

## Investigación

# Nueva proposición para predecir datos de retención obtenidos mediante cromatografía de gases de hidrocarburos derivados de las naftas

Edward Cornwell<sup>1</sup> and Gianni Cordano<sup>2</sup><sup>1</sup> Department of Inorganic and Analytical Chemistry.<sup>2</sup> Department of Organic Chemistry and Physical Chemistry. Faculty of Chemical and Pharmaceutical Sciences, University of Chile, Olivos 1007, Independencia, Santiago, Chile.

Recibido el 6 de diciembre del 2002; aceptado el 10 de febrero del 2003

**Resumen.** Para 64 hidrocarburos provenientes de las naftas pirolizadas o transformadas catalíticamente, se propone una ecuación de regresión que relaciona el índice de Kovats obtenido en columnas cromatográficas apolares ( $I_{\text{apolar}}$ ), un índice propuesto en este trabajo ( $C_g$ ) y el índice de conectividad químico de orden 1 tipo valencia ( $\chi_v^1$ ). El perfil estadístico de la regresión propuesta es semejante a un modelo tomado como referencia, pero presenta solo dos variables independientes de fácil obtención y un índice de Fisher superior, lo que permite una orientación en la identificación de los hidrocarburos aquí tratados a través de la ecuación propuesta con un error máximo de un 4.1 % entre los índices de Kovats experimentales y calculados.

**Palabras clave:** Un nuevo índice topológico, derivados de naftas, índice de retención.

**Abstract.** A regression equation for 64 naphtha hydrocarbons is proposed to relate a new index ( $C_g$ ), the Kovats index obtained in apolar chromatographic columns ( $I_{\text{apolar}}$ ) and the chemical connectivity index order 1 valence type ( $\chi_v^1$ ).

The statistical profile of the proposed regression is similar to the reference model but with two easy-to-obtain, independent variables and a higher Fisher index. This permits a good approach to identify the hydrocarbons under study by means of the suggested equation with a maximum error of 4.1 % between the experimental and calculated Kovats indices.

**Key words:** A new topological index, derivated of naftas, retention index, regression.

## Introducción

La relación cuantitativa entre retención cromatográfica y estructura química (QSRR), sus alcances, características y propiedades están descritas ampliamente en la literatura [1-3].

A través de esta disciplina, se proponen regresiones entre los descriptores moleculares [4] y ciertas variables dependientes que, en el caso específico de este trabajo, corresponden a parámetros de retención usados en cromatografía gaseosa denominados índices de Kovats [5, 6], con columnas apolares ( $I_{\text{apolar}}$ ) y polares ( $I_{\text{polar}}$ ) cuyas fases estacionarias son SE-30 y Carbowax-20M, respectivamente.

El presente estudio toma como referencia el trabajo de T. F. Woloszyn [6], en el que, de un total de 81 hidrocarburos provenientes de la pirólisis y ruptura catalítica de las naftas, grupo importante de hidrocarburos en la industria petroquímica [6], 67 se utilizaron para el estudio de regresión en columna SE-30 y 65 para el estudio de regresión en columna Carbowax-20M. Los 14 y 16 hidrocarburos restantes, respectivamente, no correlacionaron en ningún sistema de regresión propuesto [6].

Para el presente trabajo se seleccionaron 64 de los 81 hidrocarburos del trabajo de referencia para tratarlos con el mismo objetivo empleado en el trabajo de referencia [6]. Para el caso que se presenta, 17 de las 81 sustancias no correlacionaron en ningún modelo propuesto.

El propósito de este trabajo es encontrar una relación de regresión válida entre los parámetros de retención cromatográficos [ $I_{\text{apolar}}$ ,  $I_{\text{polar}}$ ,  $\Delta(I_{\text{apolar}} - I_{\text{polar}})$ ], muy utilizados en cro-

matografía gaseosa con propósitos de identificación, y ciertas variables independientes obtenidas a través de cálculos de fácil obtención sobre la base de las estructuras moleculares, entendiéndose por regresión válida, a aquella cuyo perfil estadístico sea aceptable de acuerdo con lo ya publicado [6]. Se exige además, que el número de grados degenerativos de las variables independientes sea mínimo, lo que implica que estas variables deben presentar un valor inequívoco con respecto a una cierta estructura molecular definida (invariantes) [1,2,7]. Esta característica no se presenta cuando se utilizan como variables independientes propiedades físicas como punto de ebullición, índices de refracción molar o constantes de equilibrio octanol-agua. En algunos estos casos, para distintas estructuras moleculares, se tienen valores iguales de las variables independientes (grados degenerativos) [1,2]. La ecuación de regresión propuesta debe tener un mínimo de variables independientes [8], para representar una clara interpretación del proceso de retención y debe tener un perfil estadístico tal que contenga un valor significativo (alto) en la constante de Fisher, lo que implica un alto poder de predicción de la ecuación de regresión. En consecuencia, el valor de coeficiente de correlación ( $r$ ) o de determinación ( $R^2$ ) debe tender al valor 1 acompañados de bajos valores de desviación estándar (d.e) [9]. Por otra parte, la ecuación de regresión debe presentar una validación externa significativa [6].

