

Similitud molecular empleando Índices Híbridos

Molecular similarity using Hybrid Indices

Aurelio Antelo-Collado,^I Juan L. Paneque-Pérez,^{II} María C. Hernández-Govea,^{III} Ramón Carrasco-Velaz^{IV}

I Universidad de las Ciencias Informáticas, Carretera a San Antonio de los Baños km 2½, La Habana, Cuba. E-mail: aaantelo@uci.cu

II Universidad de las Ciencias Informáticas, La Habana, Cuba. E-mail: jlpaneque@uci.cu

III Universidad de las Ciencias Informáticas, La Habana, Cuba. E-mail: mcgovea@uci.cu

IV Universidad de las Ciencias Informáticas, La Habana, Cuba. E-mail: rcarrasco@uci.cu

RESUMEN

Se presenta un método para la detección de semejanza entre moléculas basado en macheo inexacto de grafos. Se parte del grafo molecular completo ponderado en sus vértices por propiedades químico-físicas particionadas sobre los mismos, se reduce el grafo por el procedimiento CALEDE que define Centros Descriptores o fragmentos de primer orden, los cuales son subgrafos ponderados por la suma de los valores de los vértices ponderados individualmente a su vez, y se construyen fragmentos denominados de segundo orden que incluyen la distancia entre los centros de masas de ambos centros descriptores. Se presenta el método de búsqueda aplicado a una base de datos de más de 300 moléculas con sus respectivas estructuras en tres dimensiones. Esos compuestos se encuentran evaluados como anticancerígenos en la base de datos de compuestos del NCBI-USA. En el experimento computacional se encuentra que, en dependencia de la función de similitud empleada, es posible detectar compuestos que a pesar de poseer diferente topología, poseen valores de las propiedades empleadas para el macheo lo cual sugiere la presencia de potenciales farmacóforos como hallazgo relevante, lo cual constituiría un novedoso enfoque para el diseño computacional de fármacos.

Palabras Clave: similitud molecular, índices híbridos, CALEDE.

ABSTRACT

A method for detecting similarity between molecules based on inexact matching graph is presented. We start from a complete molecular graph vertices weighted by several hybrid indices. The molecular graph is reduced by CALEDE procedure, which define descriptors centers or first order fragments. These fragments are subgraphs weighted with the sum of values of the vertices weighted with the hybrid indices. It also define second order fragments by including the distance between the centers of mass of both descriptors centers. The search method applied to a database of over 300 molecules with their respective threedimensional structures is presented. These compounds are reported in the NCBI-USA database of compounds which were evaluated in anticancer tests. In the computational experiment, depending on the similarity function used, is possible to detect compounds that despite having different topology have property values suggesting the presence of potential pharmacophore. It suggest the possibility to use this approach as a novel approach for computational drug design.

KeyWords: molecular similarity, hybrid índices, CALEDE.

INTRODUCCIÓN

Se ha demostrado que los índices híbridos¹ propuestos por Carrasco y cols. para la descripción simultánea de propiedades químico-físicas y la estructura química permiten describir y representar una molécula o fragmento de ella de tres formas diferentes utilizando un procedimiento computacional y matemáticamente sencillo. Permite también comparar estructuras diferentes a partir de la utilización de dichos índices, lo cual ha abierto ha abierto una nueva puerta para el trabajo de búsquedas intensas en bases de datos.

En las bases de datos de moléculas orgánicas es posible encontrar grandes diferencias estructurales. Cuando el objeto de la base es el almacenamiento de moléculas biológicamente activas, se encuentra entonces que la elevada diferencia estructural entre los compuestos revela dos diferentes paradojas estructurales; compuestos estructuralmente semejantes presentan actividades distintas, o compuestos estructuralmente diferentes presentan similar actividad. Con los nuevos descriptores híbridos se pretende contribuir a la resolución de este problema.

Son múltiples y diferentes los métodos para la realización de búsquedas en bases de datos de estructuras químicas. La utilización de medidas de semejanza o similitud entre grafos (macheo exacto de grafos), es uno de los procedimientos más empleados históricamente en esta área de la química grafo teórica. En el presente trabajo se presentan los procedimientos desarrollados para el establecimiento de los criterios de similitud diferencia entre fragmentos o moléculas utilizando los índices híbridos, en un macheo inexacto de grafos.

