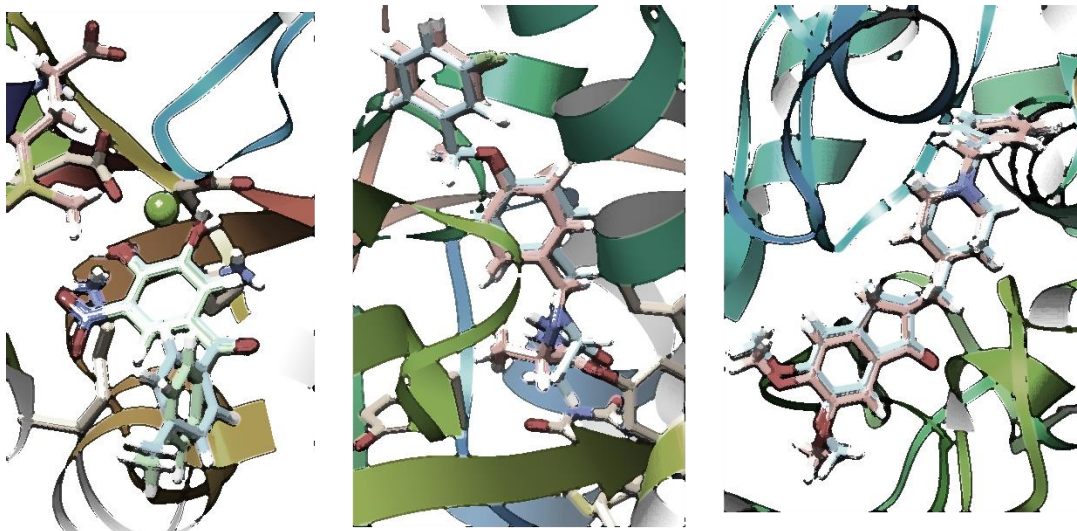


Simulaciones proteína con ligandos co-cristalizados. Guía para realizar “re-docking”



Luis Felipe Hernández Ayala
México, 2023

Índice general

Sección	Pág.
1. Introducción	3
2. Descarga e instalación de software y proteína COMT.	4
3. Re-docking	4
3.1. Descarga de la proteína COMT PDB Id: 3S68	5
3.2. Limpieza de la proteína y creación de archivos .pdb del inhibidor y la proteína	6
3.3. Visualizando los archivos .pdb en Chimera	11
3.4. Re-docking con ADV en la interfaz de Chimera	12
3.5. Cálculo de RMSD entre las estructuras simuladas y la experimental	17
4. Agradecimientos	21

Manual de re-docking

Luis Felipe Hernández Ayala* hdz.ayala@yahoo.com.mx

*Consejo Nacional de Humanidades, Ciencia y Tecnología - Universidad Autónoma Metropolitana, Unidad Iztapalapa.

1. Introducción

El proceso de re-docking consiste en realizar la simulación de un acoplamiento proteína-ligando y comparar las conformaciones obtenidas con información estructural experimental. En resumen, se realiza el acoplamiento de un ligando inhibidor conocido de una enzima y se compara con su estructura obtenida por diferentes técnicas experimentales (difracción de Rayos X, criomicroscopía electrónica, resonancia magnética nuclear, etc.). Este método es utilizado comúnmente para evaluar la precisión del protocolo de acoplamiento molecular, determinar las dimensiones y coordenadas de la *gridbox* y en cierto modo, validar la confiabilidad de las simulaciones. Por lo anterior, resulta fundamental realizar un estudio de re-docking antes de computar las simulaciones de las moléculas de interés. Es importante señalar que no siempre existe información estructural de algunas proteínas y en esos casos, el re-docking no resulta viable y algún otro protocolo de validación debe ser utilizado.

Tres son los factores que deben tomarse en cuenta para la elección del sistema proteína ligando para el re-docking:

1. La similitud química del ligando co-cristalizado con las moléculas a estudiar
2. El tamaño de la molécula. El tamaño del ligando co-cristalizado debe ser similar al de las moléculas a evaluar.
3. El número de enlaces rotables. Entre más enlaces rotables posea el ligando co-cristalizado, evaluar la precisión del método puede ser más difícil.

A veces debemos conformarnos con cumplir con 2 de los tres factores enunciado, debido a que la información experimental de ligandos co-cristalizados es limitada y la similitud química no es comparable cuando se estudian moléculas novedosas. Afortunadamente existen gran cantidad de información disponible en las bases de datos de las que podemos echar mano para realizar este tipo de estudios.

Como se mencionó anteriormente, el centro y coordenadas de la caja o *gridbox* obtenidas en el *re-docking* se utilizarán para realizar las simulaciones de los sistemas que se desean estudiar. Adicionalmente la validación del método se realizará calculando un parámetro denominado RMSD, que nos dice qué tan diferente es la conformación obtenida por la simulación comparada con la determinada experimentalmente. Un valor aceptable

de RMSD debe ser menor a 2.5 Å, sin embargo, un valor menor a 1.0 Å sería un excelente indicador de la confiabilidad de los resultados de las simulaciones.

2. Descarga de software y proteína COMT.

Para realizar el re-docking se utilizará el software de acceso libre USCF Chimera y Autodock Vina. La instalación de los programas queda fuera del propósito del manual, pero no hay más que descargar los archivos ejecutables (.exe) y seguir las instrucciones en pantalla. Los links de descarga se muestran a continuación:

USCF Chimera:

<https://www.cgl.ucsf.edu/chimera/download.html>

Autodock Vina 1.1

<https://vina.scripps.edu/downloads/>

De la carpeta de instalación de Autodock Vina se realiza una copia del ejecutable Vina.exe. Para este estudio la carpeta de trabajo está ubicada en C:\docking, el archivo ejecutable se llama “Vina”, por lo que la ruta completa será:

C:\docking\Vina

Por otro lado, la enzima utilizada para ejemplificar el proceso de re-docking es la Catecol-O-metil transferasa (COMT) co-cristalizada con el inhibidor Tolcapone, PDB ID: 3S68.

3. Re-docking

El proceso consiste en tres pasos:

1. Generación de los archivos de la proteína, *comt.pdb* y del ligando, *tcw.pdb*. Para el re-docking, el ligando es el inhibidor co-cristalizado, es decir, el Tolcapone.
2. Re-docking entre la COMT y el Tolcapone.
3. Cálculo de RMSD.

3.1 Descarga de la proteína COMT PDB Id: 3S68

USCF Chimera es capaz de descargar proteínas de la base de datos PDB RSC. Para ello, ejecutar *File / Fetch by ID...*, se abre un cuadro de dialogo, en éste se escribe el ID de la proteína (3S68) y se da clic sobre el botón *Fetch*, con esta instrucción se visualiza en pantalla la proteína COMT. Este proceso se describe en las Figuras 1 y 2.



Figura 1. Descargando la proteína COMT (3S68) con el comando Fetch de Chimera.

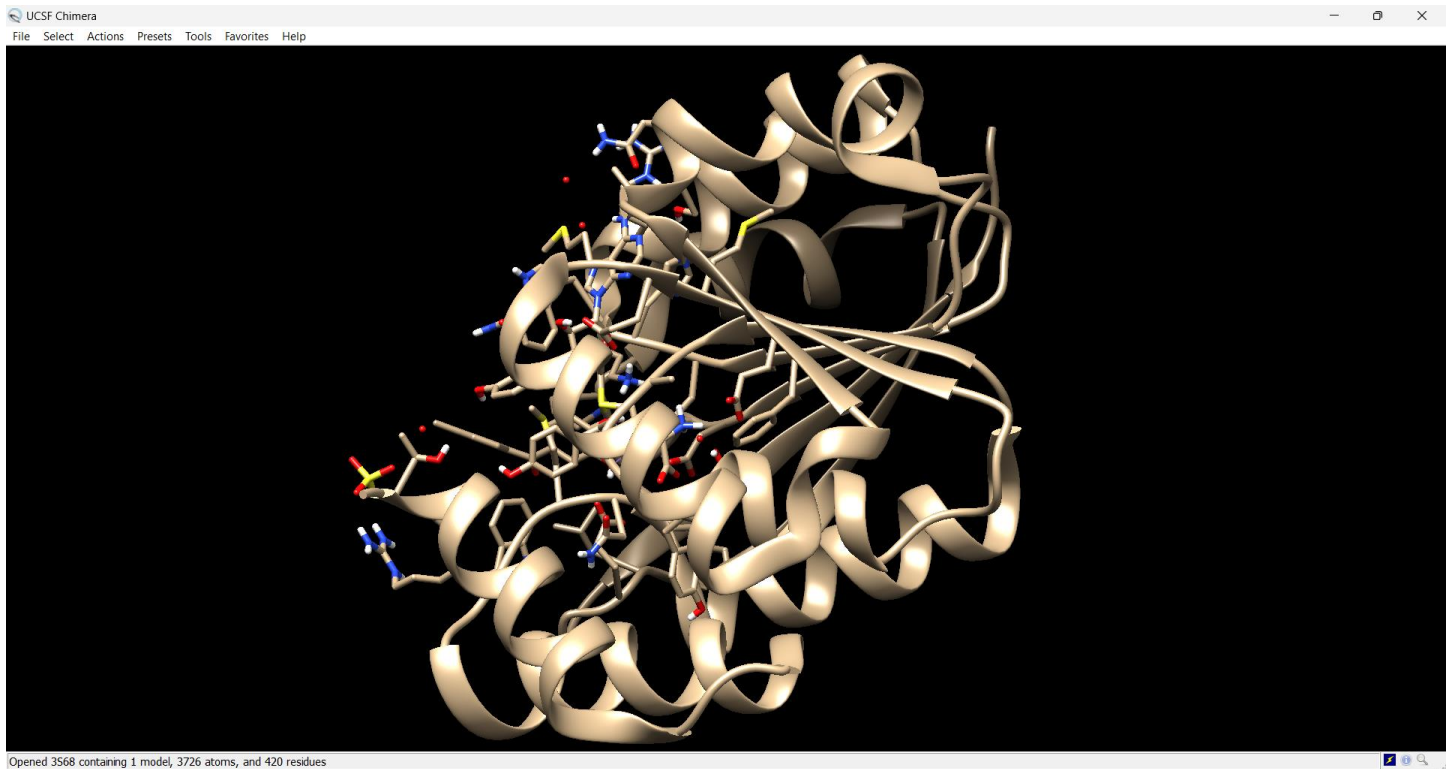


Figura 2. Proteína COMT (3S68) desplegada en la interfaz de Chimera. Además de la estructura proteica se pueden observar las estructuras de los cofactores de la COMT, el inhibidor y algunas otras especies.

3.2 Limpieza de la proteína y creación de archivos .pdb del inhibidor y la proteína

Primero se crearemos el archivo del ligando y lo llamaremos tcw.pdb, para ello seleccionaremos el ligando e invertiremos la selección y finalmente eliminaremos todas las especies, exceptuando al ligando y finalmente crear el archivo correspondiente.