En estas condiciones, se obtendrá un instrumento particularmente útil en el análisis cualitativo de orientación diseñado para el grupo de sustancias en estudio.

Como variable independiente se utilizó el índice topológico de conectividad químico orden 1 de valencia ( $\chi_v^1$ ) [10-14], el que contiene información molecular referente al tamaño de los grupos químicos unidos a la estructura principal y tiene relación directa con el grado de ramificación, molecular [10] y además se le considera como uno de los más aceptados entre los índices topológicos [15].

Como segunda variable independiente se utilizó el mismo algoritmo involucrado en la definición de  $\chi_v^1$ , pero los valores de  $d_i$  y  $d_j$  se definen en forma distinta. Sobre la base de los resultados promisorios obtenidos en trabajos anteriores [1, 2], en este trabajo se proponen valores que corresponden a las densidades de carga sustentadas en el átomo de carbono del hidrocarburo. En este trabajo esta variable (índice topológico) se define como  $C_g$ .

La variable independiente  $\chi_v^1$  presenta casos degenerativos y esto se debe a que, en la proposición de este índice topológico, tanto los valores de  $d_i$  como los de  $d_j$  contemplan números del dominio de los naturales, asignando valores fijos a los distintos tipos de carbono independientemente de su entorno molecular. Así los carbonos [-CH<sub>3</sub>, -CH<sub>2</sub>-, >CH-, >C<] en todas las moléculas presentan los valores (1, 2, 3, 4) respectivamente [12, 13] Para el caso de los valores de  $d_i$  y  $d_j$  definidos por números provenientes de las densidades de carga en el carbono que dependen del tipo de los otros carbonos e hidrógenos circundantes en la molécula, se obtienen números del dominio de los reales con baja probabilidad de degeneración, alto poder discriminativo de carbonos con el mismo número de hidrógenos, pero distintos entornos moleculares, así, definidos los carbonos en el algoritmo utilizado para calcular  $C_g$ , pasan a llamarse  $C_i$  y  $C_j$  correspondientes a las densidades de carga soportadas por los carbonos  $i, j$  respectivamente.

En este trabajo se estudian las variables [ $I_{apolar}$ ,  $I_{polar}$  y  $\Delta I$  ( $I_{apolar} - I_{polar}$ )] como variables dependientes, donde la última expresión se ha definido como una relación directamente proporcional a la estructura molecular [5]. Cabe hacer notar que estos parámetros, por sí solos, son herramientas eficientes para el análisis cualitativo. Estas tres variables dependientes y las dos variables independientes consideradas anteriormente, abren la posibilidad de analizar 9 sistemas de regresión y de evaluarlas bajo los criterios de  $r$ ,  $R^2$ , d.e y  $F$ . El análisis realizado condujo a la proposición de una expresión de regresión óptima donde participa como variable dependiente  $I_{apolar}$  y como variables independientes  $C_g$  y  $\chi_v^1$  lográndose una diferencia ( $I_{apolar\ experimental} - I_{apolar\ calculado}$ ) con errores que fluctúan entre 0 y 4,11 % para el conjunto de las 64 sustancias estudiadas.

El modelo óptimo señalado anteriormente se sometió a la prueba de validez externa [6]. Para ello, del grupo de 64 hidrocarburos, 5 de ellos, elegidos al azar, se excluyeron del grupo y, en esas condiciones, se obtuvo la ecuación de regresión con los restantes hidrocarburos (59). Interpolando los valores de  $C_g$  y  $\chi_v^1$  de las 5 sustancias de prueba, se obtuvo para ellas, valores de ( $I_{apolar\ experimental} - I_{apolar\ calculado}$ ) muy semejantes a los obtenidos cuando estas sustancias forman parte del

conjunto de 64 sustancias a través del cual se obtiene la expresión de regresión inicial. Esto indica un alto poder predictivo de la ecuación principal propuesta [6]. En este trabajo, a diferencia del modelo anterior [6], se utiliza la mitad del número de variables independientes en la ecuación de correlación óptima que relacionan  $I_{apolar}$  versus la variable dependiente con lo cual disminuye el efecto de un aumento del estadígrafo  $R^2$  debido al alto número de variables independiente [7, 8]. Para la relación anterior, ambos trabajos presentan un  $R^2$  y d.e similares, sin embargo, este trabajo presenta un estadígrafo  $F$  mayor para un muy similar número de casos, un número de variables independientes equivalentes a la mitad. El modelo anterior para la ecuación de regresión óptima [6], utiliza descriptores de laboriosa obtención (sendas de tercer orden, sendas de subestructuras) y requiere de software muy especializados y de alto costo para el cálculo del tercer eje mayor molecular.

