Properties of New Orthogonal Graph Theoretical Invariants in Structure−Property Correlations

Oswaldo Araujo[†] and Daniel A. Morales*[,‡]

Departamentos de Matemática y Química, Facultad de Ciencias, Universidad de los Andes, Mérida 5101, Venezuela

Received August 8, 1997

A recently proposed procedure for orthogonalization of graph theoretical invariants, based on Galois theory, is applied to the elucidation of structure−property relations. The properties of the new procedure are described on a correlation of the boiling points versus the new orthogonal graph theoretical invariants for the series of heptane isomers. The differences of this orthogonalization procedure with that of Randić's are clearly stated.

1. INTRODUCTION

Orthogonal graph theoretical invariant is a new concept that permits the evaluation of the role of individual descriptors in multivariate regression analysis of structure−property relationships.[1−5] The resulting regression equations are stable under the inclusion or exclusion of orthogonal descriptors.

The importance of this concept is clearly appreciated when one considers that in many studies of structure−property relationships it has been shown that, in general, many molecular descriptors are collinear or near collinear, in the sense that they overlap the same kind of molecular information (molecular shape, molecular size, branching, etc.).[6] In this sense, a regression analysis of the structure−property relationship using several nonorthogonal molecular descriptors will not allow us to ascertain the contribution of a single aspect of the structure concept to the molecular property being studied.

The use of orthogonal descriptors in structure−property relationships provides a clear interpretation of the multivariate regression equations in terms of the contribution of a single aspect of the molecular structure concept to the molecular property in concern. Originally, Randić introduced the concept of orthogonal graph theoretical invariants as *analogous to orthogonal vectors*[1−4] and defined the collinear or orthogonality of those vectors in terms of the correlation coefficient obtained from a linear regression between two such vectors. If the correlation coefficient was 1 or 0, the vectors were considered collinear or orthogonal, respectively. Thus, the correlation coefficient had the role of defining some kind of scalar product. However, in that procedure it was not clear which was the working vector space.

Recently, we have put those vague notions in a sound theoretical framework based on Galois theory.[7,8] We showed that some graph theoretical invariants are vectors of the vector space $Q(\sqrt{2}, \sqrt{3})$, over the field Q of rational numbers. On that vector space then, a symmetric bilinear form can be defined that allows us to use the Gram−Schmidt orthogonalization procedure.[9] In this way, we identify the connectivity indices ${}^{i}\chi$ as vectors in the vector space $Q(\sqrt{2}, \sqrt{3})$ and define new orthogonal graph theoretical invariants ${}^{i}\Omega$ in that space, by the Gram−Schmidt procedure.

The purpose of this paper is to show the practical applications of our procedure in studies of structure−property relations. Several properties of the new procedure and of the new orthogonal descriptors are described on a multivariate analysis of the boiling points (the property) for heptane isomers. The similarities and differences with Randić's procedure are outlined. It is shown that in our procedure, the functional dependence property−structure can be constructed using only the information from the initial multivariate regression analysis with the nonorthogonal graph theoretical invariants and results of the Gram−Schmidt procedure. At variance with Randić's approach, the coefficients of the function $P({}^{i}\Omega)$ are not constants. However, for a given isomer, they are well defined functions of the structural characteristics of that isomer.

2. GRAM−SCHMIDT PROCEDURE AND ORTHOGONAL BASIS

Let V denote a finite dimensional vector space over the field of real numbers or the complex field, with inner product $\langle , \rangle$. Let $\{e_1, e_2, ..., e_n\}$ be an orthogonal basis for V. Then, the following theorem is a generalization of the Gram−Schmidt orthogonalization process.

Theorem 1. *Let V be a finite-dimensional vector space over a field of characteristic zero, and let f be a symmetric bilinear form on V. Then, there is an ordered basis for V in which f is represented by a diagonal matrix.*

If we are given an arbitrary basis $\{u_1, u_2, ..., u_n\}$ of V which we wish to replace by an orthogonal basis, we proceed as follows. Let

[†] Departamento de Matemática.
[‡] Departamento de Química.

10.1021/ci970062h CCC: $15.00
Published on Web 10/02/1998

$$v_1 = u_1$$

$$v_2 = u_2 - \frac{f(u_2, v_1)}{f(v_1, v_1)} v_1$$

$$\vdots$$

$$v_n = u_n - \frac{f(u_n, v_{n-1})}{f(v_{n-1}, v_{n-1})} v_{n-1} - \ldots - \frac{f(u_n, v_1)}{f(v_1, v_1)} v_1 \quad (1)$$

Then, the new set $\{v_1, v_2, \ldots, v_n\}$ is an orthogonal basis.

The Gram−Schmidt orthogonalization procedure[9] transforms the nonorthogonal basis to a new orthogonal basis and therefore **redistributes** the structural information originally contained in the old nonorthogonal vectors into the new orthogonal ones.