El proceso anterior se logra ejecutando la opción *Select / Residue / TCW*, se iluminará la molécula Tolcapone (abreviada como TCW en la estructura descargada), esto indica que el ligando ha sido seleccionado (Figura 3).

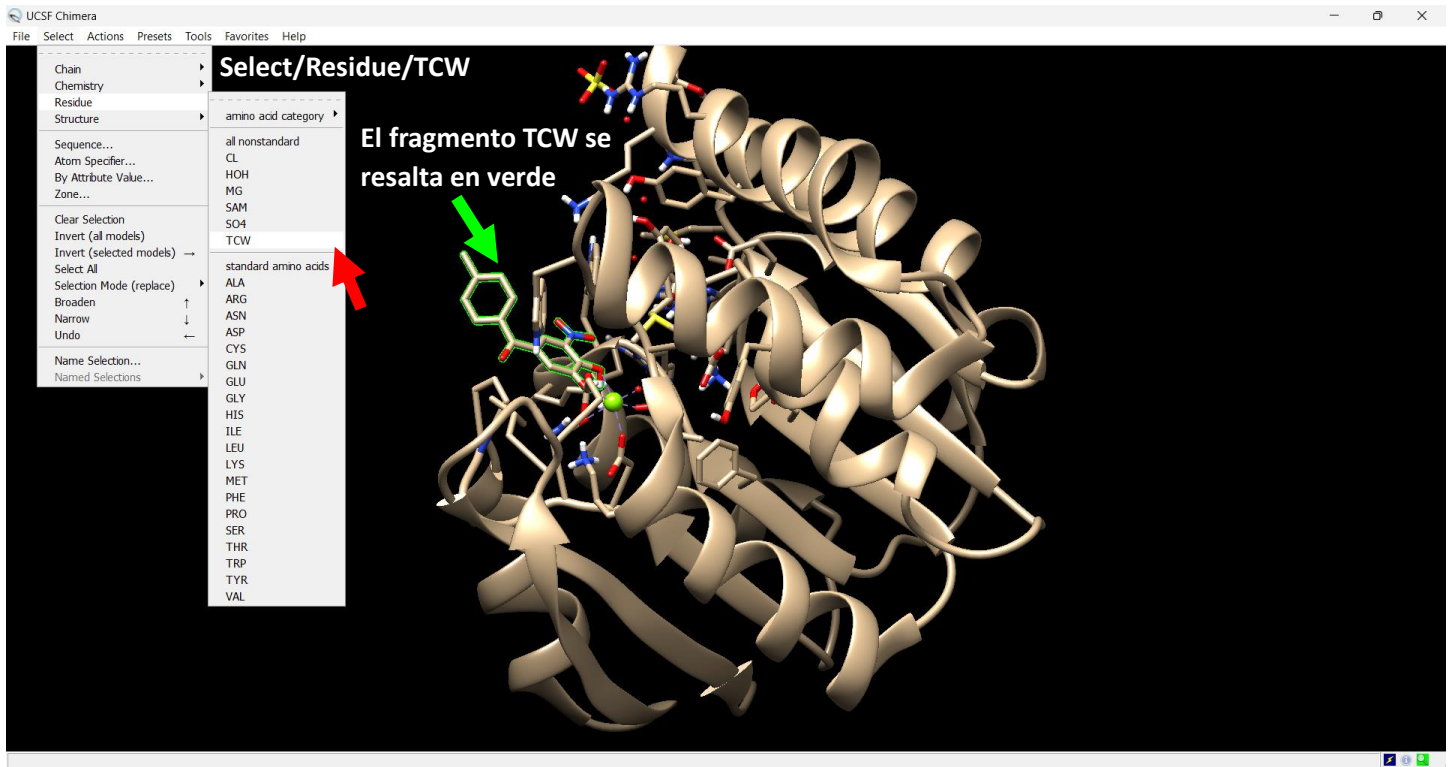


Figura 3. Selección del inhibidor Tolcapone (TCW).

Para invertir la selección, ejecutamos *Select/Invert (all models)* para seleccionar las otras especies (Figura 4).

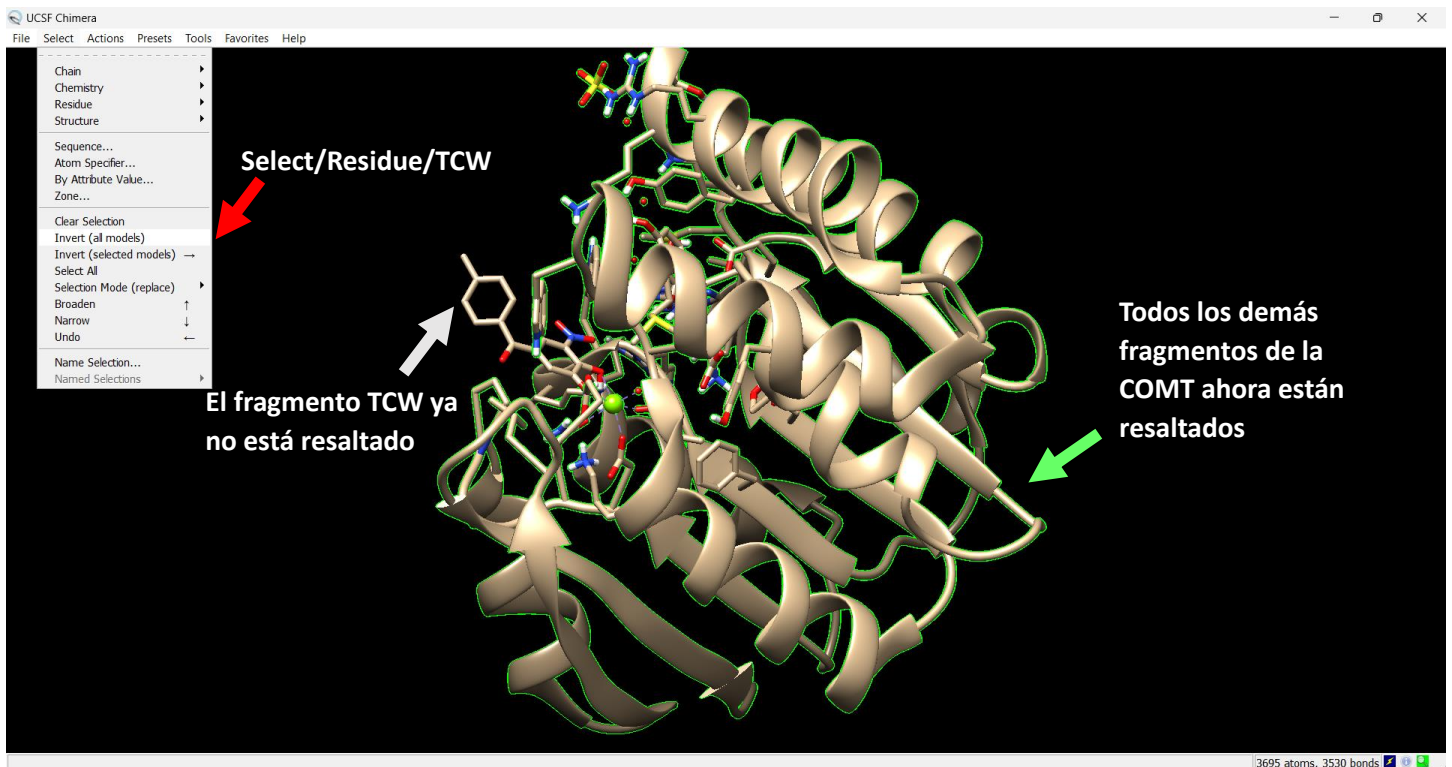


Figura 4. Selección de las estructuras de la COMT excepto Tolcapone (TCW).

Ejecutando *Actions/ Atoms/Bonds / delete* borramos todo, excepto TCW. En la Figura 5 se observa la pantalla actual en USCF Chimera.

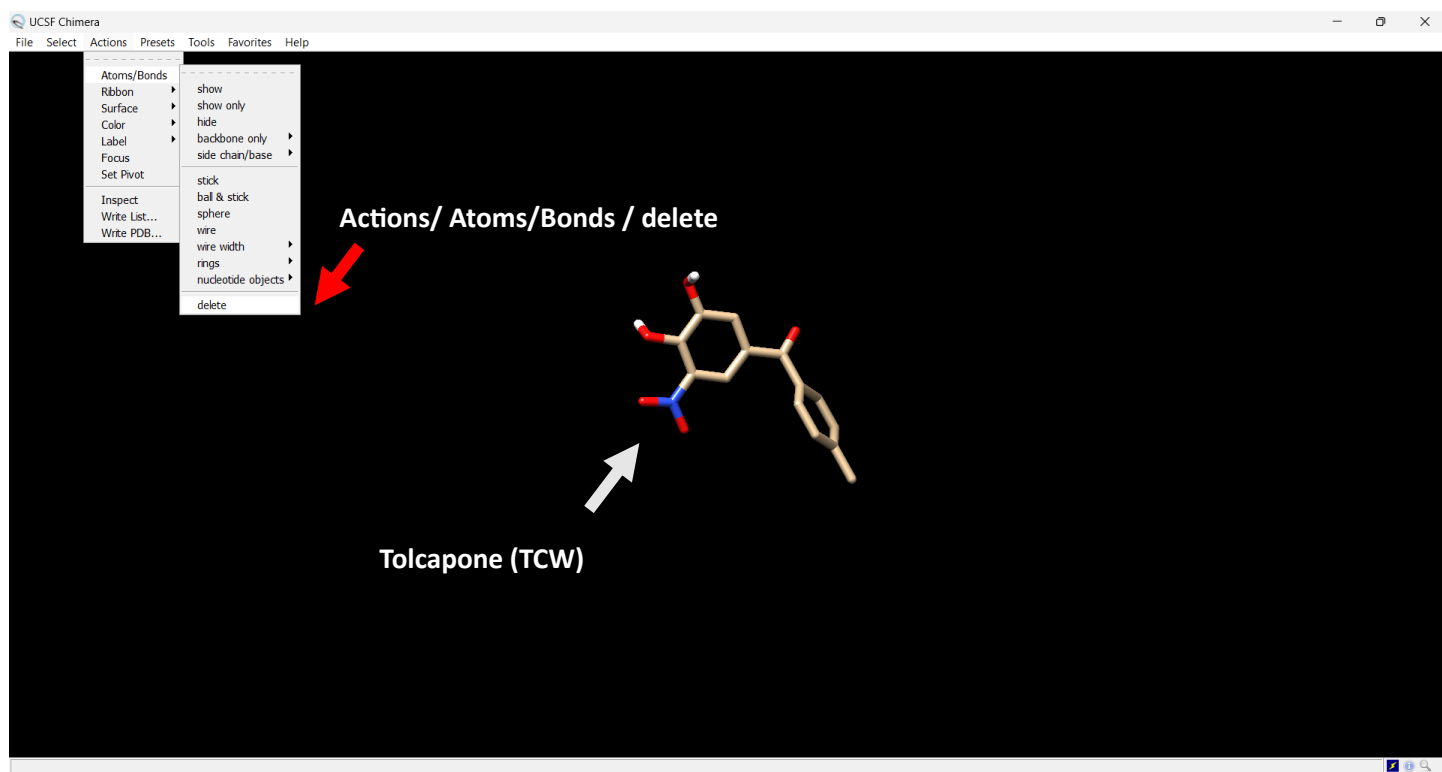


Figura 5. Inhibidor Tolcapone (TCW). La imagen corresponde a la interfaz una vez que se ejecutó el comando borrar.