Las variables  $C_g$  y  $\chi_v^1$  tomadas como vectores tienden a la ortogonalidad entre sí, esto implica que cada una de ellos describe diferentes propiedades de la estructura molecular que definen y solo duplican información en pequeña proporción. Esto se evalúa mediante la correlación entre ambas variables obteniéndose un coeficiente de correlación bajo [7].

Se evaluó la utilidad de las variables dependientes como la ecuación óptima propuesta como herramienta de orientación para el análisis cualitativo de los constituyentes del grupo de sustancias bajo estudio y otros hidrocarburos no pertenecientes a este grupo. Para ello se consideraron dos situaciones. La primera, al considerar una muestra problema que contiene solo elementos del grupo en estudio, la concordancia de los parámetros de retención, en particular la diferencia de estas retenciones ( $I_{apolar} - I_{polar}$ ) en columnas de alta eficiencia tanto de SE-30 como de Carbowax-20M, (con un margen de error de 3 unidades de índice de Kovats [6], proporciona elementos de juicio analíticos para la identificación con una baja posibilidad de error. En la segunda situación, se supone la existencia de hidrocarburos distintos a los 64 hidrocarburos bajo estudio, se supone además que no se dispone de los valores de las variables dependientes obtenidas en las condiciones cromatográficas utilizadas para obtener la ecuación óptima que en este trabajo se plantea. Al calcular los índices  $C_g$  y  $\chi_v^1$  sobre la base de la estructura de estos hidrocarburos, el parámetro de retención  $I_{apolar\ calculado}$ , se podrá obtener por interpolación en la ecuación óptima propuesta en este trabajo y comparar con lo señalado en la bibliografía [16]. Bajo estas circunstancias, el proceso analítico tiene un carácter de orientación para la identificación de estos hidrocarburos. Para su identificación con una baja probabilidad de error, se deben contemplar además métodos analíticos específicos, como una análisis cromatográfico gaseoso equipado con espectrómetro de masas. Como ejemplo de este segundo caso se utilizó el 4-metiloctano.

