

# OPTIMIZACIÓN DEL EXPONENTE IMPLÍCITO EN EL ALGORITMO CORRESPONDIENTE AL ÍNDICE DE CONECTIVIDAD DE VALENCIA TRIDIMENSIONAL CORREGIDO $\Lambda(q)$ . APLICACIÓN A HIDROCARBUROS SATURADOS

*E. Cornwell*

Departamento de Química Inorgánica y Analítica, Facultad de Ciencias Químicas y Farmacéuticas. Universidad de Chile. Casilla 233, Santiago, Chile.

(Recibido: Junio 18, 2001 - Aceptado: Diciembre 26, 2001)

## RESUMEN

En este trabajo se demuestra que el uso del algoritmo correspondiente al índice de conectividad de valencia tridimensional corregido  $[\Lambda(q)]$ , utilizado a un valor de exponente óptimo, produce un valor máximo para el índice de correlación ( $r$ ) correspondiente a la regresión lineal entre los logaritmos de los tiempos de retención relativos de un conjunto de 30 hidrocarburos saturados tomados como modelo. Este hecho particular induce a considerar la optimización del exponente de este índice topológico u otro índice de estructura similar en cualquier otra relación de regresión donde estos índices estén implícitos.

**PALABRAS CLAVES:** Índice de conectividad de valencia tridimensional, exponente, optimización.

## SUMMARY

In this work one demonstrates that the use of the algorithm corresponding to the corrected three-dimensional valence connectivity index  $[\Lambda(q)]$ , used at a value of optimal exponent, produces a maximum value for the correlation index ( $r$ ) corresponding to the linear regression between the logarithm of the relative retention times of a set of 30 hydrocarbons taken as a model. This particular fact, lead us to consider optimization of the exponent of this topological index or another index of similar structure in any other relation of regression where these indices are implicit.

**KEY WORDS:** Three-dimensional valence connectivity index, exponent, optimization.

## INTRODUCCIÓN

Desde 1817 existe un concepto fundamental en química como también en otras disciplinas científicas, este concepto se refiere al siguiente postulado: la forma precede a la función<sup>1)</sup> Desde el punto de vista químico esto significa una dependencia entre la estructura molecular y su distribución electrónica versus las propiedades moleculares<sup>2)</sup>. Las disciplinas que se preocupan de esta dependencia se denominan: relación cuantitativa entre estructura y propiedades químicas (QSPR), relación cuantitativa entre actividad biológica y estructura química (QSAR) y relación cuantitativa entre retención y estructura química (QSRR)<sup>3)</sup>, la retención se refiere a retención cromatográfica. La estructura molecular y la distribución electrónica se pueden caracterizar mediante diferentes descriptores, para el caso de la disciplina QSRR<sup>3)</sup> que es la disciplina que trata este trabajo, cabe mencionar los descriptores no específicos como: masa molecular, índice de refracción, volumen molecular, energía total, acceso de superficies por solventes. Aquellos de carácter netamente electrónicos como: momento dipolar, exceso de carga atómica, energía orbital, superdelocalización, carga superficial parcial. Aquellos relacionados con la geometría molecular como: momento de inercia, energía de ángulos tensionados, relación de largo a ancho molecular. Parámetros fisicoquímicos como: constantes hidrofóbicas, constantes de Hammett, parámetros de solubilidad, constantes de equilibrio, puntos de ebullición. Parámetros topológicos caracterizados por los llamados índices topológicos.

Entre las propiedades moleculares se pueden citar entre otras: propiedades biológicas, farmacológicas, fisicoquímicas y de retención cromatográficas<sup>2,3)</sup>

Los índices topológicos son un elemento de la disciplina denominada teoría de los grafos químicos que se basa en la representación de una molécula como un grafo simple donde los átomos se simbolizan por puntos (vértices) y los enlaces covalentes por líneas (puentes). El principal objetivo de esta aproximación es encontrar una manera de representar la estructura molecular como un número simple y reducir el problema de la relación estructura-propiedad a un problema de correlación matemática entre dos series de números, un de ellos representando la propiedad y la otra serie, representando la estructura química<sup>4)</sup>

