



Universidad Zaragoza



Grado en Química

Trabajo de final de grado



Víctor Silva Gallardo

Directores del trabajo

Prof. Dr. Pedro
Merino

Prof. Dr. Tomás
Tejero

Departamento de Química Orgánica

Grupo de Química Biológica y
Computacional

Zaragoza, Septiembre de 2019

Diseño computacional de inhibidores de glicosiltransferasas

Resumen

La química computacional es una de las distintas ramas de la química dedicada al estudio de estructuras y propiedades de un sistema químico basándose en la modelación y simulación proporcionadas por un ordenador. A medida que la potencia de los ordenadores ha ido aumentando, se han podido tratar problemas de creciente dificultad y complejidad. De entre los distintos problemas que la química computacional nos puede ayudar a resolver, nos centraremos en el estudio estructural de una proteína llamada 6EYT y de su interacción con una serie de ligandos propuestos, que actuarán como inhibidores de dicha proteína.

El objetivo será entonces ver si estos nuevos ligandos son capaces de competir con el ligando original por el centro activo de la proteína, si son capaces de unirse y mantener una interacción superior a la del ligando de partida.

Todo esto se realizará de manera computacional, utilizando nuestro conocimiento químico para interpretar los resultados que el programa nos proporcione, y dirigir éstos últimos a la síntesis experimental de inhibidores de glicosiltransferasas.

Índice de contenidos

1.- Introducción	
Métodos computacionales	1
Enzima de trabajo.....	2-6
Antecedentes del grupo.....	6
2.- Objetivos del trabajo	
Uso del software <i>Maestro</i> en <i>Schrödinger</i>	7
Interacción enzima-ligando.....	7-9
3.- Método y obtención de resultados	
Preparación de la proteína.....	9-10
<i>Glide</i> y generación de la <i>Grid</i>	11
Preparación de los ligandos.....	12
Incorporación de los ligandos.....	13
Lanzamiento del Docking.....	14
<i>Scoring function</i>	15
4.-Discusión de los resultados del <i>docking</i>	
Ligando original.....	17-18
Ligando 2OH trans.....	18-19
Ligando 2OH cis.....	19-20
Ligando 3OH trans.....	20-21
Ligando CicloO.....	21-23
5.- Conclusiones.....	24
6.- Bibliografía.....	25
7.- Anexos	

Introducción

Desde el descubrimiento de la diastasa en 1833 por los químicos franceses Anselme Payen y Jean-François Persoz, el estudio y caracterización de enzimas en distintos organismos ha ido creciendo a medida que el conocimiento de las mismas ha permitido su uso en campos como el de la alimentación, la industria farmacéutica y aplicaciones biomédicas. En ésta introducción vamos a desglosar los tres apartados básicos que explican los fundamentos de éste trabajo.

Método computacional

Los métodos computacionales son una potente herramienta de investigación, ya que no solo complementan el trabajo experimental propiamente dicho, sino que permiten ir más allá de los límites del laboratorio. La aplicación de este tipo de programas informáticos es enormemente amplia, y pueden ser usados para la resolución de estructuras moleculares, obtención de espectros, cálculo de energías...

Para la realización de todas las simulaciones, cálculos y posterior obtención de resultados se utilizará un software denominado *Schrödinger*, programa informático que consta de distintas herramientas que permiten facilitar el procesamiento de todos estos datos.

Una de ellas se llama "Maestro"¹, que será el software con el que el programa va a realizar los cálculos de *docking* de todas las interacciones entre la proteína y los ligandos de trabajo.

Por otro lado, existe otra herramienta denominada *Pymol* (integrada en Maestro) muy atractiva visualmente y que permite obtener representaciones tridimensionales que se utilizarán a menudo durante el proyecto.

Los programas informáticos que permiten resolver este tipo de problemas están divididos en tres grandes grupos²:

- Ab initio: utilizan rigurosamente la mecánica cuántica. Son más precisos aunque el coste (tanto económico como en tiempo) es mayor. Además están limitados a moléculas relativamente pequeñas (100 átomos)
- Semi-empíricos: utilizan los métodos ab initio complementados con información empírica que el usuario puede introducir en el programa. Son más baratos y no necesariamente menos precisos. Además permiten resolver moléculas de unos 10000 átomos
- Mecánica clásica: tienen la limitación de utilizar campos de fuerza sin considerar las propiedades electrónicas (ni se forman ni se rompen enlaces)

Maestro cuenta con distintas herramientas basadas en unos métodos u otros, que serán descritos más adelante. Por ahora, bastará con decir que todas estas herramientas serán de utilidad para la creación y preparación de los ligandos, el *screening* de la proteína de trabajo y el estudio posterior para determinar las interacciones entre la enzima y los ligandos.

Enzima de trabajo

Conviene dedicar un apartado que relacione y explique la parte más cercana a la biología celular del trabajo. Iremos desde los aspectos más generales hasta llegar a lo más específico, a saber, la proteína que se ha utilizado.

¿Qué son las enzimas?

Las enzimas son macromoléculas orgánicas de naturaleza proteica (es decir, constituidas a base de subunidades más pequeñas denominadas aminoácidos) que se encuentran en las células trabajando como catalizadores. Al igual que las proteínas, una vez que los aminoácidos se unen entre sí mediante enlaces peptídicos, la enzima adquiere una conformación particular para poder ser activa. Sin embargo, lo que diferencia estructuralmente a una proteína como el colágeno (componente más abundante de la piel y los huesos) de una enzima como la amilasa (presente en la saliva y que degrada el almidón) es la presencia de un “hueco” en su estructura, el cual se denomina centro activo. Este centro activo es el lugar donde otra molécula, en adelante sustrato, se introducirá y quedará retenido, mediante una serie de interacciones que comentaremos más adelante. Se forma así el denominado complejo Enzima/Sustrato, que, tras un cambio en la conformación de la enzima, dará lugar a nuevo complejo, ahora llamado complejo Enzima/Producto, para finalmente liberar el/los producto(s) y dejar a la enzima disponible para una nueva reacción³. En general, las enzimas cumplen las siguientes características:

- Ser catalítica, es decir, aumentar la velocidad a la que se produce la reacción. Aunque no existe un consenso todavía, la velocidad aumenta unas 10^9 veces.
- Alta especificidad: la gran mayoría de las enzimas tienen afinidad por un solo sustrato, o un solo tipo de sustrato. Sin embargo, algunas enzimas, como las encontradas en el complejo citocromo P450, se encargan de transformar más del 50% de los xenobióticos que entran al organismo.
- Gran sensibilidad a condiciones como la temperatura y el pH del medio.

Se estima que unas 4000 de las reacciones que ocurren en el interior de las células están catalizadas⁴.

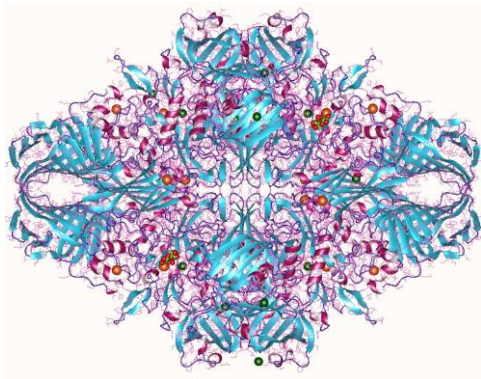


Figura 1. La lactasa es una enzima formada por cuatro subunidades. Mutaciones en ésta enzima causan la intolerancia a la lactosa

Mecanismo de acción

Hemos citado arriba que las enzimas poseen una gran especificidad, ya que son capaces de distinguir estructuras químicas muy similares basándose en la quimioselectividad, regioselectividad y la estereoespecificidad. Para justificar este hecho, en 1894, Emil Fischer⁵ propuso un modelo que explicase esta gran afinidad de las enzimas por sustratos concretos. Habitualmente llamado el modelo de “llave-cerradura”, la idea en síntesis es que el sustrato posee una geometría espacial que encaja en el centro activo de la enzima. Si bien el modelo explicaba por qué un sustrato era reconocido por su enzima, no justificaba por qué se formaba el complejo Enzima/Sustrato, es decir, qué conducía a este complejo a estabilizarse y progresar en la reacción.

Sería más de medio siglo después, en 1958, cuando Daniel Koshland⁶ propuso el denominado modelo de “ajuste inducido”, en el cual es el centro activo el que adopta la conformación más adecuada para el sustrato sólo cuando este se encuentra presente. Una vez formado el completo Enzima/Sustrato, el cambio en la conformación da lugar al producto final.