The theorem just presented can now be applied to the study of the structure−property relation, with the structure modeled by the connectivity indexes.

3. THE ORTHOGONAL CONNECTIVITY BASIS

We know that the dimension of the vector space $Q(\sqrt{2}, \sqrt{3})$ is four and that $\{1, \sqrt{2}, \sqrt{3}, \sqrt{6}\}$ is a basis of this space. On the other hand, we have shown that the connectivity indices are vectors of $Q(\sqrt{2}, \sqrt{3})$, a result related to the valence of the carbon atom.[7,8]

Now, let us suppose that we have chosen four connectivity indices ${}^i\chi$, where $i = 1, 2, 3, 4$, which are linearly independent. Our aim is then to study how well our selected basis spans the property space.[10]

Let $u = a_1 + a_2\sqrt{2} + a_3\sqrt{3} + a_4\sqrt{6}$ and $v = b_1 + b_2\sqrt{2} + b_3\sqrt{3} + b_4\sqrt{6}$ be any two vectors of $Q(\sqrt{2}, \sqrt{3})$. Define $f: Q(\sqrt{2}, \sqrt{3}) \times Q(\sqrt{2}, \sqrt{3}) \rightarrow Q$ by $f(u, v) = a_1b_1 + 2a_2b_2 + 3a_3b_3 + 6a_4b_4$. Then, the inner product $\langle u, v \rangle$ of two vectors u and v is given by $\langle u, v \rangle = f(u, v)$.

Let $\{{}^1\Omega, {}^2\Omega, {}^3\Omega, {}^4\Omega\}$ be an orthogonal basis for $Q(\sqrt{2}, \sqrt{3})$ (which there exists by the theorem).

Because the property P is a vector on $Q \subset Q(\sqrt{2}) \subset Q(\sqrt{2}, \sqrt{3})$,[11] it can be expressed in one and only one way as

$$P = \sum_{i=1}^{n} b_i \, {}^i\Omega \quad (2)$$

where $b_i \in Q$, $i = 1, \ldots n$.

Taking the inner product $\langle P, {}^j\Omega \rangle$ of the vectors

$$P = p_1 + 0\sqrt{2} + 0\sqrt{3} + 0\sqrt{6}$$

$${}^j\Omega = {}^ja_1 + {}^ja_2\sqrt{2} + {}^ja_3\sqrt{3} + {}^ja_4\sqrt{6}, \qquad j = 1, \ldots, n \quad (3)$$

and using the bilinear properties of the inner product, we get

$$\langle P, {}^j\Omega \rangle = \sum_{i=1}^{4} b_i \langle {}^i\Omega, {}^j\Omega \rangle$$

$$= \sum_{i=1}^{4} b_i f({}^i\Omega, {}^j\Omega)$$

$$= b_j f({}^j\Omega, {}^j\Omega), \; 1 \le j \le n \quad (4)$$

where the last relation is obtained by the orthogonalization condition. Thus,[12]

$$b_j = \frac{\langle P, {}^j\Omega \rangle}{f({}^j\Omega, {}^j\Omega)}$$

On the other hand, from the definition of the inner product, we obtain

$$\langle P, {}^j\Omega \rangle = f(P, {}^j\Omega)$$

$$= p_1 {}^ja_1 \quad (5)$$

and

$$\langle {}^j\Omega, {}^j\Omega \rangle = f({}^j\Omega, {}^j\Omega)$$

$$= ({}^ja_1)^2 + 2({}^ja_2)^2 + 3({}^ja_3)^2 + 6({}^ja_4)^2, \; j = 1, \ldots, n \quad (6)$$

Finally, combining eqs 4−6 we get

$$b_j = \frac{p_1 {}^ja_1}{({}^ja_1)^2 + 2({}^ja_2)^2 + 3({}^ja_3)^2 + 6({}^ja_4)^2}$$

$$b_j \in Q \; (j = 1, \ldots, n) \quad (7)$$

The b_j values give the contribution of the orthogonal descriptor ${}^j\Omega$ to the property P. On the other hand, it is obvious that

$$\sum_{i=1}^{n} \frac{{}^ja_1}{({}^ja_1)^2 + 2({}^ja_2)^2 + 3({}^ja_3)^2 + 6({}^ja_4)^2} \, {}^j\Omega = 1 \quad (8)$$

Thus, we see that we can calculate the relative contribution of each orthogonal descriptor in the relation structure−property directly from the algebra of vector spaces without using multiple regression analysis, as in the case of Randić's approach. Of course, the absolute values will need the knowledge of the value of the property for a given isomer, as in the case of Randić's approach. On the other hand, the coefficients of the relation property−structure are stable. Because we are representing the value of the property exactly by means of an orthogonal basis, any new descriptor that we introduce will be dependent on the others and its coefficient will be zero, so that the coefficients of the relation will not change. Thus, the algebra of vector spaces indicates to us how to do things right (exactly) in modeling the structure−property relation.