La teoría de los grafos es una rama de las matemáticas discretas relacionada con la topología y la combinatoria y trata de las maneras como se conectan los objetos y la consecuencia de estas conexiones. La teoría de los grafos químicos implica el uso como objeto de análisis la estructura molecular. Se entiende como grafo a un conjunto de objetos aplicados a sí mismos y la relación binaria entre estos elementos, los grafos son objetos unidimensionales pero se pueden proyectar a espacios de dimensiones superiores. En química, los grafos se caracterizan en dos dimensiones, puntos (vértices) y líneas (puentes) que unen estos puntos, en química se suele designar esta representación como diagrama. En química los grafos pueden ser usados para representar una variedad de objetos químicos tales como: moléculas, reacciones químicas, cristales, etc. El rasgo común en los sistemas químicos es la presencia de puntos y conexiones entre ellos, los puntos pueden ser: átomos, electrones, moléculas, fragmentos moleculares, intermediarios etc. Las conexiones pueden representar enlaces, etapas de reacción, fuerzas de van der Waals. En forma particular los grafos moleculares son grafos químicos que representan estructuras moleculares donde los vértices son los átomos individuales y los puentes corresponden a los enlaces entre ellos<sup>5)</sup>

La resolución de un grafo molecular a un número mediante un proceso topológico debe ser inequívoca con la estructura molecular que se refiere, si este es el caso, se habla de un invariante, de lo contrario se hace referencia de un grado degenerativo. Según sea el algoritmo topológico se obtienen índices del índice topológico distintos<sup>6)</sup>

Uno de los primeros índices topológicos fue el de M. Randić<sup>7)</sup> el cual utiliza la expresión (1) como algoritmo de cálculo de una molécula.

$$\chi = \sum_k [ (d_i d_j)^{-1/2} ]_k \quad (1)$$

El elemento  $d_i$  (como  $d_j$ ) es un número asignado al átomos  $i$  del grafo molecular y es igual a la suma de los elementos  $a_{ij}$ ,  $a_{ji}$  de una fila o columna de la matriz cuadrada de adyacencia correspondiente al grafo de la estructura molecular que se caracteriza<sup>7)</sup>.

Este índice topológico  $\chi$  ha sido modelo de otros índices posteriores que utilizan en forma básica el mismo algoritmo descrito por la ecuación (1) y difieren entre sí, en cuanto a la definición del elemento  $d_i$  siendo el exponente constante y de un valor igual a  $-1/2$ . El desarrollo de las distintas modificaciones, aplicaciones, ventajas y limitaciones de otros índices topológicos derivados de  $\chi$  se analizan en la literatura<sup>9-31)</sup>. Como resultado de ello se han obtenido índices más específicos y extensivos a todo tipo de molécula orgánica.

En este trabajo se analiza, como varía el grado de correlación entre una variable dependiente (log del tiempo de retención relativo cromatográfico de un conjunto de hidrocarburos saturados) versus la variable independiente (índice topológico de E. Estrada) al utilizar distintos valores del exponente en el índice de conectividad de valencia tridimensional corregido, cuyo autor es E. Estrada<sup>8)</sup>  $[\Lambda(q)]$ . Este índice se basa en la expresión de la ecuación (1), pero se define en forma distinta los términos  $d_i$  correspondientes a cada átomo del grafo molecular, manteniendo constante el exponente en  $-1/2$ .

En la literatura se menciona el uso de exponentes distintos a  $-1/2$  en los algoritmos relacionados al índice topológico de M. Randić, como lo es  $-1/3$  entre otros valores, esto sin explicación razonable de su aplicación<sup>21,33)</sup>. El propósito de un estudio sistemático para obtener un exponente dado con el cual se obtenga una correlación máxima debida a este efecto entre la variable dependiente e independiente, no solo es inobjetable, si no que, el uso de un exponente así optimizado puede producir un cambio en la dirección de una mayor correlación que la obtenida por una redefinición de los elementos  $d_i$  y  $d_j$  del algoritmo en estudio.

Cabe hacer notar que en las definiciones actuales de los elementos  $d_i$  y  $d_j$  concurren métodos complejos como se verá a continuación.

