New molecular descriptors based upon the Euler equations for chemical graphs

Pablo R. Duchowicz,* Michael J. Bucknum, and Eduardo A. Castro

INIFTA, División Química Teórica, Departamento de Química, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, Diag. 113 y 64, Suc. 4, C.C. 16, (1900) La Plata, Argentina
E-mail: pabloducho@yahoo.com.ar

The Euler equations for the chemical graphs are an extension of such equations for the polyhedra. These equations admit several potential forms of molecular descriptors that can be used in the characterizations of the properties of polycyclic aromatic hydrocarbons (PAH) in a typical Quantitative Structure Property–Activity Relationship (QSPR-QSAR). In this paper we describe the nature of these Euler relations for hydrocarbon graphs and the descriptors they admit, applying them to predict 37 boiling points (BP), 26 n-octanol/water partition coefficients ($\log(k_{ow})$), and 47 retention time indexes (RI) for reversed-phase liquid chromatography analysis. Final results suggest that these new descriptors can be used to complement others in a QSPR-QSAR study.

KEY WORDS: QSPR theory, polycyclic aromatic hydrocarbons, Euler equations, Schlaefli indexes

1. Introduction

The polyhedra are described in one way by their symmetry properties. Such symmetry analyses of the polyhedra have proven to be useful for the elucidation of the electronic structure of molecular fragments, or molecules, that possess vestiges of the shape and symmetry of the various polyhedra. These symmetry aspects of molecular electronic structure have been elucidated, in one particular example of the subject, by F.A. Cotton in his classic text on group theory applied to molecules which was published originally in the 1960's [1]. Alternatively, the topology of the various polyhedra, taken in the form of a relationship between their numbers of vertices, V, edges, E, and faces, F, has already been cast by Euler in his famous 1758 paper to the Saint Petersburg Academy [2].

The familiar Euler relation, which is applicable to any of the innumerable polyhedra, reads as follows:

$$V - E + F = 2. \tag{1}$$

*Corresponding author.

It is said from this result that the Euler characteristic for the sphere, from which all of the polyhedra are derived, is just equal to 2.

One can identify other topological parameters, in addition to the primary ones of V, E, and F that can be derived from these primary parameters, and that furnish additional useful information. Therefore, one can identify the polygonality, n, of the various polyhedra, that is defined as the averaged number of sides of the polygonal faces in the polyhedron considered. This is defined mathematically as

$$n = \frac{2E}{F},$$ (2)

and because each edge is shared by two faces in all of the polyhedra, this relationship is rigorous for them. In a similar way, one can identify a characteristic connectivity, p, of the polyhedra, that is defined as the averaged number of edges (or sides) meeting at each polygonal vertex of a given polyhedron. In analogy to n, the value of the connectivity is given

$$p = \frac{2E}{V},$$ (3)

and because each edge terminates at two vertices in each and every polyhedron, this relationship is rigorous over all of them as well.

The German mathematician, Schlaefli, pointed out these definitions in the 19th century [3–16]. The significance of these so-called Schlaefli symbols (n, p), the ordered pair of numbers for each polyhedron formed from the values of the polygonality, n, and the connectivity, p, for each of them, was stated by him through the so-called Schlaefli relationship shown below in equation (4). This Schlaefli relationship, can be derived from the Euler relation, shown in (1), by simple substitution

$$\frac{1}{n} - \frac{1}{2} + \frac{1}{p} = \frac{1}{E}.$$ (4)

Quite some time later, in the 1970's, it was the crystallographer Wells who pointed out a further significance to the Schlaefli symbols (n, p) as an ordered pair of numbers that specified an identity and relative location for the mapping of all of the innumerable polyhedra in a kind of Cartesian space called a Schlaefli space. The Schlaefli space is the space of n and p. Wells mapping of some of the polyhedra, termed the regular polyhedra, or the Platonic solids, which he described in his important monograph of 1977 [15], is shown below in figure 1.

This new topological mapping of these objects, has become increasingly important in the characterization of the various polyhedra, as their scope has been vastly expanded by the discovery of the innumerable 3-connected fullerenes (n ranges between 5 and 6, while $p = 3$ for the various fullerenes)[17], and the innumerable, so-called topologically irregular polyhedra [18–21] that possess

	p						
n	3	4	5	6	7	8	$\cdots$
3	t	o	i	(3, 6)			
4	c	(4, 4)					
5	d						
6	(6, 3)						
7							
8							
:							

3D Polyhedra and 3D nets

Figure 1. Partial topology map of the polyhedra and 2D regular structures.

more than one type of polygonal vertex, and simultaneously more than one type of polygonal face. This newly discovered class of polyhedra have been called the Wellsean polyhedra by Bucknum et al. [22] as the name comes as an extension of the concept and terms for the semi-regular Archimedean and Catalan polyhedra, in which, respectively, the polygonality, n, and the connectivity, p, is fractional or irregular in the Schlaefli symbol (n, p).