SOFTWARE Y PROCEDIMIENTOS

Datos

El conjunto de datos de trabajo es el ensayo AID941 del NCBI.² Está formado por un conjunto de 330 moléculas. El formato de fichero de datos estructurales utilizado de las moléculas fue el *.mol o *.sdf. Las estructuras químicas se visualizan empleando el Jmol, visualizador de Java de código abierto que realiza la representación gráfica tridimensional de alto rendimiento sin requerimientos de hardware, ya que solo precisa de la instalación de la Máquina Virtual de Java.

Descriptores topográficos

Como descriptores topográficos a utilizar se emplearon los índices de Estado Refracto, Electro y Lipotopográficos para Átomos desarrollados por Carrasco y cols.¹ Para el cálculo de estos índices se utilizó la librería Chemistry Development Kit (CDK),³ una librería de código abierto programada en Java que está destinada para la realización de cálculos en química computacional, la química y la bioinformática. La reducción del grafo molecular se realizó con el algoritmo CALEDE,⁴ definido por Carrasco y cols. e implementado en la librería CDK. Mediante CALEDE se generan fragmentos moleculares de segundo orden definidos como dos Centros Descriptores (CD) más la distancia euclidiana entre los centros de masas de los respectivos CD. Los Centros Descriptores de CALEDE (fragmentos de primer orden) son agrupaciones típicas de grafos como ciclos o anillos, estrellas de orden 3 y 4, los halógenos, oxígeno, nitrógeno y los grupos metilo, metileno y metino.

Funciones de similitud

Como funciones de similitud se emplearon las de Sørensen, Dice-Sørensen, Tanimoto, Jaccard, Ruzicka, Cze-kanowski y Soergel, incluidas en la revisión de Sung⁵ por considerarse que son las que mejor se adaptan al cálculo de similitud molecular, basados en vectores de valores.

RESULTADOS Y DISCUSIÓN

Cálculo de los umbrales de similitud a utilizar

Un elemento a tener presente cuando se realizan cálculos de semejanza mediante el empleo de funciones de similitud, lo constituye el grado de error permisible que se utilizará. Es por ello, que antes de realizar las pruebas para validar los algoritmos implementados, se hace indispensable determinar el umbral de máxima diferencia o valor de mínima similitud de cada una de las funciones seleccionadas, el cual se estará empleando en las búsquedas que se realizarán y constituirán las propuestas a umbrales de esta investigación, aunque el criterio de determinación puede cambiar, por lo que queda a decisión del investigador establecer el umbral que desee para sus experimentos. Para determinar los umbrales se escogieron arbitrariamente once fragmentos moleculares de segundo orden, pertenecientes a moléculas presentes en el ensayo AID941⁶, con los cuales se realizaron búsquedas para ajustar los límites deseados, cada uno de los fragmentos seleccionados se examinaron con los fragmentos presentes en cada una de las 330 moléculas contenidas en el propio ensayo AID941. En los experimentos realizados se considera un resultado correcto cuando los fragmentos obtenidos pertenecen al mismo tipo de fragmento seleccionado e incorrecto en caso contrario.

En la tabla 1 se muestra el procedimiento realizado para el coeficiente de Tanimoto. Inicialmente se realiza la búsqueda para el fragmento compuesto por un anillo y un heteroátomo de oxígeno, tomando un umbral lo suficientemente amplio, para que se obtuvieran resultados correctos e incorrectos, obteniéndose 76 fragmentos del mismo tipo del fragmento seleccionado y 1 de otro tipo (con un valor de 0.1996 de diferencia), por lo que es necesario disminuir el umbral hasta 0.199. Con este nuevo umbral, se realiza una nueva búsqueda, tomando un fragmento compuesto por un anillo y un heteroátomo de nitrógeno, obteniéndose 33 bien y 69 mal (con valores entre 0.0911 y 0.1987 de diferencia), por lo que se hace necesario un nuevo ajuste del umbral de forma tal que sea menor que el menor valor de diferencia obtenido dentro de los resultados incorrectos, fijándose el nuevo umbral hasta 0.091. A continuación, y con este nuevo valor, se realiza nuevamente la búsqueda, pero con un fragmento compuesto por un anillo y un heteroátomo de cloro, encontrándose 2 fragmentos correctos, por lo que no es necesario disminuir el umbral. Este procedimiento se realiza para cada uno de los 7 restantes fragmentos, modificando el valor si se encuentran resultados incorrectos y manteniéndolo igual en otro caso.