Para crear el archivo *tcw.pdb*, ejecutamos *File / Save PDB*, se abre un cuadro de dialogo donde debemos nombrar el archivo y seleccionar la carpeta destino, una vez realizado esto se da clic en *Save* y habremos creado el archivo correspondiente al ligando (Figura 6). Es necesario destacar que esta forma de crear el archivo permite guardar las coordenadas atómicas tridimensionales originales (experimentales). Esto es importante ya que serán utilizadas como la posición inicial (semilla) del acoplamiento en el re-docking.

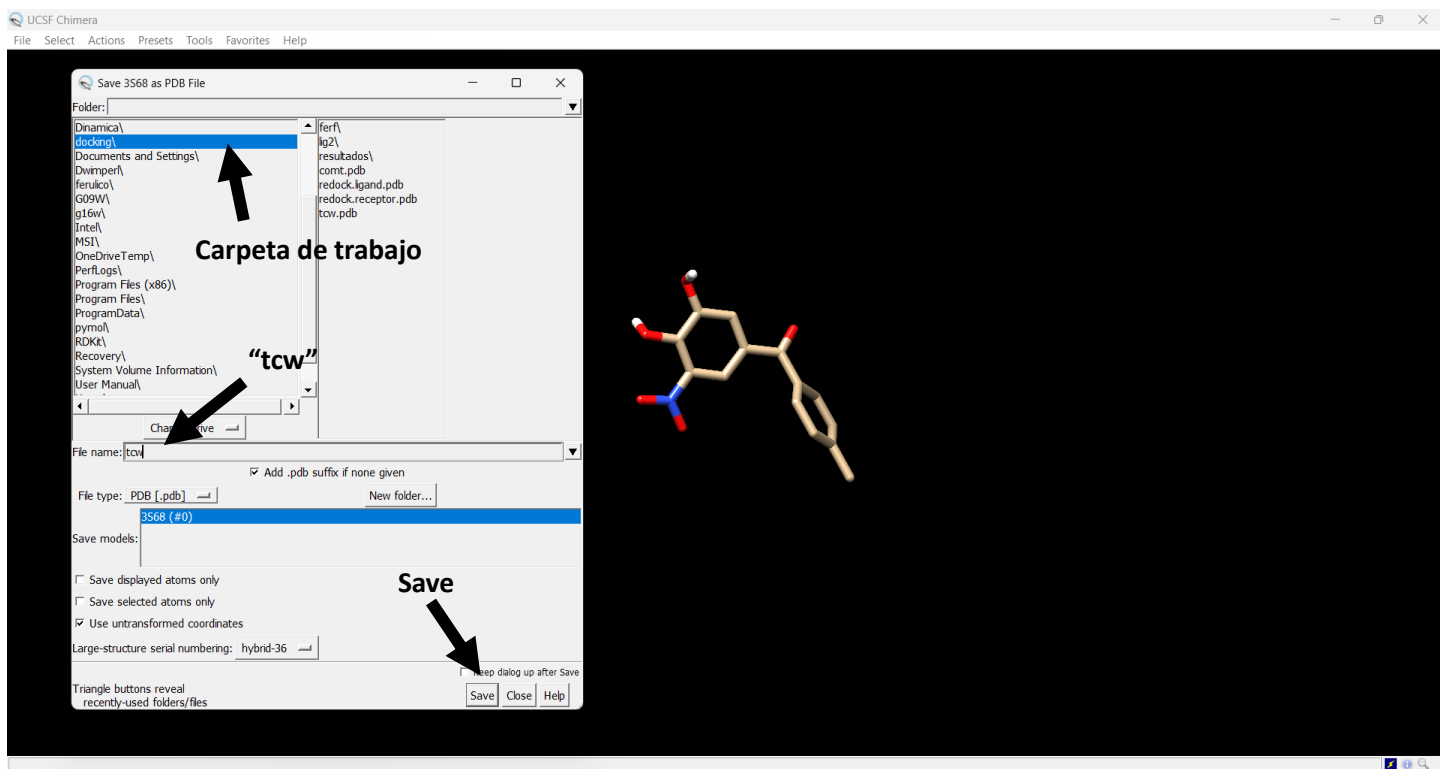


Figura 6. Salvando el archivo *tcw.pdb*.

Para crear el archivo de la proteína (*comt.pdb*) haremos un procedimiento similar al descrito anteriormente. Cerramos sesión con el comando *File / Close Session* para limpiar la interfaz de UCSF Chimera. Descargaremos una vez más la proteína con el comando *Fetch ID...*, una vez abierta borraremos uno a uno los fragmentos no relevantes (FNR), para esto se utilizan los comandos *Select* y *Action*. Los fragmentos que se eliminarán de la estructura original son: Tolcapone (TCW), agua (HOH), iones cloruro (CL) y iones sulfato (SO4). Para cada uno de ellos, ejecutaremos *Select / Residue / FNR* para seleccionar estructura en cuestión. Después, con el comando *Action / Atoms/Bonds / delete*, borraremos dicha especie. El proceso se ejemplifica con borrando las moléculas de agua en las Figuras 7 y 8. Sin embargo recordar que se deben borrar todos los FNR. Una vez más, es necesario recordar que algunos fragmentos son relevantes para la acción de la proteína y se denominan cofactores. En nuestro particular caso, las especies el átomo de magnesio (MG) y la S-adenosilmetionina (SAM) son los cofactores esenciales en la función de la COMT.

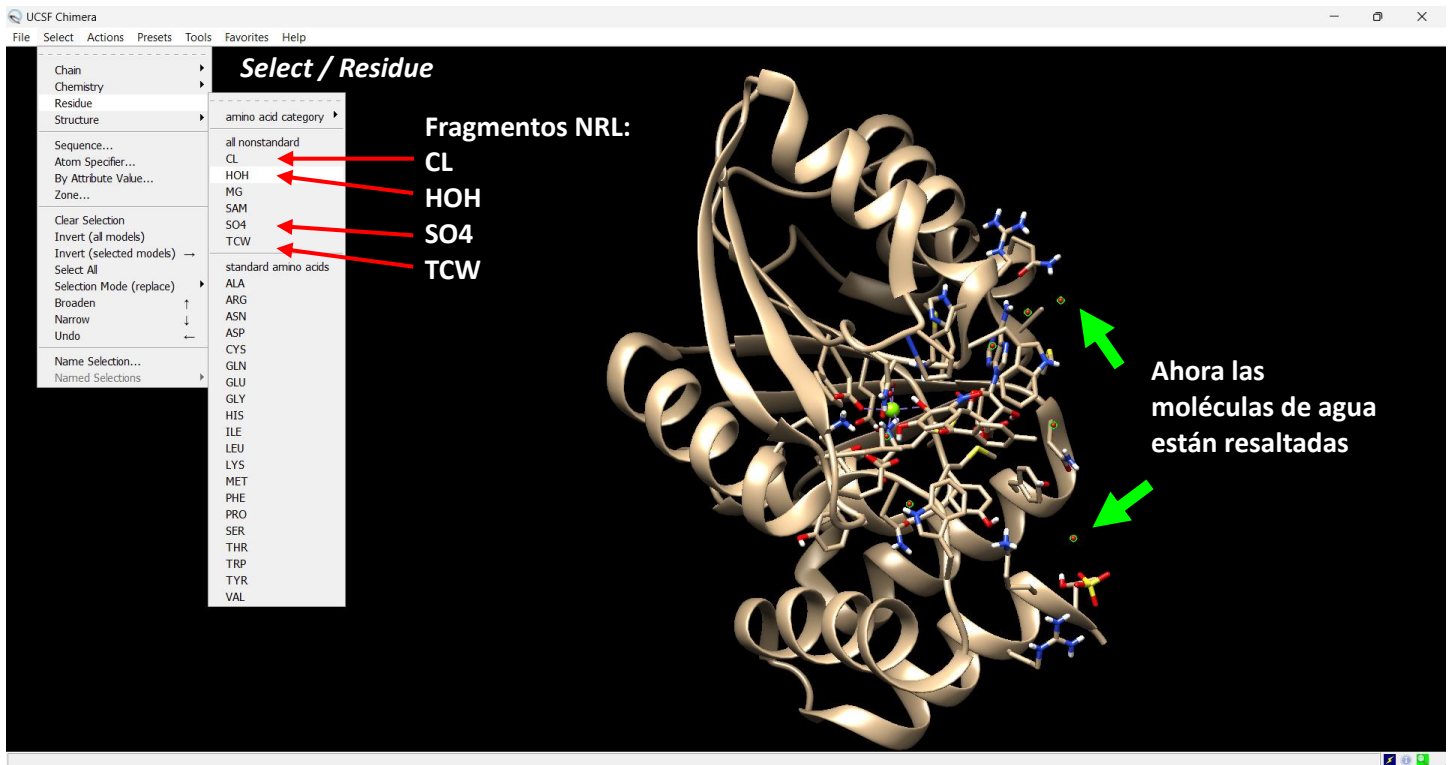


Figura 7. Con el comando *Select* es posible seleccionar cualquier fragmento. En esta figura se han seleccionado las moléculas de agua (HOH) representadas por esferas rojas. Aunque no se note con claridad, las moléculas de agua están resaltadas con color verde.