El descriptor tridimensional de E. Estrada<sup>8)</sup> se define por la siguiente expresión: (2)

$$\Lambda(q) = \sum_k [d^c(q_i) d^c(q_j)]_k^{-1/2} \quad (2)$$

El valor de  $q$  para un átomo  $i$  es  $q_i$  y corresponde a la densidad de carga electrónica existente en el átomo  $i$  y se usa como valor asignado al átomo de carbono para los grafos moleculares en este trabajo. Su magnitud se evalúa mediante la química cuántica usando la expresión (3)

$$q_i = Z_i - \sum \mu^i (P,S) \mu \quad (3)$$

Donde  $Z_i$  es la carga nuclear,  $P$  es la matriz de densidad,  $S$  es la matriz de superposición y el término de la sumatoria es la población de Mulliken. (el número de electrones en cada orbital atómico  $\mu$ ). Los conceptos y características de estas consideraciones se indican en la literatura<sup>8,34)</sup> y se dan por explicadas en el trabajo tomado como base<sup>8)</sup>

El cálculo de la densidad de carga electrónica de conectividad  $d(q_i)$  en este caso correspondiente a un átomo de carbono  $i$ , se calcula restando al valor de  $q_i$  al número de hidrógenos unidos a aquel átomo de carbono  $i$  ( $h_i$ ). Este cálculo se efectúa mediante la ecuación (4)

$$D_i(q) = q_i - h_i \quad (4)$$

Como en un grafo molecular existen diferentes átomos de hidrógeno, la densidad de carga de conectividad de carga corregida se evalúa mediante la ecuación (5)

$$d^c(q_i) = q_i - \sum_j q_{hj} \quad (5)$$

Donde  $q_{hj}$  es la densidad de carga electrónica del átomo  $j$  unida al átomo  $i$

La densidad de carga electrónica se obtiene por el método PM3<sup>35)</sup> La optimización geométrica mediante el método Broyden-Goldfarb-Shanon<sup>36)</sup> utilizando el software MOPAC versión 6.0<sup>37)</sup> y el cómputo del índice se realizó por el sistema MODEST versión 1.0<sup>38)</sup>

Los resultados de los valores de carga electrónica correspondientes a distintos tipos de átomos unidos a hidrógeno u otros tipos de átomos fueron publicados por el autor del índice topológico en estudio<sup>8)</sup> Para el caso del presente trabajo donde solo se consideran 4 distintos tipos de átomos de carbono en cuanto a la cantidad de átomos de hidrógeno unidos a él, los valores calculados<sup>8)</sup> correspondientes se detallan en la Tabla I.

**Tabla I** Valores de los parámetros  $d_i$  para diferentes grupos atómicos

grupo	$d^c(q)$
-CH <sub>3</sub>	1.2208
-CH <sub>2</sub> -	2.2015
-CH<	3.1349
>C<	4.0413

## DESARROLLO

Se demuestra que, para un conjunto de sustancias tomadas como modelo, la correlación entre una propiedad física que las caracteriza y el índice topológico  $\Lambda(q)$  corregido, ocurre con un índice de correlación de valor máximo si el exponente del índice topológico  $\Lambda_o(q)$  concurre con un valor óptimo. Esta consideración no se encuentra en los trabajos de esta índole comúnmente publicados<sup>9,31</sup>.

Para la demostración se utilizó como modelo 30 hidrocarburos saturados. Se utilizó como variable dependiente los valores del logaritmo de los tiempos de retención relativos ( $\log T_{rr}$ ) al n-heptano medidos en columna apolar mediante un proceso cromatográfico en fase gaseosa<sup>32</sup>, ver Tabla II y se utilizó como variable independiente el índice  $\Lambda_o(q)$  corregido<sup>9</sup> modificado a un valor óptimo del exponente. El proceso para obtener este exponente óptimo consistió en tomar como referente el valor implícito en el índice  $\Lambda(q)$ , y sobre la base de ello, se eligieron 27 valores de exponentes distintos al original, ver Tabla III. Se utilizó la ecuación (2) conteniendo estos exponentes mencionados y manteniendo los otros parámetros en condición de su definición original, ver Tabla I, consecuencia de ello se obtuvo 27 ecuaciones distintas para  $\Lambda(q)$ , con lo cual se calcularon 27 índices topológicos distintos para cada grafo molecular perteneciente a cada hidrocarburo saturado.

