Molecular Clustering of Phenylurea Herbicides: Comparison with Sulphonylureas, Pesticides and Persistent Organic Pollutants

Francisco Torrens*, Gloria Castellano**

* Institut Universitari de Ciència Molecular, Universitat de València, Edifici d'Instituts de Paterna, P. O. Box 22085, E-46071 València, Spain
** Facultad de Veterinaria y Ciencias Experimentales, Universidad Católica de Valencia San Vicente Mártir, E-46001 València, Spain

Keywords: Periodic law, Periodic property, Periodic table, Molecular classification, Information entropy

ABSTRACT. Chromatographic retention times of phenylurea herbicides are modelled by structure–property relationships. Properties are hydration free energy and dipole. Bioplastic evolution is an evolutionary perspective conjugating the effect of acquired characters and relations that emerge among evolutionary indeterminacy, morphological determination and natural selection principles. Classification algorithms are proposed based on information entropy and production. Phenylureas are classified by Cl2, O2 and N2 presence; their different behaviour depends on the number of Cl atoms. When applying procedures to moderate-sized sets, excessive results appear compatible with data and suffer a combinatorial explosion; however, the equipartition conjecture allows a selection criterion resulting from classification between hierarchical trees. Information entropy permits classifying compounds and agrees with principal component analyses. Phenylureas periodic table shows that those in the same group present similar properties; phenylureas also in the same period, maximum resemblance. Classification extends to phenylureas, sulphonylureas, pesticides and persistent organic pollutants.

1. INTRODUCTION

Phenylurea herbicides are used in formulations and for nonagricultural use. Sørensen et al. (2008) detected residues as water contaminants. Diuron and linuron are substituted ureas, which are soluble in water, migrate in soil and enter food chain. Both are a toxicological risk to humans and wildlife. The former, which is used in cotton growing and fruit crops, is the third most hazardous pesticide for groundwater resources. Both are applied on railways to maintain quality and provide safer working environment, which leads to groundwater contamination because of leaching potential (Cederlund et al., 2007). Phenylureas enter environment via pathways, e.g., spray drift, runoff from treated fields and leaching into groundwater; they penetrate into soil where they are subjected to micro-organism action and degradation (Canonica & Laubscher, 2008). They are unstable photochemically but they persist in water for several days or weeks depending on temperature and pH. Pesticide pollution of water reservoirs became frequent (Berrada et al., 2004). Phenylurea residues were found in water sources, processed products and crops. In India, most soft-drink bottling plants used surface water from canals and rivers, which present pesticide contamination. Water treatment was insufficient for removal of pesticides, which were above permissible limits. Evidence was provided in a 2003 Centre for Science and Environment (CSE) report, which found several pesticide residues in many soft drinks; findings were confirmed by Joint Parliamentary Committee. In 2006, CSE conducted tests and found pesticides. Sensitive, selective and efficient methods for herbicide analysis are designed. Analytical methods are high-performance liquid chromatography (LC)–ultraviolet (HPLC-UV), solid-phase microextraction (SPME)-HPLC, diode array, immunosorbent trace enrichment and HPLC, LC–mass spectrometry (MS), gas chromatography (GC)–MS, capillary electrophoresis (Chicharro et al., 2005), photochemically induced fluorescence (FL) (Gil-García et al., 2008) and derivative spectrophotometry. Sherma (2005) reviewed thin-layer chromatography (TLC) for herbicide analysis. Muñoz de la Peña et al. (2003) and Cserháti et al. (2004) reported solid-phase extraction (SPE) of phenylurea herbicides.
Lucena et al. (2005), Pan et al. (2005), Mumin et al. (2006), García-Reyes et al. (2008), Yoshioka & Ichihashi (2008) and Cao et al. (2009) informed soft-drink SPE. As polar and degradable-pesticide use becomes widespread, more sensitive analytical methods are developed for residual analysis in matrices. The HPLC showed advantages over GC because it is used for simultaneous analysis of thermally unstable nonvolatile, polar and neutral species without a derivative step. Because of thermally unstable nature of phenylurea herbicides, GC direct application is not possible and derivatization is needed. The HPLC with UV/FL detection was preferred over GC. Kaur et al. (2012) described a simple and sensitive HPLC-UV method for analysis of phenylurea herbicides (monuron, diuron, linuron, metazachlor, metoxuron), which involved single-step preconcentration by SPE. Phenylurea herbicidal action is based on ability to inhibit photosynthesis. Blasco & Picó (2009) and de Aragão Umbuzeiro (2012) revised pesticides occurrence and removal efficiency in sewage treatment plants, and Moganti et al. (2008) and Barco-Bonilla et al. (2010) reported them from Spanish, Mediterranean, Brazilian and other rivers. Navarro et al. (2006) and Navarro-Ortega et al. (2010) informed transport of organic persistent microcontaminants associated with suspended particulate material in Ebro River Basin. Milian (2012) interviewed Barceló reporting that wastewater provides the only real data on drug consumption. Jadán et al. (2012) reported bioaccessibility of Hg and Se from swordfish.