As an illustration of eq 8, consider the linear alkanes. For them it is known that

$$^{\nu}\chi = \frac{n-\nu-2}{2^{(\nu+1)/2}} + \frac{1}{2^{\nu/2-1}} \quad (9)$$

In this case, $\{^1\chi, ^2\chi\}$ is a nonorthogonal basis for $Q(\sqrt{2})$ and use of eq 1 will generate an orthogonal basis $\{^1\Omega, ^2\Omega\}$. Thus,

$$^1\chi = \frac{n-3}{2} + \sqrt{2}$$

$$^2\chi = 1 + \frac{n-4}{4}\sqrt{2}$$

$$^1\Omega = {}^1\chi = \frac{n-3}{2} + \sqrt{2}$$

$$^2\Omega = -\frac{n^2-7n+4}{n^2-6n+17} + \frac{(n-3)(n^2-7n+4)}{4(n^2-6n+17)}\sqrt{2}$$

Then, from eq 7 we obtain

$$b_1 = \frac{\frac{n-3}{2}p_1}{\left(\frac{n-3}{2}\right)^2 + 2.1^2}$$

$$b_2 = \frac{-\frac{n^2-7n+4}{n^2-6n+17}p_1}{\left(-\frac{n^2-7n+4}{n^2-6n+17}\right)^2 + 2\left(\frac{(n-3)(n^2-7n+4)}{4(n^2-6n+17)}\right)^2}$$

and, consequently,

$$\sum_{i=1}^{2} \frac{b_j}{p_1} {}^j\Omega = 1$$

4. THE PROBLEM OF PREDICTION

As stated in the preceding section our procedure is self-contained; that is, we do not need to use correlation procedures either to construct the orthogonal descriptors or to express the structure−property relation. Given a certain molecular graph, we can calculate exactly the orthogonal basis from the knowledge of the nonorthogonal basis. Moreover, we can express the property in terms of this orthogonal basis, as expressed by eq 2. However, one can argue that the modeling needs the values of the property in question. This is so, as is the case of the modeling using regression analysis. We cannot expect that graph theory will allow us to model a certain property without using experimental information. Our approach permits us to say, for instance, that the value of 98.4 °C for the boiling point of *n*-heptane can be written as (see eqs 2 and 7)

$$98.4 = \sum_{i=1}^{2} \frac{98.4\,{}^j a_1}{({}^j a_1)^2 + 2({}^j a_2)^2 + 3({}^j a_3)^2 + 6({}^j a_4)^2} {}^j\Omega$$

$$= 98.4\frac{2}{2^2 + 2.1^2} \cdot {}^1\Omega + 98.4\frac{-\frac{1}{6}}{\left(-\frac{1}{6}\right)^2 + 2\left(\frac{1}{6}\right)^2} \cdot {}^2\Omega$$

$$= 32.8\,{}^1\Omega - 2\,{}^2\Omega$$

which indicates the different contributions of the orthogonal connectivity indexes to the property.

Now how could one model the property of an isomer for which we only know its molecular graph? We simply use the result of the regression analysis for a certain number of compounds and predict from it the value of the property for the unknown. This value can then be modeled using the orthogonal descriptors that can always be calculated from the known structure and eq 1, as in the example just discussed of *n*-heptane. It is important to note that in Randić's approach, the regression equation for the orthogonal descriptors−property relation can be constructed knowing only the coefficients of the regression analysis using the nonorthogonal descriptors. Thus, the correlation analysis between property versus orthogonal descriptors does not need to be done.

5. OUR ORTHOGONALIZATION METHOD VERSUS THAT OF RANDIC

One can construct the function $P({}^i\Omega)$, $i = 1, 2, 3, 4$ directly from eq 2 using only the ${}^j\Omega$ obtained from the orthogonalization process, not invoking regression procedures and directly using the experimental or predicted data for the property.

In our case, the coefficients of $P({}^i\Omega)$ are not constants (however, they are well defined) but depend on structural properties of the corresponding isomer.

Our procedure has a sound theoretical foundation.[7,8] We can exactly model, using the algebra of vector spaces, the property in terms of a set of orthogonal graph theoretical invariants. However, to make predictions we have to resort to the standard procedure of multivariate regression analysis. But, once we know the predicted value of the property, we can "decompose" this value in terms of contributions from the different orthogonal descriptors. Because the ${}^i\Omega$ values corresponding to a given isomer (structure) are orthogonal themselves, we get an interpretation of the terms involved in the $P({}^i\Omega)$ function.