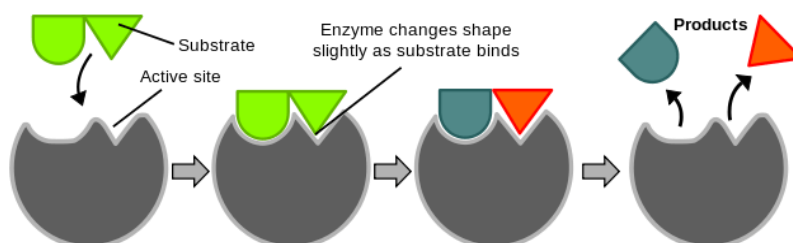


Figura 2. Idea en síntesis del modelo llave-cerradura

¿Por qué estudiar las glicosiltransferasas?

Las glicosiltransferasas son una familia de enzimas que catalizan la transferencia de un fragmento glicosídico de un azúcar activado a un glicósido nucleofílico que actúa como donador. Han sido clasificadas debido a su homología de secuencia en 96 familias diferentes dentro de la CAZy (*carbohydrate Active EnZyme data-base*), dado que tanto el mecanismo de acción como el pliegue previo a su función se mantiene constante dentro de dicho grupo⁷.

De las numerosas enzimas presentes en los microorganismos, las glicosiltransferasas han sido estudiadas por distintos motivos: por ejemplo, se sabe que mutaciones en estas enzimas causan numerosas enfermedades como el cáncer y problemas neurodegenerativos, o que el grupo sanguíneo (según el sistema ABO) viene determinado por el tipo de glicosiltransferasas expresadas por el organismo⁸.

La transferencia de ese residuo de azúcar a un grupo aceptor está también íntimamente ligada a ciertas funciones biológicas, como lo son por ejemplo los distintos mecanismos de señalización celular, desarrollo embrionario ... En general, a cualquier función celular en la que intervenga una glicoproteína, pues la síntesis de las mismas depende, entre otras biomoléculas, de las glicosiltransferasas

Son pues un conjunto de biomoléculas de interés para la biología molecular y para la química orgánica, dado que el entendimiento de su funcionamiento y consecuencias puede llevar a la síntesis de una amplia variedad de fármacos para tratar problemas médicos de distinta índole.

Estructura de la proteína de trabajo

En particular, la proteína con la que este trabajo se ha realizado recibe el nombre de 6EYT, que es una glicosiltransferasa de tipo A (GT-A) aislada a partir de una cepa de la *Salmonella*⁹.

En su forma libre, la proteína se encuentra unida a UDP, N-Acetil glucosamina y manganeso, adoptando el plegamiento habitual de la familia tipo A. Además, es capaz de hidrolizar el UD-NAcGlucosamina sin necesidad de un sustrato para la proteína.

Esta estructura cristalina se trató de obtener mediante difracción de rayos X sin demasiado éxito, ya que los cristales mostraban cierta anisotropía, eran no isoformas y además presentaban poca simetría. Se optó entonces por recurrir a las instalaciones Diamond Light Source, donde se pudo resolver la estructura.

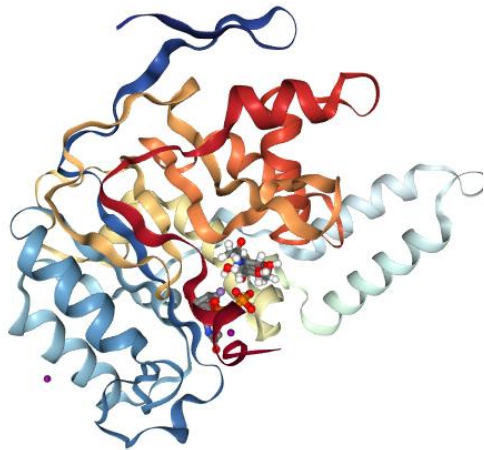


Figura 3. Proteína 6EYT utilizada en los cálculos

La estructura de la 6EYT corresponde a un dímero, esto es, está formada por dos cadenas de proteína. Se ha observado que en la mayoría de los casos, las glicosiltransferasas son dependientes de metales (iones metálicos se entiende) y en particular, los fosfatos de la cadena se encuentran coordinados a metales divalentes.

En nuestro caso el metal consiste en un átomo de manganeso (Mn^{2+}), ya que se observó mediante una calorimetría de titulación isoterma la preferencia del factor Ssek3 por este catión. El Mn^{2+} se coordina adoptando una geometría octaédrica al oxígeno del carboxilato de la unidad D228 y D235 (poner a qué especies corresponde), también al difosfato del UDP así como al hidroxilo de la unidad S327.

El ligando natural que se encuentra en la enzima es el de UDP-GlcNAc, que está coordinado al centro activo de la proteína mediante distintas interacciones. Por ejemplo, el fragmento aromático del ligando (uracilo) forma una interacción *pi-stacking* con el fragmento TRP52 y F190. Puentes de hidrógeno entre los grupos OH de la cadena fosfatada y residuos polares de los aminoácidos son también habituales.

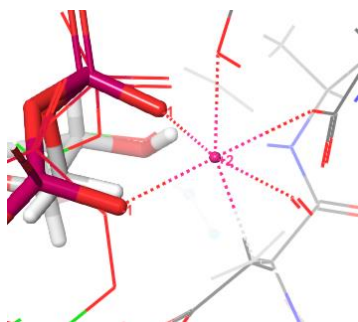


Figura 4. Ion Mn^{2+}

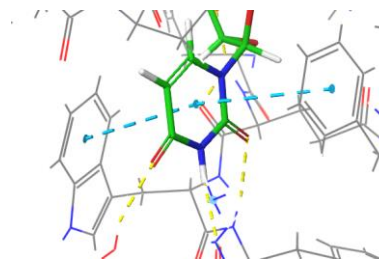


Figura 5. Interacción π - π entre el uracilo (coloreado) y el residuo TRP 52 y PHE 190

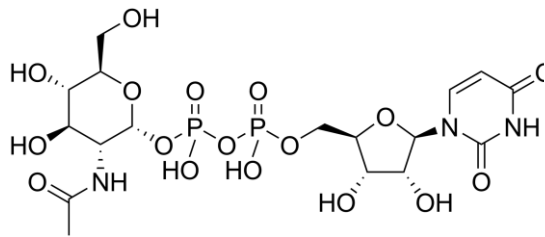


Figura 6. Estructura del UDP-GlcNAc

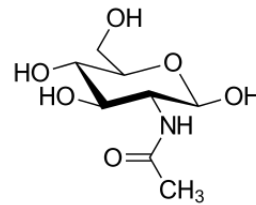
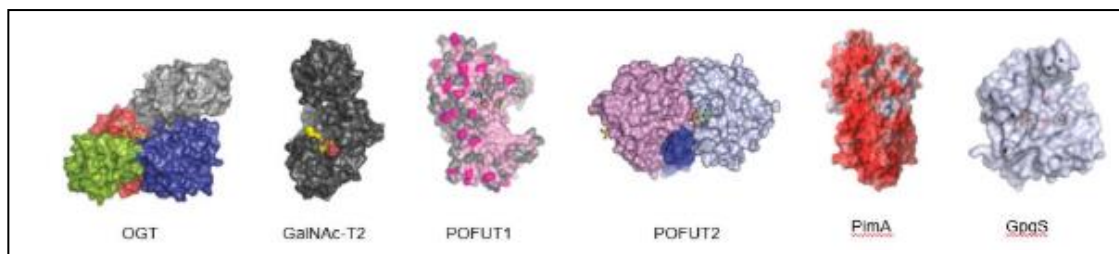


Figura 7. N-Acetilglucosamina (abreviada como GlcNAc)

Antecedentes del grupo

El grupo de investigación de Química Biológica y Computacional de la Universidad de Zaragoza se dedica, entre otros proyectos, al estudio en detalle de las glicosiltransferasas y sus posibles inhibidores, habiendo trabajado con distintas proteínas como la OGT, GalNAc-T2, POFUT1, POFUT 2, PimA y GpgS, y, tras este trabajo, la 6EYT¹⁰.

En particular, la búsqueda de análogos de los sustratos naturales en una serie de glicosiltransferasas. Numerosos trabajos previos trataban de modificar la estructura del pirofosfato o del carbohidrato, no obstante, el acierto de este estudio fue modificar la subunidad de la ribosa y del nucleótido, obteniéndose así prometedores resultados



Objetivos del trabajo

Como ya se ha citado con anterioridad, la final del trabajo va a ser el estudio de la interacción entre la enzima y una serie de ligandos que hemos propuesto. Idealmente, estos nuevos ligandos deberían actuar con el centro activo de la proteína y ser capaces de, al menos, competir con el ligando natural por dicho centro. El resultado idóneo sería obtener ligandos que se unan al sitio catalítico de manera más fuerte que el original

Sin embargo, existe otro objetivo previo a este, a saber, aprender a trabajar con una serie de herramientas informáticas que nos den los resultados que perseguimos.