**Tabla 1.** Parámetros de retención cromatográficas, índices de retención, residuos y errores.

	Nombre de la Sustancia	SE-30	Carbowax 20M		$C_g$	$\chi_1^v$	$I_{\text{apolar}}$	Residuo	% Error
		$I_{\text{apolar}}$	$I_{\text{polar}}$	$\eta$			$I_{\text{apolar}}$ Calculado		
1	1,2,3,4-tetrahidronaftaleno	1129.8	1490.4	360.6	105.14011	4.03446	1108.1	21.7	1.92
2	tolueno	759.9	1049.5	289.6	60.66075	2.41068	771.7	-11.8	-1.55
3	<i>o</i> -xileno	882.2	1185.7	303.5	78.36070	2.82735	885.0	-2.8	-0.32
4	<i>m</i> -xileno	859.6	1145.3	285.7	76.19098	2.82137	875.2	-15.6	-1.81
5	<i>p</i> -xileno	860.8	1138.9	278.1	74.18480	2.82137	866.6	-5.8	-0.67
6	1,2,3-trimetilbenceno	1006.0	1325.4	319.4	96.03163	3.24402	998.1	7.9	0.79
7	1,2,4-trimetilbenceno	977.5	1277.6	300.1	92.45077	3.23803	982.2	-4.7	-0.48
8	1,3,5-trimetilbenceno	956.2	1242.2	286.0	92.91167	3.23205	983.7	-27.5	-2.88
9	1,2,3,4-tetrametilbenceno	1127.5	1461.6	334.1	111.90107	3.66068	1103.5	24.0	2.13
10	1,2,4,5-tetrametilbenceno	1096.4	1406.8	310.4	109.15409	3.65470	1091.2	5.2	0.47
11	1,2,3,5-tetrametilbenceno	1099.5	1416.5	317.0	111.29554	3.65470	1100.4	-0.9	-0.08
12	etilbenceno	851.4	1131.9	280.5	69.23028	2.97134	858.8	-7.4	-0.87
13	1-metil-2-etilbenceno	964.6	1258.4	293.8	85.64329	3.38801	966.5	-1.9	-0.20
14	1-metil-3-metilbenceno	949.0	1224.9	275.9	84.42477	3.38203	960.7	-11.7	-1.23
15	1-metil-4-metilbenceno	950.6	1233.4	282.8	82.74701	3.38203	953.5	-2.9	-0.31
16	1,2-dimetil-3-etilbenceno	1083.5	1394.8	311.3	102.70853	3.80468	1077.0	6.5	0.60
17	1,2-dimetil-4-etilbenceno	1066.1	1357.2	291.1	98.87777	3.79869	1060.1	6.0	0.56
18	1,3-dimetil-2-etilbenceno	1069.6	1372.1	302.5	103.31507	3.80468	1079.6	-10.0	-0.93
19	1,3-dimetil-4-etilbenceno	1061.7	1350.0	288.3	100.01562	3.59159	1046.4	15.3	1.44
20	1,3-dimetil-5-etilbenceno	1042.6	1319.8	277.2	100.91243	3.79271	1068.3	-25.7	-2.46
21	1,4-dimetil-2-etilbenceno	1059.0	1343.5	284.5	99.85636	3.79869	1064.3	-5.3	-0.50
22	1,2-dietilbenceno	1043.5	1324.0	280.5	95.80382	3.94867	1060.3	-16.8	-1.61
23	1,3-dietilbenceno	1034.8	1297.3	262.5	92.57491	3.94269	1046.0	-11.2	-1.08
24	1,4-dietilbenceno	1040.1	1305.2	265.1	91.22100	3.94269	1040.1	0.0	0.00
25	<i>n</i> -propilbenceno	941.8	1210.1	268.3	76.11232	3.47134	933.1	8.7	0.92
26	1-metil-2-propilbenceno	1048.3	1327.7	279.4	92.40000	3.88801	1040.3	8.0	0.76
27	1-metil-3-metilbenceno	1034.8	1301.0	266.2	90.87184	3.88203	1033.2	1.6	0.15
28	1-metil-4-metilbenceno	1038.1	1301.0	262.9	89.20153	3.88203	1026.0	12.1	1.17
29	1-metil-4-isopropilbenceno	1009.5	1268.8	259.3	97.46083	3.76474	1051.0	-41.5	-4.11
30	1,2-dimetil-3-propilbenceno	1166.5	1458.6	292.1	109.12489	4.30468	1149.4	17.1	1.47
31	1,2-dimetil-4-propilbenceno	1149.5	1435.8	286.3	105.24128	4.29869	1132.2	17.3	1.51
32	1,3-dimetil-2-propilbenceno	1152.2	1451.6	299.4	109.74327	4.30468	1152.0	0.2	0.02
33	1,3-dimetil-4-propilbenceno	1143.7	1429.0	285.3	106.81680	4.29869	1138.9	4.8	0.42
34	1,3-dimetil-5-propilbenceno	1126.0	1406.2	280.2	107.51561	4.29271	1141.4	-15.4	-1.37
35	1,4-dimetil-2-propilbenceno	1140.0	1415.0	275.0	106.37778	4.29869	1137.1	2.9	0.25
36	<i>n</i> -butilbenceno	1039.6	1309.1	269.5	82.12462	3.97134	1003.7	35.9	3.45
37	isobutilbenceno	994.3	1241.3	247.0	85.95716	3.82719	1007.2	-12.9	-1.30
38	sec-butilbenceno	997.0	1248.1	251.1	90.95261	3.89206	1034.5	-37.5	-3.76
39	alilbenceno	932.6	1263.2	330.6	78.72694	3.08073	909.3	23.3	2.50
40	1,3-divinilbenceno	1091.3	1541.0	449.7	111.01054	3.21521	1059.8	31.5	2.89
41	1,4-divinilbenceno	1100.0	1554.2	454.2	113.18000	3.21521	1069.1	30.9	2.81
42	4-(metileno)etilbenceno	965.9	1454.3	488.4	92.27234	3.01829	961.8	4.1	0.42
43	3-(metileno)etilbenceno	960.5	1450.9	490.4	94.31918	3.01829	970.6	-10.1	-1.05
44	estireno	877.8	1255.1	377.3	78.48547	2.60761	865.8	12.0	1.37
45	2-metilestireno	977.5	1342.2	364.7	103.86228	3.02427	1012.0	-34.5	-3.53
46	3-metilestireno	977.0	1348.1	371.1	91.94029	3.01829	960.4	16.6	1.70
47	4-metilestireno	980.5	1348.6	368.1	92.27234	3.01829	961.8	18.7	1.91
48	2,4-dimetilestireno	1075.4	1440.5	365.1	118.87591	3.43496	1113.2	-37.8	-3.51
49	2,5-dimetilestireno	1078.7	1432.2	353.5	117.80142	3.43496	1108.6	-29.9	-2.77
50	2,3-dimetilestireno	1100.3	1485.1	384.8	121.03336	3.44094	1123.0	-22.7	-2.06
51	4-etilestireno	1073.2	1431.2	358.0	100.47660	3.57895	1047.2	26.0	2.42
52	3-etilestireno	1065.8	1423.7	357.9	101.75108	3.57895	1052.7	13.1	1.23
53	$\alpha$ -metilestireno	966.4	1320.9	354.5	100.56996	3.01424	997.0	-30.6	-3.17
54	<i>cis</i> - $\beta$ -metilestireno	975.3	1324.3	349.0	87.55078	3.11004	949.8	25.5	2.61
55	2-metil-trans- $\beta$ -metilestireno	1098.3	1464.0	365.7	113.28414	3.52671	1097.5	0.8	0.07
56	dipenteno	1019.8	1200.0	180.2	83.11657	4.00920	1011.4	8.4	0.82
57	indano	1015.7	1355.9	340.2	91.68093	3.28446	983.1	32.6	3.21
58	4-metilindano	1120.6	1467.9	347.3	110.47300	3.95112	1123.5	-2.9	-0.26
59	5-metilindano	1112.2	1446.9	334.7	106.60997	3.69514	1083.9	28.3	2.54
60	<i>trans</i> -hexahidroindano	949.8	1059.3	109.5	73.08453	4.13299	979.4	-29.6	-3.12
61	<i>cis</i> -hexahidroindano	980.9	1107.3	126.4	73.08453	4.13299	979.4	1.5	0.15
62	<i>cis</i> -decahidronaftaleno	1081.5	1223.0	141.5	81.67184	4.46633	1046.1	35.4	3.27
63	<i>trans</i> -decahidronaftaleno	1041.7	1160.9	119.2	81.67184	4.46633	1046.1	-4.4	-0.42
64	tetrahidrociclopentadieno	1077.6	1243.4	165.8	86.37918	4.94949	1109.6	-32.0	-2.97

