

Table of Periodic Properties of Fullerenes Based on Structural Parameters

Francisco Torrens*

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

Received August 26, 2003

The periodic table (PT) of the elements suggests that hydrogen could be the origin of everything else. *The construction principle is an evolutionary process that is formally similar to those of Darwin and Oparin.* The Kekulé structure count and permanence of the adjacency matrix of fullerenes are related to structural parameters involving the presence of contiguous pentagons p , q and r . Let p be the number of edges common to two pentagons, q the number of vertices common to three pentagons, and r the number of pairs of nonadjacent pentagon edges shared between two other pentagons. Principal component analysis (PCA) of the structural parameters and cluster analysis (CA) of the fullerenes permit classifying them and agree. A PT of the fullerenes is built based on the structural parameters, PCA and CA. The periodic law does not have the rank of the laws of physics. (1) The properties of the fullerenes are not repeated; only, and perhaps, their chemical character. (2) The order relationships are repeated, although with exceptions. The proposed statement is the following: *The relationships that any fullerene p has with its neighbor $p + 1$ are approximately repeated for each period.*

1. INTRODUCTION

The periodic table (PT) of the elements¹ suggests that hydrogen (H) could be the origin of everything else. In addition to H, i.e., to protons, there are neutrons, which can be considered as the element zero of such a PT. *The construction or aufbau principle is an evolutionary process applied to atomic species.* Inspired by the PT of the elements, several PTs can be suggested: (1) in physics, PTs of subatomic particles and atomic clusters, (2) in chemistry, relativistic effects² and PTs of molecules, e.g., alkanes,^{3,4} acyclic hydrocarbons,^{5,6} polycyclic aromatic hydrocarbons (PAH),^{7,8} benzenoid hydrocarbons,⁹ fullerenes,¹⁰ C₆₀ fullerenes¹¹ and drugs,¹² as well as organic conversions and drug design via a genetic algorithm,¹³ (3) in biochemistry, PTs of amino acids, nucleotides, proteins,^{14,15} metabolic pathways¹⁶ and networks, cellular signaling networks, phosphorylation pathways, apoptosis and disease,¹⁷ antibiotics selection and protein interaction maps,¹⁸ (4) in molecular biology, PTs of aminoacyl RNA synthetases, genomes,¹⁹ proteomes, carbohydrates, transcriptomes, metabolomes, lipidomes, *inborn errors of metabolism*,²⁰ molecular phylogenetic trees^{21,22} and genetic codes,^{23,24} (5) in biology, PTs of cells and embryonic stem cell lines, and (6) in pharmacy, pharmacogenomes. The efforts of chemists have uncovered many principles of molecular structure and simple chemical reactions. However, the molecules one is able to synthesize from scratch are quite simple. As the differences between chemistry and physics, and between chemistry and biology, disappear, multidisciplinary approaches are leading to the fascinating *new* area of nanoscience and nanotechnology, in which ingenious strategies for the creation of molecules with complex prespecified structures and novel function are being developed: molecules that do things.

Experimental results showed a direct correlation between carbon-ring consumption and fullerene formation in buffered high-temperature carbon vapors.²⁵ With the actual knowledge of carbon-structure formation kinetics, the evolution of carbon rings to form fullerenes involves the participation of intermediate carbon polycycles (CP). Since a general classification of CPs as a function of the number of atoms readily becomes extremely large, it is necessary to put some restrictions in the analysis of the most probable structures that can be present in reaction pathways. The basic assumptions for a classification must take into account that the most stable CPs should have the maximum number of bonds, minimum number of dangling bonds and minimum strain energy. These conditions cannot be simultaneously satisfied and may be approximated by the hypothesis that CPs have armchairlike borders and similar quantity of peripheral and interior C atoms to attain minimal strain-energy structures. CP selection also must be acknowledged for low-reactivity structures, which will not reduce their size by the interaction with other carbon species. Using these classification criteria, Pasqualini and López identified a manageable set of CPs.²⁶ When more than one structure was possible for a defined number of atoms, they performed ab initio energy calculations and identified the most stable structures. They selected the most probable CPs with between 14 and 35 atoms. They optimized the structures of neutral CPs using density functional theory with the hybrid functional B3LYP and the basis set 6-31G*. The resulting classification of CPs showed that they have a bowl-shaped appearance. The stability and characteristics of these CPs support the hypothesis that fullerenes are formed by the addition reaction of two of these structures. This conclusion is in accordance with the experimental discovery that fullerenes in flames are formed by the addition reaction of two PAHs.²⁷

In previous works, the calculation of Kekulé structure count and the permanent of adjacency matrices²⁸ was applied to fullerenes with different structural parameters involving the presence of contiguous pentagons.²⁹ Principal component analysis (PCA) of the structural parameters was performed.^{30–32} The aim of the present report is to analyze the interdependence between the structural parameters and to classify fullerenes. Section 2 presents the computational methods. Section 3 discusses the calculation results for the fullerenes and the recommended format for their PT, which is based on the structural parameters. Section 4 summarizes the conclusions.

2. COMPUTATIONAL METHODS

Often, in quantitative structure–property relationship (QSPR) studies, the data file contains less than one hundred objects and several thousands of X -variables. In fact, there are so many X -variables that no one can discover by inspection patterns, trends, clusters, etc. in the objects. *Principal components analysis* (PCA) is a technique extremely useful to summarize all the information contained in the \mathbf{X} -matrix and put it in a form understandable by human beings.³³ PCA works by decomposing the \mathbf{X} -matrix as the product of two smaller matrices \mathbf{P} and \mathbf{T} . The loading matrix (\mathbf{P}), with information about the variables, contains a few vectors, the so-called Principal Components (PC), which are obtained as linear combinations of the original X -variables. The score matrix (\mathbf{T}), with information about the objects, is such that each object is described in terms of their projections onto the PCs, instead of the original variables.

