

# Evaluación de la capacidad bloqueante de nutraceuticos sobre proteínas de entrada virales

**Santiago Treceño Boto**

Máster universitario en bioinformática y bioestadística  
Bioinformática farmacéutica

**Melchor Sánchez Martínez**  
**María Jesús Marco Galindo**

4 de junio de 2018



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

## FICHA DEL TRABAJO FINAL

<b>Título del trabajo:</b>	<i>Evaluación de la capacidad bloqueante de nutraceuticos sobre proteínas de entrada virales</i>
<b>Nombre del autor:</b>	<i>Santiago Treceño Boto</i>
<b>Nombre del consultor/a:</b>	Melchor Sánchez Martínez
<b>Nombre del PRA:</b>	<i>María Jesús Marco Galindo</i>
<b>Fecha de entrega (mm/aaaa):</b>	06/2018
<b>Titulación:</b>	<i>Máster universitario en bioinformática y bioestadística</i>
<b>Área del Trabajo Final:</b>	<i>Bioinformática farmacéutica</i>
<b>Idioma del trabajo:</b>	<i>Castellano</i>
<b>Palabras clave</b>	<i>Proteínas de entrada virales, screening computacional, nutraceuticos.</i>
<b>Resumen del Trabajo (máximo 250 palabras):</b>	
<p>Los virus reconocen a sus células huésped gracias a proteínas de su superficie que son capaces de unirse a determinados glúcidos que se encuentran formando parte del glicocáliz celular. Estas proteínas de reconocimiento constituyen una diana para el desarrollo de medicamentos antivirales de amplio espectro, ya que los elementos celulares a los que se unen son reconocidos por una gran cantidad de tipos virales diferentes, algunos de ellos muy lejanos entre sí desde el punto de vista evolutivo. Para la obtención de compuestos con posibles efectos antivirales se ha llevado a cabo un estudio de los sitios de unión, y mediante técnicas de biología estructural tales como el acoplamiento molecular (“docking”) o el cribado virtual de estructuras (“virtual screening”) se han obtenido compuestos con posible actividad antiviral contra los receptores estudiados. Con los potenciales fármacos obtenidos, se han llevado a cabo análisis comparativos y de viabilidad respecto a sus propiedades farmacocinéticas, para comprobar si estos compuestos podrían usarse como medicamentos. En definitiva, con estos métodos se han obtenido una serie de compuestos que podrían actuar satisfactoriamente contra las proteínas de entrada virales. Pese a ello, sería necesaria una mayor profundización en la aproximación computacional al problema planteado. Además, la realización de estudios experimentales con los compuestos obtenidos sigue siendo fundamental para el desarrollo final de los fármacos, aunque las herramientas computacionales nos permiten orientar el proceso acortando los tiempos de ejecución.</p>	

**Abstract (in English, 250 words or less):**

Viruses recognize their host cells using proteins on their surface that are able to bind certain carbohydrates that are part of the cellular glycocalyx. These recognition proteins are targets for the development of broad spectrum antiviral drugs, since the cellular elements to which they bind are recognized by many different viral types, some of them very far from each other from the point of view of the evolution. To obtain compounds with possible antiviral effects, a study of the binding sites has been carried out, and through structural biology techniques such as molecular docking or virtual screening, compounds with presumable antiviral activity against the studied receptors have been obtained. With these potential drugs, comparative and viability analyzes have been carried out regarding their pharmacokinetic properties, to test out if these compounds could be used as medicines. In short, with these methods we have obtained a series of compounds that could act properly against viral entry proteins. Despite this, it would be necessary to deep in the computational approach to the problem posed. Besides, conducting experimental studies with the compounds obtained is still essential for the final development of the drugs, although the computational tools allow us to orient the process by shortening the execution times.

## Índice

1. Introducción.....	1
1.1 Contexto y justificación del Trabajo.....	1
1.2 Objetivos del Trabajo.....	2
1.3 Enfoque y método seguido.....	3
1.4 Planificación del Trabajo .....	3
1.4.1 Tareas.....	3
1.4.2 Calendario inicial.....	4
1.4.3 Hitos.....	5
1.5 Breve sumario de productos obtenidos .....	5
1.6 Breve descripción de los otros capítulos de la memoria .....	6
2. Materiales y métodos .....	7
2.1 Caracterización de los sitios de unión .....	7
2.2 Evaluación de la conservación .....	7
2.3 Docking de referencia.....	7
2.4 Screening virtual.....	8
2.5 Evaluación de los resultados del screening.....	8
Resultados .....	9
3. Conclusiones.....	19
4. Glosario .....	21
5. Bibliografía .....	22
6. Anexos .....	24
6.1 Alineamientos.....	24
6.2 Código empleado .....	26

## Lista de figuras

Figura 1. Planificación del proyecto .....	5
Figura 2. Hits del screening .....	13,14
Figura 3. Mapa de similitud I .....	15
Figura 4. Mapa de similitud II .....	16
Figura 5. Alineamiento de los receptores de heparán sulfato .....	24
Figura 6. Alineamiento de los receptores de ácido siálico.....	25
Tabla 1. Características de los receptores .....	10
Tabla 2. Características de los sitios de unión .....	10
Tabla 3. Docking de referencia .....	11
Tabla 4. Resultados del screening .....	14
Tabla 5. Resultados del análisis ADMET .....	17

# 1. Introducción

## 1.1 Contexto y justificación del Trabajo

Los virus son organismos acelulares cuya replicación depende de la capacidad de estos para infectar células de otros organismos, u otros virus. Una vez que las partículas virales funcionales (viriones) infectan al huésped, emplean su maquinaria molecular para sintetizar las macromoléculas de las que se compone cada uno de los nuevos viriones, los cuales pueden terminar destruyendo la célula y liberándose al medio circundante.

El ciclo de replicación de los virus comienza con el reconocimiento de las células huésped mediante la unión a receptores específicos de su superficie (tropismo viral), a través de los cuales se induce la entrada del virus al citoplasma. Una vez dentro del huésped se libera el material genético del virus al interior celular y se produce la replicación del material genético y la síntesis de proteínas virales. Por último, se ensamblan los nuevos viriones y se liberan al exterior continuando así con el ciclo viral.

En este trabajo se realizará una búsqueda de nutraceuticos, que sean capaces de unirse a proteínas de la cápsida (en el caso de virus sin envuelta) o a glicoproteínas de la envuelta (en el caso de virus envueltos), que participen en el tropismo viral. Las proteínas diana del estudio son las proteínas de entrada virales, que están implicadas en la unión del virus a elementos extracelulares del huésped, como los componentes del glicocálix.

Actualmente se están desarrollando numerosas vacunas y antivirales, algunos ya en proceso de aprobación por la FDA (Food and Drug Administration)<sup>1</sup>, para luchar contra numerosas pandemias a nivel mundial causadas por virus surgidas durante los últimos años, como ha sido el caso de la gripe A, el ébola, el zika o el síndrome respiratorio de oriente medio (MERS). A pesar de ello la industria farmacéutica no dispone de muchos antivirales, aunque su alto valor añadido hace que estos medicamentos sean muy rentables económicamente. Esto se debe a la dificultad de encontrar dianas efectivas en los virus y a la

enorme capacidad de los mismos para mutar evitando el efecto de muchos antivirales<sup>2</sup>, la mayoría de ellos muy específicos. Por ello, es necesario encontrar sustancias que tengan actividad antiviral de amplio espectro para que sean efectivas frente a distintas especies y subtipos de virus. La elección de los nutraceuticos respecto a otras moléculas se debe a un enfoque práctico, ya que su utilización está menos regulada y los requisitos para su producción y comercialización son menos estrictos.

Algunos nutraceuticos como ciertos oligosacáridos de la leche humana (HMO) han demostrado recientemente ser efectivos contra algunas infecciones virales, al mimetizar la estructura de los residuos de ácido siálico presentes en la membrana plasmática, y que utilizan algunos virus para reconocer al hospedador<sup>3-6</sup>. Además, se están desarrollando antivirales de amplio espectro basados en nanopartículas con análogos de heparán sulfato, diana de muchas proteínas de entrada virales, que mimetizan la estructura de estos proteoglicanos de membrana<sup>7</sup>. Los polímeros de ácidos nucleicos también han demostrado tener una actividad antiviral de amplio espectro al evitar la unión de los virus a las células huésped al mimetizar la estructura del heparán sulfato<sup>8</sup>.

Sin embargo, es necesario seguir profundizando en la búsqueda de compuestos que puedan funcionar como antivirales de amplio espectro, ya que muchos virus utilizan varios receptores celulares distintos para el reconocimiento del huésped<sup>9,10</sup>. Además, es necesario comprobar si existen similitudes entre las proteínas de entrada virales de diferentes especies de virus para poder ampliar el uso de este tipo de compuestos a una mayor cantidad de virus diferentes. En este trabajo, la búsqueda de nutraceuticos está focalizada en encontrar compuestos análogos tanto de los receptores de ácido siálico como de los receptores de heparán sulfato, ya que son las dos vías de entrada más extendidas entre los virus.

