

# UN NUEVO VECTOR EN LA ESTRUCTURACION DEL INDICE TOPOLOGICO DE H. P. SCHULTZ

*E. Cornwell*

Facultad de Ciencias Químicas y Farmacéuticas, Universidad de Chile, Casilla 233, Santiago, Chile.

## RESUMEN

Un nuevo vector ( $V_c$ ) se propone para estructurar el índice de H. P. Schultz (2-MTI).

Este vector contiene elementos que corresponden a la densidad de carga electrónica soportada en el átomo de carbono de la molécula y da origen a un nuevo índice topológico denominado  $C_{or}$ .

El índice  $C_{or}$  presenta una mayor correlación con los logaritmos de los tiempos de retención relativos de 30 hidrocarburos saturados que los índices de H. P. Schultz, H. Wiener, A. T. Balaban:

## SUMMARY

A new vector ( $V_c$ ) is proposal for structure the topological index of H. P. Schultz (2-MTI).

This vector contain elements that correspond to the electron charge density soported on the carbon atoms of the molecule and originate the new topological index denominated  $C_{or}$ .

The  $C_{or}$  index present a bigger correlation with the logarithm of the relative retention time of 30 saturated hydrocarbons than the H. P. Schultz, H. Wiener and A.T. Balaban index.

**KEY WORDS:** Topological index, electron charge density, logarithm of relative retention time, saturated hydrocarbons, polinomial correlations.

## INTRODUCCION

La teoría de los grafos en química tiene como objetivo primordial encontrar formas de caracterizar la estructura química por un número. Toda estructura química se puede reducir a un objeto matemático por medio de transformaciones de su grafo químico a un grafo teórico invariante. Se entiende por grafo una estructura que representa una molécula y se define por las siguientes relaciones: Sea el grafo igual a  $G$ , si  $G = \{V, E\}$  donde  $V = \{v_i / i = 1, 2, \dots, n\}$ , este subconjunto representa los vértices (átomos) y  $E = \{e_j / j = 1, 2, \dots, n\}$ , este subconjunto representa las uniones (enlaces) de la molécula. Los invariantes, son números independiente de la asignación numérica que se le atribuyen a los vértices y uniones (átomos y enlaces en la estructura molecular)<sup>1,2)</sup>. Entre los diversos invariantes existentes, tenemos aquellos que son producto de operaciones matemáticas entre vectores, matrices de adyacencia y matrices de distancia<sup>3)</sup>. Estos números contienen importante información de la estructura de la molécula y conjuntamente con otros índices son usados en las disciplinas que se ocupan de las relaciones entre la actividad biológica y

la estructura química (QSAR), propiedad y estructura química (QSPR) y tiempos de retención y estructura química (QSRR)<sup>10-12</sup>. Entre los invariantes, los más utilizados son los llamados índices topológicos, los cuales tienen en su conceptualización consideraciones topológicas en la conformación de las estructuras matemáticas anteriormente citadas, como ejemplo de ellos se pueden indicar los índices topológicos de H. Wiener (W)<sup>13-14</sup>, A.T. Balaban (J)<sup>15</sup>, H. P. Schultz (2-MTI)<sup>3,6,8,9</sup> y otros<sup>5,7</sup>. En la actualidad muchos autores buscan distintas maneras de obtener estos invariantes, esto con el fin de obtener mejores descriptores (invariantes) en cuanto contengan la máxima información de una estructura molecular, el número de estos supera los 120 publicados, sin embargo muchos de ellos duplican información equivalente y otros se relacionan en forma trivial<sup>1</sup>. Un camino interesante para encontrar un nuevo índice topológico es explorar las manipulaciones algebraicas de las matrices teóricas de los grafos<sup>1</sup>. El índice 2-MTI es un ejemplo de ello<sup>3,6,8,9</sup>.