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E} \quad (1)$$

The information not contained in these matrices remains as an *unexplained X-variance* in a residual matrix (\mathbf{E}), which has exactly the same dimensionality as the original \mathbf{X} -matrix.

Each PC_i is a new coordinate expressed as a linear combination of the old features x_j :

$$\text{PC}_i = \sum_j b_{ij} x_j \quad (2)$$

The new coordinates PC_i s are called scores or factors, while coefficients b_{ij} are called loadings. The scores are ordered according to their information content with respect to the total variance among all objects. The score-score plots show the positions of compounds in the new coordinate system, while loading-loading plots show the position of features that represent compounds in the new coordinate system. PCs have two interesting properties. (1) They are extracted in decreasing order of importance. The first PC always contains more information than the second does, the second more than the third, etc. (2) Each PC is orthogonal to each other. There is absolutely no correlation between the information contained in different PCs.

Cluster analysis (CA) encompasses a number of different classification algorithms.³⁴ A general question in many areas of an inquiry is how to organize the observed data into meaningful structures. Conceptually, the approach used by CA to address this problem can well be described by the

saying *birds of a feather flock together*. Since its initial introduction, many CA algorithms have been invented. They belong to two categories: hierarchical cluster analysis (HCA) and nonhierarchical (partitioned) cluster analysis (NHCA).³⁵ HCA rearranges objects in a tree-structure. In HCA, either the database is divided successively until a predetermined number of clusters have been created, or members are successively grouped together until the predetermined number of clusters has been assembled. In either case, a *dendrogram* (binary tree) is created that maps N members in one cluster to N members in N clusters. In NHCA, a nearest-neighbor list is created and used to assemble members into related clusters. An example of this is the Jarvis-Patrick NHCA algorithm, which has been widely used to cluster chemical structures and structural databases. In HCA, each object is initially assumed a lone cluster. A distance matrix is built, generally calculating the Euclidean distance between all the objects and scanned for the minor values. The corresponding objects are clustered together and treated as a single cluster. Successive iterations lead to the total clustering of all objects, generating a dendrogram with the objects clustered together according to their similarity level.

Correct CA results rely on the following: (1) proper structure representation, (2) suitable data normalization and (3) carefully selected CA algorithms and proper parameter settings. Data normalization is the basis for comparing experiments with large series when experimental conditions may not be identical. Normalization ensures that the experimental quality of data is comparable. The commonly used normalization functions follow: (1) linear normalization: $x'_i = X'_{\min} + (X'_{\max} - X'_{\min})(x_i - X_{\min})/(X_{\max} - X_{\min})$, (2) ratio normalization: $x'_i = x_i / \sum_{i=1}^n |x_i|$, and (3) Z-score normalization: $x'_i = (x_i - \bar{x})/\delta$, where δ is the standard deviation.

3. CALCULATION RESULTS AND DISCUSSION

The structural features involving adjacent pentagons are encoded by the parameters p , q and r . The p and q enumerate, respectively, the number of edges common to two pentagons and the number of vertices common to three pentagons.³⁶ The variable r counts the number of pairs of nonadjacent pentagon edges shared with two other pentagons.³⁷ The q and r complement each other by counting both possible arrangements of three contiguous pentagons. However, there are close relationships between p and q as well as between p and r , e.g., the minimum structure with $q = 1$ needs $p = 3$, and the minimum structure with $r = 1$ requires $p = 2$. The values for $\{p, q, r\}$ are listed in Table 1. Much chemical graph-theory has revolved around the adjacency matrices \mathbf{A} . The determinant of the 3×3 matrix $[a \ b \ c, d \ e \ f, g \ h \ i]$ is $aei - ahf - dbi + dhc + gb f - gec$. The permanent of this matrix, $\text{per}(\mathbf{A})$, is the sum of the same six terms. A motivation for the consideration of the Kekulé structure count K is that K is never zero for fullerenes.³⁸ $\text{Per}(\mathbf{A})$ is bounded below by K^2 . As $\text{per}(\mathbf{A})$ and K increase exponentially with system size their logarithms are used. Cash selected a group of 27 fullerenes (in Table 1) to correlate $\ln[\text{per}(\mathbf{A})]/\ln K$, $\ln K$ and $\ln[\text{per}(\mathbf{A})]$ with $\{p, q, r\}$. Despite his good results, three remarks can be made. (1) The $\{p, q, r\}$ include some

