un solo compuesto, no es necesario hacer un filtro, por lo que nos saltaremos esta opción.
- En este ejercicio generaremos todos los posibles estados de ionización en un pH de  $7.0 \pm 2.0$  usando Epik (otra distribución de Schrödinger).
- En este ejemplo no es necesario pero es buena costumbre siempre seleccionar la opción de “Desalt”.
- Seleccione la opción de “Generate Tautomers”.
- Con respecto a la estereoquímica, generaremos todas las posibles combinaciones (lo cual no es común), estableciendo el máximo de estructuras a 32.

- Este cálculo generará un único archivo con todas las moléculas ya listas para hacerles el *docking*. Este archivo puede ser *\*.maegz* o *\*.sdf*. Ambos formatos pueden almacenar la misma información pero el primero es un formato binario que solo puede leer Maestro mientras que el segundo corresponde a un archivo de texto que se puede abrir con la mayoría de programas en química computacional. Por estas razones se acostumbra a exportarlo como *\*.sdf*.
- Haga click en **Run**. El resultado, aunque corresponda a un único compuesto, contiene varias moléculas. Analice las diferencias.

## 4 Docking

Una vez se han preparado el receptor y el ligando, podemos proceder a hacer la inserción de uno en el otro. Esto se hará con la misma herramienta con la que se generó la cuadrícula: *Glide*.

- Para hacer el *docking* debemos abrir la ventana correspondiente para lo cual debe buscar en *Tasks* la herramienta *Ligand Docking*. Al hacer click debe abrir una ventana con 6 pestañas.
- Lo primero que se debe especificar es la cuadrícula que se usará para hacer el *docking*. Esto se hace en la parte superior de la ventana donde dice **Receptor Grid**. Busque en el navegador el archivo *\*.zip* generado en la sección 6.
- En la pestaña **Ligands** se deben elegir las moléculas que se insertarán en el receptor. Estos ligandos se pueden especificar como un archivo, como entradas seleccionadas y como entradas incluidas. Seleccione la opción de su preferencia pero sea consistente. e.g. si elige “Workspace (Included Entries)” asegúrese de que todas las moléculas se muestren en el espacio de trabajo.
- Dependiendo de los recursos computacionales disponibles, es posible que se quiera no evaluar todos los ligandos disponibles sino solo los primeros *N*. Seleccionar la casilla *End* indica que se evaluarán todas las moléculas del archivo.
- Seleccione la casilla *Use input partial charges* por razones discutidas anteriormente. Mantenga el resto de opciones en esta pestaña con sus valores predeterminados.
- En la pestaña de **Setting** se debe especificar todo lo concerniente al *docking*. La casilla *Precision* es donde se especifica la función de *scoring* a usar. En este caso usaremos *SP*.
- En la casilla *Ligand Sampling* se especifica si se desea hacer *docking* con el ligando flexible o rígido. En este caso lo dejaremos flexible.
- El resto de opciones las dejaremos en sus valores predeterminados pero asegúrese de marcar con un chulo las opciones “Add Epik state penalties to docking score”, “Reward intramolecular H-bonds” y “Enhance planarity of conjugated pi groups”
- En la parte inferior del recuadro de **Advanced Settings** se puede especificar si se desea que los átomos de los ligandos puedan participar en puentes H con hidrógenos aromáticos, halógenos como aceptores o halógenos como donores.
- En la pestaña de **Constraints** se mostrarán las restricciones que se definieron durante la generación de la cuadrícula. Usted deberá hacer el cálculo de docking 2 veces. La primera vez sin ninguna restricción y la segunda vez exigiendo un puente de hidrógeno con al menos uno de los átomos de oxígeno seleccionados en la sección 6. Compare los resultados.
- En la pestaña **Output** se especifican los aspectos técnicos de lo que retornará el cálculo. En este caso mantendremos casi todas las opciones con sus valores predeterminados.
- En la opción “Write at most: # poses per ligand”. Cambie el número a 3. Esto hará que el programa imprima las 3 mejores poses del ligando en el receptor (para cada molécula).

- Seleccione la opción “Write per-residue interaction scores” y pídalo para residuos hasta 5 Å del centro de la cuadrícula. Con esto *Glide* imprimirá la energía de interacción con cada uno de los residuos que satisfaga esta condición.
- Haga click en *Run*. El archivo más importante que se genera con este tipo de cálculos es un archivo *\*\_pv.maegz*.
- Importe (si no se hizo automáticamente) el archivo mencionado anteriormente. La primera entrada corresponde al receptor y las demás son las moléculas en sus respectivas poses.
- Para mantener el receptor fijo dele *click derecho* → *Show in Workspace* → *Fix within project*. Luego para moverse entre las moléculas, solo inclúyalas en el espacio de trabajo. Así podrá ver más rápidamente todos los complejos predichos.
- Para ver propiedades como las interacciones por residuos, puede abrir la tabla del proyecto (símbolo # en la parte superior derecha de Maestro). Ahí se mostrarán todas las propiedades de todas las entradas del proyecto. Dentro de estas propiedades se encuentran dos tituladas “Glide Score” y “Docking score”. ¿Cuál es la diferencia entre estas dos columnas? En la tabla también encontrará las contribuciones energéticas de cada uno residuo de los residuos especificados.

