

A Chemical Index Inspired by Biological Plastic Evolution: Valence-Isoelectronic Series of Aromatics

Francisco Torrens*

Institut Universitari de Ciència Molecular, Universitat de València,
Dr. Moliner 50, E-46100 Burjassot, València, Spain

Received September 23, 2003

Plastic evolution is a new perspective of the evolutionary process conjugating the effect, on one hand, of the acquired characters and, on the other, of the relationships that emerge among the principle of evolutionary indeterminacy, the principle of morphologic determination, and natural selection. Plastic evolution is applied to design the coordination index I_c . I_c is used to characterize the valence-isoelectronic series of cyclopentadiene, benzene, toluene, and styrene and compared to charge indices for dipole moment. The parameters needed to calculate I_c are minus the standard enthalpy of formation T , molecular surface area S , and molecular weight W . With the exception of the O heteromolecule, I_c allows differentiating quantitatively the styrene series as well as the cyclopentadiene from the toluene series. I_c improves the multivariable regression equations for dipole moment and is a predictive index when it is used together with the topological charge indices. The correction introduced in the dipole moment is produced in the correct direction. Inclusion of the heteroatom in the π -electron system is beneficial for the description of the dipole moment, owing to either the role of additional p and/or d orbitals provided by the heteroatom or the role of *steric* factors in the π -electron conjugation. Principal component analysis groups the chemical indices in three classes, which are in agreement with cluster analysis.

1. INTRODUCTION

A mammal's resting metabolism can be predicted reasonably well from body weight. Rubner showed that the standard metabolism of dogs was related to their body weight.¹ Small dogs have a higher metabolic rate per kilogram than large ones, the metabolic rate being approximately proportional to surface area ($2/3$ power of body weight). Kleiber extended this analysis to mammals and birds ranging in size from the ringdove to the steer.² He found that the standard metabolic rate is more closely related to the $3/4$ power of body weight than to the $2/3$ power. His relationship, $M = 3 W^{0.75}$ (M , standard metabolism in $\text{kcal}\cdot\text{h}^{-1}$; W , body weight in kg) was invaluable comparing a multitude of metabolism-dependent parameters of mammals of different weights and served as a model for many anatomical and physiological comparisons.³

Zuntz built a treadmill and measured the energy to move a horizontal meter and to climb 1 m on an inclined treadmill for humans, horses, and dogs.⁴ Slowtsoff made measurements on dogs.⁵ Both concluded that the cost of moving a horizontal meter is approximately proportional to $W^{2/3}$, while the cost of climbing 1 m is nearly the same per kg. Brody,⁶ Wilkie,⁷ and Hemmingsen⁸ concluded that a maximum sustained metabolism is proportional to $W^{0.75}$. Tucker^{9,10} reconsidered the energy cost of travelling 1 km in birds and mammals, concluding that the minimum cost of flying 1 km for birds is proportional to $W^{0.77}$. From data on running mammals, he concluded that the cost of running 1 km is not a straight

line on logarithmic coordinates. It decreases with increasing speed.¹¹ A minimum cost is approached, which is a constant for each animal. Taylor and Raab¹² correlated the weight and the cost of running in mammals.¹³ The maximal metabolic rate induced by exercise scales¹⁴ with $M^{0.86}$. During locomotion, a significant fraction of the electrical activity of a number of limbs muscles occurs while they undergo little or no length change or are being stretched before shortening.¹⁵ For quadrupedal mammals, the amount of energy required to move a unit body weight a unit distance, $E_{\text{trans}}\cdot W^{-1}$, decreases^{16–19} in proportion to body weight, $W^{-0.31}$. Increasing muscle force requirements by 24% by loading resulted in proportional increases in cross-sectional area of muscles.²⁰ Running downhill causes structural damage in the deep slow-twitch extensor muscles of the limbs by a mechanical etiology.²¹ The peak muscle stress is about $1/3$ of the maximum isometric stress at the *preferred speeds*, which is consistent with the idea that these are mechanically equivalent speeds.²² The energy cost per kg per stride at most equivalent speeds is about the same for all animals but increases with speed.²³ The energy cost of locomotion of the African elephant is predicted remarkably well by allometric relationships.²⁴ The cost of muscular force production determines the energy cost of running.^{25,26}

The exponents of allometries evolved from constraints associated with foraging.²⁷ The bivariate relationship between the peak oxygen uptake $\dot{V}_{\text{O}_2\text{ peak}}$ and W in men showed that the W exponent is close to unity.²⁸ In small mammals, an evolutionary increase in daily movement distances was achieved increasing the running speed, without increases in

* Corresponding author e-mail: torrens@bohr.quifis.uv.es.

energy cost.²⁹ An analysis of $M \propto W^\alpha$ showed³⁰ little evidence for rejecting $\alpha = 2/3$ except for mammals with $W \geq 10$ kg. A multiple-causes model of allometry was reported.³¹ Body-size dependence of resting energy cost was attributed to the nonenergetic homogeneity of a fat free mass.³²

In earlier publications, topological charge indices were applied to the molecular dipole moment of hydrocarbons,³³ the valence-isoelectronic series of aromatics,³⁴ and phenyl alcohols.³⁵ A fractal-like index of phenyl alcohols³⁶ and 4-alkylanilines³⁷ was computed. In a previous report, the coordination index I_c was successfully applied to phenyl alcohols and 4-alkylanilines.³⁸ The present article applies I_c to a valence-isoelectronic series of aromatics. The objective of this study is the validation of the usefulness of I_c by means of its capability to distinguish between members of the styrene series as well as its interest as the predictive index for molecular dipole moments as compared to topological charge indices. Section 2 presents the computational method. Section 3 discusses the results. Section 4 summarizes the conclusions.