gases mimics and OP derivatives in gas phase via a colorimetric probe array. Rowe (2014) selected the right drug candidate for accurately predicting in vivo hepatotoxicity. Oppermann & Egelkraut-Holtus characterized and quantitatively analyzed wine via spectroscopic methods. Kappes et al. (2014) informed on As poison. In earlier publications Torrens (1999) presented aqueous coefficient calculations for chemicals and drugs. His group reported a novel approach to predict aquatic toxicity from molecular structure (Castillo-Garit, et al., 2008), atom-based quadratic indices to predict aquatic toxicity of benzene derivatives to Tetrahymena pyriformis (Castillo-Garit, et al., 2009), comparative study to predict toxic modes of action of phenols from molecular structures (Brito-Sánchez et al. 2013), atom-based quadratic indices to predict aquatic toxicity of benzene derivatives to T. pyriformis (Castillo-Garit et al., submitted for publication-a) and machine learning-based models to predict modes of toxic action of phenols from molecular structures (Castillo-Garit et al., submitted for publication-b). They informed on polymer bisphenol-A, incorporation of silica (SiO2) nanospheres into epoxy–amine materials, polymer nanocomposites (NCs) (Torrens Zaragozá, 2011), diethylstilbestrol and thermal/mechanical properties of epoxy–SiO2 NCs (Torrens & Castellano, 2013). They described retention-time QSRRs of phenylurea herbicides (Torrens & Castellano, 2012a, 2014a) and pesticides (Torrens & Castellano, 2014b) by bioplastic evolution. A simple computerized algorithm is proposed for establishing chemical structure–biosignificance relation. Starting point is to use information entropy for pattern recognition. Entropy is formulated based on similarity matrix between two biochemical species. As entropy is weakly discriminating for classification, more powerful concepts of entropy production and equipartition conjecture are introduced. The aim of the present report is to review the properties that distinguish the phenylurea structures according with retention times. This study applies chemical index to phenylureas. The goal of this work is index usefulness validation via capability to distinguish phenylureas, and interest as predictive index for retention time as compared with hydration free energy and dipole moment. The following section describes the computational method. The next section illustrates and discusses the calculation results. Finally, the last section summarizes our conclusions.

2. RESEARCH METHOD

The problem in classification studies is to define similarity indices when several criteria of comparison are involved (Iordache, 2011, 2012, 2014). The first step in quantifying similarity for phenylureas is to list main chemical characteristics of molecules. Vector of properties $\vec{1} = \langle i_1, i_2, \ldots, i_k, \ldots \rangle$ should be associated with every urea $i$, whose components correspond to different molecular features in a hierarchical order according to expected retention importance. If characteristic $m$-th is chromatographically more significant for retention than $k$-th then $m < k$. Components $i_k$ are either “1” or “0” according to whether a similar characteristic of rank $k$ is either present or absent in urea $i$ compared to reference. Analysis includes three characteristics in urea molecules: presence of two Cl/O/N atoms (cf. Fig. 1). It is assumed that chemical characteristics of a urea molecule can be ranked according to their contribution to retention in the following order of decaying importance: Cl$_2$ > O$_2$ > N$_2$. Index $i_1$ = 1 denotes Cl$_2$, $i_2$ = 1 means O$_2$ and $i_3$ = 1 signifies N$_2$. In linuron, numbers of atoms are \{Cl$_2$, O$_2$, N$_2$\}; obviously, its vector is $<111>$. Table 1 contains vectors associated with ureas.
Let us denote by $r_{ij}$ ($0 \leq r_{ij} \leq 1$) the similarity index of two phenylureas associated with vectors $\vec{i}$ and $\vec{j}$, respectively. Similitude relationship is characterized by similarity matrix $\mathbf{R} = [r_{ij}]$.

Similarity index between two ureas $\vec{i} = <i_1,i_2,\ldots,i_k,\ldots>$ and $\vec{j} = <j_1,j_2,\ldots,j_k,\ldots>$ is defined as:

$$r_{ij} = \sum_k t_k (a_k)^k \quad (k = 1, 2, \ldots)$$

(1)

where $0 \leq a_k \leq 1$ and $t_k = 1$ if $i_k = j_k$ but $t_k = 0$ if $i_k \neq j_k$. The definition assigns a weight $(ak)^k$ to any property involved in the description of molecule $i$ or $j$. The grouping algorithm uses the stabilized matrix of similarity obtained applying max–min composition rule $o$ defined by:

$$(\mathbf{R} o \mathbf{S})_{ij} = \max_k \left[ \min_k (r_{ik}, s_{jk}) \right]$$

