IV Conferencia Científica Internacional
Universidad de las Ciencias Informáticas

Temática: seleccionar la temática a partir de las líneas temáticas de los talleres

3DFrag-MCP: Método basado en Propiedad Máxima Común para obtener Fragmentos Relevantes a una actividad biológica.

Aurelio Antelo-Collado [1], Ramón Carrasco-Velar[1*], Nicolás Garcia-Pedrajas[2], Gonzalo Cerruela-Garcia[2]

[1] University of Informatics Science, Carretera San Antonio de los Baños Km. 2 1/2, Boyeros, La Habana, Cuba, Havana, Cuba. aantelo@uci.cu
[2] Department of Computing and Numerical Analysis, University of Cordoba. Campus de Rabanales, Albert Einstein Building, E-14071 Córdoba, Spain.

[*] Autor para correspondencia: rcarrasco@uci.cu

Resumen

El establecimiento de relaciones estructura-actividad y la identificación de similitud molecular es un ejercicio muy frecuente en la química-medicinal. La cuantificación de la similitud, caracteriza el grado de semejanza entre pares de moléculas. Existen múltiples métodos de similitud molecular entre pares de moléculas pequeñas en 3D, la mayoría de los cuales demandan mucho procesamiento de cálculo por utilizar la superposición de grafos, lo que los hace ineficientes frente a volúmenes de datos muy grandes. Tampoco permiten obtener diferencias en propiedades locales y globales pues no emplean propiedades químico-físicas. Por lo tanto, es aconsejable desarrollar un método de similitud molecular que esté basado en descriptores atómicos topográficos híbridos ya que poseen mayor contenido de información al estar ponderados por propiedades químico-físicas lo que les permite identificar estructuras diferentes con propiedades similares y viceversa. En la presente tesis se propone el concepto de Propiedad Máxima Común (MCPhd) como nuevo enfoque para cuantificar la similitud molecular utilizando descriptores grafo-teóricos topográficos híbridos que derivaron hacia un nuevo método de similitud molecular entre ligandos basado en el concepto de MCPhd para cuantificar la similitud entre dos compuestos. Esto es extensible a la similitud de subgrafos de grafos moleculares para obtener fragmentos relevantes a una actividad biológica denominado 3DFrag-MCP. Teniendo en cuenta que resulta frecuente el tener que trabajar

con series de datos numerosas y desbalanceadas, se propone 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. Todos los métodos propuestos se validaron con otros reportados en la literatura con resultados más ventajosos.

Palabras clave: Similitud molecular, propiedad máxima común, descriptores atómicos topográficos híbridos, grafo reducido hibrido.

Abstract
The establishment of structure-activity relationships and the identification of molecular similarity is a very frequent exercise in medicinal chemistry. Similarity quantification characterizes the degree of similarity between pairs of molecules. There are multiple methods of molecular similarity between pairs of small molecules in 3D, most of which require a lot of computational processing because they use the superposition of graphs, which makes them inefficient in the face of very large volumes of data. They also do not allow obtaining differences in local and global properties because they do not use chemical-physical properties. Therefore, it is advisable to develop a molecular similarity method based on hybrid atomic topographic descriptors since they have a higher information content as they are weighted by chemical-physical properties, which allows them to identify different structures with similar properties and vice versa. In the present thesis, the Maximum Common Property (MCPhd) concept is proposed as a new approach to quantify molecular similarity using hybrid grapho-theoretic topographic descriptors that derived towards a new method of molecular similarity between ligands based on the MCPhd concept to quantify the similarity between two compounds. This is extensible to the similarity of subgraphs of molecular graphs to obtain fragments relevant to a biological activity called 3DFrag-MCP. Considering the frequent need to work with large and unbalanced data sets, a new form of chemical graph reduction called Hybrid Reduced Graph (HRG) is proposed, based on descriptor centers weighted by chemical-physical properties and the distance between them. All the proposed methods were validated with others reported in the literature with more advantageous results.

Keywords: *Molecular similarity, maximum common property, hybrid topographic atomic descriptors, hybrid reduced graph.*

Introducción

El método que se presenta propone un enfoque diferente para obtener fragmentos relevantes a una actividad biológica. Está basado en similitud de grafos moleculares, y es llamado 3DFrag-MCP (*3D Fragments of Maximum Common Property*). Difiere de otros métodos reportados, en la forma de identificar los subgrafos similares en contraste con lo que se conoce rigurosamente como similitud molecular o similitud química [1].

En 3DFrag-MCP, al igual que en otros métodos como MCS [2-7], ISIDA [8], SHAFTS [9,10] y SL-align [11], la exploración del espacio de búsqueda se lleva a cabo mediante la comparación de pares de compuestos químicos o moléculas. Su funcionamiento parte de una estructura tridimensional (3D) obtenida a partir de su optimización con cualquier enfoque químico-cuántico tal como los métodos de similitud molecular 3D, SHAFTS y SL-align, pero se diferencia de ellos en que utiliza un nuevo concepto denominado Propiedad Máxima Común $(MCPhd)$ [12], para cuantificar la similitud entre dos compuestos químicos. $(MCPhd)$ está inspirado en el trabajo de Willett et. al. [2] sobre *Subgrafos Máximo Común* (MCS), pero en vez de utilizar isomorfismo de grafos, utiliza el valor de las propiedades físico-químicas de los compuestos expresadas por descriptores atómicos topográficos híbridos para cuantificar dicha similitud

Para cuantificar la similitud entre dos compuestos químicos, el método 3DFrag-MCP utilizando la función o coeficiente de Tanimoto Tc_{MCPhd} [11], el cual surge del coeficiente de Tanimoto basado en la MCS denominado Tc_{MCS} [14, 15], definido para dos grafos moleculares A y B según la expresión:

$$Tc_{MCS} = \frac{|MCS(A,B)|_b}{|A|_b + |B|_b - |MCS(A,B)|_b} \quad (1)$$

donde $|A|_b$ es el número de enlaces del grafo molecular A, $|B|_b$ es el número de enlaces del grafo molecular *B* y $|MCS(A,B)|_b$ es el número de enlaces del MCS de A y B. Al sustituir el concepto MCS en la Ecuación 1 por MCP se obtiene el coeficiente de Tanimoto basado en MCP como se muestra en la ecuación:

$$Tc_{MCPhd} = \frac{|MCPhd(A,B)|_b}{|A|_b + |B|_b - |MCPhd(A,B)|_b} \quad (2)$$

donde $|A|_b$ es el número de átomos pesados del grafo molecular A, $|B|_b$ el número de átomos pesados del grafo molecular *B* y $|MCPhd(A,B)|_b$ el menor número de átomos pesados entre los fragmentos con el mayor MCP entre los grafos moleculares *A* y *B*.

Por último, el nuevo método permite realizar evaluaciones y comparaciones teniendo en cuenta no sólo la estructura, sino también propiedades asociadas a la naturaleza electrostática, estérica y lipofílica de la molécula, ayudando a comprender la analogía o proximidad existente entre la estructura de las moléculas y ciertas propiedades físico-químicas de las moléculas representadas por los descriptores híbridos con respecto a la actividad biológica.

Materiales y métodos o Metodología computacional

El método 3DFrag-MCP está compuesto de seis pasos ejecutados en forma lineal (Figura 1) y para su funcionamiento se parte de una colección de compuestos o grafos moleculares en 3D pertenecientes a un ensayo, con actividad biológica reportada (Activo o Inactivo). Es importante destacar que el primer paso denominado análisis del equilibrio de clases, sólo se aplica si el ensayo se encuentra desbalanceado, es decir, que el porcentaje de los grafos moleculares activos presentes en la muestra supera en 1,5 o más a los grafos inactivos o viceversa.