In 1997, Bucknum et al., extended this Schlaefli space to include the three-dimensional structures [23]. As Wells had already abundantly implied in his work, [15] the Schlaefli symbols (n, p) could be identified for the two-dimensional (2D) and three-dimensional (3D) periodic tessellations, by the identification of the corresponding Wells point symbol. The Wells point symbol was an encoding developed by him for individual structures, that included a compact and rigorous summary of the constituent connectivities of a structure, in their various fundamental polygonal circuitries, according to the proper stoichiometry or ratio of connectivities in the structure. A generic example of a Wells point symbol for a binary structure would be $(A^a)_x (B^b)_y$, where the exponents identify the vertices as being "a" and "b" connected and the bases "A" and "B" are the constituent sizes of the polygonal circuits, about the a-connected and b-connected vertices, respectively, within the structure. Finally, the subscripts "x" and "y" specify the relative stoichiometry of the structure with there being "x/y" a-connected vertices for every b-connected vertex [22].

Therefore, with the identification of the various n-sided polygonal circuits in a structure, about the constituent p-connected vertices, and the specification of their stoichiometry, it thus becomes possible to calculate the corresponding Schlaefli symbols (n, p) for any arbitrary polyhedron or 2D or 3D structure, as

n \ p	3	4	5	6	7	8	⋯
3	t	o	i	(3,6)	(3,7)	(3,8)	
4	c	(4,4)	(4,5)	(4,6)	(4,7)	(4,8)	
5	d	(5,4)	(5,5)	(5,6)	(5,7)	(5,8)	
6	(6,3)	(6,4)	(6,5)	(6,6)	(6,7)	(6,8)	
7	(7,3)	(7,4)	(7,5)	(7,6)	(7,7)	(7,8)	
8	(8,3)	(8,4)	(8,5)	(8,6)	(8,7)	(8,8)	
⋮							

Figure 2. Complete topology map for the regular structures.

has been shown previously by the authors [22]. Although, strictly speaking, the Schlaefli symbols (n, p) for the 2D or 3D periodic tessellations do not follow any simple Schlaefli-like relation, like that shown in equation (2) for the polyhedra, they nonetheless are useful for identifying the position and relative location of various structures in an expanded Schlaefli space as shown in figure 2.

In 2005, Bucknum et al., showed that the ratio n/p, the so-called topological form index for structures, could be used to correlate the elementary polygonal circuit area of the polyhedra and the 2D and 3D tessellations with each other through a simple quadratic equation in n/p [24]. It is not clear whether the Wells' conjecture, developed out of this modeling work, implies that there might be an analog of the Schlaefli relation for the polyhedra, as shown in equation (2), existing out there for the various 2D and 3D structures in terms of their Schlaefli symbols (n, p).

In the present communication, we extend these ideas based upon the use of an Euler relation and corresponding Schlaefli relation for the polyhedra, to that of the simple graphs as outlined, in this instance for chemical graphs, by their structural formulae. We treat as an important test case for the proposed concepts, polyaromatic hydrocarbons (PAH's) as chemical graphs, as these are potentially the simplest case and involve no weighting scheme for heteroatoms, and so they are an important test case for the model. The validity of the proposed scheme is judged on the basis of the QSPR-QSAR theory, where the reliability of the analogous topological quantities V, E, F, n and p, as identified in the preceding discussion for the polyhedra and 2D and 3D structures, which can be transferred as rigorous definitions to the various chemical graphs, can be assessed by the goodness of fit they provide to various properties of PAH such as BP, $\log(K_{ow})$, and RI.

2. Graph theory and molecules

In this section we introduce the analogous Equations involving the topological parameters V, E, F, n and p for graphs. The Euler relation for graphs is given in equation (5):

$$V - E + F = 1. \tag{5}$$

This relationship is entirely analogous to that for the polyhedra, shown as equation (1), only the characteristic for the polyhedra is 2, while that for the graphs is 1. The explanation for this difference is that in moving from the polyhedra to the graphs, one is, in fact, removing a single face from the polyhedra (the so-called infinite bounded face) to generate the corresponding graph. Therefore, the quantity in equation (1) identified as the number of faces, F, is reduced by 1, and so the right-hand side of equation (1) is reduced by 1 as well, to get the result in equation (5) for the graphs.

One can see that just as in the polyhedra, in the graphs, one can identify V, E, and F as descriptors that characterize the graph. For a typical hydrocarbon like benzene, as shown on the left in figure 3, one can polygonalize the chemical graph, as shown on the right in figure 3, to generate the corresponding chemical graph with the indicated values of $V = 12$, $E = 18$, and $F = 7$. In a more general illustration of the polygonalization process, shown in figure 4, an aufbau diagram of the building up of Buckminsterfullerene from its progenitors is described with a connection halo drawn around each member in the sequence, see ref. [25] for more details.

This polygonalization procedure has therefore been described previously for some chemical graphs by Bucknum et al. [25], but in their earlier paper on chemical graphs, Bucknum et al. identify different descriptors for the graphs than are described here with the use of the Euler equations for graphs. In particular, Bucknum et al. identify a characteristic value of n as an average over the polygon sizes inscribed within the so-called "connection halo" about the molecular structure, like that shown in figure 3. So, in the benzene molecule for example,
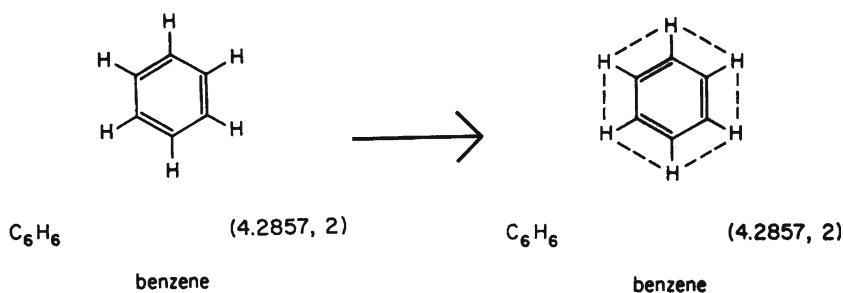

