

## RELACIONS ESTRUCTURA-PROPIETAT PER UN CONJUNT D'HIDROCARBURS A PARTIR DE NOUS DESCRIPTORS TRIDIMENSIONALS DERIVATS DE LA SEMBLANÇA MOLECULAR QUÀNTICA

M. Lobato, E. Besalú i R. Carbó

Institut de Química Computacional. Universitat de Girona. 17071. Girona, Catalunya (Espanya)

---

### RESUM

En aquest article es defineixen uns nous índexs tridimensionals per a la descripció de les molècules a partir de paràmetres derivats de la Teoria de la Semblança Molecular i de les distàncies euclidianes entre els àtoms i les càrregues atòmiques efectives. Aquests índexs, anomenats 3D, s'han aplicat a l'estudi de les relacions estructura-propietat d'una família d'hidrocarburs, i han demostrat una capacitat de descripció de tres propietats de la família (temperatura d'ebullició, temperatura de fusió i densitat) molt més acurada que quan s'utilitzen els índexs 2D clàssics.

### RESUMEN

En este artículo se definen unos nuevos índices tridimensionales, aplicados a la descripción de las moléculas, a partir de parámetros derivados de la Teoría de la semejanza Molecular y de las distancias euclídeas entre átomos y cargas atómicas efectivas. Estos índices, llamados 3D, se han aplicado al estudio de las relaciones estructura-propiedad de una familia de hidrocarburos, demostrando una capacidad de descripción de tres propiedades de la familia (temperatura de ebullición, temperatura de fusión, y densidad) mucho más precisa que cuando se utilizan los índices 2D clásicos.

### ABSTRACT

In this work, some new tridimensional molecular descriptors are described, making use of parameters derived from Molecular Similarity Theory, and also euclidean distances between atoms and effective atomic charges. These descriptors, here called 3D indices, are applied to the study of Structure-Property Relationships of a family of hydrocarbons and the properties Boiling Point, Melting Point and Density. The use of this 3D indices gives better results, in all cases, than the ones obtained when 2D topological indices are used.

**Keywords:** Molecular Descriptors, Quantum Molecular Similarity, Molecular Integrals, Graph-Theoretical Indices, QSAR, QSPR.

---

## INTRODUCCIÓ

Darrerament, a l'Institut de Química Computacional s'han fet estudis sobre la Semblança Molecular i la seva aplicació a relacions estructura-activitat i estructura-propietat, com també dels fonaments teòrics de dites relacions. Les referències més recents es poden trobar en (1-3). També s'ha estudiat com, gràcies a una reformulació i extensió de les definicions dels índexs topològics, es pot arribar a establir un pont entre la Connectivitat Molecular i la Semblança Molecular Quàntica (4). Aquí

es presenta la definició i l'estudi de nous índexs i l'aplicació a una família de compostos, i es comparen els resultats de relacions estructura-propietat que s'obtenen amb la metodologia clàssica i amb la metodologia desenvolupada al nostre laboratori.

## DESCRIPTORS MOLECULARS

Les molècules es poden descriure de diferents maneres. Una de les més senzilles seria fer ús de la seva fórmula molecular. També es pot descriure la molècula, dins el marc de la Topologia Molecular, com un graf format per un conjunt de punts (àtoms) i vèrtexs (enllaços). A partir d'aquesta descripció es poden definir una sèrie d'índexs moleculars (topològics), que s'utilitzen en estudis de relacions estructura-activitat i estructura-propietat (8). Aquests estudis es basen en la suposició que la *topologia* de la molècula és la que dóna lloc a les seves propietats. Avui en dia, encara els índexs utilitzats més regularment són alguns dels primers que es van definir: l'índex de Wiener (5) i l'índex de Connectivitat de Randić (6), o bé modificacions d'aquests. Tots aquests índexs es basen en una descripció de les molècules a partir del seu graf d'hidrògens suprimits (l'esquelet molecular), on s'han menyspreat els àtoms d'hidrogen. Aquesta eliminació es justifica a partir del fet que, en un principi, es van definir per a l'estudi d'hidrocarburs, on els hidrògens contribueixen poc a les propietats moleculars i no determinen la seva estructura (7), a causa de la seva petita mida. Els índexs de connectivitat generalitzats (8) es basen en sumes de contribucions de subgrafs de diferents tipus al graf d'hidrògens suprimits de la molècula. Avui en dia, es poden trobar a la bibliografia (9) més de cent vint índexs d'ús comú, incloent-hi els d'informació estadística i els *ad hoc* (10).