## **1.2 Objetivos del Trabajo**

- Caracterización de los sitios de unión virales para el ácido siálico y el heparán sulfato.



- Estudio de la conservación evolutiva de los sitios de unión virales.
- Búsqueda de nutraceuticos que impidan el reconocimiento y la unión del virus a la célula huésped.

### **1.3 Enfoque y método seguido**

Los virus son capaces de reconocer muchos componentes celulares distintos en los huéspedes que van a infectar, en algunos casos muy específicos. Por eso en este trabajo se centra la atención en los receptores de ácido siálico y de heparán sulfato, ya que están ampliamente distribuidos entre muchas especies de virus y son poco específicos. Se podría haber optado por elegir otros receptores como nectinas, integrinas...<sup>10</sup>, pero esto haría que los compuestos encontrados no tuvieran un amplio espectro de actuación.

En PDB, la mayoría de estructuras virales corresponden a viriones completos, proteínas de la cápsida o glicoproteínas de la envuelta. En todos los casos, se trata de proteínas muy grandes por lo que se opta por trabajar con los sitios de unión reduciendo así el tamaño. Si otros recursos computacionales estuvieran disponibles, se podría realizar el estudio con las proteínas completas.

### **1.4 Planificación del Trabajo**

#### **1.4.1 Tareas**

En primer lugar, se han obtenido los sitios de unión de las proteínas, a partir de las estructuras contenidas en PDB (Protein Data Bank<sup>11</sup>) correspondientes a proteínas de entrada virales con ligando cristalográfico. Utilizando los ligandos cristalográficos se ha comprobado por docking que los compuestos diana se unen adecuadamente sus sitios de unión.

Mediante alineamientos de secuencias y alineamientos estructurales se ha determinado si existen secuencias o estructuras conservadas entre distintas especies de virus conocidas.

Para la realización del screening de nutraceuticos, se tiene que elaborar una librería de moléculas en base a la cual se realiza el screening. Esta librería será obtenida de bases de datos de moléculas catalogadas como nutraceuticos aprobados, aunque también sería posible realizar librerías que contuvieran otros compuestos diferentes, o realizar estudios de similitud para intentar obtener nuevos compuestos a partir de las dianas virales. Los compuestos obtenidos a partir del screening se validarán mediante docking. Si las proteínas de entrada virales no son conservadas el screening se hará sobre varias proteínas virales, y si se conservan simplemente se comprobarán los compuestos encontrados para una estructura consenso en el resto de las proteínas. Con los compuestos encontrados, se llevará a cabo una evaluación de sus propiedades farmacocinéticas para determinar la idoneidad de su posible uso para la industria farmacéutica.

#### **1.4.2 *Calendario inicial***

1. Caracterización de los sitios de unión (20/03/18 – 13/04/18)
  - Detección de los sitios de unión (20/03/18 – 29/03/18)
  - Validación por docking (30/03/18 – 13/04/18)
2. Estudio de conservación evolutiva (14/04/18 – 23/04/18)
  - Búsqueda de elementos conservados (14/04/18 – 17/04/18)
  - Validación por docking (18/04/18 – 23/04/18)
  - Clustering (14/04/18 – 23/04/18)
3. Búsqueda de nutraceuticos (24/04/18 – 21/05/18)
  - Screening (24/04/18 – 10/05/18)
  - Validación por docking (11/05/18 – 16/05/18)
  - Evaluación de las propiedades farmacocinéticas (16/05/18 – 21/05/18)
4. Finalización de la memoria (21/05/18 – 31/05/18)
5. Revisión de la memoria (31/05/18 – 05/06/18)

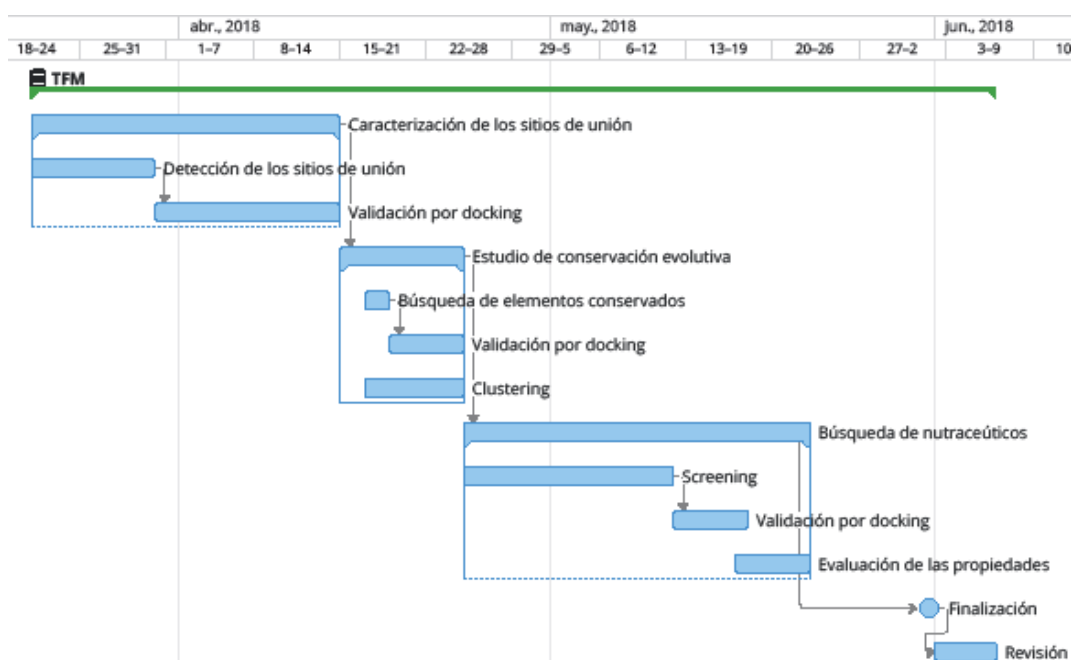


Figura 1. Diagrama de Gantt que detalla la planificación del proyecto.

### 1.4.3 Hitos

En este estudio hay dos puntos críticos principales que son fundamentales e imprescindibles para la consecución de los objetivos. El primero consiste en lograr la caracterización de los sitios de unión del ácido siálico y el heparán sulfato en al menos un virus con envuelta y otro sin envuelta, lo cual se ha logrado antes de llegar a la mitad del tiempo disponible para el desarrollo del trabajo.

El segundo hito y el más importante, es la obtención de nutraceuticos que muestren capacidad para unirse a las proteínas de entrada virales y de esta forma, sean capaces de bloquear el reconocimiento y unión de los virus a las células huésped. Este hito se ha visto algo retrasado, debido a problemas de compatibilidad con las librerías frente a las que se ha realizado el screening, pero también se ha cumplido de forma satisfactoria.

### 1.5 Breve resumen de productos obtenidos

Como resultado del trabajo realizado en este estudio se han obtenido una serie de compuestos químicos que podrían unirse a las proteínas de entrada virales provocando que los virus fueran incapaces de reconocer las células huésped.

Para llegar al resultado final se han desarrollado scripts y se han utilizado programas cuyos códigos se pueden encontrar en el anexo de la memoria.

### **1.6 Breve descripción de los otros capítulos de la memoria**

En el cuerpo de la memoria se detallan los materiales computacionales utilizados para el desarrollo del proyecto, así como los resultados obtenidos a partir de la realización del mismo. Por razones de espacio, muchos de los datos y el código con el que se ha realizado parte del proyecto se encuentran disponibles en el anexo de la memoria.

## 2. Materiales y métodos

### 2.1 Caracterización de los sitios de unión

Las estructuras estudiadas se han elegido entre las estructuras disponibles en PDB que contienen un ligando cristalográfico igual o similar al que se unen biológicamente (ácido siálico y heparán sulfato), intentando evitar duplicidades en la medida de lo posible, para tener proteínas de unión representativas de diferentes grupos de virus. Dichas estructuras tienen un ligando (cocrystal) que está unido al sitio de unión objetivo para nuestro estudio.

Para la caracterización de los sitios de unión se ha desarrollado un script de Python 2.7.14 (Jupyter Notebook 5.4.1), que utilizando el paquete Bio.PDB de la librería Biopython 1.70<sup>12</sup>, detecta los residuos que estén separados del ligando por menos de 6 Å. Estos residuos serían los componentes del sitio de unión y sus cercanías. A partir de las coordenadas de los residuos y del ligando, se han calculado las coordenadas del grid de interacción y el tamaño del mismo. Además, se ha extraído la estructura y la secuencia de esta región de unión al ligando.