Tabla 1. Ajuste del umbral para el coeficiente de Tanimoto

	Fragmento	Umbral	Bien		Mal			Molecula
			Cant	Max	Cant	Min	Max	
Tanimoto	Anillo-Heteroatomo[O]	0,200	76	0,1999	1	0,1996	0,1996	701067.mol
	Anillo-Heteroatomo[N]	0,199	33	0,1955	69	0,0911	0,1987	815515.mol
	Anillo-Heteroatomo[Cl]	0,091	2	0,0130	0			16189701.mol
	Anillo-Heteroatomo[Br]	0,091	5	0,0910	0			203396.mol
	Anillo-Heteroatomo[S]	0,091	5	0,0850	0			1253074.mol
	Anillo-Metilo	0,091	4	0,0850	1	0,0790	0,0790	664134.mol
	Anillo-Heteroatomo[F]	0,078	22	0,0670	0			675202.mol
	Anillo-Anillo	0,078	3	0,0680	0			665132.mol
	Anillo-Cluster	0,078	2	0,0500	0			661810.mol
	Cluster-Heteroatomo[N]	0,078	3	0,0767	0			1005277.mol
	Cluster-Cluster	0,078	1	0,0000	0			661810.mol

El método descrito anteriormente se realizó para cada una de las funciones de similitud y de distancia seleccionadas para esta investigación, tomando como base los fragmentos seleccionados. Cabe destacar que en el caso de las funciones de similitud el ajuste de umbral se realizó tomando un valor mayor que el mayor valor de similitud obtenido dentro de los resultados incorrectos y en el caso de las funciones de distancia, como se explicó anteriormente en el caso del coeficiente de Tanimoto. En las tablas 2-3- 4-5-6-7 se muestran los valores obtenidos en los experimentos realizados.

Tabla 2. Ajuste del umbral para el índice de Jaccard

	Fragmento	Umbral	Bien		Mal			Molecula
			Cant	Min	Cant	Min	Max	
Jaccard	Anillo-Heteroatomo[O]	0,950	168	0,9503	17	0,9510	0,9779	701067.mol
	Anillo-Heteroatomo[N]	0,978	13	0,9782	6	0,9820	0,9946	5336330.mol
	Anillo-Heteroatomo[Cl]	0,995	2	0,9998	0			16189701.mol
	Anillo-Heteroatomo[Br]	0,995	4	0,9990	0			1487485.mol
	Anillo-Heteroatomo[S]	0,995	2	0,9971	0			3243696.mol
	Anillo-Metilo	0,995	3	0,9952	0			664134.mol
	Anillo-Heteroatomo[F]	0,995	20	0,9958	0			2136935.mol
	Anillo-Anillo	0,995	10	0,9967	0			2136935.mol
	Anillo-Cluster	0,995	1	1,0000	0			3243630.mol
	Cluster-Heteroatomo[N]	0,995	1	1,0000	0			5336330.mol
	Cluster-Cluster	0,995	1	1,0000	0			5336330.mol

Tabla 3. Ajuste del umbral para el coeficiente de Dice-Sørensen

	Fragmento	Umbral	Bien		Mal			Molecula
			Cant	Min	Cant	Min	Max	
Dice-Sørensen	Anillo-Heteroatomo[O]	0,980	134	0,9806	6	0,9801	0,9888	701067.mol
	Anillo-Heteroatomo[N]	0,989	12	0,9894	6	0,9909	0,9973	5336330.mol
	Anillo-Heteroatomo[Cl]	0,997	5	0,9974	0			2905495.mol
	Anillo-Heteroatomo[Br]	0,997	4	0,9994	0			203396.mol
	Anillo-Heteroatomo[S]	0,997	4	0,9970	0			649054.mol
	Anillo-Metilo	0,997	7	0,9975	0			649054.mol
	Anillo-Heteroatomo[F]	0,997	28	0,9971	0			5014184.mol
	Anillo-Anillo	0,997	1	1,0000	4	0,9972	0,9976	649054.mol
	Anillo-Cluster	0,998	2	0,9990	0			661810.mol
	Cluster-Heteroatomo[N]	0,998	2	0,9992	0			11835305.mol
	Cluster-Cluster	0,998	1	1,0000	0			661810.mol