En el nostre Institut s'ha treballat en la definició de nous índexs (4), que podem anomenar índexs 3D, derivats no del graf molecular (bidimensional) sinó de la geometria tridimensional de les molècules. Les principals diferències entre aquests nous índexs 3D i els utilitzats clàssicament són que els primers utilitzen la informació que proporciona l'estructura tridimensional, tenen en compte la contribució dels hidrògens a la molècula, i diferencien els àtoms que formen la molècula; és a dir, s'utilitza més informació química.

L'estructura d'aquests nous índexs es basa en la utilització de paràmetres derivats de les integrals que es fan servir típicament en la Teoria de la Semblança Molecular (11): integrals de recobriment, de repulsió electrònica i gravitacionals. També s'utilitzen les distàncies tridimensionals i les càrregues atòmiques efectives, que en aquest treball s'han calculat mitjançant la metodologia AM1 implementada al programa AMPAC v5.0 (12). La **Taula 1** descriu els diferents paràmetres que s'han utilitzat en la definició dels índexs 3D. Per calcular aquests paràmetres, es representa cada àtom mitjançant una funció gaussiana normalitzada de tipus *1s*. Com a exponents de les funcions es poden utilitzar els derivats de l'optimització de l'energia dels àtoms aïllats (13), o bé els desenvolupats al nostre Institut. Aquests darrers exponents s'obtenen de tal manera que reproduueixen el 95% de la densitat atòmica dins una esfera centrada en l'origen i que té com a radi el radi covalent de l'àtom ( $R_c$ ).

Així, s'han obtingut com a solució de l'equació següent :

$$\int_0^{R_c} |\psi|^2 d\tau = 0.95 \quad (1)$$

Paràmetre	Definició
$d_{ij}$	Distància tridimensional euclidiana entre els àtoms $i$ i $j$
$S_{ij}^v$	Integral de solapament entre els orbitals corresponents als àtoms $i$ i $j$
$R_{ij}$	Integral de repulsió entre els orbitals corresponents als àtoms $i$ i $j$
$G_{ij}$	Integral gravitacional entre els orbitals corresponents als àtoms $i$ i $j$
$v_i$	Suma dels elements de la matriu de solapament corresponent a la fila $i$ (valència de solapament)
$vr_i$	Suma dels elements de la matriu d'integrals de repulsió corresponent a la fila $i$ (valència de repulsió)
$vg_i$	Suma dels elements de la matriu d'integrals gravitacionals corresponent a la fila $i$ (valència gravitacional)
$q_i$	Càrrega efectiva de l'àtom $i$

Taula 1. Paràmetres que s'utilitzen en el càlcul dels índexs 3D de descripció molecular.

on  $\Psi = N \exp(-\alpha r^2)$  és la funció gaussiana  $1s$  associada a cada àtom de la qual es parlava anteriorment i  $N$  és el seu factor de normalització. El plantejament de l'equació (1) ens porta a resoldre numèricament l'equació següent:

$$\sqrt{\frac{2}{\pi}} \alpha R_c \exp(-2 \alpha R_c^2) + \operatorname{erfc}(R_c \sqrt{2 \alpha}) - \frac{1}{40} = 0 \quad (2)$$

on  $\operatorname{erfc}(a)$  és la funció d'error complementària per l'escalar  $a$ . La solució de l'equació (2) porta a l'obtenció del valor de l'exponent  $\alpha$  per un radi covalent prefixat.

Aquests exponents descriuen més bé el volum atòmic, i això millora, en certs casos, els càlculs de relacions estructura-propietat respecte a utilitzar els exponents basats en l'energia. En aquest treball es presenten els resultats amb els dos tipus d'exponents. A la Taula 2 es poden veure els valors dels exponents pels àtoms de carboni i d'hidrogen.

Àtom	Exponents nous	Exponents anteriors
H	6.87270	0.39017
C	0.93476	0.70085

Taula 2. Exponents de funcions GTO, nous (defïnits en el text) i antics (d'optimització d'energies, ref. 13) del carboni i l'hidrogen.