2. COMPUTATIONAL METHOD

The morphology-functionality relationship in organisms lies in that the former is a material support of the latter, which is the dynamic effect of morphology in the context of a physical environment—living matter interaction. Morphology, functionality, energy cost, and vital viability are four mutually influenced concepts. When a morphology is rather functional, it accomplishes its work with minimal energy cost, and the vital viability of the organ or organism is maximal. Quantifying these concepts involves defining an index called *functional coordination index* I_c . It is formulated as the quotient between the work accomplished by the morphology T and a *morphologic index* I_m representative of it, according to the following expression:

$$I_c = \frac{T}{I_m} \quad (1)$$

The greater the work T accomplished by a concrete morphology I_m , the greater I_c .

For an organism, Ruiz-Bustos proposed I_m as the ratio between its morphologic surface area S and body weight W .³⁹

$$I_m = \frac{S}{W} \quad (2)$$

The substitution of eq 2 in (1) results in

$$I_c = \frac{T}{S/W} = \frac{W \cdot T}{S} \quad (3)$$

At the same time, the expression of T by its equivalence in classical mechanics gives

$$T = \frac{W \cdot x \cdot d^2x}{dt^2} \quad (4)$$

Substituting eq 4 in (3) finally results in

$$I_c = \frac{W^2 \cdot x \cdot d^2x}{S \cdot dt^2} \quad (5)$$

The I_c is greater if the following conditions are fulfilled. (1) The greater the body weight at equal travelled time and space, the greater I_c . (2) The I_c is proportional to the space travelled in the shortest possible time. (3) The smaller the body surface, the greater I_c and the coordination between a function and its morphology requires a smaller energy cost.

Topological charge indices allowed evaluating the dipole moment μ for hydrocarbons.⁴⁰ They are used in this study to correlate the experimental μ . I_c will be used as a complementary index to improve the fit. The connectivity or *adjacency matrix* of the hydrogen-suppressed molecular graph \mathbf{A} has elements A_{ih} , which represent the bonds between the atoms corresponding to vertices i and h . For either $i = h$ or i not bonded to h , i bonded to h by a simple, aromatic, double, and triple bond, the element A_{ih} takes the value 0, 1, 1.5, 2, and 3, respectively. The *Coulombian matrix* \mathbf{Q} has elements $Q_{hj} = 0$ for $h = j$, otherwise $Q_{hj} = 1/D_{hj}^2$, where D_{hj} is the topological *distance* (minimum number of bonds) between vertices h and j . The matrix \mathbf{M} is defined as the following product: $\mathbf{M} = \mathbf{A} \times \mathbf{Q}$. Therefore,

$$M_{ij} = \sum_{h=1}^N A_{ih} Q_{hj} \quad (6)$$

where N is the number of vertices. The charge term between vertices i and j is defined as $C_{ij} = M_{ij} - M_{ji}$. The molecular dipole moment μ is defined as the sum of the edge dipoles $\bar{\mu}_{ij}$:

$$\mu = \sum_{i=1}^{N-1} \sum_{j=i+1}^N \mu_{ij} = \frac{1}{2} \sum_{i=1}^{N-1} \sum_{j=i+1}^N A_{ij} C_{ij}$$

The *vector semisum charge-transfer index* μ_{vec} is defined as the norm of μ :

$$\mu_{\text{vec}} = N(\mu) = (\mu_x^2 + \mu_y^2 + \mu_z^2)^{1/2} \quad (7)$$

The μ_{vec} represents the molecular dipole moment calculated by the vector semisum of \bar{C}_{ij} for all pairs of adjacent vertices in the molecular graph. The *vector semisum valence charge-transfer index* μ_{vec}^V is calculated by the former procedure with a modified matrix \mathbf{A}' , whose main diagonal values describe the electronegativity of heteroatoms.

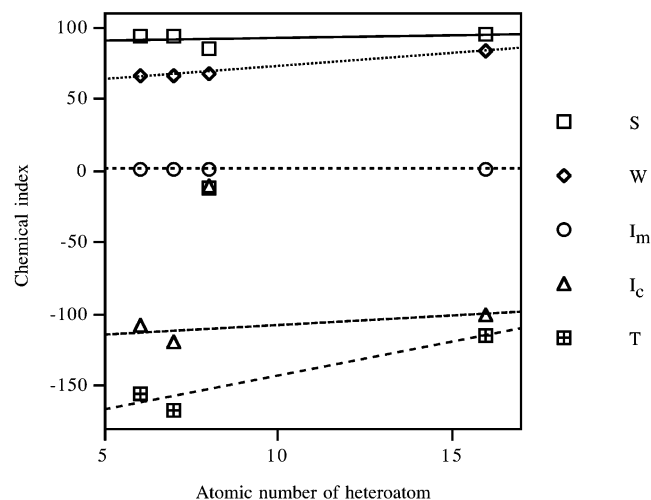
3. CALCULATION RESULTS AND DISCUSSION

The application of the coordination index to the chemical characterization of molecules requires adapting the variables T , S , and W in eq 3. Now, T is defined as minus the standard enthalpy of formation in $\text{kJ} \cdot \text{mol}^{-1}$, S as the molecular surface area in \AA^2 , and W as the molecular weight in $\text{g} \cdot \text{mol}^{-1}$. Table 1 gives the chemical indices for the characterization of the valence-isoelectronic series of cyclopentadiene, benzene, toluene, and styrene. In general, the chemical indices vary with the atomic number Z of the heteroatom. The only exception results in I_m , which is nearly constant and for the four series takes the values 1.21 ± 0.08 , 1.24 ± 0.07 , 1.26 ± 0.14 , and $(1.26 \pm 0.10) \text{ mol} \cdot \text{\AA}^2 \cdot \text{g}^{-1}$, respectively. The average values for the four series as well as for the homologous series of phenyl alcohols and 4-alkylanilines are