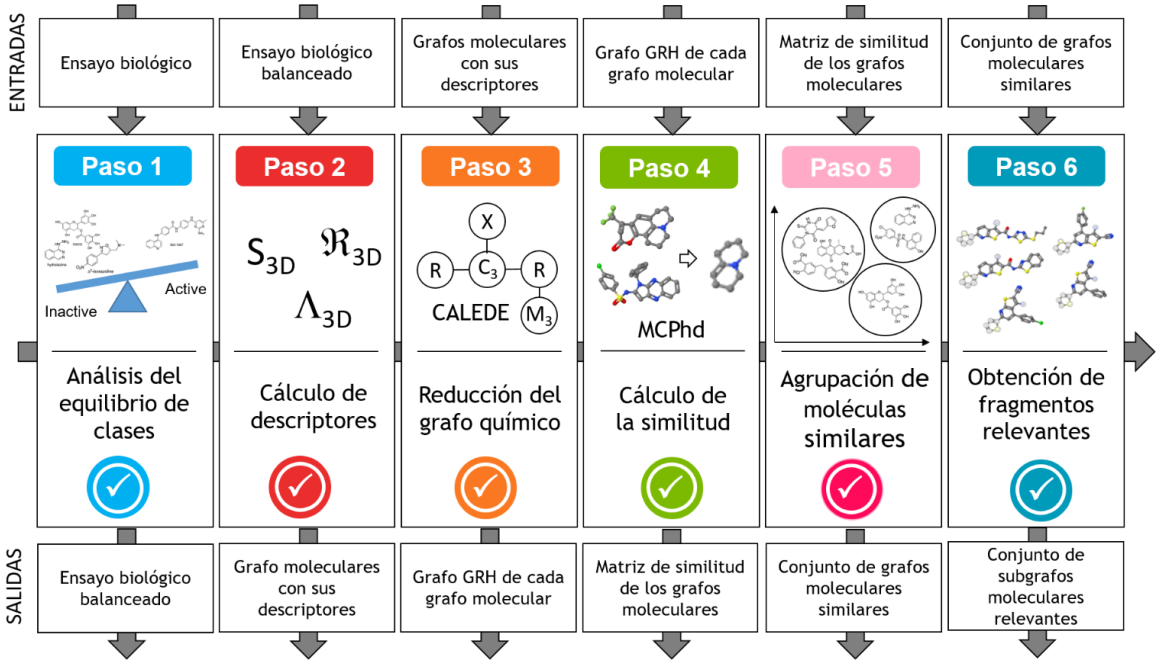


Figura 1. *Pasos del método 3DFrag-MCP*.

A continuación, se presenta una breve explicación de los pasos del método:

1. Análisis del equilibrio de clases: este paso solo se realiza si el ensayo presenta problemas de desequilibrio de clases y su objetivo final es lograr un balance entre las clases en el ensayo, empleando un método de selección de características.
2. Cálculo de descriptores: el objetivo es obtener grafos moleculares con sus vértices ponderados por propiedades físico-químicas expresadas con los valores de los descriptores atómicos topográficos híbridos propuestos.

3. Reducción del grafo químico: su objetivo es obtener una representación de los grafos moleculares con menos aristas y/o vértices que mantenga las características principales o relevantes del grafo original. A los grafos obtenidos se le ponderan sus nuevos vértices con el valor total de las propiedades físico-químicas expresadas con los valores totales de los descriptores atómicos topográficos híbridos propuestos.
4. Cálculo de la similitud: el objetivo es obtiene una matriz de similitud entre todos los grafos moleculares reducidos logrados en el paso anterior. La cuantificación del valor de similitud de todos contra todos se obtiene con el coeficiente de similitud de Tanimoto.
5. Obtención de grafos moleculares similares: su objetivo es obtener los grafos moleculares similares que pertenecen a la clase activa a partir de la matriz de similitud entre los grafos moleculares del paso anterior utilizando un método de agrupamiento.
6. Obtención de subgrafos moleculares relevantes: el objetivo es obtener los subgrafos moleculares que pueden ser relevantes a la actividad biológica evaluada en el ensayo, a partir de los grafos moleculares similares que pertenecen a la clase activa obtenidos en el paso anterior.

Paso 1. Análisis del equilibrio de clases

Aunque no constituye objetivo de esta presentación se brinda una descripción muy breve sobre este paso. El algoritmo se basa en la construcción de una combinación de selectores de características [16] siguiendo un enfoque similar al utilizado para la construcción de conjuntos de clasificadores de boosting [17]. De este modo, se aplican ciclos de selección de características (FS) para centrarse en las instancias que se han considerado más problemáticas. En cada una de los ciclos ejecutado

s se obtienen un subconjunto de características adecuadas para clasificar las instancias más problemáticas. Para combinar las soluciones de las distintas rondas se aplica un proceso de validación [18].

Paso 2. Cálculo de descriptores

Los descriptores atómicos híbridos utilizados en esta sección son la continuación lógica de los Índices del Estado Electrotopológico ($\mathcal{S}$) [26], Refractotopológico ($\mathfrak{R}$) y Lipotopológico (Λ) [13] para átomos, pero incorporándole para transformarlos en topográfico, información estructural tridimensional (3D) obtenida a partir de la estructura

optimizada con cualquier enfoque químico-cuántico. Esta transformación es un procedimiento muy sencillo: solo basta con sustituir la distancia topológica por la correspondiente euclidiana. Por lo tanto, el Índice del Estado Electrotopológico original de Kier y Hall se transforma en Electrotopográfico, el Refractotopológico en el índice Refractotopográfico y el índice Lipotopológico en el Lipotopográfico. Los índices Refracto y Lipo se desarrollan a partir de la teoría del grafo químico y la partición de la refractividad atómica o la lipofílidad atómica, tal como la definen Ghose y Crippen [27, 28, 29].

Los descriptores empleados se implementaron en la biblioteca CDK (*Chemical Development Kit*) en su versión 2.1 [30]. Como resultado de este paso se obtienen los grafos moleculares pertenecientes al ensayo con sus vértices ponderados por propiedades físico-químicas expresadas con los valores de los descriptores topográficos S_{3D}, $\Re_{3D}$ y Λ_{3D}.

Paso 3. Reducción del grafo químico.

La forma de reducción del grafo molecular propuesta se basa en el procedimiento desarrollado por Avindon et al. [31] y utiliza el concepto de Centro Descriptor (DC) definido por Carrasco et al [13].

En la Figura 4, se muestran los DCs que se proponen utilizar en este paso. Estos son los anillos de diferentes órdenes (R_n), clústeres de orden 3 y 4 (C_3 y C_4, respectivamente), heteroátomos como halógenos, amino, etc. (X), grupos terminales como el metilo (M_3), el metileno (M_2) y metino (M) y (CH_2) para los carbonos que no pertenecen a ninguno de los DC anteriores.

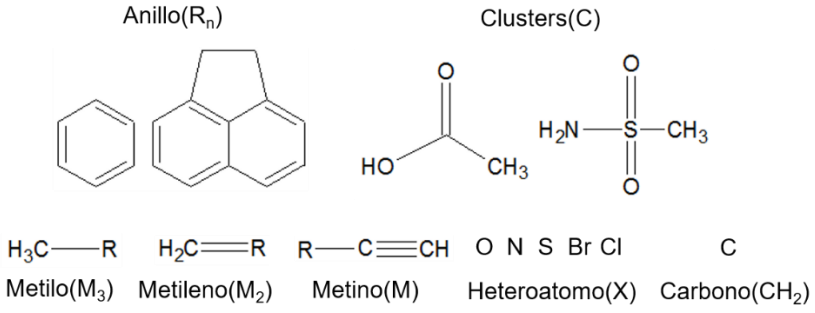


Figura 2. Ejemplo de Centros Descriptores utilizados en la reducción del grafo químico.