En este trabajo se reemplaza el vector  $V_{(1 \times n)}$  propio del índice topológico 2-MTI<sup>3</sup> por el vector  $V_c$  cuyos elementos propuestos por el autor como  $f_i$  corresponden a las densidades de carga electrónica asignada a cada átomo de carbono de la molécula que se describe.

He utilizado el vector  $V_c$  en vez del vector  $V$  para modificar el índice topológico 2-MTI, ya que el primero presenta mayor poder de discriminación de los distintos tipos de carbono presentes en un hidrocarburo saturado de acuerdo al grado de ramificación y largo de cadena de la molécula, así para un carbono del tipo -C H el elemento  $C_i$  del vector  $V$  es igual a 3 independiente del grado de ramificación y largo de cadena. Para el mismo tipo de carbono anterior, el elemento  $f_i$  del vector  $V_c$  presenta distintos valores de acuerdo a los parámetros moleculares indicados anteriormente. Esta cualidad indica que el índice  $C_{or}$  derivado de  $V_c$  será más susceptible a traducir diferencias estructurales finas que su predecesor 2-MTI. Este tipo de análisis que realizo, está de acuerdo con los señalados a situaciones semejantes aplicadas a matrices para generar índices que den cuenta de isómeros cis y trans en alquenos<sup>2</sup>.

Todos los índices topológicos señalados, tienen como elemento común la matriz cuadrada de distancia  $[D]_{(n \times n)}$ , por ello se utilizaron como testigos de comparación los índices que contienen en su definición este elemento.

## DESARROLLO Y CONCLUSIONES

Para probar la hipótesis planteada en la introducción, se correlacionaron para 30 hidrocarburos saturados de bajo peso molecular los logaritmos de los tiempos de retención relativos al n-heptano obtenidos mediante cromatografía de gases (GLC)<sup>16</sup> con los índices topológicos 2-MTI, W, J respectivamente, ver Tabla I. Los resultados de las regresiones se indican en la Tabla II. Estas regresiones se analizaron en base a los parámetros estadísticos siguientes: índice de correlación ( $r$ , o  $R^2$  si se trata de regresiones polinomiales) y el índice de Fisher ( $F$ )<sup>17</sup>. La búsqueda de los modelos de regresiones óptimas involucraron las siguientes expresiones matemáticas:

$$1. y = A + B \cdot X$$

$$2. y = A + B \cdot \ln X$$

$$3. Y = A \cdot B^{e^X}$$

$$4. Y = A \cdot X^B$$

$$5. Y = A + B_1 \cdot X + B_2 \cdot X^2$$

Se encontró como modelo óptimo de correlación la expresión matemática número 5. La búsqueda del modelo óptimo se basó en los parámetros estadísticos inherentes a cada ecuación. Además, se analizaron para las correlaciones que involucran los índices topológicos  $C_{or}$ , W, 2-MTI las correlaciones lineales entre los valores calculados de logaritmo de tiempos de retención relativos mediante las ecuaciones de regresión pertinentes y los valores experimentales correspondientes.