Table 1. Chemical Indices for the Characterization of the Valence-Isoelectronic Series of Aromatic Molecules

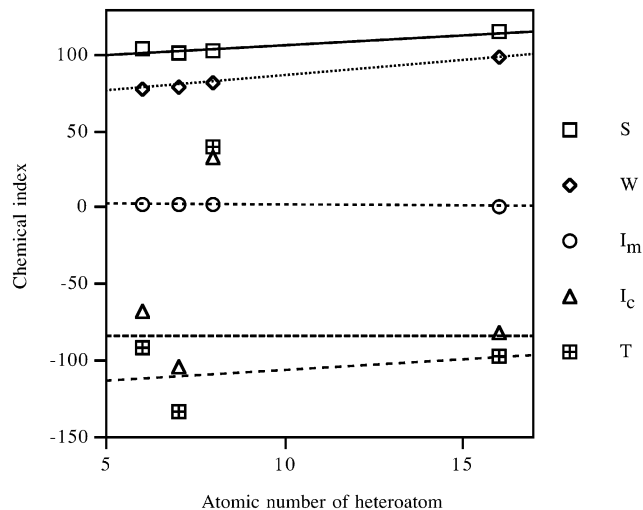
series	molecule	Z^a	W^b	T^c	S^d	I_m^e	I_c^f
cyclopentadiene	cyclopentadiene	6	66	-154.8	94.49	1.432	-108.1
	pyrrole	7	67	-166.6	94.26	1.407	-118.4
	furan	8	68	-12.1	85.32	1.255	-9.7
	thiophene	16	84	-114.6	96.07	1.144	-100.2
	all series	6-16	66-84	-166.6-(-12.1)	85.32-96.07	1.258 ± 0.14	-118.4-(-9.7)
benzene	benzene	6	78	-91.9	104.34	1.338	-68.7
	pyridine	7	79	-133.8	101.80	1.289	-103.8
	4 <i>H</i> -pyrane	8	82	40.1	102.28	1.247	32.2
	4 <i>H</i> -thiopyrane	16	98	-97.0	114.87	1.172	-82.8
	all series	6-16	78-98	-133.8-40.1	101.80-114.87	1.239 ± 0.07	-103.8-32.2
toluene	toluene	6	92	-60.0	124.26	1.351	-44.4
	aniline	7	93	-89.6	127.38	1.370	-65.4
	phenol	8	94	93.4	120.96	1.287	72.6
	thiophenol	16	110	-107.4	128.06	1.164	-92.2
	all series	6-16	92-110	-107.4-93.4	120.96-128.06	1.264 ± 0.10	-92.2-72.6
styrene	styrene	6	104	-161.4	137.61	1.323	-122.0
	benzylideneamine	7	105	-167.9	135.68	1.292	-130.0
	benzaldehyde	8	106	37.7	126.97	1.198	31.5
	C ₆ H ₅ -CH=SiH ₂	14	120	-184.0	147.39	1.228	-149.8
	C ₆ H ₅ -CH=PH	15	122	-221.0	141.12	1.157	-191.0
	thiobenzaldehyde	16	122	-213.0	139.69	1.145	-186.1
	C ₆ H ₅ -CH=GeH ₂	32	165	-301.6	143.30	0.868	-347.3
	C ₆ H ₅ -CH=AsH	33	166	-272.9	141.15	0.850	-321.0
	C ₆ H ₅ -CH=Se	34	169	-278.6	141.03	0.834	-333.9
	C ₆ H ₅ -CH=SnH ₂	50	211	-367.4	154.99	0.735	-500.1
	C ₆ H ₅ -CH=SbH	51	213	-341.9	142.10	0.667	-512.5
	C ₆ H ₅ -CH=Te	52	218	-273.9	150.06	0.688	-397.9
	C ₆ H ₅ -CH=PbH ₂	82	299	-441.6	155.57	0.520	-848.8
	C ₆ H ₅ -CH=BiH	83	300	-394.6	146.43	0.488	-808.5
	C ₆ H ₅ -CH=Po	84	300		137.78	0.459	
phenyl alcohols ^g	partial series	6-16	104-122	-213.0-37.7	126.97-139.69	1.207 ± 0.08	-186.1-31.5
	all series	6-84	104-300	-441.6-37.7	126.97-155.57	0.843 ± 0.4	-848.8-31.5
4-alkylanilines ^g	all series		94-192	93.4-297.0	128.07-274.30	1.385 ± 0.05	68.6-207.9
	all series		93-177	-89.6-81.1	127.17-247.75	1.386 ± 0.015	-65.5-58.0

^a Z : atomic number of heteroatom. ^b W : molecular weight ($\text{g}\cdot\text{mol}^{-1}$). ^c T : minus standard enthalpy of formation ($\text{kJ}\cdot\text{mol}^{-1}$). ^d S : molecular surface area (\AA^2). ^e I_m : morphologic index ($\text{mol}\cdot\text{\AA}^2\cdot\text{g}^{-1}$). ^f I_c : coordination index ($\text{kJ}\cdot\text{g}\cdot\text{mol}^{-2}\cdot\text{\AA}^{-2}$). ^g Homologous series.

**Figure 1.** Variation of indices for the valence-isoelectronic series of cyclopentadiene vs atomic number of heteroatom.

also listed. Notice that I_m ranges overlap for most of the series. The corresponding interpretation is that the series of molecules with overlapping I_m ranges have a similar form. The only series than can be distinguished quantitatively by I_m is styrene (or benzene) from phenyl alcohols (or 4-alkylanilines).