La nueva forma de reducción del grafo químico se muestra en la Figura 5. Los nuevos vértices del grafo reducido, denominados DCs, se muestran en la Figura 5A, identificados por círculos y en la Figura 5B, se muestra el grafo reducido que se obtiene al aplicar la nueva forma de reducción de un grafo molecular. Como se puede apreciar se logra realizar una reducción de un grafo de 37 vértices con 39 aristas a uno con 8 vértices y 6 aristas manteniendo las características y propiedades principales.

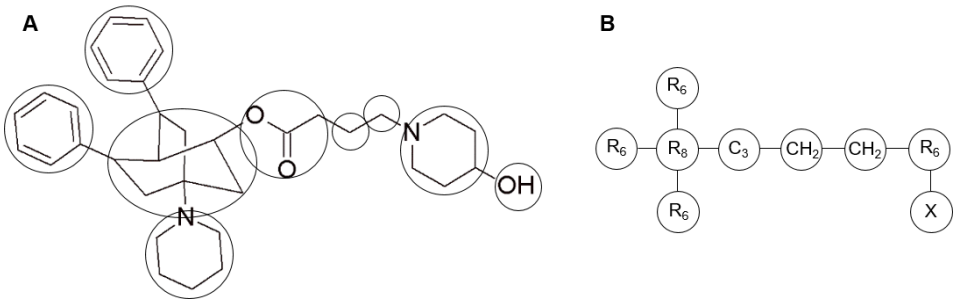


Figura 3. Nueva forma de reducción del grafo químico.

A este nuevo grafo molecular reducido se le incorpora información químico-física a través de los descriptores híbridos topográficos previamente calculados, convirtiéndolo en un grafo reducido híbrido [13] según la Ec. 12.

$$\phi_{total} = \sum_{i=1}^{n} \phi v_i \qquad (3)$$

donde ϕ representa alguno de los índices híbridos (S_{3D}, $\mathcal{R}_{3D}$, Λ_{3D}), ϕv_i el valor del índice híbrido seleccionado en el vértice o átomo i perteneciente al DC y n la cantidad de vértices o átomos que conforman el DC.

Las aristas entre los nuevos vértices se calculan de la distancia euclidiana entre los centros de masas de los CD's. Con esas distancias se construye la matriz de distancias topográficas del nuevo grafo (Fig. 6)

	DC_1	DC_2	DC_3	...	DC_n
DC_1	$d_E(DC_1, DC_1)$	$d_E(DC_1, DC_2)$	$d_E(DC_1, DC_3)$	...	$d_E(DC_1, DC_n)$
DC_2	$d_E(DC_2, DC_1)$	$d_E(DC_2, DC_2)$	$d_E(DC_2, DC_3)$	...	$d_E(DC_2, DC_n)$
DC_3	$d_E(DC_3, DC_1)$	$d_E(DC_3, DC_2)$	$d_E(DC_3, DC_3)$	...	$d_E(DC_3, DC_n)$
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
DC_n	$d_E(DC_n, DC_1)$	$d_E(DC_n, DC_2)$	$d_E(DC_n, DC_3)$	...	$d_E(DC_n, DC_n)$

Figura 6. Matriz de distancia topográfica entre DCs.

Como resultado final del paso se obtienen los nuevos grafos reducidos *HRG* con sus respectivos DCs ponderados con el valor total de alguno de los índices híbridos S_{3D}, $\mathcal{R}_{3D}$ y Λ_{3D} y su matriz de distancia entre los DCs.

Paso 4. Cálculo de la similitud

Para cuantificar el valor de similitud entre dos compuestos químicos o moléculas se propone en este paso utilizar el concepto de Propiedad Máxima Común ($MCPhd$) [11]. Este paso inicia con la entrada de un conjunto $C = \{HRG_1, HRG_2, HRG_3, \dots, HRG_n\}$ de grafos moleculares reducidos a los cuales se le cuantifica el valor de la similitud de todos contra todos, para obtener una matriz de similitud, utilizando el método SimilarityMCPhd basado en el concepto de $MCPhd$ que se muestra en la Figura 1.

El método necesita para su funcionamiento los siguientes parámetros: dos grafos HRG (HRG_1 y HRG_2), el índice híbrido (i) utilizado para cuantificar la similitud y (u) como umbral de similitud. A continuación, se presenta su descripción con la ayuda de en un ejemplo que se muestra en la Figura 2.

Inicia obteniendo del conjunto de grafos $C = \{HRG_1, HRG_2, HRG_3, \dots, HRG_n\}$ la lista de pares de subgrafos $L_{(g_{GHR_1}, g_{GRH_2})} = \left\{\left(g_{HRG_1\,1}, g_{HRG_2\,1}\right), \left(g_{HRG_1\,2}, g_{HRG_2\,2}\right), \dots, \left(g_{HRG_1\,n}, g_{HRG_2\,n}\right)\right\}$ con mayores valores de MCP. Esta lista se obtiene a partir del método FragmentMCPhd realizando los siguientes pasos:

1. A partir de los grafos HRG_1 y HRG_2, se obtiene la matriz de similitud entre los DCs empleando el coeficiente de Tanimoto para datos continuos [32] con el valor total del índice atómico topográfico utilizado (Ecuación 12), y la matriz de distancia entre los DCs utilizando la distancia Euclidiana. Ver paso B la Figura 2.

$$T_{cont}(DC_1, DC_2) = \frac{\sum \phi_{total_{i\,DC_1}} \cdot \phi_{total_{i\,DC_2}}}{\sum \left(\phi_{total_{i\,DC_1}}\right)^2 + \sum \left(\phi_{total_{i\,DC_2}}\right)^2 - \sum \phi_{total_{i\,DC_1}} \cdot \phi_{total_{i\,DC_2}}} \qquad 13$$

2. Seguidamente, utilizando la matriz de similitud entre los DCs, se crea una lista $L_{(CD_{HRG_1}, CD_{HRG_2})} = \left\{\left(DC_{HRG_1\,1}, DC_{HRG_2\,2}\right), \left(DC_{HRG_1\,1}, DC_{HRG_2\,2}\right), \dots, \left(DC_{HRG_1\,n}, DC_{HRG_2\,n}\right)\right\}$ de pares de DCs $\left(DC_{HRG_1\,i}, DC_{HRG_2\,j}\right)$ pertenecientes a los grafos HRG_1 y HRG_2 que el valor cuantificado de la similitud perteneciente a la celda C_{ij} sea mayor o igual al umbral de similitud (u) introducido como

parámetro (paso C-a del Anexo I). Además, utilizando las matrices de distancia de los DCs obtenidos en el paso anterior, para cada par $\left(DC_{HRG_1{}_i}, DC_{HRG_2{}_j}\right)$, se construye otra lista $L_{\left(CD_{HRG_1{}_i}, CD_{HRG_2{}_j}\right)} = \left\{\left(DC_{HRG_1}{}_1, DC_{HRG_2}{}_2\right), \left(DC_{HRG_1}{}_1, DC_{HRG_2}{}_2\right), \ldots, \left(DC_{HRG_1}{}_n, DC_{HRG_2}{}_n\right)\right\}$ con los pares de DCs $\left(\left(DC_{HRG_1{}_i}, DC_{HRG_2{}_j}\right), \left(DC_{HRG_1{}_i}, DC_{HRG_2{}_j}\right)\right)$ que están a una distancia de Canberra [33] menor o igual a 0,15 (Paso C-b Figura 8).