In Randić's approach, each element of structure (each ${}^i\Omega^*$) contributes *equally* to the molecular property, for any isomer.[1−4] In our approach, each element of structure (each ${}^i\Omega$) contributes *differently* to the molecular property for each isomer. However, in opposition with the regression equation approach, which uses the nonorthogonal descriptors and where each coefficient in the correlation varies randomly with the introduction of a new descriptor, the effect produced by the introduction of a new orthogonal descriptor to the property function is very well defined and can be calculated ab initio. Furthermore, in Randić's approach, the ${}^i\Omega^*$ are obtained from *global correlations* that involve all isomers; that is, his ${}^i\Omega^*$ contain structural information about all isomers. In contrast, in our case, each ${}^i\Omega$ *is* calculated individually for each isomer by the Gram−Schmidt orthogonalization procedure, without resorting to regression analysis. This difference is the reason why in Randić's approach the coefficients of $P({}^i\Omega^*)$ are constants whereas in our case they are, in general, well-defined functions of the individual structural characteristics, obtained through the ratio of bilinear forms.

Table 1. Values of $^{\nu}\chi$ for the Isomers of Heptane

isomer	$^{1}\chi$	$^{2}\chi$	$^{3}\chi$	$^{4}\chi$
n-heptane	$2+\sqrt{2}$	$1+{}^{3}/_{4}\sqrt{2}$	${}^{1}/_{2}+{}^{1}/_{2}\sqrt{2}$	${}^{1}/_{2}+{}^{1}/_{8}\sqrt{2}$
3-ethylpentane	${}^{3}/_{2}\sqrt{2}+{}^{1}/_{2}\sqrt{6}$	${}^{1}/_{2}\sqrt{3}+{}^{1}/_{2}\sqrt{6}$	$\sqrt{3}$	${}^{1}/_{2}\sqrt{3}$
3-methylhexane	${}^{1}/_{2}+\sqrt{2}+{}^{1}/_{3}\sqrt{3}+{}^{1}/_{3}\sqrt{6}$	${}^{1}/_{2}+{}^{1}/_{3}\sqrt{3}+{}^{1}/_{2}\sqrt{6}$	${}^{1}/_{2}\sqrt{3}+{}^{1}/_{4}\sqrt{6}$	${}^{1}/_{6}\sqrt{3}+{}^{1}/_{6}\sqrt{6}$
2-methylhexane	$1+{}^{1}/_{2}\sqrt{2}+{}^{2}/_{3}\sqrt{3}+{}^{1}/_{6}\sqrt{6}$	${}^{1}/_{2}+{}^{1}/_{4}\sqrt{2}+{}^{1}/_{2}\sqrt{3}+{}^{1}/_{3}\sqrt{6}$	${}^{1}/_{4}\sqrt{2}+{}^{1}/_{3}\sqrt{3}+{}^{1}/_{12}\sqrt{6}$	${}^{1}/_{4}\sqrt{6}$
2,3-dimethylpentane	${}^{1}/_{3}+{}^{1}/_{2}\sqrt{2}+\sqrt{3}+{}^{1}/_{6}\sqrt{6}$	$1+{}^{1}/_{6}\sqrt{2}+{}^{1}/_{3}\sqrt{3}+{}^{1}/_{3}\sqrt{6}$	${}^{2}/_{3}+{}^{1}/_{2}\sqrt{2}+{}^{1}/_{6}\sqrt{6}$	${}^{1}/_{3}\sqrt{2}$
2,4-dimethylpentane	${}^{4}/_{3}\sqrt{3}+{}^{1}/_{3}\sqrt{6}$	${}^{1}/_{6}\sqrt{2}+{}^{2}/_{3}\sqrt{3}+{}^{2}/_{3}\sqrt{6}$	${}^{2}/_{3}\sqrt{2}$	${}^{2}/_{3}\sqrt{2}$
3,3-dimethylpentane	$1+{}^{3}/_{2}\sqrt{2}$	${}^{3}/_{4}+{}^{3}/_{2}\sqrt{2}$	${}^{1}/_{2}+\sqrt{2}$	${}^{1}/_{4}$
2,2-dimethylpentane	$2+{}^{3}/_{4}\sqrt{2}$	${}^{9}/_{4}+{}^{3}/_{4}\sqrt{2}$	1	${}^{3}/_{4}$
2,2,3-trimethylbutane	${}^{3}/_{2}+{}^{5}/_{6}\sqrt{3}$	${}^{3}/_{2}+{}^{7}/_{6}\sqrt{3}$	$\sqrt{3}$	0

6. A FORMALIZATION OF RANDIC'S METHOD ALONG THE LINES OF THE PRESENT APPROACH

We now describe how Randić's procedure can be expressed in the terminology and methodology that we describe. Consider the heptane isomers. In applying his orthogonalization procedure, Randić considers $^{\nu}\chi$ as a column vector with nine components, where each component represents an isomer. This viewpoint can be formalized easily starting from the fact that, as we have shown, $^{\nu}\chi$ is an element of the Q-vector space $Q(\sqrt{2}, \sqrt{3})$.