The variation of the indices for the cyclopentadiene series with Z is illustrated in Figure 1. The descriptors more

**Figure 2.** Variation of indices for the valence-isoelectronic series of benzene vs atomic number of heteroatom.

sensitive to Z are in the order $T > W \sim I_c > S > I_m$. Linear correlations have been observed between I_m , S , W , and Z as well as nearly linear correlations between I_c , T , and Z (only the $Z = 8$ O heteromolecule is separated from the fitting line).

The change of the indices for the benzene series vs Z is shown in Figure 2. The most sensitive descriptors are

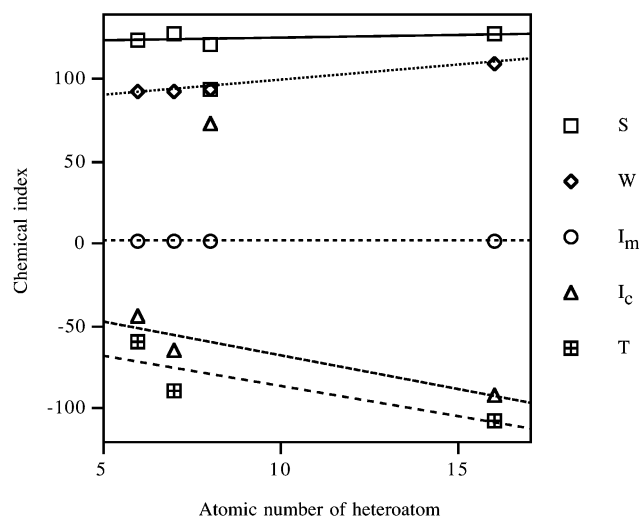


Figure 3. Variation of indices for the valence-isoelectronic series of toluene vs atomic number of heteroatom.

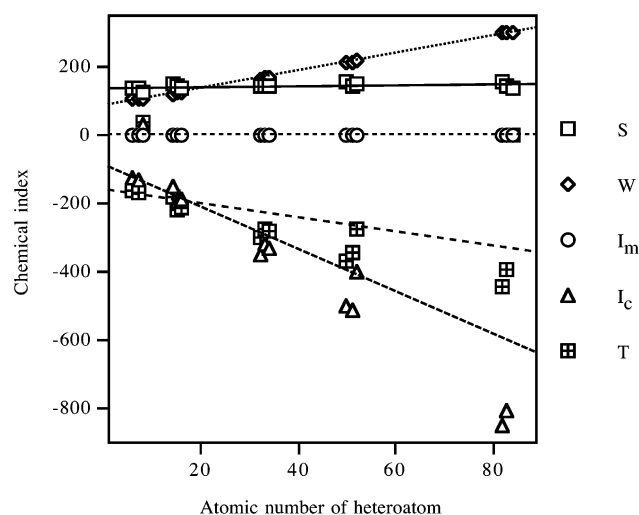


Figure 4. Variation of indices for the valence-isoelectronic series of styrene vs atomic number of heteroatom.

$W \sim T \sim S > I_c > I_m$. The I_c , I_m , T , S , and W follow linear or nearly linear correlations with Z .

The variation of the indices for the toluene series vs Z is displayed in Figure 3. The most sensitive descriptors are $I_c > T > W > S > I_m$. The I_c , I_m , T , S , and W present linear or nearly linear correlations with Z .

The change of the indices for the styrene series vs Z is presented in Figure 4. The most sensitive descriptors are $I_c > T > W > S > I_m$. The I_c allows differentiating quantitatively the styrene (cyclopentadiene, benzene, or 4-alkylaniline) from the phenyl alcohol series. The I_c , I_m , T , S , and W reveal linear or nearly linear correlations with Z . On average for the four series the most sensitive descriptors are $I_c > T > W > S > I_m$. If the O heteromolecules are removed, I_c permits differentiating quantitatively the styrene and phenyl alcohol series from the others as well as the cyclopentadiene from the toluene and the cyclopentadiene (or benzene) from the 4-alkylaniline series.

The experimental dipole moment for the C, N, O, and S heteromolecules in the four series has been correlated as a

function of the chemical indices. The best linear fit results

$$\mu = 6.07 - 3.74I_m \quad \text{MAPE} = 68.54\% \quad \text{AEV} = 0.9000 \quad (8)$$

where MAPE is the mean absolute percentage error and AEV is the approximation error variance. All other models with greater MAPE and AEV have been discarded. The best quadratic model results

$$\mu = 3.69 - 1.47I_m^2 \quad \text{MAPE} = 68.46\% \quad \text{AEV} = 0.8996 \quad (9)$$

and AEV decreases 0.04%. The inclusion of I_m or I_c does not improve the fit.

The variance can be decreased if the fit is limited to fewer molecules. In particular if only the C, N, and S heteromolecules are considered the best linear fit results

$$\mu = -1.18 + 1.05\mu_{\text{vec}}^V - 0.0162I_c \quad \text{MAPE} = 39.54\% \quad \text{AEV} = 0.3551 \quad (10)$$

and AEV decreases 61%. The best quadratic model results

$$\mu = -0.479 + 1.14\mu_{\text{vec}}^V + 0.0000759I_c^2 \quad \text{MAPE} = 35.62\% \quad \text{AEV} = 0.2799 \quad (11)$$

and AEV diminishes 69%. If I_c/Z is included in the model the best linear regression results

$$\mu = -0.745 - 0.595\mu_{\text{vec}}^V + 1.27\mu_{\text{vec}}^V - 0.0195I_c + 0.0560I_c/Z \quad (12)$$

$$n = 12 \quad r = 0.871 \quad s = 0.594 \quad F = 5.5 \quad \text{MAPE} = 33.43\% \quad \text{AEV} = 0.2489$$

and AEV drops 72%. The best quadratic model does not improve the correlation. The inclusion of I_m does not improve the fit. Although charge indices are, in principle, more adequate to predict μ , these equations (specially 8–10), which show rather poor correlation, if any, are kept with the only purpose to compare with equations below in which morphologic indices improve charge indices.