Table 1. Values of p , q and r Counts for Fullerenes

fullerene	K	per(A)	$\ln[\text{per(A)}]/\ln K$	p	q	r	q/p	r/p
C ₂₀ (I _h)	36	1392	2.0199	30	20	30	0.6667	1.0000
C ₂₄ (D _{6d})	54	4692	2.1192	24	12	36	0.5000	1.5000
C ₂₆ (D _{3h})	63	8553	2.1853	21	8	30	0.3810	1.4286
C ₂₈ (T _d)	75	15705	2.2378	18	4	24	0.2222	1.3333
C ₂₈ (D ₂)	90	16196	2.1540	20	8	24	0.4000	1.2000
C ₃₀ (C _{2v}) I	107	29621	2.2034	17	4	20	0.2353	1.1765
C ₃₀ (C _{2v}) II	117	30053	2.1651	18	6	20	0.3333	1.1111
C ₃₀ (D _{5h})	151	31945	2.0672	20	10	20	0.5000	1.0000
C ₃₂ (D ₃)	144	55140	2.1968	15	2	18	0.1333	1.2000
C ₃₂ (C ₂) I	151	55705	2.1780	16	4	16	0.2500	1.0000
C ₃₂ (C ₂) II	168	57092	2.1375	17	6	16	0.3529	0.9412
C ₃₂ (D ₂)	184	58384	2.1045	18	8	15	0.4444	0.8333
C ₃₄ (C _{3v})	195	103665	2.1902	15	3	15	0.2000	1.0000
C ₃₄ (C _s)	196	104484	2.1896	15	3	16	0.2000	1.0667
C ₃₄ (C ₂) I	204	103544	2.1714	14	2	14	0.1429	1.0000
C ₃₄ (C ₂) II	212	107720	2.1632	17	6	16	0.3529	0.9412
C ₃₆ (D _{6h})	272	192528	2.1706	12	0	12	0.0000	1.0000
C ₃₆ (D _{2d})	288	192720	2.1489	12	0	12	0.0000	1.0000
C ₃₆ (C _{2v})	312	197340	2.1231	13	2	10	0.1538	0.7692
C ₃₆ (D _{3h})	364	207924	2.0764	15	6	6	0.4000	0.4000
C ₃₈ (C _{2v})	360	366820	2.1768	14	2	14	0.1429	1.0000
C ₃₈ (C _{3v})	378	363300	2.1572	12	1	9	0.0833	0.7500
C ₃₈ (D _{3h})	456	411768	2.1116	18	8	18	0.4444	1.0000
C ₄₀ (D _{5d}) I	562	515781	2.0775	10	0	10	0.0000	1.0000
C ₄₀ (T _d)	576	704640	2.1185	12	4	0	0.3333	0.0000
C ₄₀ (D _{5d}) II	701	803177	2.0750	20	10	20	0.5000	1.0000
C ₄₄ (T)	864	2478744	2.1775	12	4	0	0.3333	0.0000
C ₄₄ (D _{3h})	960	2436480	2.1416	9	2	0	0.2222	0.0000
C ₆₀ (I _h)	12500	395974320	2.0986	0	0	0		
C ₇₀ (D _{5h})	52168			0	0	0		
C ₇₂ (D _{6d})				0	0	0		
C ₇₄				0	0	0		
C ₇₆ (D ₂)				0	0	0		
C ₇₈ (D _{3h})				0	0	0		
C ₈₀ (I _h)	140625			0	0	0		
C ₈₀ (D ₂)				0	0	0		
C ₈₂ (C _s)				0	0	0		
C ₈₄ (D ₂)				0	0	0		
C ₈₆ (C ₂)				0	0	0		
C ₈₈ (D ₂)				0	0	0		
C ₉₀ (C _{2v})				0	0	0		
C ₉₂ (C ₂)				0	0	0		
C ₉₄ (C _s)				0	0	0		
C ₉₆ (C _{3h})				0	0	0		
C ₉₈				0	0	0		
C ₁₀₀				0	0	0		
C ₁₂₀ (I _h)	21956126972			0	0	0		
C ₁₄₀ (I _h)	2167239697			0	0	0		
C ₁₈₀ (I _h)	1389029765625			0	0	0		
C ₂₄₀ (I _h)	21587074966666816			0	0	0		

redundant information. (2) The error of some fitted parameters is large. (3) Nonlinear effects of $\{p,q,r\}$ can affect $\ln[\text{per(A)}]/\ln K$, $\ln K$ or $\ln[\text{per(A)}]$. Therefore, a different strategy is used. (1) The $\{p,q,r\}$ is expanded with the ratios $\{q/p, r/p\}$. (2) Smaller superpositions of (p,q) and of (p,r) are sought. (3) Not all the counts $\{p,q,r\}$ are necessarily retained. (4) Nonlinear correlations are allowed.

The best linear correlation of $\ln[\text{per(A)}]/\ln K$ with $\{p,q,r\}$ results

$$\ln[\text{per(A)}]/\ln K = 2.14 - 0.0108q + 0.00364r$$

$$n = 29; R = 0.721; s = 0.036; F = 14.1; \text{MAPE} = 1.21\%; \text{AEV} = 0.4803 \quad (3)$$

where MAPE is the mean absolute percentage error and AEV

is the approximation error variance. There are several fullerenes with the same set $\{p,q,r\}$. Equation 3 explains 95% of the correlation coefficient of the means ($n = 24$, $R = 0.757$). There are general *degeneracy* problems with trying to fit per(A) and K with $\{p,q,r\}$. Whole families of fullerenes have the same values for $\{p,q,r\}$, yet with widely varying values of per(A) and K , e.g., bucky-tubes with fixed fullerene caps but of varying length are fullerenes all with the same values of $\{p,q,r\}$, while the values of per(A) and K increase without bound as the length of the tubes are increased. Many large fullerenes have $p = q = r = 0$, although the values for per(A) and K increase exponentially with the number of sites N of the fullerene. As N has not been included in the correlations, the applicability of the fits is restricted to smaller fullerenes ($N < 70$). On the other hand, the best quadratic correlation of $\ln[\text{per(A)}]/\ln K$ with

quadratic functions of $\{p, q, r\}$ gives

$$\begin{aligned}\ln[\text{per}(\mathbf{A})]/\ln K &= 2.13 + 0.0515z_{41} \\ z_{41} &= 0.225z_{31} + 1.20z_{32} \\ z_{31} &= -1.16 + 0.232q \\ z_{32} &= 1.05z_{22} - 0.875z_{21}z_{22} \\ z_{21} &= 1.22 - 0.0983r + 0.00277qr \\ z_{22} &= -0.726z_{11} - 0.921z_{12} \\ z_{11} &= -1.16 + 0.232q \\ z_{12} &= 1.22 - 0.0983r + 0.00277qr \quad (4) \\ \text{MAPE} &= 0.87\%; \text{AEV} = 0.2432\end{aligned}$$

and AEV decreases 49%. No superposition of $p-q$ or $p-r$ is observed in eqs 3 and 4. This diminishes the risk of collinearity in the fit given the close relationships $p-q$ and $p-r$.³⁹ If $\{q/p, r/p\}$ are included in model 4, the best linear fit results

$$\begin{aligned}\ln[\text{per}(\mathbf{A})]/\ln K &= 1.88 + 0.0361p - 0.0490q + \\ &\quad 0.00953r + 0.0497q/p - 0.253r/p \\ n &= 28; R = 0.941; s = 0.019; F = 34.2; \text{MAPE} = \\ &\quad 0.66\%; \text{AEV} = 0.1558 \quad (5)\end{aligned}$$

and AEV decreases 68%. Equation 5 explains 98% of the R of the means ($n = 23, R = 0.956$). The best quadratic model does not improve the results.