**Tabla I.** Valores de los logaritmos de los tiempos de retención relativos y los distintos invariantes

Substancias	log trr	Wiener	Balaban	2-MTI	Cor
n-propano	-2.056	4	1.6330	16	2.3280
2-metilpropano	-1.688	9	2.3238	36	4.3980
2-metilbutano	-1.115	18	2.5395	68	7.9310
n-butano	-1.498	10	1.9747	38	4.7030
2,2-dimetilpropano	-1.437	16	3.0237	64	6.9760
n-pentano	-0.9830	21	2.1906	74	8.7460
2,2-dimetilbutano	-0.8197	28	3.1685	106	11.5010
2,3-dimetilbutano	-0.658	29	2.9935	108	12.0440
2-metilpentano	-0.64	32	2.6272	118	13.1700
3-metilpentano	-0.572	31	2.7542	114	12.8460
n-hexano	-0.487	35	2.3391	128	14.3620
2,2-dimetilpentano	-0.369	46	3.1545	170	18.0630
2,4-dimetilpentano	-0.345	48	2.9532	176	18.9820
2,2,3-trimetilbutano	-0.311	42	3.5412	156	16.5800
3,3-dimetilpentano	-0.216	44	3.3604	162	17.3880
2-metilhexano	-0.164	52	2.6783	190	20.4020
2,3-dimetilpentano	-0.148	46	3.1442	168	18.1030
3-metilhexano	-0.12	50	2.8318	182	19.7860
3-etilpentano	-0.074	48	2.9923	174	19.1620
2,2,4-trimetilpentan	-0.063	66	3.3889	242	24.9050
n-heptano	0.0000	56	2.4475	204	21.9600
2,2-dimetilhexano	0.0881	71	3.1118	260	26.8600
2,2,3,3-tetrametilbutano	0.0973	58	4.0204	214	22.1320
2,5-dimetilhexano	0.1342	74	2.9278	270	28.1160
2,4-dimetilhexano	0.1486	71	3.0988	258	27.2590
2,2,3-trimetilpentan	0.1617	63	3.6233	230	24.5520
3,3-dimetilhexano	0.1965	67	3.3734	244	25.5950
2,3,4-trimetilpentan	0.2368	65	3.4642	236	24.9980
2,3,3-trimetilpentan	0.2629	62	3.7083	226	23.6690
2,3-dimetilhexano	0.2822	70	3.1708	254	26.9150

El índice topológico 2-MTI<sup>(3)</sup> para una molécula se obtiene de las ecuaciones a1, a2, a3.

$$Q = ( [A]_{(n \times n)} + [D]_{(n \times n)} ) \quad a1$$

$$V_{(1 \times n)} * Q = [ C1, C2, \dots, Cn ]_{(1 \times n)} \quad a2$$

$$2\text{-MTI} = \sum_{i=1}^n C_i \quad a3$$

Donde  $[A]_{(n \times n)}$  es la matriz cuadrada de adyacencia que le corresponde a una estructura de hidrocarburo particular.  $[D]_{(n \times n)}$  es la matriz de distancia correspondiente al mismo hidrocarburo. Los elementos de la matriz de adyacencia  $[A]_{(n \times n)}$  son:  $a_{ij} = 0$  para el caso que los átomos  $i, j$  no sean adyacentes, en caso contrario  $a_{ij} = 1$ , el valor de los elementos de la diagonal de la matriz cumplen con la relación  $a_{ii} = 0$ . Los elementos de la matriz de distancia  $[D]_{(n \times n)}$  son:  $a_{ii} = 0$ , los elementos  $a_{ij}$  serán numéricamente iguales a la suma de los enlaces existentes entre los átomos de carbono  $i$  y  $j$  por el camino mas corto, a cada enlace se le asigna un valor igual a 1 y el número de sumandos será igual al número de enlaces químicos entre los átomos de carbono  $i, j$ . La medida de distancia se realiza topológicamente a través de los enlaces, diferente es el caso en un tratamiento topográfico donde las distancias se miden a través del trazo lineal entre los átomos  $i$  y  $j$ .

**Tabla II.** Parámetros estadísticos de las correlaciones entre los logaritmos de los tiempos de retención relativos y los distintos índices topológicos en estudio. Modelo de correlación  $Y = A + B1*X + B2*X^2$ .

Indice	A	Error	B1	Error	B2	Error	R <sup>2</sup>	F
2-MTI	-2.2384	±0.0804	0.0174	±0.0012	-3.1389*10 <sup>-5</sup>	±3.8628*10 <sup>-6</sup>	0.97307	487.7161
C <sub>or</sub>	-2.3553	±0.0814	0.1702	±0.0112	-0.0028	±3.4223*10 <sup>-4</sup>	0.97735	582.5289
W	-2.1991	±0.0729	0.0626	±0.0039	-4.1225*10 <sup>-4</sup>	±4.7264*10 <sup>-5</sup>	0.97634	557.1021
J	-5.8919	±1.6708	3.0083	±1.1862	-0.3775	±0.2072	0.58450	18.9919

Variable dependiente es el logaritmo del tiempo de retención relativo, variable independiente, los diversos índices indicados en la primera columna.

