Inter*net* Electronic Journal of
Molecular Design

July 2006, Volume 5, Number 7, Pages 364–375

Editor: Ovidiu Ivanciuc

Special issue dedicated to Professor Lemont B. Kier on the occasion of the 75th birthday

The Evolution of the Valence Delta in Molecular Connectivity Theory

Lionello Pogliani [1]

[1] Dipartimento di Chimica, Università della Calabria, Via P. Bucci – cubo 14 C, I–87036 Arcavacata di Rende (CS), Italia

The Evolution of the Valence Delta in Molecular Connectivity Theory [#]

Lionello Pogliani [1],*

[1] Dipartimento di Chimica, Università della Calabria, Via P. Bucci – cubo 14 C, I–87036 Arcavacata di Rende (CS), Italia

Abstract
Motivation. The evolution of a fundamental parameter in Kier and Hall molecular connectivity theory, the valence delta δ^v, is followed from its beginnings till its most recent definitions. This parameter, which was initially based on quantum concepts (atomic number, and valence electrons), is now solely based on concepts belonging to general and complete graphs, like multiple edges, and self–connections of pseudographs, and order and regularity of complete graphs. Furthermore, the hydrogen content of a molecule, which was implied in hydrogen suppressed chemical graphs, is now directly encoded into the valence delta. These two new features of the valence delta allow to graph differentiate among atoms of any type. The model quality of the newly defined valence delta is tested with three properties of three different classes of compounds: side–chain molecular volume, and isoelectric point of amino acids, and boiling points of amines plus boiling points of alcohols. The model of the composite class [amines + alcohols], which seems to behave as a new property underlines the peculiar character of the new valence delta.

Method. A new valence delta number, which includes hydrogen perturbation, is defined. The defined hydrogen perturbation makes no use of any new graph concept because it is the ratio between two different valence delta numbers. The ratio is always smaller than one and can be fine–tuned thanks to an exponential parameter.

Results. A model of three properties with and without hydrogen perturbation is presented. The model, which is achieved with different values for the exponential parameter, underlines the advantages and characteristics of this kind of perturbation.

Conclusions. The hydrogen perturbation throughout the present model study is not only property–dependent, but also *N*–dependent, *i.e.*, dependent on the number of studied compounds.

Keywords. Valence delta; general graphs; complete graphs; hydrogen perturbation; connectivity indices; model computations.

1 INTRODUCTION

Kier and Hall condensed their studies about that branch of chemical graph theory known as molecular connectivity theory throughout two prestigious books [1,2]. Since then molecular connectivity theory has undergone a considerable development [3–27, among many others]. Here

---

[#] Dedicated to Professor Lemont B. Kier on the occasion of the 75th birthday.
* Correspondence author; fax: 0039–0984–492044; E–mail: lionp@unical.it.

we will focus our attention on the development a key parameter in molecular connectivity underwent in recent years, *i.e.*, the development of the valence delta number, normally written as δ^v. The first general definition of the valence delta, due to Kier and Hall (KH), valid for any type of atom, either second–row atoms or higher–row atoms, was the following, [1]:

$$^{KH1}\delta^v = \frac{Z^v - h}{Z - Z^v - 1} \tag{1}$$

where Z is the atomic number, Z^v is the number of valence electrons, and h is the number of hydrogen atoms suppressed, which means that electrons in bonds with hydrogen atoms are excluded. The valence delta was considered a count of the valence electrons, weighted by the number of core electrons, and the suppressed hydrogen atom were viewed as the zero–valued reference structure in counting non–valence electrons. For second–row atoms Eq. (1) gives Eq. (2):

$$^{KH2}\delta^v = Z^v - h \tag{2}$$

This definition has a more palatable electronic character. In fact, it equals the sum of the sigma (σ), pi (π), and lone–pair (n) electrons and excludes electrons in bonds with hydrogen atoms:

$$^{KH2}\delta^v = \sigma + \pi + n - h \tag{3}$$

Table 1. The δ, $^{KH1}\delta^v$, and $^{KH2}\delta^v$ values, the *I*–State values, the X_{KH}, electronegativities, the ^RX_M,[a] electronegativities, the δ^v(ps), and the δ^v(6) values of Eq. (6)