Tabla 4. Ajuste del umbral para el coeficiente de Sørensen

	Fragmento	Umbral	Bien		Mal			Molecula
			Cant	Max	Cant	Min	Max	
Sørensen	Anillo-Heteroatomo[O]	0,200	243	0,1999	41	0,1214	0,1988	2136882.mol
	Anillo-Heteroatomo[N]	0,120	229	0,1199	14	0,0825	0,1188	11839287.mol
	Anillo-Heteroatomo[Cl]	0,080	1	0,0000	0			727631.mol
	Anillo-Heteroatomo[Br]	0,080	3	0,0788	0			1893291.mol
	Anillo-Heteroatomo[S]	0,080	24	0,0775	1	0,0706	0,0706	2377612.mol
	Anillo-Metilo	0,070	81	0,0700	4	0,0484	0,0683	5771267.mol
	Anillo-Heteroatomo[F]	0,048	21	0,0422	0			6260437.mol
	Anillo-Anillo	0,048	4	0,0430	0			665132.mol
	Anillo-Cluster	0,048	6	0,0358	2	0,0471	0,0476	1215664.mol
	Cluster-Heteroatomo[N]	0,047	5	0,0456	0			1005277.mol
	Cluster-Cluster	0,047	2	0,0079	0			12005721.mol

Tabla 5. Ajuste del umbral para el coeficiente de Soergel

	Fragmento	Umbral	Bien		Mal			Molecula
			Cant	Max	Cant	Min	Max	
Soergel	Anillo-Heteroatomo[O]	0,300	180	0,2999	12	0,2165	0,2997	2136883.mol
	Anillo-Heteroatomo[N]	0,217	246	0,2167	14	0,1524	0,2123	11839287.mol
	Anillo-Heteroatomo[Cl]	0,153	1	0,0000	0			727631.mol
	Anillo-Heteroatomo[Br]	0,153	3	0,1460	0			1893291.mol
	Anillo-Heteroatomo[S]	0,153	27	0,1525	1	0,1319	0,1319	2377612.mol
	Anillo-Metilo	0,130	79	0,1296	4	0,0924	0,1279	5771267.mol
	Anillo-Heteroatomo[F]	0,092	21	0,0809	0			6260437.mol
	Anillo-Anillo	0,092	4	0,0824	0			665132.mol
	Anillo-Cluster	0,092	6	0,0691	2	0,0900	0,0909	1215664.mol
	Cluster-Heteroatomo[N]	0,090	5	0,0873	0			1005277.mol
	Cluster-Cluster	0,090	2	0,0157	0			12005721.mol

Tabla 6. Ajuste del umbral para el coeficiente de Ruzicka

	Fragmento	Umbral	Bien		Mal			Molecula
			Cant	Min	Cant	Min	Max	
Ruzicka	Anillo-Heteroatomo[O]	0,750	101	0,7534	1	0,7835	0,7835	2136883.mol
	Anillo-Heteroatomo[N]	0,784	238	0,7842	14	0,7877	0,8476	11839287.mol
	Anillo-Heteroatomo[Cl]	0,848	1	1,0000	0			727631.mol
	Anillo-Heteroatomo[Br]	0,848	3	0,8526	0			1893291.mol
	Anillo-Heteroatomo[S]	0,848	26	0,8500	1	0,8681	0,8681	2377612.mol
	Anillo-Metilo	0,869	83	0,8690	5	0,8690	0,9076	5771267.mol
	Anillo-Heteroatomo[F]	0,908	21	0,9191	0			6260437.mol
	Anillo-Anillo	0,908	4	0,9176	0			665132.mol
	Anillo-Cluster	0,908	6	0,9309	2	0,9091	0,9100	1215664.mol
	Cluster-Heteroatomo[N]	0,910	5	0,9127	0			1005277.mol
	Cluster-Cluster	0,910	2	0,9843	0			12005721.mol