Los elementos del vector V del índice 2-MTI para un hidrocarburo saturado corresponden tanto a los elementos  $a_{ii}$  de la matriz de adyacencia al cuadrado,  $[A]_{(n \times n)}^2$ , como al valor numérico asignado a cada carbono de un hidrocarburo,  $d_i^v$  calculado en base a la ecuación a4, concepto con el cual se define el índice de conectividad químico<sup>18</sup>.

$$d_i^v = (Z^v - h) / (Z - Z^v - 1) \quad a4$$

Donde  $Z^v$  implican los electrones de valencia, Z el número atómico, h los átomos de hidrógeno unidos al átomo de carbono bajo consideración. De este modo, los cuatro diferentes tipos de átomos de carbono encontrados en los hidrocarburos saturados se tipifican mediante los valores de un conjunto numérico compacto perteneciente a los naturales, este conjunto es { 4, 3, 2, 1 }, correspondientes a los carbonos ( -C-, -CH-, -CH<sub>2</sub>-, -CH<sub>3</sub> ) así un tipo de carbono se define por un solo número, independiente del grado de ramificación y número de carbonos de la estructura, en consecuencia el vector V como parte del invariante 2-MTI le confiere menor poder discriminativo que el vector propuesto  $V_c$  que a su vez es elemento constitutivo de índice topológico propuesto  $C_{or}$ .

El vector  $V_c$  lo definido con elementos de los números reales positivos, este hecho le confiere mayor poder discriminativo, así un tipo de carbono, como ejemplo CH tendrá numéricamente un valor distinto de acuerdo al grado de ramificación del hidrocarburo como al número de carbonos existentes en la molécula. El vector  $V_c$  se define por la ecuación a5.

$$V_c = [f_1, f_2, \dots, f_n]_{(1 \times n)} \quad a5$$

Cada elemento del vector definido  $f_n$  corresponde a la densidad de carga electrónica soportada en cada átomo de carbono de la molécula de hidrocarburo saturado, esta característica vista como elemento numérico, es levemente distinta para átomos de carbonos iguales pero insertos en estructuras con ramificaciones y tamaños moleculares distintas, por lo tanto el vector  $V_c$  presentará mayor poder de discriminación. El método de cálculo de las densidades de carga electrónica para estos fines se puede realizar por cualquier aproximación de la química cuántica ( semiempírica o *ab initio* ) y se recomienda para los hidrocarburos el método AM1 como lo indica E. Estrada<sup>2</sup>, se adjunta al método el proceso iterativo de Polak-Ribiere bajo la condición de convergencia de 0.1 Kcal/ (A° mol). Este proceso se llevó a cabo mediante el software Hyperchem<sup>19</sup>

La obtención del índice topológico  $C_{or}$  se realizó a través de las ecuaciones a1, a5, a6 y a7

$$V_{c(1 \times n)} * Q = [E_1, E_2, \dots, E_n]_{(1 \times n)} \quad a6$$

$$C_{or} = \sum_{i=1}^n E_i \quad a7$$

Donde el término  $C_{or}$  es un escalar correspondiente a un invariante descriptor de una molécula de hidrocarburo saturado específico.