First, recall the definition of direct sum of two vector spaces.[9] Let U and V be vector spaces over a field K. The (external) direct sum $U + V$ of U and V is the vector space whose underlying set is the Cartesian product $U \times V$ of U and V, and with vector space operations defined by

$$(u, v) + (u_1, v_1) = (u + u_1, v + v_1)$$

$$\alpha(u, v) = (\alpha u, \alpha v) \tag{11}$$

for all vectors and scalars involved.

It is easily checked, and we omit the details, that $U + V$ is a vector space. The following theorem is also easily proved.

Theorem 2. *Let U and V be finite dimensional vector spaces with basis* $\{u_1, u_2, ..., u_m\}$, $\{v_1, v_2, ..., v_n\}$, *respectively. Then:*

1. $\{(u_1, 0), (u_2, 0), ..., (u_m, 0), (0, v_1), (0, v_2), ..., (0, v_n)\}$ *is a basis of* $U + V$.

2. $\dim(U + V) = \dim U + \dim V$.

Set $E = Q(\sqrt{2}, \sqrt{3})$. Then $E^9 = E + ...(seven\ times) + E$ and $\dim E^9 = 36$ because $\dim E = 4$. Thus, if x belongs to E^9, $x = (x_1, ..., x_{36})$ with $x_i \in E$. In particular, $^{\nu}\chi \in E^9$ means $^{\nu}\chi = (^{\nu}\chi_1, ..., ^{\nu}\chi_{36})$ where each component represents an isomer. For instance, in Table 1, $^{1}\chi$, $^{2}\chi$, $^{3}\chi$, and $^{4}\chi$ are vectors of E^9 whose components are *n*-heptane, 3-ethylpentane, 3-methylhexane, 2-methylhexane, 2,3-dimethylpentane, 2,4-dimethylpentane, 3,3-dimethylpentane, 2,2-dimethylpentane, and 2,2,3-trimethylbutane.

Let $f: E^9 \times E^9 \rightarrow Q$ be defined by $f(x, y) = \sum_{i=1}^{36} x_i y_i$, where $x = (x_1, ..., x_{36})$ and $y = (y_1, ..., y_{36})$.

It is easily shown that f is a symmetric bilinear form on the pair of vector spaces (E^9, E^9).

Let S be a subspace of E^9 such that $\{^{1}\chi, ^{2}\chi, ^{3}\chi, ^{4}\chi\}$ is a nonorthogonal basis. Then, by using the Gram−Schmidt orthogonalization process, we obtain an orthogonalized basis $\{^{1}\Omega, ^{2}\Omega, ^{3}\Omega, ^{4}\Omega\}$ of the Q-vector space S.

Example. For *n*-heptane and 3-ethylpentane we have (cf. Table 1)

$$^{1}\chi_1 = 2 + \sqrt{2}$$

$$^{1}\chi_2 = \frac{3}{2}\sqrt{2} + \frac{1}{2}\sqrt{6}$$

Therefore,

$$^{1}\chi = (^{1}\chi_1, ^{1}\chi_2)$$

$$= \left(2 + \sqrt{2}, \frac{3}{2}\sqrt{2} + \frac{1}{2}\sqrt{6}\right)$$

$$^{2}\chi_1 = 1 + \frac{3}{4}\sqrt{2}$$

$$^{2}\chi_2 = \frac{1}{2}\sqrt{3} + \frac{1}{2}\sqrt{6}$$

Hence,

$$^{2}\chi = (^{2}\chi_1, ^{2}\chi_2)$$

$$= \left(1 + \frac{3}{4}\sqrt{2}, \frac{1}{2}\sqrt{3} + \frac{1}{2}\sqrt{6}\right)$$

We have $f: E^2 \times E^2 \rightarrow Q$

$$f(^{1}\chi, ^{2}\chi) = 2 + \frac{3}{4} + \frac{1}{4} = 3$$

Thus, $^{1}\chi$ and $^{2}\chi$ are not orthogonal.

Let S be a subspace of E^2 generated by $\{^{1}\chi, ^{2}\chi\}$. Because they are linearly independent, the set $\{^{1}\chi, ^{2}\chi\}$ is a nonorthogonalized basis of S.