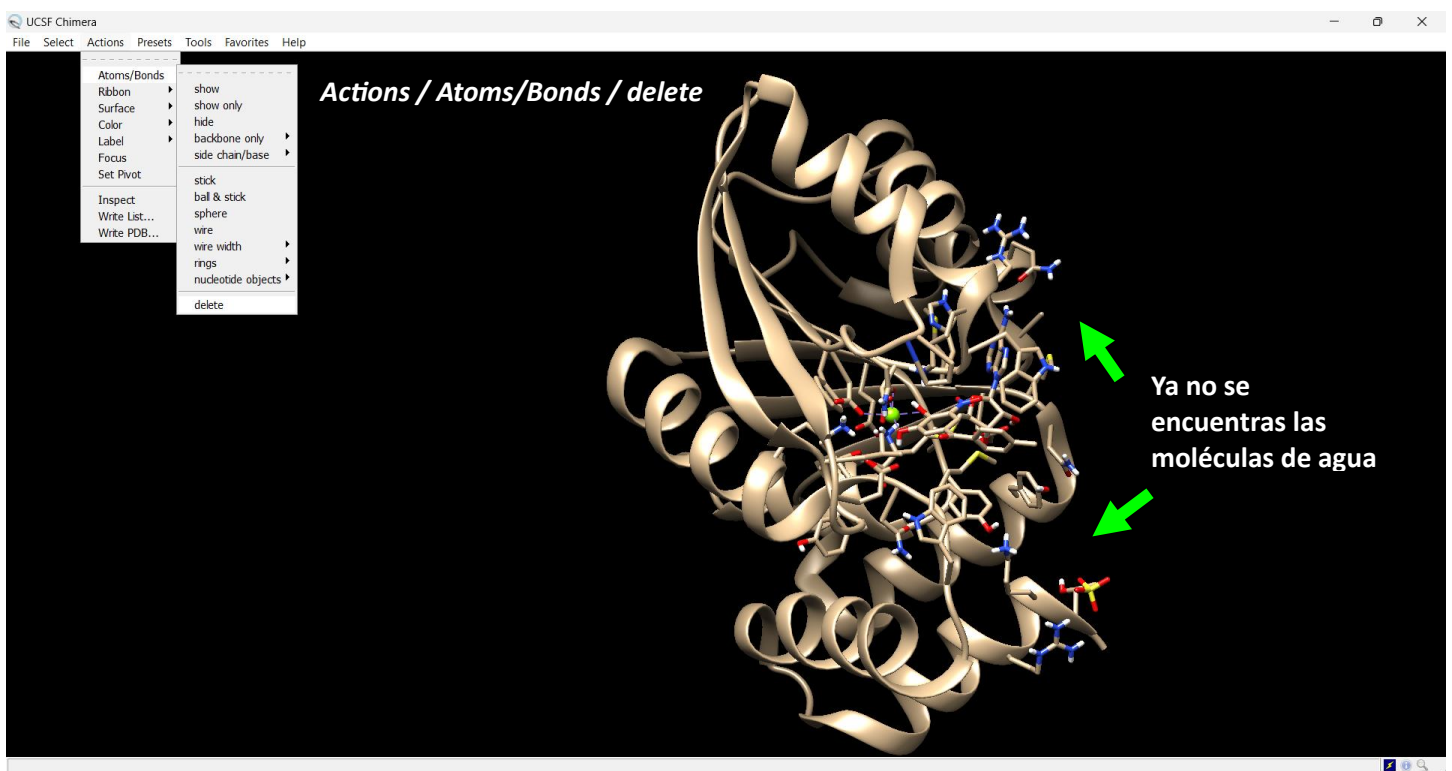


Figura 8. Borrando las moléculas de agua con el comando *Action / ... / delete*. Es necesario eliminar todos los fragmentos NRL.

Al finalizar el proceso de limpieza queda la estructura proteica, el magnesio y la S-adenosilmetionina como se muestra en la Figura 9. El proceso de guardado del archivo *comt.pdb* es análogo al descrito para crear el archivo *tcw.pdb*. Con este proceso habremos creado los archivos necesarios para el re-docking, ambos en formato .pdb.

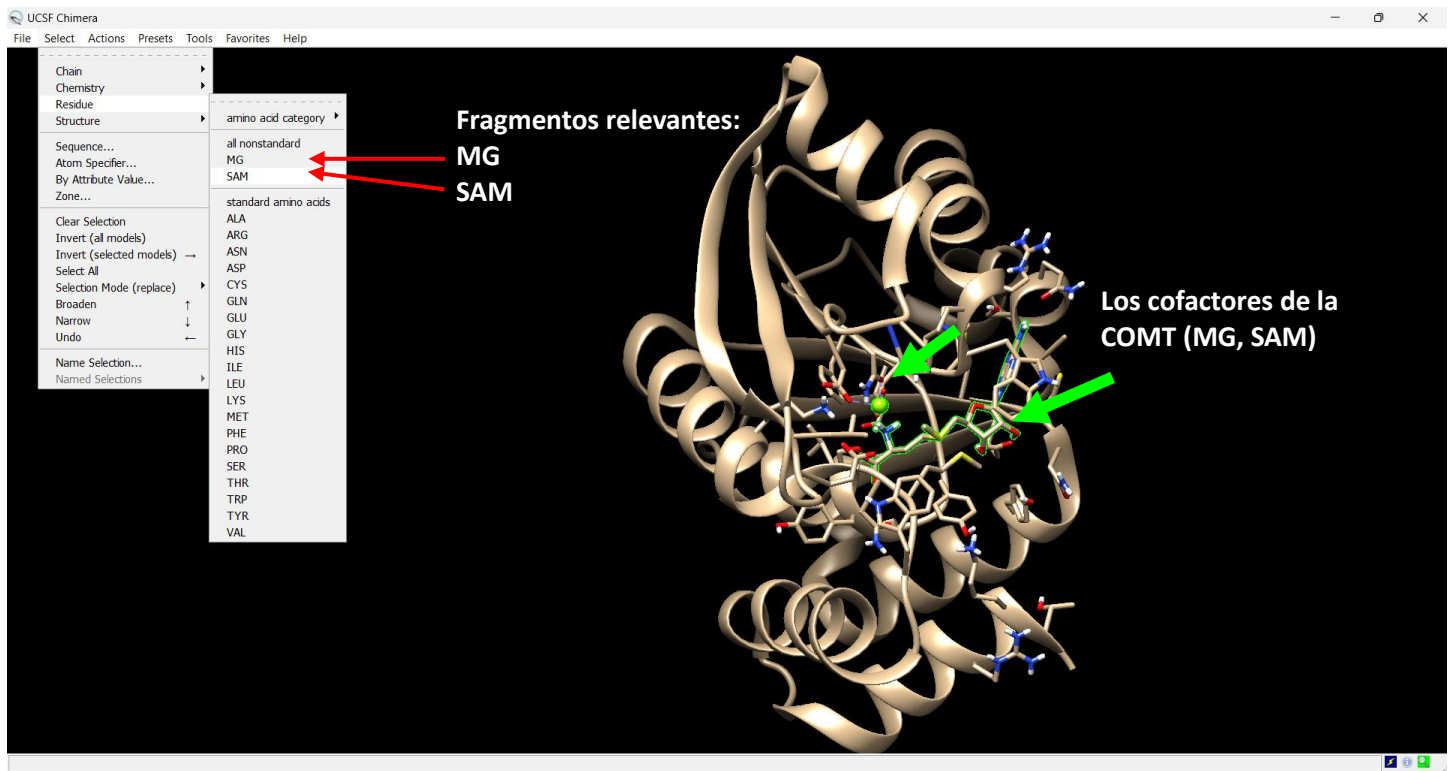


Figura 9. Estructura de la proteína después de borrar todos los FNR. El archivo *comt.pdb* debe contener la parte proteica y los cofactores, el magnesio (II) (MG) y la S-adenosilmetionina (SAM).

3.3. Visualizando los archivos .pdb en Chimera

Con el archivo *comt.pdb* abierto, ejecutamos *File / Open...*, se abre un cuadro de dialogo, en el cual aparece la carpeta de trabajo y los archivos con los que USCF Chimera es capaz de trabajar. Seleccionamos el archivo *tcw.pdb* y damos clic en *Open*. El Tolcapone ahora se muestra en color azul. Ahora pueden observarse las dos estructuras en la misma ventana (Figura 10), el paso siguiente es realizar el re-docking.

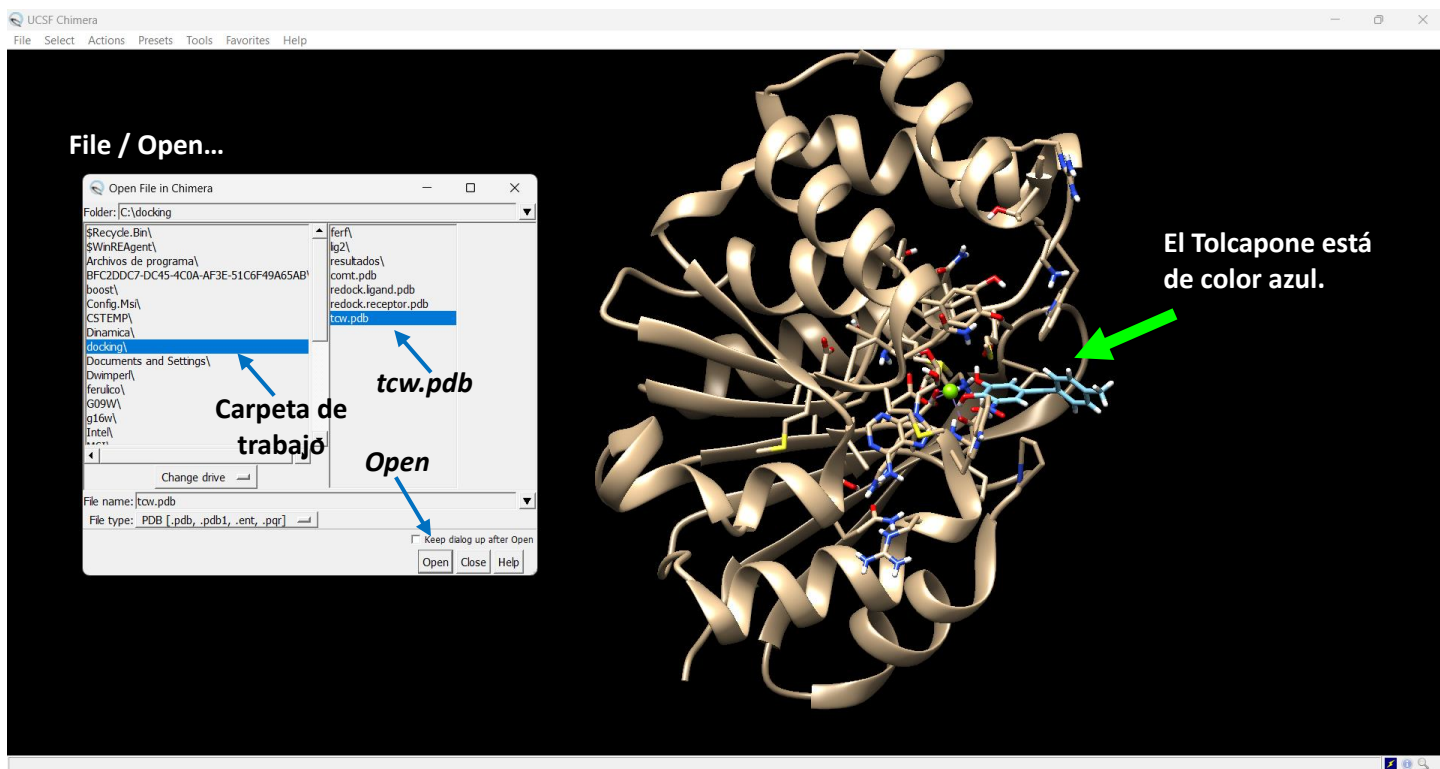


Figura 10. Archivos *comt.pdb* y *tcw.pdb* abiertos en la interfaz de Chimera.