3. Finalmente, se seleccionan de todas las listas $\left\{L_{\left(CD_{HRG_1{}_i}, CD_{HRG_2{}_j}\right)}, L_{\left(CD_{HRG_1{}_i}, CD_{HRG_2{}_j}\right)}, \ldots, L_{\left(CD_{HRG_1{}_n}, CD_{HRG_2{}_m}\right)}\right\}$, las de mayor tamaño y se construyen para cada una de ellas los pares de subgrafos $\left(g_{HRG_1{}_i}, g_{HRG_2{}_j}\right)$, donde $g_{HRG_1{}_i} = \left\{DC_{HRG_1}{}_1, DC_{HRG_1}{}_2, \ldots, DC_{HRG_1}{}_n\right\}$, los cuales se devuelven en otra lista $L_{(g_{HRG_1}, g_{HRG_2})} = \left\{\left(g_{HRG_1}{}_1, g_{HRG_2}{}_1\right), \left(g_{HRG_1}{}_2, g_{HRG_2}{}_2\right), \ldots, \left(g_{HRG_1}{}_n, g_{HRG_2}{}_n\right)\right\}$.

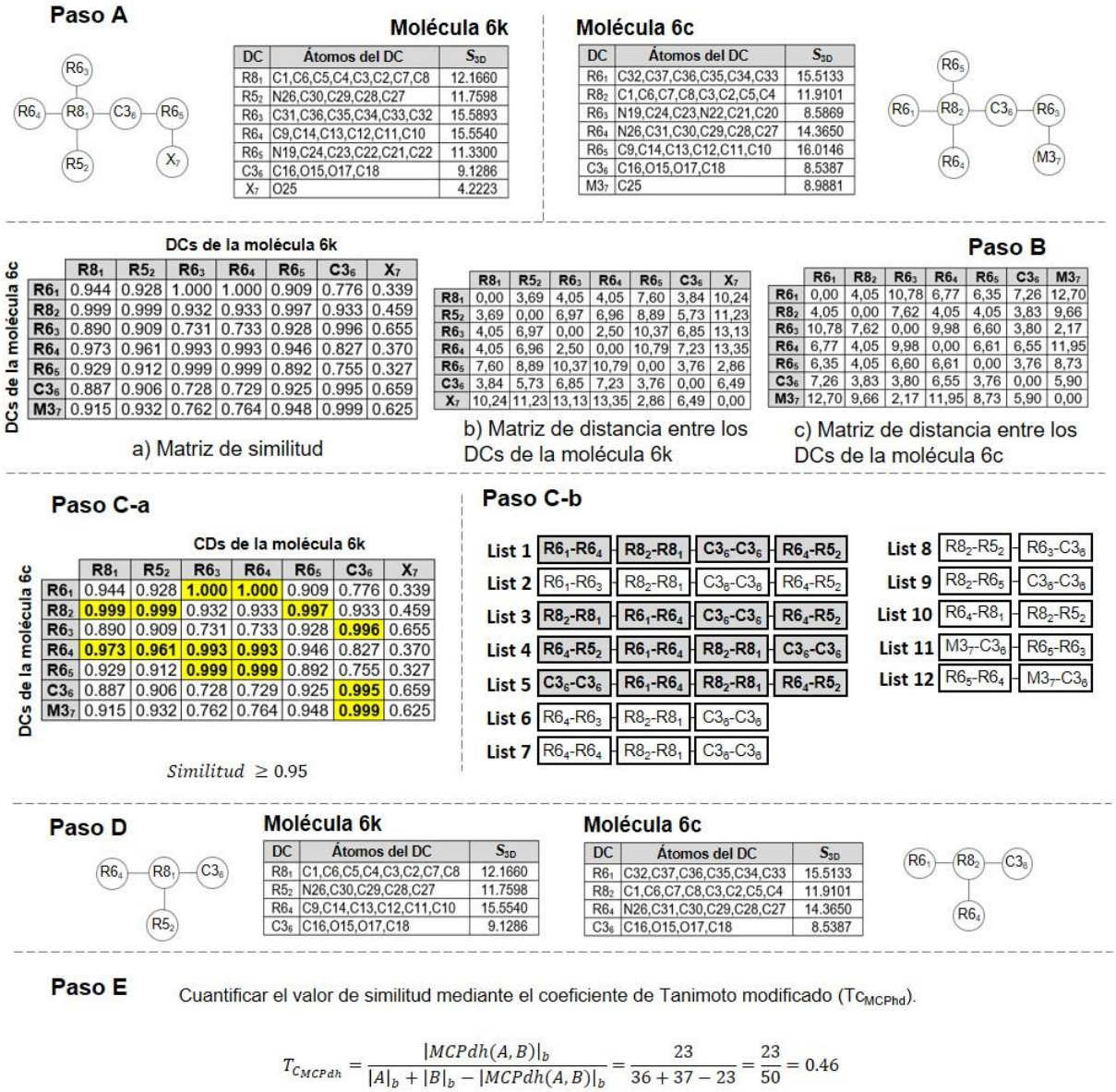
```
Method: MCPhd(G1, G2, u, i)

 INPUT: Two graph G1 and G2, the similarity thershold (u),
        the coefficient of similarity (f) and index (i)
 EXIT: Two similar fragments (f1 and f2) and quantification of
        similarity

 1.- listFrag ← fragmentPMC(G1, G2, u, i)
 2.- FOREACH frag IN listFrag
 3.-    f1 ← frag.getFragMol1()
 4.-    f2 ← frag.getFragMol2()
 5.-    IF f1 ≠ null and f2 ≠ null THEN
 6.-       atomf1 ← getHeavyAtomsMCP(f1)
 7.-       atomf2 ← getHeavyAtomsMCP(f2)
 8.-       c ← min(atomf1, atomf2)
 9.-       a ← getHeavyAtoms(G1)
10.-       b ← getHeavyAtoms(G2)
11.-       index ← c/(a+b-c)
12.-       listFrag ← index
13.-    END IF
14.- NEXT frag
15.- RETURN Max(listFrag)
```

Figura 4. Método MCPhd para el cálculo de la similitud molecular.

Con esta lista $L_{(g_{HRG_1}, g_{HRG_2})}$ de pares de subgrafos $\left(g_{HRG_{1\,i}}, g_{HRG_{2\,j}}\right)$ obtenidos y los grafos G_1 y G_2, se procede a cuantificar la similitud entre los dos grafos moleculares, utilizando la función o coeficiente de Tanimoto Tc_{MCPhd}.

A cada par de subgrafos $\left(g_{HRG_{1\,i}}, g_{HRG_{2\,j}}\right)$ de la lista de pares de subgrafos $L_{(g_{HRG_1}, g_{HRG_2})}$ se le cuantifica la similitud asignándole a $|MCPhd(A,B)|_b$ el número más pequeño de átomos pesados pertenecientes a los subgrafos $g_{HRG_{1\,i}}$ y $g_{HRG_{2\,j}}$, mientras que a $|A|_b$ y $|B|_b$ se les asigna el número de átomos pesados pertenecientes a cada gráfico G_1 y G_2, respectivamente. Estos valores se sustituyen en la función de similitud para obtener la cuantificación de la similitud de los gráficos G_1 y G_2. Finalmente, se obtiene el par de subgrafo $\left(g_{HRG\,1}, g_{HRG\,2}\right)$ que presenta el mayor valor de similitud (paso D de la Figura 8). Como resultado del método se obtienen el valor de la similitud entre los grafos moleculares G_1 y G_2 como se muestra en el paso E de la Figura 8.

Por último, en este paso se utiliza el método para el cálculo de la similitud entre dos grafos moleculares G_1 y G_2 descrito anteriormente para calcular la similitud por pares entre todos los grafos moleculares pertenecientes al ensayo y obtener como resultado una matriz de similitud.