Even, the variance can be decreased if the fit is limited to fewer series. For example for the first 14 molecules in the styrene series the best linear fit results

$$\mu = 3.16 + 0.103I_c/Z \quad \text{MAPE} = 42.96\% \quad \text{AEV} = 0.7783 \quad (13)$$

and AEV decreases 14%. The best quadratic model does not improve the correlation. If only the first 12 molecules are considered the best linear regression results

$$\mu = 3.19 + 0.105I_c/Z \quad \text{MAPE} = 43.95\% \quad \text{AEV} = 0.7601 \quad (14)$$

and AEV decreases 16%. The best quadratic model does not improve the fit. If only the first 9 molecules are included the best linear regression results

$$\mu = 17.1 - 20.3\mu_{\text{vec}} - 6.75I_m \quad \text{MAPE} = 40.86\% \quad \text{AEV} = 0.6526 \quad (15)$$

Table 2. Cross-Validation Correlation Coefficient in a Leave-*n*-Out Procedure for Styrene Series

<i>n</i>	μ vs I_m (8)	μ vs μ_{vec}, I_c (10)	μ vs $\mu_{\text{vec}}, \mu_{\text{vec}}^V, I_c, I_c/Z$ (12)	μ vs I_c/Z (13)	μ vs I_c/Z (14)	μ vs μ_{vec}, I_m (15)	μ vs μ_{vec}, I_m (17)
1	0.063	0.514	0.584	0.170	0.174	0.005	0.766
2	0.062	0.514	0.573	0.146	0.147	0.030	0.769
3	0.061	0.513	0.566	0.116	0.111	0.063	0.754
4	0.060	0.513		0.075	0.060	0.074	
5	0.059	0.512		0.016	0.015	0.010	
6	0.057	0.512		0.073	0.129	0.669	
7	0.055	0.515		0.214	0.263		
8	0.052	0.536		0.432	0.431		
9	0.049	0.442		0.753	0.973		
10	0.044			0.891	0.346		
11	0.038			0.606			
12	0.034			0.516			
13	0.097						
14	0.386						

and AEV decreases 27%. The best quadratic model results

$$\mu = 1.96 + 1.69z_{11}z_{12}$$

$$z_{11} = 1.67 + 0.00861I_c$$

$$z_{12} = 4.55 - 11.1\mu_{\text{vec}}I_m \quad \text{MAPE} = 42.22\% \quad \text{AEV} = 0.6280 \quad (16)$$

and AEV diminishes 30%. If only the first 6 molecules are incorporated the best linear regression results

$$\mu = 37.6 - 25.1\mu_{\text{vec}} - 22.0I_m \quad (17)$$

$$n = 6 \quad r = 0.935 \quad s = 0.710 \quad F = 10.4 \quad \text{MAPE} = 21.75\% \quad \text{AEV} = 0.1257$$

and AEV drops 86%. The best quadratic model results

$$\mu = 1.98 - 0.00743z_{31} + 1.42z_{32}$$

$$z_{31} = -1.55 + 0.584I_c/Z + 0.00581I_c(I_c/Z) - 0.000119I_c^2$$

$$z_{32} = 0.976z_{22} + 0.0548z_{21}^2 - 0.0341z_{22}^2$$

$$z_{21} = 0.326 + 0.0277I_c + 0.000149I_c^2$$

$$z_{22} = 1.39z_{11} - 0.519z_{12} - 0.68I_{z_{11}z_{12}} + 0.714z_{12}^2$$

$$z_{11} = 213 - 348\mu_{\text{vec}} - 221I_m + 460\mu_{\text{vec}}^2 + 81.4I_m^2$$

$$z_{12} = -1.55 + 0.584I_c/Z + 0.00581I_c(I_c/Z) - 0.000119I_c^2 \quad (18)$$

$$\text{MAPE} = 0.14\% \quad \text{AEV} < 0.0001$$

and AEV decreases 100%. Due to the complexity of eq 18, a Fortran program has been written to calculate μ . The program is available from the author via the Internet (Francisco.Torrens@uv.es).

The correlation coefficient found between the *cross-validated* representatives and the property values R_{cv} has been calculated with the leave-*n*-out procedure.⁴¹ The procedure furnishes a new method for selecting the best set of descriptors according to the criterion of maximization of the value of R_{cv} . The R_{cv} calculations for the styrene series are

given in Table 2 for $1 \leq n \leq 14$. In general for smaller n , R_{cv} decreases with n . In particular, the method μ vs $\{\mu_{\text{vec}}, I_m\}$ (eq 17) gives the greatest R_{cv} . The corresponding interpretation is that $\{\mu_{\text{vec}}, I_m\}$ is the most predictive set of descriptors for modeling μ .