3.4. Re-docking con ADV en la interfaz de Chimera

Chimera posee muchas opciones para realizar diferentes procedimientos, una de estas opciones permite realizar simulaciones de acoplamiento molecular mediante ADV, para esto vamos a ejecutar el menú *Tools / Surface/Binding Analysis / Autodock Vina* y se desplegará un cuadro de texto con las opciones del acoplamiento (Figura 11). Este cuadro de diálogo será utilizado hasta el final de la presente guía y debemos llenar o declarar todos los campos para realizar el re-docking.

Lo primero es nombrar el archivo que se generará después del acoplamiento y declarar su ubicación (se recomienda la misma carpeta de trabajo), esta acción se realiza dando clic en el botón *Browse*, se abre un cuadro de texto como el de la Figura 12. Nombraremos el archivo como *redock.pdbqt*, la extensión del archivo se genera automáticamente si la opción *Add .pdbqt suffix if none given* está habilitada. También, la ubicación de guardado es la carpeta de trabajo (C: \docking), ésta no cambiará en ningún momento a menos que nosotros se lo indiquemos al programa. Salvamos el archivo oprimiendo el botón *Set Output Location*.

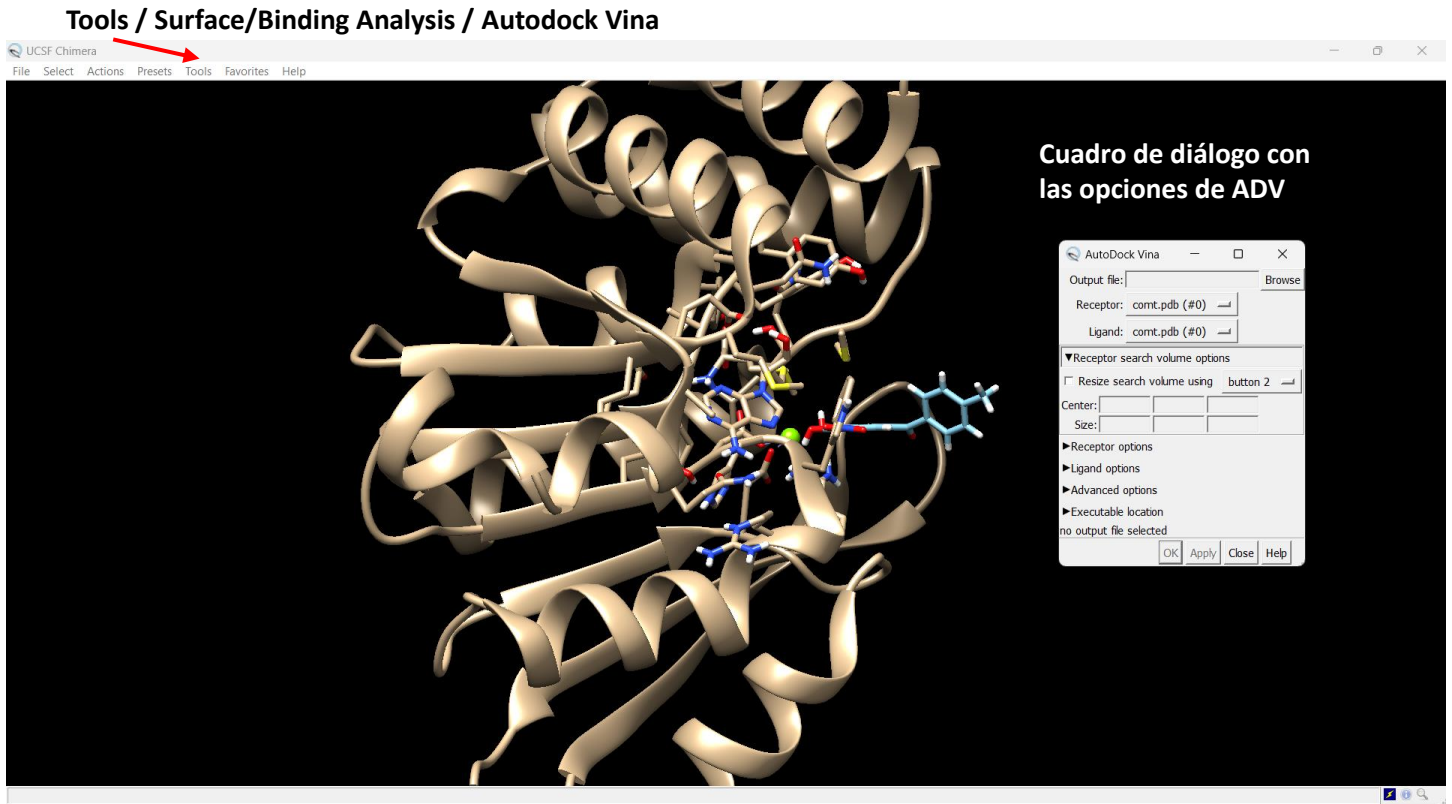


Figura 11. Cuadro de dialogo con las opciones de ADV.

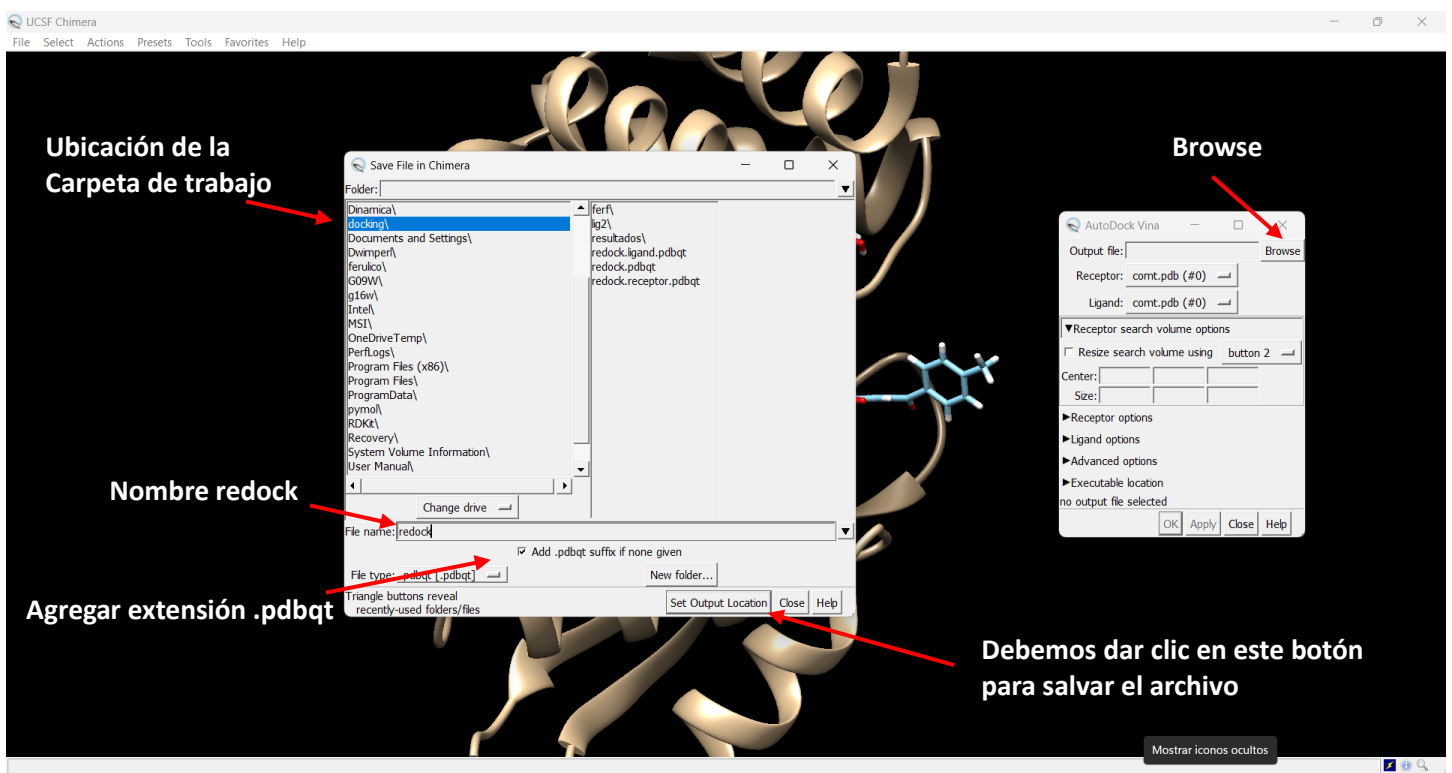


Figura 12. Opciones para nombrar y guardar el archivo de salida del re-docking, *redock.pdbqt*.

Una vez salvado el archivo debe aparecer la ubicación y nombre en el recuadro *Output file* del cuadro de opciones. Con las opciones *Receptor* y *Ligand* le indicamos al programa qué estructura es la proteína (COMT) y qué estructura es el ligando (Tolcapone). Apretando el botón al lado de la palabra Receptor podemos elegir el receptor entre las estructuras (archivos) abiertos en el software. En este caso el receptor está declarado correctamente, sin embargo, el ligando no. Apretando el botón correspondiente, el programa nos da las opciones *comt.pdb (#0)* y *tcw.pdb (#1)*, como es de esperarse, elegimos la segunda de ellas. En la Figura 13 se presenta la descripción anterior. Esta numeración que el software hace de las estructuras en pantalla es importante y más adelante le daremos un uso al calcular el RMSD.