Tabla 7. Ajuste del umbral para el coeficiente de Czekanowski

	Fragmento	Umbral	Bien		Mal			Molecula
			Cant	Min	Cant	Min	Max	
Czekanowski	Anillo-Heteroatomo[O]	0,850	123	0,8503	1	0,8786	0,8786	2136883.mol
	Anillo-Heteroatomo[N]	0,879	238	0,8790	14	0,8812	0,9175	11839287.mol
	Anillo-Heteroatomo[Cl]	0,918	1	1,0000	0			727631.mol
	Anillo-Heteroatomo[Br]	0,918	3	0,9205	0			1893291.mol
	Anillo-Heteroatomo[S]	0,918	26	0,9189	1	0,9294	0,9294	2377612.mol
	Anillo-Metilo	0,930	81	0,9300	4	0,9317	0,9516	5771267.mol
	Anillo-Heteroatomo[F]	0,952	21	0,9578	0			6260437.mol
	Anillo-Anillo	0,952	4	0,9570	0			665132.mol
	Anillo-Cluster	0,952	6	0,9642	2	0,9524	0,9529	1215664.mol
	Cluster-Heteroatomo[N]	0,953	5	0,9544	0			1005277.mol
	Cluster-Cluster	0,953	2	0,9921	0			12005721.mol

En la tabla 8 se presenta el resumen de los umbrales para cada una de las 7 funciones de similitud/distancia seleccionadas para esta investigación.

Tabla 8. Umbrales calculados para las funciones de similitud/distancia

Función	Umbral
Jaccard	0,995
Tanimoto	0,078
Dice-Sorensen	0,998
Sørensen	0,047
Soergel	0,09
Ruzicka	0,91
Czekanowski	0,953

Es necesario aclarar que los valores de umbral calculados, fueron obtenidos a partir del algoritmo de búsqueda de fragmentos simples implementado y teniendo en cuenta las tres propiedades químico-físicas para cada uno de los fragmentos analizados, estos valores pueden cambiar para las búsquedas en las que solo se tenga en cuenta uno de las propiedades.

Utilización de la fragmentación y los índices híbridos

No todas las moléculas cumplen el principio de Johnson y Maggiora⁴ de que compuestos estructuralmente semejantes exhiben propiedades semejantes. Como contradicción a este postulado existen las paradojas estructurales, expresadas en que moléculas similares estructuralmente tienden a exhibir propiedades biológicas diferentes y moléculas diferentes estructuralmente tienden a exhibir propiedades biológicas similares. Por lo antes planteado, el cotejo exacto de grafos no constituye una elección a la hora de realizar búsquedas de similitud molecular. La necesidad de realizar comparaciones entre moléculas y encontrar un método para determinar similitud molecular, nos condujo al empleo de descriptores híbridos ponderados por propiedades químico-físicas, específicamente el índice de estado Refractotopográfico para átomos, el índice de estado Electrotopográfico para átomos y el índice de estado Lipotopográfico para átomos, los cuales reportados anteriormente.

En la figura 1 se muestra el comportamiento del valor del índice del estado refractotopológico. Puede verse la diferencia de valor del índice, dada la diferente distribución espacial de estos compuestos, los cuales son considerados isómeros de posición por su diferente distribución en la estructura. Obsérvese como los átomos externos varían su valor de una estructura a la otra y como disminuye apreciablemente el valor del átomo situado geométricamente al centro. Esta variación está dada por la influencia que ejercen el resto de los átomos sobre los restantes de la molécula.

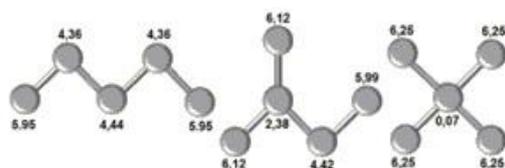


Fig. 1. Valores del índice del estado Refractotopográfico de átomos en fragmentos moleculares

Otro ejemplo que ilustra lo planteado anteriormente lo constituyen los fragmentos mostrados en la figura 2, donde se muestra que los valores del Índice del Estado Refractotopográfico de los átomos que conforman a los fragmentos moleculares varían al cambiar uno de los átomos presentes en los mismos, evidenciándose que el índice revela el comportamiento de las propiedades químico-físicas y cómo esta varía, no solo en dependencia del tipo de átomo y de la estructura que presenten los fragmentos, sino del efecto que ejercen los átomos vecinos. Este hecho sugiere el empleo de los índices para estudios de similitud molecular.