(2)

where $\mathbf{R} = [r_{ij}]$ and $\mathbf{S} = [s_{ij}]$ are matrices of equal type and $(\mathbf{R} o \mathbf{S})_{ij}$ is element $(i,j)$-th of matrix $\mathbf{R} o \mathbf{S}$. When applying max–min composition rule iteratively so that $\mathbf{R}(n+1) = \mathbf{R}(n) o \mathbf{R}$ an integer $n$ exists such that: $\mathbf{R}(n) = \mathbf{R}(n+1) = \ldots$ The resulting matrix $\mathbf{R}(n)$ is called stabilized similarity matrix. Stabilization importance lies in the fact that it generates partition into disjoint classes. The stabilized matrix is designated by $\mathbf{R}(n) = [r_{ij}(n)]$. The grouping rule follows: $i$ and $j$ are assigned to the same class if $r_{ij}(n) \geq b$. The class of $i$ noted $\vec{i}$ is set of species $j$ that satisfies rule: $r_{ij}(n) \geq b$. Matrix of classes is:

$$\mathbf{R}(n) = [r_{ij}] = \max_{st} (r_{st}) \quad (s \in \vec{i}, \ t \in \vec{j})$$

(3)

where $s$ stands for any index of species belonging to class $\vec{i}$ (similarly for $t$ and $\vec{j}$). Rule (3) means finding largest similarity index between species of two different classes. Information entropy $h$ associated with similarity matrix $\mathbf{R}$ results:
Every hierarchical tree corresponds to entropy dependence on grouping level and diagram $h-b$ is obtained. The equipartition conjecture of entropy production is proposed as selection criterion among different variants resulting from classification among hierarchical trees. The best dendrogram is that in which entropy production is most uniformly distributed. The equipartition line is:

$$h_{eq} = h_{max}$$  

(5)

Since classification is discrete, a way of expressing equipartition would be a regular staircase function. The best variant is chosen that minimizing deviation square sum:

$$SS = \sum_{h} (h - h_{eq})^2$$  

(6)

Learning procedures similar to stochastic methods are implemented. Consider a given partition into classes as good from practical observations, which corresponds to the reference similarity matrix $S = [s_{ij}]$ obtained for an arbitrary number of fictious properties. Consider also the same set of species as in the good classification and actual properties. The degree of similarity $rij$ is computed [Eq. (1)] giving matrix $R$. The number of properties for $R$ and $S$ differs. The learning procedure consists in finding classification results for $R$ as close as possible to the good classification. The new similarity matrix is obtained [Eq. (1)]. The distance between partitions characterized by $R$ and $S$ results:

$$D = -\sum_{ij} (1 - r_{ij}) \ln \frac{1 - r_{ij}}{1 - s_{ij}} - \sum_{ij} r_{ij} \ln \frac{r_{ij}}{s_{ij}} \quad (\forall 0 \leq r_{ij}, s_{ij} \leq 1)$$  

(7)


3. RESULTS AND ANALYSIS

For phenylureas, HPLC retention times $R_t$ were taken from Kaur et al. Metoxuron was reference $R_t^o$ because of least retention (cf. Table 1); ratios $(R_t - R_t^o)/R_t^o$ were calculated. Molecular dipole $\mu$ was computed with Molecular Orbital Package–Austin model 1 (MOPAC-AM1). For linuron, an alternative conformation with $1.1\text{kJ} \cdot \text{mol}^{-1}$ energy difference was used with higher dipole $\mu^*$. Standard Gibbs free energy of hydration $\Delta G_{hyd}^o$ was calculated via program SCAP (Torrens & Castellano, 2012b).

Table 1. Vector property (Cl$_2$O$_2$N$_2$), retention, dipole and hydration energy for phenylureas

<table>
<thead>
<tr>
<th>Molecule</th>
<th>$R_t$ (min)</th>
<th>$R_t^o$ (min)</th>
<th>$(R_t - R_t^o)/R_t^o$</th>
<th>$\mu^*$ (D)</th>
<th>$\Delta G_{hyd}^o$ (kJ/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metoxuron N$_2$OCl$_2$&lt;011&gt;</td>
<td>4.30</td>
<td>0.00</td>
<td>0.000000</td>
<td>2.872</td>
<td>-16.6</td>
</tr>
<tr>
<td>Monuron N$_2$OCl$_2$&lt;001&gt;</td>
<td>4.90</td>
<td>0.60</td>
<td>0.13953</td>
<td>4.612</td>
<td>-21.0</td>
</tr>
<tr>
<td>Diuron N$_2$OCl$_2$&lt;101&gt;</td>
<td>7.25</td>
<td>2.95</td>
<td>0.68605</td>
<td>4.622</td>
<td>-22.7</td>
</tr>
<tr>
<td>Metazachlor N$_2$OCl$_2$&lt;000&gt;</td>
<td>8.68</td>
<td>4.38</td>
<td>1.01860</td>
<td>4.749</td>
<td>-23.2</td>
</tr>
<tr>
<td>Linuron N$_2$OCl$_2$&lt;111&gt;</td>
<td>12.40</td>
<td>8.10</td>
<td>1.88372</td>
<td>3.999 (5.077)</td>
<td>-30.3</td>
</tr>
</tbody>
</table>

