A new chemical index inspired by biological plastic evolution

Francisco Torrens
Institut Universitari de Ciência Molecular, Universitat de València, Dr. Moliner 50, E-46100 Burjassot (València), Spain
Received 10 January 2003

This paper presents an application of plastic evolution to design a new chemical index (the coordination index, \( I_e \)) that permits molecular characterization. \( I_e \) is applied to homologous series of phenyl alcohols and 4-alkylanilines, and compared to charge indices for dipole moment. The parameters needed to calculate \( I_e \) are minus the standard enthalpy of formation \( T \), molecular surface area \( S \) and molecular weight \( W \). The correlations between \( I_e, T, S, W \) and the number of C atoms in the alkyl chain indicate a homogeneous molecular structure of the phenyl alcohols and 4-alkylanilines.

Introduction

In earlier publications, molecular topological charge-transfer indices were applied to the calculation of the molecular dipole moment of hydrocarbons, valence-isoelectronic series of benzene and styrene, and homologous series of phenyl alcohols and 4-alkylanilines. The fractal dimension of phenyl alcohols and 4-alkylanilines was computed. This work presents an application of plastic evolution to design a new chemical index (the co-ordination index \( I_e \)) that permits molecular characterization. Index \( I_e \) is applied to phenyl alcohols and 4-alkylanilines (percutaneous enhancers of transdermal-delivery drugs).

Computational method

The morphology–functionality relationship in organisms lies in that the former is material support of the latter and in that this is the dynamic effect of morphology, all this immersed in the context defined by the physical environment–living matter interaction. Morphology, functionality, energetic cost and vital viability are four mutually influenced concepts. Therefore, when a particular morphology is rather functional, it accomplishes its work with minimal energetic cost and the vital viability of the organ or organism is maximal. Quantifying these concepts involves defining an index called functional coordination index \( I_e \). 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 expression:

\[
I_e = \frac{T}{I_m}
\]

In line with this, the greater the work \( T \) accomplished by a concrete morphology, expressed through its index \( I_m \), the greater the functional coordination index \( I_e \), between both.

As morphologic index \( I_m \) for an organism, Ruiz-Bustos proposed the ratio between its morphologic surface area \( S \) and body weight \( W \), both parameters with important role in the physiological processes of the morphology:

\[
I_m = \frac{S}{W}
\]

The substitution of Eq. (2) in (1) results:

\[
I_e = \frac{T}{S/W} = \frac{W \cdot T}{S}
\]

At the same time, the expression of \( T \) by its equivalence in classical mechanics gives:

\[
T = \frac{W \cdot x \cdot d^2 x}{dt^2}
\]

Substituting Eq. (4) in (3) finally results in:

\[
I_e = \frac{W^2 \cdot x \cdot d^2 x}{S \cdot dt^2}
\]

In accordance with Eq. (5), the coordination index \( I_e \), between morphology and function is greater if the following conditions are fulfilled: (i) The greater the body weight at equal travelled time and space, the greater the functional coordination index \( I_e \); (ii) the coordination index is proportional to the space travelled in the shortest possible time; and (iii) As the
body surface is smaller, at the same remaining conditions, \( L \) is greater and the coordination between a function and its morphology requires smaller energetic cost.

Topological charge indices showed a good ability to evaluate the charge transfer between pairs of atoms and the global charge transfer, as demonstrated by the good correlation achieved with the dipole moment for a set of heterogeneous hydrocarbons. For a given graph, they are defined as follows:

\[
G_k = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} [C_{ij}]^2 \delta(k_{ij}) \\
J_k = \frac{G_k}{N-1} 
\]

where \( N \) is the number of vertices in the hydrogen-suppressed molecular graph, and \( C_{ij} \) is the charge term between vertices \( i \) and \( j \), defined as \( C_{ij} = M_{ij} - M_{ji} \), with \( M_{ij} \) and \( M_{ji} \) being elements of the \( N \times N \) matrix \( M \), obtained as the product of two matrices: \( M = A \cdot Q \). Therefore,

\[
M_{ij} = \sum_{h=1}^{N} A_{ih} Q_{hj} 
\]