DC	Átomos del DC	S_{3D}
$R8_1$	C1,C6,C5,C4,C3,C2,C7,C8	12.1660
$R5_2$	N26,C30,C29,C28,C27	11.7598
$R6_3$	C31,C36,C35,C34,C33,C32	15.5893
$R6_4$	C9,C14,C13,C12,C11,C10	15.5540
$R6_5$	N19,C24,C23,C22,C21,C22	11.3300
$C3_6$	C16,O15,O17,C18	9.1286
X_7	O25	4.2223

DC	Átomos del DC	S_{3D}
$R6_1$	C32,C37,C36,C35,C34,C33	15.5133
$R8_2$	C1,C6,C7,C8,C3,C2,C5,C4	11.9101
$R6_3$	N19,C24,C23,N22,C21,C20	8.5869
$R6_4$	N26,C31,C30,C29,C28,C27	14.3650
$R6_5$	C9,C14,C13,C12,C11,C10	16.0146
$C3_6$	C16,O15,O17,C18	8.5387
$M3_7$	C25	8.9881

	$R8_1$	$R5_2$	$R6_3$	$R6_4$	$R6_5$	$C3_6$	X_7
$R6_1$	0.944	0.928	1.000	1.000	0.909	0.776	0.339
$R8_2$	0.999	0.999	0.932	0.933	0.997	0.933	0.459
$R6_3$	0.890	0.909	0.731	0.733	0.928	0.996	0.655
$R6_4$	0.973	0.961	0.993	0.993	0.946	0.827	0.370
$R6_5$	0.929	0.912	0.999	0.999	0.892	0.755	0.327
$C3_6$	0.887	0.906	0.728	0.729	0.925	0.995	0.659
$M3_7$	0.915	0.932	0.762	0.764	0.948	0.999	0.625

	$R8_1$	$R5_2$	$R6_3$	$R6_4$	$R6_5$	$C3_6$	X_7
$R8_1$	0,00	3,69	4,05	4,05	7,60	3,84	10,24
$R5_2$	3,69	0,00	6,97	6,96	8,89	5,73	11,23
$R6_3$	4,05	6,97	0,00	2,50	10,37	6,85	13,13
$R6_4$	4,05	6,96	2,50	0,00	10,79	7,23	13,35
$R6_5$	7,60	8,89	10,37	10,79	0,00	3,76	2,86
$C3_6$	3,84	5,73	6,85	7,23	3,76	0,00	6,49
X_7	10,24	11,23	13,13	13,35	2,86	6,49	0,00

	$R6_1$	$R8_2$	$R6_3$	$R6_4$	$R6_5$	$C3_6$	$M3_7$
$R6_1$	0,00	4,05	10,78	6,77	6,35	7,26	12,70
$R8_2$	4,05	0,00	7,62	4,05	4,05	3,83	9,66
$R6_3$	10,78	7,62	0,00	9,98	6,60	3,80	2,17
$R6_4$	6,77	4,05	9,98	0,00	6,61	6,55	11,95
$R6_5$	6,35	4,05	6,60	6,61	0,00	3,76	8,73
$C3_6$	7,26	3,83	3,80	6,55	3,76	0,00	5,90
$M3_7$	12,70	9,66	2,17	11,95	8,73	5,90	0,00

List 1	$R6_1$-$R6_4$	$R8_2$-$R8_1$	$C3_6$-$C3_6$	$R6_4$-$R5_2$
List 2	$R6_1$-$R6_3$	$R8_2$-$R8_1$	$C3_6$-$C3_6$	$R6_4$-$R5_2$
List 3	$R8_2$-$R8_1$	$R6_1$-$R6_4$	$C3_6$-$C3_6$	$R6_4$-$R5_2$
List 4	$R6_4$-$R5_2$	$R6_1$-$R6_4$	$R8_2$-$R8_1$	$C3_6$-$C3_6$
List 5	$C3_6$-$C3_6$	$R6_1$-$R6_4$	$R8_2$-$R8_1$	$R6_4$-$R5_2$
List 6	$R6_4$-$R6_3$	$R8_2$-$R8_1$	$C3_6$-$C3_6$	
List 7	$R6_4$-$R6_4$	$R8_2$-$R8_1$	$C3_6$-$C3_6$	

List 8	$R8_2$-$R5_2$	$R6_3$-$C3_6$	
List 9	$R8_2$-$R6_5$	$C3_6$-$C3_6$	
List 10	$R6_4$-$R8_1$	$R8_2$-$R5_2$	
List 11	$M3_7$-$C3_6$	$R6_5$-$R6_3$	
List 12	$R6_5$-$R6_4$	$M3_7$-$C3_6$	

DC	Átomos del DC	S_{3D}
$R8_1$	C1,C6,C5,C4,C3,C2,C7,C8	12.1660
$R5_2$	N26,C30,C29,C28,C27	11.7598
$R6_4$	C9,C14,C13,C12,C11,C10	15.5540
$C3_6$	C16,O15,O17,C18	9.1286

DC	Átomos del DC	S_{3D}
$R6_1$	C32,C37,C36,C35,C34,C33	15.5133
$R8_2$	C1,C6,C7,C8,C3,C2,C5,C4	11.9101
$R6_4$	N26,C31,C30,C29,C28,C27	14.3650
$C3_6$	C16,O15,O17,C18	8.5387

$$T_{C_{MCPdh}} = \frac{|MCPdh(A,B)|_b}{|A|_b + |B|_b - |MCPdh(A,B)|_b} = \frac{23}{36 + 37 - 23} = \frac{23}{50} = 0.46$$

Figura 5. Ejemplo del funcionamiento del método $MCPhd$.

Paso 5. Obtención de las moléculas similares

El objetivo es obtener las moléculas similares clasificadas como activas y para lograrlo es necesario un método de agrupamiento (Clustering), técnica de aprendizaje no supervisado, que pueda encontrar y clasificar los elementos que comparten características semejantes juntos en un mismo grupo, separado de los otros grupos con los que no comparten características [34]. En este trabajo, seleccionamos el algoritmo k-Means [35] por ser un algoritmo lineal muy conocido por su eficiencia y de un costo computacional muy bajo [36]. Además, por su amplia aplicación en el campo del diseño de fármaco y en el cribado de bases de datos de compuestos químicos [37- 46].

Uno de los problemas para ejecutar el algoritmo k-Means es encontrar, el parámetro de entrada, número óptimo de clusters para un conjunto de datos dado. Existen varios métodos disponibles para identificar el número óptimo de clusters para un conjunto de datos dado, para este paso se proponen los métodos del codo (Elbow), la silueta media (Silhouette) y la estadística de la brecha (Grap).

El objetivo de este paso es lograr agrupar los grafos moleculares presentes en el ensayo a partir de la matriz de similitud obtenida en el paso 2.3.4 y obtener los grafos moleculares clasificados en clústeres donde el 100% de los grafos agrupados sean activos a la actividad biológica en estudio. Como resultado final del paso es devolver las listas $\{L_1, L_2, \dots, L_n\}$ de compuestos evaluados como activo ante la actividad biológica en estudio.

Paso 6. Obtención de los fragmentos relevantes

De la lista de grafos moleculares $\{L_1, L_2, \dots, L_n\}$s obtenida en el paso 2.3.5 y utilizando el coeficiente de Tc_{MCPhd} definido en el paso 2.3.4 se procede a seleccionar cuales pueden ser los posibles subgrafos moleculares $\{g_1, g_2, \dots, g_n\}$ relevantes a la actividad biológica en estudio. Para ello, se determina para cada lista L_i la lista de subgrafos $L_{g_i} = \{g_1, g_2, \dots, g_n\}$ que presenten el mayor valor de MCP al comparar cada para de grafos (G_i, G_{i+1}) de la lista L_i. Luego, a todos los subgrafos pertenecientes a las listas $\{L_{g_1}, L_{g_2}, \dots, L_{g_n}\}$ se les calcula el valor total del índice híbrido topográfico ϕ_{total} utilizando la Ecuación 9. Seguidamente, se obtiene la lista de subgrafos o fragmentos relevantes $Lg_{relevante}$ a partir de cada lista de subgrafos L_{g_i}. Para ello, se buscan los subgrafos estructuralmente iguales dentro de cada lista L_{g_i} contando las veces que se repiten y se inserta en la nueva lista $Lg_{relevante}$ de forma descendente por la cantidad de veces que se repiten. Finalmente se devuelve como resultado de este paso la lista de subgrafos $Lg_{relevante} = \{g_1, g_2, \dots, g_n\}$ como los fragmentos relevantes a la propiedad

biológica estudiada junto con el valor mínimo (ϕ_{min}) y máximo (ϕ_{max}) del valor total del índice atómico topográfico ϕ_{total}.