For linuron, an alternative conformation with $1.1\text{kJ} \cdot \text{mol}^{-1}$ is used ($\mu^*$ between parentheses).

The $(R_t - R_t^o)/R_t^o$ correlated with the dipole moment $\mu$. The fit turned out to be:

$$\left( R_t - R_t^o \right)/R_t^o = -0.338 + 0.260 \mu, \quad n = 5 \quad r = 0.268 \quad s = 0.843 \quad F = 0.2$$

MAPE = 70.12% \hspace{1em} \text{AEV} = 0.9282  

(8)

where $n$ is the number of points, $r$ Pearson correlation coefficient, $s$ standard deviation, $F$ Fischer ratio, MAPE mean absolute percentage error and AEV approximation error variance. Linuron alternative conformation improved results:

$$\left( R_t - R_t^o \right)/R_t^o = -1.97 + 0.620 \mu^*, \quad n = 5 \quad r = 0.709 \quad s = 0.617 \quad F = 3.0$$

MAPE = 51.17% \hspace{1em} \text{AEV} = 0.4977  

(9)
and AEV decayed by 46%. Both dipoles presented fit and AEV dropped by 83%. Hydration free energy $\Delta G_{\text{hydr}}^o$ improved fit and AEV decreased by 91%. Additional fitting parameters were tested: partition coefficients, free energies of solvation and water→organic-solvent transfer, molecular weight, volume, surface, globularity and rugosity, total and hydrophobic/philic solvent-accessible surfaces, molecular fractal dimension and that for external atoms, numbers of Cl, total atoms and cycles, and differential formation enthalpy, etc.; however, results did not better. Dipole inclusion allowed fit and AEV decayed by 93%. Dipole $\mu^*$ addition permitted fitting and AEV dropped by 94%. Both dipoles improved fit and AEV decreased by 98.6%. Chemical indices required variables $T$, $S$ and $W$: $T$ is minus standard formation enthalpy, $S$, molecular surface and $W$, molecular weight. They allowed $I_m = S/W$ and $I_c = T/I_m$. Molecular surface was computed with program TOPO (Torrens, 2003). $\{I_m,I_c\}$ bettered fit and AEV decayed by 78%. Correction of $\{\mu^*,\Delta G_{\text{hydr}}^o\}$ with $I_c$ bettered fit and AEV dropped by 98.7%. Simplified analysis was obtained by hierarchical QSPR (HQSPR): variables were split into logical blocks (electronic $\{\mu,\mu^*\}$, solvation $\{\Delta G_{\text{hydr}}^o\}$, plastic $\{I_m,I_c\}$); every block was represented by its most important variable: $\mu^*$, $\Delta G_{\text{hydr}}^o$ or $I_c$.

Pearson correlation coefficient matrix $R$ was calculated between pairs of vector properties $<i_1,i_2,i_3>$ of five ureas. Intercorrelations are illustrated in partial correlation diagram (PCD), which contains high ($r \geq 0.75$), medium ($0.50 \leq r < 0.75$), low ($0.25 \leq r < 0.50$) and zero ($r < 0.25$) partial autocorrelations. Pairs of ureas with higher partial correlations show similar vector property. However, the results should be taken with care because Entry 4 with constant vector $<000>$ and Entry 5 with constant vector $<111>$ show null standard deviation, causing greatest partial correlations $r = 1$ with any urea, which is an artefact. With the equipartition conjecture, upper triangle resulted:

$$
R = \begin{bmatrix}
0.875 & 0.625 & 0.125 & 0.500 & 0.375 \\
0.875 & 0.375 & 0.750 & 0.125 \\
0.875 & 0.250 & 0.625 \\
0.875 & 0.250 & 0.000 \\
0.875 & \end{bmatrix}
$$

Some correlations are high, e.g., $R_{2,4} = 0.750$. They are illustrated in partial correlation diagram, which contains one high (cf. Fig. 2, red lines), three medium (orange), three low (yellow) and three zero (black) partial correlations. Three out of four high partial correlations of Entry 4 are corrected: its correlation with Entry 1 is medium, its correlation with Entry 3 is low and its correlation with Entry 5 is zero partial correlation. All four high partial correlations of Entry 5 are corrected: correlation with Entry 3 is medium, correlation with Entry 1 is low, and correlations with Entries 2 and 4 are zero partial correlations. In all, 6 out of 7 high partial correlations of Entries 4 and 5 are corrected.
Grouping rule in case with equal weights $a_k = 0.5$ for $0.51 \leq b_1 \leq 0.62$ allows classes:

$C-b_1 = (1,2)(3,5)(4)$

Three groupings are obtained with associated entropy $h-R-b_1 = 4.70$ matching to $<i_1,i_2,i_3>$ and $C-b_1$; the binary taxonomy of Table 1 separates clusters 1, 2 and 3 with two, two and one ureas, respectively. Ureas 3 and 5 with two Cl atoms and great retention are grouped into the same class. Compounds belonging to the same grouping appear highly correlated in the partial correlation diagram (Fig. 2); however, $C-b_1$ results should be taken with care because class (4) with only one substance could be an outlier. At level $b_2$ with $0.13 \leq b_2 \leq 0.50$ the set of groupings is:

$C-b_2 = (1,2,4)(3,5)$

Two classes result and entropy decays to $h-R-b_2 = 2.08$ matching to $<i_1,i_2,i_3>$ and $C-b_2$ dividing clusters: 1 and 2 with three and two ureas. Again, ureas with Cl2 and great retention are grouped into the same class (3,5). Ureas belonging to the same cluster appear highly correlated in the partial correlation diagram (Fig. 2). Table 2 shows a comparative analysis of set containing 1–5 classes in agreement with partial correlation diagram (Fig. 2).
From previous partial correlation diagram (Fig. 2) and set of five classifications (Table 2) we suggest to split data into three groupings:

(1)(2,4)(3,5)

Ureas dendrogram (cf. Fig. 3) shows different behaviour depending on the number of Cl atoms. Once more, ureas with Cl$_2$ and great retention are grouped into class (3,5).

The illustration of the classification above in a radial tree (cf. Fig. 4) shows the different behaviour of ureas depending on the number of Cl atoms. The same classes above are recognized in qualitative agreement with partial correlation diagram and dendrogram (Figs. 2 and 3). One more time, ureas with Cl$_2$ and great retention are grouped into class (3,5).
Program SplitsTree allows examining cluster analysis (CA) data. Based on *split decomposition* it takes as input a *distance matrix* and produces as output a graph, which represents relations between taxa. For ideal data, graph is a tree whereas less ideal data will give rise to a tree-like net, which could be interpreted as a possible evidence for conflicting data. As split decomposition does not attempt to force data onto a tree it can provide a good indication of how tree-like are given data. Splits graph for five ureas (cf. Fig. 5) reveals conflicting relation between class 3 and groupings 1 and 2 because of interdependences. It indicates spurious relation from base-composition effects. It shows different urea behaviour depending on number of Cl atoms in qualitative agreement with partial correlation diagram and binary/radial trees (Figs. 2–4).
A principal components analysis (PCA) was performed for urea vector properties (Shaw, 2003). Importance of PCA factors $F_1$–$F_3$ for $\{i_1,i_2,i_3\}$ is collected in Table 3. Factor $F_1$ explains 56% variance (44% error); $F_{1/2}$, 83% variance (17% error); $F_{1-3}$, 100% variance (0% error).

Table 3. Importance of principal component analysis factors for phenylurea vector property

<table>
<thead>
<tr>
<th>Factor</th>
<th>Eigenvalue</th>
<th>Percentage of variance</th>
<th>Cumulative percentage of variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_1$</td>
<td>1.556666667</td>
<td>55.56%</td>
<td>55.56%</td>
</tr>
<tr>
<td>$F_2$</td>
<td>0.833333333</td>
<td>27.77%</td>
<td>83.33%</td>
</tr>
<tr>
<td>$F_3$</td>
<td>0.300000000</td>
<td>16.67%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

The PCA factor loadings are shown in Table 4.

Table 4. Principal component analysis loadings for the vector property of phenylureas

<table>
<thead>
<tr>
<th>Property</th>
<th>$F_1$</th>
<th>$F_2$</th>
<th>$F_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i_1$</td>
<td>0.53432248</td>
<td>0.70710678</td>
<td>0.46291005</td>
</tr>
<tr>
<td>$i_2$</td>
<td>0.53452248</td>
<td>-0.70710678</td>
<td>0.46291005</td>
</tr>
<tr>
<td>$i_3$</td>
<td>0.65346387</td>
<td>0.00000000</td>
<td>0.76592895</td>
</tr>
</tbody>
</table>

*Factors greater than 0.7 are boldfaced.