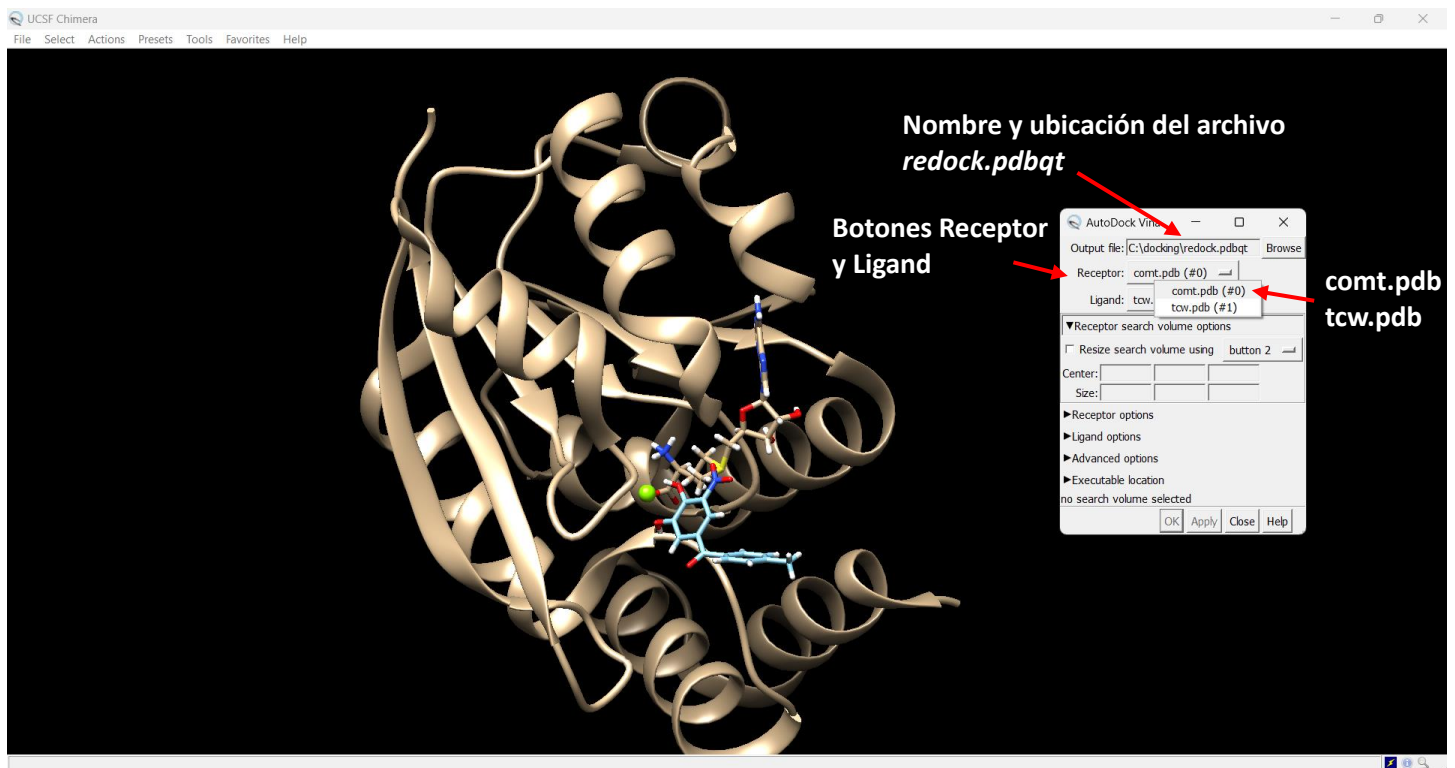


Figura 13. Elijiendo las estructuras del ligando (TCW) y la proteína (COMT)

La parte más importante es la determinación de las coordenadas y el tamaño de la *gridbox*. Este procedimiento se realizará iterativamente, hasta conseguir con una simulación que arroje la conformación más parecida a la experimental (valor de RMSD más pequeño). Con los cuadros *Center* definiremos las coordenadas (x, y, z, en ese orden) y con *Size*, el volumen de la *gridbox*. Modificando sistemáticamente estos parámetros, buscaremos la simulación que mejor reproduzca la conformación del inhibidor co-cristalizado con la proteína. Para facilitar la ubicación de la caja y visualizarla mejor, es recomendable ocultar la estructura secundaria para mostrar solo las estructuras de la SAM, el Mg(II) y el Tolcapone. Para ello se seleccionarán todos los aminoácidos utilizando la función *Select / Residue / Standard amino acids* y se ocultarán con la aplicación de la función *Actions*

/ *Atoms/Bonds / Hide* y finalmente *Actions / Ribbon / Hide*, con lo anterior lograremos esconder la proteína y presentar la pantalla de Chimera como se muestra en la Figura 14.

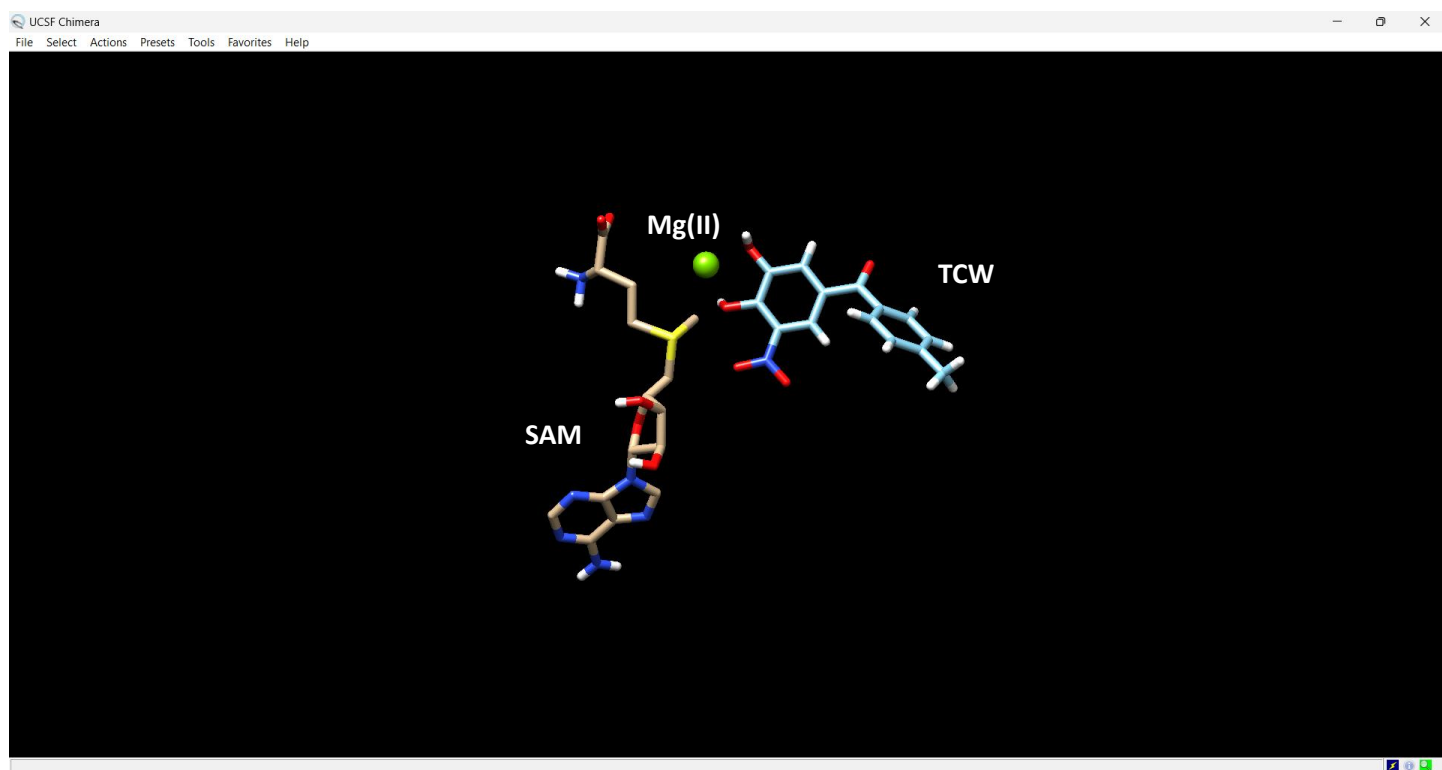


Figura 14. Estructuras del ligando (TCW), la SAM y el Mg(II).

En un primer intento utilizaremos las coordenadas $x = -11.7$, $y = 42.5$ y $z = 65.0$ con un volumen de $16 \times 16 \times 16$ unidades. Una vez definidas las dimensiones aparece la *gridbox* como un cubo virtual con las aristas de color verde. Debemos también escoger las opciones del receptor (*Receptor options*) y las del ligando (*Ligand options*). Las opciones avanzadas (*Advanced Options*) se pueden mantener por default, pero en mi particular punto de vista, es recomendable llevarlas al máximo, en cualquier caso, se deben mantener las mismas condiciones para las moléculas que se desean probar ya en el proceso del acoplamiento con las moléculas a estudiar. Por último, es necesario escribir o buscar la ruta del archivo ejecutable de ADV (*vina.exe*), en el recuadro *Executable location*, en este caso es *C:\docking\vina.exe*. En la Figura 15 se muestran todas las opciones, en el manual *Simulaciones de acoplamiento molecular proteína-ligando con software de libre acceso*, se explicaron detalladamente estas opciones por lo que es adecuado revisarlo para aclarar cualquier duda. Las opciones que tienen que ver con los cofactores (*non-standard residues*) se mantienen como *false* debido a su importancia en el sitio activo de la enzima que estamos trabajando, y por ello debemos tomarlos en cuenta para la simulación.

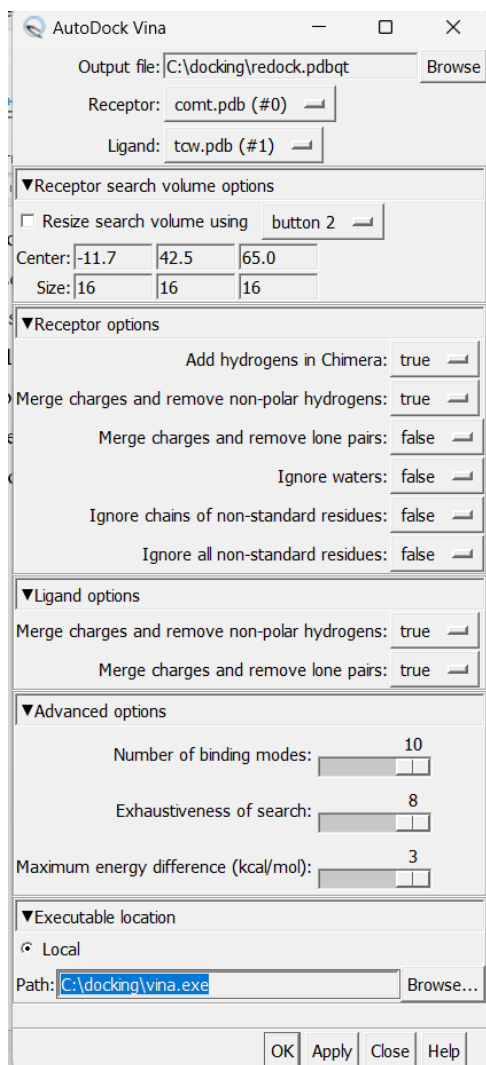


Figura 15. Parámetros del primer intento de re-docking entre la COMT y el Tolcapone.

Con el botón *Apply* iniciamos la simulación del re-docking. En la esquina inferior izquierda de la pantalla, el programa comenzará a dictar varias opciones de escritura y si todo está correcto y completo presentará el mensaje *running*, si no es así mostrará una pantalla con los errores que deberemos arreglar. Después de algunos segundos o minutos quizá, aparecerá una nueva ventana con la leyenda *View Dock - C:\docking\redock.pdbqt* con los resultados del *re-docking* y una nueva estructura de color rojo, esta es la mejor pose que encontró el programa. Esta pantalla se muestra en la Figura 16. La velocidad de la simulación depende entre otras cosas del tamaño de la *gridbox*, el número de enlaces rotables en el ligando y los recursos de la computadora. En este caso ADV encontró que la mejor pose tiene un score de -7.6 Kcal/mol, también se presentan los valores de RMSD entre las diversas conformaciones obtenidas, sin embargo, este valor no es el de interés debido a que se calcula respecto a la mejor conformación encontrada (0.0 Å) y el valor que buscamos es la desviación entre las conformaciones experimental (azul) y roja (simulada).