Analizando los resultados de las ecuaciones presentadas en la Tabla II, se puede deducir que : el índice topológico propuesto  $C_{or}$  presenta un mayor índice de correlación e índice de Fisher que los índices de H. P Schultz, (2-MTI), H. Wiener, (W) y A. T. Balaban (J) en cuanto a las correlaciones con los logaritmos de los tiempos de retención relativos de 30 hidrocarburos saturados livianos, siendo este último índice (J) el menos eficiente como descriptor. Cabe hacer notar, que la ecuación de correlación para el índice  $C_{or}$  en los términos A, B1, B2 presenta un mayor error que las ecuaciones de correlación de los índices 2-MTI, W en esos mismos términos. Sin embargo, un estudio de las correlaciones lineales que responden al modelo  $y = m \cdot x + n$  realizadas para estos índices topológicos donde se correlacionan los logaritmos de los tiempos de retención relativos experimentales<sup>16)</sup> versus los logaritmos de los tiempos de retención relativos calculados a través de las ecuaciones de correlación propuestas para cada índice, indican que para el caso del índice propuesto  $C_{or}$  los valores de r y F son mayores como se indica en la Tabla III. Esto lleva implícito que, para el caso del uso del índice propuesto  $C_{or}$  las diferencias entre los valores experimentales y calculados serán menores que para el caso del uso del de los demás índices bajo estudio . Esto justifica el uso del índice topológico  $C_{or}$  como descriptor como así mismo sugiere un estudio de la modificación de la matriz de distancia en cuanto al uso de distancias atómicas físicas en vez de topológicas para estudiar luego la integración de ambas modificaciones en un solo índice.

Tabla III. Parámetros estadísticos de las correlaciones.

Índice	r	F
$C_{or}$	0.9888	1205.88
2-MTI	0.9864	1011.58
W	0.9760	1155.42

## REFERENCIAS

1. E. Estrada, L. Rodríguez, A. Gutiérrez. *Natch*, **35**, 145 (1997).
2. E. Estrada. *J. Chem. Inf. Comput. Sci.*, **35**, 708 (1995).
3. H. P. Schultz. *J. Chem. Inf. Comput. Sci.*, **29**, 227 (1989).
4. E. Estrada, L. Rodríguez. *Natch*, **35**, 157 (1997).
5. Y. Yang, L. Xu, C. Hu. *J. Chem. Inf. Comput. Sci.*, **34**, 1140 (1994).
6. H. P. Schultz, E. B. Schultz, T. P. Schultz. *J. Chem. Inf. Comput. Sci.*, **30**, 27 (1990).
7. Y. Yao, L. Xu, Y. Yang, X. Yuan. *J. Chem. Inf. Comput. Sci.*, **33**, 590 (1993).
8. H. P. Schultz, T. P. Schultz. *J. Chem. Inf. Comput. Sci.*, **33**, 240 (1993).
9. H. P. Schultz, E. B. Schultz, T. P. Schultz. *J. Chem. Inf. Comput. Sci.*, **33**, 863 (1993).
10. E. J. Ariëns. *Drug Desing*, Vol III. Editorial Academic Press, New York y Londres (1972).
11. R. Kaliszan. *Anal Chem.*, **64**, 619 A (1992).
12. M. Randic, A. F. Kleiner, L. M. DeAlba. *J. Chem. Inf. Comput. Sci.*, **34**, 277 (1994).
13. Y. Lukovits. *J. Chem. Inf. Comput. Sci.*, **31**, 503 (1991).
14. B. Mohar. *J. Chem. Inf. Comput. Sci.*, **33**, 153 (1993).
15. Z. Mihalic, N. Trinajstic. *J. Chem. Inf. Comput. Sci.*, **9**, 701 (1992).
16. P. H. Desty, S. Goldup, E. Y. Swanton. *Gas Chromatography*. Editorial Academic Press. New York (1962).
17. Microcal Origin 4.00 Software Inc. One Roundhouse Plaza. Northampton. Ma.01060. USA. (1995).
18. L. B. Kier, L. H. Hall. *J. Pharm. Sci.*, **72**, 1170 (1983).
19. Hyperchem, Hypercube and Autodesk, Inc. Developed by Hypercube Inc. (1993).