The PCA $F_1$–$F_3$ profile for vector property is listed in Table 5. For $F_1$ and $F_3$, variable $i_3$ shows the greatest weight in the profile; however, $F_1$ and $F_3$ cannot be reduced to two variables $\{i_1,i_3\}$ without 29% and 21% errors, respectively. For $F_2$, variables $\{i_1,i_2\}$ present 100% weight and $F_2$ is reduced to both variables with 0% error. Factors $F_{1-3}$ can be considered as linear combinations of $\{i_1,i_3\}$, $\{i_1,i_2\}$ and $\{i_1,i_3\}$ with 29%, 0% and 21% errors.
Table 5. Profile of PCA factors for phenylurea vector property. Percentages ≥50% are bold

<table>
<thead>
<tr>
<th>Factor</th>
<th>Percentage of ( t_1 )</th>
<th>Percentage of ( t_2 )</th>
<th>Percentage of ( t_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( F_1 )</td>
<td>28.57</td>
<td>28.57</td>
<td>42.85</td>
</tr>
<tr>
<td>( F_2 )</td>
<td>59.00</td>
<td>59.00</td>
<td>0.00</td>
</tr>
<tr>
<td>( F_3 )</td>
<td>21.43</td>
<td>21.43</td>
<td>57.14</td>
</tr>
</tbody>
</table>

The PCA \( F_2-F_1 \) scores plot for ureas (cf. Fig. 6) shows different behaviour depending on number of Cl atoms. It distinguishes three groupings: class 1 with one urea (\( F_1 \gg F_2 \), bottom), grouping 2 with two ureas (\( F_1 < F_2 = 0 \), left) and class 3 with two ureas (\( F_1 \approx F_2 > 0 \), top).

Figure 6. Principal component analysis \( F_2 \) vs. \( F_1 \) scores plot for the phenylurea herbicides

From PCA factor loadings of ureas \( F_2-F_1 \) loadings plot (cf. Fig. 7) depicts the three properties. In addition, as a complement to the scores plot (Fig. 6) for the loadings (Fig. 7) it is confirmed that ureas in class 3 located on the top present a contribution of Cl\(_2\) situated on the same position (Fig. 6). The urea in grouping 1 on the bottom shows more pronounced contribution of O\(_2\). Two clusters of properties are clearly distinguished in the loadings plot: class 1 \{Cl\(_2\)\} (\( 0 < F_1 < F_2 \), Fig. 7, top) and grouping 2 \{O\(_2\),N\(_2\)\} (\( F_1 \gg F_2 \), bottom).
Instead of five ureas in the $\mathbb{R}^3$ space of three vector properties, consider three properties in the $\mathbb{R}^5$ space of five ureas. The upper triangle of matrix $R$ between pairs of properties resulted:

$$R = \begin{pmatrix} 0.969 & 0.344 & 0.219 \\ 0.969 & 0.594 \\ 0.969 \end{pmatrix}$$

Some correlations are relatively high, e.g., $R_{2,3} = 0.594$. All are represented in partial correlation diagram, which contains one medium (cf. Fig. 8, orange), one low (yellow) and one zero (black) partial correlations.

Dendrogram for vector properties (cf. Fig. 9) separates first Cl$_2$ (class 1) and then O$_2$ from N$_2$ (grouping 2), in agreement with PCA loadings plot and partial correlation diagram (Figs. 7 and 8).
The radial tree for the vector properties (cf. Fig. 10) separates the same two classes as PCA loadings plot, partial correlation diagram and dendrogram (Figs. 7–9).
Figure 10. Radial tree for the vector properties corresponding to phenylurea herbicides

Splits graph for properties (cf. Fig. 11) reveals no conflicting relation between vector components. It is in agreement with PCA loadings plot, partial correlation diagram and binary/radial trees (Figs. 7–10).
A PCA was performed for vector properties. Factor $F_1$ explains 61% of variance (39% error), factors $F_{1/2}$ account for 83% of variance (17% error), factors $F_{1-3}$ rationalize 96% of variance (4% error), etc. In PCA $F_2$–$F_1$ scores plot the same two clusters of properties are distinguished: class 1 \{Cl$_2$\} ($0 > F_1 > F_2$, cf. Fig. 12, bottom) and grouping 2 \{O$_2$,N$_2$\} ($F_1 < F_2$, top) in agreement with PCA loadings plot, partial correlation diagram, binary/radial trees and splits graph (Figs. 7–11).