## Desarrollo

Los índices topológicos de conectividad orden 1 tipo valencia [12-14] se calcularon utilizando la expresión (1)

$$\chi_v^1 = \sum [(d_i d_j)^{-1/2}]^k \quad (1)$$

Los elementos  $d_i$  como  $d_j$  corresponden a números asignados al átomo  $i$  ( $j$ ) del grafo molecular. Estos elementos se calcularon mediante la siguiente expresión (2)

$$d_{(i,j)} = (Z^v - h) / (Z - Z^v - I) \quad (2)$$

donde  $Z^v$  es el número de electrones de valencia,  $Z$  es el número total de electrones y  $h$  es el número total de hidrógenos unidos al átomo  $i$  ( $j$ ).

El índice topológico  $C_g$  se calculó con el mismo algoritmo expresado en (1) pero con definiciones distintas para  $d_i$ , ( $d_j$ ) que pasan a denominarse  $C_i$ , ( $C_j$ ). El cálculo numérico de  $C_i$ , ( $C_j$ ) se realizó evaluando la densidad de carga de cada átomo de carbono del hidrocarburo. Estas cargas se calculan mediante el método AM1 de la química cuántica y se le adjunta el proceso iterativo de Polak-Ribiere bajo las condiciones de convergencia de 0.1 Kcal / (A° mol). Este proceso se llevó a cabo mediante el software Hyperchem [17]. El algoritmo para el cálculo del índice topológico  $C_g$  se realizó mediante la relación (3).

$$C_g = \sum [(C_i C_j)^{-1/2}]_k \quad (3)$$

Las 64 sustancias en estudio, sus parámetros de retención cromatográficos [5] experimentales ( $I_{\text{apolar}}$ ,  $I_{\text{polar}}$ ,  $\Delta I$ ), los índices topológicos utilizados  $C_g$ ,  $\chi_v^1$ , los valores de los  $I_{\text{apolar}}$  calculados ( $I_{\text{apolar-calculados}}$ ) y el error porcentual sobre la diferencia en valor absoluto de ( $I_{\text{apolar-experimental}} - I_{\text{apolar-calculado}}$ ) se señalan en la Tabla 1. Para las 64 sustancias bajo estudio se calcularon ambos índices topológicos mediante el software Excel 2000 aplicando las relaciones (1) y (3), respectivamente, o en forma alternativa mediante un programa diseñado para la calculadora Hewlett Packard 48GX\*.

Las regresiones que involucran a las variables de retención cromatográficas (variables dependientes) con los índices topológicos (variables independientes) responden a los modelos matemáticos  $y = mx + n$ ;  $y = bx + cz$ . Analizadas todas las combinaciones posibles entre las variables dependientes e independientes, se obtuvieron 9 expresiones de correlación que se señalan en la Tabla 2.