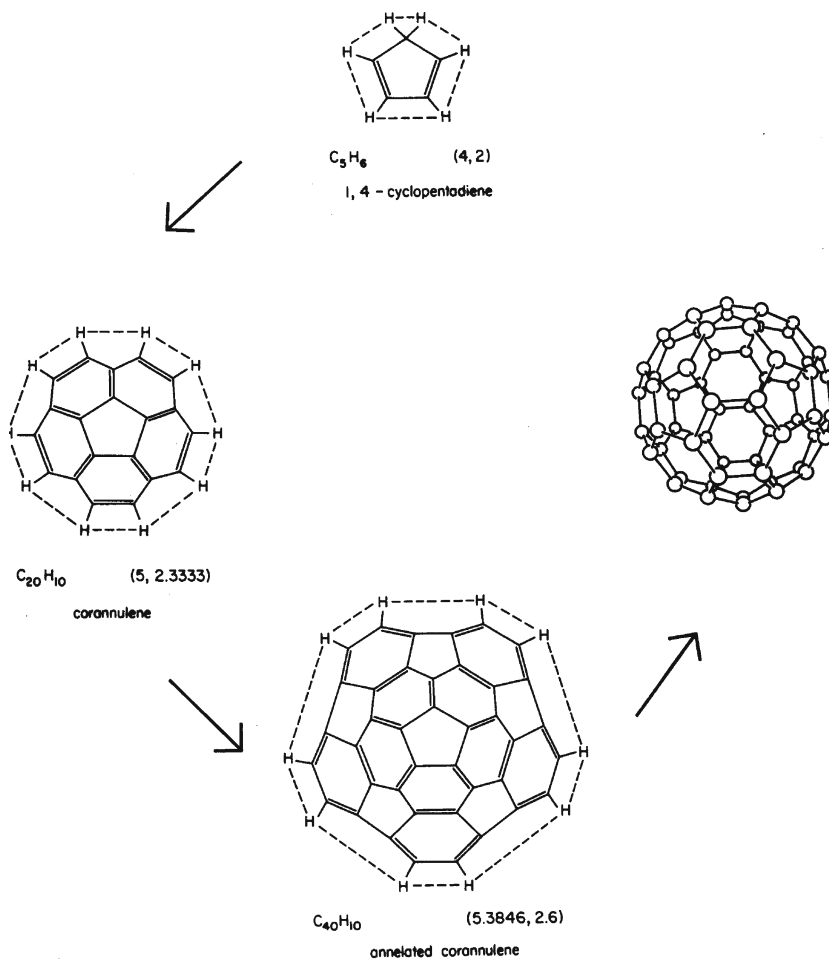Figure 3. Polygonalization of the benzene moiety.

Figure 4. Aufbau molecular structures for icosahedral fullerene with connection halos drawn around each member.

the parameter n is identified as $n = 4.2814$, which is just the average over the polygon sizes within the connection halo in benzene. Unfortunately, this value of n, that will be identified as n_{polygon} hereafter, is not equal to a simple expression like the ordinary value of n for the graphs constructed from a ratio of the number of edges in the graph, E, to the number of faces in the graph, F. In stark contrast, in the polyhedra we have the relation, $n_{\text{polygon}} = n$, for every polyhedron. Likewise, in the earlier paper on chemical graphs by Bucknum et al., they identify a connectivity, p, as equal to the average over the chemical valences in the chemical graph. In this scheme the H atoms would have a connectivity of 1 while the C atoms in aromatic chemical graphs would have a connectivity of 3. This connectivity index is identified as p_{valence} hereafter, and is given as $p_{\text{valence}} = 2$ in the benzene chemical graph shown in figure 3. But, as we will see

below, in this paper we will describe, in addition to $n_{polygon}$ and $p_{valence}$, another set of values of (n, p) based on the Euler equations, which will also be used as descriptors in our eventual QSPR-QSAR analysis.

Turning to equation (5), we see that the Euler characteristic for graphs is 1. This simply and elegantly means that any valid set of integers for a graph, in the form of the triplet (V, E, F), will have that combination so specified by equation (5). Therefore, no matter how we divide up space into compartments of a bounded figure in the form of a graph, the equality in (5) will always rigorously hold. Furthermore, we can substitute the Schlaefli identities of $n = 2E/F$ and $p = 2E/V$, as given above in the polyhedron analysis in equations (2) and (3), into equation (5) to obtain a Schlaefli relation for graphs, including the polygonalized chemical graphs of Bucknum et al. [25], that is entirely analogous to that shown in equation (4) for the polyhedra. This Schlaefli relation for graphs is shown in equation (6) and relates the characteristic values of (n, p) for graphs to their corresponding number of edges, E. These edges are shown to be inscribed within the connection halo, in the case of chemical graphs like that shown for benzene in figure 3.

