

# **TP1: Cribado virtual de compuestos**

**Docentes a cargo:**

**Dr. Juan Pablo Bustamante**

**Dra. María Cecilia Gómez**

**Diseño & Descubrimiento de Drogas**

## 1- Introducción

El *virtual screening* o cribado es una técnica computacional utilizada para realizar búsquedas de compuestos químicos que actúen como potenciales ligandos de unión a blancos de drogas.

SIRT1 es una de las 7 sirtuinas codificadas en el genoma de mamíferos. Las diferentes variantes tienen que ver con distintas localizaciones subcelulares así como también de sus actividades químicas. Es una proteína NAD-dependiente que desacetila una gran variedad de sustratos incluyendo p53, NF-κB, FOXO y PGC-1α, entre otros, involucrados en procesos celulares que varían desde metabolismo energético hasta supervivencia celular. Cabe destacar que dicha proteína se ha visto sobreexpresada en diferentes tumores. Como podrá darse cuenta la proteína SIRT1 está implicada en una gran variedad de enfermedades humanas y por ende constituye un prometedor blanco terapéutico<sup>1,2</sup>.

El dominio catalítico de la SIRT1 está compuesto por un extenso subdominio de unión a NAD<sup>+</sup>, que presenta el plegamiento de Rossmann (láminas β paralelas conectadas por hélices α), y un subdominio menor compuesto por un módulo helicoidal y un sitio de unión a Zn<sup>2+</sup>, estos dos últimos están insertos en el subdominio de NAD<sup>+</sup>. Adicionalmente hay una región denominada segmento regulatorio C-terminal (CTR), que se ubica en el borde inferior del dominio de unión de NAD<sup>+</sup> en una región hidrofóbica de la proteína. Dicho segmento, a su vez, está dividido en dos regiones, la C-terminal es la menor de las dos regiones, mientras que la mayor corresponde al N-terminal e incluye un giro β que complementa la lámina β central del plegado Rossmann con dos láminas adicionales.

La proteína se puede encontrar en dos formas, una denominada *open* en la cual el cofactor NAD<sup>+</sup> no está presente (forma *apo* de la enzima), y otra denominada *close* en la cual el cofactor se encuentra completamente encapsulado y se forma un sitio de unión con un túnel hidrofóbico que conduce el sustrato a la región interior del sitio activo. En la forma *apo* el dominio menor de la SIRT1 sufre una rotación de ~25° acompañada por un desplazamiento de ~15 Å de los residuos del dominio produciendo una apertura entre los dominios. La región de unión a NAD<sup>+</sup> y la de unión de CTR permanecen prácticamente sin cambios.

## 2- Procedimiento

**NOTA 1:** Cree una carpeta para el TP1 y trabaje siempre desde allí.

### 2.A- Obtención y análisis de las estructuras receptor y ligandos

2.A.a - Obtención del receptor: Descargue desde la base de datos Protein Data Bank (<http://www.rcsb.org/>) el pdb correspondiente a la proteína deacetilasa nad-dependiente Sirtuin-1 de humano o SIRT-1 (pdb ID 4KXQ). Fíjese la información brindada en la página web

sobre esta macromolécula: tamaño, tipo de estructura, estado cuaternario, ligandos y la información que considere pertinente. Ayúdese mirando el paper asociado a dicho pdb<sup>1</sup> y **redacte un resumen sobre el sistema comentando esa información.**

Observe con un editor de texto el archivo pdb, intente identificar la información brindada en la introducción y con la ayuda de VMD observe los distintos dominios y moléculas que se encuentren en la estructura. **Haga una imagen en la cual identifique las cadenas, dominios y el sitio de zinc.**

2.A.b - Obtención de ligandos: Para obtener el set de compuestos que usaremos para el presente estudio de cribado, vaya a la página de ZINC (<http://zinc.docking.org/>) y descargue el dataset de DrugBank (Catalogs → “Search for catalogs”: DrugBank → Seleccionar el grupo DrugBank-approved. Luego en la sección “Interesting Substance Subsets” seleccione “Browse All” y una vez adentro “get total” bajo el formato sdf).