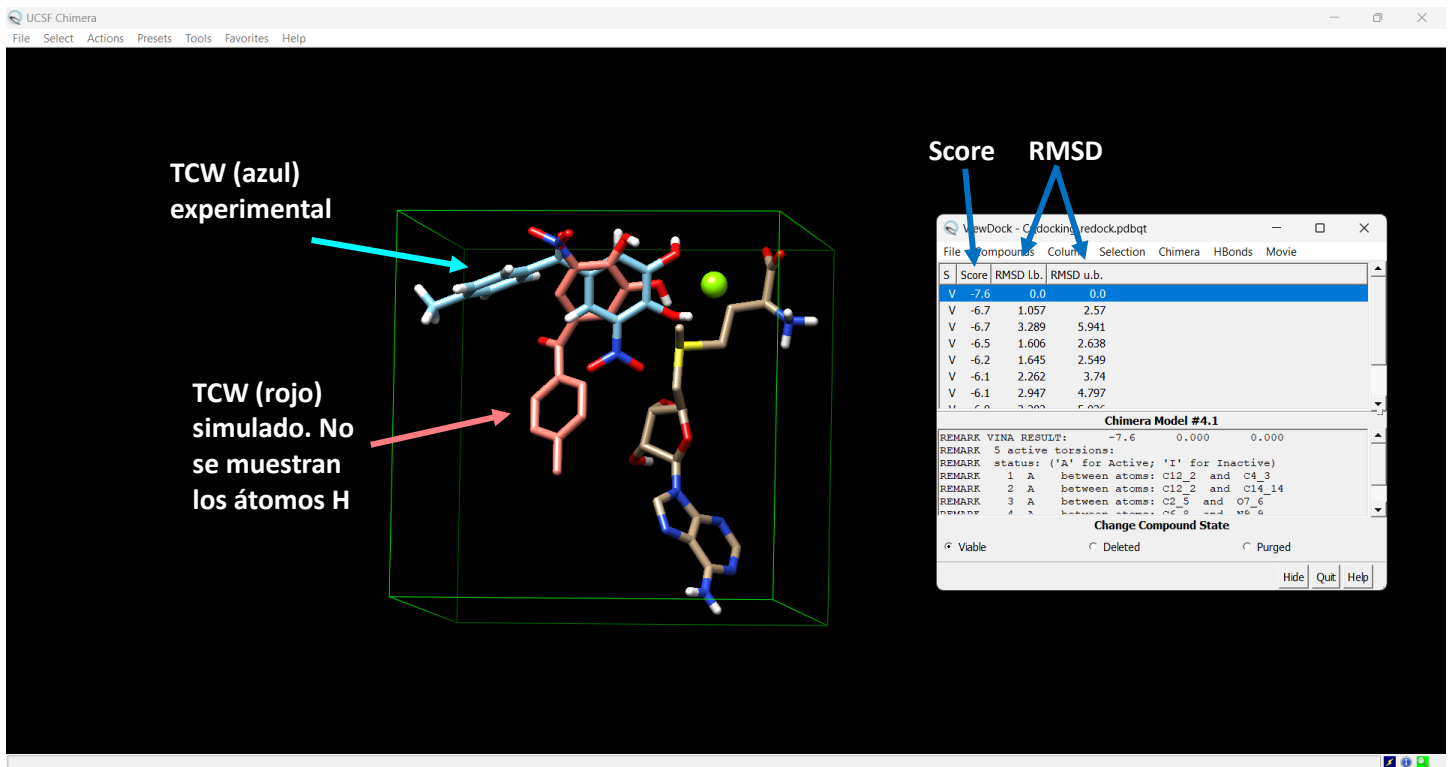


Figura 16. Estructura simulada (roja) y experimental (azul) de TCW. A la derecha de la pantalla se muestra el cuadro de resultados del re-docking.

3.5. Cálculo de RMSD entre las estructuras simuladas y la experimental.

Para el cálculo del RMSD vamos a utilizar la línea de comandos de Chimera, Para desplegarla ejecutamos el comando *Tools / General control / Comand Line*, esta línea de comandos aparece abajo en la pantalla de Chimera (Figura 17). En la figura 16 puede observarse que las estructuras del ligando difieren en los átomos de hidrógeno (en la estructura simulada no se muestran), para calcular el RMSD es necesario mostrarlos, se utiliza la función *Add Hydrogens* con el comando *Tools / Structure Editing / Add H*. En el cuadro de dialogo que se muestra al ejecutar el comando, se selecciona sólo la estructura simulada marcada como *Docked tcw.pdb (#3.1)*. Las demás opciones se mantienen por default excepto *Residue-name-based...* que se tiene que cambiar por *Unspecified* ya que no tenemos algún aminoácido en la estructura de Tolcapone. Al oprimir el botón *OK*, aparecen los átomos de hidrógeno en la estructura simulada. Todas las opciones y la estructura obtenida se muestran en la Figura 17.

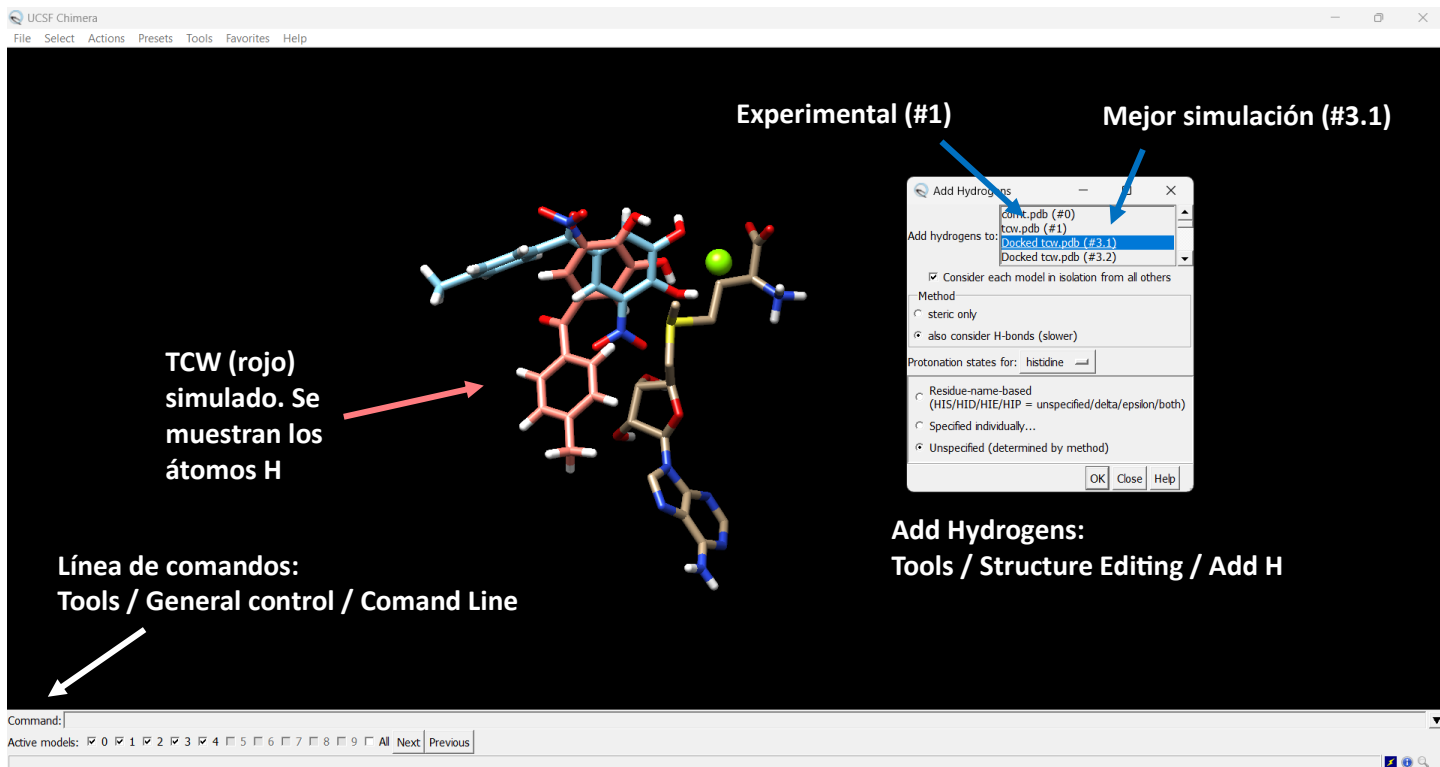


Figura 17. Desplegando la línea de comandos de Chimera y adicionando los átomos de hidrógeno a la estructura simulada

El cálculo de RMSD se efectúa escribiendo en la línea de comandos *rmsd #1 #3.1* (con espacios entre la palabra *rmsd* y cada número de estructura). Al presionar la tecla *Enter* se realiza la comparación átomo por átomo de las estructuras experimental y simulada (mejor pose). El valor de RMSD calculado y el número de átomos comparados, 5.805 Å y 31 átomos respectivamente, aparecen bajo la línea de comandos (Figura 18). Cabe resaltar que se pueden comparar cualquier par de estructuras, siempre y cuando contengan el mismo número de átomos, cambiando los números de las estructuras en la línea de comando. Por ejemplo, se pueden comparar las estructuras de un segundo docking con la instrucción *rmsd #1 #4.1*. El número de estructura puede constatarse con la *opción Select / Chain A*.