The matrix \( A \) is called the connectivity or adjacency matrix; its elements \( A_{ij} \) represent the bonds between the atoms corresponding to vertices \( i \) and \( h \). The element \( A_{ij} \) takes the value “0” if either \( i = h \) or \( i \) is not bonded to \( h \); it takes the value “1” if \( i \) is bonded to \( h \) by a single bond, 1.5 if the bond is aromatic, “2” for double bonds and “3” for triple bonds. The matrix \( Q \) is known as the Coulombian matrix; its element \( Q_{hj} = 0 \) for \( h = j \), otherwise \( Q_{hj} = \sqrt{D_{ij}} \), where \( D_{ij} \) is the minimum number of bonds (topological distance) between vertices \( h \) and \( j \). In Equation (6), \( \delta \) represents the Kronecker delta function: \( \delta(\alpha, \beta) = 1 \) if \( \alpha = \beta \), and \( \delta(\alpha, \beta) = 0 \) if \( \alpha \neq \beta \), and \( D_{ij} \) is the topological distance between vertices \( i \) and \( j \). The \( G_k \) represent the overall sum of the absolute values of the charge terms \( C_{ij} \) for every pair of vertices \( i \) and \( j \) situated at a topological distance \( k \). The valence topological charge indices \( G_k \) and \( V_k \) are calculated following the former procedure with a modified matrix \( A^* \), whose main diagonal values describe the electronegativity of the heteroatoms.

The algebraic semisum charge-transfer index \( \mu_{alg} \) is defined as

\[
\mu_{alg} = \frac{1}{2} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} A_{ij} C_{ij} 
\]

The \( \mu_{alg} \) represent the molecular dipole moment calculated by algebraic semisum of \( C_{ij} \) for all pairs of adjacent vertices in the molecular graph. An edge-to-edge analysis of \( \mu \) suggests that each edge dipole moment \( \mu_{ij} \) connecting vertices \( i \) and \( j \) can be evaluated as

\[
\mu_{ij} = \frac{1}{2} A_{ij} C_{ij} 
\]

Each edge dipole can be associated with a vector \( \mu_{ij} \) in space lying in the edge connecting vertices \( i \) and \( j \), with direction from \( j \) to \( i \). The molecular dipole moment vector \( \tilde{\mu} \) results the vector sum of the edge dipole moments as

\[
\tilde{\mu} = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \tilde{\mu}_{ij} = \frac{1}{2} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} A_{ij} \tilde{C}_{ij} 
\]

The vector semisum charge-transfer index \( \mu_{vec} \) is defined as the norm of \( \tilde{\mu} \):

\[
\mu_{vec} = N(\tilde{\mu}) = \left( \mu_x^2 + \mu_y^2 + \mu_z^2 \right)^{1/2} 
\]

The \( \mu_{vec} \) represent the molecular dipole moment calculated by vector semisum of \( \tilde{C}_{ij} \) for all pairs of adjacent vertices in the molecular graph.

Principal component analysis (PCA) is used to transform a number of potentially correlated variables into the same number of independent variables that then can be ranked based upon their contributions for explaining the whole data set. The transformed variables that can explain all the information in the data are called principal components (PCs). The first PC accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. The components having minor contribution to the data set may be discarded without losing too much information. If the number of PCs is less than four, then the multidimensional data can be graphed in two- or three-dimensional space.

Results and Discussion

The application of the coordination index to the chemical characterization of molecules requires...
Table 1 — Chemical indices for the characterization of a homologous series of phenyl alcohols.