Uso del software Maestro

El software consta de un conjunto de opciones que a priori puede resultar desbordante. No es la finalidad exponer aquí todo el conjunto de herramientas que se han utilizado, pero sí algunas de ellas que consideramos más importantes así como una descripción general de la interfaz del programa.

- PrepWizard o Asistente de preparación: es la primera herramienta que utilizaremos y que nos permite tratar la proteína, ya que en su forma cruda existen estados de oxidación anómalos, protonaciones inadecuadas y residuos no deseados.
- Receptor Grid Generator: herramienta que nos permite seleccionar el centro activo de la enzima, fijarlo y posteriormente indicar al programa que lleve los ligandos propuestos allí.
- Epik y Macromodel: son dos herramientas complementarias que sirven para realizar el tratamiento previo de los ligandos una vez que los cargamos al programa. Nos permiten generar los estados de protonación adecuados al pH que le indiquemos, así como adecuar las conformaciones o poses de cada ligando de acuerdo a la disposición del centro activo.
- Docking: el paso clave del proceso, donde se utilizan una serie de criterios matemáticos y estadísticos para predecir la orientación óptima de una molécula cuando se une a otra para dar lugar a un complejo estable.

Con estas herramientas se podrán tratar todas las moléculas que metamos en el programa para dejarlas en las condiciones adecuadas. Aprender a utilizar cada una de ellas cuando es necesaria ha sido una parte considerable del trabajo, así como interpretar, utilizando la formación química adquirida hasta ahora, los resultados para elaborar las posteriores conclusiones.

Interacción enzima-ligando

Lo primero que nos podemos cuestionar es la naturaleza química de los ligandos, en particular, su estructura química. Dada la complejidad de la química orgánica, las posibilidades para proponer moléculas son en la práctica ilimitadas.

En la bibliografía existe una amplia diversidad de moléculas capaces de actuar como inhibidores para glicosiltransferasas, aunque en nuestro caso, así como en el de otras muchas enzimas de esta naturaleza, todos los ligandos propuestos comparten una estructura análoga que consiste en:

- Un glúcido (sitio A)
- Una ribosa (sitio B)
- Un nucleótido de uracilo (sitio C)
- Un ribonucleótido (sitio D)

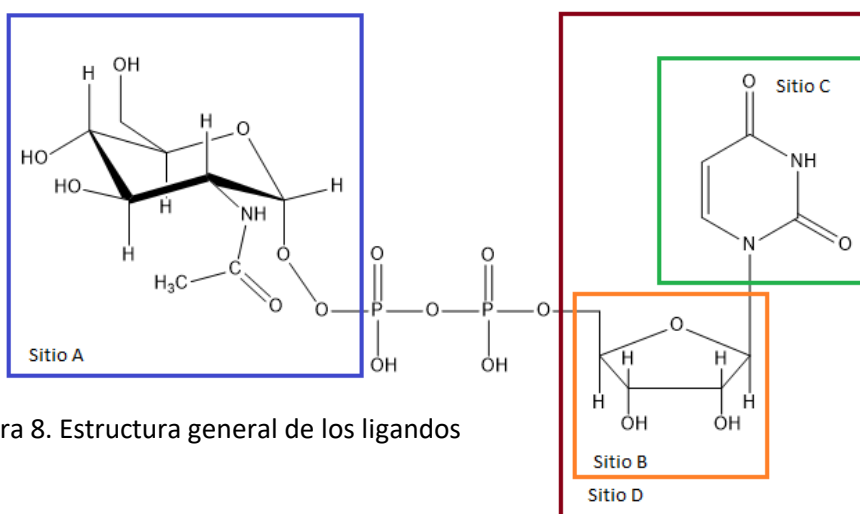


Figura 8. Estructura general de los ligandos

Los estudios estructurales y computacionales han revelado que la unión del sustrato a la enzima se produce principalmente a través de la subunidad del nucleótido, mientras que el residuo azucarado posee más libertad para participar en interacciones enzima-sustrato.

Cabe añadir que se pueden realizar distintas modificaciones sobre el ligando base. Por ejemplo, se puede modificar la subunidad de la ribosa (quitando alguno de los los grupos hidroxilo en los carbonos 2' y 3') o modificar la unidad de la base nitrogenada (sustituyendo el oxígeno por azufre y obteniendo tioderivados). Son posibles otras modificaciones de diferente complejidad que no abordaremos en este trabajo, por lo que nos limitaremos a modificar el azúcar manteniendo el resto de la estructura fija.

En numerosas ocasiones las cadenas fosfatadas son sustituidas por grupos menos polares, ya que la alta polaridad de ambos grupos fosfato ocasiona dificultades para que el compuesto cruce la membrana plasmática. Uno de los ligandos con los que probaremos el *docking* consistirá en sustituir los fosfatos por una cadena hidrocarbonada.

Así pues, tanto el fragmento nucleosídico como la cadena fosfatada serán estructuras constantes en todos los ligandos que expondremos a continuación, ya que las glicosiltransferasas se encargarán de transferir el grupo azucarado del ligando, mientras que el resto de la molécula se encuentra coordinada al centro activo de la proteína.

Lo que conseguimos con este estudio es reducir en gran medida el número de compuestos que habría que preparar en el laboratorio, y de los posibles ligandos que realmente fuesen exitosos, llevar a cabo una selección que nos proporcione el mejor de todos ellos.

Dado el tiempo que conlleva preparar toda una serie de ligandos, lo laborioso que resulta obtenerlos y además el coste económico que supone, no solo el sintetizarlos sino después realizar los ensayos experimentales y estudios posteriores con la enzima, la idea de poder llevar a cabo todos estos pasos mediante simulaciones resulta muy atractiva.

De esta manera, se reduce así en tiempo y en coste económico el trabajo para proponer y diseñar futuros inhibidores tanto de la 6EYT como de otras proteínas.

Método y obtención de resultados

Existe un cierto protocolo consistente y bien desarrollado que permite obtener resultados fiables y que vamos a exponer a continuación¹¹.

Preparación de la proteína

El primer paso es conseguir un fichero informático que contenga la estructura cristalina de la proteína. Este archivo, denominado PDB (por sus siglas en inglés Protein Data Bank) se descarga de una página web que aloja cientos de archivos similares.

Una vez se abre el archivo con el programa Maestro se observa algo como esto:

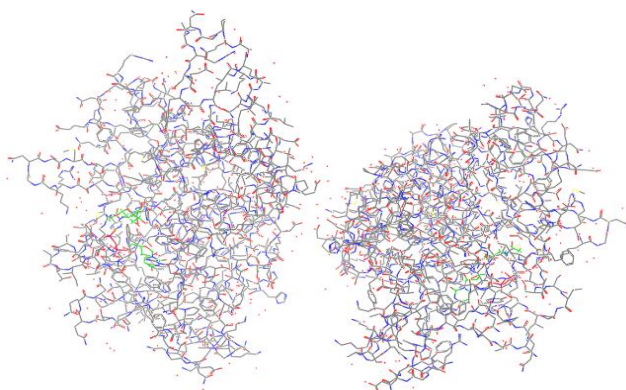


Figura 9. Proteína sin tratar

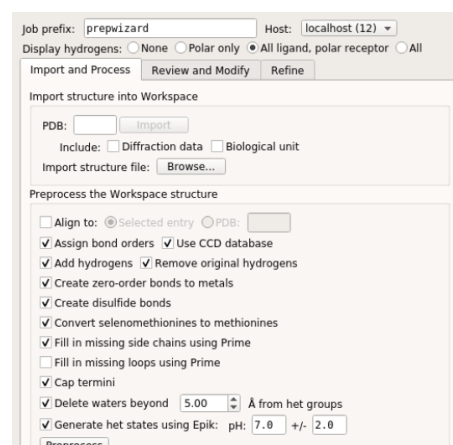


Figura 10 Asistente de preparación

Esta es la forma de la proteína cruda, es decir, sin haber sido tratada. No podemos comenzar a trabajar con la proteína en esta forma porque los cálculos no van a ser exactos y los resultados distarán mucho de los reales.

Como hemos citado con anterioridad, vemos que la proteína tiene dos subunidades claramente diferenciadas, que llamaremos cadenas A y B. La cadena A será la cadena con la que vamos a trabajar, que contiene el centro activo de la proteína.

Para realizar el tratamiento necesario, el programa cuenta con una herramienta llamada *PrepWizard* (Asistente de preparación), que nos permite, entre otras cosas, quitar los hidrógenos originales para que el programa los ponga donde debe, añadir los N y C terminales (*Cap termini*).