### 2.2 Evaluación de la conservación

Las estructuras de las cadenas de las proteínas que interaccionan con los ligandos de interés han sido comparadas mediante un servidor web de alineamiento múltiple de secuencias proteicas (PROMALS3D<sup>13</sup>). El output del programa se ha extraído en el formato especial: "Alineamiento coloreado con estructura secundaria, conservación e información de secuencia consenso", utilizando los valores por defecto.

### 2.3 Docking de referencia

Para realizar los docking moleculares de los receptores objeto de estudio a sus ligandos cristalográficos se ha utilizado el servidor web MTiAutoDock 1.0<sup>14</sup>, especificando en el input los archivos en formato ".mol2" del ligando y en

formato “.pdb” del receptor. Como centro del sitio de unión se han dado los residuos que interaccionan con el ligando según PDBsum<sup>15</sup>, que son los que se encuentran a menos de 4 Å de distancia, en una lista que el servidor utiliza para calcular los parámetros de la cavidad de unión.

Los archivos en formato “.mol2” de los ligandos protonados a pH fisiológico se han obtenido a partir de archivos “.sdf” utilizando babel 2.3.2 <sup>16</sup>.

## **2.4 Screening virtual**

Para obtener compuestos químicos capaces de unirse a los sitios de unión de las proteínas del estudio se ha realizado un screening virtual utilizando el servidor web MTiOpenScreen<sup>14</sup>, basado en AutoDock Vina. En el input se han especificado los receptores en formato “.pdb” y la librería de compuestos a utilizar por el screening, en este caso la librería “Diverse-lib”. Como centro del sitio de unión se ha dado la lista de residuos que lo componen, de la misma manera que para la realización del docking.

La librería “Diverse-lib” está conformada por compuestos obtenidos a partir de la base de datos de PubChem BioAssay, que han sido filtrados eliminando los compuestos tóxicos o con características farmacológicas no deseadas, y asegurando la diversidad química en la librería<sup>14</sup>.

Utilizando los mejores “hits” obtenidos con el screening, se ha realizado un docking de comprobación empleando MTiAutoDock 1.0 con el ligando protonado a pH fisiológico en formato “.mol2”, el receptor en formato “.pdb”, y la lista de residuos que conforman el sitio de unión del receptor.

## **2.5 Evaluación de los resultados del screening**

Con los “hits” del screening anterior se han obtenido imágenes en dos dimensiones de las moléculas obtenidas. Para ello, se ha realizado en primer lugar la transformación a formato “.sdf” de dos dimensiones de los resultados del screening, en formato “.mol2”, utilizando babel 2.3.2.

A partir de las moléculas del output del screening en formato “.sdf”, se ha empleado la librería RDKit 2017.09.3 de Python 2.7.14 (Jupyter Notebook 5.4.1), para realizar dibujos en dos dimensiones de las mismas y mapas de similitud.

A partir de los SMILES (“Simplified Molecular Input Line Entry Specification”) de los primeros hits del screening, se ha llevado a cabo también un análisis ADMET (“Absorption, Distribution, Metabolism, Excretion and Toxicity”) utilizando el servidor web pkCSM<sup>17</sup>. La finalidad del ADMET es comprobar que las características farmacocinéticas de los compuestos obtenidos a partir del screening son compatibles con su posible uso como fármacos.

## Resultados

En la siguiente tabla se disponen los datos correspondientes a la información acerca de las características del receptor para todas las estructuras del estudio. Como se puede ver en la Tabla 1, se dispone de dos estructuras correspondientes a virus que usan el heparán sulfato como ligando de unión. El ligando cristalográfico utilizado para caracterizar la unión es el ácido glucurónico (IDS), que es uno de los componentes principales del heparán sulfato junto con la N-acetilglucosamina. Sin embargo, es más específico que esta última para el heparán sulfato, por ello se ha decidido escoger ese ligando. Ambas estructuras corresponden a proteínas de la cápsida de virus no envueltos.

Para los virus que utilizan el ácido siálico como ligando de unión, se dispone de cinco estructuras. Dos de ellas corresponden a proteínas de la cápsida de virus no envueltos y las otras tres corresponden a glicoproteínas de la envuelta de virus envueltos. El ligando utilizado para caracterizar la unión es el ácido siálico (SIA) o derivados oligoméricos del mismo.

id_PDB	Especie	Tipo	Ligando_SIA	Ligando_HS
1ZBA	Foot-and-mouth disease virus	No envuelto	n/a	IDS
2XBO	Equine rhinitis A virus	No envuelto	SIA	n/a
3CL5	Bovine coronavirus	Envuelto	SIO	n/a
4BSH	Influenza A virus	Envuelto	SIA	n/a
4C7W	Murine coronavirus	Envuelto	SIO	n/a
5BNN	Enterovirus D68	No envuelto	4U0	n/a
5W1O	Human papillomavirus	No envuelto	n/a	IDS

Tabla 1: Características de las proteínas utilizadas como receptores

En la Tabla 2 se exponen las características de las cavidades de unión de cada una de las estructuras. Como se puede ver en la tabla los residuos que conforman los sitios de unión difieren mucho en número y en características. Los residuos más abundantes entre todos los sitios de unión son la serina, la glicina y la leucina. Respecto a los tipos de aminoácidos hay mayor número de aminoácidos cargados o polares, que apolares. Cabe destacar que no hay ningún aminoácido azufrado en los sitios de unión de ninguna de las estructuras, a pesar de que la cisteína es un residuo muy abundante en los sitios de unión de las proteínas.

Coordenada X	Coordenada Y	Coordenada Z	Tamaño grid	Residuos
32,3700027	22,0305004	144,5144958	15x20x15	KVT
-8,8210001	231,3265076	224,4355011	16x15x16	ETTAQL
22,0819988	13,7814999	19,5340004	21x18x20	KFLSNTLL
36,1360016	-35,6805000	36,0774994	23x20x19	NGATSAHHSGKLQSGR
-8,6429996	-4,2124996	2,9839997	19x18x17	HSKGSSSYL
-42,4189987	-34,5804977	129,4620056	20x27x24	APNIIGNQLDGPDIG
24,0789986	13,9980001	4,3150001	18x23x15	VPKVQYRDGDMSI

Tabla 2: Caracterización de los sitios de unión

En cuanto a los alineamientos realizados entre las secuencias de las cadenas en las que se encuentra el ligando, se han detectado ciertas estructuras secundarias conservadas como hélices alfa y hojas beta. Por el contrario, la identidad de secuencia entre las distintas proteínas es casi nula. Además, la longitud de las secuencias es muy diferente en cada caso, por lo que encontrar similitudes es muy complicado en este tipo de proteínas. En el anexo se adjuntan las imágenes de los alineamientos múltiples.



De forma preliminar, se han realizado docking estructurales utilizando los ligandos cristalográficos para que los valores de afinidad de los mismos sirvan como referencia para los resultados del screening, donde se buscan compuestos que sean capaces de unirse potencialmente a los sitios de unión elegidos.

En la Tabla 3 se muestran los valores de afinidad (energía libre) de la mejor pose encontrada por MTiAutoDock 1.0, para cada una de las estructuras. Estos valores servirán más adelante como valor de referencia.

<b>id_PDB</b>	<b>Energía</b>	<b>Torsiones</b>
1ZBA	-7,97	8
2XBO	-8,37	10
3CL5	-7,52	13
4BSH	-9,33	10
4C7W	-8,65	13
5BNN	-12,89	22
5W1O	-7,30	8

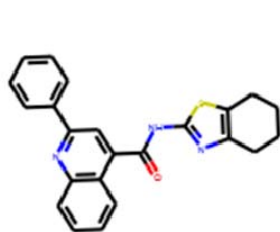
Tabla 3: Energía libre del acoplamiento y número de torsiones para la mejor pose obtenida por MTiAutoDock 1.0

A partir de la realización del screening virtual, se ha obtenido una lista de posibles fármacos que se unen potencialmente a los sitios de unión de las proteínas objeto de estudio. En la Figura 2 se presentan las estructuras y los nombres de los tres primeros hits para las proteínas estudiadas. El código numérico del nombre de los ligandos corresponde al código SID de PubChem<sup>18</sup> para esas moléculas. Como se puede ver en la imagen todas las estructuras son parecidas, en el sentido de que tienen estructuras similares que se repiten en la mayoría de estos hits, como es el caso de los anillos aromáticos, los heterociclos y heteroátomos muy electronegativos. Todo ello puede favorecer las interacciones de los ligandos con el sitio de unión donde se había visto que había una mayoría de residuos polares.