Resultados y discusión

Se empleó un conjunto de 36 compuestos pertenecientes a la serie 4-aminobiciclo[2.2.2]octan-2-y14-aminobutanoatos (Anexo 3) reportados por Weis et al. [47], evaluados contra la cepa K-1 multirresistente de *Plasmodium falciparum y* con el objetivo de demostrar la aplicabilidad del método a un caso real reportado en la literatura. Como el método 3DFrag-MCP para su funcionamiento utiliza la estructura tridimensional (3D) de los compuestos y la muestra a utilizar están en 2D, se utiliza el servicio de generación rápida de estructuras 3D con CORINA clásico [48], para logar obtener la estructura 3D de cada compuesto.

Otro análisis importante, es que a la muestra seleccionada no hoy que aplicarle ningún método de equilibrio de clases ya que se encuentra balanceados los compuestos en 17 activos, 18 inactivos y uno no evaluado, por lo tanto, no hay que aplicar el primer paso del método. Además, para obtener los fragmentos relevantes a la cepa K-1 multirresistente de *Plasmodium falciparum* se utilizará la propiedad electrostática a través del Índice de Estado Electrotopográfico para átomos (S_{3D}).

A los 36 compuestos 3D, se le calculan los descriptores atómicos S_{3D} como se muestra en el Anexo 4. Seguidamente se realiza una reducción del grafo molecular de cada compuesto y se le calcula el valor total ϕ_{total} del índice atómicos S_{3D} de cada DC perteneciente al nuevo grafo HRG.

Con los nuevos grafos HRG de los 36 compuestos se procede a cuantificar la similitud molecular a través del coeficiente Tc_{MCPhd} por pares de compuestos utilizando el índice S_{3D}, obteniéndose la matriz de similitud. A partir de esa matriz y aplicando un método de agrupamiento no jerárquico, en este caso el método K-Means, se obtienen las moléculas similares clasificadas como activas. Primeramente, para aplicar el agrupamiento es necesario determinar la cantidad de clústeres (k) óptimos, para ello utilizamos los métodos del codo, la silueta y la brecha. La Figura 9, muestra que al aplicar los tres métodos la cantidad de clústeres óptimos es 6.

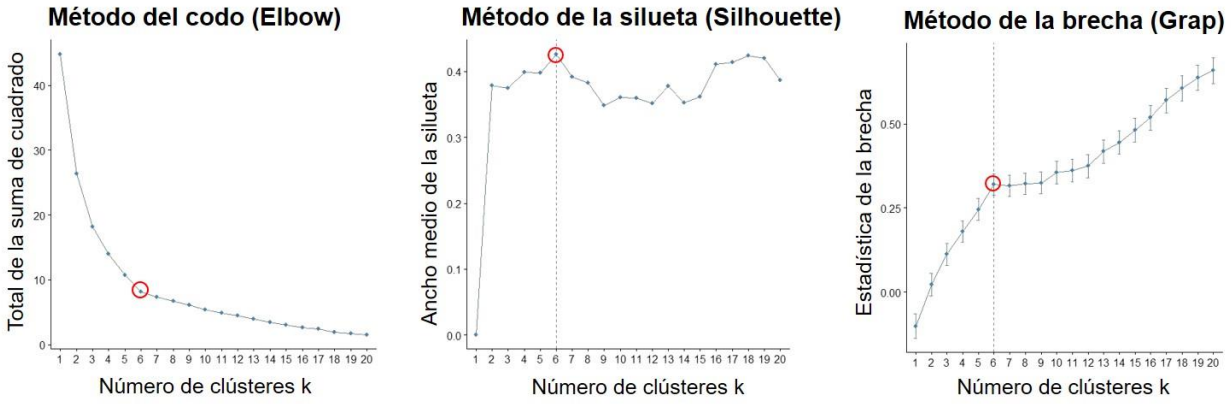


Figura 9. Número óptimo de clústeres por los métodos Elbow, Silhouette y Gap.

Los resultados del agrupamiento se muestran en la Tabla , donde se puede apreciar que los clústeres 5 y 6 agrupan el 100% de los compuestos activos e inactivos respectivamente. Los 7 compuestos activos agrupados en el clúster 5 (7c, 7h, 7i, 8c, 8f, 8i, 8f) presentan los valores de IC_{50} (0.06, 0.28, 0.26, 0.05, 0.09, 0.18 y 0.09) respectivamente, mientras que los 7 compuestos inactivos agrupados en el clúster 6 (6a, 6d, 6g, 6j, 7a, 7b, 7g) presentan los valor de IC_{50} (0.35, 0.71, 0.52, 2.16, 0.46, 0.55 y 0.40) respectivamente y el resto de los clústeres agruparon compuestos tanto activos como inactivos. Esto demuestra que, utilizando la matriz de similitud a partir del índice topográfico S_{3D} se logran agrupar correctamente 7 compuestos activos representando el 41,1% de los 17 presentes en la muestra. Además, de esos 7 compuestos, 4 de ellos (8c, 7c, 8f y 8l), el 57.1% de ese clúster, presentan los valores más pequeños de IC_{50}, es decir, los compuestos más activos del ensayo.

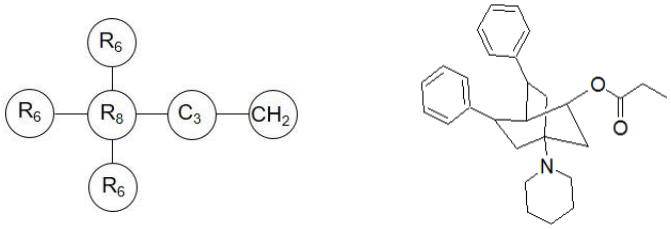
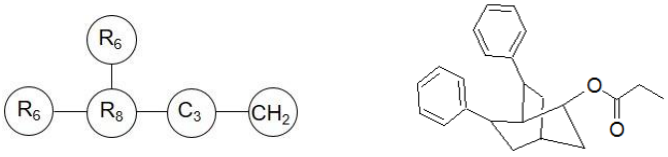
Tabla 1. Resultados del algoritmo K-Means con 6 clústeres.