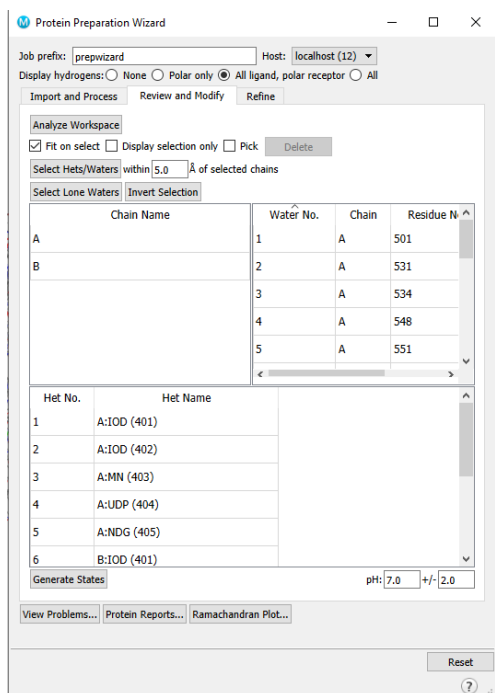


Figura 11

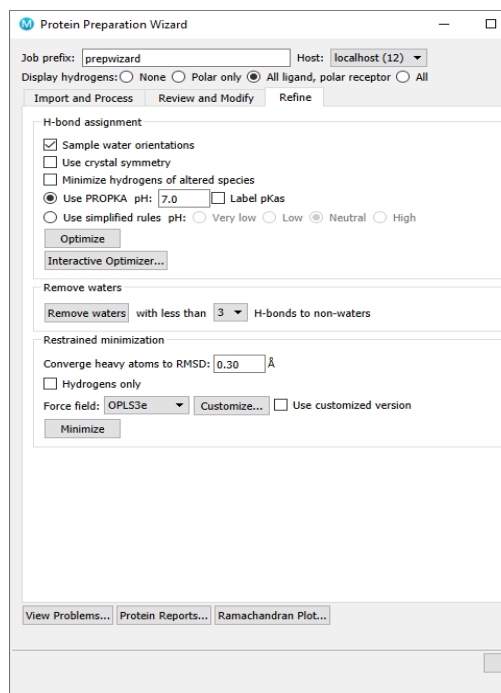


Figura 12

Desde aquí podemos eliminar todos los elementos que no serán necesarios para los cálculos, desde la cadena B de la proteína hasta las aguas residuales, átomos de yodo...

Finalmente, minimizaremos la proteína para que la configuración adoptada sea la que el programa utilice para realizar los cálculos. La herramienta *EPIK* que se encuentra implementada en Maestro se encargará de protonar de forma correcta los aminoácidos, además de eliminar las posibles aguas que no se hayan quitado antes. El campo de fuerza OPLS añadirá las cargas.

Tras el refinamiento de la proteína, podemos superponer la proteína original con la que hemos obtenido después del tratamiento para ver los posibles cambios conformacionales.

Glide y generación de la Grid

El siguiente paso es generar la *Grid*, que nos permite fijar la zona de inhibición de la proteína, esto es, el centro activo¹². Para ello, el programa utiliza la herramienta *Receptor Grid Generation (Glide)*. El software *Glide* que ya viene incorporado en Maestro utiliza una serie de criterios que permiten buscar distintas posiciones para un ligando en un sitio de unión (centro activo).

A medida que el ligando se va acercando a este centro activo va adquiriendo distintas conformaciones de mínima energía, que después son examinadas para poderse orientar de manera adecuada.

El programa genera un cubo como el mostrado en la imagen () con un eje de coordenadas en el interior que indica la orientación del mismo. En el origen de coordenadas se sitúa el ligando original que viene unido a la proteína, ya que ahí es donde se deberán unir los nuevos ligandos que carguemos al programa.

Las dimensiones de la caja son acordes con el tamaño del ligando que queremos introducir.

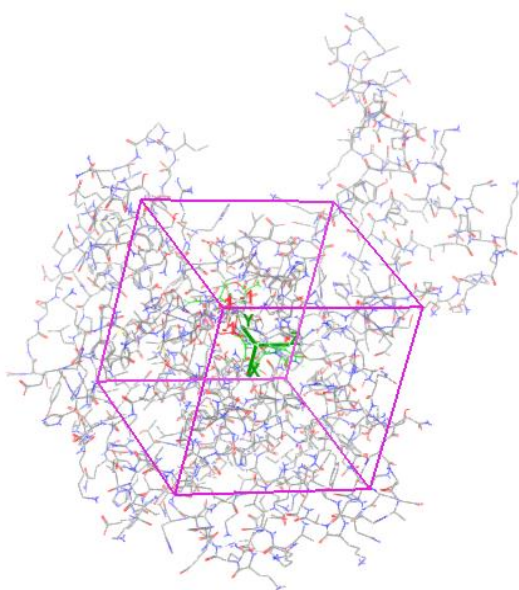


Figura 13

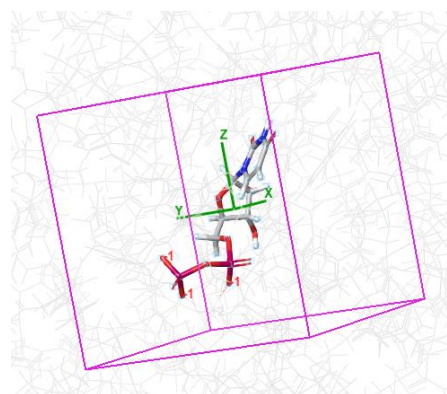


Figura 14

La *Grid* ha sido centrada en la ribosa (sitio A). Una vez tenemos la *Grid* generada, vamos a insertar los ligandos en el programa y tratarlos para realizar el *docking*.

Preparación de los ligandos

La obtención de las estructuras de cada ligando se obtienen dibujándolas con ciertos programas informáticos. Aunque *Schrödinger* trae la opción de poder dibujarlas in situ, es más sencillo utilizar otro tipo de software como *GaussianView* o, en este caso, *ChemDraw Professional*.

Una vez dibujadas (ver anexos), obtenemos el código *SMILES* para poder cambiar el formato de los archivos creados a otro que Maestro entienda (PDB).

Al igual que la proteína, los ligandos no pueden ser introducidos directamente en el programa para lanzar el *docking*, sino que requieren una preparación para adecuarlos a las condiciones en que hemos dejado la proteína.

Para el tratamiento de los ligandos Maestro nos ofrece distintas opciones. En particular, dado que los ligandos con los que estamos trabajando contienen residuos azucarados, utilizaremos dos herramientas integradas en el programa: *EPIK* y *Macromodel*.

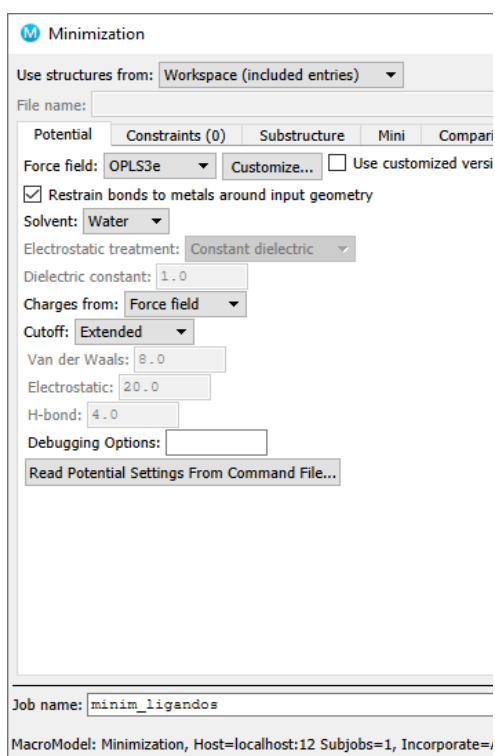


Figura 15

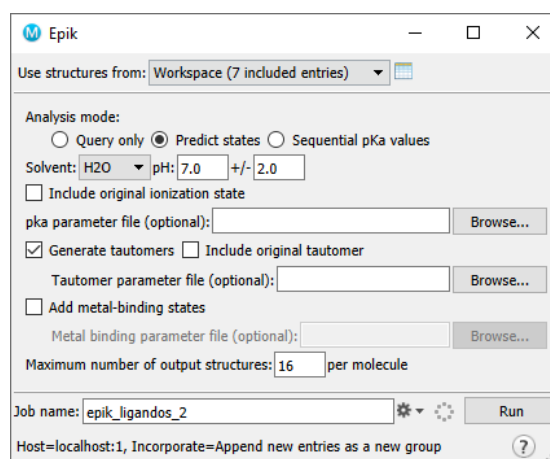


Figura 16

La herramienta *EPIK*, que hemos utilizado antes, nos permite generar los estados de protonación adecuados, eligiendo el rango de pH que consideremos apropiado.

