

Temática: Quimioinformática, QSAR y QSRP

## ***3DFrag-MCP (Relevant 3D fragments with maximum common property value)***

**Aurelio Antelo Collo <sup>1\*</sup>, Ramón Carrasco Velar <sup>2</sup>, Nicolás García Pedrajas <sup>3</sup>, Gonzalo Cerruela Gracia <sup>4</sup>**

<sup>1</sup> Universidad de las Ciencias Informática. [aantelo@uci.cu](mailto:aantelo@uci.cu)

<sup>2</sup> Universidad de las Ciencias Informática. [carrasco@uci.cu](mailto:carrasco@uci.cu)

<sup>3</sup> Universidad de Córdoba. [malgapen@uco.es](mailto:malgapen@uco.es)

<sup>4</sup> Universidad de Córdoba. [gcerruela@uco.es](mailto:gcerruela@uco.es)

\* Autor para correspondencia: [aantelo@uci.cu](mailto:aantelo@uci.cu)

---

### **Resumen**

Se presenta un método para obtener fragmentos relevantes a una actividad biológica basado en similitud de grafos moleculares, llamado 3DFrag-MCP (*Relevant 3D fragments with maximum common property value*). El mismo, propone un enfoque diferente a otros métodos reportados en el estado del arte, en la forma de identificar los subgrafos similares en contraste con lo que se conoce rigurosamente como similitud molecular o similitud química. Además, utiliza un nuevo método de similitud molecular basado en el concepto de Propiedad Máxima Común (MCPhd) para cuantificar la similitud entre dos moléculas que utiliza descriptores grafo-teóricos topográficos híbridos y una nueva forma de reducción del grafo químico denominada Grafo Reducido Híbrido (HRG), basada en centro descriptores ponderados por las propiedades químico-físicas y la distancia entre ellos.

El método permite realizar evaluaciones y comparaciones teniendo en cuenta no sólo la estructura, sino también otras propiedades asociadas a la naturaleza electrostática, estérica y lipofílica de la molécula, lo cual ayuda a comprender la relación existente entre las propiedades físico-químicas de las moléculas con respecto a la respuesta o actividad biológica.

**Palabras clave:** fragmentos relevantes, similitud molecular, propiedad máxima común, descriptores topográficos híbridos.

.

### **Abstract**

*A method to obtain fragments relevant to a biological activity based on molecular graph similarity, called 3DFrag-MCP (Relevant 3D fragments with maximum common property value), is presented. It proposes a different approach to other methods reported in the state of the art, in the way of identifying similar subgraphs in contrast to what is rigorously known as molecular similarity or chemical similarity. In addition, it uses a new molecular similarity method based on the Maximum Common Property concept (MCPhd) to quantify the similarity between two molecules using hybrid topographic graph-theoretic graph descriptors and a new form of chemical graph reduction called Hybrid Reduced Graph (HRG), based on center descriptors weighted by chemical-physical properties and the distance between them.*

*The method allows evaluations and comparisons taking into account not only the structure, but also other properties associated with the electrostatic, steric and lipophilic nature of the molecule, which helps to understand the relationship between the physicochemical properties of the molecules with respect to the biological response or activity.*

**Keywords:** *relevant fragments, molecular similarity, maximum common property, hybrid topographic descriptors.*

---

## Introducción

La interacción entre las ciencias de la vida y el avance de la tecnología impulsa un ciclo continuo de crecimiento de los datos químicos; estos datos suelen almacenarse en bases de datos abiertas o parcialmente abiertas [1]. Por otra parte, los avances informáticos de las dos últimas décadas han permitido el uso extensivo del cribado virtual (VS) para utilizar la información presente en esas bases de datos en el descubrimiento de fármacos [2]. En los últimos años, los métodos de VS, como técnica computacional, se encuentran entre los más populares en la investigación farmacéutica para obtener las moléculas que presentan mayor probabilidad de unirse a una diana de una determinada enfermedad, típicamente un receptor de proteína o enzima [3],[4],[5].

Una de las categorías es el cribado virtual basado en el ligando (LBVS), el cual dispone de métodos precisos que permiten comparar la molécula activa conocida con las presentes en las bases de datos públicas (Chempider, ChEMBL, PubChem, Drugbank y ZINC, entre otras). Entre los métodos más empleados se encuentra la búsqueda de similitudes; en este enfoque se exploran las bases de datos a fin de identificar otras moléculas con estructuras similares a una o más moléculas activas de referencia y, por el principio de similitud de propiedad (SPP), es de esperar que con actividades similares [6],[7].

A pesar de que se ha desarrollado un número no despreciable de métodos que permiten cuantificar la similitud basados en el principio SPP, algunos autores [8],[9],[10] han cuestionado la hipótesis de que estructuras similares implican actividades similares, existiendo preocupación por el efecto de acantilados (o pozos) de actividad (activity cliffs), en los que moléculas estructuralmente similares tienen una diferencia significativa en la potencia [11],[7],[12]. Además, se ha demostrado que pequeños cambios químico-estructurales en una molécula conducen a cambios notables en las propiedades físico-químicas y, en consecuencia, en la actividad bioquímica [13],[14],[15]. Por lo que se puede afirmar

que moléculas estructuralmente similares pueden tener actividad biológica diferente, así como moléculas estructuralmente diferentes puedan tener actividad biológica similares, dando paso a las paradojas estructurales.

Es por ello, que determinar los fragmentos responsables de la actividad biológica en una colección de moléculas se hace muy complejo. Una alternativa, consistiría en utilizar propiedades químico-físicas que permitan entender las relaciones entre la estructura química y la actividad biológica como la electronegatividad expresada en el índice atómico S-State introducidos por Kier y Hall [16] y la refractividad molecular expresada en el índice atómico R-State publicado por Carrasco y colaboradores [17], ambos índices utilizados posteriormente por Adhikari y colaboradores [18] y Tarun y colaboradores [19] en estudios de estructura actividad (QSAR). Estos enfoques logran combinar los principios fisicoquímicos con la estructura de los grafos moleculares y también contienen información topológica sobre la estructura química del entorno atómico, dando una visión distinta para entender las relaciones entre la estructura química y la actividad biológica.

El método 3DFrag-MCP que se presenta, para cuantificar la similitud molecular utiliza el grafo con todos los átomos de hidrógeno, ponderando los átomos diferentes de hidrógeno, por el valor de ciertas propiedades físico-químicas de las moléculas asociadas a la naturaleza electrostática, estérica y lipofílica, que se ven expresadas en descriptores atómicos topográficos híbridos (DATH) índice de Estado Electrotopográfico para átomos ( $S_{3D}$ ), índices de Estado Refractotopográfico para átomos ( $\mathfrak{R}_{3D}$ ) y de Estado Lipotopográfico para átomos ( $\Lambda_{3D}$ ) reportados por Carrasco Carrasco y colaboradores [20]. Además, al igual que otros métodos como MCS [21],[22],[23],[24], ISIDA [25], SHAFTS [26],[27] y LS-align [28], la exploración del espacio de búsqueda se lleva a cabo mediante la comparación de pares de moléculas. Su funcionamiento parte de una estructura tridimensional (3D) obtenida a partir de su optimización con cualquier enfoque químico-cuántico similar a otros métodos de similitud molecular 3D como SHAFTS y LSalign, pero se diferencia en que utiliza para cuantificar la similitud entre dos moléculas en un nuevo concepto denominado Propiedad Máxima Común (MCP) [29].

## Materiales y métodos

El método 3DFrag-MCP, está compuesto de cinco pasos ejecutados en forma consecutiva (Figura 1) y su funcionamiento parte de una colección de moléculas o grafos moleculares tridimensionales (3D), a los cuales se les determinó el nivel de una actividad biológica dada (Activo o Inactivo).