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Number of C atoms in alkyl chain (n)</th>
<th>W (g·mol⁻¹)</th>
<th>T (J·mol⁻¹)</th>
<th>S (Å²)</th>
<th>Iₙ (mol·Å²·g⁻¹)</th>
<th>I (J·g⁻¹·mol⁻²·Å⁻³)</th>
<th>J/ₙ (J·g⁻¹·mol⁻²·Å⁻³)</th>
<th>μₑ₀ (D)</th>
<th>μₑ₀ (D)</th>
<th>μ₀μₑ₀ (D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>phenol</td>
<td>0</td>
<td>94</td>
<td>93.4</td>
<td>128.07</td>
<td>1.362</td>
<td>68.6</td>
<td>—</td>
<td>1.400</td>
<td>0.737</td>
<td>2.431</td>
</tr>
<tr>
<td>1-Benzyl alcohol</td>
<td>1</td>
<td>108</td>
<td>130.8</td>
<td>142.33</td>
<td>1.318</td>
<td>99.3</td>
<td>99.28</td>
<td>1.700</td>
<td>0.589</td>
<td>2.348</td>
</tr>
<tr>
<td>2-Phenyl-1-ethanol</td>
<td>2</td>
<td>122</td>
<td>153.7</td>
<td>166.60</td>
<td>1.366</td>
<td>112.6</td>
<td>56.29</td>
<td>1.590</td>
<td>0.700</td>
<td>2.257</td>
</tr>
<tr>
<td>3-Phenyl-1-propanol</td>
<td>3</td>
<td>136</td>
<td>188.7</td>
<td>188.18</td>
<td>1.384</td>
<td>136.3</td>
<td>45.45</td>
<td>1.640</td>
<td>0.573</td>
<td>2.519</td>
</tr>
<tr>
<td>4-Phenyl-1-butanol</td>
<td>4</td>
<td>150</td>
<td>214.9</td>
<td>209.74</td>
<td>1.398</td>
<td>153.7</td>
<td>38.42</td>
<td>1.345b</td>
<td>0.702</td>
<td>2.249</td>
</tr>
<tr>
<td>5-Phenyl-1-pentanol</td>
<td>5</td>
<td>164</td>
<td>243.8</td>
<td>231.30</td>
<td>1.410</td>
<td>172.8</td>
<td>34.57</td>
<td>1.626b</td>
<td>0.573</td>
<td>2.519</td>
</tr>
<tr>
<td>6-Phenyl-1-hexanol</td>
<td>6</td>
<td>178</td>
<td>268.3</td>
<td>252.71</td>
<td>1.420</td>
<td>189.0</td>
<td>31.50</td>
<td>1.346b</td>
<td>0.702</td>
<td>2.250</td>
</tr>
<tr>
<td>7-Phenyl-1-heptanol</td>
<td>7</td>
<td>192</td>
<td>297.0</td>
<td>274.30</td>
<td>1.429</td>
<td>207.9</td>
<td>29.69</td>
<td>1.634b</td>
<td>0.573</td>
<td>2.518</td>
</tr>
</tbody>
</table>

* Taken from Reference 10.
* Calculations carried out with program MOPAC-AM111.

The experimental dipole moments for the phenyl alcohols have been correlated as a function of the adaptable variables T, S and W in Eq. (3). Now, T is defined as minus the standard enthalpy of formation in kJ·mol⁻¹, S as the molecular surface area in Å² and W as the molecular weight in g·mol⁻¹. Table 1 shows the chemical indices for the characterization of a homologous series of phenyl alcohols. In particular, Iₙ is nearly constant (1.39 ± 0.05 mol·Å²·g⁻¹). However, Iₙ increases with the number of C atoms in the alkyl chain n.

The variation of the indices for the phenyl alcohols with n is illustrated in Fig. 1. In particular, Iₙ is the only descriptor that remains constant along the series. Linear correlations have been observed between Iₙ, T, S, W and n. The descriptors more sensitive to n are in the order $T > S > Iₙ > W$.

Table 2 shows the chemical indices for a homologous series of 4-alkylanilines. In particular, Iₙ is almost constant (1.386 ± 0.015 mol·Å²·g⁻¹). Notice that this range lies inside that for the phenyl alcohols. The corresponding interpretation is that both series of molecules have similar form. On the other hand, Iₙ increases with n.