Por otro lado, la herramienta *Macromodel* nos permite minimizar los ligandos y dejarlos en la conformación de menor energía, utilizando un campo de fuerza OPLS3e.

Tras este tratamiento previo, Maestro nos devuelve una entrada en el fichero indicando cómo han quedado los ligandos (disposición espacial, estado de protonación...).

En la figura 17 podemos observar como algunos ligandos apenas cambian de conformación, mientras que otros modifican su orientación para ajustarse mejor al centro activo.

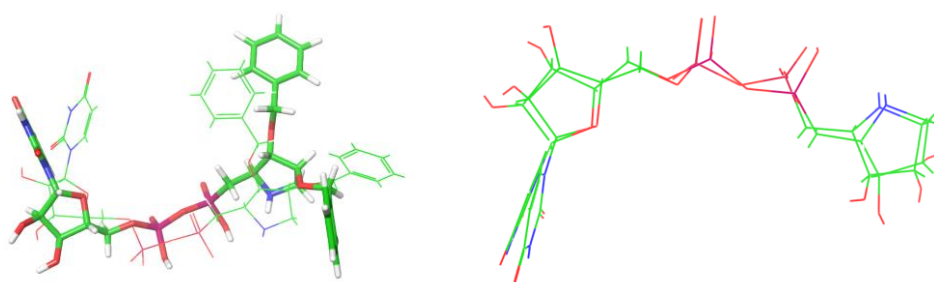


Figura 17. Disposición de los ligandos tras el *Macromodel*

Incorporación de los ligandos

Como ya hemos comentado anteriormente, la estructura de los ligandos presenta un fragmento común que se debe mantener, ya que el ribonucleósido es la *moiety* donde se produce y mantiene la interacción con el centro activo de la proteína.

En los seis primeros ligandos hemos modificado el residuo del azúcar para ver cómo interacciona éste con el centro activo, veremos si cambiando la polaridad, impedimentos estéricos y grupos funcionales podemos lograr una unión mejor al sitio catalítico de la proteína.

Respecto a los dos últimos ligandos, hemos optado por modificar en uno la cadena fosfatada, sustituyéndola por una cadena carbonada de menor polaridad. Además, incluimos también el ligando original, ya que realizaremos un *docking* con la proteína original sin el ligando para ver el *score* entre el ligando cristalizado y la enzima.

Lanzamiento del Docking

El *docking*¹³ es paso fundamental en la simulación computacional. El objetivo que nos permite cumplir esta herramienta es predecir el modo en el que un ligando (sustrato) se va a orientar espacialmente (conocido comúnmente como las “poses”) para interactuar después con el sitio catalítico de la proteína en cuestión.

Para ello, necesitamos siempre dos tipos de información: la correspondiente a la estructura de la proteína o molécula que va a acoger al sustrato y la información necesaria sobre el sitio activo de la misma.

Existe una amplia diversidad de algoritmos disponibles en los que basar los cálculos de *docking*, clasificados de acuerdo a la serie de criterios que utilizan para predecir la conformación y después valorar la interacción mediante una *scoring function*¹³, tema que abordaremos más adelante.

Algoritmo	Objetivo
Local Move Monte Carlo (LMMC)	Búsqueda estocástica
LUDI	Basado en fragmentos
Genetic algorithm	Búsqueda estocástica
Molecular dynamics	Para refinar tras el docking

Tabla 1. Algunos modelos para realizar el docking

Maestro incorpora distintos métodos de docking que podemos elegir de acuerdo a la necesidad del proyecto. Optamos por elegir el método XP (extra-precision), que, aunque necesita mayores tiempos que el método SP (standard-precision) es más potente y capaz de discriminar con mayor fiabilidad.

Aunque normalmente este método es más adecuado usarlo como paso posterior a un docking con SP, debido a que hemos trabajado con ocho ligandos en total no es necesario este paso previo.

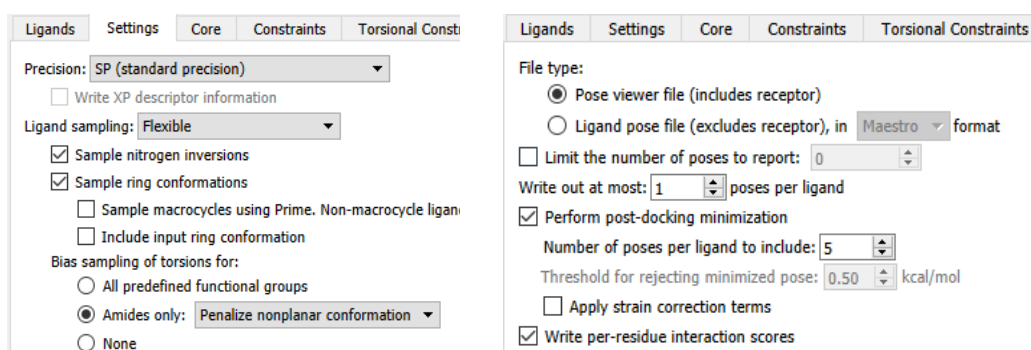


Figura 18

Maestro nos permite incorporar todos los ligandos al mismo tiempo y lanzar el *docking* con todos ellos, lo que resulta menos laborioso que hacerlo de uno en uno. Una vez ha acabado, el programa crea una nueva entrada.

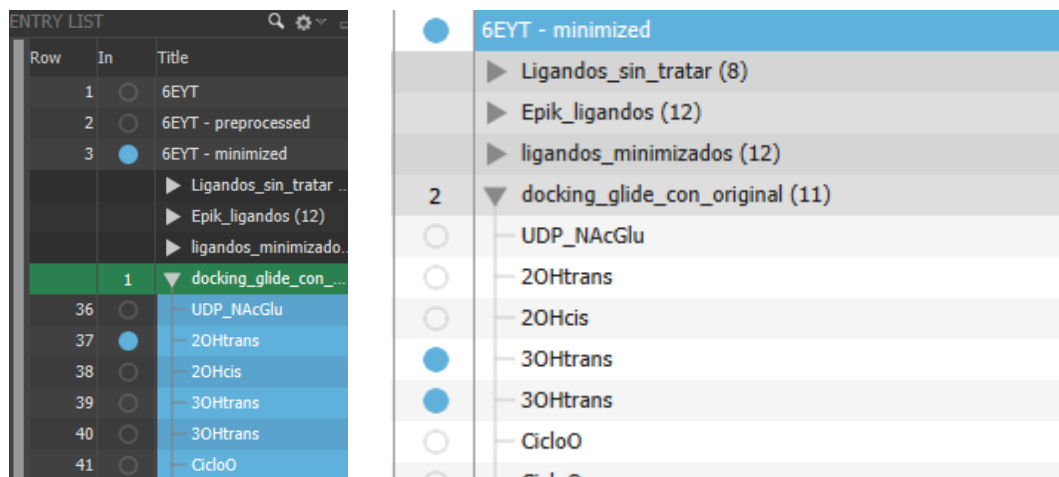


Figura 19

Scoring function

Un aspecto muy importante referente al propio *docking* es la *scoring function*, que nos devolverán en forma de valores numéricos la calidad de la interacción producida. Aunque existen cuatro tipos de funciones, aquí presentamos tres de ellas¹⁵:

- **Force field scoring function:** tiene en cuenta las interacciones físicas como las fuerzas electrostáticas, de Van der Waals, fuerzas de torsión de los enlaces e incluso las entalpías de solvatación
- **Empirical scoring function:** esta función considera las interacciones iónicas, puentes de hidrógeno y los fenómenos de superficie de contacto entre la enzima y el ligando. Los coeficientes de la función se obtienen mediante métodos de regresión lineal. Suelen ser más rápidos que los anteriores (la Glide que hemos utilizado está basada en éste método).
- **Knowledge-based scoring function:** utiliza grandes bases de datos (como el Protein Data Bank) desde donde toma observaciones estadísticas. Aquellas interacciones más habituales encontradas en estos repositorios de datos son a las que confiere mayor puntuación.

Discusión de los resultados del docking

Los cálculos que han sido realizados como se describe en el apartado de los métodos computacionales pueden visualizarse desde la *Project Table* (figura 20) del programa. Una vez ahí, podemos indicar al software que nos clasifique los ligandos de acuerdo al criterio que queramos, por ejemplo, podemos ordenarlos de acuerdo a su energía, a la puntuación de las poses individuales y de su *glide score* (tabla 2). Este último es uno de los dos criterios que utilizaremos en el trabajo para valorar lo bien que el ligando interactúa con el centro activo, comparándolo con el ligando natural que venía cristalizado, siendo el otro los mapas de interacciones que se describen más abajo.