The recommended format of phenylureas periodic table (PT, cf. Table 6) shows that they are classified first by $i_1$, $i_2$ and, finally, by $i_3$. Vertical groups are defined by <$i_1$,i$_2$> and horizontal periods, by <$i_3$>. Periods of four units are assumed; e.g., group g00 stands for <$i_1$,i$_2$> = <00>: <000> (Cl$_1$,O$_1$,N$_3$) and <001> (Cl$_1$,O$_1$,N$_2$), etc. Ureas in the same column appear close in partial correlation diagram, binary/radial trees, splits graph and PCA scores (Figs. 2–6). Can et al. (2013) determined sulphonylurea and phenylurea-herbicide toxicities with quantitative structure–toxicity relationships (QSTRs). Table 6 includes 27 sulphonylurea and phenylurea herbicides: metazachlor is similar to sulphonylureas, monuron matches most phenylureas and diuron goes with phenylurea neburon. Wang & Telepchak (2013) determined pesticides in red wine by QuEChERS extraction,
rapid mini-cartridge cleanup and LC–MS–MS detection. Table 6 comprises eight pesticides: (1) metazachlor corresponds thiabendazole, pyrimethanil, cyprodinil, pyrazophos and chlorpyrifos; (2) monuron goes with diazinone. Cabrera & Altmaier (2013) reported high-resolution and ultratrace analysis of pesticides via SiO2 monoliths. Table 6 takes in six new pesticides: (1) metazachlor ties in with metamitron, carbofuran and prometryne; (2) monuron fits carbetamide and isoproturon; (3) methamidophos agrees with metolachlor. Forster & Altmaier (2013) informed qualitative LC–MS analysis of pesticides via monolithic SiO2 capillaries and potential for pesticide assay in kidney. Table 6 contains two novel pesticides: metazachlor tallies chlorfenvinphos and monuron equals pencycuron. Nold (2009) provided analytical standards for persistent organic pollutants (POPs). Bruggemann & Annoni (2014) discussed average heights in partially ordered sets. Table 6 embraces five POPs: metazachlor equals lindane, 2,2',4,4',5-pentabromodiphenylether(BDE-99), chlordecone, pentachlorobenzene and perfluorooctane sulphonate (PFOS).

Table 6. Periodic properties table of phenyl/sulphonylurea herbicides, pesticides and POPs

<table>
<thead>
<tr>
<th>Period</th>
<th>g00</th>
<th>g01</th>
<th>g10</th>
<th>g11</th>
</tr>
</thead>
<tbody>
<tr>
<td>p0</td>
<td>Metazachlor</td>
<td>Methamidophos</td>
<td>Carbenzazim</td>
<td>Metolachlor*</td>
</tr>
<tr>
<td></td>
<td>Amidosulfuron</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Benisulfuron-methyl</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chlorimuron-ethyl</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ethametsulfuron-methyl</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Flazasulfuron</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Metsulfuron-methyl</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nicosulfuron</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Oxyfluoruron</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pyrazosulfuron-ethyl</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thifenasulfuron-methyl</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Triflusulfuron</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trinexsulfuron-Methyl</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trifloxysulfuron-Na</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Triflumuron-methyl</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Azinphosmethyl</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chlorotrocarbure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Idenocarbure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thiabendazole*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pyrimethanil</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cyprodinil</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pyrazophos</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chlorpyrifos</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Metamitron*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carbofuran*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Prometryne*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chlorfenvinphos**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lindane***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2,2',4,4',5-Pentabromodiphenylether (BDE-99)***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chlordecone***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pentachlorobenzene***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Perfluorooctane sulphonate (PFOS)***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p1</td>
<td>Monuron</td>
<td>Meturon</td>
<td>Diuron</td>
<td>Neburon</td>
</tr>
<tr>
<td></td>
<td>Buturon</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chlorotrocarbure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dannuron</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fenuron</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Methylisuron</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Flumeturon</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Siduron</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Diazinone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carbetamide*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Isoproturon*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pencycuron**</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Regular: phenylurea herbicides.  
*Italics*: sulpho/sulphonylurea herbicides.  
**Bold**: pesticides.  
**Bold***: persistent organic pollutants.
Property $P$ variation (cf. Fig. 13) of vector $<i_1,i_2,i_3>$ is expressed in the decimal system $P = 10^2i_1 + 10i_2 + i_3$ vs. structural parameters $\{i_1,i_2,i_3\}$ for ureas. Property was not used in PT (Table 6) development and serves to validate it. Most points appear superimposed. Results show a parameters hierarchy: $i_1 > i_2 > i_3$ in agreement with property PT with vertical groups defined by $\{i_1,i_2\}$ and horizontal periods, by $\{i_3\}$.