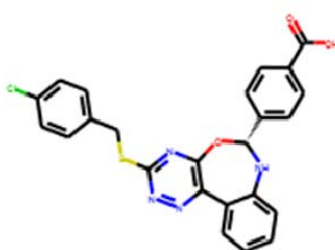
En la Tabla 4 se muestran los resultados del docking de comprobación realizado con los primeros hits del screening. Como es posible apreciar, la energía libre de interacción es similar en la mayoría de los casos por lo que es esperable que estos ligandos pudieran unirse adecuadamente a sus respectivos receptores. Sin embargo, el dato de energía libre obtenido directamente del resultado del screening tiende a ser más favorable que el obtenido a partir del docking de comprobación. Esto puede ser debido a que se utilizan algoritmos diferentes en cada caso. MTiAutoDock utiliza AutoDock 4 como programa de docking, y MTiOpenScreen utiliza AutoDock Vina, que es un programa más moderno y que ofrece mejores prestaciones respecto a su predecesor<sup>19</sup>. Pese a ello, los valores que ofrecen ambos programas son bastante similares en las estructuras objeto del estudio, a excepción de los hits de 5BNN, los cuales presentan una gran disparidad posiblemente por algún fallo de AutoDock 4. Por tanto, para esta estructura solo se va a tener en cuenta los datos de energía libre aportados por AutoDock Vina.

Con el fin de obtener compuestos con actividad antiviral, sería más conveniente obtener moléculas que tuvieran una energía libre de interacción más baja que los ligandos naturales. Esto permitiría a dichas moléculas unirse más fuertemente a las proteínas de entrada virales y bloquear su unión a los ligandos naturales actuando como inhibidores competitivos. En este estudio, se obtienen energías más bajas que las de referencia para las estructuras 3CL5, 4C7W y 5W1O. Las dos primeras corresponden a virus envueltos que se unen al ácido siálico, mientras que la última corresponde a un virus sin envuelta que se une al heparán sulfato. Las energías de interacción de los ligandos obtenidos para estas proteínas son más bajas que el valor de referencia en los valores dados por los dos programas utilizados (AutoDock 4 y AutoDock Vina). En el caso contrario, cuando las energías de interacción son mayores que la energía de referencia, ambos programas también coinciden lo que nos da una idea aproximada de cómo se van a unir estos compuestos respecto a los ligandos naturales de los receptores estudiados. Cabe destacar, además, que ninguno de los hits del screening tiene una estructura similar a los ligandos naturales, lo cual puede ser debido a las características de la librería empleada.

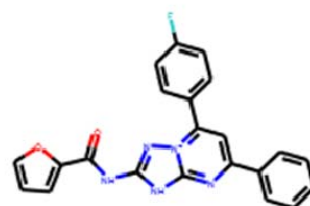
## 1ZBA



17506350\_Intermediate

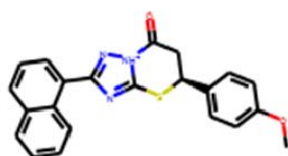


24275565\_Intermediate

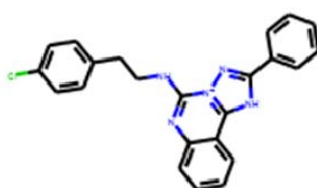


24394295\_Intermediate

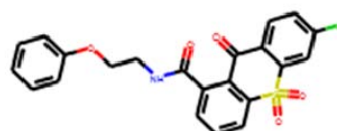
## 2XBO



24364895\_Intermediate

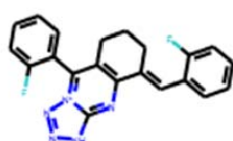


24387334\_Intermediate

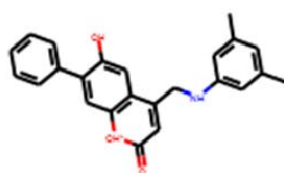


81065966\_Accepted

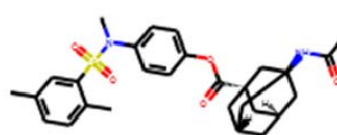
## 3CL5



24343513\_Accepted

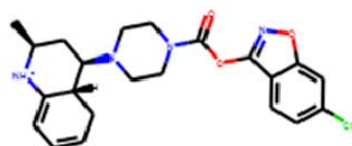


17512937\_Intermediate

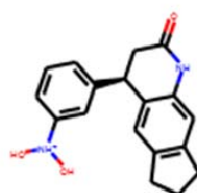


56321763\_Intermediate

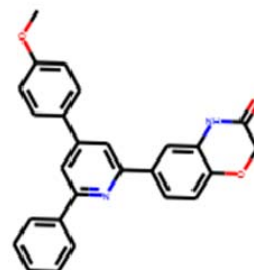
## 4BSH



125269061\_Intermediate

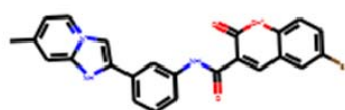


24812561\_Intermediate

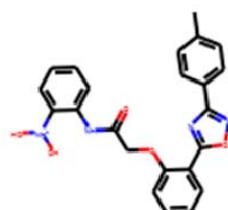


17386798\_Intermediate

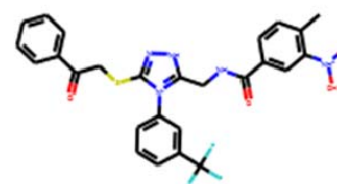
## 4C7W



17479619\_Intermediate

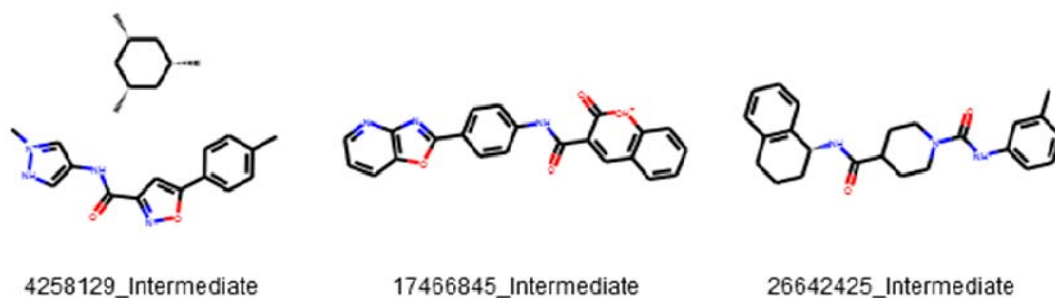


24393894\_Intermediate



24299280\_Intermediate

## 5BNN



## 5W1O

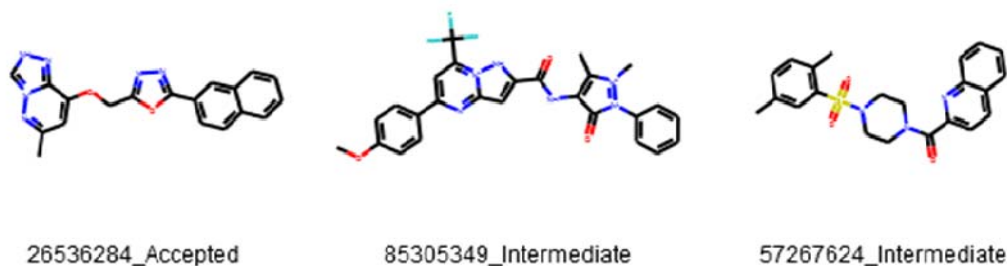


Figura 2: Estructuras y denominación de los tres mejores hits obtenidos a partir del screening con MTiOpenScreen sobre la librería Diverse-lib.

id_PDB	Hit	Energía AutoDock	Energía Vina
1zba	17506350_Intermediate	-7,35	-7,60
1zba	24275565_Intermediate	-7,34	-7,40
1zba	24394295_Intermediate	-7,32	-7,40
2xbo	24364895_Intermediate	-6,37	-6,30
2xbo	24387334_Intermediate	-6,37	-6,10
2xbo	81065966_Accepted	-6,14	-6,00
3cl5	24343513_Accepted	-7,61	-8,20
3cl5	17512937_Intermediate	-7,62	-8,10
3cl5	56321763_Intermediate	-7,61	-8,00
4bsh	125269061_Intermediate	-8,52	-9,10
4bsh	24812561_Intermediate	-8,55	-9,00
4bsh	17386798_Intermediate	-8,50	-8,80
4c7w	17479619_Intermediate	-8,89	-9,90
4c7w	24393894_Intermediate	-9,82	-9,30
4c7w	24299280_Intermediate	-9,32	-9,20
5bnn	4258129_Intermediate	-3,48	-8,90
5bnn	17466845_Intermediate	-3,48	-8,80
5bnn	26642425_Intermediate	-3,48	-8,80
5w1o	26536284_Accepted	-7,93	-9,20
5w1o	85305349_Intermediate	-7,79	-9,20
5w1o	57267624_Intermediate	-7,83	-9,20

Tabla 4: Energía libre del acoplamiento obtenida por MTiAutoDock 1.0 y por MTiOpenScreen para los tres mejores hits del screening.