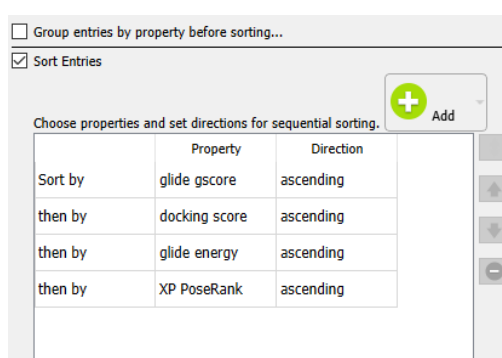


Figura 20

Ligando	Glide gscore (puntuación)
UDP NAcGlu	-22.566
2 OH trans	-20.475
2 OH cis	-18.920
3 OH trans (conf 1)	-18.788
3 OH trans (conf 2)	-18.714
CicloO (conf 1)	-18.311
CicloO (conf 2)	-18.185
2OBn	-12.655
3BnO	-4.244
3BnO carbonado	-0.618

Tabla 2

La tabla 2 es la entrada generada por el programa tras haberle pedido el *score*. Los números correspondientes a cada ligando vienen a agrupar una serie de factores, desde términos de interacción lipófila-lipófila, puentes de hidrógeno, penalización de enlaces que rotan o no, fuerzas de Van der Waals, electrostáticas...entre otras. Las funciones citadas arriba (*scoring functions*) conceden a cada uno de estos parámetros un valor determinado, y evalúan la presencia o no de todas estas contribuciones generando un resultado final, esto es, la *glide gscore* para cada ligando.

Así, todos esos valores nos dan una idea la afinidad entre el centro activo de la enzima y el ligando que aproximemos. En el apartado Incorporación de los ligandos ya indicamos que haríamos el *docking* con el ligando original. Este es el motivo por el que lo hemos hecho, de lo contrario, no tendríamos un valor de referencia, y no podríamos medir la calidad de nuestros resultados sin algo con qué compararlo.

Maestro nos permite crear nuevos criterios para ordenarlos, y uno interesante es el del peso molecular. Es interesante porque podemos después crear un gráfico de dispersión en el que representemos la puntuación (glide score) frente a la masa de cada ligando.

Para ello, creamos un eje de coordenadas donde colocamos, en el eje de ordenadas la puntuación de la glide (glide gscore), y en el eje de abscisas ponemos las masas moleculares (molecular weight).

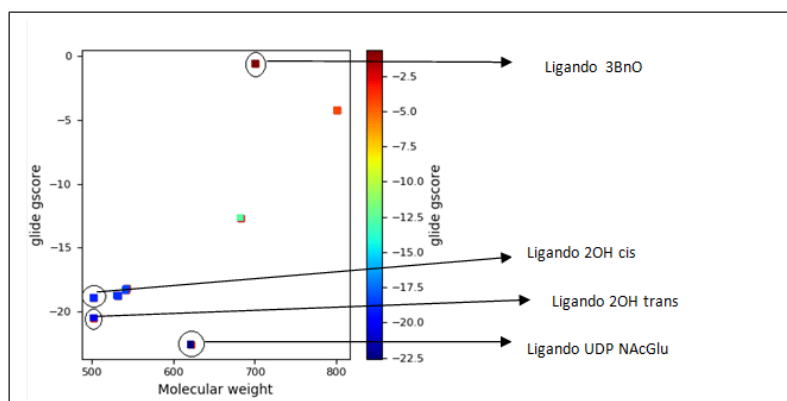


Figura 21

Debe observarse que a pesar de que el *docking* fue realizado con la entrada de los ligandos minimizados (que incluía los ocho ligandos originales más las poses generadas de mínima energía) algunos de los cálculos originan valores tan similares entre sí que aparecen superpuestos en la figura 21.

Observamos entonces que cuanto mayor es el valor del *gscore* en términos absolutos (o cuanto más negativo) mejor es la interacción que el programa concede al ligando.

De todos los ligandos propuestos, vemos que los que mejor puntuación adquieren en la tabla 2 son los ligandos 2OHtrans, 2OHcis, 3OHtrans y el CicloO.

Estudiaremos de manera más exhaustiva cómo interaccionan dichos ligandos con los residuos del centro activo de la 6EYT, además de considerar el ligando natural, que tomaremos como referencia para ver qué cambios son los que han llevado a una unión más o menos intensa.

Ligando original: interacción con el centro activo

Dado que los sitios B y C de la figura 8 son constantes a lo largo de todos los ligandos, estudiaremos una sola vez las interacciones que se producen en tal caso y después abordaremos los residuos azucarados del resto de los ligandos.

El resto de interacciones son las correspondientes al sitio D del ligando, ya comentadas en el apartado anterior.

Puentes de hidrógeno		Interacciones $\pi-\pi$	Puentes salinos	Contactos	
TYR 75		TRP 52	ARG 335	SER 332	TYR 224
SER 333		PHE 190	Mn	SER 333	PHE 190
TRP 334			ASP 226	TRP 334	GLN 51
ARG 335				ARG 335	TRP 52
ASN 259				ASP 228	PHE 53
PHE 53				ALA 227	GLU 54
ASP 191				ASP 226	ARG 55
GLN 51				TYR 228	
ASP 191					
ASN 259					

Es el ligando que mejor compete con el ligando original, además Maestro solo generó una pose para el mismo.

Ligando 2OH cis

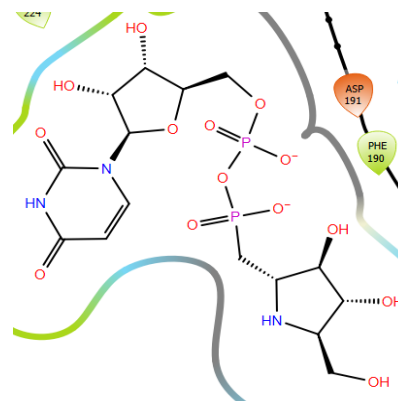
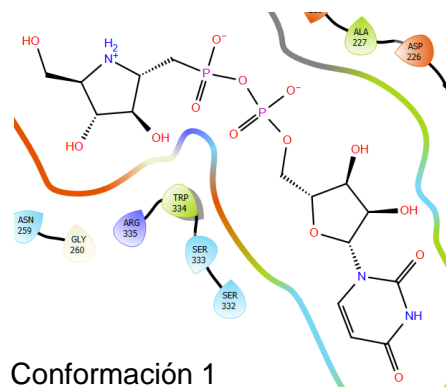
Aquí las interacciones por puente de hidrógeno son notablemente diferentes. Vemos la formación de enlaces de hidrógeno entre los dos grupos hidroxilo de la pirrolidina y los residuos ASP 191 y ASN 259, mientras que en el isómero trans solo uno de los hidroxilos los formaba. Además, ninguno de los protones unidos al átomo de hidrógeno de la pirrolidina forma ahora puentes de hidrógeno, ya que no tiene ningún residuo de aminoácidos cerca (ver anexos). Las interacciones $\pi-\pi$ son las mismas que en el ligando original, y el número de contactos con los aminoácidos adyacentes es algo menor, lo que deja expuesta una parte del ligando.

Puentes de hidrógeno	Interacciones $\pi-\pi$	Puentes salinos	Contactos	
TYR 75	TRP 52	ARG 335	SER 332	TYR 224
SER 333	PHE 190	Mn	SER 333	PHE 190
TRP 334		ASP 226	TRP 334	GLN 51
ARG 335			ARG 335	TRP 52
GLN 51			ASP 228	PHE 53
PHE 53			ALA 227	GLU 54
ARG 55			ASP 226	ARG 55
GLN 51			TYR 228	
ASP 191				
ASN 259				

Ligando 3 OH trans

El caso de este ligando requiere un tratamiento adicional, ya que Maestro nos ha generado dos salidas a la hora de minimizarlo y, por tanto, dos posibles configuraciones a la hora de realizar el *docking*. Vemos que en una de las poses el ligando adopta una configuración similar a la del ligando original, mientras que en la otra adquiere una disposición más alargada.

También es interesante notar que en una de las dos conformaciones el nitrógeno de la pirrolidina está protonado y en otra no, lo que permite a la conformación 1 formar un puente salino adicional con la asparagina 325 (ver anexo)



Aunque la pirrolidina consta ahora de tres grupos hidroxilos, solo dos de ellos forman enlaces de hidrógeno (con GLU 258 y ASN 259). Uno de los hidroxilos de la ribosa forma un enlace de hidrógeno con GLN 51. Vemos que se forma un puente salino adicional respecto del ligando 2OH. La puntuación que el programa le otorga es algo menor que para el ligando 2OH (tanto el cis como el trans).