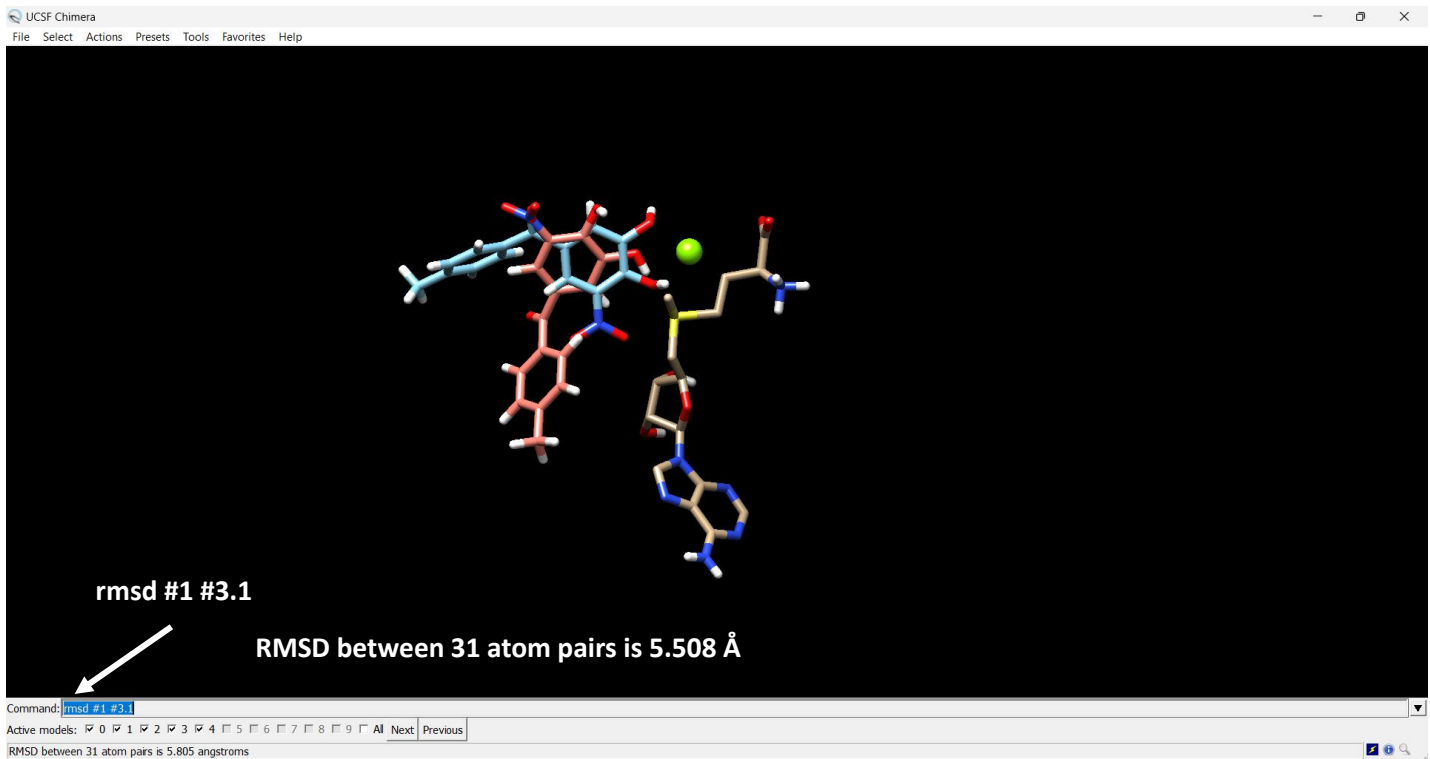


Figura 18. Cálculo de RMSD.

Modificando sistemáticamente las dimensiones y el centro de la caja se pueden obtener mejores valores de RMSD. Por ejemplo, disminuyendo las dimensiones y centrándola en Mg(II) y el TCW se obtiene un valor de 2.9 Å.

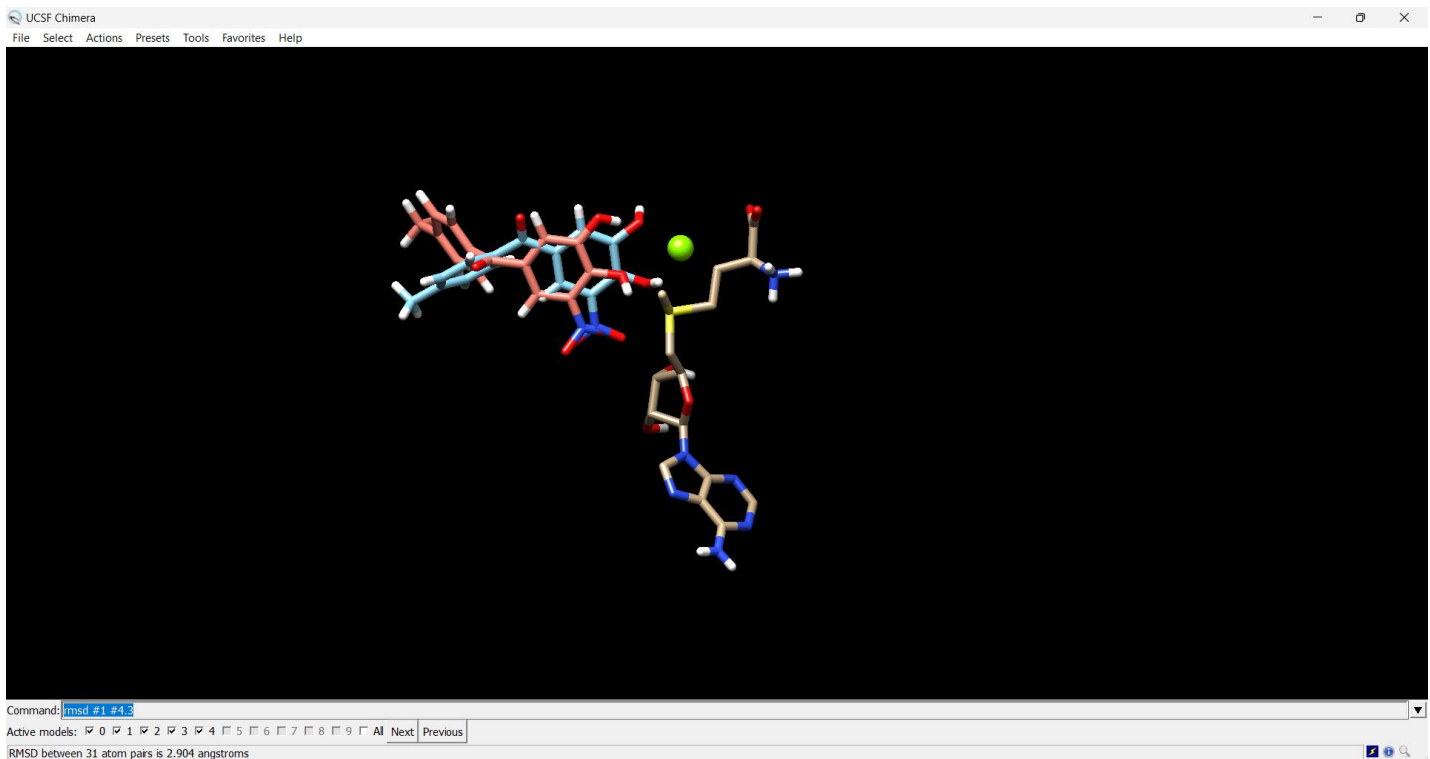


Figura 18. Cálculo de RMSD= 2.9 Å. Gridbox x= 13.7, y= 38.5 y z= 65.0; 15 x 15 x 15.

Es importante mencionar que las dimensiones de la caja pueden disminuirse hasta un cierto límite y será el marcado por el tamaño de las moléculas que queremos acoplar. Los valores aceptables de RMSD son de 2.5 Å con un óptimo de > 1.0 Å. Una vez conseguidos estos valores debemos utilizar las mismas coordenadas del centro y tamaño de caja para realizar las simulaciones de las moléculas de estudio.

En el proceso de re-docking, Chimera y ADV crean varios archivos como el los `pdb` y `pdbqt` del receptor y ligando, así como el archivo con las coordenadas de la caja con extensión `.conf` y el ya mencionado archivo de salida. Estos archivos pueden ser de utilidad en el proceso de docking como tal y se recomienda guardarlos. En la Figura 19 se muestran la carpeta de trabajo con los archivos mencionados.

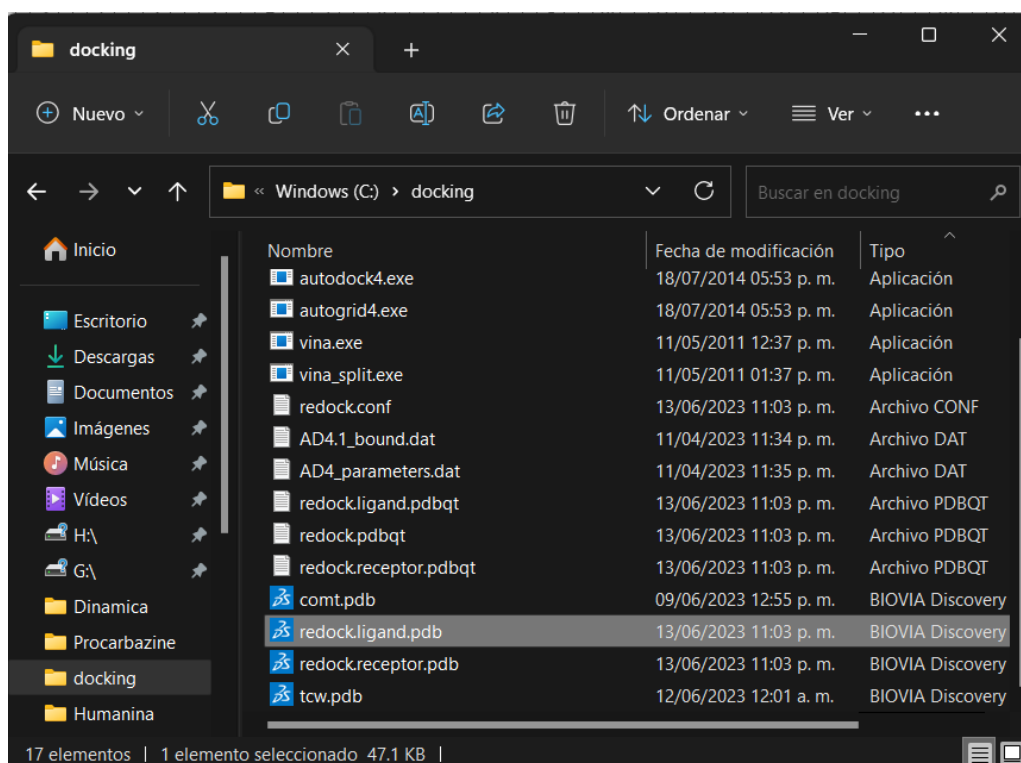


Figura 19. Archivos generados en el proceso de re-docking.

Como conclusión podemos decir que el proceso de re-docking facilita la obtención de la *gridbox* en las simulaciones de acoplamiento molecular, al tiempo que valida la precisión y la confiabilidad de los resultados de los protocolos utilizados. Por último, los resultados de este proceso (valores de RMSD) se reportan comúnmente junto con los valores de afinidades, obtenidos para las moléculas objetivo, en los trabajos relacionados con el acoplamiento molecular. Por ello, es de suma importancia conocer su ejecución.

Agradecimientos

El autor de la guía (CVU: 385566) agradece al Consejo Nacional de Humanidades Ciencia y Tecnología (CONAHCYT) por el apoyo económico otorgado en el marco de la Estancias Posdoctorales por México 2022-1 en la modalidad académica. También se agradece a la Universidad Autónoma Metropolitana Unidad Iztapalapa por el acceso a la Infraestructura para la realización de este proyecto.