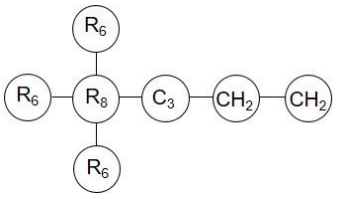
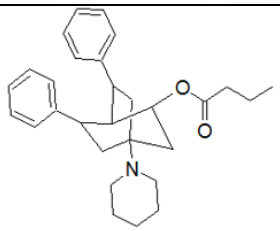
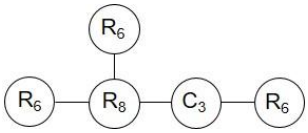
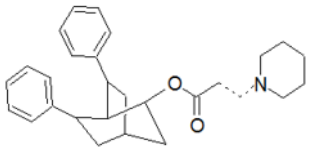
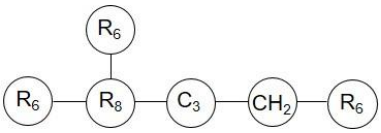
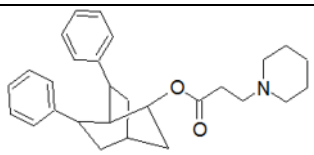
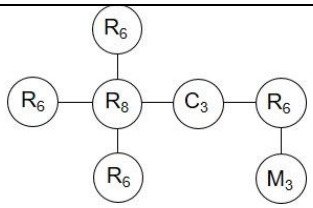
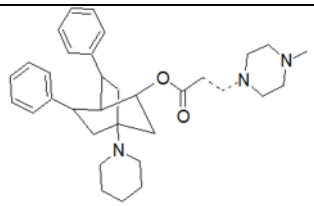
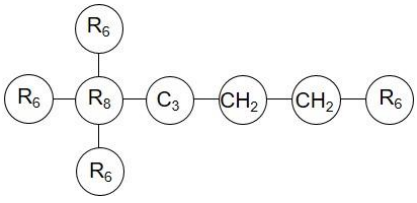
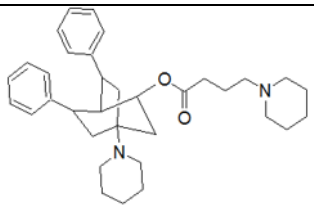
		Compuestos del cluster	Cantidad de moléculas		Total	% de moléculas	
			Activas	Inactivas		Activas	Inactivas
Clústeres	1	6b, 6h, 7b, 7e, 7k	1	4	5	20	80
	2	6c, 6f, 6c, 6l, 7f, 7l	4	1	6	66,6	16,6
	3	8a, 8d, 8g, 8j,7j	2	3	5	40	60
	4	6e, 6k, 6b, 8e, 8h, 8k	3	3	6	50	50

	5	7c,7h, 7i, 8c, 8f, 8i, 8l	7	0	7	100	0
	6	6a, 6d, 6g, 6j, 7a, 7b, 7g	0	7	7	0	100

Estos siete compuestos agrupados en el clúster 5 como activos se comparan por pares, obteniéndose de cada comparación los subgrafos o fragmentos g1 y g2 que presenten el mayor valor de similitud en la propiedad electrostática representada por el índice S_{3D} utilizando el coeficiente Tc_{MCPhd}, es importante señalar que solo hay que realizar 21 comparaciones de las 42 posibles, porque el coeficiente utilizado para cuantificar la similitud es bilateral. A cada fragmento obtenido se le calcula el valor total del índice híbrido topográfico ϕ_{total} y se guarda en una lista $L_g = \{g_1, g_2, g_3, \dots, g_n\}$. Luego, se buscan los fragmentos estructuralmente iguales dentro de la lista L_g contando las veces que aparecen en los 7 compuestos (nivel de importancia) y se insertan en una nueva lista $Lg_{relevante}$ ordenados de forma descendentes por su nivel de importancia. Finalmente, se devuelven los fragmentos de la lista $Lg_{relevante}$ como los fragmentos relevantes a la actividad biológica en estudio, como se muestra en la Tabla .

Tabla 2. Fragmentos relevantes a la cepa K-1 multirresistente de Plasmodium falciparum.

		$S_{3D_{total}}$	
Subgrafos relevantes	**Compuestos**	**Mínimo**	**Máximo**
	7c = 68,955 7i = 69,260 8c = 68,418 8f = 68,769 8i = 68,679 8l = 68,750	68,418	69,260
	7c = 54,626 7h = 55,037 8c = 54,097 8f = 54,383 8l = 54,383	54,097	55,037

		8c = 71,631 8f = 71,996 8i = 71,906 8l = 72,006	71,906	72,006
		7h = 65037 7i = 64,867 8i = 64,497	64,497	65037
		7h = 68,365 7i = 68,211	68,211	68,365
		7c = 83,085 8c = 82,720	82,720	83,085
		8f = 82,583 8l = 83,077	82,583	83,077

La identificación de fragmentos relevantes a una actividad biológica dada, como la mostrada en el ejemplo, sienta las bases para la minería de grafos en bases de datos de compuestos químicos que faciliten el hallazgo de nuevas propiedades en compuestos conocidos, al identificar fragmentos dentro de esas moléculas, que cumplen determinadas características químico-físicas evidenciadas por los descriptores híbridos

Conclusiones

Se propone el concepto de Propiedad Máxima Común como nuevo enfoque para cuantificar la similitud molecular utilizando descriptores grafo-teóricos en 3D llamados híbridos al estar ponderados por propiedades químico-físicas.

Se propone una nueva forma de reducción del grafo químico denominada Grafo Reducido Híbrido (HRG), basada en Centro Descriptores ponderados por descriptores híbridos y la distancia entre ellos.

Se propone, además, un nuevo método para identificar fragmentos relevantes a una actividad biológica denominado 3DFrag-MCP, el cual está basado en similitud de grafos moleculares lo cual allana el camino de desarrollo de nuevas entidades biológicamente activas o de detectar esos fragmentos en entidades conocidas con lo cual se pueden desarrollar nuevos fármacos.

Agradecimientos

Los autores desean expresar su agradecimiento a la Universidad de las Ciencias Informáticas y al Departamento de Ciencias de la Computación y Análisis Numérico de la Universidad de Córdoba por el financiamiento de esta investigación.

Referencias

1. Maggiora G, Vogt M, Stumpfe D, Bajorath J. (2013). Molecular Similarity in Medicinal Chemistry. J. Med. Chem 57:3186-3204. doi:10.1021/jm401411z
2. 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 doi.org/10.1023/A:1021271615909
3. Y. Cao, T. Jiang, T. Girke. (2008). A maximum common substructure–based algorithm for searching and predicting drug–like compounds, Bioinformatics 24 i366–i374.
4. E. Duesbury, J. D. Holliday, P. (2015). Willett, Maximum common substructure–based data fusion in similarity searching, J. Chem. Inf. Model. 55 222–230.
5. H. C. Ehrlich, M. Rarey. (2011). Maximum common subgraph isomorphism algorithms and their applications in molecular science: A review, WIREs Comput. Mol. Sci. 1 68–79.
6. I. Koch. (2001). Enumerating all connected maximal common subgraphs in two graphs, Theor. Comput. Sci. 250 1–30.
7. J. J. McGregor. (1982). Backtrack search algorithms and the maximal common subgraph problem, Software Pract. Exper. 12 23–34.

8. Ruggiu, F.; Gilles, M.; Varnek, A.; Horvath, D. (2010). ISIDA Property-Labelled Fragment Descriptors. Mol. Inform. 29, 855–868.
9. Liu, X., Jiang, H., Li, H. (2011). SHAFTS: a hybrid approach for 3D molecular similarity calculation. 1. Method and assessment of virtual screening. Journal of chemical information and modeling, 51(9), 2372-2385 doi: 10.1021/ci200060s.
10. Xiaofeng Liu, Honglin Li. (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):3564-74. doi: 10.1021/jm200139j
11. Hu, J., Liu, Z., Yu, D. J., Zhang, Y. (2018). LS-align: an atom-level, flexible ligand structural alignment algorithm for high-throughput virtual screening. Bioinformatics, 34(13), 2209-2218.
12. Antelo-Collado, A., Carrasco-Velar, R., García-Pedrajas, N., Cerruela-Garcia, G. (2020). Maximum common property: a new approach for molecular similarity. *J Cheminform* **12,** 61 doi.org/10.1186/s13321-020-00462-3
13. Carrasco R, Prieto JO, Antelo A, Padrón JA, Cerruela G, Maceo ´AL, Alcolea R, Silva LG. (2013). Hybrid Reduced Graph For SAR Studies. SAR and QSAR in Environmental Research 24:201-214. doi:10.1080/1062936X.2013.764926
14. Maggiora GM, Shanmugasundaram V. (2004). Molecular Similarity Measures. Methods Mol. Biol. 275:1-50. doi: 10.1385/1-59259-802-1:001
15. Zhang B, Vogt M, Maggiora GM. Bajorath J. (2015). Design Of Chemical Space Networks Using A Tanimoto Similarity Variant Based Upon Maximum Common Substructures. J. Comput Aided Mol Des 29:937-950. doi:10.1007/s10822-015-9872-1
16. Antelo-Collado A, Carrasco-Velar R, García-Pedrajas N, Cerruela-García G. (2021). Effective Feature Selection Method for Class-Imbalance Datasets Applied to Chemical Toxicity Prediction. J Chem Inf Model. 61(1):76-94. doi: 10.1021/acs.jcim.0c00908
17. Pérez-Rodríguez, J.; Haro-García, A. d.; Romero del Castillo, J. A.; García-Pedrajas, N. (2018). A general framework for boosting feature subset selection algorithms. Inf. Fusion. 44, 147−175.
18. de Haro-García, A.; Cerruela-García, G.; García-Pedrajas, N. (2020). Ensembles of Feature Selectors for dealing with Class-Imbalanced Datasets: A proposal and comparative study. Inf. Sci. 540,89− 116.
19. Song, Q.; Ni, J.; Wang, G. (2011) A fast clustering-based feature subset selection algorithm for high-dimensional data. IEEE Trans. Knowl. Data Eng. 25,1−14.