Figure 13. Variation of property $P(p)$ of phenylurea herbicides vs. counts $\{i_1,i_2,i_3\}$

Property $P$ change of vector $<i_1,i_2,i_3>$ in base 10 (cf. Fig. 14) vs. the number of group in PT, for ureas, reveals minima corresponding to compounds with $<i_1,i_2>$ ca. $<00>$ (group g00) and maxima with $<i_1,i_2>$ ca. $<11>$ (group g11). For group 1, period 2 is superimposed on 1. Periods $p0$ and $p_1$ represent rows 1 and 2, respectively, in Table 6. Corresponding function $P(i_1,i_2,i_3)$ denotes two periodic waves clearly limited by minima and maxima, which suggest a periodic behaviour that recalls the form of a trigonometric function. For $<i_1,i_2,i_3>$, a minimum is clearly shown. The distance in $<i_1,i_2,i_3>$ units between each pair of consecutive minima is four, which coincides with urea sets in successive periods. The minima occupy analogous positions in the curve and are in phase. The representative points in phase should correspond to elements in the same group in PT. For minima, $<i_1,i_2,i_3>$ coherence exists between the two representations; however, consistency is not general. Wave comparison shows two differences: period 1 is incomplete and period 2 is somewhat staircase-like. Most characteristic points of the plot are minima, which lie about group g00. Values $<i_1,i_2,i_3>$ are repeated as periodic law (PL) states.
An empirical function $P(p)$ reproduces different $<i_1,i_2,i_3>$ values. Minimum of $P(p)$ has meaning only if it is compared with former $P(p-1)$ and later $P(p+1)$ points needing to fulfil:

$$\begin{align*}
P_{\text{min}}(p) &< P(p-1) \\
P_{\text{min}}(p) &< P(p+1)
\end{align*}$$

Order relations (10) should repeat at intervals equal to period size and are equivalent to:

$$\begin{align*}
P_{\text{min}}(p) - P(p-1) &< 0 \\
P(p+1) - P_{\text{min}}(p) &> 0
\end{align*}$$

As relations (11) are valid only for minima, others more general are desired for all $p$ values. The differences $D(p) = P(p+1) - P(p)$ are calculated assigning every value to phenylurea $p$:

$$D(p) = P(p+1) - P(p)$$

Instead of $D(p)$, $R(p) = P(p+1)/P(p)$ can be taken allocating them to urea $p$. If PL were general, elements in the same group in analogous positions in different periods would satisfy:

$$\begin{align*}
either & D(p) > 0 \text{ or } D(p) < 0 \\
and either & R(p) > 1 \text{ or } R(p) < 1
\end{align*}$$

However, results show that this is not the case so that PL is not general existing some anomalies; e.g., $D(p)$ variation vs. group number (cf. Fig. 15) presents a lack of coherence between $<i_1,i_2,i_3>$ Cartesian and PT representations. If consistency were rigorous, all points in each period would present the same sign. In general, a trend exists to give $D(p) > 0$, especially for the lower groups.
Change of $R(p)$ vs. group number (cf. Fig. 16) confirms the lack of constancy between Cartesian and PT charts. If steadiness were exact all points in each period would show $R(p)$ either lesser or greater than one. A trend exists to give $R(p) > 1$, especially for the lower groups.

4. CONCLUSION

From the present results and discussion the following conclusions can be drawn.

1. Objective was to develop a structure–property model for qualitative and quantitative prediction of phenylureas retention. Results contribute to pesticide residue prediction in food and environmental samples. Program SCAP allows hydration and solvation free energies. Hydration free energy and dipole differentiated ureas. An alternative conformation above global minimum was used with higher dipole. Parameters needed for co-ordination index are molar formation enthalpy, weight and surface area. Morphological and co-ordination indices improved correlations for retention. Co-ordination index was predictive for retention when used with others. Correlation between molecular area and weight points to a homogeneous molecular structure of ureas, and
ability to predict and tailor herbicide properties; the latter is nontrivial in environmental toxicology. A hierarchical quantitative structure–property relation provided properties simplified analyses.

2. Criteria selected to reduce the analysis to a manageable quantity of phenylureas referred to structural characteristics related to presence of two O, N and, especially, Cl atoms. Classification agreeded with principal component analyses. Program MolClas is a simple, reliable, efficient and fast procedure for molecular categorization based on equipartition conjecture of entropy production. It was written not only to analyze the equipartition conjecture of entropy production but also to explore the world of molecular clustering.

3. Periodic law does not satisfy physics-law status: (a) phenylureas retentions are not repeated; perhaps chemical character; (b) order relations are repeated with exceptions. The analysis forces the statement: The relations that any compound \( p \) has with its neighbour \( p + 1 \) are approximately repeated for each period. Periodicity is not general; however, if substance natural order is accepted the law must be phenomenological. Retention is not used in periodic-table generation and serves to validate it. Other-property analysis would give insight into possible generality of periodic table. Periodic classification was extended to phenylureas, sulphonylureas, pesticides and persistent organic pollutants.

Work is in progress on the quantitative structure–antioxidant activity models of isoflavonoids from Dalbergia parviflora.

ACKNOWLEDGEMENTS

One of us, F. T., thanks support from the Spanish Ministerio de Economía y Competitividad (Project No. BFU2013-41648-P) and EU ERDF.

REFERENCES