Puentes de hidrógeno	Interacciones $\pi-\pi$	Puentes salinos	Contactos	
SER 332	TRP 52	ARG 335	SER 332	TYR 224
SER 333	PHE 190	Mn	SER 333	PHE 190
ARG 55		ASP 226	TRP 334	GLN 51
ARG 335		ASP 325	ARG 335	TRP 52
GLN 51			ASP 228	PHE 53
PHE 53			ALA 227	GLU 54
ARG 55			ASP 226	ARG 55
ASN 259			GLY 260	ASN 259
GLU 258			GLU 258	SER 249
TYR 75			ARG 194	

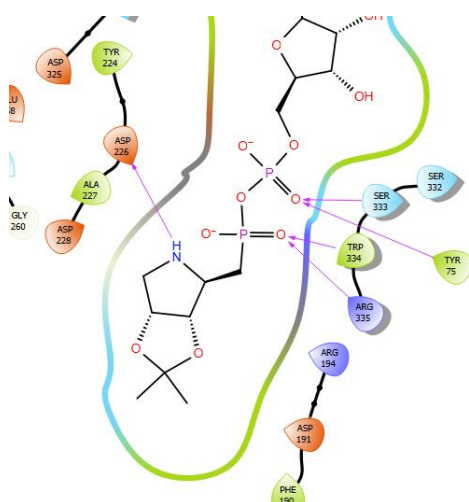
Ligando CicloO

Para este ligando también se nos han generado dos salidas en el fichero, si bien es cierto que ambas encajan considerablemente peor en el centro activo que los otros tres presentados arriba. El número de interacciones por puente de hidrógeno es menor, ya que solo se forma un puente entre el protón unido covalentemente al nitrógeno de la pirrolidina y el residuo ASP 226. Existe puente de hidrógeno adicional entre uno de los oxígenos del fosfato 1 y TYR 75.

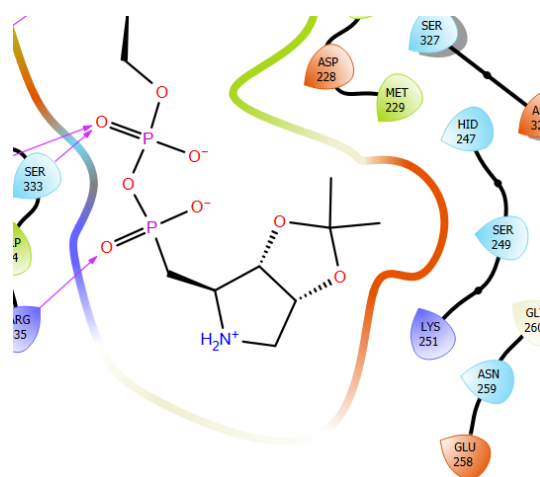
Por otro lado, tanto los puentes salinos como las interacciones $\pi-\pi$ son las mismas que en el ligando original, no se produce ninguna más. Existen dos motivos por el que el ligando presenta menores interacciones entre el sitio A y el entorno del centro activo. La primera es la inferior polaridad de la pirrolidina modificada, ya que no puede actuar como donador de hidrógenos para formar puentes (salvo por el protón del nitrógeno). A su vez, aunque podría aceptar hidrógenos por los átomos de oxígeno

del éter cíclico, no existen cerca aminoácidos polares capaces de entrar a formar parte de tales interacciones.

El segundo motivo es de carácter configuracional, esto es, la disposición en el espacio del ligando hace que el sitio A se encuentre algo más alejado del centro activo de lo que debería, por lo que una parte del ligando no está en contacto con ningún residuo.



Configuración 1



Configuración 2

La configuración 1 puede formar un puente de hidrógeno adicional con el protón del nitrógeno de la pirrolidina, sin embargo, la configuración 2 no puede, y tal como se ve en la tabla de *scoring* (tabla 2) tiene una puntuación menor.

Puentes de hidrógeno	Interacciones $\pi-\pi$	Puentes salinos	Contactos	
ARG 55	TRP 52	ARG 335	SER 332	ARG 55
SER 333	PHE 190	Mn	SER 333	PHE 190
TRP 334			TRP 334	GLN 51
ARG 335			ARG 335	TRP 52
GLN 51			GLU 54	PHE 53
PHE 53			TYR 224	TYR 75
ASP 226			ARG 55	GLU 258
TYR 75			ASN 259	GLY 260
			ASP 191	ARG 194
			TYR 224	TYR 75

En cuanto al resto de ligandos (3BnO, 3BnO carbonado y 2OBn) se ha optado por no estudiarlos a fondo debido a que la puntuación obtenida dista mucho del *gscore* del ligando original y también de los otros ligandos anteriormente expuestos.

Así pues, utilizando tanto los entornos químicos de los inhibidores en cada docking correspondiente como nuestro conocimiento químico, conocemos de manera cualitativa aspectos sobre la estructura química que deben tener los ligandos para considerar prepararlos en el laboratorio.

Si a esto le sumamos la información cuantitativa que Maestro nos da, a través de la *glide gscore* principalmente, podemos crear una idea muy aproximada sobre cómo dirigir la síntesis de inhibidores.

Conclusiones

De los distintos ligandos propuestos, observamos que el ligando 2OH trans es el que mejor interactúa con el centro activo de la enzima, descartando el ligando original. Este ligando introduce un puente de hidrógeno adicional en el residuo azucarado, donde el OH en trans interactúa con el nitrógeno 2 de la glutamina, lo que le aporta una estabilidad adicional frente al ligando natural. Se produce también un puente salino con la asparagina, mientras las interacciones propias de la cadena fosfatada, la ribosa y del UDP son iguales, lo que indica que el sitio D (figura 8) se mantiene donde debe estar, y es el azúcar el que cambia de posición para ser transferida.

Podrían considerarse otras modificaciones tanto en el nucleósido como en la cadena fosfatada, sin embargo hemos visto que disminuir la polaridad de dicha cadena no mejora en absoluto los resultados del *docking* ni la puntuación obtenida en el *glide score*.

Los datos indican pues, que la mejor alternativa es mantener el nucleótido y la ribosa tal como el ligando natural las presenta, y dejando la cadena fosfatada intacta, modificar solamente el residuo azucarado, jugando tanto con la polaridad como con el tamaño. Así, tenemos ligandos como el cicloO, en los que el azúcar presenta una muy baja polaridad en comparación a otros ligandos, y la puntuación del *gscore* es inferior a otros donde se mantienen las interacciones hidrofílicas.

Estos resultados, obtenidos gracias al uso de la química computacional, nos han permitido estudiar la capacidad inhibitoria de una familia de ligandos de estructura concreta, ahorrándonos un coste tanto económico como en tiempo.

Bibliografía

1. <https://www.schrodinger.com/training/maestro11/workshops/an-introduction-to-maestro-11>
2. <http://rsc.anu.edu.au/~agilbert/gilbertspace/uploads/Chem3108.pdf>
3. <https://biologydictionary.net/enzyme-substrate-complex/>
4. <http://book.bionumbers.org/how-many-reactions-do-enzymes-carry-out-each-second/>
5. Koshland, D. E., The key–lock theory and the induced fit theory. *Angew. Chem.Int. Ed.* 1995, 33 (23-24), 2375-2378.
6. <http://www.tulane.edu/~biochem/lecture/gbch701/koshland2.pdf>
7. <https://www.cazypedia.org/index.php/Glycosyltransferases>
8. Structural basis for the glycosyltransferase activity of the *Salmonella* effector SseK3. Esposito D, Günster RA, Martino L, El Omari K, Wagner A, Thurston TLM, Rittinger K.
9. Inhibition of glycosyltransferase activities as the basis for drug development. Schutzbach J¹, Brockhausen I
10. Nucleoside Diphosphate Sugar Analogues that Target Glycosyltransferases, *Asian J. Org. Chem.* 2016, 1413-1427
11. Flexible Ligand Docking with GlideMatthew P. Repasky Mee Shelley Richard A. Friesner
12. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. Richard A. FriesnerJay L. BanksRobert B. MurphyThomas A. HalgrenJasna J. KlicicDaniel T. Mainz Matthew P. RepaskyEric H. KnollMee ShelleyJason K. PerryDavid E. ShawPerry FrancisPeter S. Shenkin
13. Meng XY, Zhang HX, Mezei M, Cui M. Molecular docking: a powerful approach for structure-based drug discovery. *Curr Comput Aided Drug Des.* 2011 Jun;7(2):146-57. Review.
14. Huang SY, Grinter SZ, Zou X. Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. *Phys Chem Chem Phys.* 2010 Oct 28;12(40):12899-908
15. N. Huang, C. Kalyanaraman, J. J. Irwin and M. P. Jacobson, *J. Chem. Inf. Model.*, 2006,46, 243-253
16. Molecular docking. Morris GM, Lim-Wilby M. Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA, USA.
17. Glycomimetics Targeting Glycosyltransferases: Synthetic, Computational and Structural Studies of Less-Polar Conjugates, Ghirardello, M.; de las Rivas, M.; Lacetera, A.; Delso, I.; Lira-Navarrete, E.; Tejero, T.; Martin-Santamaria, S.; Hurtado-Guerrero, R.; Merino, P.