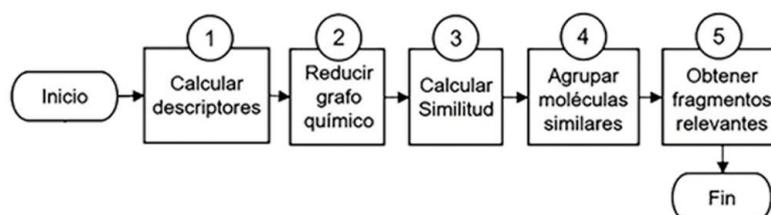


Figura 1. Método 3DFrag-MCP para obtener fragmentos relevantes a una actividad biológica.

**Paso 1.- Cálculo de descriptores:** el objetivo es ponderar los vértices (átomos) de los grafos moleculares 3D, pertenecientes al ensayo molecular, con propiedades físico-químicas expresadas con los valores de los descriptores DATH propuestos. Como resultado de este paso se obtienen los grafos moleculares en 3D pertenecientes al ensayo con sus vértices ponderados por propiedades físico-químicas asociadas a la naturaleza electrostática, estérica y lipofílica expresadas con los valores de los descriptores DATH  $S_{3D}$ ,  $\mathfrak{R}_{3D}$  y  $\Lambda_{3D}$ .

**Paso 2.- Reducción del grafo químico:** su objetivo es obtener una representación de los grafos moleculares 3D (del paso anterior) más abstracta y reducida (menos aristas y vértices) pero que mantenga las características principales o relevantes del grafo original. Para ello se emplean los centros descriptores (DC) que se muestran en la Figura 2.

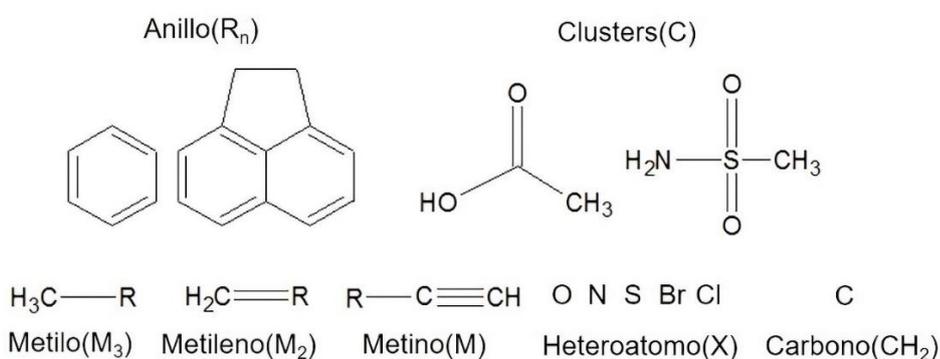


Figura 2. Ejemplos de DCs utilizados en la nueva forma de reducción del grafo molecular.

A este nuevo grafo molecular reducido se le incorpora información químico-física a través de los descriptores DATH, expresadas con los valores totales (Ecuación 1) convirtiéndolo en un grafo reducido híbrido (HRG) [10].

$$\phi_{total} = \sum_{i=1}^n \phi v_i \quad (1)$$

donde  $\phi_{total}$  representa el valor de alguno de los descriptores DATH ( $S_{3D}$ ,  $\mathfrak{R}_{3D}$ ,  $\Lambda_{3D}$ ) del DC,  $\phi v_i$  el valor del descriptor DATH seleccionado en el vértice o átomo  $i$  perteneciente al DC y  $n$  la cantidad de vértices o átomos que conforman el DC.

Como resultado final del paso se obtienen los nuevos grafos HRG con sus respectivos DCs ponderados con el valor total de alguno descriptores DATH  $S_{3D}$ ,  $\mathfrak{R}_{3D}$  y  $\Lambda_{3D}$  y su matriz de distancia entre los DCs.

**Paso 3.- Cálculo de la similitud:** el objetivo es obtener una matriz de similitud entre todos los grafos moleculares 3D reducidos logrados en el paso anterior. La cuantificación del valor de similitud de cada uno de los grafos con respecto a los restantes del ensayo se obtiene empleando el método de similitud molecular Propiedad Máxima Común (MCPHd)

reportado por Antelo y colaboradores [29]. El cual emplea para su ejecución grafos reducidos HRG, descriptores DATH y un umbral de similitud.

**Paso 4.- Obtención de grafos moleculares similares:** su objetivo es lograr agrupar los grafos moleculares similares clasificados como activos a partir de la matriz de similitud obtenida en el paso 3 empleando un método de agrupamiento (Clustering), técnica de aprendizaje no supervisado, que consigue encontrar y clasificar los elementos que comparten características semejantes dentro de un mismo grupo, separado de los otros grupos con los que no comparten dichas características [30].

Se seleccionó el algoritmo DBSCAN (Density-Based Spatial Clustering of Applications with Noise) [31] por ser un algoritmo de agrupamiento basado en el concepto de densidad, generando los conjuntos se formarán en las áreas del espacio más densamente pobladas [32]. Además, por su amplia aplicación en el campo del diseño de fármacos y en el cribado de bases de datos de moléculas [33], [34], [35], [36].

Como resultado final de este paso se obtiene una lista  $\{L_1, L_2, \dots, L_n\}$  de grafos moleculares clasificados en clústeres donde el 100% de los grafos agrupados sean activos a la actividad biológica estudiada.

**Paso 5.- Obtención de subgrafos moleculares relevantes:** el objetivo es identificar los subgrafos moleculares 3D que pueden ser relevantes a la actividad biológica evaluada en el ensayo, a partir de los grafos moleculares 3D similares que pertenecen a la clase activa obtenidos en el paso anterior.

Para ellos se utiliza el conjunto  $L = \{L_1, L_2, L_3, \dots, L_n\}$  de grafos HRG obtenidos en el paso 4 para determinar los fragmentos moleculares relevantes  $\{g_1, g_2, \dots, g_n\}$  a una actividad biológica dada, ejecutando el algoritmo que se muestra en la Figura 3. En el mismo se cuantifica el valor de la similitud por pares de grafos HRG, pertenecientes a cada lista, utilizando el método MCPhd basado en el concepto de MCP.

El algoritmo obtiene para cada lista  $L_i$  el conjunto de subgrafos o fragmentos moleculares  $L_{g_i} = \{g_1, g_2, \dots, g_N\}$  que presenten el mayor valor de MCP al comparar cada par de grafos HRG  $(G_i, G_j)$ . Seguidamente, a cada uno de los fragmentos moleculares  $g_N$  pertenecientes al conjunto de fragmentos moleculares  $L_{g_i}$ , se les calcula el valor total del descriptor DATH  $\phi_{total}$  y se obtienen un nuevo conjunto de fragmentos moleculares, denominado relevantes  $L_{g_{relevante}} = \{g_1, g_2, \dots, g_N\}$  de acuerdo a un nivel de importancia que se determina por la cantidad de veces que aparece en el conjunto de fragmentos moleculares  $L_{g_i}$ . A cada fragmento molecular relevante se le determina el valor total del descriptor DATH mínimo ( $\phi_{min}$ ) y máximo ( $\phi_{max}$ ).

Finalmente, se devuelve como resultado de este paso, el conjunto de fragmentos moleculares relevantes  $L_{g_{relevante}} = \{g_1, g_2, \dots, g_N\}$  con sus respectivos valores totales del descriptor DATH mínimo ( $\phi_{min}$ ) y máximo ( $\phi_{max}$ ).