Para valorar de forma objetiva la similitud que existe entre los primeros hits para algunos de los receptores, se han realizado mapas de similitud entre los ligandos obtenidos para proteínas correspondientes a virus relacionados o con características comunes.

En la Figura 3, se presenta un mapa de similitud por el método de Morgan<sup>20</sup> para los primeros hits de los dos receptores que corresponden a virus que se unen a heparán sulfato para su entrada en la célula. En estos mapas se compara la distribución de cada uno de los átomos y sus interacciones con el resto, de forma que se establecen átomos que tienen un gran peso en la similitud entre dos moléculas (verdes) ya que la diferencia entre ellas aumenta si se eliminan, y átomos que si son eliminados provocan que la diferencia entre las moléculas disminuya (rosas)<sup>21</sup>. De esta forma, se puede observar que la mayor parte de la estructura de los ligandos es compartida (verde), por lo que las interacciones que se forman en el sitio de unión de ambas proteínas puede que también sea similar.

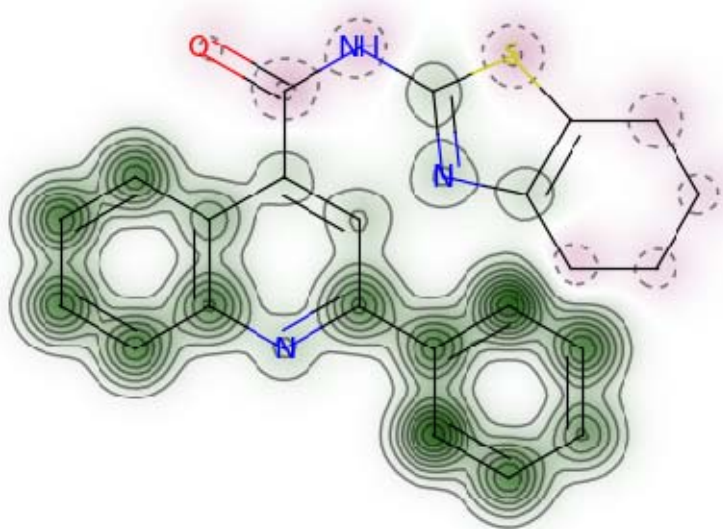


Figura 3: Mapa de similitud para los primeros hits de dos proteínas de entrada viral que unen heparán sulfato, 1ZBA y 5W1O.

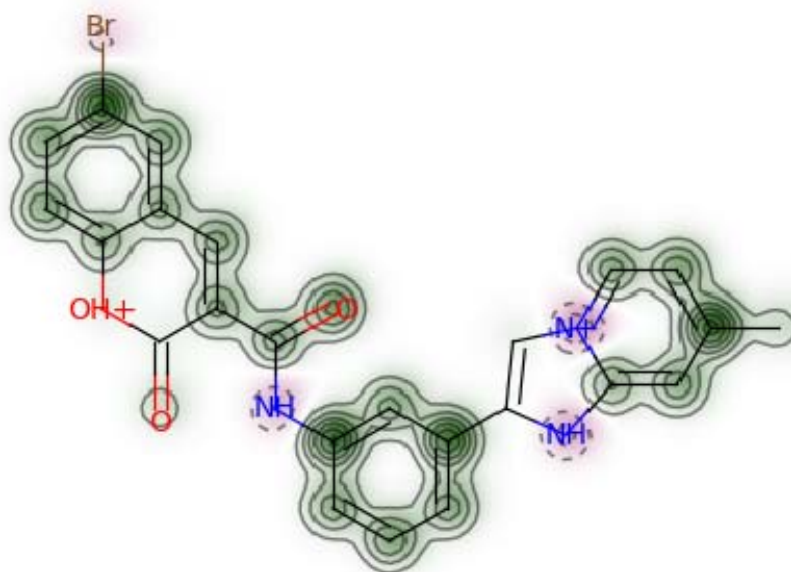


Figura 4: Mapa de similitud para los primeros hits de dos proteínas de entrada virales que unen ácido siálico, 4BSH y 4C7W.

Como también se puede ver en la Figura 4, en el caso de receptores que unen ácido siálico, también existe una similitud bastante alta entre componentes estructurales de los ligandos. Por ello, las interacciones que se forman entre los receptores virales y sus targets celulares podrían tener características similares, que permitieran desarrollar antivirales de amplio espectro en el futuro.

Utilizando las moléculas obtenidas como resultado del screening se han llevado a cabo análisis ADMET para determinar la idoneidad de las mismas para ser usadas como fármacos. Uno de los aspectos más importante para esto es la toxicidad potencial de los compuestos encontrados en el estudio. Como se puede ver en la Tabla 4, hay muchos compuestos que no cumplirían todos los controles de toxicidad e incluso algunos resultarían mutagénicos (Toxicidad AMES). Dos de los indicadores de toxicidad más importantes son los ensayos de toxicidad en *T. pyriformis* y en *P. promelas*. En el primer test se considera que un compuesto es tóxico si tiene un valor mayor a  $-0,5 \log \mu\text{g/L}$ . En el segundo se considera tóxico si tiene un valor menor a  $-0,3 \log \text{mM}$ .

Molécula	1ba-1	1ba-2	1ba-3	2bo-1	2bo-2	2bo-3	3d15-1	3d15-2	3d15-3	4bsh-1	4bsh-2	4bsh-3	4c7w-1	4c7w-2	4c7w-3	5bnn-1	5bnn-2	5bnn-3	5w1o-1	5w1o-2	5w1o-3
Solubilidad en agua (log mol/L)	-4,17	-3,13	-2,87	-4,21	-3,98	-5,39	-2,98	-4,06	-5,60	-4,12	-4,95	-4,75	-2,98	-5,16	-4,48	-4,07	-3,73	-4,32	-2,88	-3,07	-4,44
Permeabilidad Caco2 (log cm/s)	1,30	0,58	0,89	1,09	1,81	1,11	1,04	0,66	0,90	1,42	0,88	1,30	1,40	0,85	1,01	0,61	1,31	1,42	1,33	1,29	1,37
Absorción intestinal (%)	91,78	77,77	85,85	95,24	93,67	96,01	83,57	93,69	100,00	90,80	93,18	94,32	93,02	95,48	98,29	93,42	100,00	88,00	99,46	96,79	95,01
Permeabilidad cutánea (log Kp)	-2,74	-2,74	-2,74	-2,74	-2,74	-2,74	-2,73	-2,74	-2,78	-2,74	-2,73	-2,74	-2,74	-2,74	-2,74	-2,82	-2,74	-2,91	-2,74	-2,74	-2,77
P-glicoproteína sustrato	Si	Si	Si	No	Si	Si	Si	Si	No	No	Si	Si	Si	Si	Si	Si	No	Si	No	No	No
P-glicoproteína I inhibidor	Si	No	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si
P-glicoproteína II inhibidor	Si	No	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si
VDss (humano) (log l/kg)	-0,29	-2,14	0,23	-0,34	-0,26	-0,59	1,10	-0,80	-0,05	-0,01	0,25	-0,39	0,00	-0,60	-0,83	0,79	0,30	0,60	0,09	0,28	-0,21
Fracción desligada (humano)	0,13	0,04	0,17	0,24	0,10	0,06	0,00	0,01	0,00	0,02	0,00	0,29	0,08	0,18	0,15	0,06	0,17	0,03	0,21	0,16	0,00
Permeabilidad BBB (log BB)	0,25	-0,94	-0,45	0,76	0,63	-0,60	0,00	0,17	-0,83	-0,81	-0,01	-0,01	-0,92	-0,82	-1,57	-0,35	-0,67	0,12	-1,22	-1,74	-0,53
Permeabilidad CNS (log PS)	-1,57	-2,76	-2,05	-1,79	-1,67	-2,28	-1,55	-1,51	-2,17	-1,83	-1,89	-1,68	-1,55	-2,09	-2,27	-1,75	-1,94	-1,74	-3,82	-2,35	-2,23
CYP2D6 sustrato	No	No	Si	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
CYP3A4 sustrato	Si	No	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si
CYP1A2 inhibidor	Si	Si	Si	Si	Si	No	Si	Si	No	Si	Si	Si	Si	Si	No	Si	Si	No	Si	No	Si
CYP2C19 inhibidor	Si	No	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	No	No	Si
CYP2C9 inhibidor	Si	Si	No	Si	Si	Si	No	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	No	No	Si
CYP2D6 inhibidor	No	No	Si	No	Si	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
CYP3A4 inhibidor	Si	No	No	No	Si	Si	Si	Si	Si	No	No	No	Si	Si	Si	Si	Si	Si	Si	Si	No
Limpieza total (log mL/min/kg)	0,19	-0,17	0,45	0,20	0,12	0,13	-0,05	0,84	0,06	0,26	0,14	0,39	0,43	0,05	0,27	0,05	0,62	0,71	0,08	-0,14	0,95
OCT2 renal sustrato	No	No	Si	No	No	No	Si	No	No	No	No	No	Si	No	No	Si	No	Si	No	No	No
Toxicidad AMES	No	No	Si	Si	No	No	Si	Si	No	No	Si	No	Si	No	No	No	No	Si	Si	No	No
Max, dosis tolerada (humano) (log mg/kg/día)	0,25	0,72	0,28	0,75	0,40	0,43	0,28	0,22	-0,16	0,69	-0,11	0,70	0,43	0,78	0,62	-0,45	0,18	-0,64	0,69	0,64	0,20
hERG I inhibidor	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
hERG II inhibidor	Si	No	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si
Toxicidad oral aguda en rata (LD50) (mol/kg)	2,74	2,94	2,65	2,75	2,64	2,40	2,58	2,53	2,71	2,48	2,10	2,44	2,74	2,51	2,60	2,61	3,20	2,18	2,77	2,76	2,61
Toxicidad oral crónica en rata (LOAEL) (log mg/kg/día)	0,78	2,19	1,40	0,52	0,57	1,11	0,87	0,79	1,82	1,01	2,02	0,59	0,63	0,73	0,10	1,19	1,25	1,37	1,22	-0,06	1,69
Hepatotoxicidad	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si
Sensibilización cutánea	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
Toxicidad en <i>T. pyriformis</i> (log µg/L)	0,29	0,29	0,29	0,29	0,29	0,32	0,29	0,29	0,31	0,30	0,53	0,29	0,29	0,29	0,29	0,31	0,29	0,70	0,29	0,29	0,46
Toxicidad en <i>P. promelas</i> (log mM)	0,04	-1,49	-0,15	-0,97	-3,79	-0,43	-0,68	0,92	-1,56	-2,05	1,16	-1,01	-1,16	-4,66	-1,71	-0,17	0,32	1,15	-4,84	-0,32	-1,40