## 2.B- Preparación de archivos de docking:

### 2.B.a – Preparación del receptor

Vamos a preparar el archivo del receptor para el docking. Para no modificar el archivo original y asegurarnos de no perder la información relevante del sistema que nos provee dicho archivo, haga una copia del mismo bajo el nombre 4kxq\_cristal.pdb que va a ser el archivo que si vamos a modificar.

Teniendo en cuenta la discusión sobre el modelo a simular realizaremos con ayuda del VMD las siguientes modificaciones sobre el archivo generado anteriormente:

1. Eliminación de la información sobre el método cristalográfico del archivo PDB.
2. Eliminación de las conformaciones alternativas a la A en los residuos que posean más de una conformación.
3. Eliminación de aguas cristalográficas, hidrógenos y moléculas co-cristalizadas (excepto el átomo de Zn).
4. Deje sólo los residuos correspondientes a la cadena A.

Guarde la estructura con dichas modificaciones bajo el nombre 4kxq\_proteina\_sH\_ZAFF.pdb

Como ya habrá notado, nuestra proteína presenta un sitio metálico. El centro es un átomo de Zn coordinado a 4 cisteínas. Reconozca el número de residuo de las cisteínas coordinadas al metal y editando el PDB cambie el nombre del residuo CYS por CY1: este procedimiento es una forma de advertirle al tLeap que para construir la topología considere los parámetros de una cisteína desprotonada en el azufre que se une al átomo de Zn. (Residuos de cisteína coordinados al Zn 371 374 395 398). Al átomo de Zn renómbrelo como residuo ZN1 sin modificar el tipo de átomo.

Una vez que tengamos nuestro receptor con todos los cambios realizados ingrese al programa tLeap. Cargue el campo de fuerzas para proteínas (ff99SB). Para tener la información de parámetros y conectividades del sitio metálico vamos a precisar de un campo de fuerzas

adicional llamado ZAFF. Cargue dicho campo de fuerzas dentro de su sesión de tLeap con los siguientes comandos:

```
> loadamberprep ZAFF/ZAFF.prep  
> loadamberparams ZAFF/ZAFF.frcmod
```

Cargue la estructura PDB corregida y observe la salida del programa para ver si hubo algún error o warning en el proceso

```
> SIR=loadpdb 4kxq_proteina_sH_ZAFF.pdb  
> check SIR
```

Como habrá notado en la salida del programa tenemos tres advertencias en la estructura. La primera es sobre un enlace entre el átomo C del residuo 510 y el átomo de Zn del residuo 511 que no posee parámetros. Ese enlace fue declarado por el programa únicamente porque ambos residuos son consecutivos en el pdb que cargamos, pero no debe ser tenido en cuenta porque como ya visualizamos anteriormente, el Zn sólo está coordinado a átomos de S de cisteínas. Utilice el comando **deleteBond** para eliminar el enlace y vuelva a verificar la estructura.

```
> deleteBond SIR.510.C SIR.511.ZN  
> check SIR
```

Otra forma de evitar la formación del enlace es poner una línea con la palabra TER luego de los átomos de la proteína, es decir, antes del átomo de Zn.

Para eliminar el siguiente *Warning* que nos dice que el sistema tiene una carga neta de -3, agregue contraiones de Na<sup>+</sup> que compensen dicha carga

```
> additions SIR Na+ 0
```

Compruebe nuevamente la estructura e informe el resultado en el informe. Guarde la estructura con el comando

```
> savepdb SIR SIRT_close_dock.pdb
```

Salga del programa tLeap y abra AutoDockTools (adt)

Cargue la estructura pdb generada en el paso anterior yendo a File → Read Molecule. Luego asigne a los átomos el tipo AD4 y calcule las cargas Kollman del receptor

Edit → Atoms → Assign AD4 type

Edit → Charges → add kollman charges.