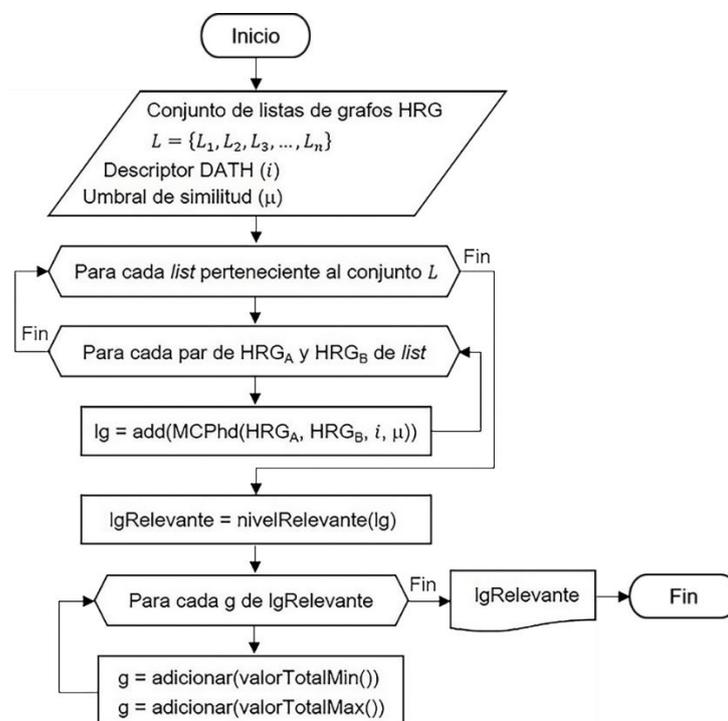


Figura 3. Algoritmo para obtener los fragmentos moleculares relevantes a una actividad biológica.

Se utilizaron dos conjuntos de datos. El primero, un ensayo con 36 moléculas perteneciente a la serie de 4-aminobenciclo[2.2.2]octan-2-yl-4-aminobutanoatos, utilizada por Weis y colaboradores [37], conformada por 12 butanoatos, 12 etanoatos y 12 propanoatos (ver Anexo 1), los cuales fueron evaluados contra la cepa K-1 multirresistente de *Plasmodium falciparum* y la cepa *Trypanosoma brucei rhodesiense* dando como resultado 17 moléculas activas, 18 inactivas y una no clasificada. El segundo, un ensayo de 51 moléculas 2-arilbenzoxazoles utilizados por Jana y colaboradores [38] para evaluar la inhibición de la transferencia de ésteres de colesterol mediada por la proteína de transferencia de ésteres de colesterol (CETP), el cual presenta 40 moléculas activas y 11 inactivas (Ver Anexo 2). Para convertir la estructura de cada molécula a 3D, se empleó el programa OpenBabel [39].

## Resultados y discusión

Los resultados del agrupamiento para el primer y segundo ensayo se muestran en la Tabla 1 y 2. Las moléculas (7c, 7h, 7i, 8c, 8f, 8i, 8f) y (Cpd27, Cpd29, Cpd31, Cpd35, Cpd37, Cpd39, Cpd40) agrupadas como activas en el primer y segundo ensayo respectivamente, son utilizadas para obtener el conjunto de fragmentos moleculares relevantes  $L_{g_i} = \{g_1, g_2, \dots, g_N\}$  que presenten el mayor valor de similitud en las propiedades electrostática y lipofílica respectivamente, representadas por los descriptores DATH  $S_{3D}$  y  $\Lambda_{3D}$  en los ensayos correspondientes, utilizando el coeficiente  $T_{C_{MCP}}$ .

Seguidamente, a cada fragmento molecular perteneciente al conjunto  $L_{g_i}$  se le determina el valor total mínimo ( $\phi_{min}$ ) y máximo ( $\phi_{max}$ ) del descriptor DATH, como se muestra en la Figura 2 y 3.

Tabla 1. Resultados del algoritmo DBSCAN con 5 clústeres para las 36 moléculas de la serie de 4-aminobicio[2.2.2]octan-2-yl-4-aminobutanoatos .

Clústeres	Moléculas del clúster	Cantidad de moléculas		Total	% de moléculas	
		Activas	Inactivas		Activas	Inactivas
1	6a, 6d, 6g, 6j, 7a, 7d, 7g, 7j	0	8	8	0	100
2	6b, 6c, 6f, 6i, 6l, 7b, 7e, 7f, 7k, 7l, 6h	5	5	11	41,3	41,3
3	6e, 6k, 8b, 8e, 8h, 8k	3	3	6	50	50
4	<b>7c, 7h, 7i, 8c, 8f, 8i, 8l</b>	<b>7</b>	<b>0</b>	<b>7</b>	<b>100</b>	<b>0</b>
5	8a, 8d, 8g, 8j	2	2	4	50	50

Tabla 2. Resultados del algoritmo DBSCAN con 5 clústeres para las 51 moléculas 2-arilbenzoxazoles.

Clústeres	Moléculas del clúster	Cantidad de moléculas		Total	% de moléculas	
		Activas	Inactivas		Activas	Inactivas
1	Cpd10, Cpd11, Cpd12, Cpd15, Cpd18, Cpd19, Cpd24, Cpd51, Cpd8, Cpd9	6	4	10	60	40
2	Cpd13, Cpd2, Cpd3, Cpd4, Cpd44, Cpd45, Cpd46, Cpd49, Cpd5	6	3	9	66,7	33,3
3	Cpd1, Cpd14, Cpd28, Cpd30, Cpd32, Cpd33, Cpd36, Cpd38, Cpd41	8	1	9	88.9	11.1
4	Cpd16, Cpd42, Cpd43, Cpd47	3	1	4	75	25
5	<b>Cpd27, Cpd29, Cpd31, Cpd35, Cpd37, Cpd39, Cpd40</b>	<b>7</b>	<b>0</b>	<b>7</b>	<b>100</b>	<b>0</b>

Los resultados obtenidos por Weis en el primer ensayo, demostraron que los butanoatos presentan una mayor actividad antiplasmódica que sus análogos etanoatos y propanoatos, porque la mayoría muestran una mejor selectividad. Además, que las moléculas más prometedoras de su serie son los 6c, 7c y 8c y las moléculas más activas y selectivas de toda la serie de éteres son los butanoatos 8f y 8l. Por último, concluyó que las moléculas más activas de todas las series poseían un grupo piperidino o un grupo terminal 4-metilpiperazina NR<sup>3</sup>R<sup>4</sup> en el extremo de la cadena.

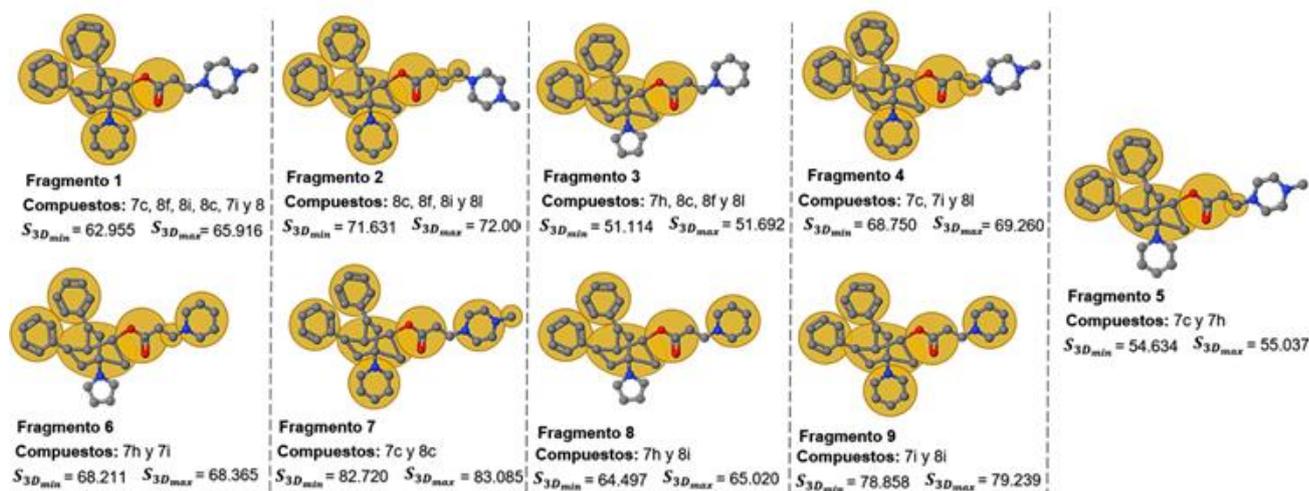


Figura 2. Fragmentos moleculares relevantes a la cepa K-1 multirresistente de Plasmodium falciparum y la cepa Trypanosoma brucei rhodesiense.