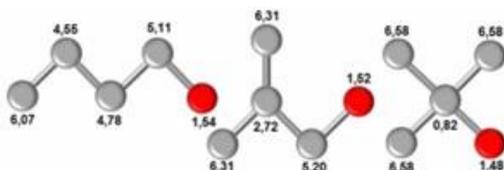


Fig. 2. Valores del índice del estado refractotopográfico de los átomos en fragmentos moleculares con un átomo de oxígeno (en rojo)

En la figura 3 se muestran fragmentos estructuralmente iguales, que difieren debido a la existencia de diferentes átomos pesados (F, Cl, Br, I), cuyos valores del índice del estado refractotopográfico son distintos, y la influencia de cada uno de estos sobre el resto provoca que estos presenten valores diferentes en sus propiedades, consecuentemente presentan actividad biológica diferente. Por lo antes planteado el empleo de los índices de Carrasco y cols. constituye uno de los ejes de esta investigación.

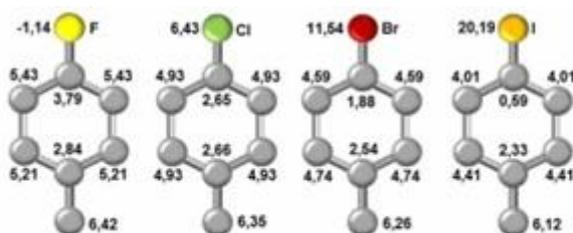
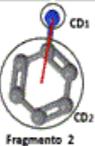
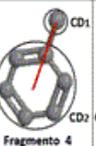
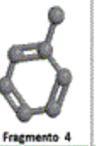
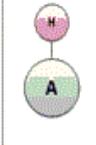
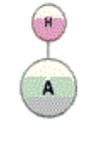
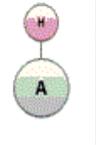
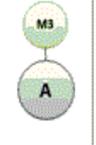
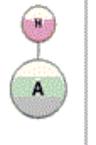


Fig. 3. Fragmentos moleculares estructuralmente similares

Otro elemento fundamental lo constituye la fragmentación molecular utilizando los centros descriptores de CA-LEDE. Como se aprecia en la tabla 9, los subgrafos mostrados son topológicamente idénticos y se han definido como fragmentos de segundo orden al estar formados por un anillo y un heteroátomo, excepto el cuarto que está formado por un anillo y un metilo.

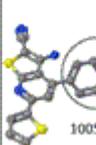
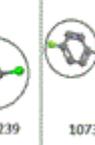
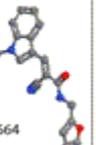
Tabla 9. Valores de los índices que presentan fragmentos moleculares de segundo orden topológicamente idénticos y topográficamente similares

				
Fragmento 1	Fragmento 2	Fragmento 3	Fragmento 4	Fragmento 5
				
Fragmento 1	Fragmento 2	Fragmento 3	Fragmento 4	Fragmento 5
CD1: Heteroátomo ETPG : -0,1620 RTPG : 6,7320 LTPG : 1,4640	CD1: Heteroátomo ETPG : 1,8693 RTPG : 8,2311 LTPG : -0,7067	CD1: Heteroátomo ETPG : -5,9919 RTPG : 13,6820 LTPG : -1,5306	CD1: Heteroátomo ETPG : 9,5916 RTPG : 8,2640 LTPG : 0,8924	CD1: Heteroátomo ETPG : 4,9754 RTPG : -0,4700 LTPG : -0,5504
CD2: Anillo ETPG : 14,5466 RTPG : 23,8616 LTPG : 1,7429 Dia: 3,1172	CD2: Anillo ETPG : 14,7585 RTPG : 30,5061 LTPG : 2,2154 Dia: 2,6166	CD2: Anillo ETPG : 13,6019 RTPG : 30,3202 LTPG : 3,3098 Dia: 3,1235	CD2: Anillo ETPG : 6,4621 RTPG : 24,8518 LTPG : 2,3127 Dia: 2,2642	CD2: Anillo ETPG : 12,5159 RTPG : 30,9916 LTPG : 3,3593 Dia: 2,1945
				

Debe acotarse que las propiedades químicas de un grupo metilo difieren apreciablemente a las de un heteroátomo, aunque sea comparable en cuanto a tamaño con respecto al bromo o el yodo.

Otro de los criterios que permite validar la utilización de la fragmentación propuesta, en los estudios de similitud molecular se evidencia en la tabla 10.