Anexos

Algunas definiciones de interés

Glicósido: son compuestos formados por la unión entre un azúcar simple (como la glucosa) y otra molécula mediante un enlace glicosídico

Enlace glicosídico: es un tipo de enlace covalente por el que se unen un azúcar y otra molécula, que puede ser o no un carbohidrato. Si bien el enlace más habitual incluye un átomo de oxígeno (enlace O-glicosídico) otros átomos como el de nitrógeno, azufre o carbono también ocurren de forma natural.

Glicosiltransferasas: son un conjunto de enzimas que catalizan la formación del enlace glicosídico para generar un glicósido, utilizando azúcares fosfatados (activados) como donores y grupos nucleofílicos como aceptores.

Efactor: son proteínas secretadas por bacterias patogénicas al interior de la célula a la que infectan.

Docking: método de la química computacional que tiene por finalidad predecir la conformación de una molécula y su posterior unión con otra para formar un complejo estable.

Scoring function: son funciones matemáticas usadas para predecir la afinidad de unión entre dos moléculas (enzima y sustrato) tras haberse realizado un docking.

Anexos

Tabla de los aminoácidos de acuerdo a su estructura

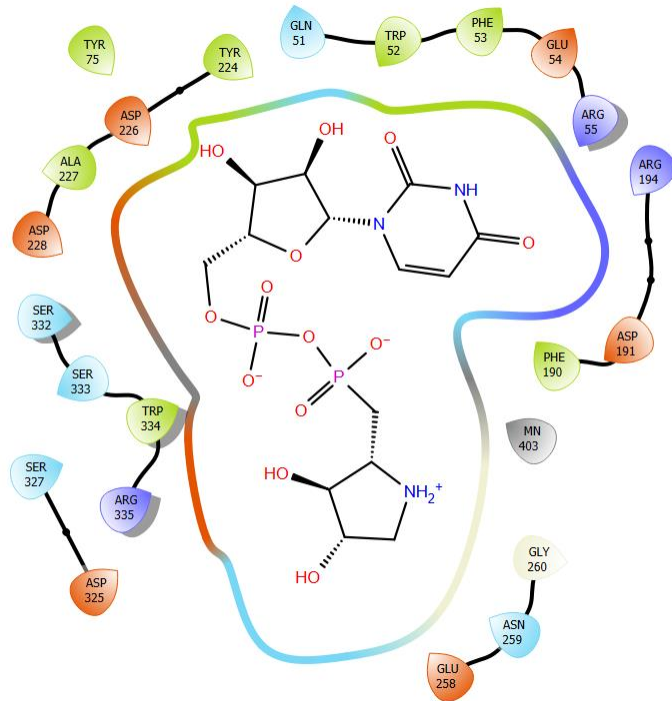
SMALL			NUCLEOPHILIC			
	Glycine (Gly, G) MW: 75.07	Alanine (Ala, A) MW: 89.09		Serine (Ser, S) MW: 105.09, pK _a ~ 16	Threonine (Thr, T) MW: 119.1, pK _a ~ 16	Cysteine (Cys, C) MW: 121.2, pK _a = 8.18
HYDROPHOBIC						
	Valine (Val, V) MW: 117.1	Leucine (Leu, L) MW: 131.2	Isoleucine (Ile, I) MW: 131.2	Methionine (Met, M) MW: 149.2	Proline (Pro, P) MW: 115.1	
AROMATIC				ACIDIC		
	Phenylalanine (Phe, F) MW: 165.2	Tyrosine (Tyr, Y) MW: 181.2, pK _a = 10.46	Tryptophan (Trp, W) MW: 204.2		Aspartic Acid (Asp, D) MW: 133.1, pK _a = 3.9	Glutamic Acid (Glu, E) MW: 147.1, pK _a = 4.07
AMIDE			BASIC			
	Asparagine (Asn, N) MW: 132.1	Glutamine (Gln, Q) MW: 146.1		Histidine (His, H) MW: 155.2, pK _a = 6.04	Lysine (Lys, K) MW: 146.2, pK _a = 10.79	Arginine (Arg, R) MW: 174.2, pK _a = 12.48

No solo la polaridad es importante a la hora de observar las interacciones ente el inhibidor y los residuos que le rodean. Otros aspectos como los puentes salinos, interacciones electrostáticas o los apilamientos- π pueden producir uniones intensas.

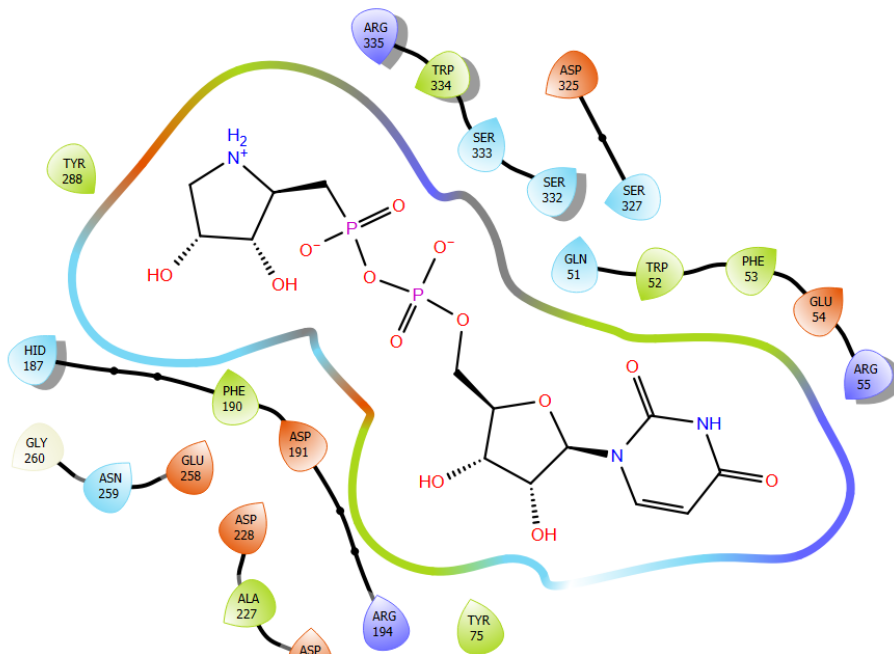
Anexos

Entorno químico de los ligandos estudiados durante el *docking*.

Ligando 2OH trans

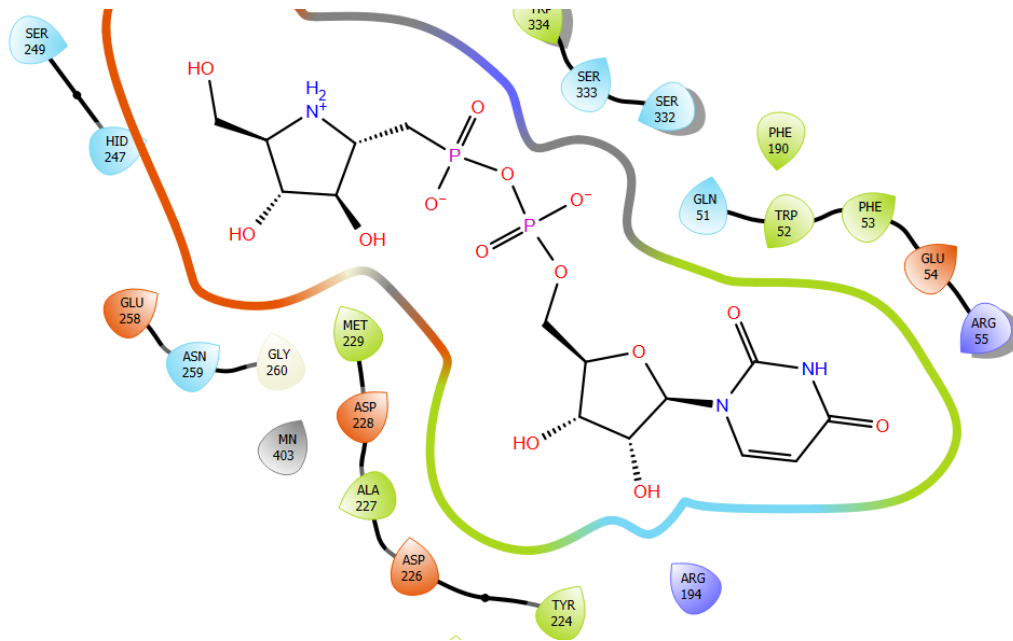


Ligando 2OH cis



Anexos

Ligando 3OH trans



Ligando CicloO