Atom	δ	$^{KH1}\delta^v$	$^{KH2}\delta^v$	I	X_{KH}	^RX_M (eV) [b]	δ^v(ps)	δ^v(6) (q=1)
= O	1	6	6	7	1.25	9.09	6	6
− O −	2	6	6	3.5	1.00	7.27	6	6
≡ N	1	5	5	6	1.00	7.10	5	5
= N −	2	5	5	3	0.75	4.89	5	5
> N −	3	5	5	2.0	0.5	3.56	5	5
≡ C −	2	4	4	2.5	0.5	2.41	4	4
> C =	3	4	4	1.667	0.25	0.81	4	4
> C <	4	4	4	1.25	0.00	0.00	4	4
− Cl	1	0.78	7	4.11	0.67	3.56	7	2.33 (p=2)
= S	1	0.67	6	3.667	0.55	2.90	6	2 (p=2)
− S −	2	0.67	6	1.833	0.44	2.16	6	2 (p=2)

[a] ^RX_M : rescaled to the >C< value (7.98 eV). [b] eV = electron volts.

For higher–row atoms a correction ($1/N^2$, N = principal quantum number) was introduced at the level of the $^{KH2}\delta^v - \delta$ difference, where δ is the connectivity degree of each vertex–encoding atom in simple chemical graphs. Later on [2] and within the frame of the electrotopological state, a $(2/N)^2$ $^{KH2}\delta^v$ was introduced at the level of the *I*–State index for any type of atom:

$$I = \frac{(2/N)^2 \, ^{KH2}\delta^v + 1}{\delta} \tag{4}$$

With the definition of the *I*–State index, valid for any type of atom, the importance of Eq. (1) faded away. In Table 1 are given, for different atoms, the δ values, the $^{KH1,2}\delta^v$ values (from Eqs. (1) and (2)), the *I*–State values, the Kier and Hall electronegativity, X_{KH}, and the rescaled Mulliken–Jaffe

electronegativity values, $^{R}X_{M}$. This table tells that a relation exists between X_{KH} and $^{R}X_{M}$. These values were taken from Ref. [2]. The Mulliken–Jaffe electronegativities have been rescaled to $X_{M} = 7.98$ (eV) for the >C< atom, while the Kier and Hall electronegativities are defined in the following way:

$$X_{KH} = (^{KH2}\delta^{v} - \delta)/N^{2} \qquad (5)$$

2 MATERIALS AND METHODS

From its beginnings molecular connectivity theory [3] recognized that the δ values were dependent on graph concepts, and could be derived from the simple graph of the hydrogen suppressed (HS) molecule. The same was not true for the valence delta, which was strongly determined by quantum concepts like Z^{v} and Z, and later on by N. For alkanes things are in order as for them we have: $\delta^{v} = \delta$. The given definitions of the valence delta and of the simple delta show that the hydrogen atoms are implied, even if they originate from HS–graphs. In fact, they are subtractive entities. Actually the fact that here and there more than one attempt has been made, even by Kier and Hall [2], to codify the hydrogen atoms underlines the complexity of the problem.