En función de sus perfiles estadísticos  $r$ ,  $R$  (según involucre una o más variables independientes), de  $F$ , se propone la siguiente expresión como la correlación óptima (4).

$$I_{\text{apolar}} = 4.2885 (\pm 0.1945) * C_g + 89.6514 (\pm 5.1021) * \chi_v^1 + 295.4778 (\pm 22.7568) \quad (4)$$

$$R^2 = 0.9451$$

$$\text{d.e} = 20.1,$$

$$F = 525.2, N = 64$$

**Tabla 2.** Parámetros estadísticos de las regresiones estudiadas.

Variables dependientes	Variables independientes	$r$ o $R$	d.e	$F$
$I_{\text{apolar}}$	$C_g, \chi_v^1$	0.97216	20.2	525.2
$I_{\text{polar}}$	$C_g, \chi_v^1$	0.90952	49.1	146.0
$I_{\text{polar}}$	$C_g, \chi_v^1$	0.90523	49.8	281.4
$\Delta I$	$C_g, \chi_v^1$	0.87358	37.2	98.3
$I_{\text{apolar}}$		0.81688	49.2	124.3
$I_{\text{apolar}}$		0.71276	59.8	64.0
$\Delta I$		-0.58802	61.3	32.8
$\Delta I$		0.48052	66.3	18.6
$I_{\text{polar}}$		0.13841	116.0	1.2

Los símbolos  $I$ ,  $\Delta I$ ,  $\chi_v^1$ ,  $r$ ,  $R$ ,  $s.d$ ,  $F$  significan: Índice de Kovats, diferencia de los índices de Kovats entre columna de carácter polar (Carbowax 20-M) y apolar (SE-30), coeficiente de correlación, coeficiente de determinación, desviación estándar e índice de Fisher respectivamente.

A través esta regresión se calculó  $I_{\text{apolar-calculado}}$  lo que permite evaluar los residuos ( $I_{\text{apolar-experimental}} - I_{\text{apolar-calculado}}$ ) y el porcentaje de error sobre este residuo tomado como valor absoluto. Se obtuvo un error máximo de un 4.11 % y un mínimo de 0 % correspondientes a las sustancias 1-metil-4-isopropilbenceno y 1,4-dietilbenceno, respectivamente (Tabla 1).

Para las 64 sustancias estudiadas, la función  $I_{\text{apolar-calculado}}$  versus  $I_{\text{apolar-experimental}}$  presenta un  $r = 0.9722$ , un d.e = 18 y un coeficiente de Fisher = 1067.37. Los residuos calculados ( $I_{\text{apolar-experimental}} - I_{\text{apolar-calculado}}$ ) presentan una distribución gaussiana con una desviación estándar de 19.05 y un promedio de -0.68 unidades de índice de Kovats.

La evaluación de la duplicidad de información molecular contenida en los índices topológicos  $C_g$  y  $\chi_v^1$ , se realizó mediante la correlación entre ambos índices, obteniéndose un  $r = 0.24731$ , un d.e = 0.50 y un  $F = 4.040$ . Este bajo valor en la correlación indica que ambos índices tienden a la ortogonalidad [7], lo que está comprobado por el hecho de que el producto punto del primer vector por el vector ortogonal al primero tiende a [18, 19]. El valor obtenido fue de  $-1.60 \times 10^{-8}$  por lo tanto la duplicidad de información en  $C_g$  y  $\chi_v^1$  es mínima.

La relación  $I_{\text{apolar}}$  versus  $C_g$  no presenta una correlación significativa, como así mismo la relación  $I_{\text{apolar}}$  versus  $\chi_v^1$ , como se aprecia en los valores del coeficiente de correlación en la Tabla 2. Sin embargo, en la ecuación (4) ambos índices presentan una correlación alta, debido al hecho de que ambos contribuyen a abrir un espacio en el vector  $C_g$  (que no correlaciona con  $I_{\text{apolar}}$ ) con lo cual se produce una alta correlación para la relación  $I_{\text{apolar}}$  versus  $C_g$  y  $\chi_v^1$ , superior a la relación  $I_{\text{apolar}}$  versus  $\chi_v^1$ . Este fenómeno se produce con cierta frecuencia y está tratado en la literatura [7]

El índice topológico  $C_g$  presenta un grado degenerativo menor que el índice topológico  $\chi_v^1$ . Los grupos de sustancias (42, 47), (60, 61) y (62, 63) indicadas en la Tabla 1, presentan valores iguales del índice  $C_g$  para estructuras moleculares distintas (en los dos últimos grupos, las duplas de sustancias difieren en sus formas *cis* y *trans*).

**Tabla 3.** Parámetros de evaluación externa de la regresión propuesta, ecuación (4).

Sustancias usadas en la evaluación externa	$I_{\text{apolar-experimental}}$	$I_{\text{apolar-calculado(5)}}$	A % error (5)	B % error (4)
<i>p</i> -xileno	860.8	867.3	0.75	0.67
estireno	877.8	866.5	1.28	1.36
1,3,5-trimetilbenceno	956.2	984.3	2.94	2.87
sec-butilbenceno	997.0	1035.3	3.84	3.76
dipenteno	1019.8	1012.3	0.73	0.83
1,4-dimetil-2-etilbenceno	1059.0	1064.9	0.56	0.50
1,3-dimetil-4-etilbenceno	1061.7	1046.9	1.40	1.44
1,4-dimetil-2-propilbenceno	1140.0	1137.8	0.19	0.26

Los valores numéricos entre paréntesis identifican la ecuación de regresión utilizada. Los promedios de la columna A y B son 1.46, 1.46 y sus varianzas 1.62 y 1.52, respectivamente.