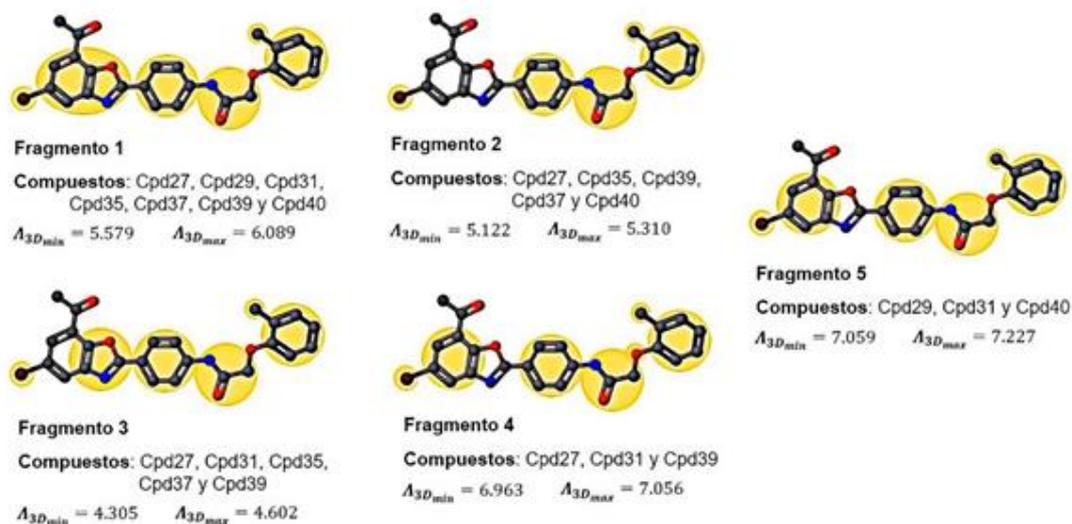


Figura 3. Fragmentos moleculares relevantes a la inhibición de la transferencia de ésteres de colesterol mediada por la proteína de transferencia de ésteres de colesterol (CETP).

Como resultado de aplicar el método 3DFrag-MCP al primer ensayo, que se muestran en la **Error! Reference source not found. 1**, se obtienen siete moléculas activas agrupados en el clúster cuatro (7c, 7h, 7i, 8c, 8f, 8i y 8l), mientras que ocho moléculas inactivas se agruparon en el clúster uno, el resto de los clústeres agruparon grafos moleculares tanto activos como inactivos. Esto demuestra que, utilizando la matriz de similitud a partir del descriptor DATH  $S_{3D}$  se logran agrupar correctamente siete moléculas activas representando el 41,1% de las activas presentes en la muestra. Además,

cuatro de ellas (8c, 7c, 8f y 8l), equivalentes al 57.1% de ese clúster, presentan los valores más pequeños de propiedad  $IC_{50}$ , es decir, las moléculas más activas del ensayo. Las siete moléculas activas agrupados están conformados por el 57,1% de butanoatos (8c, 8f, 8i y 8l) y el 42,8% de propanoatos (7c,7h y 7i). Este resultado confirma lo demostrado experimentalmente por Weis, que los butanoatos presentan una mayor actividad antiplasmódica.

Además, entre las moléculas agrupadas como activas se encuentran dos de las tres moléculas más prometedoras de su serie (moléculas 7c y 8c) y las moléculas más activas y selectivas de toda la serie de éteres (moléculas 8f y 8l). Por último, de los nueve fragmentos relevantes obtenidos a la cepa K-1, el 44,4% presentan un grupo piperidino y el 55,6% un grupo terminal 4-metilpiperazina ( $NR^3R^4$ ) en el extremo de la cadena, es decir, el 100% de los fragmentos relevantes encontrados presentan un grupo piperidino o 4-metilpiperazino que son los más activos de todas las series según Weis.

Si se comparan los resultados experimentales obtenidos por Weis con los alcanzados con el método 3DFrag-MCP utilizando el descriptor  $DATH S_{3D}$  se puede apreciar que se obtienen resultados semejantes. Esto resultados demuestran la eficacia del método propuesto.

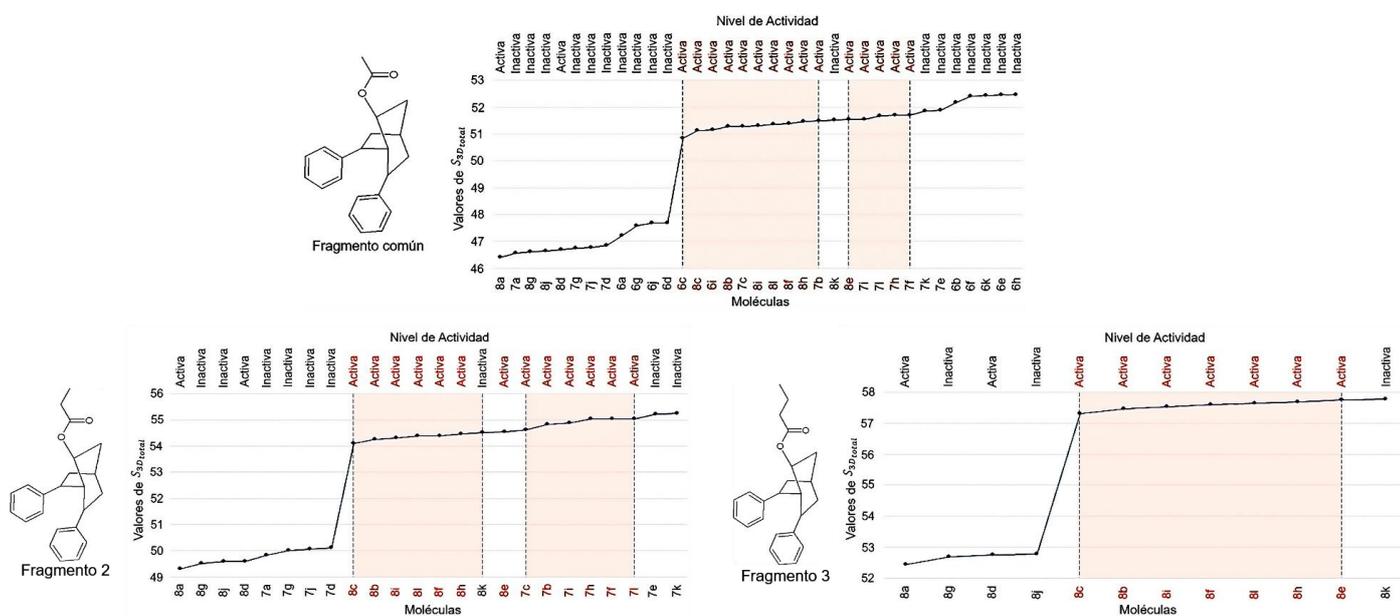


Figura 4. Valores totales del descriptor  $DATH S_{3D}$  del fragmento común, 2 y 3 en cada uno de las moléculas evaluadas frente a la cepa K-1 multirresistente de Plasmodium falciparum y la cepa Trypanosoma brucei rhodesiense.