**Tabla II.** Parámetros de retención para los distintos hidrocarburos saturados

Substancias	Trr	logTrr
n-propano	0.0088	-2.0555
2-metilpropano	0.0205	-1.6883
2-metilbutano	0.0768	-1.1146
n-butano	0.0318	-1.4976
2,2-dimetilpropano	0.0366	-1.4365
n-pentano	0.1040	-0.9830
2,2-dimetilbutano	0.1550	-0.8097
2,3-dimetilbutano	0.2200	-0.6576
2-metilpentano	0.2290	-0.6402
3-metilpentano	0.2680	-0.5719
n-hexano	0.3260	-0.4868
2,2-dimetilpentano	0.4280	-0.3686
2,4-dimetilpentano	0.4520	-0.3449
2,2,3-trimetilbutano	0.4890	-0.3107
3,3-dimetilpentano	0.6080	-0.2161
2-metilhexano	0.6850	-0.1643
2,3-dimetilpentano	0.7120	-0.1475
3-metilhexano	0.7580	-0.1203
3-etilpentano	0.8430	-0.0742
2,2,4-trimetilpentano	0.8660	-0.0625
n-heptano	1.0000	0.0000
2,2-dimetilhexano	1.2250	0.0881
2,2,3,3-tetrametilbutano	1.2510	0.0973
2,5-dimetilhexano	1.3620	0.1342
2,4-dimetilhexano	1.4080	0.1486
2,2,3-trimetilpentano	1.4510	0.1617
3,3-dimetilhexano	1.5720	0.1965
2,3,4-trimetilpentano	1.7250	0.2368
2,3,3-trimetilpentano	1.8320	0.2629
2,3-dimetilhexano	1.9150	0.2822

**Tabla III.** Distintos valores de los exponentes (-n) utilizados y los valores de los índices de regresión obtenidos.

Exponente n*(-1)	Índice de correlación r
0.66670	0.98767
0.52632	0.99233
0.50505	0.99263
0.50025	0.99269
0.50000	0.99269
0.47371	0.99290
0.47000	0.99292
0.46948	0.99292
0.46729	0.99293
0.46083	0.99295
0.45455	0.99297
0.44843	0.99297
0.44643	0.99298
0.44563	0.99298
0.44543	0.99298
0.44524	0.99298
0.44444	0.99298
0.44248	0.99297
0.41000	0.99286
0.34483	0.99210
0.33333	0.99190
0.25000	0.98989
0.20000	0.98827
0.16667	0.98705
0.14286	0.98610
0.12500	0.98536
0.11111	0.98477

Se correlacionaron linealmente los logaritmos de los tiempos de retención relativos de los 30 hidrocarburos saturados versus los distintos tipos de índices topológicos obtenidos  $\Lambda_n(q)$ , de ello se obtiene 27 ecuaciones lineales con sus respectivos índices de correlación (r)

La correlación lineal entre el conjunto de valores de r y los exponentes (n),  $rf(n)$ , responde a una correlación lineal expresada por el modelo (5). Con este modelo se obtiene un índice de determinación ( $R^2$ ) alto.

$$y = A + BX + CX^2 + DX^3 \quad (5)$$

La ecuación particular para  $rf(n)$  para los valores de la Tabla II corresponde a la ecuación (6)

$$r = 0.9797 - 0.0478n - 0.0138n^2 + 0.0600n^3 \quad (6)$$

$$R^2 = 0.9994$$

$$SD = 9.25 \cdot 10^{-6}$$

$$N = 27$$

La ecuación particular para  $rf(n)$  presenta una alta correlación ( $R^2$ ) y baja desviación estándar (SD), lo que implica un modelo muy eficiente.