The principal component analysis (PCA)⁴² for the chemical indices of the C, N, O, and S heteromolecules results in eight factors F_1 – F_8 , which are linear combinations of $\{W, T, S, I_m, I_c, \mu_{\text{experiment}}, \mu_{\text{vec}}, \mu_{\text{vec}}^V\}$. In particular, for the styrene valence-isoelectronic series, the coefficients for factor F_1 are

$$F_1 = -0.275W + 0.433T - 0.446S + 0.021I_m + 0.440I_c + 0.214\mu_{\text{experiment}} - 0.334\mu_{\text{vec}} + 0.432\mu_{\text{vec}}^V \quad (19)$$

The coefficients for factor F_2 are

$$F_2 = 0.474W + 0.098T - 0.088S - 0.597I_m + 0.036I_c + 0.483\mu_{\text{experiment}} + 0.377\mu_{\text{vec}} + 0.158\mu_{\text{vec}}^V \quad (20)$$

The use of only the first factor F_1 explains 61% of the variance and gives a relative error of 39%. Moreover, the combined use of the first two factors, F_1 and F_2 , explains 96% of the variance reducing the relative error to 4%. The profile of factor F_1 is $\{7.56, 18.77, 19.86, 0.04, 19.34, 4.59, 11.16, 18.68\}\%$. Therefore, F_1 cannot be reduced to four variables $\{T, S, I_c, \mu_{\text{vec}}^V\}$ without making a relative error of 23% (the sum of the percentages for the four remaining variables). On the other hand, the profile of factor F_2 is $\{22.50, 0.95, 0.77, 35.60, 0.13, 23.34, 14.22, 2.48\}\%$. For factor F_2 the most important variables are $\{W, I_m, \mu_{\text{experiment}}\}$. In some way, factor F_1 could be considered as a linear combination of $\{T, S, I_c, \mu_{\text{vec}}, \mu_{\text{vec}}^V\}$ with a relative error of 12%. However, factor F_2 can be expressed as a linear combination of $\{W, I_m, \mu_{\text{experiment}}, \mu_{\text{vec}}\}$ with a relative error of 4%.

PCA F_2 vs F_1 plot for the chemical indices of the four valence-isoelectronic series is exhibited in Figure 5. Chemical indices $\{I_m, \mu_{\text{experiment}}, \mu_{\text{vec}}, \mu_{\text{vec}}^V\}$ appear superposed. Three classes of chemical indices are clearly distinguished: class 1 $\{W, S\}$ (under the bisector, $F_1 > F_2 > 0$, right of Figure 5), class 2 $\{T, I_c\}$ (above the bisector, $F_1 < F_2$, left), and class 3 $\{I_m, \mu_{\text{experiment}}, \mu_{\text{vec}}, \mu_{\text{vec}}^V\}$ (below the bisector, $0 \approx F_1 > F_2$, bottom).

The cluster analysis (CA)⁴³ radial tree for $\{W, T, S, I_m, I_c, \mu_{\text{experiment}}, \mu_{\text{vec}}, \mu_{\text{vec}}^V\}$ of the C, N, O, and S hetero-

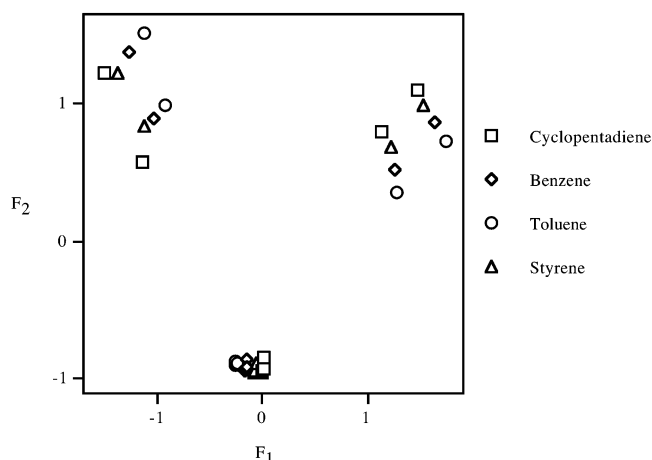


Figure 5. Principal component analysis F_2 vs F_1 plot for the chemical indices of valence-isoelectronic series.

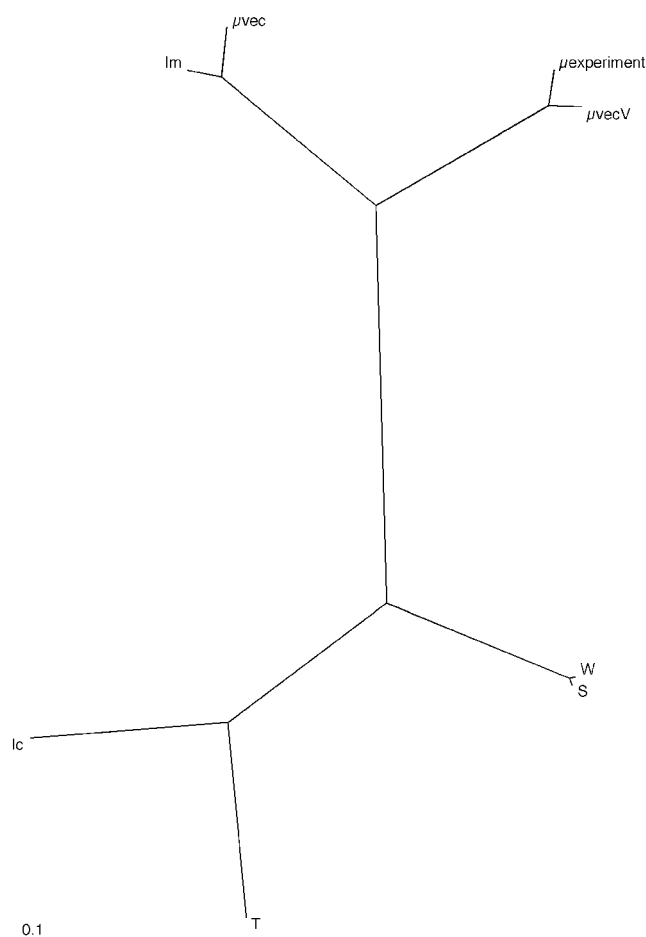


Figure 6. Radial tree graph for the chemical indices of valence-isoelectronic series.

molecules in the valence-isoelectronic series separates first both chemical indices in class 1 $\{W, S\}$ (Figure 6 right), then both indices in class 2 $\{T, I_c\}$ (bottom), and finally the 4 indices in class 3 $\{I_m, \mu_{\text{experiment}}, \mu_{\text{vec}}, \mu_{\text{vec}}^V\}$ (top). The classes correspond to PCA (Figure 5).