Las cargas de Kollman representan valores fijos para cada átomo de los aminoácidos derivados del potencial electrostático obtenido por mecánica cuántica.

En caso que al asignar el tipo AD4 a los átomos le aparezca algún error en la consola de ADT, corrija con un editor de textos la línea de los azufres de las cisteínas CY1 de manera que los caracteres de SG queden alineados con los otros átomos. Luego deberá volver a cargar la molécula en ADT.

Guarde el archivo del receptor con la extensión pdbqt

## 2.b – Preparación de los ligandos

Dado que el cribado consiste en usar muchos potenciales ligandos a blanco, preparar cada ligando individualmente sería muy engorroso, por lo cual usaremos algunos scripts que automatizan el proceso. Hemos descargado de la página web de ZINC un solo archivo en formato sdf que contiene los compuestos (Identifique cuantos compuestos tiene nuestro set de ligandos). Precisamos primero separar este archivo en archivos mol2 individuales de cada compuesto. Abra el archivo virtual\_drugbank.sh para entender que hace en cada paso, anótelos en el informe y luego ejecútelo.

**NOTA 2:** Verifique que las ubicaciones y los nombres de los archivos utilizados por el script se corresponden con sus archivos o modifique convenientemente el script para que utilice los archivos que correspondan correctamente.

```
#!/bin/sh
mkdir ../VirtualScreening
cd ../VirtualScreening
mkdir Ligands_drugbank
cd Ligands_drugbank
csplit --prefix=tmp -n4 -ks ../dbap.sdf '$$$$/1' '{*}'
```

```
for file in $(ls)
do
zid=$(grep ZINC $file)
filename="$zid".sdf
mv $file $filename
done
```

Realizaremos ahora una optimización por medio de un campo de fuerzas de las estructuras de los ligandos. Cree una carpeta con el nombre mol2 dentro de Ligands\_drugbank y ejecute desde esta última carpeta el comando

```
obabel -i sdf *.sdf -o mol2 -O mol2/.mol2 --gen3d -m
```

Esta operación va a demorar un tiempo considerable (dependiendo también de los recursos computacionales, en mi PC fueron como 7 hs) ya que se van a realizar optimizaciones sobre cada uno de los compuestos de nuestro set. También recibirá algunas advertencias sobre el proceso de optimización pero no son relevantes para la actividad que nosotros vamos a realizar.

Una vez obtenidas las estructuras optimizadas se generarán los archivos pdbqt individuales que son los archivos que precisa autodock vina para llevar a cabo el cribado. Cree en la carpeta donde tiene los archivos .sdf una carpeta nueva con el nombre pdbqt y desde la carpeta mol2 ejecute el siguiente comando:

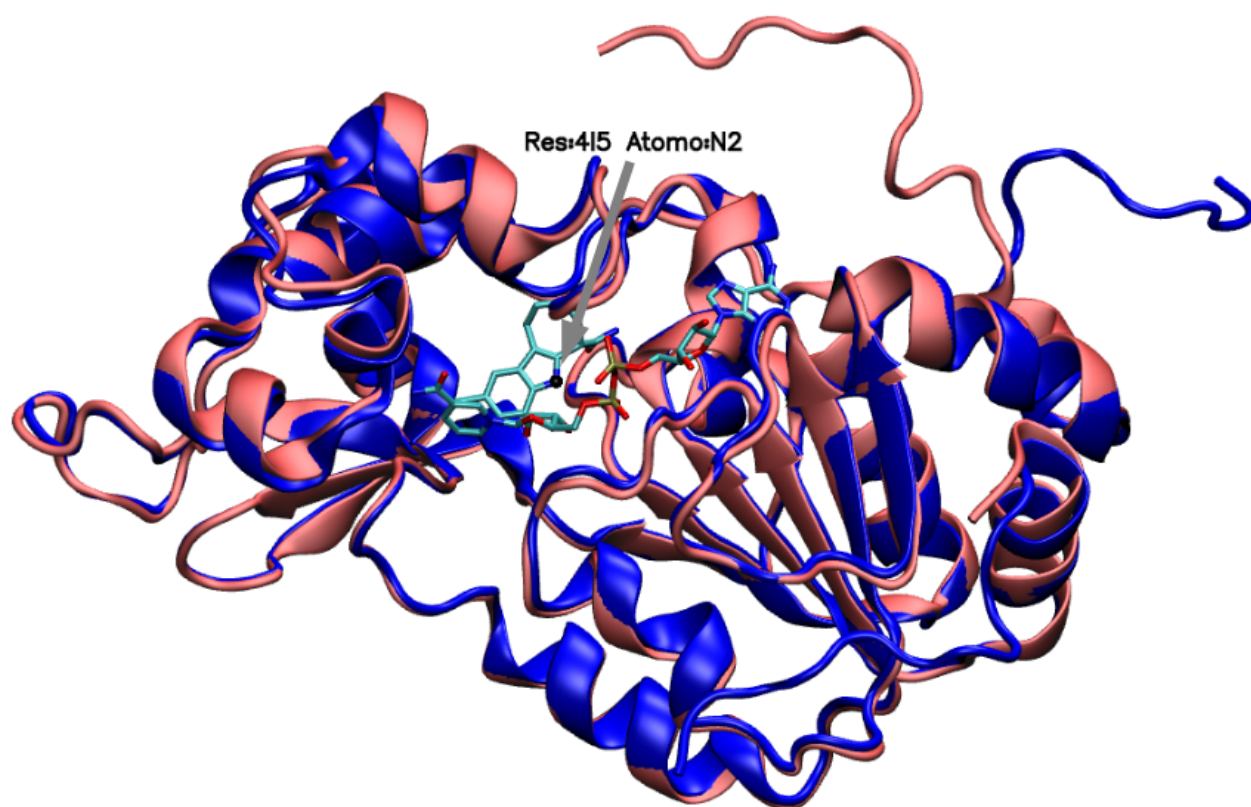
```
obabel *.mol2 -O ../pdbqt/*.pdbqt --AddPolarH
```

Babel<sup>2</sup> es un programa que se utiliza para cambiar formatos de archivos. Por lo general reconoce el tipo según la extensión de los mismos, pero sino se puede indicar con comandos. En este caso como no se indicó ningún prefijo para los archivos de salida se va a copiar el nombre de cada .sdf. La siguiente opción indica al programa que debe agregar los hidrógenos polares. **¿A qué piensa que se debe que solo se le pidan agregar los hidrógenos polares?**

## 2.c - Configuración de la grilla

Para realizar el cribado no precisamos calcular las grillas de afinidad de los átomos pero si indicar el centro y las dimensiones (en Angstroms) a AutoDock Vina y un parámetro que indica la exhaustividad en la búsqueda.

Descargue del PDB la estructura depositada bajo el código 4I5I correspondiente a la SIRT1 unida a un ligando y **evalúe en qué posición centraría la grilla si quisiera obtener el posicionamiento del ligando reportado**. Tenga en cuenta que cuando quiera utilizar la información del centro de la grilla encontrada para la estructura 4I5I en nuestro receptor 4KXQ (utilice el pdb guardado con tleap), las proteínas deben estar alineadas para tener el mismo sistema de ejes de coordenadas. Puede superponerlas yendo, en VMD, a Extensions -> Analysis -> MultiSeq. (Cargue primero la estructura que desea utilizar como referencia para la superposición, es decir la que quiere conservar sus coordenadas, en nuestro caso SIRT\_close\_dock.pdb).



**Fig. 1 - Superposición de estructuras SIRT\_close\_dock.pdb (rosado) con la estructura cristalográfica 4I5I (azul). Cofactor NAD y compuesto 4I5 en licorice correspondientes a la estructura 4I5I.**

Para determinar las dimensiones de la caja puede medir a partir del átomo usado como centro la distancia aproximada hacia el átomo más extremo en cada dimensión, o de manera automática:

TK console:

```
set sel [atomselect 1 "resname XXX and chain A"]
measure minmax $sel
```

XXX debe ser reemplazado por el nombre del ligando del cual queremos conocer sus dimensiones.