The variation of the indices for the 4-alkylanilines with n is shown in Fig. 2. In particular, Iₙ is constant. Linear correlations have been observed between Iₙ, T, S, W and n. The descriptors more sensitive to n are $T > Iₙ > S > W$. The comparison with the phenyl alcohols shows that Iₙ allows differentiating quantitatively the phenyl alcohols from the 4-alkylanilines. Notice that the data for 4-alkylanilines are closer to their fitting line than those for the phenyl alcohols are (Fig. 1). This fact agrees with the better correlation coefficient for the formers.

The experimental dipole moments for the phenyl alcohols have been correlated as a function of the chemical indices (Table 1). The fit to a constant results:

$$\mu = 1.54 \quad \text{MAPE = 8.38%} \quad \text{AEV = 1.0000}$$

The mean absolute percentage error (MAPE) is 8.38% and the approximation error variance (AEV) is 1.0000.

If μₑ₀ is included in the model of phenyl alcohols, the best linear regression for μ results:
Table 2—Chemical indices for the characterization of a homologous series of 4-alkylanilines.

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Number of C atoms in alkyl chain (n)</th>
<th>W g mol⁻¹</th>
<th>T J mol⁻¹</th>
<th>S Å²</th>
<th>I₁ g mol⁻¹</th>
<th>I₂ J g mol⁻² Å⁻²</th>
<th>I₃ J g mol⁻² Å⁻²</th>
<th>μExperiment D</th>
<th>μvec D</th>
<th>μvecv D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aniline</td>
<td>0</td>
<td>93</td>
<td>-89.6</td>
<td>127.17</td>
<td>1.367</td>
<td>-65.5</td>
<td>-</td>
<td>1.560</td>
<td>0.655</td>
<td>0.792</td>
</tr>
<tr>
<td>4-Methylaniline</td>
<td>1</td>
<td>107</td>
<td>-58.1</td>
<td>147.28</td>
<td>1.376</td>
<td>-42.2</td>
<td>-42.24</td>
<td>1.640</td>
<td>0.079</td>
<td>1.526</td>
</tr>
<tr>
<td>4-Ethylaniline</td>
<td>2</td>
<td>121</td>
<td>-33.9</td>
<td>167.29</td>
<td>1.383</td>
<td>-24.5</td>
<td>-12.25</td>
<td>1.492</td>
<td>0.385</td>
<td>1.339</td>
</tr>
<tr>
<td>4-Propylaniline</td>
<td>3</td>
<td>135</td>
<td>-0.6</td>
<td>187.56</td>
<td>1.389</td>
<td>-0.4</td>
<td>-0.13</td>
<td>1.502</td>
<td>0.087</td>
<td>1.533</td>
</tr>
<tr>
<td>4-Butylaniline</td>
<td>4</td>
<td>149</td>
<td>28.0</td>
<td>207.56</td>
<td>1.393</td>
<td>20.1</td>
<td>5.03</td>
<td>1.511</td>
<td>0.398</td>
<td>1.333</td>
</tr>
<tr>
<td>4-Pentylaniline</td>
<td>5</td>
<td>163</td>
<td>52.4</td>
<td>227.60</td>
<td>1.396</td>
<td>37.6</td>
<td>7.51</td>
<td>1.489</td>
<td>0.093</td>
<td>1.533</td>
</tr>
<tr>
<td>4-Hexylaniline</td>
<td>6</td>
<td>177</td>
<td>81.1</td>
<td>247.75</td>
<td>1.400</td>
<td>58.0</td>
<td>9.66</td>
<td>1.496</td>
<td>0.395</td>
<td>1.332</td>
</tr>
</tbody>
</table>

* * Taken from Reference 10.
* * Calculations carried out with program MOPAC-AM1[11].

\[ \mu = 2.61 - 1.68 \mu_{vec} \]  
\[ n = 8 \quad r = -0.831 \quad s = 0.088 \quad F = 13.4 \quad \text{MAPE = 3.98%} \quad \text{AEV = 0.3100} \]

and AEV decreases 69%. The negative sign of the \( \mu_{vec} \) coefficient is due to the opposite trend of \( \mu_{vec} \) and the \( \mu_{\text{experiment}} \). All other models with greater MAPE and AEV have been discarded. The best non-linear model for \( \mu \) does not improve the results. The inclusion of \( \mu_{vec} \) does not provide better correlations.