Para el caso del índice topológico  $\chi_v^1$ , existen 13 grupos de sustancias con iguales valores del índice para distintas estructuras, de los cuales 11 están constituidos por duplas y 2 por tres sustancias, véase Tabla 1.

Los valores de  $I_{\text{apolar-calculado}}$  obtenidos mediante la ecuación de regresión (4) presentan valores iguales solo en los casos de los isómeros *cis* y *trans*, véase Tabla 1. Esta propiedad discriminativa de las estructuras moleculares esta implícita en la definición anterior de  $C_i$  y  $C_j$ .

La relación óptima, ecuación (4), fue sometida a una validación externa según el procedimiento señalado en la literatura [6]. Estadísticamente, este tipo de prueba es más significativo que la de validación interna [6]. Para ello, se eligieron al azar 8 compuestos del conjunto de 64 sustancias bajo estudio. Sobre la base del conjunto restante (56 sustancias), excluidas las 8 elegidas, se realizó la correlación  $I_{\text{apolar}}$  versus  $C_g$  y  $\chi_v^1$  ecuación (5). Basado en esta ecuación, se calculó el residuo ( $I_{\text{apolar-experimental}} - I_{\text{apolar-calculado-ecuación-5}}$ ) para las 8 sustancias. Este residuo se compara con el obtenido mediante la ecuación (4) donde se consideran las 64 sustancias. Se debe señalar que, los promedios de ambos errores porcentuales no son estadísticamente diferentes, lo que da un nivel de confianza de un 95 %.

Empleando la ecuación (5) se realizó la correlación entre  $I_{\text{apolar-calculado}}$  versus  $I_{\text{apolar-experimental}}$ , dando lugar a la ecuación (6) con un  $r = 0.9800$ , semejante al obtenido en el trabajo que se tomó como modelo [6], lo que implica que el modelo que corresponde a la correlación (4) presenta un grado de robustez estadística comparable con la proposición de la literatura [6]. El número de sustancias elegidas al azar, es coincidente con el utilizado en el trabajo comparativo [6], y es suficiente para tener una baja probabilidad de que ocurra un coeficiente de correlación alto debido al azar.

Las ecuaciones (5), (6) se detallan a continuación:

$$I_{\text{apolar}} = 4.2737(\pm 0.2039)C_g + 89.9497(\pm 5.3185)\chi_v^1 + 296.4728(\pm 25.3569)$$

$$R^2 = 0.9419$$

$$d.e = 20.39$$

$$F = 429.87$$

$$N = 56$$

$$I_{\text{apolar-calculado-ecuación-5}} = 0.9723(\pm 0.0845) * I_{\text{apolar-experimental}} + 32.964(\pm 8.0500)$$

$$r = 0.9801$$

$$d.e = 20.42$$

$$F = 146.161$$

$$N = 8$$

Las 8 sustancias elegidas al azar, sus  $I_{\text{apolar-calculado-ecuación-5}}$ ,  $I_{\text{apolar-experimental}}$  y los porcentajes de error del residuo tomado como valor absoluto, obtenidos a través de la ecuación 4 y 5 se señalan en la Tabla 3

La orientación en el análisis de identificación de las 64 sustancias utilizadas en este trabajo, se basa en la comparación de los índices de Kovats recogidos en la literatura [6] tanto para columnas de SE-30 como de Carbowax-20M y sus diferencias [ $(I_{\text{apolar-experimentales}} - I_{\text{apolar-experimental}})$ ] con los obtenidos en el laboratorio en idénticas condiciones, con un margen de error de 3 unidades de índice de Kovats.

Para el caso de hidrocarburos distintos a los estudiados, se procede planteando una hipótesis como se indica en el siguiente ejemplo. Si se sospecha que una muestra contiene 4-metiloctano, se calculan sus índices topológicos  $C_g$  y  $\chi_v^1$  (52.56930 y 4.00920, respectivamente) y al interpolarlos en la ecuación (4) se obtiene un valor de  $I_{\text{apolar-calculado}}$  igual a 880.3. En la literatura [16] se señala que, en similares condiciones cromatográficas, el índice experimental es de 860.0, lo que implica un error de estimación de un 2.33 % que está dentro del error encontrado entre  $I_{\text{apolar-experimental}} - I_{\text{apolar-calculado}}$  para las 64 sustancias bajo estudio. Un método confirmatorio de la estructura definitiva exige de otros análisis como análisis cromatográfico equipado con espectrómetro de masas que permite mediante la superposición de espectros con patrones certificados tener evidencias de la estructura con una alta probabilidad de acierto, además se deben confirmar los resultados de identidad con los antecedentes que aportan la espectroscopia de absorción en el infrarrojo y la espectroscopia de resonancia magnética nuclear para el protón como para carbono 13. Por otra parte, mediante este mecanismo de interpolación se puede determinar aproximadamente el orden de elusión para un conjunto de hidrocarburos distintos a los estudiados. Se debe señalar que el proceso de interpolación debe ser rea-

lizado mediante una sustancia a la vez, de lo contrario se puede producir una pérdida de la robustez estadística de la ecuación.