Como estamos trabajando con muchos archivos, vamos a crear otro script con la configuración de la grilla que va a ser la misma para todos los compuestos. Cree el archivo conf.txt siguiendo el modelo a continuación, complete con los datos de la grilla que obtuvo anteriormente y en la opción *exhaustiveness* seleccione 20:

```
receptor = receptor.pdbqt
```

```
center_x =
center_y =
center_z =
```

```
size_x =
size_y =
size_z =
exhaustiveness =
```

## 2.C - Ejecución del Docking y análisis

Ahora que ya contamos con todos los archivos de entrada de las estructuras receptor y ligandos, y el archivo de configuración de la grilla, estamos en condiciones de llevar adelante nuestro cribado.

Nuevamente recurriremos al uso de un script que automatice el procedimiento, llámelo `vina_screen_local.sh` y ejecútelo dónde ubicó los archivos `.pdbqt`:

```
#!/bin/bash

for f in $(ls ZINC*.pdbqt); do
    nombre_lig=$(basename $f .pdbqt)
    echo Procesando el ligando $nombre_lig
    mkdir ${nombre_lig}
    vina --config conf.txt --ligand $f --out ${nombre_lig}/out.pdbqt --log ${nombre_lig}/log.txt
done
```

**¿Entiende qué realiza en cada paso? Coméntelo en el informe.**

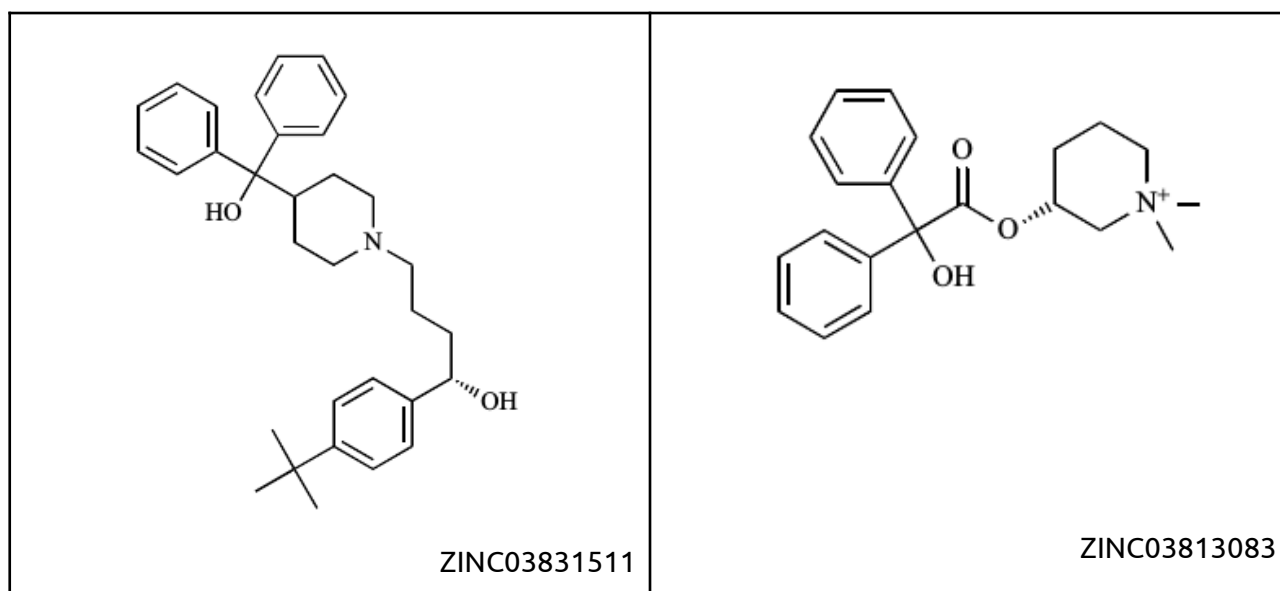
**NOTA 2:** Debido a la demanda computacional de correr este script con todo el set de compuestos y el tiempo limitado de TP, llegado a este punto solicite a los docentes los archivos correspondientes para el análisis