4. CONCLUSIONS

Plastic evolution allows the design of the new chemical indices I_m and I_c . They are applied to the valence-isoelectronic series of aromatics and compared to the charge indices

for the dipole moment. From the preceding results the following conclusions can be drawn.

1. The parameters needed to calculate the index I_c are the molecular weight W , minus standard enthalpy of formation T , and the molecular surface area S .

2. With the exception of the O heteromolecule, I_c allows differentiating quantitatively the styrene series as well as the cyclopentadiene from the toluene series.

3. Linear and quadratic correlation models have been obtained for μ . The I_m and I_c ³⁹ have improved the multivariable regression equations for μ . The I_c index is shown a predictive index for μ when it is used together with the topological charge indices. The cross-validation leave- n -out procedure shows that $\{\mu_{\text{vec}}, I_m\}$ is the most predictive set of descriptors for μ (eq 17) according to the criterion of maximization of R_{cv} .

4. The behavior of μ_{vec}^V is intermediate between μ_{vec} and $\mu_{\text{experiment}}$, and so the correction introduced with respect to μ_{vec} is produced in the correct direction.

5. Inclusion of the heteroatom in the π -electron system is beneficial for the description of the dipole moment, owing to either the role of additional p and/or d orbitals provided by the heteroatom or the role of steric factors in the π -electron conjugation. The analysis of both electronic and steric factors in μ caused by the presence of the heteroatom shows that the electronic factor dominates over the steric one.

6. PCA allows grouping the chemical indices in three classes: $\{W, S\}$, $\{T, I_c\}$, and $\{I_m, \mu_{\text{experiment}}, \mu_{\text{vec}}, \mu_{\text{vec}}^V\}$. Therefore, I_m is closer to μ than I_c is. The similarity between chemical indices has been compared with the CA of these indices. CA is in agreement with PCA.

ACKNOWLEDGMENT

I wish to thank Dr. E. Besalú for providing me with several versions of his full linear leave-many-out program before publication. The author acknowledges financial support from the Spanish MCT (Plan Nacional I+D+I, Project No. BQU2001-2935-C02-01).

REFERENCES AND NOTES

- (1) Rubner, M. Ueber den Einfluss der Koerpergroesse auf Stoff- und Kraft-wechsel. *Z. Biol.* **1883**, *19*, 535–562.
- (2) Kleiber, M. Body Size and Metabolism. *Hilgardia* **1932**, *6*, 315–353.
- (3) Kleiber, M. *The Fire of Life*; Wiley: New York, 1961.
- (4) Zuntz, N. Ueber den Stoffverbrauch des Hundes bei Muskelarbeit. *Arch. Ges. Physiol.* **1897**, *68*, 191–211.
- (5) Slowtsoff, B. *Arch. Ges. Physiol.* **1903**, *95*, 158.
- (6) Brody, S. *Bioenergetics and Growth*; Reinhold: New York, 1945.
- (7) Wilkie, D. R. The Work Output of Animals: Flight by Birds and by Manpower. *Nature (London)* **1959**, *183*, 1515–1517.
- (8) Hemmingsen, A. M. Energy Metabolism as Related to Body Size and Respiratory Surfaces, and its Evolution. *Rep. Steno Mem. Hosp. Nord. Insulin Lab.* **1960**, *9*(2), 7–110.
- (9) Tucker, V. A. The Energetics of Bird Flight. *Sci. Am.* **1969**, *220*(5), 70–78.
- (10) Tucker, V. A. Energetic Cost of Locomotion in Animals. *Comput. Biochem. Physiol.* **1970**, *34*, 841–846.
- (11) Taylor, C. R.; Schmidt-Nielsen, K. *Physiologist* **1969**, *12*, 372.
- (12) Taylor, C. R.; Raab, J. L. *Federation Proc.* **1970**, *29*, 265.
- (13) Taylor, C. R.; Schmidt-Nielsen, K.; Raab, J. L. Scaling of Energetic Cost of Running to Body Size in Mammals. *Am. J. Physiol.* **1970**, *219*, 1104–1107.
- (14) Taylor, C. R.; Maloiy, G. M. O.; Weibel, E. R.; Langman, V. A.; Kamau, J. M. Z.; Seeherman, H. J.; Heglund, N. C. Design of the Mammalian Respiratory System. III. Scaling Maximum Aerobic