Esta función presenta un par ordenado crítico (n,r) máximo, cuyo valor numérico se obtiene derivando la ecuación (6) y haciendo esta derivada igual a cero y obteniendo el valor crítico n y r por los métodos corrientes matemáticos. La derivada corresponde a la siguiente expresión (7).

$$-0.0478 - 0.0276n + 0.1801n^2 = 0 \quad (7)$$

Siendo el primer miembro, la derivada (dr/dn) de la ecuación (6)

El exponente óptimo es igual a -0.44434 y r máximo se obtuvo utilizando este exponente en la relación de correlación  $\log Trr f(\Lambda_o(q))$ , para dicha relación el coeficiente de correlación r tiene un valor de 0.9930

La ecuación de regresión que contiene el par ordenado crítico que genera la regresión óptima se representa por la ecuación (8)

$$\begin{aligned} \log Trr &= -3.3062 + 1.0017\Lambda_o(q) & (8) \\ r &= 0.9930 \\ SD &= 0.0741 \\ n &= 30 \end{aligned}$$

Si se utiliza como exponente el usado en forma común en la literatura<sup>9-31)</sup> (-0.5000) para este algoritmo de este modelo, se obtiene la ecuación de regresión (9)

$$\begin{aligned} \log Trr &= -3.3527 + 1.104 \Lambda(q) & (9) \\ r &= 0.9920 \\ SD &= 0.0880 \\ N &= 30 \end{aligned}$$

En ambas situaciones (ecuaciones (8),(9)) los índices de correlación y desviación estándar son semejantes, esto obliga a un análisis de evaluación de las diferencias, para ello se usó el siguiente criterio:

Para ambas situaciones; ecuaciones (8), (9) donde se emplean los exponentes asociados al índice  $\Lambda_o(q)$  y  $\Lambda(q)$  cuyos valores son: -0.44434 y -0.5000, respectivamente, se obtuvieron las diferencias absolutas entre los tiempos de retención relativos experimentales y calculados mediante las ecuaciones (8),(9) respectivamente. Para el primer caso [exponente óptimo(-0.44434)] la mediana de las diferencias a la mitad de las puntuaciones tiene un valor de 0.0031, para el segundo caso, en las mismas condiciones de puntuación un valor de 0.0041, según el trabajo de Otto Exner<sup>39)</sup> el modelo más eficiente es aquel que presenta una menor mediana de las diferencias a la mitad de las puntuaciones. En este caso corresponde al índice corregido  $\Lambda_o(q)$  con exponente óptimo igual a -0.44434, obtenido por el procedimiento señalado en este trabajo. Se recomienda en índices topológicos que respondan a la misma estructura del propuesto por M. Randic y sus derivados utilizar esta optimización del exponente toda vez que se utilicen distintos tipos de series homólogas orgánicas, puesto que el exponente óptimo tiene valores distintos en función de que tipo de serie homólogas se trate<sup>21, 40)</sup>

Nota: f: significa, función de.  $\Lambda(q)$  significa, índice de E. Estrada,  $\Lambda_o(q)$  significa índice de E. Estrada con exponente óptimo calculado en este trabajo y  $\Lambda_n(q)$  significa índice de E. Estrada con exponente distinto a -1/2 pero distinto al óptimo.

## CONCLUSIONES

- 1.- La función de optimización estudiada es continua y tiene un solo punto crítico correspondiente al máximo de la función.
- 2.- A través del procedimiento de optimización del exponente implícito en el índice topológico  $\Lambda(q)$ , se logró una mayor correlación entre las variables en estudio, en comparación con la situación donde se utiliza como exponente el valor  $-1/2$ , esto sin alterar el significado físico de los otros elementos involucrados en dicho índice.
- 3.- El procedimiento de optimización del exponente es significativamente más simple que una redefinición de los otros parámetros implícitos en  $\Lambda(q)$ , y probablemente se obtenga mediante esta modificación un grado de correlación más significativo que aquella situación donde se mantiene el exponente en un valor igual a  $-1/2$  y se redefinen los demás parámetros implícitos en  $\Lambda(q)$ .