Otro resultado que demuestra la eficacia del método 3DFrag-MCP propuesto, se muestra la **Error! Reference source not found.** 4, donde se puede apreciar un gráfico de línea ordenado de menor a mayor, con los valores totales del descriptor  $DATH S_{3D}$  de uno de los fragmentos comunes perteneciente a los fragmentos relevantes obtenidos. En el mismo se puede observar que cuando los valores totales del descriptor  $DATH S_{3D}$  se encuentran entre los intervalos de

[50.828, 51.495] y [51.541, 51.693] las moléculas son activas contra la cepa K-1. Si se tiene en cuenta las moléculas que presentan los nueve fragmentos relevantes obtenidos al aplicar el método 3DFrag-MCP propuesto, se puede apreciar que todas caen en los intervalos descritos anteriormente. Ese mismo comportamiento se observa para el fragmento 2 y 3, pero con valores distintos de los intervalos, para el fragmento 2 de [54.097, 54.461] y [54.626, 55.052], mientras que para el fragmento 3 es de [57.311, 57.759].

De los resultados obtenidos al ser aplicado al segundo ensayo, mostrados en la Tabla 2. Se obtienen siete moléculas activas agrupados en el clúster cinco (Cpd27, Cpd29, Cpd31, Cpd35, Cpd37, Cpd39, Cpd40), de las cuales el 100%, presentan los valores más pequeños de propiedad biológica ( $IC_{50}$  con valores inferiores a  $0.38 \mu\text{M}$ ), es decir, las moléculas más activas del ensayo.

Por otro lado, los cinco fragmentos seleccionados como relevantes que se muestran en la Figura 3, presentan un átomo de bromo (Br) en  $R_2$ . Si se analizan las moléculas que conforman el ensayo se puede apreciar que el 100% (11 en total) de las moléculas que presentan el átomo de Br en  $R_2$ , que se muestran en la Tabla 3, presentan valores de  $IC_{50}$  inferiores o igual a  $0.210 \mu\text{M}$ , lo que las incluye como moléculas muy activas a la inhibición de la transferencia de ésteres de colesterol. Además, las siete moléculas que selecciona el método y utilizadas para obtener los fragmentos relevantes son las que presentan el menor valor de la propiedad  $IC_{50}$  del total de moléculas que presentan el átomo de Br en  $R_2$ .

Tabla 3. Descripción de las moléculas 2-arilbenzoxazoles que presentan en la posición  $R_2$  un átomo de Br.

Molécula	$R_1$	$R_2$	$R_3$	$R_4$	$R_5$	$R_6$	$IC_{50}$ ( $\mu\text{M}$ )	Actividad
Cpd13	H	Br	H	H	-	-	1.300	Activa
Cpd21	H	Br	H	$\text{CH}_3$	-	-	0.510	Activa
Cpd25	H	Br	H	F	-	-	0.910	Activa
<b>Cpd27</b>	<b>H</b>	<b>Br</b>	<b>H</b>	<b><math>\text{COCH}_3</math></b>	-	-	<b>0.380</b>	<b>Activa</b>
<b>Cpd29</b>	-	<b>Br</b>	-	-	<b>H</b>	<b><math>\text{CH}_3</math></b>	<b>0.059</b>	<b>Activa</b>
<b>Cpd31</b>	-	<b>Br</b>	-	-	<b><math>\text{CH}_3</math></b>	<b><math>\text{CH}_3</math></b>	<b>0.044</b>	<b>Activa</b>
Cpd33	-	Br	-	-	$\text{CH}_3$	$\text{CH}_2\text{CH}_3$	0.110	Activa
<b>Cpd35</b>	-	<b>Br</b>	-	-	<b><math>\text{CH}_3</math></b>	<b>n-Pr</b>	<b>0.200</b>	<b>Activa</b>
<b>Cpd37</b>	-	<b>Br</b>	-	-	<b><math>\text{CH}_3</math></b>	<b>i-Pr</b>	<b>0.210</b>	<b>Activa</b>
<b>Cpd39</b>	-	<b>Br</b>	-	-	<b><math>\text{CH}_3</math></b>	<b>CCH</b>	<b>0.094</b>	<b>Activa</b>
<b>Cpd40</b>	-	<b>Br</b>	-	-	<b><math>\text{CH}_3</math></b>	<b>propynyl</b>	<b>0.210</b>	<b>Activa</b>

Otro resultado que demuestra la eficacia del método 3DFrag-MCP propuesto, se presenta en la Figura 5, donde se muestra un gráfico de línea ordenado de menor a mayor, con los valores totales del descriptor DATH  $A_{3D}$  de uno de los fragmentos comunes perteneciente a los fragmentos relevantes obtenidos. En el mismo se puede observar que cuando el valor total del descriptor DATH  $A_{3D}$  se encuentra entre los intervalos de [4.118, 4.646], [4.955, 5.045], [5.259, 5.317], [5.333, 5.509] y [5.622, 6.145] las moléculas son activas a la inhibición de la transferencia de ésteres de colesterol. Al igual que en el ensayo anterior, si analizamos las moléculas que presentan los cinco fragmentos relevantes

obtenidos al aplicar el método 3DFrag-MCP, se puede apreciar que todas caen en los intervalos descritos anteriormente. Ese mismo comportamiento se observa para los fragmentos 1, 2, 3, 4 y 5.