Tabla 4. Tabla de resultados del análisis ADMET obtenida del servidor web pkCSM a partir de los tres primeros hits del screening para cada estructura. Para más información acerca de alguno de los campos de la tabla, consultar la publicación de pkCSM<sup>17</sup>.

A la vista de los resultados, todos los compuestos encontrados tendrían algún tipo de toxicidad. Esto también se refleja en el test de hepatotoxicidad, que es positivo para todas las moléculas encontradas, e indica que puede causar algún desequilibrio hepático, aunque no tiene por qué ser grave. Uno de los aspectos más relevantes respecto a la toxicidad es que los compuestos que presentan algún halógeno en su estructura obtienen peores resultados en la predicción, por lo que podrían ser menos aptos como fármacos que el resto. Sin embargo, muchos fármacos ampliamente usados en la actualidad tienen propiedades tóxicas por lo que habría que realizar estudios experimentales para determinar si se terminan rechazando algunos de estos compuestos.

Respecto al resto de propiedades farmacocinéticas, las moléculas obtenidas tienen varios puntos en común. Estos compuestos tienen baja solubilidad en agua, alta capacidad de absorción intestinal y buena permeabilidad cutánea. La permeabilidad Caco-2, que indica la capacidad para atravesar membranas celulares, es buena en la mayoría de las moléculas (mayor que 0,90 log cm/s). En otras propiedades, sobre todo en las referidas a la distribución del fármaco, su excreción y su metabolismo, existen diferencias notables entre compuestos por lo que serían necesarios otros estudios para avanzar en estos aspectos.



### 3. Conclusiones

Las herramientas disponibles actualmente para el campo de la biología estructural o de la bioinformática farmacéutica nos permiten hacer predicciones sobre las características de un posible medicamento o las interacciones que se crean entre un medicamento y su diana terapéutica. En este trabajo se han obtenido mediante screening virtual una serie de compuestos que podrían ser utilizados como antivirales para los virus estudiados, aunque esto requeriría estudios mucho más profundos en el ámbito experimental. Sin embargo, el hecho que podamos predecir la estructura de compuestos que pueden funcionar bien contra ciertas dianas puede ahorrar mucho tiempo y dinero en el ámbito de la industria farmacéutica acortando el proceso para desarrollar nuevos medicamentos.

En el caso específico del desarrollo de antivirales, se ha constatado la razón por la cual es tan difícil crear antivirales de amplio espectro, ya que las estructuras de las proteínas de entrada virales estudiadas son muy diversas y diferentes entre sí. A pesar de ello, los compuestos obtenidos en el screening para las diferentes proteínas guardan cierta similitud entre sí, tanto si pertenecen a un virus envuelto o sin envuelta. Esto podría indicar que existen algunas características comunes entre los sitios de unión para el ácido siálico o el heparán sulfato en las proteínas de entrada virales.

Respecto a la evaluación de la viabilidad de los compuestos obtenidos en el screening para ser usados como fármacos, el análisis ADMET ha puesto de relevancia las posibles contraindicaciones de muchas de estas sustancias. Particularmente, en el caso de los compuestos con heteroátomos halógenos parece que existe una probabilidad mayor de que presenten cierta toxicidad, a pesar de que la librería empleada para el screening ya había sido filtrada para eliminar compuestos no aptos para su uso como medicamentos.

Los objetivos para el trabajo se han cumplido satisfactoriamente en su mayoría. Sin embargo, el estudio de relación evolutiva de las proteínas estudiadas no ha sido posible realizarlo con éxito debido a las enormes diferencias entre las

secuencias, así como de las estructuras de dichas proteínas. Los otros objetivos referentes a la obtención de nutraceuticos que puedan actuar contra las dianas elegidas han tenido que variar ligeramente debido a problemas técnicos con el servidor web empleado, y se ha optado por ampliar el screening a un número mayor de compuestos con características químicas más diversas.

La planificación del proyecto se ha seguido según los plazos en casi todo el desarrollo del trabajo. Sin embargo, los problemas técnicos surgidos por el uso del servidor web para el docking y el screening han hecho que algunos de los hitos de la planificación se retrasaran, sobre todo en el inicio de la planificación. Además, debido a la falta de similitudes entre las proteínas del estudio, se ha tenido que eliminar de la planificación alguno de los análisis como el clustering y se ha tenido que modificar el alcance del screening, ya que se ha tenido que realizar para todas las estructuras ante la falta de una estructura consenso.

A partir de los compuestos obtenidos en este trabajo, podrían realizarse estudios experimentales para evaluar la viabilidad de los mismos para su uso como antivirales de amplio espectro. Además, desde el punto de vista meramente computacional se podría realizar el screening con librerías de moléculas propias obtenidas a partir del análisis de la similitud de los ligandos naturales con compuestos relacionados con ellos, tanto funcional como estructuralmente.

## 4. Glosario

- ADMET: Acrónimo de administración, distribución, metabolismo, excreción y toxicidad. El análisis ADMET se utiliza para predecir las propiedades farmacocinéticas de posibles fármacos.
- Docking: Acoplamiento molecular, es una técnica computacional para predecir las características de la unión entre dos moléculas.
- Hits: Resultados favorables procedentes del screening virtual.
- Input: Entrada de datos de un programa.
- Log: logaritmo decimal de un número.
- Output: Salida de datos de un programa.
- Screening: Cribado virtual que se emplea para determinar qué moléculas entre unas dadas se unen mejor a un receptor.