## 5 Búsqueda conformacional

Es posible que una molécula tenga una muy buena energía libre de enlace ( $\Delta G_{bind}$ ) pero que no sea un buen inhibidor. Esto puede ocurrir cuando el ligando requiere de demasiada energía para llegar a la pose en la que interactúa propiamente con el receptor. En otras palabras, si la energía del mínimo global es mucho más baja que la energía de la molécula en la pose de inhibición, su actividad puede ser menor que la predicha. Para verificar esto haremos los mismo cálculos que se realizaron en el taller anterior.

- Identifique la molécula con mejor energía libre de unión.
- Realice un cálculo de *Single-Point-Energy* con los mismos parámetros especificados en el taller anterior, de la molécula del inciso anterior en la pose predicha.
- Identifique la energía potencial interna en la tabla del proyecto.
- Ahora compararemos esta energía con la energía del mínimo global. Para esto vaya a *Tasks* → *Conformational Search (Macromodel)*
- Deje casi todas las opciones con sus valores predeterminados.
- En la pestaña de **CSearch** se especifica el algoritmo con el que se hará la búsqueda conformacional. En la opción *Method* seleccione “Torsional Sampling (MCMM)” este es el método explicado en clase.
- Asegúrese de seleccionar e incluir únicamente la molécula cuya búsqueda se desea hacer.
- en la opción *Number of structures to save for each search*, el 0 significa guardar todas las estructuras no redundantes. Esto puede resultar en cientos de geometrías para una misma molécula. Cambie este número a 1 para que el programa solo imprima una molécula (el mínimo global).
- Haga click en **Run**.
- Compare las geometrías y los valores de energía potencial interna cuando la molécula se encuentra en su estado ligado versus cuando está en su mínimo global.

## 6 Virtual Screening

En los cálculos anteriores se estudió una única molécula. Si tuvieramos interés en encontrar nuevos compuestos potencialmente activos sería necesario estudiar una mayor variedad de compuestos. En esta sección usaremos la cuadrícula generada en el paso pero haremos el *docking* de compuestos con estructuras disponibles en línea pero cuyas actividades desconocemos.

- Los ligandos los descargaremos de la base de datos ZINC (<http://zinc.docking.org/>). Vaya a la base de datos y haga click en las siguientes opciones:

*Subsets → Catalog → Biosynth → Downloads → All (SDF, Mid pH 6-8)*

Así debe descargarse \*.sdf.

- Importe el archivo que acaba de descargar en Maestro (Maestro admite formatos comprimidos). En total debe haber abierto 1609 compuestos.
- Por motivos de tiempo, para disminuir el número de ligandos, quédese con los primeros 200 que se importaron y elimine los demás.
- Prepare los ligandos similar a como se hizo en la sección 3. Esta vez filtre los compuestos para quedarse únicamente con aquellos que tengan un peso molecular entre 250 y 500g/mol y que tengan una carga total entre  $\{-2, 2\}$ . Además seleccione la opción “Retain specified chiralities (vary other chiral centers)”.
- Haga el *docking* de la misma manera en la que se hizo en la sección 4.
- ¿Cuál es el compuesto más tentativo para decir que va a ser un buen inhibidor? ¿Qué le dice su instinto químico? ¿Se ven las interacciones que se ven con el ligando nativo?

## 7 Re-scoring con MM-GBSA

Es posible que un ligando parezca que va a ser un muy buen inhibidor con base en la función de *scoring* *GlideSP*, pero que en verdad no lo sea. Esto puede ocurrir por varias razones, una de ellas es que esta función no es la mejor de todas y está sujeta mostrarnos falsos positivos. Una forma de reducir esta incertidumbre es usando una función de *scoring* más exigente como lo es *MM-GBSA*. Esta función usa un campo de fuerza más exacto que la familia OPLS-AA pero tarda más en evaluarse. Para usar esta función se debe tener acceso a la instancia *Prime* de Schrödinger.

- Identifique las 10 moléculas con mejor energía libre de enlace ( $\Delta G_{bind}$ ).
- Incluya el receptor en el espacio de trabajo y seleccione los 10 ligandos a estudiar.
- En el recuadro de *Tasks* busque la herramienta *MM-GBSA (Prime)*.
- Especifique que el receptor será aquel incluido en el espacio de trabajo y que los ligandos serán las moléculas seleccionadas. Esto también se puede hacer con el archivo \*\_pv.maegz que nos otorga *Glide* tras el cálculo *docking* pero en este caso ese archivo contiene todos los ligandos mientras que solo nos interesa estudiar 10.
- Asegúrese de que el modelo de solvatación esté fijado en “VSGB”. Ya que no vamos a considerar la flexibilidad del receptor, podemos ignorar la última sección del recuadro.
- Haga click en **Run**.
- El resultado de este cálculo es un archivo \*\_maegz que tiene los nuevos complejos. Durante el cálculo la geometría de los ligandos y el receptor cambiaron porque se minimizan en un de los pasos. Esto no se debe confundir con la exploración del espacio conformacional. Lo que hizo *MM-GBSA* solo fue una minimización, no se le dio flexibilidad a la proteína en ningún momento.

- ¿Cómo se comparan las nuevas geometrías con las predichas con *Glide* para estos 10 compuestos?  
¿Según *MM-GBSA* cuál es el compuesto que más probablemente sea activo? ¿El orden es el mismo que el ofrecido por *GlideSP*?