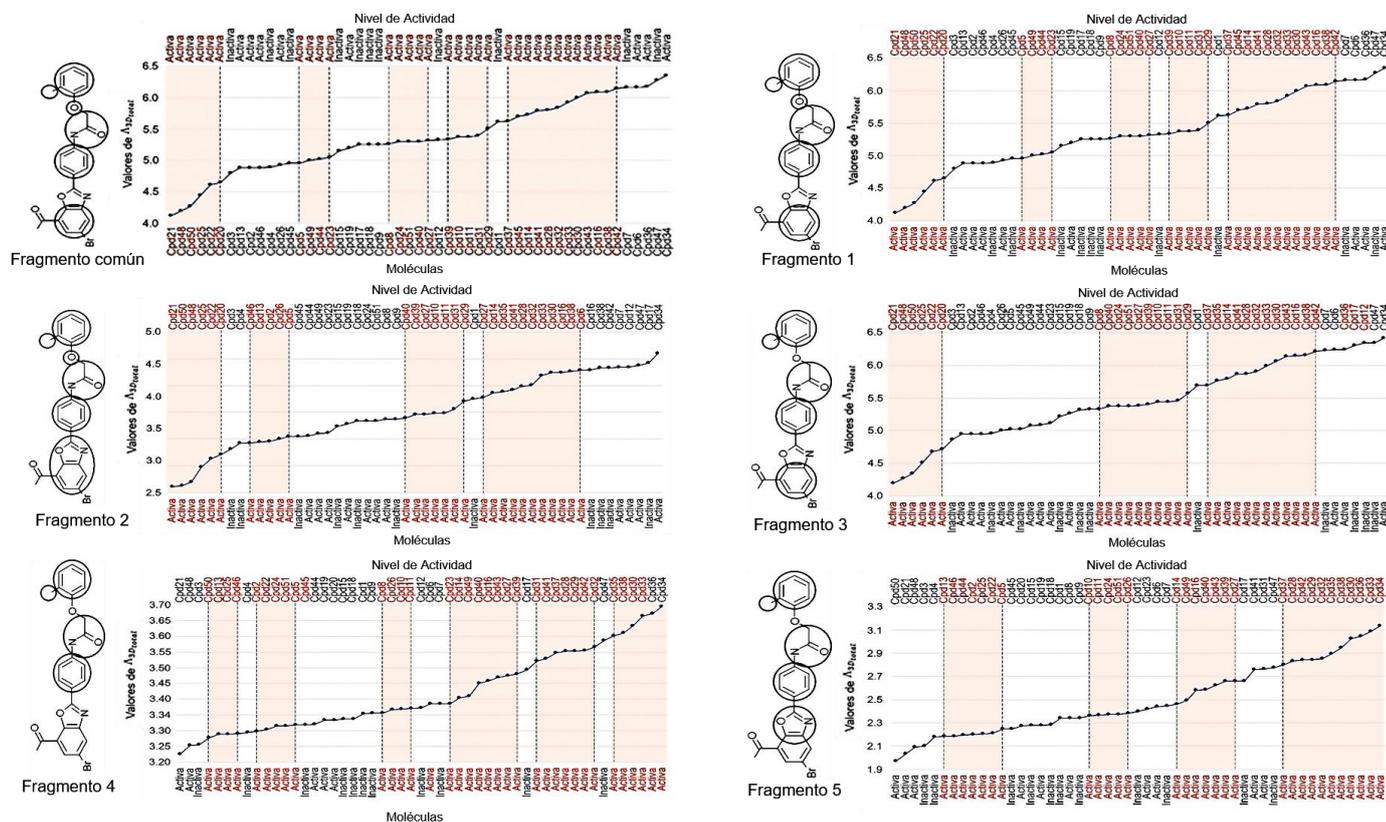


Figura 5. Valores totales del descriptor DATH  $\Lambda_{3D}$  del fragmento común, 1, 2, 3, 4 y 5 en cada uno de las moléculas evaluadas frente a la inhibición de la transferencia de ésteres de colesterol.

## Conclusiones

El método 3DFrag-MCP basado en el concepto de Propiedad Máxima se puede utilizar para encontrar fragmentos relevantes a una actividad biológica, en los que el peso de la respuesta depende de propiedades químico-físicas diferentes, lo cual no es posible lograr cuando se emplean solamente aspectos relacionados con la estructura o la conectividad entre los nodos del grafo. El mismo se aplicó a problemas reales estudiados por Weis y Jana, demostrando su utilidad para encontrar fragmentos relevantes a una actividad biológica.

## Referencias

1. Singh, N., Chaput, L., & Villoutreix, B. O. (2021). Virtual screening web servers: Designing chemical probes and drug candidates in the cyberspace. *Briefings in Bioinformatics*, 22(2), 1790-1818.

2. Rester, U. (2008). From virtuality to reality—Virtual screening in lead discovery and lead optimization: A medicinal chemistry perspective. *Curr. Opin. Drug Discov. Devel*, 11, 559-568.
3. Santibáñez-Morán, M. G., López-López, E., Prieto-Martínez, F. D., Sánchez-Cruz, N., & Medina-Franco, J. L. (2020). Consensus virtual screening of dark chemical matter and food chemicals uncover potential inhibitors of SARS-CoV-2 main protease. *RSC Adv*, 10(42), 25089-25099.
4. Bechelane Maia, E. H., Cristina Assis, L., Alves de Oliveira, T., Marques da Silva, A., & Gutterres Taranto, A. (2020). Structure-Based Virtual Screening: From Classical to Artificial Intelligence. *Frontiers in chemistry*, 8(343).
5. Jintong Du, L., BoLiu, J. Y., Xuben Hou, J. Y., & Hao, F. (2020). Structure-based virtual screening, biological evaluation and biophysical study of novel Mcl-1 inhibitors. *FutureMed. Chem*, 12(14), 1293-130.
6. Köppen, H. (2009). Virtual screening-What does it give us? *urr Opin Drug Discov Devel*, 12(3), 397-407.
7. Stumpfe, D., & Bajorath, J. (2011). Similarity searching. *WIREs Comput Mol Sci*, 1, 260-282.
8. Martin, Y., Kofron, J., & Traphagen, L. (2002). Do structurally similar molecules have similar biological activity? *J Med Chem*, 45(19), 4350-4358.
9. Kubinyi, H. (2002). Chemical similarity and biological activities. *Journal of the Brazilian Chemical Society*, 13(6), 717-726.
10. Maggiora, G., Vogt, M., Stumpfe, D., & Bajorath, J. (2014). Molecular Similarity in Medicinal Chemistry. *J. Med. Chem*, 57(8), 3186-3204.
11. Guha, R., & Van Drie, J. H. (2008). Structure–Activity Landscape Index: Identifying and Quantifying Activity Cliffs. *J. Chem. Inf. Model*, 58(3), 646-658.
12. Cruz-Montegudo, M., Medina-Franco, J., Pérez-Castillo, Y., Nicolotti, O., Cordeiro, M., & Borges, F. (2014). Activity cliffs in drug discovery: Dr Jekyll or Mr Hyde? *Drug Discov Today*, 19(8), 1069-1080.
13. Bhandare, R., & Canney, D. (2014). Bioisosteric replacement and related analogs in the design, synthesis and evaluation of ligands for muscarinic acetylcholine receptors. *Med Chem*, 10(4), 361-375.
14. Seddon, M. P., Cosgrove, D. A., & Gillet, V. J. (2018). Bioisosteric Replacements Extracted from High-Quality Structures in the Protein Databank. *ChemMedChem*, 13(6), 607-613.
15. Ertl, P. (2020). Identification of Bioisosteric Substituents by a Deep Neural Network. *J. Chem. Inf. Model*, 60(7), 3369-3375.
16. Kier, L. B., & Hall, L. H. (s. f.). An Electrotopological-State Index for Atoms in Molecules. *Pharm Res*, 7(801–807), 1990.
17. Carrasco-Velaz, R., Padrón, J. A., & Gálvez, J. (2004). Definition of a novel atomic index for QSAR: the refractotopological state. *J Pharm Pharm Sci*, 7(1), 19-26.
18. Adhikari, N., Maiti, M., & Jha, T. (2010). Predictive comparative QSAR modelling of (phenylpiperazinyl-alkyl) oxindoles as selective 5-HT1A antagonists by stepwise regression, PCRA, FA-MLR and PLS techniques. *Eur J Med Chem*, 45(3), 1119-1127.
19. Tarun, J., Soma, S., Soumya, B., Amit Kumar, H., Nilanjan, A., & Milan Kumar, M. (2008). QSAR study on some orally active uracil derivatives as human gonadotropin–releasing–hormone receptor antagonists. *Internet Electron. J. Mol. Des*, 7(11), 234-250.