## 5. Bibliografía

1. Monto, A. S. Vaccines and antiviral drugs in pandemic preparedness. *Emerging Infectious Diseases* **12**, 55–60 (2006).
2. Irwin, K. K., Renzette, N., Kowalik, T. F. & Jensen, J. D. Antiviral drug resistance as an adaptive process. *Virus Evol.* **2**, vew014 (2016).
3. Matrosovich, M., Herrler, G. & Klenk, H. D. Sialic acid receptors of viruses. *Top. Curr. Chem.* **367**, 1–28 (2015).
4. Zocher, G. *et al.* A Sialic Acid Binding Site in a Human Picornavirus. *PLoS Pathog.* **10**, e1004401 (2014).
5. Izquierdo-Useros, N. *et al.* Sialyllactose in viral membrane gangliosides is a novel molecular recognition pattern for mature dendritic cell capture of HIV-1. *PLoS Biol.* **10**, e1001315 (2012).
6. Pandey, R. P. *et al.* Broad-spectrum neutralization of avian influenza viruses by sialylated human milk oligosaccharides: in vivo assessment of 3'-sialyllactose against H9N2 in chickens. *Sci. Rep.* **8**, 2563 (2018).
7. Cagno, V. *et al.* Broad-spectrum non-toxic antiviral nanoparticles with a virucidal inhibition mechanism. *Nat. Mater.* **17**, 195–203 (2018).
8. Vaillant, A. Nucleic acid polymers: Broad spectrum antiviral activity, antiviral mechanisms and optimization for the treatment of hepatitis B and hepatitis D infection. *Antiviral Research* **133**, 32–40 (2016).
9. Grove, J. & Marsh, M. The cell biology of receptor-mediated virus entry. *Journal of Cell Biology* **195**, 1071–1082 (2011).
10. Dimitrov, D. S. Virus entry: Molecular mechanisms and biomedical applications. *Nature Reviews Microbiology* **2**, 109–122 (2004).
11. Berman, H. M. *et al.* The protein data bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
12. Cock, P. J. A. *et al.* Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
13. Pei, J., Kim, B. H. & Grishin, N. V. PROMALS3D: A tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **36**, 2295–2300 (2008).
14. Labbé, C. M. *et al.* MTiOpenScreen: A web server for structure-based

- virtual screening. *Nucleic Acids Res.* **43**, W448–W454 (2015).
15. Laskowski, R. A. *et al.* PDBsum: A Web-based database of summaries and analyses of all PDB structures. *Trends in Biochemical Sciences* **22**, 488–490 (1997).
  16. O’Boyle, N. M. *et al.* Open Babel: An Open chemical toolbox. *J. Cheminform.* **3**, 33 (2011).
  17. Pires, D. E. V., Blundell, T. L. & Ascher, D. B. pkCSM: Predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *J. Med. Chem.* **58**, 4066–4072 (2015).
  18. Kim, S. *et al.* PubChem Substance and Compound databases. *Nucleic Acids Res.* **44**, D1202–D1213 (2016).
  19. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–61 (2010).
  20. Rogers, D. & Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **50**, 742–754 (2010).
  21. Riniker, S. & Landrum, G. A. Similarity maps - a visualization strategy for molecular fingerprints and machine-learning methods. *J. Cheminform.* **5**, 43 (2013).

## 6. Anexos

### 6.1 Alineamientos

A continuación, se presentan los resultados del alineamiento de secuencias correspondiente a las proteínas de los virus que se unen a heparán sulfato (Figura 5) y de los virus que se unen al ácido siálico (Figura 6).

#### Colored PROMALS3D alignment (sequences in aligned order)

```

Conservation:
5wlo_chainC_p002 1 --AVVSTDEYVARTNIYYHAGTSRL LAVGHPYFPIKKPNNNKILVPKVSGLQYRVFRIHLDPDNKFGFPD 68
1zba_chain1_p001 1 TTTTGE----- 6
Consensus_aa:
Consensus_ss:
..hhsp.....
hhh eee eeee eeeee eeeee eeeeeee eeeeeee

Conservation:
5wlo_chainC_p002 69 TSFYNPDTQRLVWACVGVEVGRGQPLGVGISGHPLLNKLDLTENASAYAANAGVDNR ECISM DYKQTQLC 138
1zba_chain1_p001 7 -----SA-----DPVTTTVENYGGDTQ----- 23
Consensus_aa:
Consensus_ss:
.....s.....ssshshsppt.s.hp.....
eeeeeeeeee eeeee eeee

Conservation:
5wlo_chainC_p002 139 LIGCKPPIGEHWGKGSPECTNVAVNPGDCPP--L--ELINTVIQ---DGDMDVTGFG--AMDFTTLQANKS 199
1zba_chain1_p001 24 -----VQRHHHTDVGFI MDRFVKINSLSPTHVIDLMQTHKH GIVGA--LL--R- 67
Consensus_aa:
Consensus_ss:
.....h.p..chss..l..chlphs.b...cs..l.ps@...hhsh..Lb..+.
eee eee eeeeeeeeeeeee eeeee hhh hhhhh

Conservation:
5wlo_chainC_p002 200 EVPLDICTSICKYPDIYKMSPEPYGDSLFFYLRRQMFVRHLFN RAGTVGENVPDDL YIKGSGSTANLAS 269
1zba_chain1_p001 68 AAT-YYFS--DLEI-VV--R--HDGNLTWVPN----- 91
Consensus_aa:
Consensus_ss:
.hs..hho.....h..h..c...Dt..hh@l.p.....
hhhhe ehhhhhe e eehhhhhhhhhhhhhhhhh eee

Conservation:
5wlo_chainC_p002 270 SNYFPTPSGSMVTSDAQIFNKPY--WLQR-AQGHNNGICWGNQLFVTVDVTTR--ST--NMSLCAAISTS 332
1zba_chain1_p001 92 -----GAPEAALSNTSNPTAYNKAPFTRLAL-----PYTAPHRVLTAVYDG----- 132
Consensus_aa:
Consensus_ss:
.....hscsh...hpp.h..pssshh..sb...shshs+...h...t.....
eeee eee e eee eeee eeeeeeeeeeeee hhhhhhhheeeee

Conservation:
5wlo_chainC_p002 333 ETTYKNTNFKEYLRHG---EYDLQFIFQLCKITLTADVMTYIHSMNSTILEDWNGGGSGAEDPL---- 394
1zba_chain1_p001 133 -----TNKYSTQLPASFNYGAIQ-----AQAI-----HELLV 159
Consensus_aa:
Consensus_ss:
.....hp+.....b.hhQ.....tbs.....h....
hhhhhhhhh eeeeeeeeeeee hhhhhhhh hhhh eeeee

Conservation:
5wlo_chainC_p002 395 -----KK--YTFWEVNLKEKFSADLDQFFLGRKFLQAGL 427
1zba_chain1_p001 160 RMKRAELYCPRLLAIKVTSQDRYKQKIIAPA----- 191
Consensus_aa:
Consensus_ss:
.....hhhhcVs.p-+@p.c.l...s.....
eee ee eeeee hhh hhhhhhhh

```

Figura 5. Alineamiento de secuencias proteicas y comparación de estructuras secundarias de las proteínas implicadas en la unión a heparán sulfato.