Tanto las regresiones como la comparación estadística de promedios se realizó mediante el Software Origin 4.0 [20].

- Programa para Hewlett Packard 48GX para calcular(\*):
  - $C_g$  y  $\chi_v^1$ : Programa = <<  $\rightarrow$  X Y << XY \* .5^ INV  
 $\Sigma$  LIST>> >>; X e Y vectores

## Conclusiones

a. Para las 64 sustancias estudiadas, el índice topológico propuesto ( $C_g$ ), presenta un grado degenerativo significativamente menor que el índice de conectividad químico ( $\chi_v^1$ ), y una correlación mayor con los índices de Kovats tanto en fase estacionaria polar como apolar. Sin embargo, el índice  $\chi_v^1$  presenta una mejor correlación que el índice propuesto ( $C_g$ ), con la variable dependiente correspondiente a la diferencia de los índices señalados anteriormente.

b. La regresión óptima propuesta se configura con una variable dependiente que corresponde al índice de Kovats en columna apolar y dos variables independientes, la primera basada en una proposición original tratada en este trabajo y la segunda obtenida de la literatura. Prácticamente ambas no son redundantes en la información molecular y, utilizadas en forma contributiva, producen una regresión con altos índice de determinación y de Fischer lo que se traduce en un poder de predicción significativo para la variable dependiente.

c. El fenómeno de retención cromatográfico de las sustancias bajo estudio en columna apolar, queda determinado por dos contribuciones, la distribución de cargas en los distintos tipos de carbonos de la molécula ( $C_g$ ) y por la estructura molecular ( $\chi_v^1$ ).

d. La regresión óptima propuesta presenta solo 2 variables independientes que involucran cálculos de fácil obtención en comparación con la proposición tomada como modelo que contiene 4 variables independientes. La obtención de estas últimas variables requiere de software de alto costo o procesos laboriosos. Para la regresión propuesta, se obtuvo un perfil estadístico similar al del modelo.

e. Se obtuvo una regresión estadísticamente válida, que con carácter de orientación, es útil en la identificación de hidrocarburos distintos a los 64 hidrocarburos provenientes de la pirólisis de las naftas.

## Bibliografía

1. Cornwell, E. *Bol. Soc. Chil. Quím.* **2000**, *45*, 649-653.
2. Cornwell, E. *Bol. Soc. Chil. Quím.* **2002**, *47*, 53-60.
3. Kalizan, R. *Anal. Chem.* **1992**, *64*, 619A-A631A.
4. Mihalic, Z.; Trinajstic, N. *J. Chem. Educ.* **1992**, *69*, 701-712.
5. Report for Analytical Chemistry. *Anal. Chem.* **1964**, *36*, 31A-40A.
6. Woloszyn, T. F.; Jurs, P. C. *Anal. Chem.* **1993**, *65*, 582-587.
7. Randic, M. J. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 672-687.
8. Toplis, J. G.; Edwards, R. P. *J. Med. Chem.* **1979**, *22*, 1238-1244.
9. Morgan, E. *Chemometric Experimental Design*. John Wiley and Sons, **1991**.
10. Randic, M. J. *Am. Chem. Soc.* **1975**, *97*, 6609-6615.
11. Estrada, E. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1022-1025.
12. Kier, L. B.; Hall L. H. *J. Pharm. Sci.* **1983**, *72*, 11701173.
13. Sabljic, A, Horvatic, D. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 292-295.
14. Kier, L. B.; Hall, L. H, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 792-795.
15. Khadikar, P.V.; Karmokar, S.; Agrawal, V. K. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 934-949
16. Zweig, G. Ph. D.; Sherma, J Ph.D *Handbook of Chromatography*. CRC Press, Vol I, **1972**.
17. Hyperchem, Hypercube and Autodesk, Inc. Developed by Hypercube Inc. **1993**.
18. Perry, W. L. *Álgebra Lineal con Aplicación*. Mc. Graw Hill, **1990**.
19. Randic, M. J. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 686-692
20. Microcal Origin 4.00 Software Inc. One Roundhouse Plaza. Northampton. Ma. 01060. USA **1995**.