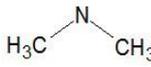
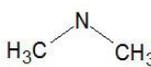
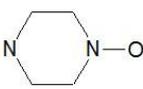
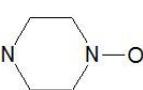
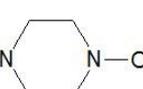
20. Carrasco, R., Prieto, J., Antelo, A., Padrón, J., Cerruela, G., Maceo, A., Alcolea, R., & Silva, L. (2013). Hybrid Reduced Graph For SAR Studies. *SAR and QSAR in Environmental Research*, 24, 201-214.
21. Raymond, J. W., & Willett, P. (2002). Maximum common subgraph isomorphism algorithms for the matching of chemical structures. *J Comput Aided Mol Des*, 16, 521-533.
22. Duesbury, E., Holliday, J., & Willett, P. (2015). Maximum Common Substructure-Based Data Fusion in Similarity Searching. *J. Chem. Inf. Model*, 55(2), 222-230.
23. Ehrlich, H. C., & Rarey, M. (2011). Maximum common subgraph isomorphism algorithms and their applications in molecular science: A review. *WIREs Comput Mol Sci*, 1, 68-79.
24. Ina, K. (2001). Enumerating all connected maximal common subgraphs in two graphs. *Theoretical Computer Science*, 250(1-2), 1-30.
25. Ruggiu, F., Marcou, G., Varnek, A., & Horvath, D. (2010). ISIDA Property-Labelled Fragment Descriptors. *Mol Inform*, 29(12), 855-868.
26. Liu, X., Jiang, H., & Li, H. (2011). SHAFTS: a hybrid approach for 3D molecular similarity calculation. 1. Method and assessment of virtual screening. *J. Chem. Inf. Model*, 51, 2372-2385.
27. Xiaofeng, L., & Honglin, L. (2011). SHAFTS: A Hybrid Approach for 3D Molecular Similarity Calculation. 2. Prospective Case Study in the Discovery of Diverse p90 Ribosomal S6 Protein Kinase 2 Inhibitors to Suppress Cell Migration. *J. Med. Chem*, 54(10), 3564-3574.
28. Hu, J., Liu, Z., Yu, D., & Zhang, Y. (2018). LS-align: An atom-level, flexible ligand structural alignment algorithm for high-throughput virtual screening. *Bioinformatics*, 34(13), 2209-2218.
29. Antelo-Collado, A., Carrasco-Velaz, R., García-Pedrajas, N., & Cerruela-García, G. (2020). Maximum common property: A new approach for molecular similarity. *Journal of Cheminformatics*, 12(1), 61.
30. Ieva, C., Gotlieb, A., Kaci, S., & Lazaar, N. (2018). Discovering Program Topoi via Hierarchical Agglomerative Clustering. *IEEE Transactions on Reliability, Institute of Electrical and Electronics Engineers*, 67(3), 758-770. <https://doi.org/10.1109/TR.2018.2828135>
31. Khan, K., Rehman, S. U., Aziz, K., Fong, S., & Sarasvady, S. (2014). DBSCAN: Past, present and future. In *The fifth international conference on the applications of digital information and web technologies (ICADIWT 2014)*. IEEE, 232-238.
32. Ester, M., Kriegel, H., Sander, J., & Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise.
33. Stafford, K. A., Anderson, B. M., Sorenson, J., & van den Bedem, H. (2022). AtomNet PoseRanker: Enriching Ligand Pose Quality for Dynamic Proteins in Virtual High-Throughput Screens. *J. Chem. Inf. Model*, 65(2), 1178-1189.
34. Ghemrawi, R., Khair, M., Hasan, S., Aldulaymi, R., AlNeyadi, S. S., Atatreh, N., & Ghattas, M. A. (2022). The Discovery of Potent SHP2 Inhibitors with Anti-Proliferative Activity in Breast Cancer Cell Lines. *Int. J. Mol. Sci*, 23(8), 4468.
35. Homology Modeling and Molecular Dynamics Simulations of Trypanosoma cruzi Phosphodiesterase b1. (2022). *Chemistry & Biodiversity*, 19(1), e202100712.

36. Arba, M., Paradis, N., Wahyudi, S. T., Brunt, D. J., Hausman, K. R., Lakernick, P. M., Singh, M., & Wu, C. (2022). Unraveling the binding mechanism of the active form of Remdesivir to RdRp of SARS-CoV-2 and designing new potential analogues: Insights from molecular dynamics simulations. *Chem Phys Lett*, 799, 139638.
37. Weis, R., Seebacher, W., Brun, R., Kaiser, M., Saf, R., & Faist, J. (2013). 4-Aminobicyclo[2.2.2]octan-2-yl 4-aminobutanoates with antiprotozoal activity. *Chemical Monthly*, 145(2), 311-317.
38. Dhritiman, J., Amit, H., Nilanjan, A., Milan, M., Chanchal, M., & Tarun, J. (2011). Chemometric modeling and pharmacophore mapping in coronary heart disease: 2-arylbenzoxazoles as cholesteryl ester transfer protein inhibitors. *MedChemComm*, 2(9), 840-852.
39. O'Boyle, N. M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, R. G. (2011). Open Babel: An open chemical toolbox. *J Cheminform*, 3(33).

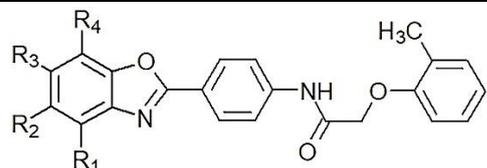
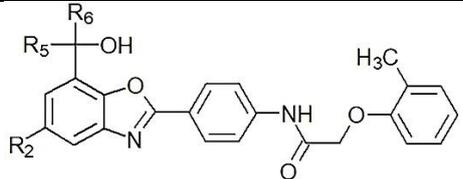
## Anexos

**Anexo 1.** Moléculas del ensayo 4-aminobicyclo[2.2.2]octan-2-yl 4-aminobutanoates.

		<b>6</b>	<b>7</b>	<b>8</b>	<b>Actividad</b>		
a	NR <sup>1</sup> R <sup>2</sup>		NR <sup>3</sup> R <sup>4</sup>		6a	IC <sub>50</sub> = 0.35	Inactivo
					7a	IC <sub>50</sub> = 0.46	Inactivo
					8a	IC <sub>50</sub> = 0.24	Activo
b	NR <sup>1</sup> R <sup>2</sup>		NR <sup>3</sup> R <sup>4</sup>		6b	IC <sub>50</sub> = 0.54	Inactivo
					7b	IC <sub>50</sub> = 0.26	Activo
					8b	IC <sub>50</sub> = 0.12	Activo
c	NR <sup>1</sup> R <sup>2</sup>		NR <sup>3</sup> R <sup>4</sup>		6c	IC <sub>50</sub> = 0.106	Activo
					7c	IC <sub>50</sub> = 0.06	Activo
					8c	IC <sub>50</sub> = 0.05	Activo
d	NR <sup>1</sup> R <sup>2</sup>		NR <sup>3</sup> R <sup>4</sup>		6d	IC <sub>50</sub> = 0.71	Inactivo
					7d	IC <sub>50</sub> = 0.55	Inactivo
					8d	IC <sub>50</sub> = 0.19	Activo
e	NR <sup>1</sup> R <sup>2</sup>		NR <sup>3</sup> R <sup>4</sup>		6e	IC <sub>50</sub> = 0.70	Inactivo
					7e	IC <sub>50</sub> = 0.37	Inactivo
					8e	IC <sub>50</sub> = 0.19	Activo