A more precise graph definition of the valence delta should forget everything about the quantum concepts (Z^{v}, Z, and N) and be solely based on the number of connections and self–connections an atom–encoding vertex has in a HS–pseudograph, as has recently been shown [20]. A pseudograph or general graph is a graph which allows for multiple connections and self–connections (or loops). The δ^{v} values of the second row atoms can, in fact, be derived in a pure 'graph way' considering the general graphs of the HS–suppressed molecules.
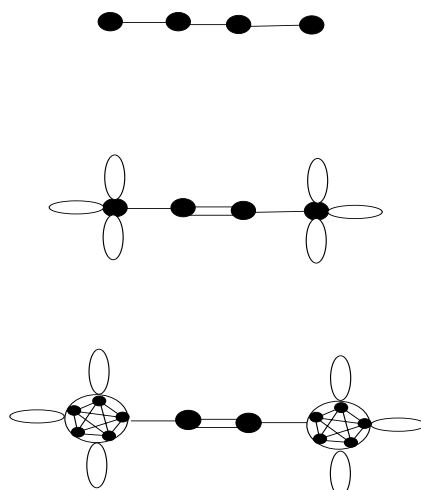
Figure 1. The HS graph (top) and HS–pseudograph of FCH=CHF (middle, loops encode the fluorine non–bonding electrons); (bottom) the HS–pseudograph+K_{P}–graph for BrCH=CHBr. The carbon and bromine cores are encoded with K_{1}, and K_{5} complete graphs, respectively.

Figure 1 shows the HS simple graph (top) of the FHC=CHF molecule, from which the connectivity δ values of each vertex can be derived. The HS general graph of Figure 1–middle is used to derive the connectivity δ^v values of the same molecule (for the bottom graph see later on). Let us call the valence delta of a second row atom which can be encoded with a vertex in a HS–pseudograph as δ^v(ps). Thus Figure 1, middle, tell us that for the carbon atoms δ^v(ps) = 3, while for the fluorine atoms, δ^v(ps) = 7 (loops count twice). The electron–core of any second–row and higher–row atom can now be encoded with complete K_p graphs. A graph is considered complete if every two of its vertices are adjacent, which means that every complete graph is also regular, *i.e.*, all its vertices have the same degree. Thus, the different electron cores are encoded by the aid of different complete K_p graphs, where p is the order of the graph and $r = p - 1$ is its regularity. A pure graph definition of δ^v is the following [16–20]:

$$\delta^v = \frac{q\delta^v(ps)}{(pr+1)} \tag{6}$$

Here, q is an optimization parameter, which, normally, equals 1 or p, and $p = 1, 2, 3, ...$ For second row atoms $p = 1$. In most QSPR studies [21–25, 28, 29] p, which is a kind of principal 'graph' number, resulted to be odd–valued (1, 3, 5…). In the middle graph of Figure 1 the electron cores of the two carbon atoms as well as of the two fluorine atoms are encoded with K_1 complete graphs. In the bottom graph of Figure 1 the core electrons of the two bromine atoms are encoded with odd–valued K_5 complete graphs. Second row atoms, whose core electrons are encoded with K_1 complete graphs, have $\delta^v = \delta^v$(ps). Throughout the last two columns of Table 1 are collected the δ^v(ps), and the δ^v values calculated with Eq. (6). If the Kier and Hall electronegativity and the *I*–State index are now defined as, $X_{KH} = [\delta^v(ps) - \delta] / N^2$, and $I = [\delta^v + 1] / \delta$, then no changes can be detected at the X_{KH} level, and only minor changes at the *I* level (see Table 1).

The suppressed hydrogen atoms can be read as subtractive entities. In fact, the HS general graphs of $CH_2=CH_2$, –CH=CH–, and >C=C< are: •=•, –•=•–, and >•=•<, respectively. Nevertheless, graphs, •=•, –•=•–, >•=•< can also be read as systems with no hydrogen atoms. Furthermore, the subtractive method of implying the hydrogen atoms does not always work fine. For instance, in alkanes for carbon atoms in groups like, –CH$_3$, –CH$_2$–, and >CH–, the equality $\delta^v = \delta^v$(ps) = δ 'sounds strange'. The problem becomes even more serious with the first atom of the following two set of compounds, {LiF, BeHF, BH$_2$F, CH$_3$F}, which have all the same δ^v = 1 value. A similar problem arises with the C atoms in FHC=CHF, and HC≡CH, with the N and O atoms in >N–, and –O–H, and with the O and F atoms in –O–, and F–H. A way out of this burrow could consider the hydrogen atoms as a perturbation term to be added to q in Eq. (6). This perturbation should be (*i*) smaller than one, and possibly even smaller, should avoid (*ii*) the overlapping problem, *i.e.*, that a δ^v(–CH$_2$–) overlaps in value with δ^v(>CH–). This perturbation (*iii*) should allow retrieving Eq. (6) for completely substituted heteroatoms, and (*iv*) should also have a $(p \cdot r + 1)^{-1}$ dependence on the