If \( I_m \) is included in the model of phenyl alcohols, the best linear regression for \( \mu \) results:

\[ \mu = 4.79 - 1.74 \mu_{vec} - 1.54 I_m \]  
\[ n = 8 \quad r = -0.915 \quad s = 0.070 \quad F = 12.8 \quad \text{MAPE = 2.76%} \quad \text{AEV = 0.1635} \]

and AEV decreases 84%. The best non-linear model for \( \mu \) results:

\[ \mu = -36.5 + 82.1 \mu_{vec} + 20.2 I_m - 27.0 \mu_{vec}^2 - 35.4 \mu_{vec} I_m \]  
\[ \text{MAPE = 1.15%} \quad \text{AEV = 0.0399} \]

and AEV decreases 96%. The inclusion of \( I \) or \( I_{vec} \) does not improve the results.

The experimental dipole moments for the 4-alkylanilines (Table 2) have been correlated. The fit to a constant results:

\[ \mu = 1.53 \]  
\[ \text{MAPE = 2.73%} \quad \text{AEV = 1.0000} \]

The inclusion of \( \mu_{vec} \) in the linear model of 4-alkylanilines does not improve the correlations. However, the best non-linear regression for \( \mu \) results:

\[ \mu = 6.00 - 3.22 I_m \]  
\[ n = 7 \quad r = -0.674 \quad s = 0.045 \quad F = 4.2 \quad \text{MAPE = 1.83%} \quad \text{AEV = 0.5463} \]

and AEV decreases 45%. The inclusion of \( I \) does not improve the results.
If $l_i/n$ is included in the model of the last 6 4-alkylanilines in Table 2, the best linear regression for $\mu$ results:

$$\mu = 1.51 - 0.002681 \cdot l_i/n$$  \hspace{1cm} (18)

where $n = 6$, $r = -0.900$, $s = 0.028$, and $F = 17.1$.

MAPE = 1.17%  \hspace{1cm} AEV = 0.1894

and AEV decreases 81%. If two indices are included in the model, the best linear regression for $\mu$ results:

$$\mu = 1.49 + 0.00103l_i - 0.00443l_i/n$$  \hspace{1cm} (19)

where $n = 6$, $r = -0.953$, $s = 0.023$, and $F = 14.8$.

MAPE = 0.93%  \hspace{1cm} AEV = 0.0923

and AEV decreases 91%. If four indices are included in the model, the best linear regression for $\mu$ results:

$$\mu = 0.604 - 2.13 \cdot \mu_{exp} - 3.25 \cdot \mu_{vec} + 4.36l_i - 0.00360l_i/n$$  \hspace{1cm} (20)

where $n = 6$, $r = -0.963$, $s = 0.035$, and $F = 3.2$.

MAPE = 0.69%  \hspace{1cm} AEV = 0.0671

and AEV decreases 93%.

PCA for the chemical indices results in eight factors $F_1$-$F_8$, which are linear combinations of $W$, $T$, $S$, $I_m$, $I_c$, $\mu_{exp}$, $\mu_{vec}$ and $\mu^v$. For the phenyl alcohols, the coefficients for factor $F_1$ are:

$$F_1 = 0.448W + 0.448T + 0.448S + 0.410I_m + 0.447I_c$$

$$-0.017\mu_{exp} - 0.168\mu_{vec} + 0.019\mu^v$$  \hspace{1cm} (21)

The coefficients for factor $F_2$ are:

$$F_2 = -0.026W - 0.009T - 0.355S - 0.156I_m$$

$$-0.001I_c + 0.583\mu_{exp} - 0.566\mu_{vec} + 0.560\mu^v$$  \hspace{1cm} (22)

The use of only the first factor $F_1$ explains 62% of the variance and gives a relative error of 38%. Moreover, the combined use of the first two factors, $F_1$ and $F_2$, explains 94% of the variance, reducing the relative error to 6%.