El siguiente paso sería ordenar los resultados de cada docking por energía y quedarnos con los compuestos que presenten la mejor afinidad por nuestro blanco. En este caso lo que vamos a hacer es analizar algunos compuestos puntuales. Diríjase a la carpeta **subset** y visualice cada una de las conformaciones de los distintos potenciales ligandos. El archivo .pdbqt de salida de cada docking entre los ligandos y el receptor contiene las múltiples conformaciones posibles para el ligando correspondiente. Con el siguiente comando puede obtener archivos pdb individuales corriéndolo dentro de cada carpeta de ligandos.

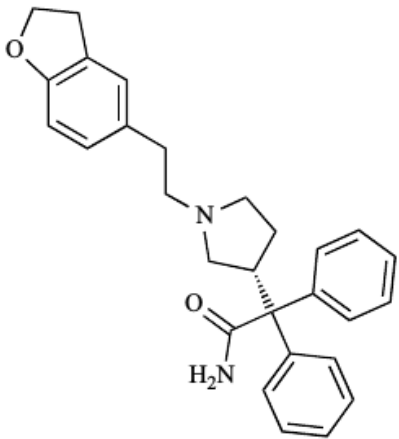
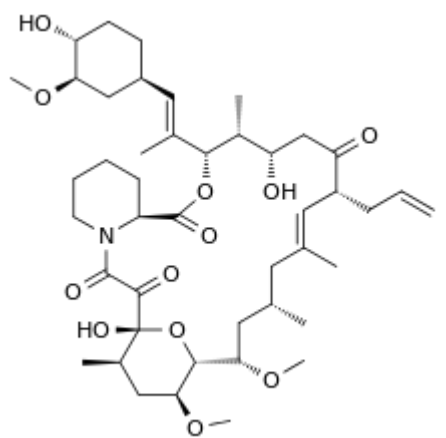
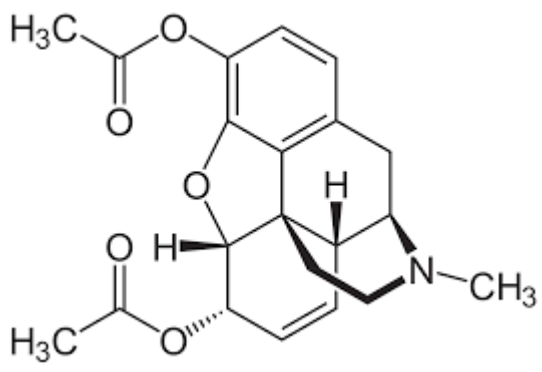
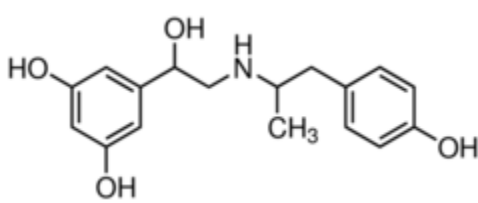
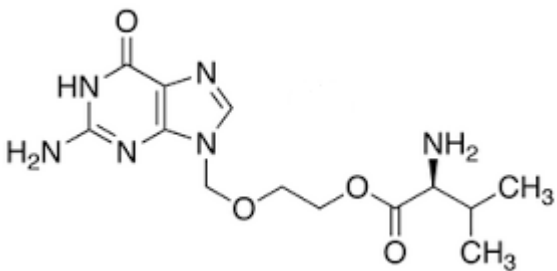
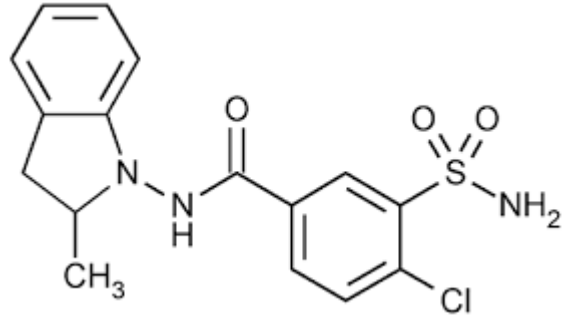
```
obabel out.pdbqt -O out.pdb -m
```