# **Colored PROMALS3D alignment (sequences in aligned order)**

```

Conservation:
4c7w_chainA_p005 1 -----FNEPLNIVSHLNDWFLFGDAR-SDC-----TPKLD 30
3c15_chainA_p003 1 -----FDNPPNTNVVSHLNGDWFLFGDARSDCNHVVTNPNR--- 35
4bsh_chainA_p004 1 DKICLGHHAVSNGTKVNTLTERGVEVVNATETVERTNVPRICS-KG--- 45
5bnn_chainC_p002 1 -----GVPTYLLP-----GSGQFLTTD-----DHS--SAPVLPFCFN 29
2xbo_chainI_p001 1 -----VTNVG-----EDGEP-----GETEPRH- 17
Consensus_aa:
Consensus_ss:
.....sss.h.p.p.....

Conservation:
4c7w_chainA_p005 31 WLDL---DPKLCNSGRISAK--SGNSL-FRSFHFIF-DFYNYSGEQDVIFYEGVH-----GFKCLAYGD 87
3c15_chainA_p003 36 -NYSYMDLNPALCDSGKISSK--AGNSI-FRSFHFIF-DFYNYTGEQQIIFYEGVNFPTFYHAFKCTTSGS 100
4bsh_chainA_p004 46 -----KRT-VDLGQCGLLGTITGFPQCDQFLEFSADLIIE--RREGSD--VCY----- 88
5bnn_chainC_p002 30 -PTP-----EMH-IPG-----QIRN-----RREGSD--VCY----- 42
2xbo_chainI_p001 18 -ALS-----PVD-----MHVHTD----- 29
Consensus_aa:
Consensus_ss:
.....C.S.....hp.ps.....
e eee eeeee

Conservation:
4c7w_chainA_p005 88 NKRWMNGNKARFYARVYEKMAQYRSLSF--VNVSYAYGGNAKPTSICKD---KTL-TLNNPTFISKESN 150
3c15_chainA_p003 101 NDIMW-QNKGLFYTQVYKNMAVYRSITF--VNVYPVYNGSAQSTALCKS---GSL-VLNNPAYIAREAN 162
4bsh_chainA_p004 89 ----PGK-FVNEEALRQILRESGGIDK--ETMGFTYS-GIRT-NGATSACRRSGSSFYAEMKWLNSNTD 148
5bnn_chainC_p002 43 -----MLEMIQVESMMEINNTDGANG-----MERLRVDI--- 71
2xbo_chainI_p001 30 -----VSFLDRFFDV-----ETLELS--- 46
Consensus_aa:
Consensus_ss:
.....l.b...sh.hs.s.shps.....p..hl..p..
e hhhhhhhh eeeee eee eeeee

Conservation:
4c7w_chainA_p005 151 YVDYYYES-EANFTLQ-----GCDEFIVPLCVFN---GHSKSSSDPANKYYTD 195
3c15_chainA_p003 163 FGDYYKVV-EADFYLS-----GCDEYIVPLCIFN---GKFLS---NTKYDYD 202
4bsh_chainA_p004 149 NAA-FPQM-TKSYKNT-----RKDPALIIWGIHH--SGSTTEQTKL---YSGGNK 191
5bnn_chainC_p002 72 SVQADLDQLLENIPLDIQLDGPLRNTLVGNISRYTTHWSGSLEMTFMF--CGSFMATG-----KL 129
2xbo_chainI_p001 47 ---NLTGSPATHVLDPFPGST--AQLAWARLLNTCTYFFSDLELSIQFKFTTTPSSVGEG---FV 102
Consensus_aa:
Consensus_ss:
.s...p...s@s...l.htlh...sp..s.....p..
ee eee eeeeeee ee

Conservation:
4c7w_chainA_p005 196 SQSYYNMDTG-----VLYGFNSTLDVGNVTQNPGLDLTCRYLAL--- 234
3c15_chainA_p003 203 SQYYFNKDTG-----VIYGLNSTETIT---TGDFDNCHYLVL--- 236
4bsh_chainA_p004 192 LITVGS--SN-----YQSFVSPGARFPQVNGQSGRIDFHWLML--- 228
5bnn_chainC_p002 130 ILCYTP--PGGSCPTT-----RETA--MLGTHIVWDFG-----LQSSITLIIPWISGS 173
2xbo_chainI_p001 103 WVKWFF--VGAPTKTDAWQLEGGGNSVRIQQLAVAGMSPVVFVKIAG-----SRSQACGFSVPYTSMW 164
Consensus_aa:
Consensus_ss:
.b.h.s..ss.....h.shs.s.s.....phshhhl.h...
eeeeee eeeee eeeeeee

Conservation:
4c7w_chainA_p005 235 -TF-----GNVKAVSLE-----YLLSLPSKAI---CLRKPKSFM 264
3c15_chainA_p003 237 -PS-----GNVLAISNE-----LLLTVPTKAI---CLNKRKDF 266
4bsh_chainA_p004 229 -NP-----NDTVTFSEFNG-----AFIAPDRAS--FLRG-K--S 255
5bnn_chainC_p002 174 HYRMFNS-----DAKSTNANVGYVTCFMQTNLIVPSESSDTCSLIGFIAAK-DDFSLR--- 225
2xbo_chainI_p001 165 RVVPVFYNGWGAPTKKATYNWLPGAHFGSILLTS--DAH-----DKGGCY--LRYRFPFRANMY--- 219
Consensus_aa:
Consensus_ss:
.ss.....ss..h.....hh...A...h.....
eeeeee eee e eeee

Conservation:
4c7w_chainA_p005 265 PVQVVVD-----SDNMTAVACQLPYCFRNTSA-----FHFRQLLSGLLYNVSCIAQQGAF 314
3c15_chainA_p003 267 PVQVVDSRWNNARQSDNMTAVACQPPYCYFRNSTNYVGVYDINHGDAFTSILSGLLYDSPCFSQQGVF 336
4bsh_chainA_p004 256 MGI-----QSSV-----QVDANCEGDCYHSGGTI 279
5bnn_chainC_p002
2xbo_chainI_p001
Consensus_aa:
Consensus_ss:
ee e

Conservation:
4c7w_chainA_p005 315 VYNNVSSSWPAY-GYGHCPT---AANIIGYMAPVCIIYDSDPLVPR----- 354
3c15_chainA_p003 337 RYDNVSSVWPLY-SYGRCP---AAD----- 358

```

Figura 6. Alineamiento de secuencias proteicas y comparación de estructuras secundarias de las proteínas implicadas en la unión a ácido siálico.

## 6.2 Código empleado

Para la caracterización de los sitios de unión de los sitios de unión se utilizó un código de Python que se presenta resumido y en forma de ejemplo a continuación:

```
from Bio.PDB import *
PDB_name = 'XXXX'
ligand_name = 'H_XXX'
chain = 'X'

# Importación de los datos estructurales
pdbl = PDBList()
pdbl.retrieve_pdb_file(PDB_name)
parser = MMCIFParser()
structure = parser.get_structure(PDB_name, 'C:\\Users\\SANTI\\' +
PDB_name[1:3].lower() + '\\\\' + PDB_name.lower() + '.cif')

# Obtención de los datos estructurales del ligando
ligand=[]
for residue in structure.get_residues():
    tags = residue.get_full_id()
    if tags[3][0] != " ":
        ligand.append(tags[3])
ligand = [res for res in ligand if res[0] == Ligand_name]
i = structure[0]['1']
center_residues = [i[res] for res in ligand]
center_atoms = Selection.unfold_entities(center_residues, chain)

# Búsqueda de átomos cercanos a los átomos del ligando
atom_list = [atom for atom in structure.get_atoms()]
ns = NeighborSearch(atom_list)
nearby_residues = {res for center_atom in center_atoms
                    for res in ns.search(center_atom.coord, 6, 'R')}
nearby_residues = [res for res in nearby_residues]
nearby_atoms = Selection.unfold_entities(nearby_residues, 'A')

# Cálculo de las coordenadas del sitio de unión y el tamaño de la caja
x = [a.get_coord()[0] for a in nearby_atoms]
y = [a.get_coord()[1] for a in nearby_atoms]
z = [a.get_coord()[2] for a in nearby_atoms]
box_center = [(min(x)+max(x))/2, (min(y)+max(y))/2, (min(z)+max(z))/2]
box_size = [abs(min(x)-max(x)), abs(min(y)-max(y)), abs(min(z)-max(z))]
print box_center
print box_size

# Obtención de la secuencia de aminoácidos del sitio de unión
from Bio.Seq import *
from Bio.Alphabet import *
parser = PDBParser()
ppb = PPBuilder()
structure_bs = parser.get_structure('BS_' + PDB_name, 'BS_' + PDB_name +
'.pdb')
seq = Seq("",generic_protein)
for pp in ppb.build_peptides(structure_bs):
```



```
seq += pp.get_sequence()
print seq
```

En la realización de los diferentes dockings, se genera un output que corresponde la actuación realizada por el programa si fuese llevada a cabo en un entorno local. Como ejemplo se presenta uno de los outuput, cuya estructura general corresponde a todos los dockings realizados con MTiAutoDock 1.0:

```
/service/MTiAutoDock-dev.py
receptor_file : receptor.pdb
gridcenter_residues: _2_THR_134__,_2_ARG_135__,_1_LYS_193__.
receptor_format: pdb
ligand_format : mol2
ligand_file : ligands_in.mol2
wdir : .
demo : None
gridspacing : 0.375
gridpts : 30,59,40
gridcenter_coordinates: 33.6633703704,20.8205185185,138.826407407
verbose : False
```

Dado que en ocasiones se ha utilizado babel en la línea de comandos de Linux para la conversión de los formatos de los ligandos, se presenta a continuación un ejemplo de los comandos empleados:

```
# Conversión y protonación de los ligandos para el docking
obabel -isdf ligand.sdf -omol2 -O ligand.mol2 -p 7.4
# Conversión de los ligandos a estructuras bidimensionales para ser
procesados por RDKit
obabel -imol2 ligand.mol2 -osdf -O ligand.sdf -d --gen2D
```

Para el análisis estructural de los hits y la presentación gráfica de los ligandos realizada con RDKit en Python se presentan los fragmentos de código más relevantes como ejemplo:

```
from __future__ import print_function
from rdkit import Chem
from rdkit.Chem import PandasTools
from rdkit.Chem import AllChem
from rdkit.Chem import Draw
from rdkit.Chem.Draw import SimilarityMaps

#Generación de las imágenes de los hits
PDBid = 'xxxx'
```

```

suppl = Chem.SDMolSupplier(PDBid + '.sdf')
ms = [x for x in suppl if x is not None]
for m in ms: tmp=AllChem.Compute2DCoords(m)
img=Draw.MolsToGridImage(ms[:9:3],molsPerRow=3,subImgSize=(200,200),legends=[
x.GetProp('_Name') for x in ms[:9:3]])
img.save(PDBid + '.png')

#Creación de los mapas de similitud
s_ligands1 = Chem.SDMolSupplier('xxxx.sdf')
f_ligands1 = [x for x in s_5bnn if x is not None]
s_ligands2 = Chem.SDMolSupplier('xxxx.sdf')
f_ligands2 = [x for x in s_5w1o if x is not None]
fig, maxweight = SimilarityMaps.GetSimilarityMapForFingerprint(f_ligands1[0],
f_ligands2[0], SimilarityMaps.GetMorganFingerprint)

```