20. Hall, M. Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning. International Conference on Machine Learning, 2000.
21. Domingo, C.; Watanabe, O. (2000). MadaBoost: A Modification of AdaBoost. Proceedings of the 13th Annual Conference on Computational Learning Theory, pp 180-189.
22. Ting, K. M. (2000). A comparative study of cost-sensitive boosting algorithms. Proceedings of the 17th International Conference on Machine Learning: San Francisco, USA, pp 983−990.
23. Webb, G. I. (2000). MultiBoosting: A Technique for Combining Boosting and Wagging. Mach. Learn. 40, 159−196.
24. Galar, M.; Fernandez, A.; Barrenechea, E.; Bustince, H.; Herrera, F. (2011). A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. IEEE Trans. Syst. Man Cybern. C Appl. Rev. 42, 463−484.
25. Nikolaou, N.; Edakunni, N.; Kull, M.; Flach, P.; Brown, G. (2016). Cost-sensitive boosting algorithms: Do we really need them? Mach. Learn. 104, 359−384.
26. Kier LB, Hall LH. (1990). An Electrotopological-State Index for Atoms in Molecules. Pharm Res 7:801-807. doi:10.1023/A:1015952613760.
27. A.K. Ghose and G.M. Crippen, Atomic physicochemical parameters for three-dimensional structure-directed quantitative structure-activity relationships I. Partition coefficients as a measure of hydrophobicity, J. Comput. Chem, 7 (1986), pp. 565–577.
28. A.K. Ghose and G.M. Crippen, Atomic physicochemical parameters for three-dimensional structure- directed quantitative structure-activity relationships. 2. Modeling dispersive and hydrophobic interactions, J. Chem. Inf. Comp. Sci, 7 (1987), pp. 21–35.
29. A.K. Ghose, A. Pritchett, and G.M. Crippen, Atomic physicochemical parameters for three- dimensional structure directed quantitative structure-activity relationships III: Modeling hydrophobic interactions, J. Comput. Chem, 9 (1988), pp. 80–90.
30. Willighagen EL, Mayfield JW, Alvarsson J et. al. (2017) The Chemistry Development Kit (CDK) V2.0: Atom Typing, Depiction, Molecular Formulas, And Substructure Searching. J Cheminf 9:33. doi:10.1186/s13321-017-022.
31. Avidon VV, Pomerantsev IA, Golender VE, Rozenblit AB (1982) Structure-activity relationship oriented languages for chemical structure representation. J. Chem. Inf. Comp. Sci 22:207-214.

32. Steffen, A., Kogej, T., Tyrchan, C., Engkvist, O. (2009). Comparison of molecular fingerprint methods on the basis of biological profile data. Journal of chemical information and modeling, 49(2), 338-347.
33. Lance GN, Williams WT (1966) Computer programs for hierarchical polythetic classification ("similarity analysis"). Computer Journal, 9:60–64.
34. C. Ieva , A. Gotlieb , S. Kaci , N. Lazaar , Discovering program topoi via hierarchical agglomerative clustering, IEEE Trans. Reliab. PP (99) (2018) 1–13.
35. J.B. MacQueen, Some methods for classification and analysis of multivariate ob- servations, in: Proc. of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, 1967, pp. 281–297 . 1
36. H. Xiong , J.J. Wu , J. Chen , k-means clustering versus validation measures: a data-distribution perspective, IEEE Trans. Syst. Man Cybern. Part B 39 (2) (2009) 318–331
37. Richard G. Lawson and Peter C. Jurs, "Cluster analysis of acrylates to guide sampling for toxicity testing," Journal of Chemical Information and Computer Sciences, vol. 30, no. 2, pp. 137-144, 1990.
38. Richard E. Higgs, Kerry G. Bemis, Ian A. Watson, and James H. Wikel, "Experimental Designs for Selecting Molecules from Large Chemical Databases," Journal of Chemical Information and Computer Sciences, vol. 37, no. 5, pp. 861-870, 1997.
39. N Brown et al., "A chemoinformatics analysis of hit lists obtained from high-throughput affinity-selection screening," Journal of biomolecular screening, vol. 11, no. 2, p. 123-130, 2006.
40. Faisal Saeed, Naomie Salim, Ammar Abdo, and Hamza Hentabli, "Combining Multiple Clusterings of Chemical Structures Using Cumulative Voting-based Aggregation Algorithm," in Proceedings of the 5th Asian Conference on Intelligent Information and Database Systems - Volume Part II, Berlin, Heidelberg, pp. 178-185, 2013.
41. E. Fersini, I. Giordani, E. Messina, and F. Archetti, "Relational clustering and Bayesian networks for linking gene expression profiles and drug activity patterns," in Bioinformatics and Biomedicine Workshop, 2009. BIBMW 2009. IEEE International Conference, pp. 20- 25, 2009.
42. Denis M. Bayada, Hans Hamersma, and Vincent J. van Geerestein, "Molecular Diversity and Representativity in Chemical Databases," Journal of Chemical Information and Computer Sciences, vol. 39, no. 1, pp. 1-10, 1999.

43. Vincent J. van Geerestein, Hans Hamersma, and Steven P. van Helden, "Exploiting Molecular Diversity: Pharmacophore Searching and Compound Clustering," in Computer-Assisted Lead Finding and Optimization.: Verlag Helvetica Chimica Acta, pp. 157-178, 2007.
44. Thorndike, R.L. Who belongs in the family?. Psychometrika 18, 267–276 (1953). doi.org/10.1007/BF02289263.
45. David J Wild and C.John Blankley, "VisualiSAR: a Web-based application for clustering, structure browsing, and structure-activity relationship study ," Journal of Molecular Graphics and Modelling , vol. 17, no. 2, pp. 85-89, 1999.
46. Noviandy, Teuku Rizky & Maulana, Aga & Sasmita, Novi & Suhendra, Rivansyah & Muslem, Muslem & Idroes, Ghazi & Paristiowati, Maria & Helwani, Zuchra & Yandri, Erkata & Rahimah, Souvia & Muhammad, & Irvanizam, Irvanizam & Idroes, Rinaldi. (2021). The implementation of K-Means clustering in kovats retention index on gas chromatography. IOP Conference Series: Materials Science and Engineering. 1087. 012051. 10.1088/1757-899X/1087/1/012051.
47. Weis R, Seebacher W, Brun R, Kaiser M, Sat R, Faist J (2013) 4-Aminobicyclo[2.2.2]octan-2-yl 4-aminobutanoates with antiprotozoal activity. Monatsh Chem. doi:10.1007/s00706-013-1116-2.
48. Fast 3D Structure Generation with CORINA Classic (2020). https://www.mn-am.com/online demos/corina demo. Accessed 18 Feb 2020.