**¿Considera que los parámetros de la grilla fueron elegidos correctamente? Justifique su respuesta.**

Utilice el archivo grilla\_1.pdb provisto por la cátedra para visualizar la grilla utilizada. Analice la diversidad de estructuras de los ligandos con ayuda de la Fig. 1 y optimice, con los compuestos seleccionados de la carpeta **subset**, los parámetros de la grilla de manera que podamos obtener una configuración que sea válida para un mayor número de ligandos.





 <p>ZINC01996117</p>	 <p>ZINC85537027</p>
 <p>ZINC04097183</p>	 <p>ZINC00057320</p>
 <p>ZINC01530713</p>	 <p>ZINC00601305</p>

**Fig. 2: Diversidad de ligandos**

Realice varios dockings con algunos de los compuestos de la carpeta hasta que considere que los parámetros de la caja sean adecuados. Discuta con sus compañeros y docentes, y fundamente luego en el informe, **cuándo consideró que los parámetros de la caja fueron los adecuados para la confección de la grilla**. Ayuda: en función de qué lo decidió, qué observó, qué comparó. **Redacte en el informe la estrategia elegida por usted para optimizar los parámetros de la grilla y los criterios para seleccionar el mejor de ellos.**

Luego de esto, realizaremos el análisis de los resultados de la carpeta Cribado\_Final/Resultados\_Docks, los cuales han sido obtenidos con los parámetros:

```
receptor = receptor.pdbqt
```

```
center_x = 24.4
```

```
center_y = -23.7
```

```
center_z = 8
```

```
size_x = 17
```

```
size_y = 20
```

```
size_z = 15
```

```
exhaustiveness = 20
```

Utilice el siguiente comando para obtener el nombre de los primeros 7 resultados de docking ordenados por la energía de unión (observe si alguno coincide con el trabajo de Balasundaram Padmanabhan *et al.*<sup>3)</sup>.

```
grep " 1 " */log.txt | sort -k 3 -n | head -n7
```

Cree una carpeta llamada Seleccionados dentro de Resultados\_Docks y copie en ella los resultados de los ligandos seleccionados. Cree un script llamado split\_conformaciones.sh para obtener los pdbs individuales de cada ligando

```
#!/bin/bash
```

```
for f in $(ls ZINC*/out.pdbqt); do
```

```
    b=$(dirname $f)
```

```
    cd $b
```

```
    babel -i pdbqt out.pdbqt -o pdb $b.pdb -m
```

```
    cd ..
```

```
done
```