En el programa de càlcul dels índexs 3D que aquí es presenten s'utilitza un sistema de bucles aniuats (14) per tal de generar-los a partir de combinar els elements que apareixen a la Taula 1. L'estructura dels índexs 3D que es generen segueix la forma

$$\gamma = \sum_i^{Nat} \sum_j^{Nat} K_i \cdot K_j \quad (3)$$

on  $\gamma$  és l'índex 3D calculat,  $Nat$  és el nombre d'àtoms de la molècula i  $k_i$ ,  $k_j$  són les diferents contribucions, tal i com s'han especificat anteriorment i es recullen a la **Taula 1**. Aleshores es poden definir totes les possibilitats de combinació de les contribucions definides, de manera que cada índex 3D deriva d'una d'aquestes combinacions. Això dona lloc a tota una sèrie d'índexs 3D amb valor únic per cada molècula, que s'anomenen els seus descriptors, i que s'utilitzen per a l'estudi de relacions estructura-activitat i estructura-propietat. Per realitzar aquest treball s'han utilitzat un total de vint índexs 2D clàssics i un total de setanta-un índexs 3D, generats segons l'equació (3).

### RELACIONS ESTRUCTURA-ACTIVITAT I ESTRUCTURA-PROPIETAT

Tots els paràmetres anteriorment descrits es poden utilitzar en un programa de regressió multilíneal per tal de correlar els seus valors (que, de fet, són una codificació numèrica de l'estructura de les molècules) amb una propietat fisicoquímica, que dona a relacions estructura-propietat, o bé amb la seva activitat biològica, tal com la toxicitat, potència respecte a un cert analgèsic, etc. per establir relacions estructura-activitat. Aleshores, es poden fer estudis anàlegs amb els índexs 2D clàssics o bé amb els índexs 3D, tot fent la regressió entre la propietat o activitat i la representació de les molècules, que és expressada pel seu conjunt d'índexs. Això es pot expressar de forma matricial com la resolució de l'equació:

$$A X = \Pi \quad (4),$$

on  $X$  és la matriu que representa el conjunt de coeficients que s'han de determinar. En aquesta equació, la matriu  $A$  està formada per vectors columna que s'obtenen en calcular el valor de cada índex per totes les molècules; l'element  $a_{ij}$  correspon al valor de l'índex  $j$  per la molècula  $i$ . La matriu  $\Pi$  es forma agrupant per columnes les diferents propietats de la família que es vol descriure.

L'equació completa (4) es pot reduir per tal de poder estudiar una sola propietat de la família i amb un nombre d'índexs determinat. Aleshores el que es resol és una equació del tipus

$$A x = \pi \quad (5),$$

on  $x$  i  $\pi$  ara són vectors columna extrets, respectivament, de les matrius  $X$  i  $\Pi$  de coeficients i propietat.

Per tal de distingir quins són els paràmetres que millor descriuen una propietat determinada, es resol l'equació (5) per diferents formes de la matriu  $A$ , que s'obtenen d'escollir diferents conjunts d'índexs de manera que no es treballa amb la matriu de tots els índexs, sinó amb una reduïda, que conté alguns d'ells, sempre procurant que el seu nombre sigui significativament inferior al nombre de molècules en estudi. En aquest treball es fa el càlcul de la regressió multilíneal primer prenent un conjunt d'un sol índex; després, amb totes les parelles possibles; amb grups de tres, etc. fins a un límit marcat per l'entrada del programa, que pels resultats aquí presentats és de quatre. El conjunt de paràmetres recollits en el vector  $x$  que dona de sortida el programa és el considerat posteriorment per fer prediccions del valor de la propietat per molècules que no han intervingut en la regressió, de les quals podem

calcular els índexs coneixent la seva estructura.

De tots els conjunts possibles amb un nombre determinat de variables, es tria la millor regressió com aquella que dona una millor estimació del coeficient de predicció ( $Q^2$ ) (15), atès que el que interessa de les relacions estructura-propietat és la capacitat dels paràmetres emprats per a fer prediccions.

### RESULTATS I DISCUSSIÓ

Tota la metodologia abans descrita ha estat aplicada a una família d'hidrocarburs formada per alcans ramificats, sense ramificar, alquens i compostos cíclics. Per aquests compostos es coneixen els seus valors de temperatura d'ebullició, temperatura de fusió i densitat. Les molècules de la família, com també les seves propietats, es poden veure a la **Taula 3**.