Tabla 10. Resultado de la búsqueda de fragmentos en diferentes posiciones

			
CD1: Heteroátomo [-0,1620 6,7320 1,4640]	CD1: Heteroátomo [1,8775 -3,7556 0,5748]	CD1: Heteroátomo [-0,5302 7,6128 1,0270]	CD1: Heteroátomo [0,0643 6,0677 1,4694]
CD2: Anillo [14,5466 23,8616 1,7429]	CD2: Anillo [13,1968 29,5066 1,8547]	CD2: Anillo [12,6794 26,6897 2,5196]	CD2: Anillo [15,1821 26,9642 1,8999]
			
1005239	1073564	1090196	1253074

En la misma se muestra la realización de cuatro búsquedas tomando varios fragmentos que estructuralmente son idénticos, pero como se evidencia las propiedades químico-físicas de cada uno de ellos son diferentes, este cambio se debe fundamentalmente a la posición espacial con respecto al resto de los átomos que conforman a las moléculas a las cuales pertenecen, debido a que en algunas el fragmento se encuentra en un extremo y en otros casos se encuentra al centro, por lo que la influencia sobre el fragmento específico varía en los diferentes casos.

En la tabla se muestra la cantidad de resultados encontrados para cada una de las posiciones, según se muestra en la molécula en la parte inferior de la tabla.

Los criterios planteados anteriormente sugieren que para realizar estudios de similitud molecular no se debe tener en cuenta solamente métodos que trabajen a partir de la estructura de las moléculas, sino que es razonablemente aconsejable, explorar alternativas como la propuesta en este trabajo, mediante el empleo de una forma de fragmentación de las moléculas, que aporte información no solo estructural, sino además de propiedades asociadas a esas estructuras como la que brindan el algoritmo CALEDE de fragmentación y los descriptores híbridos

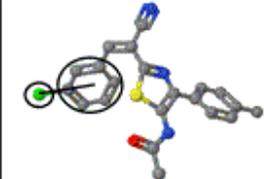
ponderados con propiedades químico físicas particionadas sobre los vértices del grafo molecular.

Búsquedas de fragmentos simples

Los fragmentos de segundo orden son, como se explicó anteriormente, los subgrafos del grafo reducido de CALEDE que están formados por dos Centros Descriptores. A partir de estos fragmentos se forma un vector formado por los valores de los índices Electrotopográfico, Refractotopográfico y Lipotopográfico de cada Centro Descriptor y la distancia entre los centros de masas de cada uno de ellos. Esta representación en vectores de números reales, permite la aplicación de funciones matemáticas, que determinen la semejanza entre ellos.

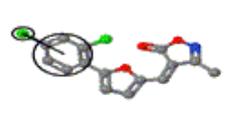
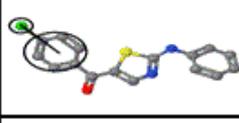
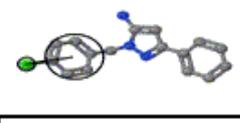
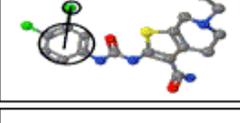
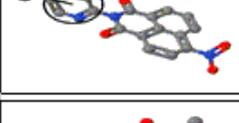
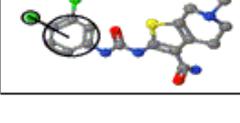
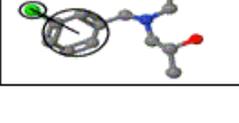
En la tabla 11 se muestra el fragmento A6-HCl1 perteneciente a la molécula 12005721 del ensayo AID941, en la misma aparecen los valores de los y la distancia entre los centros descriptores que lo forman.

Tabla 11. Descripción del fragmento A6-HCl1 perteneciente a la molécula 12005721

		Anillo6	Cl1
	ETPG	14,498	-0,227
	RTPG	24,123	6,856
	LTPG	1,720	1,499
	Dist	3,123	

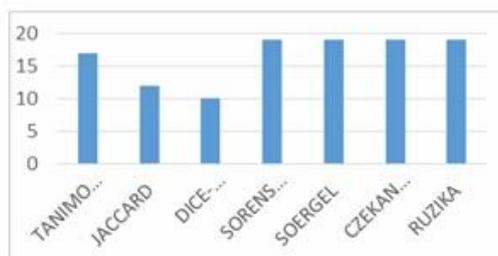
Tomando como objetivo el fragmento mostrado en la tabla 11, se realizaron varias ejecuciones del algoritmo implementado para la búsqueda de fragmentos similares aplicando cada una de las funciones de similitud seleccionadas. En la tabla 12 se muestran los fragmentos en común que fueron encontrados por cada una de las funciones de similitud.