$$\frac{1}{n} - \frac{1}{2} + \frac{1}{p} = \frac{1}{2E}. \tag{6}$$

Equation (6) is entirely rigorous for all the innumerable graphs, and their consequent realization as chemical graphs, by inscribing a connection halo about the periphery of any hydrocarbon molecule known, of all the various hydrocarbons. Therefore, one simply calculates the values of (n, p) from the primary indexes (V, E, F) according to the Schlaefli identities $n = 2E/F$ and $p = 2E/V$. One can, also, separately calculate $n_{polygon}$ and $p_{valence}$ for these structures, as we have done in this paper, to get a less mathematically rigorous set of descriptors for chemical graphs, if they are perhaps more chemically intuitive. The mathematical basis for the new set of molecular descriptors labeled as V, E, F, n and p for chemical graphs, as described in this paper, is therefore the Euler equations for graphs, shown here as equations (5) and (6). The alternative descriptors of $n_{polygon}$ and $p_{valence}$ are simply added on here in this instance to be consistent with the earlier contribution by Bucknum et al. with regard to chemical graph theory [25].

3. Results and discussion

The data set for BP (°C), $\log(k_{ow})$, and RI of PAH was collected from the literature [26,27]. These three properties reflect the environmental impact of PAH, since these compounds are known to be potent pollutants resulting from combustion processes in which carbon fuel is not completely converted to CO or CO_2, such as the burning of wood and coal, exhaust of gasoline and diesel from combustion engines, and other sources [28].

Table 1 displays the numerical values of the seven molecular descriptors discussed previously: $n, p, n_{\text{polygon}}, p_{\text{valence}}, E, F, V$. The first thing that can be noted is that the new descriptors are degenerate, having the same values for various PAH isomers of the same size, i.e. $E = 36$ for anthracene and phenanthrene. This will probably cause the same numerical prediction of the property for any two isomers considered.

When modeling 37 BP the best relationship found, in terms of the smallest value of the standard deviation for the model (S), leads to the following statistics:

$$BP = -712.206 + 301.860 \cdot p_{\text{valence}} + 10.809 \cdot E$$
$$N = 37, \quad R = 0.9941, \quad S = 12.980\,°C, \quad F = 1432.679 \tag{7}$$
$$R_{\text{loo}} = 0.9929, \quad S_{\text{loo}} = 13.667\,°C.$$

As can be seen from equation (7), the relationship has good predictive ability in the homologous set, according to the "leave-one-out cross validation" (loo) parameters, and this trend is also manifested in the predictions of the model, which are shown in table 2. Among the 37 chemicals, there is only one outlier of the model exceeding $2S$: compound indeno[1,2,3-cd]fluoranthene.

The present subset of descriptors was also able to explain 26 octanol–water partition coefficients of the PAH's, given as $\log(k_{\text{ow}})$, leading to the following equations which have no-outliers exceeding $2S$

$$\log(K_{\text{ow}}) = -7.812 + 1.523 \cdot n + 0.113 \cdot E$$
$$N = 26, \quad R = 0.9884, \quad S = 0.198, \quad F = 487.648 \tag{8}$$
$$R_{\text{loo}} = 0.9850, \quad S_{\text{loo}} = 0.215.$$

This relationship again performed well in the prediction stage and the fitted values are presented in table 3.

For the case of 47 chromatographic retention indexes, given as RI, the model found is

$$RI = -8.289 + 1.500 \cdot n + 0.085 \cdot E$$
$$N = 47, \quad R = 0.9366, \quad S = 0.333, \quad F = 157.279 \tag{9}$$
$$R_{\text{loo}} = 0.9307, \quad S_{\text{loo}} = 0.337,$$

having statistics that are slightly worse than the other property correlations, but also enabling us to establish a parallelism between the structure and the property (see table 4). Equation (9) has three outliers: dibenzo[c,g]phenanthrene, dibenzo[b,def]chrysene, and naphtho[2,1,8-qra] naphthacene. The predictions achieved with the three models are also represented graphically in figures 5–7.

From the above models it can be concluded that the most important molecular descriptor is always E, the number of edges in the constituent, polygonalized molecular structures, i.e. the chemical graphs, since it works better than the other

Table 1
Numerical values for the new Euler-based molecular descriptors proposed in the present study.