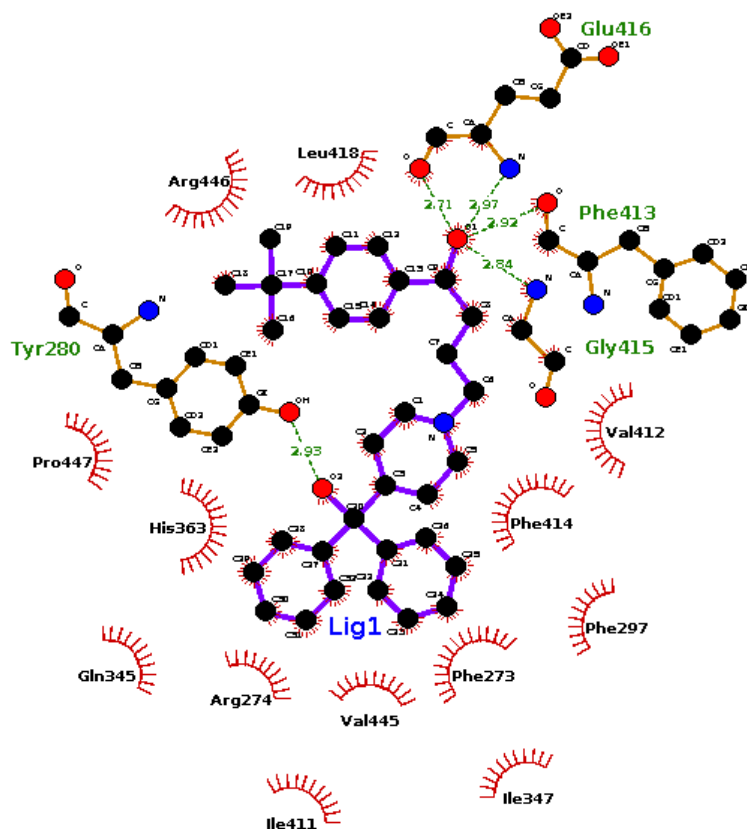
¿Cuántas poses observa para cada uno de los compuestos?

¿Con qué ligando obtuvo el docking de menor energía? ¿Y el de mayor energía?

**Visualice las diferentes conformaciones de cada uno junto al receptor y escriba en el informe sus conclusiones al respecto. ¿Cuáles son las interacciones que más afectan los resultados de los dockings?. Genere capturas de por lo menos 3 complejos (al menos uno de los que coinciden con el paper mencionado) e identifique las interacciones que detecta en cada una de ellas y los átomos y residuos involucrados en dicha interacción.**

### Análisis con LigPlus de los ligandos

Una herramienta muy útil y sencilla para visualizar las interacciones entre el ligando y el receptor es el archivo de java LigPlus. Deberá generar un archivo pdb individual con el receptor y cada uno de los ligandos para la entrada del programa. En la Fig. 3 se muestra un ejemplo de la salida para uno de los ligandos seleccionados con el receptor del cristal 4kxq.



**Fig.3:** Salida del programa LigPlus para el complejo 4kxq-ZINC03831511

En la Fig. 3 pueden verse los residuos de la proteína que establecen interacciones hidrofóbicas con el ligando como es el caso de Phe414, His363, Gln345 y los enlaces puente hidrógeno junto con sus distancias como con la Tyr280, Phe413. Realice dicho análisis con 3 de los ligandos. **¿Se corresponden las interacciones que había supuesto en su análisis con VMD? Comente brevemente las coincidencias/diferencias que obtuvo entre ambos análisis.**

Basado en un análisis estructural final, **¿puede identificar cuáles de ellos son los más estabilizados, es decir, los de menor energía de unión? ¿En qué basa su razonamiento?. Escriba las respuestas en el informe.**

Como habrá notado hay diferentes tipos de interacciones entre los ligandos y la proteína, interacciones hidrofóbicas, electrostáticas, polares, etc, y no necesariamente se encuentran presentes las interacciones con los mismos residuos en cada complejo. **¿Por qué considera que ocurre esto?** Habrá encontrado soluciones con valores de afinidad energética muy próximos que tengan interacciones diferentes con la proteína, esto nos da la oportunidad de sugerir compuestos nuevos, modificando los actuales, que maximicen las interacciones combinando grupos que favorecen a la estabilidad de los compuestos para nuestro receptor en estudio. En

este sentido, el siguiente trabajo práctico estará focalizado en optimizar tanto los ligandos como diferentes parámetros de la estrategia de docking.

## Referencias

- 1- Andrew M. Davenport *et al.*- "Structural and Functional Analysis of Human SIR" - JMol Biol. **2014** 426,3,: 526–541.
- 2- Balasundaram Padmanabhan, *et al.* - "Identification of New Inhibitors for Human SIRT1: An *in-silico* Approach" - Medicinal Chemistry, **2016**, 12, 347-361
- 3- [http://openbabel.org/wiki/Main\\_Page](http://openbabel.org/wiki/Main_Page)