There are powerful exact computational approaches for K , which are fairly reasonable or general. For arbitrary chemical graphs, enumeration via Heilbronner recursion is feasible up to ~ 90 atoms. Better efficiency is found with Kasteleyn's method, which is applicable to all planar graphs. This involves the evaluation of the determinant of a *signed* adjacency matrix \mathbf{A}' ,⁴⁰ where the method is neatly extended to deal with conjugated circuit counts, using the inverse of \mathbf{A}' . The method was applied to fullerenes of up to 980 atoms⁴¹ and infinite translationally symmetric networks.⁴² For $\ln K$ alone, the best linear correlation results are

$$\begin{aligned}\ln K &= 20.3 - 1.45p + 0.941q + 0.243r \\ n &= 35; R = 0.811; s = 4.630; F = 19.8; \text{MAPE} = \\ &\quad 24.31\%; \text{AEV} = 0.3426 \quad (6)\end{aligned}$$

Equation 6 explains 83% of the R of the means ($n = 24, R = 0.975$). Quadratic models do not improve the results. If $\{q/p, r/p\}$ are included in the model, the best linear fit results are

$$\begin{aligned}\ln K &= 10.2 - 0.407p + 0.238q + 0.792q/p + 0.320r/p \\ n &= 28; R = 0.931; s = 0.339; F = 28.7; \text{MAPE} = \\ &\quad 3.77\%; \text{AEV} = 0.1627 \quad (7)\end{aligned}$$

and AEV decreases 53%. Equation 7 explains 97% of the R

of the means ($n = 23, R = 0.957$). The best quadratic correlation gives

$$\begin{aligned}\ln K &= 9.09 - 0.245p - 0.00449p^2 + \\ &\quad 0.293p \cdot q/p \quad \text{MAPE} = 3.97\%; \text{AEV} = 0.1524 \quad (8)\end{aligned}$$

and AEV decreases 56%.

For $\ln[\text{per}(\mathbf{A})]$ alone, the best linear correlation results are

$$\begin{aligned}\ln[\text{per}(\mathbf{A})] &= 20.2 - 0.660p + 0.383q \\ n &= 29; R = 0.949; s = 0.757; F = 118.5; \text{MAPE} = \\ &\quad 4.05\%; \text{AEV} = 0.0988 \quad (9)\end{aligned}$$

Equation 9 explains 97% of the R of the means ($n = 24, R = 0.977$). The best quadratic correlation gives

$$\begin{aligned}\ln[\text{per}(\mathbf{A})] &= 20.0 - 0.666p + 0.616q - \\ &\quad 0.00850pq \quad \text{MAPE} = 3.91\%; \text{AEV} = 0.0871 \quad (10)\end{aligned}$$

and AEV decreases 12%. The inclusion of $\{q/p, r/p\}$ does not improve the results.

In modeling it is essential to determine the complexity of the model to avoid overfitting. The predictive capability of the resulting model depends on the quality of the data (the more and better the data available, the more accurate prediction is possible) and on the number k of significant latent variables necessary. Cross-validation is a practical and reliable method for testing this significance. In principle, the so-called *leave-one-out* approach consists of developing a number of models with one sample omitted at the time. After developing each model, the omitted data are predicted, and the differences between actual and predicted $y \equiv \ln[\text{per}(\mathbf{A})]/\ln K$ values are calculated. The performance of the model (its predicted ability) can be given by Pearson's correlation coefficient between the experimental and predicted y values. This cross-validated correlation coefficient is denoted by R_{cv} . Leave-one-out selects the best set of descriptors according to the criterion of maximization of R_{cv} . The procedure can be generalized omitting many (n) samples at the time, in the leave-many-out (leave- n -out) procedure.⁴³ The R_{cv} ($1 \leq n \leq 10$) have been calculated. In particular, for $\ln[\text{per}(\mathbf{A})]/\ln K$ and $\ln K$, the sets $\{p, q, r, q/p, r/p\}$ and $\{p, q, q/p, r/p\}$, respectively, give the greatest R_{cv} for the whole range of n . However, for $\ln[\text{per}(\mathbf{A})]$, $\{p, q\}$ gives the greatest R_{cv} . The corresponding interpretation is that $\{p, q, r, q/p, r/p\}$ is more predictive than $\{q, r\}$ or $\{p, q, r\}$ for modeling $\ln[\text{per}(\mathbf{A})]/\ln K$, that $\{p, q, q/p, r/p\}$ is more predictive than $\{p, q, r\}$ for $\ln K$, and that $\{p, q\}$ is more predictive than $\{p, q, r\}$ for $\ln[\text{per}(\mathbf{A})]$.