Name	n	p	$n_{polygon}$	$p_{valence}$	E	F	V
Naphthalene	5.400	3.000	3.400	2.111	27.000	10.000	18.000
Anthracene	5.538	3.000	4.769	2.167	36.000	13.000	24.000
Phenanthrene	5.538	3.000	4.769	2.167	36.000	13.000	24.000
Naphthacene	5.625	3.000	4.875	2.200	45.000	16.000	30.000
Benz[a]anthracene	5.625	3.000	4.875	2.200	45.000	16.000	30.000
Chrysene	5.625	3.000	4.875	2.200	45.000	16.000	30.000
Triphenylene	5.625	3.000	4.875	2.200	45.000	16.000	30.000
Pyrene	5.571	3.000	4.857	2.231	39.000	14.000	26.000
Benzo[c]phenanthrene	5.625	3.000	4.875	2.200	45.000	16.000	30.000
Perylene	5.647	3.000	4.941	2.250	48.000	17.000	32.000
Benzo[a]pyrene	5.647	3.000	4.941	2.250	48.000	17.000	32.000
Benzo[e]pyrene	5.647	3.000	4.941	2.250	48.000	17.000	32.000
Picene	5.684	3.000	4.947	2.222	54.000	19.000	36.000
Pentaphene	5.684	3.000	4.947	2.222	54.000	19.000	36.000
Benzo[b]chrysene	5.684	3.000	4.947	2.222	54.000	19.000	36.000
Dibenz[a,h]anthracene	5.684	3.000	4.947	2.222	54.000	19.000	36.000
Dibenz[a,j]anthracene	5.684	3.000	4.947	2.222	54.000	19.000	36.000
Benzo[b]triphenylene	5.684	3.000	4.947	2.222	54.000	19.000	36.000
Benzo[c]chrysene	5.684	3.000	4.947	2.222	54.000	19.000	36.000
Pentacene	5.684	3.000	4.947	2.222	54.000	19.000	36.000
Dibenzo[c,g] phenanthrene	5.684	3.000	4.947	2.222	54.000	19.000	36.000
Benzo[a]naphthacene	5.684	3.000	4.947	2.222	54.000	19.000	36.000
Dibenzo[b,def]chrysene	5.700	3.000	5.000	2.263	57.000	20.000	38.000
Dibenzo[def,mno] chrysene	5.667	3.000	5.000	2.294	51.000	18.000	34.000
Dibenzo[a,j]naphthacene	5.727	3.000	5.000	2.238	63.000	22.000	42.000
Dibenzo[a,l]naphthacene	5.727	3.000	5.000	2.238	63.000	22.000	42.000
Dibenzo[a,c] naphthacene	5.727	3.000	5.000	2.238	63.000	22.000	42.000
Dibenzo[e,l]naphthacene	5.700	3.000	5.000	2.263	57.000	20.000	38.000
Dibenzo[de,gr] naphthacene	5.700	3.000	5.000	2.263	57.000	20.000	38.000
Dibenzo[g,p]chrysene	5.727	3.000	5.000	2.238	63.000	22.000	42.000
Benzo[c]picene	5.727	3.000	5.000	2.238	63.000	22.000	42.000
Benzo[ghi]perylene	5.667	3.000	5.000	2.294	51.000	18.000	34.000
Dibenzo[b,k]chrysene	5.727	3.000	5.000	2.238	63.000	22.000	42.000
Dibenzo[c,l]chrysene	5.727	3.000	5.000	2.238	63.000	22.000	42.000
Benzo[b]perylene	5.700	3.000	5.000	2.663	57.000	20.000	38.000
Benzo[a]perylene	5.700	3.000	5.000	2.663	57.000	20.000	38.000
Dibenzo[de,mn] naphthacene	5.700	3.000	5.000	2.663	57.000	20.000	38.000
Naphtho[2,3-g]chrysene	5.727	3.000	5.000	2.238	63.000	22.000	2.000
Benzo[h]pentaphene	5.727	3.000	5.000	2.238	63.000	22.000	42.000

Table 1
(Continued)

Name	n	p	$n_{polygon}$	$p_{valence}$	E	F	V
Benzo[a]pentacene	5.727	3.000	5.000	2.238	63.000	22.000	42.000
Coronene	5.684	3.000	5.053	2.333	54.000	19.000	36.000
Naphtho[1,2,3,4-def]chrysene	5.700	3.000	5.000	2.263	57.000	20.000	38.000
Dibenzo[def,p]chrysene	5.700	3.000	5.000	2.263	57.000	20.000	38.000
Benzo[rst]pentaphene	5.700	3.000	5.000	2.263	57.000	20.000	38.000
Benzo[g]chrysene	5.684	3.000	4.947	2.222	54.000	19.000	36.000
2,3:5,6-Dibenzophenanthrene	5.684	3.000	4.947	2.222	54.000	19.000	36.000
Naphtho[2,1,8-qra]naphthacene	5.700	3.000	5.000	2.263	57.000	20.000	38.000
Dibenz[a,e]aceantrylene	5.700	3.000	5.000	2.263	57.000	20.000	38.000
Acenaphthylene	5.455	3.000	4.727	2.200	30.000	11.000	20.000
Dibenzo[a,k]fluoranthene	5.700	3.000	5.000	2.263	57.000	20.000	38.000
Naphtho[2,3-k]fluoranthene	5.700	3.000	5.000	2.263	57.000	20.000	38.000
Fluoranthene	5.571	3.000	4.857	2.231	39.000	14.000	26.000
Dibenzo[k,mno]fluoranthene	5.667	3.000	5.000	2.294	51.000	18.000	34.000
1,2-Dihydroacenaphthylene	5.455	3.000	4.727	2.200	30.000	11.000	20.000
9H-fluorene	5.500	3.000	4.417	2.182	33.000	12.000	22.000
Benzo[b]fluorene	5.600	3.000	4.867	2.143	42.000	15.000	28.000
Benzo[c]fluorene	5.600	3.000	4.867	2.143	42.000	15.000	28.000
Benzo[ghi]fluoranthene	5.600	3.000	4.933	2.286	42.000	15.000	28.000
Benzo[a]aceanthrylene	5.647	3.000	4.941	2.250	48.000	17.000	32.000
Indeno[1,2,3-cd]pyrene	5.667	3.000	5.000	2.294	51.000	18.000	34.000
Indeno[1,2,3-cd]fluoranthene	5.684	3.000	5.053	2.333	54.000	19.000	36.000
Cyclopenta[cd]pyrene	5.600	3.000	4.933	2.286	42.000	15.000	28.000
Benzo[j]fluoranthene	5.647	3.000	4.941	2.250	48.000	17.000	32.000
Benzo[k]fluoranthene	5.647	3.000	4.941	2.250	48.000	17.000	32.000
Benzo[a]fluorene	5.500	3.034	4.750	2.207	44.000	16.000	29.000
Dibenz[e,k]acephenanthrylene	5.700	3.000	5.000	2.263	57.000	20.000	38.000
Benzo[b]fluoranthene	5.647	3.000	4.941	2.250	48.000	17.000	32.000
Benzene	5.143	3.000	4.286	2.000	18.000	7.000	12.000