Molècula	Temp. ebullició (K)	Temp. fusió (K)	Densitat (g/ml)
Isobutà	261.30	113.60	0.5490
2,2,3-Trimetilbutà	353.90	248.80	0.6901
3-Etil-2-metilpentà	388.60	158.00	0.7193
2,2,3-Trimetilpentà	383.30	160.70	0.7161
2,2,4-Trimetilpentà	372.20	165.60	0.6919
3-Metilheptà	392.10	152.50	0.7058
2-Metiloctà	415.80	192.90	0.7107
4-Metilnonà	438.70	171.40	0.7323
3,3-Dimetilpentà	359.10	138.60	0.6936
2,2-Dimetilheptà	405.70	160.00	0.7105
2,3-Dimetilbutà	331.00	144.50	0.6616
2,2-Dimetil-3-etilpentà	406.80	173.70	0.7438
Pentadecà	543.60	283.00	0.7685
Hexà	342.00	178.00	0.6603
Octà	398.70	216.20	0.7025
a-Pinè	429.00	218.00	0.8582
Camfè	431.00	324.00	0.8790
E-Decalina	468.00	242.60	0.8699
Etilciclohexà	404.80	161.70	0.7880
Propilciclopentà	404.20	155.70	0.7763
Butilciclohexà	454.50	198.30	0.7992
1-Metilciclopentè	348.50	145.80	1.4367
Trifeniletilè	493.00	345.00	1.0373
Cicloheptè	388.00	217.00	0.8228
Z-Ciclooctè	411.40	261.00	0.8472

**Taula 3.** Família d'hidrocarburs estudiada: molècules i propietats de cadascuna d'elles (temperatura d'ebullició, temperatura de fusió i densitat). Totes les dades són de la referència 16.

Els resultats de la regressió multilíneal entre les diferents propietats i els índexs 2D clàssics estan tabulades de manera que es presenten diferents paràmetres estadístics:

$Q^2$ : Coeficient de predicció aproximat. Ens dona indicació de la capacitat predictiva del model de regressió.

$r^2$ : Coeficient de correlació lineal.

$s_N$ : Desviació estàndard.

$p$ : Significança estadística.

Aquests resultats es presenten en les taules 4, 5, i 6

Nombre d'índexs	$Q^2$	$r^2$	$s_N$	$p$
1	0.859	0.902	17.27	< 0.001
2	0.913	0.949	12.39	< 0.001
3	0.940	0.957	11.44	< 0.001
4	0.965	0.976	8.53	< 0.001

*Taula 4. Resultats de la regressió multilíneal entre la propietat temperatura d'ebullició i la matriu d'índexs 2D clàssics.*

Nombre d'índexs	$Q^2$	$r^2$	$s_N$	$p$
1	0.461	0.568	37.86	< 0.01
2	0.526	0.629	35.07	< 0.005
3	0.628	0.763	28.01	< 0.001
4	0.610	0.769	27.65	< 0.001

*Taula 5. Resultats de la regressió multilíneal entre la propietat temperatura de fusió i la matriu d'índexs 2D clàssics.*

Nombre d'índexs	$Q^2$	$r^2$	$s_N$	$p$
1	0.079	0.168	0.15	< 0.05
2	0.080	0.229	0.48	< 0.02
3	0.082	0.294	0.14	< 0.01
4	0.128	0.368	0.13	< 0.01

*Taula 6. Resultats de la regressió multilíneal entre la propietat densitat i la matriu d'índexs 2D clàssics.*

Els resultats indiquen que la temperatura d'ebullició està molt ben descrita pels índexs 2D topològics clàssics, mentre que la temperatura de fusió no dona tan bons resultats. Pel que respecta a la densitat, que és una propietat més dependent de les

interaccions moleculars, s'obtenen uns resultats poc acurats.

Els resultats obtinguts amb els índexs 3D nous que aquí es descriuen són els que apareixen a les taules 7, 8 i 9.

**Taula 7**

Nombre d'índexs	Q <sup>2</sup>	r <sup>2</sup>	s <sub>N</sub>	p
1	0.871	0.918	15.78	< 0.001
	<i>0.889</i>	<i>0.915</i>	<i>16.01</i>	<i>&lt; 0.001</i>
2	0.932	0.957	11.41	< 0.001
	<i>0.941</i>	<i>0.961</i>	<i>10.80</i>	<i>&lt; 0.001</i>
3	0.979	0.988	5.94	< 0.001
	<i>0.976</i>	<i>0.985</i>	<i>6.69</i>	<i>&lt; 0.001</i>
4	0.991	0.994	4.30	< 0.001
	<i>0.985</i>	<i>0.993</i>	<i>4.52</i>	<i>&lt; 0.001</i>

**Taula 7.** Resultats de la regressió multilíneal entre la propietat temperatura d'ebullició i la matriu d'índexs 3D, calculats utilitzant els exponents d'optimització de volum (normal) i els d'optimització d'energia (en cursiva).