The upper triangle of the 3×3 symmetrical correlation matrix \mathbf{R} for $\{p, q, r\}$ results is

$$\mathbf{R} = \begin{pmatrix} 1.000 & 0.831 & 0.892 \\ & 1.000 & 0.742 \\ & & 1.000 \end{pmatrix}$$

High correlation is observed for the pairs $p-r$ and $p-q$. The correlation increases in the order $R_{qr} \ll R_{pq} < R_{pr}$. The upper triangle of the 5×5 correlation matrix \mathbf{R} for

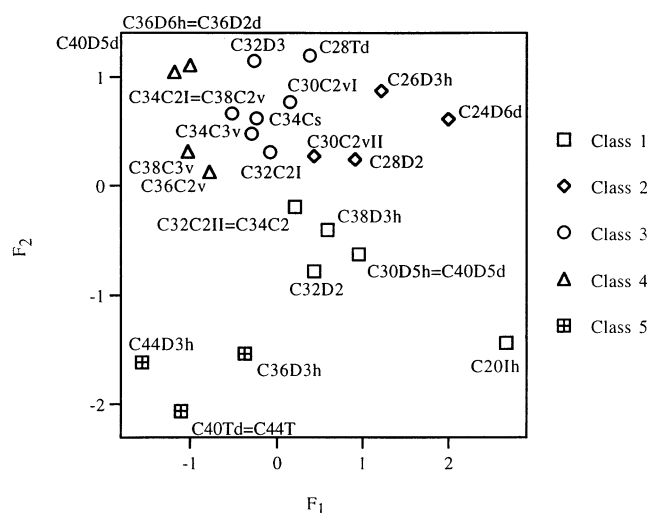


Figure 1. PCA F_2 vs F_1 plot for the fullerenes.

$\{p, q, r, q/p, r/p\}$ results is

$$R = \begin{pmatrix} 1.000 & 0.929 & 0.857 & 0.805 & 0.542 \\ & 1.000 & 0.635 & 0.934 & 0.225 \\ & & 1.000 & 0.457 & 0.875 \\ & & & 1.000 & 0.029 \\ & & & & 1.000 \end{pmatrix}$$

High correlation is obtained for $(q, q/p)$, (p, q) , $(r, r/p)$ and for (p, r) . The correlation between q/p and r/p (0.029) is 20 times smaller than that between q and r (0.635). The correlation increases in the order $R_{qr} \ll R_{pr} < R_{pq}$. In the PCA PC_2 – PC_1 (F_1 – F_2) plot (Figure 1), the fullerenes with the same set of $\{p, q, r, q/p, r/p\}$ values, belonging to classes 1, 3, 4 and 5, appear superposed. Five classes of fullerenes are clearly distinguished: class 1 with seven-members ($F_1 \gg F_2$, middle right of Figure 1), class 2 with 4 units ($F_1 > F_2 > 0$, top right), class 3 with 8 units ($F_1 < F_2$, top), class 4 with 5

units ($F_1 \ll F_2$, top left) and class 5 with 4 units ($0 > F_1 > F_2$, bottom left). In general, fullerene isomers belong to the same class. The exceptions are the isomers of C_{28} , C_{30} , C_{32} , C_{34} , C_{36} , C_{38} and C_{40} , which fit in two or three classes. To classify the last 22 fullerenes in Table 1, PCA was repeated with $\{p, q, r\}$, which grouped them in class 5 close to C_{44} (D_{3h}).

The radial tree for the fullerenes relating to $\{p, q, r, q/p, r/p\}$ in Figure 2 separates first the 7 fullerenes in class 1 [C_{20} (I_h)– C_{34} (C_2) II, Figure 2 bottom right], then the 4 fullerenes in class 2 [C_{24} (D_{6d})– C_{30} (C_{2v}) II, bottom], the 8 fullerenes in class 3 [C_{28} (T_d)– C_{38} (C_{2v}), left], the 5 fullerenes in class 4 [C_{40} (D_{5d}) I– C_{38} (C_{3v}), top] and the 4 fullerenes in class 5 [C_{36} (D_{3h})– C_{44} (T), right]. The classes correspond to PCA (Figure 1). To classify C_{60} – C_{240} , the radial tree was repeated for $\{p, q, r\}$, resulting in their inclusion in a new branch connected to C_{44} (D_{3h}). C_{70} – C_{240} appear superposed to C_{60} .

The recommended format for a PT of fullerenes is presented in Table 2. Fullerenes are classified first by p , then by q and, finally, by r . Periods of 10 p units are assumed. Group p_0 stands for $p = 0, 10, \dots$, subgroup q_0 for $q = 0, 10, \dots$, and subsubgroup r_0 for $r = 0, 10, \dots$. The fullerenes in the same column (group) of Table 2 appear close in PCA and tree (Figures 1 and 2).

The variation of $P \equiv \ln[\text{per}(\mathbf{A})]/\ln K$ vs p in Figure 3 suggests discontinuities for fullerenes with p values of 0, 10, 20 and 30 (group p_0). The distance in p units between each pair of consecutive discontinuities is 10, a value that coincides with those given previously for the fullerene sets belonging to the same period in PT. Splitting the p range by these discontinuities, the three periods in PT can be fitted to parabolic relationships. The first period can be interpolated as