Table 2
Observed and predicted BP (°C) for 37 polyaromatic hydrocarbons (PAH's) according to the Euler-based scheme.

No.	Name	BP exp.	BP pred.	Diff.[a]
1	Naphthalene	218	216.4	1.55
2	Anthracene	340	330.3	9.61
3	Phenanthrene	338	330.3	7.61
4	Naphthacene	440	437.6	2.37
5	Benz[a]anthracene	435	437.6	−2.62
6	Chrysene	431	437.6	−6.62
7	Triphenylene	429	437.6	−8.62
8	Pyrene	393	382	10.9
9	Perylene	497	485	11.9
10	Benzo[a]pyrene	496	485	10.9
11	Benzo[e]pyrene	493	485	7.92
12	Picene	519	541.5	−22.5
13	Dibenz[a,h]anthracene	535	541.5	−6.53
14	Dibenz[a,j]anthracene	531	541.5	−10.5
15	Benzo[b]triphenylene	535	541.5	−6.53
16	Dibenzo[b,def]chrysene	596	586.2	9.74
17	Dibenzo[def,mno]chrysene	547	530.7	16.2
18	Benzo[ghi]perylene	542	530.7	11.2
19	Coronene	590	574.9	15
20	Naphtho[1,2,3,4-def]chrysene	592	586.2	5.74
21	Dibenzo[def,p]chrysene	595	586.2	8.74
22	Benzo[rst]pentaphene	594	586.2	7.74
23	Acenaphthylene	270	275.5	−5.58
24	Fluoranthene	383	382	0.936
25	1,2-Dihydroacenaphthylene	274	275.5	−1.58
26	9H-fluorene	294	302.5	−8.52
27	Benzo[b]fluorene	398	388	9.96
28	Benzo[c]fluorene	406	388	17.9
29	Benzo[ghi]fluoranthene	422	430.9	−8.99
30	Indeno[1,2,3-cd]pyrene	534	530.7	3.25
31	Indeno[1,2,3-cd]fluoranthene	531	574.9	−43.9
32	Cyclopenta[cd]pyrene	439	430.9	8
33	Benzo[j]fluoranthene	480	485	−5.07
34	Benzo[k]fluoranthene	481	485	−4.07
35	Benzo[a]fluorene	403	428.8	−25.8
36	Benzo[b]fluoranthene	481	485	−4.07
37	Benzene	80.1	85.8	−5.7

[a]Diff.: Difference experimental-predicted property.

Table 3
Observed and predicted $\log(k_{OW})$ for 26 polyaromatic hydrocarbons (PAH's) according to the Euler-based scheme.

N	Name	$\log(k_{OW})$ exp.	$\log(k_{OW})$ pred.	Diff.
1	Naphthalene	3.33	3.45	−0.123
2	Anthracene	4.54	4.67	−0.136
3	Phenanthrene	4.55	4.67	−0.126
4	Naphthacene	5.96	5.82	0.136
5	Benz[a]anthracene	5.91	5.82	0.0861
6	Chrysene	5.84	5.82	0.0161
7	Triphenylene	5.45	5.82	−0.373
8	Pyrene	5.14	5.06	0.0749
9	Benzo[c]phenanthrene	5.84	5.82	0.0161
10	Perylene	6.3	6.19	0.103
11	Benzo[a]pyrene	6.3	6.19	0.103
12	Dibenz[a,h]anthracene	6.75	6.93	−0.181
13	Benzo[b]triphenylene	7.19	6.93	0.258
14	Pentacene	7.19	6.93	0.258
15	Benzo[a]naphthacene	6.81	6.93	−0.121
16	Benzo[ghi]perylene	6.87	6.56	0.304
17	Coronene	6.75	6.93	−0.181
18	Acenaphthylene	4	3.87	0.126
19	Fluoranthene	5.22	5.06	0.154
20	1,2-Dihydroacenaphthylene	3.92	3.87	0.0467
21	9H-fluorene	4.18	4.28	−0.1
22	Benzo[b]fluorene	5.75	5.44	0.302
23	Benzo[k]fluoranthene	6	6.19	−0.196
24	Benzo[a]fluorene	5.4	5.52	−0.127
25	Benzo[b]fluoranthene	5.8	6.19	−0.396
26	Benzene	2.13	2.05	0.0739

six descriptors studied for any of the three physicochemical properties considered in the QSPR-QSAR analysis. This fact can be explained in terms of E being a composite of (n, p) as referenced in the Schlaefli relation, thus providing a global index of the structural content of both descriptors.