Now, using the Gram−Schmidt orthogonalization process, we have

$$^{1}\Omega = ^{1}\chi$$

$$^{2}\Omega = ^{2}\chi - \frac{f(^{2}\chi, ^{1}\chi)}{f(^{2}\chi, ^{2}\chi)}\,^{1}\chi$$

$$= ^{2}\chi - \frac{16}{11}\,^{1}\chi$$

because

$$f(^{2}\chi, ^{2}\chi) = 1 + \frac{9}{16} + \frac{1}{4} + \frac{1}{4} = \frac{33}{16}$$

In this case, $\dim_Q E^2 = 8$ and the set $\{(1, 0), (\sqrt{2}, 0), (\sqrt{3}, 0), (\sqrt{6}, 0), (0, 1), (0, \sqrt{2}), (0, \sqrt{3}), (0, \sqrt{6})\}$ is a basis

Table 2. Stepwise Regression Equations for the Boiling Points of Heptane Isomers using the Connectivity Indices ${}^{i}\chi$, $i = 1, 2, 3, 4$

equation	r	SE, °C
$bp = 40.4181\ {}^{1}\chi - 41.3827$	0.9994	2.760
$bp = -11.8751\ {}^{1}\chi - 15.4670\ {}^{2}\chi + 167.6221$	0.9996	2.380
$bp = 130.4733\ {}^{1}\chi + 25.7546\ {}^{2}\chi + 11.9197\ {}^{3}\chi - 416.046$	0.9998	1.817
$bp = 20.3389\ {}^{1}\chi - 8.9538\ {}^{2}\chi - 1.3729\ {}^{3}\chi - 9.3553\ {}^{4}\chi + 54.4705$	0.9999	1.150

Table 3. Calculated Boiling Points (bp) of Heptanes using Connectivity Indices ${}^{i}\chi$, $i = 1, 2, 3, 4$

isomer	exp bp	calc bp	Δ
n-heptane	98.4	97.47	+0.93
3-ethylpentane	93.5	93.33	0.17
3-methylhexane	92	92.59	−0.59
2-methylhexane	90	90.98	−0.98
2,3-dimethylpentane	89.8	88.76	+1.04
2,4-dimethylpentane	80.5	80.86	−0.36
3,3-dimethylpentane	86.1	87.28	−1.18
2,2-dimethylpentane	79.2	78.69	+0.51
2,2,3-trimethylbutane	80.9	80.43	+0.47

of E^2. Moreover, if $x \in E^2$, then $x = \sum_{i=1}^{8} \alpha_i e_i$, where $e_1 = (1, 0)$, ..., $e_8 = (0, \sqrt{6})$. On the other hand, $f(x, y) = \sum_{i=1}^{8} \alpha_i \beta_i$ with $y = \sum_{i=1}^{8} \beta_i e_i$.

Remark. We have defined the direct sum of two vector spaces but it is easy to generalize to a finite number of spaces. On the other hand, we have taken E^9 to consider Table 1 of Randić's paper.[1]

7. RESULTS AND DISCUSSION

In Table 1 we show the values of the ${}^{\nu}\chi$, $\nu = 1, 2, 3, 4$ expressed as vectors of the vector space $Q(\sqrt{2}, \sqrt{3})$ and in Table 2 we show the regression equations and the statistical parameters r (the correlation coefficient) and SE (the standard error) derived for boiling points (bp) of heptanes using the connectivity indices ${}^{\nu}\chi$, $\nu = 1, 2, 3, 4$. As observed, we obtain an increase of r and a decrease of SE each time we include an additional connectivity index into the basis. This result is evidence that each time we include an additional descriptor the property space is much better represented. The computed bp and the residuals for the family of heptane isomers using all four connectivity indices in a multivariate regression are shown in Table 3. Another aspect observed in the correlation equations is a random feature: each time a new additional descriptor is included, the corresponding coefficient changes dramatically. Thus, no interpretation of the coefficients of the correlation as representing particular features of the concept of molecular structure can be given. Randić has shown that this feature can be remedied by the introduction of orthogonal descriptors. In his approach, the introduction of orthogonal graph theoretical invariants gives stability in the coefficients of the correlation equations and, consequently, now an interpretation of the coefficients can be given as representing the weight of different aspects of the molecular structure concept (implicit in the particular indices chosen) to the property in question.

In our approach, the coefficients of the orthogonalized graph theoretical invariants in the equation of property versus structure are not constants. However, we know the exact dependence of the coefficients on the value of the property and the parameters of the orthogonalization process. On the other hand, because our set of orthogonal descriptors is a basis, any new descriptor added will be dependent on the others already present; therefore, their contribution to the relation will be zero so that the coefficients of the structure−property relation will not change by the introduction of a new descriptor. In this way we can think of eqs 2, 5, and 8 as general equations from where molecular properties of a given isomer can be computed from its molecular structure and where each coefficient of the formula, which reflects the contribution of a particular aspect of molecular structure to the property in question, can be computed ab initio.[13]

The numerical values of the symmetric bilinear form $f(a, b)$ for all nine heptane isomers are shown in Table 4. These values were used to calculate the values of the orthogonalized ${}^{\nu}\Omega$, $\nu = 1, 2, 3, 4$, which are given in Table 5. With those values and the experimental data on boiling points, we can construct the values of the coefficients in the structure−property relation. These values are shown in Table 6. In this way we can "divide" the value of the boiling point for a given isomer in contributions coming from different and nonoverlapping aspects of the molecular structure concept. In the case of an isomer for which we do not know its boiling point, we can predict its value from the equations in Table 2 and then model this value in terms of contributions from the corresponding basis of orthogonal descriptors.