$$P = 2.10 + 0.0668p - 0.00689p^2 \quad (11)$$

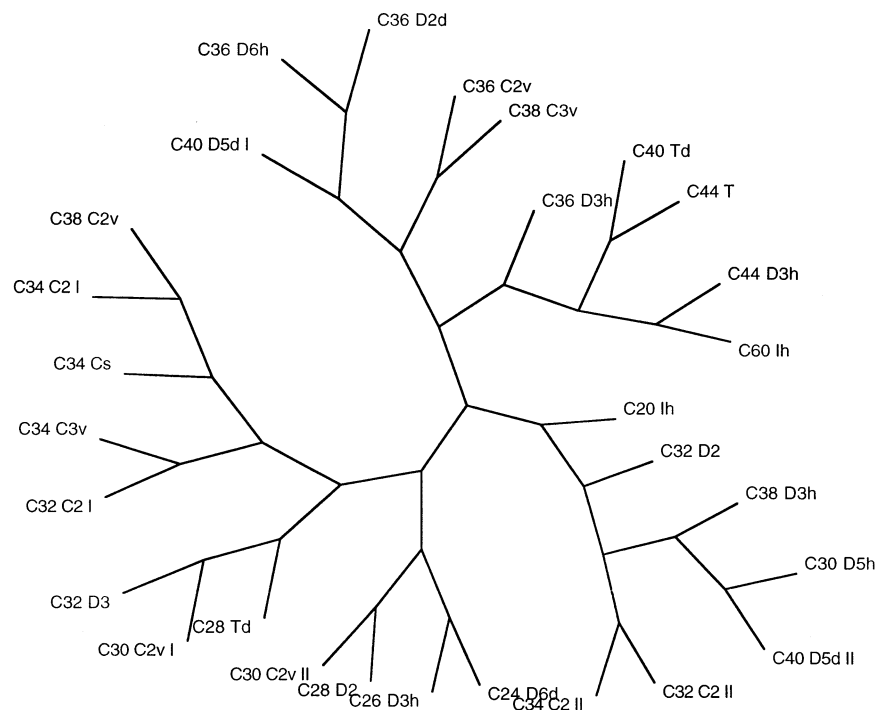
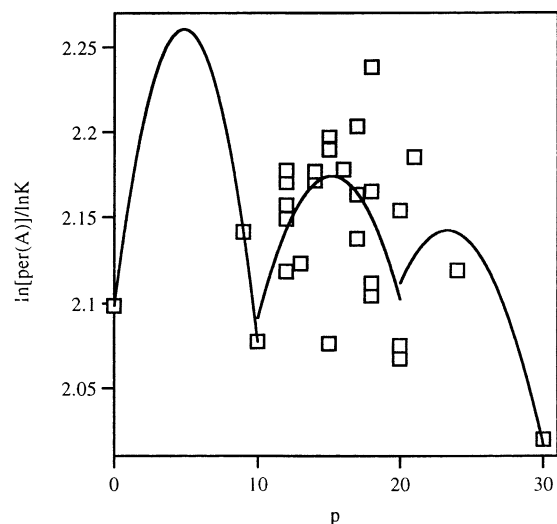


Figure 2. Radial tree graph cluster analyzing for the fullerenes.

Table 2. Recommended Format for the Table of Periodic Properties of Fullerenes Based on Structural Parameters

I																			
p0		p1		p2		p3		p4		p5		p6		p7		p8		p9	
q0	q8	q8	q0	q1	q4	q2		q2		q3	q6	q4	q4	q6	q4	q6		q8	q2
r0	r4	r0	r2	r9	r0	r0	r4	r8	r5	r6	r6	r6	r0	r6	r4	r0	r5	r8	r0
60I _h ^a																			44D _{3h}
40D _{5d} I			36D _{6h}	38C _{3v}	40T _d	36C _{2v}	34C ₂ I	32D ₃	34C _{3v}	34C _s	36D _{3h}	32C ₂ I	30C _{2v} I	32C ₂ II	28T _d	30C _{2v} II	32D ₂	38D _{3h}	
			36D _{2d}		44T		38C _{2v}							34C ₂ II					
30D _{5h}	28D ₂	26D _{3h}						24D _{6d}											
40D _{5d} II																			
20I _h																			

^a 60I_h, 70D_{5h}, 72D_{6d}, 74, 76D₂, 78D_{3h}, 80I_h, 80D₂, 82C_s, 84D₂, 86C₂, 88D₂, 90C_{2v}, 92C₂, 94C_s, 96C_{3h}, 98, 100, 120I_h, 140I_h, 180I_h, 240I_h.

**Figure 3.** Variation of $\ln[\text{per}(\text{A})]/\ln K$ vs count p .

The second period can be fitted as

$$P = 2.09 + 0.0320(p - 10) - 0.00309(p - 10)^2 \quad n = 24 \quad R = 0.553 \quad s = 0.040 \quad F = 4.6 \quad (12)$$

If the different repetitions of P are substituted by their mean, a maximum correlation coefficient of 0.920 is obtained. Therefore eq 12 with $R = 0.553$ explains 60% of the correlation coefficient of the means. The third period can be fitted as

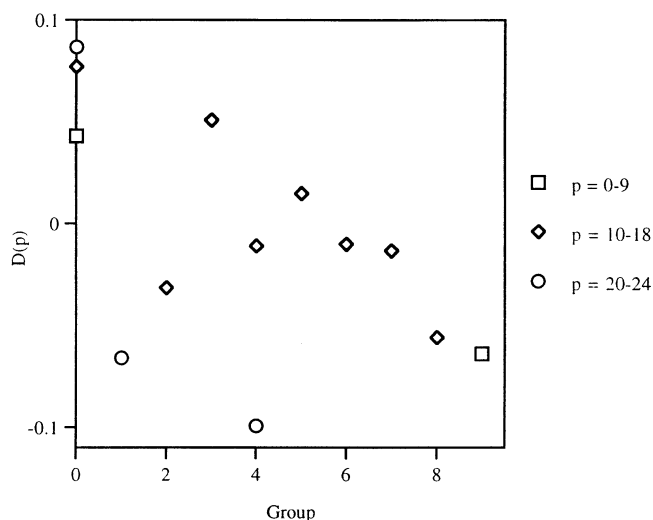
$$P = 2.11 + 0.0185(p - 20) - 0.00279(p - 20)^2 \quad n = 6 \quad R = 0.720 \quad s = 0.055 \quad F = 1.6 \quad (13)$$

Equation 13 explains 83% of the correlation coefficient of the means (0.865). The signs and magnitudes of the coefficients in eqs 11–13 are of some interest. There is agreement in the signs of the coefficients of the three fits. Moreover, there is some consistency in the values of the coefficients: 2.10 ± 0.01 , 0.0391 ± 0.0277 and -0.00426 ± 0.00263 .