The seven new variables introduced here obey many desired attributes for molecular descriptor design [29,30]; being able to correlate with at least one property, change with size and molecular ramification, being easy to compute, having a direct structural interpretation, being linearly independent of other descriptors, etc. The fact that these descriptors include degeneration for isomeric structures can be surmounted if combining such definitions of molecular descriptors with other more elaborated variables, such as geometrical or electronic ones derived from a quantum chemical calculation.

Table 4
Observed and predicted RI for 47 polyaromatic hydrocarbons (PAH's) according to the Euler-based scheme.

No.	Name	RI exp.	RI pred.	Diff.
1	Naphthalene	2	2.11	−0.116
2	Anthracene	3.2	3.09	0.108
3	Phenanthrene	3	3.09	−0.0915
4	Naphthacene	4.51	3.98	0.521
5	Benz[a]anthracene	4	3.98	0.0111
6	Chrysene	4.1	3.98	0.111
7	Triphenylene	3.7	3.98	−0.288
8	Pyrene	3.58	3.39	0.183
9	Benzo[c]phenanthrene	3.64	3.98	−0.348
10	Perylene	4.33	4.27	0.0521
11	Benzo[a]pyrene	4.53	4.27	0.252
12	Benzo[e]pyrene	4.28	4.27	0.00218
13	Picene	5.18	4.84	0.334
14	Pentaphene	4.67	4.84	−0.175
15	Benzo[b]chrysene	5	4.84	0.154
16	Dibenz[a,h]anthracene	4.73	4.84	−0.115
17	Dibenz[a,j]anthracene	4.56	4.84	−0.285
18	Benzo[b]triphenylene	4.4	4.84	−0.445
19	Benzo[c]chrysene	4.45	4.84	−0.395
20	Dibenzo[c,g]phenanthrene	4.07	4.84	−0.775
21	Benzo[a]naphthacene	4.99	4.84	0.144
22	Dibenzo[b,def]chrysene	6	5.12	0.875
23	Dibenzo[def,mno]chrysene	5.08	4.56	0.516
24	Dibenzo[de,gr]naphthacene	4.92	5.12	−0.204
25	Benzo[ghi]perylene	4.76	4.56	0.196
26	Benzo[b]perylene	5.04	5.12	−0.0847
27	Naphtho[1,2,3,4-def]chrysene	4.97	5.12	−0.154
28	Dibenzo[def,p]chrysene	4.89	5.12	−0.234
29	Benzo[rst]pentaphene	5.73	5.12	0.605
30	Benzo[g]chrysene	4.27	4.84	−0.575
31	2,3:5,6-Dibenzophenanthrene	4.33	4.84	−0.515
32	Naphtho[2,1,8-qra]naphthacene	5.87	5.12	0.745
33	Dibenz[a,e]aceantrylene	4.9	5.12	−0.224
34	Dibenzo[a,k]fluoranthene	4.9	5.12	−0.224
35	Fluoranthene	3.37	3.39	−0.0268
36	9H-fluorene	2.7	2.77	−0.0779
37	Benzo[b]fluorene	3.84	3.69	0.144
38	Benzo[c]fluorene	3.49	3.69	−0.205
39	Benzo[a]aceanthrylene	4.22	4.27	−0.0578
40	Indeno[1,2,3-cd]pyrene	4.84	4.56	0.276
41	Indeno[1,2,3-cd]fluoranthene	4.93	4.84	0.0847
42	Benzo[j]fluoranthene	4.24	4.27	−0.0378

Table 4
(Continued)

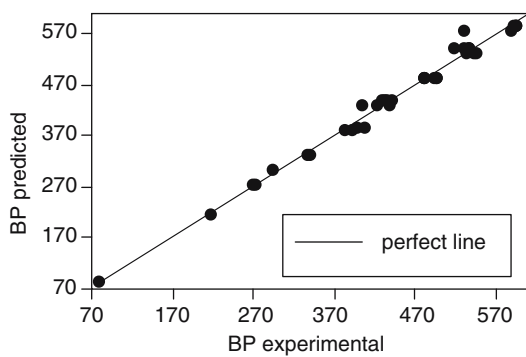| No.	Name	RI exp.	RI pred.	Diff.
43	Benzo[k]fluoranthene	4.42	4.27	0.142
44	Benzo[a]fluorene	3.72	3.71	0.00394
45	Dibenz[e,k]acephenanthrylene	5.27	5.12	0.145
46	Benzo[b]fluoranthene	4.29	4.27	0.0121
47	Benzene	1	0.962	0.037

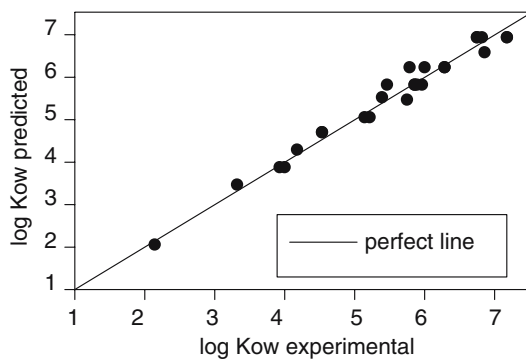Figure 5. QSAR-QSPR of boiling point data for various PAH's according to the Euler scheme.