**Taula 8**

Nombre d'índexs	Q <sup>2</sup>	r <sup>2</sup>	s <sub>N</sub>	p
1	0.416	0.508	40.39	< 0.001
	<i>0.425</i>	<i>0.523</i>	<i>39.77</i>	<i>&lt; 0.001</i>
2	0.603	0.683	32.42	< 0.001
	<i>0.620</i>	<i>0.697</i>	<i>31.73</i>	<i>&lt; 0.001</i>
3	0.680	0.762	28.12	< 0.001
	<i>0.686</i>	<i>0.770</i>	<i>27.60</i>	<i>&lt; 0.001</i>
4	0.771	0.860	21.52	< 0.001
	<i>0.713</i>	<i>0.813</i>	<i>24.92</i>	<i>&lt; 0.001</i>

**Taula 8.** Resultats de la regressió multilíneal entre la propietat temperatura de fusió i la matriu d'índexs 3D, calculats utilitzant els exponents d'optimització de volum (normal) i els d'optimització d'energia (en cursiva).

**Taula 9**

Nombre d'índexs	Q <sup>2</sup>	r <sup>2</sup>	s <sub>N</sub>	p
1	0.247	0.385	0.128	< 0.1
	<i>0.254</i>	<i>0.348</i>	<i>0.132</i>	<i>&lt; 0.2</i>
2	0.388	0.641	0.098	< 0.001
	<i>0.360</i>	<i>0.493</i>	<i>0.116</i>	<i>&lt; 0.06</i>
3	0.464	0.739	0.083	< 0.001
	<i>0.815</i>	<i>0.903</i>	<i>0.051</i>	<i>&lt; 0.001</i>
4	0.478	0.815	0.070	< 0.001
	<i>0.863</i>	<i>0.924</i>	<i>0.045</i>	<i>&lt; 0.001</i>

**Taula 9.** Resultats de la regressió multilíneal entre la propietat densitat i la matriu d'índexs 3D, calculats utilitzant els exponents d'optimització de volum (normal) i els d'optimització d'energia (en cursiva).

Com es pot apreciar, l'ús dels nous índexs 3D descriu amb una qualitat anàloga a la dels índexs 2D clàssics la temperatura d'ebullició, mentre que millora la regressió quan es tracta de l'estudi de la temperatura de fusió i de la densitat, especialment aquesta darrera, que és on els índexs 2D clàssics donen resultats més poc acurats.

### Bibliografia

1. CARBÓ, R., BESALÚ, E., AMAT, LL. i FRADERA, X. en "QSAR and Molecular Modelling: Concepts, Computational Tools and Biological Applications", Sanz, F. (ed.), pàg. 396, Prous Pub., Barcelona (1995).
2. CARBÓ, R., BESALÚ, E., AMAT LL. i FRADERA X. "Proceedings of the 2nd Girona Seminar on Molecular Similarity", R.Carbo, P.G.Mezey (ed.) *Advances in Molecular Similarity*, Vol. 1, JAI PRESS INC. Greenwich (Conn.), en premsa.
3. CARBÓ, R., BESALÚ, E., AMAT, LL. i FRADERA, X. *J. Math. Chem.* (1995) 18, 237.
4. BESALÚ, E. i CARBÓ, R. *Scientia Gerundensis* (1995) 21, 145
5. WIENER, H., *J. Am. Chem. Soc.* (1947) 69, 17.
6. RANDIC, M., *J. Am. Chem. Soc.* (1975) 97, 6609.
7. ROUVRAY, D.H., *J. Mol. Struct. (Theochem)*, (1989), 185, 187.
8. KIER, L.B. i HALL, L.H. "Molecular Connectivity in Chemistry and drug research", AC Press, Nova York (1976).
9. ROUVRAY, D.H., *J. Comp. Chem.*, (1987), 8, 470.
10. NEEDHAM, E. *et al.*, *J. Am. Chem. Soc.* (1988), 110, 4186.
11. BESALÚ, E. Tesi doctoral. Institut de Química Computacional, Universitat de Girona (1996).
12. DEWAR, M.J., *et al.*, *J. Am. Chem. Soc.*, (1985), 107, 3902.
13. CLEMENTI, E. i RAIMONDI, D.L., *J. Chem. Phys.*, (1963), 38, 2686.
14. CARBÓ, R. i BESALÚ, E., *J. Math. Chem.*, (1995), 18, 37.
15. MONTGOMERY, D.C., PECK, E. "Introduction to linear regression Analysis", 2nd. Ed., John Wiley & Sons, Nova York (1992).
16. LIDE, D.R. (ed.) "Handbook of Chemistry and Physics", 72nd. Edition, Boca Raton (1992).