- Capacity to Body Mass: Wild and Domestic Mammals. *Respir. Physiol.* **1981**, *44*, 25–37.
- (15) Goslow, G. E.; Seeherman, H. J.; Taylor, C. R.; McCutchin, M. N.; Heglund, N. C. Electrical Activity and Relative Length Changes of Dog Limb Muscles as a Function of Speed and Gait. *J. Exp. Biol.* **1981**, *94*, 15–42.
 - (16) Taylor, C. R.; Heglund, N. C.; Maloiy, G. M. O. Energetics and Mechanics of Terrestrial Locomotion. I. Metabolic Energy Consumption as a Function of Speed and Body Size in Birds and Mammals. *J. Exp. Biol.* **1982**, *97*, 1–21.
 - (17) Fedak, M. A.; Heglund, N. C.; Taylor, C. R. Energetics and Mechanics of Terrestrial Locomotion. II. Kinetic Energy Changes of the Limbs and Body as a Function of Speed and Body Size in Birds and Mammals. *J. Exp. Biol.* **1982**, *97*, 23–40.
 - (18) Heglund, N. C.; Cavagna, G. A.; Taylor, C. R. Energetics and Mechanics of Terrestrial Locomotion. III. Energy Changes of the Centre of Mass as a Function of Speed and Body Size in Birds and Mammals. *J. Exp. Biol.* **1982**, *97*, 41–56.
 - (19) Heglund, N. C.; Fedak, M. A.; Taylor, C. R.; Cavagna, G. A. Energetics and Mechanics of Terrestrial Locomotion. IV. Total Mechanical Energy Changes as a Function of Speed and Body Size in Birds and Mammals. *J. Exp. Biol.* **1982**, *97*, 57–66.
 - (20) Armstrong, R. B.; Taylor, C. R. Relationship Between Muscle Force and Muscle Area Showing Glycogen Loss During Locomotion. *J. Exp. Biol.* **1982**, *97*, 411–420.
 - (21) Armstrong, R. B.; Taylor, C. R. Glycogen Loss in Rat Muscles During Locomotion on Different Inclines. *J. Exp. Biol.* **1993**, *176*, 135–144.
 - (22) Perry, A. K.; Blickhan, R.; Biewener, A. A.; Heglund, N. C.; Taylor, C. R. Preferred Speeds in Terrestrial Vertebrates: Are They Equivalent? *J. Exp. Biol.* **1988**, *137*, 207–219.
 - (23) Heglund, N. C.; Taylor, C. R. Speed, Stride Frequency and Energy Cost per Stride: How Do They Change with Body Size and Gait? *J. Exp. Biol.* **1988**, *138*, 301–318.
 - (24) Langman, V. A.; Roberts, T. J.; Black, J.; Maloiy, G. M. O.; Heglund, N. C.; Weber, J.-M.; Kram, R.; Taylor, C. R. Moving Cheaply: Energetics of Walking in the African Elephant. *J. Exp. Biol.* **1995**, *198*, 629–632.
 - (25) Roberts, T. J.; Kram, R.; Weyand, P. G.; Taylor, C. R. Energetics of Bipedal Running I. Metabolic Cost of Generating Force. *J. Exp. Biol.* **1998**, *201*, 2745–2751.
 - (26) Roberts, T. J.; Chen, M. S.; Taylor, C. R. Energetics of Bipedal Running II. Limb Design and Running Mechanics. *J. Exp. Biol.* **1998**, *201*, 2753–2762.
 - (27) Witting, L. The Body Mass Allometries as Evolutionarily Determined by the Foraging of Mobile Organisms. *J. Theor. Biol.* **1995**, *177*, 129–137.
 - (28) Batterham, A. M.; Vanderburgh, P. M.; Mahar, M. T.; Jackson, A. S. Modeling the Influence of Body Size on $\dot{V}_{O2\text{ peak}}$: Effects of Model Choice and Body Composition. *J. Appl. Physiol.* **1999**, *87*, 1317–1325.
 - (29) Koteja, P.; Swallow, J. G.; Carter, P. A.; Garland, T., Jr. Energy Cost of Wheel Running in House Mice: Implications for Coadaptation of Locomotion and Energy Budgets. *Physiol. Biomed. Zool.* **1999**, *72*, 238–249.
 - (30) Dodds, P. S.; Rothman, D. H.; Weitz, J. S. Re-examination of the 3/4-Law of Metabolism. *J. Theor. Biol.* **2001**, *209*, 9–27.
 - (31) Darveau, C.-A.; Suarez, R. K.; Andrews, R. D.; Hochachka, P. W. Allometric Cascade as a Unifying Principle of Body Mass Effects on Metabolism. *Nature (London)* **2002**, *417*, 166–170.
 - (32) Heymsfield, S. B.; Gallagher, D.; Kotler, D. P.; Wang, Z.; Allison, D. B.; Heshka, S. Body-Size Dependence of Resting Energy Expenditure Can Be Attributed to Nonenergetic Homogeneity of Fat-Free Mass. *Am. J. Physiol. Endocrinol. Metab.* **2002**, *282*, E132–E138.
 - (33) Torrens, F. A New Topological Index to Elucidate Apolar Hydrocarbons. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 709–719.
 - (34) Torrens, F. Valence Topological Charge-Transfer Indices for Dipole Moments. *J. Mol. Struct. (THEOCHEM)* **2003**, *621*, 37–42.
 - (35) Torrens, F. Valence Topological Charge-Transfer Indices for Dipole Moments. *Molecules* **2003**, *8*, 169–185.
 - (36) Torrens, F. Fractal Dimension of Transdermal-Delivery Drug Models. *J. Sci. Libanais* (submitted for publication).
 - (37) Torrens, F. Fractal Dimension of Transdermal-Delivery Drug Models: 4-Alkylanilines. *Physica A* Submitted for publication.
 - (38) Torrens, F. A New Chemical Index Inspired by Biological Plastic Evolution. *Indian J. Chem., Sect. A* **2003**, *42*, 1258–1263.
 - (39) Ruíz-Bustos, A. *La Evolución Plástica*; Andalucía: Granada, 1994, Chapter 8.
 - (40) Gálvez, J.; García, R.; Salabert, M. T.; Soler, R. Charge Indexes. New Topological Descriptors. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 520–525.
 - (41) Besalú, E. Fast Computation of Cross-Validated Properties in Full Linear Leave-Many-Out Procedures. *J. Math. Chem.* **2001**, *29*, 191–203.
 - (42) Hotelling, H. Analysis of a Complex of Statistical Variables into Principal Components. *J. Educ. Psychol.* **1933**, *24*, 417–441.
 - (43) Tryon, R. C. *J. Chronic Dis.* **1939**, *20*, 511–524.

CI034213U