Figure 6. QSPR-QSAR of octanol–water partition coefficients in various PAH's according to the Euler scheme.
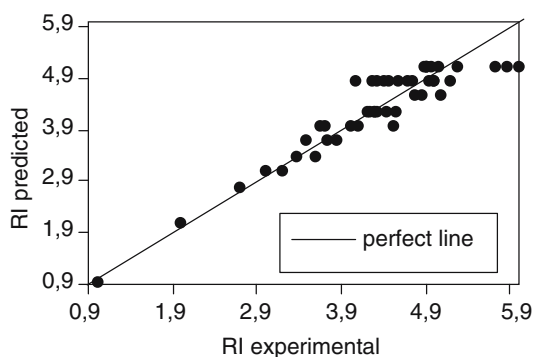
Figure 7. QSPR-QSAR of chromatographic retention indices in various PAH's according to the Euler scheme.

4. Conclusion

Although there exist a large number of molecular descriptors reported in the literature, there still exits the need for new descriptors, because each class of such descriptors encode some specific structural feature. The molecular descriptors deduced from graph considerations using the Euler relations for graphs as described in equations (5) and (6), can be employed to properly characterize physicochemical and biological activities of chemicals. In the present study, we were able to correlate BP, $\log(k_{ow})$, and RI with E, n, and $p_{valence}$ for a model set of PAH, achieving good predictions with solely 2-parameter models. Although the proposed descriptors include degeneration over isomers, they can be complemented with other types of variables to explain a given property and surpass this problem.

It is our purpose to continue exploring the present ideas based upon a topological analysis of graphs and to apply them to more sophisticated chemical graph and polyhedral structures, such as fullerenes, helicenes, etc. exhibiting different type of physicochemical properties of interest. The introduction of heteroatoms and multiple bonds in the definitions of the present descriptors should also be addressed, and derived QSPR-QSAR results will be published soon elsewhere.

References

[1] F.A. Cotton, *Chemical Applications of Group Theory*, 3rd ed. (John Wiley & Sons, Inc., New York, 1990).

[2] L. Euler, Elementa doctrinae solidorum et Demonstratio nonnularum insignium proprietatum quibus solida heddris planis inclusa sunt praedita, *Proceedings of the St. Petersburg Academy*, 1758.

[3] A.F. Wells, Acta Cryst. 7 (1954) 535.

[4] A.F. Wells, Acta Cryst. 7 (1954) 545.

[5] A.F. Wells, Acta Cryst. 7 (1954) 842.

[6] A.F. Wells, Acta Cryst. 7 (1954) 849.

[7] A.F. Wells, Acta Cryst. 8 (1955) 32.

[8] A.F. Wells, Acta Cryst. 9 (1956) 23.

[9] A.F. Wells and R.R. Sharpe, Acta Cryst. 16 (1963) 857.

[10] A.F. Wells, Acta Cryst. 18 (1965) 894.

[11] A.F. Wells, Acta Cryst. B24 (1968) 50.

[12] A.F. Wells, Acta Cryst. B25 (1969) 1711.

[13] A.F. Wells, Acta Cryst. B28 (1972) 711.

[14] A.F. Wells, Acta Cryst. B32 (1976) 2619.

[15] A.F. Wells, *Three Dimensional Nets and Polyhedra*, 1st ed. (John Wiley and Sons Inc., New York, 1977).

[16] A.F. Wells, *Further Studies of Three-dimensional Nets*, 1st ed. (American Crystallographic Association, American Crystallographic Association Press, 1979) Monograph Number 8. Note that unfortunately the author could not locate direct references to the work of Schläfli in topology, the references to Wells are indirect but refer to Schläfli's work.

[17] J. Baggott, *Perfect Symmetry: The Accidental Discovery of Buckminsterfullerene*, 1st ed. (Oxford University Press: Oxford, 1996).

[18] M. Wenninger, *Dual Models*, 1st ed. (Cambridge University Press, 1983).

[19] J.P. Frederico, *Descartes on Polyhedra: A Study of the De Solidorum Elementis*, appearing in *Sources in the History of Mathematics and Physical Sciences 4*, 1st ed. (Springer-Verlag, 1982).

[20] I. Peterson, Sci. News 160 (2001) 396.

[21] P.R. Cromwell, *Polyhedra*, 1st ed. (Cambridge University Press, 1997).

[22] M.J. Bucknum and E.A. Castro, Comm. Math. Comp. Chem. (MATCH) 54 (2005) 89.

[23] M.J. Bucknum, Carbon 35 (1997) 1.

[24] M.J. Bucknum and E.A. Castro, J. Math. Chem. 2005 in press. Available at http://www.springerlink.com

[25] M.J. Bucknum E.A. Castro, Comm Math. Comp. Chem. (MATCH) 54 (2005) 121.

[26] F.A.L. Riveiro and M.M.C. Ferreira, J. Mol. Struct. (THEOCHEM) 663 (2003) 109.

[27] M.M.C. Ferreira, Chemosphere 44 (2001) 125.

[28] A.D.P. Netto, J.C. Moreira, A.E.X.O. Dias, G. Arbilla, L.F.V. Ferreira, A.S. Oliveira and J. Barek, A Review of Methodology. Quím. Nova 23 (2000) 765.

[29] M. Randic, J. Math. Chem. 7 (1991) 155.

[30] A.T. Balaban, J. Mol. Struct. (THEOCHEM) 165 (1988) 243.