## BIBLIOGRAFIA

1. P. Seybolt, M. M. Bagal. **66**, 575 (1987).
2. E. S. Ariens, Drug Desing Vol 1 p. 271, Vol 2 p.133 Academic Press. New York (1971).
3. R. Kalizan. Anal.Chem.**64**, 619 A (1992).
4. E. Estrada. J. Chem. Inf. Comput. Sci. **35**, 701 (1995).
5. Z. Mihalic, N. Trinystic. J. Chem. Educ. **69**, 701 (1992).
6. E. Cornwell. Bol. Soc. Chil. Quím. **45**, 649 (2000).
7. M. Randic., J.A.C.S. **97**, 6609 (1975).
8. E. Estrada. J. Chem. Inf. Comput. Sci. **35**, 708 (1995).
9. G.L. Amidon., S.T. Anic. J. Pharm. Sci. **65**, 801 (1976).
10. L.B. Kier., W.J. Muray., M. Randic., L. Hall. J.Pharm. Sci. **65**, 1226 (1976).
11. L.B. Kier., L.H. Hall. J. Med. Chem. **20**, 1631 (1977).
12. L.H. Hall., L.B. Kier. J. Pharm. Sci. **67**, 1743 (1978).
13. L.B. Kier., L.H. Hall. J. Pharm. Sci. **68**, 120 (1979).
14. S.C. Basak., D.P. Gieschen, O.K. Harris., V.R. Maguson J. Pharm. Sci. **72**, 934 (1983).
15. P.J. Hansen., P.C. Jurs., J. Chem. Educ. **65**, 574 (1988).
16. A.R. Katritzky., E.V. Gordeeva, J. Chem. Inf. Comput. Sci. **33**, 835 (1993).
17. A. Sabljic., D. Horvatic. J. Chem. Inf. Comput. Sci. **33**, 292 (1993).
18. J. Galvez., R. García-Domenech, V. De Julian-Ortiz.,R. Saler. J. Chem. Inf. Comput. Sci. **34**,1198 (1994).
19. Pogliani. J. Phys. Chem. **98**, 1494 (1994).
20. A.T. Balaban, J. Chem. Inf. Comput. Sci. **35**, 339 (1995).
21. E. Estrada. J. Chem. Inf. Comput. Sci. **35**, 1022 (1995).
22. E. Estrada. J. Chem. Inf. Comput. Sci. **35**, 31 (1995).
23. D. Amic., D. Davidovic-Amic., A. Juric, B. Lucic., N. Trinajstic. J. Chem. Inf. Comput. Sci. **35**, 1034 (1995).
24. E. Estrada. J. Chem. Inf. Comput. Sci. **36**, 837 (1996).
25. V. Sharma., R. Goswami., A.K. Madan. J. Chem. Inf. Comput. Sci. **37**, 273 (1997).
26. E. Estrada., N. Guevara, J. Gutman. J. Chem. Inf. Comput. Sci. **38**, 428 (1998).
27. E. Estrada. L. Rodríguez. J. Chem. Comput. Sci. **39**, 1037. (1999).
28. E. Estrada. J. Chem. Inf. Comput. Sci. **39**, 1042 (1999).
29. L.B. Kier, L.H. Hall. J. Chem. Inf. Comput. Sci. **40**, 792 (2000).
30. D. Bonchev. J. Chem. Inf. Comput. Sci. **40**, 934 (2000).
31. J. Huuskonem. J. Chem. Inf. Comput. Sci. **40**, 773 (2000).
32. P.H. Desty., S. Goldup., E.Y. Swanton. Gas Chromatography. Editorial Academic Press. New York (1962).
33. Alterburg. Phys. Chem (Elipsis) **261**, 389 (1980).
34. Sublej,J. Semiempirical methods of Quantum Chemistry. Ellis Hardwood Ltd: West Sussex (1985) p.152.

35. J.J.P Stewart..Optimization of Parameters for Semi-Empirical Methods j. Comput. Chem. **10**, 209 (1989).
36. D. F. Shanon. J. Optim.. Ther. Appl. **46**, 87 (1985).
37. J. P.Stewart quantum Chemistry program Exchange N° 581.
38. L. Rodríguez, E, Estrada, Y Gutiérrez, Y Muñoz. MODEST (Molecular DESIng Tools) versión 1.0 Central University of Las Villas, Santa Clara, Cuba.
39. O. Exner., J. Kramosil., I. Vajoa. J. Chem. Inf. Comput. Sci. **33**, 407 (1993).
40. S.C. Basak., S. Nikoliv., N. Trinajstic., D. Amic., D. Beslo. J. Chem. Inf. Comput. Sci. **40**, 927 (2000).