The profile of factor $F_1$ is \{20.09, 20.09, 20.10, 16.82, 19.99, 0.03, 2.84, 0.04\}%.

Therefore, $F_1$ cannot be reduced to four variables ($W$, $T$, $S$ and $I_c$) without making a relative error of 20% (the sum of the percentages for the four remaining variables). On the other hand, the profile of factor $F_2$ is \{0.06, 0.01, 0.12, 2.45, 0.00, 33.95, 32.06, 31.35\}%.

For factor $F_2$, the most important variables are $\mu_{exp}$, $\mu_{vec}$ and $\mu^v$. In some way, factor $F_1$ could be considered as a linear combination of $W$, $T$, $S$, $I_m$ and $I_c$ (with a relative error of 3%). However, factor $F_2$ can be expressed as a linear combination of $\mu_{exp}$, $\mu_{vec}$ and $\mu^v$ with a relative error of 3%.

For the 4-alkylanilines, the coefficients for factor $F_1$ are:

$$F_1 = 0.407W + 0.407T + 0.412I_m + 0.408I_c$$

$$-0.280\mu_{exp} - 0.162\mu_{vec} + 0.25\mu^v$$  \hspace{1cm} (23)

The coefficients for factor $F_2$ are:

$$F_2 = 0.097W + 0.085T + 0.097S + 0.020I_m + 0.082I_c$$

$$-0.375\mu_{exp} + 0.690\mu_{vec} - 0.591\mu^v$$  \hspace{1cm} (24)

The use of $F_1$ explains 73% of the variance and gives an error of 27%. Furthermore, the use of $F_1$ and $F_2$ explains 94% of the variance, reducing the error to 6%.

The profile of $F_1$ is \{16.54, 16.59, 16.54, 16.94, 16.62, 7.85, 2.62, 6.30\}%.

The profile of $F_2$ is \{0.95, 0.72, 0.94, 0.04, 0.66, 14.09, 47.67, 34.93\}%.

The PCA $F_2$ vs. $F_1$ plot for the chemical indices of both series is illustrated in Fig. 3. Chemical indices $I_m$, $\mu_{exp}$, $\mu_{vec}$ and $\mu^v$ appear superposed in Fig. 3. Three classes of chemical indices are clearly distinguished: class 1 \{W,S\} (under the bisector,
Plastic evolution allows the design of the new chemical indices \( I_m \) and \( I_e \). They are applied to the homologous series of phenyl alcohols and 4-alkylanilines and compared to charge indices for the dipole moment. From the preceding results the following conclusions can be drawn.

(i). The parameters needed to calculate the index \( I_e \) are the molecular weight \( W \), minus standard enthalpy of formation \( T \), and molecular surface area \( S \).

(ii). The comparison of the results for phenyl alcohols and 4-alkylanilines shows that \( I_e \) allows differentiating quantitatively between phenyl alcohols and 4-alkylanilines.

(iii). Linear and non-linear correlation models have been obtained for the molecular dipole moments of phenyl alcohols and 4-alkylanilines. The new indices \( I_m \) and \( I_e \) have improved the multivariable regression equations for \( \mu \). The correlations between \( I_e \), \( I_m \), \( T \), \( S \), \( W \) and the number of C atoms in the alkyl chain point not only to a homogeneous molecular structure of the phenyl alcohols and 4-alkylanilines, but also to the ability to predict and tailor drug properties. The latter is nontrivial in pharmacology.

(iv). PCA allows grouping the chemical indices in three classes: \{ \( W, S \) \}, \{ \( T, I_e \) \} and \{ \( I_m, \mu_{\text{expt}}, \mu_{\text{sec}}, \mu^V \) \}. Therefore, \( I_m \) is closer to \( \mu \) than \( I_e \) is.

Work is in progress on the application of \( I_e \) to the valence-isoelectronic series of cyclopentadiene, benzene, toluene and styrene. This will give an insight into a possible generality of these conclusions.

Acknowledgement
The author acknowledges the financial support of the Spanish MCT (Plan Nacional I+D+I, Project No. BQU2001-2935-C02-01).

References