It is important to mention that for the prediction of the boiling point we have not made any attempt to find the best set of four graph theoretical invariants to be used in the orthogonalization process (i.e., those which will yield the smallest standard deviation). We have just chosen a maximal set of linearly independent elements of $Q(\sqrt{2}, \sqrt{3})$.

Table 4. Values of the Symmetric Bilinear Form $f(u, v)$ for the Isomers of Heptane

isomer[a]	$f({}^{2}\chi, {}^{1}\Omega)$	$f({}^{1}\Omega, {}^{1}\Omega)$	$f({}^{3}\chi, {}^{2}\Omega)$	$f({}^{2}\Omega, {}^{2}\Omega)$	$f({}^{3}\chi, {}^{1}\Omega)$	$f({}^{4}\chi, {}^{3}\Omega)$	$f({}^{3}\Omega, {}^{3}\Omega)$	$f({}^{4}\chi, {}^{2}\Omega)$	$f({}^{4}\chi, {}^{1}\Omega)$
n-heptane	7/2	6	−	−	−	−	−	−	−
3-E	3/2	6	3/2	15/8	0	−	−	−	−
3-M	19/12	13/4	119/156	307/234	1	123/1228	459/1228	11/26	1/2
2-M	25/12	3	7/72	149/432	1	−	−	−	−
2,3-MM	11/6	34/9	25/34	1427/1224	8/9	1228/4281	1876/4281	−31/612	1/3
2,4-MM	4	6	2/9	25/18	0	−	−	−	−
3,3-MM	21/4	11/2	−	−	−	−	−	−	−
2,2-MM	45/8	41/8	−	−	−	−	−	−	−
2,2,3-MMM	31/6	13/3	−	−	−	−	−	−	−

[a] 3-E, 3-ethylpentane; 3-M, 3-methylhexane; 2-M, 2-methylhexane; 2,3-MM, 2,3-dimethylpentane; 2,4-MM, 2,4-dimethylpentane; 3,3-MM, 3,3-dimethylpentane; 2,2-MM, 2,2-dimethylpentane; 2,2,3-MMM, 2,2,3-trimethylbutane.

Table 5. Values of ${}^{n}\Omega$ for the Isomers of Heptane

isomer	${}^{1}\Omega$	${}^{2}\Omega$	${}^{3}\Omega$	${}^{4}\Omega$
n-heptane	$2+\sqrt{2}$	$-1/6+1/6\sqrt{2}$	–	–
3-ethylpentane	$3/2\sqrt{2}+1/2\sqrt{6}$	$-3/8\sqrt{2}+1/2\sqrt{3}+3/8\sqrt{6}$	$3/10\sqrt{2}+3/5\sqrt{3}-3/10\sqrt{6}$	–
3-methylhexane	$1/2+\sqrt{2}+1/3\sqrt{3}+1/3\sqrt{6}$	$10/39-19/39\sqrt{2}+20/117\sqrt{3}+79/234\sqrt{6}$	$-93/307-15/614\sqrt{2}+183/614\sqrt{3}-15/307\sqrt{6}$	$-4/51+1/102\sqrt{2}-1/51\sqrt{3}+1/51\sqrt{6}$
2-methylhexane	$1+1/2\sqrt{2}+2/3\sqrt{3}+1/6\sqrt{6}$	$-7/36-7/72\sqrt{2}+1/27\sqrt{3}+47/216\sqrt{6}$	$-83/298+33/298\sqrt{2}+15/149\sqrt{3}-5/149\sqrt{6}$	–
2,3-dimethylpentane	$1/3+1/2\sqrt{2}+\sqrt{3}+1/6\sqrt{6}$	$57/68-31/408\sqrt{2}-31/204\sqrt{3}+103/408\sqrt{6}$	$85/1427+614/1427\sqrt{2}-199/1427\sqrt{3}-136/4281\sqrt{6}$	$-15/469+2/469\sqrt{2}-5/1407\sqrt{3}+8/469\sqrt{6}$
2,4-dimethylpentane	$4/3\sqrt{3}+1/3\sqrt{6}$	$1/6\sqrt{2}-2/9\sqrt{3}+4/9\sqrt{6}$	$16/25\sqrt{2}+8/225\sqrt{3}-16/225\sqrt{6}$	–
3,3-dimethylpentane	$1+3/2\sqrt{2}$	$-9/44+3/44\sqrt{2}$	–	–
2,2-dimethylpentane	$2+3/4\sqrt{2}$	$9/164-3/41\sqrt{2}$	–	–
2,2,3-trimethylbutane	$3/2+5/6\sqrt{3}$	$-15/52+9/52\sqrt{3}$	–	–