f	NR <sup>1</sup> R <sup>2</sup>		NR <sup>3</sup> R <sup>4</sup>		6f	IC <sub>50</sub> = 0.47	Inactivo
					7f	IC <sub>50</sub> = 0.25	Activo
					8f	IC <sub>50</sub> = 0.09	Activo
g	NR <sup>1</sup> R <sup>2</sup>		NR <sup>3</sup> R <sup>4</sup>		6g	IC <sub>50</sub> = 0.52	Inactivo
					7g	IC <sub>50</sub> = 0.40	Inactivo
					8g	IC <sub>50</sub> = 0.35	Inactivo
h	NR <sup>1</sup> R <sup>2</sup>		NR <sup>3</sup> R <sup>4</sup>		6h	IC <sub>50</sub> = 0.37	Inactivo
					7h	IC <sub>50</sub> = 0.28	Activo
					8h	IC <sub>50</sub> = 0.20	Activo
i	NR <sup>1</sup> R <sup>2</sup>		NR <sup>3</sup> R <sup>4</sup>		6i	IC <sub>50</sub> = 0.18	Activo
					7i	IC <sub>50</sub> = 0.26	Activo
					8i	IC <sub>50</sub> = 0.18	Activo
j	NR <sup>1</sup> R <sup>2</sup>		NR <sup>3</sup> R <sup>4</sup>		6j	IC <sub>50</sub> = 2.16	Inactivo
					7j	IC <sub>50</sub> = 2.25	Inactivo
					8j	IC <sub>50</sub> = 1.43	Inactivo
k	NR <sup>1</sup> R <sup>2</sup>		NR <sup>3</sup> R <sup>4</sup>		6k	IC <sub>50</sub> = 0.86	Inactivo
					7k	IC <sub>50</sub> = 1.91	Inactivo
					8k	IC <sub>50</sub> = 0.51	Inactivo
l	NR <sup>1</sup> R <sup>2</sup>		NR <sup>3</sup> R <sup>4</sup>		6l	IC <sub>50</sub> = nd	nd
					7l	IC <sub>50</sub> = 0.17	Activo
					8l	IC <sub>50</sub> = 0.09	Activo

**Anexo 2.** Moléculas 2-arilbenzoxazoles utilizadas para evaluar la inhibición de la transferencia de ésteres de colesterol mediada por la proteína de transferencia de ésteres de colesterol (CETP).

 <p style="text-align: center;"><b>B</b></p>					 <p style="text-align: center;"><b>C</b></p>			
Compuesto	R1	R2	R3	R4	R5	R6	IC <sub>50</sub> (mM)	Actividad
Cpd1 <sup>b</sup>	H	H	H	H	-	-	13.000	Inactivo

Cpd2 <sup>b</sup>	H	Cl	H	H	-	-	1.100	Activo
Cpd3 <sup>b</sup>	H	H	H	Cl	-	-	21.000	Inactivo
Cpd4 <sup>b</sup>	CH <sub>3</sub>	H	H	H	-	-	23.000	Inactivo
Cpd5 <sup>b</sup>	H	CH <sub>3</sub>	H	H	-	-	2.000	Activo
Cpd6 <sup>b</sup>	H	NO <sub>2</sub>	H	H	-	-	0.940	Activo
Cpd7 <sup>b</sup>	H	H	NO <sub>2</sub>	H	-	-	3.200	Inactivo
Cpd8 <sup>b</sup>	H	F	H	H	-	-	1.900	Activo
Cpd9 <sup>b</sup>	H	H	F	H	-	-	7.500	Inactivo
Cpd10 <sup>b</sup>	H	CN	CN	H	-	-	0.130	Activo
Cpd11 <sup>b</sup>	H	H	H	H	-	-	0.410	Activo
Cpd12 <sup>b</sup>	H	H	H	CN	-	-	5.200	Inactivo
Cpd13 <sup>b</sup>	H	Br	H	H	-	-	1.300	Activo
Cpd14 <sup>b</sup>	H	OCH <sub>3</sub>	H	H	-	-	0.840	Activo
Cpd15 <sup>b</sup>	H	SCH <sub>3</sub>	H	H	-	-	2.900	Inactivo
Cpd16 <sup>b</sup>	H	COCH <sub>3</sub>	H	H	-	-	1.300	Activo
Cpd17 <sup>b</sup>	H	CH(OH)CH <sub>3</sub>	H	H	-	-	3.400	Inactivo
Cpd18 <sup>b</sup>	H	CHCH <sub>2</sub>	H	H	-	-	2.800	Inactivo
Cpd19 <sup>b</sup>	H	CCH	H	H	-	-	2.000	Activo
Cpd20 <sup>b</sup>	H	CN	CH <sub>3</sub>	H	-	-	1.900	Activo
Cpd21 <sup>b</sup>	H	Br	H	CH <sub>3</sub>	-	-	0.510	Activo
Cpd22 <sup>b</sup>	H	CN	H	CH <sub>3</sub>	-	-	0.060	Activo
Cpd23 <sup>b</sup>	H	CN	H	CN	-	-	0.270	Activo
Cpd24 <sup>b</sup>	H	Cl	H	NO <sub>2</sub>	-	-	0.570	Activo
Cpd25 <sup>b</sup>	H	Br	H	F	-	-	0.910	Activo
Cpd26 <sup>b</sup>	H	CN	H	F	-	-	0.062	Activo
Cpd27 <sup>b</sup>	H	Br	H	COCH <sub>3</sub>	-	-	0.380	Activo
Cpd28 <sup>b</sup>	H	CN	H	COCH <sub>3</sub>	-	-	0.086	Activo
Cpd29 <sup>c</sup>	-	Br	-	-	H	CH <sub>3</sub>	0.059	Activo
Cpd30 <sup>c</sup>	-	CN	-	-	H	CH <sub>3</sub>	0.046	Activo
Cpd31 <sup>c</sup>	-	Br	-	-	CH <sub>3</sub>	CH <sub>3</sub>	0.044	Activo
Cpd32 <sup>c</sup>	-	CN	-	-	CH <sub>3</sub>	CH <sub>3</sub>	0.028	Activo
Cpd33 <sup>c</sup>	-	Br	-	-	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>3</sub>	0.110	Activo

Cpd34 <sup>c</sup>	-	CN	-	-	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>3</sub>	0.031	Activo
Cpd35 <sup>c</sup>	-	Br	-	-	CH <sub>3</sub>	n-Pr	0.200	Activo
Cpd36 <sup>c</sup>	-	CN	-	-	CH <sub>3</sub>	n-Pr	0.058	Activo
Cpd37 <sup>c</sup>	-	Br	-	-	CH <sub>3</sub>	i-Pr	0.210	Activo
Cpd38 <sup>c</sup>	-	CN	-	-	CH <sub>3</sub>	i-Pr	0.080	Activo
Cpd39 <sup>c</sup>	-	Br	-	-	CH <sub>3</sub>	CCH	0.094	Activo
Cpd40 <sup>c</sup>	-	Br	-	-	CH <sub>3</sub>	propynyl	0.210	Activo
Cpd41 <sup>c</sup>	-	CN	-	-	CH <sub>3</sub>	propynyl	0.160	Activo
Cpd42 <sup>b</sup>	-	H	-	-	CH <sub>3</sub>	CH <sub>3</sub>	0.440	Activo
Cpd43 <sup>b</sup>	H	CH <sub>2</sub> OH	H	H	-	-	1.300	Activo
Cpd44 <sup>b</sup>	H	CF <sub>3</sub>	H	H	-	-	0.590	Activo
Cpd45 <sup>b</sup>	H	H	CH <sub>3</sub>	H	-	-	13.000	Inactivo
Cpd46 <sup>b</sup>	H	H	H	CH <sub>3</sub>	-	-	1.700	Activo
Cpd47 <sup>b</sup>	H	H	H	COOCH <sub>3</sub>	-	-	3.500	Inactivo
Cpd48 <sup>b</sup>	H	CH <sub>3</sub>	H	CH <sub>3</sub>	-	-	0.280	Activo
Cpd49 <sup>b</sup>	H	CH <sub>3</sub>	H	COCH <sub>3</sub>	-	-	0.240	Activo
Cpd50 <sup>b</sup>	H	CF <sub>3</sub>	H	CF <sub>3</sub>	-	-	0.092	Activo
Cpd51 <sup>b</sup>	H	Cl	H	NO <sub>2</sub>	-	-	0.049	Activo
b.- Moléculas generadas a partir del compuesto B.								
c.- Moléculas generadas a partir del compuesto C.								