If a function $P(p)$ is assumed a minimum value of such a function $P(p)$, $P_{\min}(p)$ for a given value of p has meaning only if it is compared with those of the former $P(p-1)$ and later $P(p+1)$ points, needing to fulfill the following conditions:

$$\begin{aligned} P_{\min}(p) &< P(p-1) \\ P_{\min}(p) &< P(p+1) \end{aligned} \quad (14)$$

Conditions 14 are, in fact, order relationships, being these relationships precisely what, the periodic law (PL) being

**Figure 4.** Variation of $D(p) = P(p+1) - P(p)$ vs group number; $P \equiv \ln[\text{per}(\text{A})]/\ln K$.

fulfilled rigorously, should repeat at determined intervals and equal to the values of the size of the periods. Relationships 14 are equivalent to the following ones:

$$\begin{aligned} P_{\min}(p) - P(p-1) &< 0 \\ P(p+1) - P_{\min}(p) &> 0 \end{aligned} \quad (15)$$

As expressions 15 are valid only for the minima of $P(p)$, more general others are desired for all the values of p . For that, it is considered adequate to calculate the differences $P(p+1) - P(p)$, assigning each of their values to fullerene p . Naming this value $D(p)$, one has

$$D(p) = P(p+1) - P(p) \quad (16)$$

Instead of $D(p)$ the values $R(p) = P(p+1)/P(p)$ can be taken assigning them to fullerene p . If PL were general, the elements belonging to the same group will satisfy, without exceptions, an identical relationship, i.e., they will fulfill:

$$D(p) > 0 \text{ or } D(p) < 0 \quad (17)$$

$$R(p) > 1 \text{ or } R(p) < 1 \quad (18)$$

However, the results show that this is not the case, so that PL is not general, existing with some anomalies, e.g., the variation of $D(p)$ vs group number in Figure 4 presents the lack of coherence between the Cartesian and PT representations of $P \equiv \ln[\text{per}(\text{A})]/\ln K$. If coherence were rigorous, all the points in each period will have the same sign. There is a tendency in the points to give $D(p) < 0$, especially for the greater groups. However, in detail, there are anomalies that

are manifested, in which the representative points of the different fullerenes for the successive periods are not always in phase.

There remains the question of what simple structural characteristic might be fitted to per(A) for larger fullerenes, where there are many with presumably diverse values of per(A) although still $p = q = r = 0$. A possible candidate may be the pyramidalization angle θ averaged for the C atoms, which hinders the aromaticity of the fullerenes and decreases from C₆₀–I_h (12.0°), C₇₀–D_{5h} (11.2°), C₈₀–I_h (10.4°), C₉₀–C_{2v} (10.0°), C₁₀₀ (9.6°), C₂₄₀–I_h (5.8°), C₅₀₀–I_h (4.0°) to C₁₀₀₀ (2.8°).

4. CONCLUSIONS

From the preceding results the following conclusions can be drawn.

1. Several criteria have been selected to reduce the analysis to a manageable quantity of structures from the enormous set of fullerene isomers. They refer to the structural parameters related with the presence of contiguous pentagons $\{p, q, r\}$, which destabilize the structures.

2. A thorough analysis of each case fullerene must be performed to check for the most stable and least reactive possible structure.⁴⁴

3. Despite its importance in the systematization of chemistry, PL has not the rank of the laws of physics. (1) The properties of the fullerenes are not repeated; only, and perhaps, their chemical character. (2) The order relationships are repeated, although with exceptions. The analysis forces the following statement: *The relationships that any fullerene p has with its neighbor p + 1 are approximately repeated for each period.* Periodicity is not general; however, if a natural order of the fullerenes is accepted, the underlying law must be phenomenological.