Table 6. Values of the Coefficients b_j

isomer	b_1	b_2	b_3	b_4
n-heptane	$164/5$	$-984/5$	–	–
3-ethylpentane	–	–	–	–
3-methylhexane	$184/13$	$5520/307$	$-11408/153$	-736
2-methylhexane	30	$-7560/149$	-180	–
2,3-dimethylpentane	$1347/170$	$460674/7135$	$22899/1876$	$-4041/4$
2,4-dimethylpentane	–	–	–	–
3,3-dimethylpentane	$861/55$	$-1722/5$	–	–
2,2-dimethylpentane	$6336/205$	$1584/5$	–	–
2,2,3-trimethylbutane	$7281/260$	$-809/6$	–	–

8. CONCLUSIONS

This paper is centered on the application of a general theory of orthogonal molecular descriptors that we have developed previously on rigorous mathematical grounds. We have shown that, in our approach, the orthogonal molecular descriptor can be constructed directly from the molecular structure−property relation based on those orthogonal descriptors. One important feature that emerges from these equations is that the coefficients of the mathematical relation are not constants, as in the case of Randić's orthogonalization approach.[1−4] However, we know the exact way that the parameters change; thus, no random feature is introduced by our approach. Furthermore, because we are using an orthogonal basis, any new descriptor introduced will be dependent on the others and its contribution to the property will be zero.

We have discussed the differences and similarities of our approach with that of Randić and have presented a manner in which Randić's approach could be made rigorous, from a mathematical point of view, although it could make the computation of the orthogonal descriptors more tedious.

REFERENCES AND NOTES

(1) Randić, M. Orthogonal molecular descriptors. *New J. Chem.* **1991**, *15*, 517−525.

(2) Randić, M. Search for optimal molecular descriptors. *Croat Chem. Acta* **1991**, *64*, 43−54.

(3) Randić, M. Generalized molecular descriptors. *J. Math. Chem.* **1991**, *7*, 155−168.

(4) Randić, M.; Trinajstić, N. Isomeric variations in alkanes: boiling points of nonanes. *New J. Chem.* **1994**, *18*, 179−189.

(5) Klein, D. J.; Randić, M.; Babić, D.; Lučić, B.; Nikolić, S.; Trinajstić, N. Hierarchical orthogonalization of descriptors. *Int. J. Quantum Chem.* **1997**, *63*, 215−222.

(6) Morales, D. A.; Araujo, O. On the search for the best correlation between graph theoretical invariants and physicochemical properties. *J. Math. Chem.* **1993**, *13*, 95−106.

(7) Araujo, O.; Morales, D. A. An alternative approach to orthogonal graph theoretical invariants. *Chem. Phys. Lett.* **1996**, *257*, 393−396.

(8) Araujo, O.; Morales, D. A. A theorem about the algebraic structure underlying orthogonal graph invariants. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1051−1053.

(9) Curtis, C. W. *Linear Algebra: An Introductory Approach*; Allyn and Bacon: Boston, 1974.

(10) The property space can be considered as the subspace of E^n, where $E = \mathrm{Q}(\sqrt{2}, \sqrt{3})$ generated by ${}^{i}\chi$, $i = 1, 2, 3, 4$, with ${}^{i}\chi \in E$ (see, section 6) and n is the number of compounds to model.

(11) The boiling point, as any other physical property, is always a rational because there is a finite accuracy to its measurement. The boiling points are measured with a thermometer, which has a finite accuracy. Nobody could say a priori that the boiling point of *n*-heptane is $(1737/25)\sqrt{2}$, for instance. One would have to have had on hand a thermometer with an infinite accuracy, an instrument that perhaps mankind will never have.

(12) In fact, this equation for b_j can be considered as a particular case of the following result. Let V denote a finite dimensional vector space

over the field of real numbers or the complex field with an inner product $\langle , \rangle$. Let $\{e_1, e_2, ..., e_n\}$ be an orthogonal basis for V. Then, every vector $v \in V$ can be expressed uniquely in the form $v = \sum_{i=1}^{n}(\langle v,e_i\rangle/\langle e_i,e_i\rangle)e_i$.

(13) Knowing, of course, the results of the correlation using nonorthogonal descriptors.

CI970062H