Tabla 12. Fragmentos comunes para cada una de las funciones de similitud

		Anillo6	CI1			Anillo6	CI2
	ETPG	14,547	-0,163		ETPG	13,900	-0,011
	RTPG	23,862	6,732		RTPG	24,701	6,853
	LTPG	1,743	1,464		LTPG	0,617	1,414
	Dist	3,117			Dist	3,117	
		Anillo6	CI1			Anillo6	CI1
	ETPG	15,061	-0,030		ETPG	13,866	-0,104
	RTPG	24,904	6,686		RTPG	24,396	6,701
	LTPG	1,550	1,398		LTPG	1,741	1,418
	Dist	3,117			Dist	3,119	
		Anillo6	CI1			Anillo6	CI1
	ETPG	16,286	-0,001		ETPG	13,770	-0,073
	RTPG	24,618	6,692		RTPG	25,577	7,122
	LTPG	1,621	1,422		LTPG	1,577	1,541
	Dist	3,118			Dist	3,121	
		Anillo6	CI1			Anillo6	CI2
	ETPG	13,967	-0,104		ETPG	13,770	0,031
	RTPG	25,075	7,183		RTPG	25,577	6,913
	LTPG	1,585	1,541		LTPG	1,577	1,460
	Dist	3,125			Dist	3,119	
		Anillo6	CI1			Anillo6	CI1
	ETPG	12,955	-0,208		ETPG	12,955	-0,208
	RTPG	24,034	6,736		RTPG	24,034	6,736
	LTPG	2,085	1,481		LTPG	2,085	1,481
	Dist	3,118			Dist	3,118	

A partir del análisis de los resultados, se evidencia que el algoritmo implementado permite encontrar aquellos fragmentos dentro del ensayo que poseen gran semejanza en los valores de los índices con el fragmento objetivo y a la vez presentan una notable semejanza estructural.

En la figura 4 se muestra la cantidad de resultados devueltos por el algoritmo, por cada una de las funciones de similitud seleccionadas.


Fig. 4. Cantidad de resultados devueltos por el algoritmo, por cada una de las funciones de similitud seleccionadas

De las nueve funciones de similitud evaluadas, solo Jaccard y DiceSorensen devuelven un menor número de fragmentos. El procedimiento aquí presentado introduce el enfoque del macheo inexacto de grafos moleculares para grafos ponderados por propiedades químico-físicas particionadas, lo cual puede resultar una nueva vía de aproximación a la detección de regiones de moléculas involucradas directamente en la respuesta biológica, determinado por algunas propiedades químico-físicas, con lo cual se abriría una nueva opción para el diseño computacional de fármacos.

CONCLUSIONES

Se desarrolló un procedimiento de búsqueda de fragmentos moleculares basado en el método CALEDE de fragmentación de moléculas ponderadas por los índices topográficos híbridos de Carrasco et.al. y el principio del macheo inexacto de grafos. El empleo de diferentes funciones de similitud mostró la factibilidad del método empleado para extraer información de diferente naturaleza de bases de datos de estructuras químicas.

AGRADECIMIENTOS

Los autores, agradecen las facilidades brindadas por la Universidad de Ciencias Informáticas para la realización de este trabajo.

REFERENCIAS BIBLIOGRÁFICAS

1. Carrasco-Velar R, Prieto-Entenza J.O, Antelo-Collado A, Padrón-García J.A, Cerruela-García G, Maceo-Pixa Á.L, Alcolea-Núñez R, Silva-Rojas L.G. (2013). Hybrid reduced graph for SAR studies, SAR and QSAR in Environmental Research. DOI:10.1080/1062936X.2013.764926.
2. NCBI. PubChem BioAssay Database AID941. Disponible en: <http://www.ncbi.nlm.nih.gov/pcassay>
3. Steinbeck Christoph Han, Yongquan Kuhn S, Horlacher O, Luttmann E, Willighagen E. The Chemistry Development Kit (CDK): an open source Java library for Chemo and Bioinformatics. J. Chem. Inf. Comput. Sci. 2003.
4. Carrasco-Velar R, Trinchet-Almaguer D, Ortiz-Tornin S, Pérez-Durán R.E. CALEDE Lenguaje descriptor de la estructura química. QUITEL XXXIII.
5. Sung Hyuk, Cha. Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions, 2007.

Recibido: 22 de marzo de 2016.

Aprobado: 12 de mayo de 2016.