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) Mendeleev, D. *The Principles of Chemistry*; Longmans and Green: London, 1905.
- (2) Pyykkö, P. Relativistic Effects in Structural Chemistry. *Chem. Rev.* **1988**, *88*, 563–594.
- (3) Randić, M.; Wilkins, C. L. Graph Theoretical Ordering of Structures as a Basis for Systematic Searches for Regularities in Molecular Data. *J. Phys. Chem.* **1979**, *83*, 1525–1540.
- (4) Randić, M.; Wilkins, C. L. Graph-Theoretical Analysis of Molecular Properties. Isomeric Variations in Nonanes. *Int. J. Quantum Chem.* **1980**, *18*, 1005–1027.
- (5) Bytautas, L.; Klein, D. J. Formula Periodic Table for the Isomer Classes of Acyclic Hydrocarbons – Enumerative and Asymptotic Characteristics. *Croat. Chem. Acta* **2000**, *73*, 331–357.
- (6) Bytautas, L.; Klein, D. J.; Schmalz, T. G. All Acyclic Hydrocarbons: Formula Periodic Table and Property Overlap Plots via Chemical Combinatorics. *New J. Chem.* **2000**, *24*, 329–336.
- (7) Dias, J. R. A Periodic Table for Polycyclic Aromatic Hydrocarbons. *Acc. Chem. Res.* **1985**, *18*, 241–248.
- (8) Dias, J. R. A Formula Periodic Table for Benzenoid Hydrocarbons and the Aufbau and Excised Internal Structure Concepts in Benzenoid Enumerations. *J. Math. Chem.* **1990**, *4*, 17–29.
- (9) Klein, D. J.; Seitz, W. A. Pauling–Wheland Resonance Theory of Benzenoid Hydrocarbons. *J. Mol. Struct. (THEOCHEM)* **1988**, *169*, 167–181.
- (10) Liu, X.; Klein, D. J.; Schmalz, T. G.; Seitz, W. A. Generation of Carbon-Cage Polyhedra. *J. Comput. Chem.* **1991**, *12*, 1252–1259.
- (11) Liu, X.; Klein, D. J.; Seitz, W. A.; Schmalz, T. G. Sixty-Atom Carbon Cages. *J. Comput. Chem.* **1991**, *12*, 1265–1269.
- (12) Corriou, J. P.; Iordache, O.; Tondeur, D. Classification of Biomolecules by Information Entropy. *J. Chim. Phys. Phys.-Chim. Biol.* **1991**, *88*, 2645–2652.
- (13) De Julián-Ortiz, J. V. Virtual Darwinian Drug Design: QSAR Inverse Problem, Virtual Combinatorial Chemistry, and Computational Screening. *Comb. Chem. High Throughput Screen.* **2001**, *4*, 295–310.
- (14) *Proteins*; Funatsu, M., Hiromi, K., Imahori, K., Murachi, T., Narita, K., Eds.; Halsted: New York, 1972.
- (15) Tanford, C.; Reynolds, J. *Nature's Robots*; Oxford University Press: New York, 2001.
- (16) Nicholson, D. E. *Metabolic Pathways Map*; Sigma Chemical: St. Louis, MO, 1997.
- (17) Beere, H. M.; Green, D. R. *Apoptosis and Disease*; Elsevier: Amsterdam, 2002.
- (18) McCraith, S.; Holtzman, T.; Moss, B.; Fields, S. Genome-Wide Analysis of Vaccinia Virus Protein–Protein Interactions. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 4879–4884.
- (19) Kay, L. E. *The Molecular Vision of Life*; Oxford University Press: Nueva York, 1993.
- (20) Nicholson, D. E. *Inborn Errors of Metabolism*; Sigma Chemical: St. Louis, MO, 1998.
- (21) Brown, S. M. *Bioinformatics: A Biologist's Guide to Biocomputing and the Internet*; Eaton: Natick, MA, 2000.
- (22) *BioComputing: Computer Tools for Biologists*; Brown, S. M., Ed.; Eaton: Natick, MA, 2003.
- (23) Nirenberg, M.; Leder, P. RNA Code Words and Protein Synthesis. *Science* **1964**, *145*, 1399–1407.
- (24) Mehl, R. A.; Anderson, J. C.; Santoro, S. W.; Wang, L.; Martin, A. B.; King, D. S.; Horn, D. M.; Schultz, P. G. Generation of a Bacterium with a 21 Amino Acid Genetic Code. *J. Am. Chem. Soc.* **2003**, *125*, 935–939.
- (25) Von Helden, G.; Hsu, M.-T.; Kemper, P. R.; Bowers, M. T. Structures of Carbon Cluster Ions from 3 to 60 Atoms: Linear to Rings to Fullerenes. *J. Chem. Phys.* **1991**, *95*, 3835–3837.
- (26) Pasqualini, E. E.; López, M. Bowl Shaped Precursors of Fullerenes. In *Fullerenes*; Kamat, P. V., Guldi, D. M., Kadish, K. M., Eds.; ECS Proc. No. 2002-12, Electrochemical Society: Pennington, NJ, 2002; pp 805–814.
- (27) Ahrens, J.; Bachmann, M.; Baum, Th.; Griesheimer, J.; Kovacs, R.; Weilmünster, P.; Homann, K.-H. Fullerenes and Their Ions in Hydrocarbon Flames. *Int. J. Mass Spectrom. Ion Processes* **1994**, *138*, 133–148.
- (28) Torrens, F. Computing the Kekulé Structure Count for Alternant Hydrocarbons. *Int. J. Quantum Chem.* **2002**, *88*, 392–397.
- (29) Torrens, F. Computing the Permanent of the Adjacency Matrix for Fullerenes. *Internet Electron. J. Mol. Des.* **2002**, *1*, 351–359.
- (30) Torrens, F. Principal Component Analysis of Structural Parameters for Fullerenes. *Internet Electron. J. Mol. Des.* **2003**, *2*, 96–111.
- (31) Torrens, F. New Structural Parameters of Fullerenes for Principal Component Analysis. *Theor. Chem. Acc.* **2003**, *110*, 371–376.
- (32) Torrens, F. Principal Component Analysis of New Structural Parameters for Fullerenes. *Internet Electron. J. Mol. Des.* **2003**, *2*, 546–563.
- (33) Hotelling, H. Analysis of a Complex of Statistical Variables into Principal Components. *J. Educ. Psychol.* **1933**, *24*, 417–441.
- (34) Tryon, R. C. *J. Chronic Dis.* **1939**, *20*, 511–524.
- (35) Jarvis, R. A.; Patrick, E. A. Clustering Using a Similarity Measure Based on Shared Nearest Neighbors. *IEEE Trans. Comput.* **1973**, *C22*, 1025–1034.
- (36) Liu, X.; Klein, D. J.; Schmalz, T. G.; Seitz, W. A. Generation of Carbon-Cage Polyhedra. *J. Comput. Chem.* **1991**, *12*, 1252–1259.
- (37) Cash, G. G. Permanents of Adjacency Matrices of Fullerenes. *Polycycl. Arom. Compounds* **1997**, *12*, 61–69.
- (38) Klein, D. J.; Liu, X. Theorems for Carbon Cages. *J. Math. Chem.* **1992**, *11*, 199–205.
- (39) Hocking, R. R. The Analysis and Selection of Variables in Linear Regression. *Biometrics* **1976**, *32*, 1–49.
- (40) Klein, D. J.; Liu, X. Many-Body Conjugated-Circuit Computations. *J. Comput. Chem.* **1991**, *12*, 1260–1264.
- (41) Klein, D. J.; Zhu, H.; Valenti, R.; Garcia-Bach, M. A. Many-Body Valence-Bond Theory. *Int. J. Quantum Chem.* **1997**, *65*, 421–438.

- (42) Zhu, H.; Balaban, A. T.; Klein, D. J.; Živković, T. P. Conjugated-Circuit Computations on Two-Dimensional Carbon Networks. *J. Chem. Phys.* **1994**, *101*, 5281–5292.
- (43) Besalú, E. Fast Computation of Cross-Validated Properties in Full Linear Leave-Many-Out Procedures. *J. Math. Chem.* **2001**, *29*, 191–203.
- (44) Pasqualini, E. E.; López, M. Fullerene Formation Kinetics in Carbon Vapors. In *Fullerenes*; Kamat, P. V., Guldi, D. M., Kadish, K. M., Eds.; ECS Proc. No. 2002-12, Electrochemical Society: Pennington, NJ, 2002; pp 815–820.

CI030029X