order of K_p. The following algorithm for δ^v with a hydrogen perturbation f_δ agrees with all these guidelines [28, 29]:

$$\delta^v = \frac{(q + f_\delta^n)\delta^v(ps)}{(pr + 1)} \tag{7}$$

$$f_\delta = [\delta^v_m(ps) - \delta^v(ps)] / \delta^v_m(ps) = 1 - \delta^v(ps)/\delta^v_m(ps) = n_H/\delta^v_m(ps) \tag{8}$$

Here $\delta^v_m(ps)$ is the maximal $\delta^v(ps)$ value an heteroatom can have in a chemical HS general graph when all bonded hydrogen atoms are substituted by heteroatoms (to simplify matters choose a dummy monovalent K_1 X heteroatom). Eq. (7) makes no use of any new graph concept and the equality $f_\delta = n_H/\delta^v_m(ps)$ is the result of the left–side definition, which is solely based on graph concepts. The exponent n is an optimization parameter, which will allow checking how important is the hydrogen contribution, *i.e.*, the higher n the lower the importance of the perturbation caused by the hydrogen atoms. Clearly for completely substituted second–row heteroatoms Eq. (7) goes over into Eq. (6) as $f_\delta = 0$, *i.e.*, for quaternary carbon atoms (C_q) with $n_H = 0$ (and the like for the F and Br atoms in Fig. 1), $f_\delta = 0$ as $\delta^v_m(ps) = \delta^v(ps)$. The same holds for a vanishingly small perturbation, where for $n \rightarrow \infty$, $f_\delta^n \rightarrow 0$. Eq. (7) also tells that in alkanes, whose graphs and general graphs coincide, and where $\delta^v(ps) = \delta$, $p = q = 1$, and $r = 0$, relation $\delta^v = (1 + f_\delta^n)\delta$ holds. Four different values for n in f_δ will here be considered: $n = 2, 4, 6,$ and 8. Notice that for $n = 1$ CH_2F_2 and BF_3 will have the same δ^v, a situation that should clearly be avoided using $n > 1$. Groups like $-CH_3$ ($f_\delta = 0.75$), $-CH_2-$ ($f_\delta = 0.5$), and $-CH<$ ($f_\delta = 0.25$) will now have the following valence delta values: $\delta^v = (1 + 0.75^n)$, $(1 + 0.5^n)\cdot 2$, and $(1 + 0.25^n)\cdot 3$, respectively. For $-NH_2$ ($f_\delta = 2/5 = 0.4$) and $-OH$ ($f_\delta = 1/6 = 0.17$), we have, $\delta^v = (1 + 0.4^n)\cdot 3$, and $(1 + 0.17^n)\cdot 5$, respectively. Note that the values obtained with Eq. (7) for the atoms in Table 1 cannot be differentiated from the values obtained with Eq. (6). The optimal model of the properties needs the following molecular connectivity basis indices (β) only:

$$\{\beta\} = \{^0\chi, ^1\chi, \chi_t, D^v, , ^0\chi^v, ^1\chi^v, \chi^v_t \}$$

$$D^v = \Sigma_i \delta^v_i \tag{9}$$

$$^0\chi = \Sigma_i (\delta_i)^{-0.5} \tag{10}$$

$$^1\chi = \Sigma (\delta_i \delta_j)^{-0.5} \tag{11}$$

$$\chi_t = (\Pi \delta_i)^{-0.5} \tag{12}$$

For the $^1\chi$ type of indices the sums and products are over all edges of the chemical graph (σ bonds in a molecule). Replacing δ with δ^v the corresponding $^0\chi^v$, $^1\chi^v$ and χ^v_t indices are obtained. In a case a molecular connectivity higher–order term, $X = f(\chi)$ is needed to achieve an optimal model. These descriptors are obtained by the aid of a trial–and–error procedure [15] performed with a wide set of molecular connectivity indices. To check the quality of a model the following statistics will be used: the squared correlation coefficient, r^2, the standard deviation of the estimate, s, the F value, and the number of points, N. Cross–validation methods like the leave–one–out or leave–more–out

methods, will also be used for small sets of data. Sometimes these method are called jackknife, a name that stems from the method that each observation is removed (*i.e.*, cut with the knife) one at a time, or more at a time (for the higher–order *Jackknife*) in order to get a feeling for the spread of data. The leave–one–out method is normally used when there is a small number of data. The prediction coefficient, q^2, which is normally used to check the validity of the leave–one–out method, is: $q^2 = (SD - PRESS) / SD$, where $SD = \Sigma(y_i - <y>)^2$ is the squared deviation of the observed value from their mean, and $PRESS = \Sigma(y_i - y_{iloo})^2$, where y_{iloo} is a predicted value of the studied property where the prediction has been made by the leave–one–out method [28]. A value $q^2 > 0.5$ has been proposed as a satisfactory result, but not all authors agree on the validity of such a low q^2 value [29]. Anyway, our q^2 values are consistently better than 0.5. The utility of each regression parameter, c_i, (in parenthesis near c_i or collected under a vector form) will also be given for the fit–model equation, *i.e.*, $|c_i/s_i|$, where s_i is the estimated error of c_i. A high utility underlines the importance of the corresponding index, even if a low utility should be expected when the corresponding regression parameter is nearly zero. Model plots have been given following the recently published considerations about these kinds of plots [22, 23].

3 RESULTS AND DISCUSSION

Some recent results [20,21] that allow appreciate the importance of the hydrogen perturbation will here be reconsidered. Tested properties are: the side–chain volume, V, the isoelectric point, pI, of amino acids, and the boiling point, BP, of amines and alcohols. An eye will be kept to the ratio of the number of hydrogen atoms, n_H, to the number of heteroatoms n_O (subscript 'O' stays for atoms other than H atoms), *i.e.*, n_H/n_O. For instance, for C_2H_5OH we have: $n_H/n_O = 6/3 = 2$. The model of the side–chain molecular volume, and isoelectric point of amino acids, and the model of the boiling points of amines and alcohols with and without hydrogen perturbation will here be discussed. These three classes of compounds have the following n_H/n_O ratio (rounded off): 1.0 for 18 amino acids, 2.2 for 21 amines, and 2.0 for 27 alcohols. The boiling points of the composite class [amines + alcohols, $n_H/n_O = 2.1$] will allow to derive some interesting hints about the importance of the hydrogen perturbation. Notice that for the examined properties $p = 1$. As already mentioned, the following four *n* values of f_δ will here be tested: 2, 4, 6, and 8.

3.1 Amino Acids

The two properties of natural amino acids here examined, shown in table 2, have, $n_H/n_O = 1$. The values for these properties are taken form Ref. [15] (*pI*), and Ref. [1] (*V*).

3.1.1 Side–Chain Molecular Volume of Amino Acids, V

The best model, with $f_\delta = 0$, is shown by the following combination of valence basis indices, and by a *X* term:

$$\{ D^v, {}^0\chi^v \} : F = 887, r^2 = 0.992, s = 2.5, q^2 = 0.989, N = 18$$
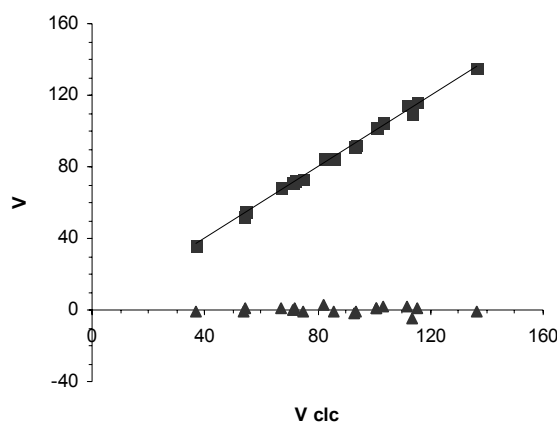$$X = [(D^v)^{1.3} + ({}^0\chi)^{2.1}] / (D^v - 0.7D) : F = 2109, r^2 = 0.992, s = 2.3, N = 18$$

For $f_\delta \neq 0$ there is a nice improvement in the model quality for $n = 6$, where we have:

$$^{f6}\{D^v, {}^0\chi^v\} : F = 1475, r^2 = 0.994, s = 1.9, q^2 = 0.993, N = 18$$
$$^{f6}X = [(D^v)^{1.4} + 1.3({}^0\chi^v)^{2.3}] / [D^v - 2.5({}^1\chi^v)^{0.5}] : F = 4009, r^2 = 0.996, s = 1.7, N = 18$$

The fit–model equation used to obtain Figure 2 is the following, where in parenthesis are the utility values (see method section):

$$V(f=6) = -0.85427D^v (8.9) + 23.5283{}^0\chi^v (36) - 7.7291 (4.0) \tag{13}$$

Figure 2. Plot of the experimental side–chain molecular volume V values, vs. calculated V values, V_{clc}, and plot of their residuals ($V_{exp} - V_{calc}$; on the x axis).

Now, consider a training set of $N = 10$ amino acids, excluding from the model: Gly, Ser, Thr, Leu, Asn, Lys, Glu, and Arg, while keeping n_H/n_O constant. The best term is similar to the previous ^{f6}X term only parameter "$1.3({}^0\chi^v)^{2.3}$" changes into "$1.2({}^0\chi^v)^{2.4}$". The statistics now are: $F = 2013$, $r^2 = 0.996$, $s = 1.6_5$, $N = 10$, and $C = [21.55006 (45), - 61.8174 (18)]$. The calculated values are collected throughout the $V(Clc)$–8–$l.o.$ column of Table 2 (right–side). The inferred values for the eight left–out amino acids are quite satisfactory.

3.1.2 Isoelectric Point, *pI*

The experimental values are shown in Table 2. As already mentioned in Ref. [15] there is only one type of descriptor, *i.e.*, an X term, which is able to satisfactorily describe this property, which for $f_\delta = 0$ it is the following:

$$X(pI) = (0.01\chi_t^v + \Delta g/g_T)({}^0\chi^v/D)^{3.2} : F = 810, r^2 = 0.980, s = 0.3, q^2 = 0.975, N = 18$$

Here, $g_T = 4$ for Arg, $g_T = 3$ for Asp, Lys, Glu, and His, and $g_T = 2$ for the remaining amino acids; Δg is the difference between the main acidic and basic functional groups, thus, it equals $\Delta g = -1$ for Lys and His, $\Delta g = -2$ for Arg,, and $\Delta g = 1$ for Glu and Asp, otherwise it is zero [15]. For $f_\delta \neq 0$ the

description shows a clear improvement for $n = 2$, but not as dramatic as in the previous V case. The descriptor is quite similar to $X(pI)$:

$$^{f2}X(pI) = (0.01\chi_t^v + \Delta g/g_T)(^0\chi^v/D)^4 : F = 873, r^2 = 0.982, s = 0.2_5, q^2 = 0.977, N = 18$$

$$pI = -1378.58\,^{f2}X(pI)\,(30) + 5.80145\,(95) \tag{14}$$

Table 2. The side–chain molecular volume, V (in Å^3), and the isoeletric point, pI, for the natural amino acids (AA), together with the calculated V values with 8–left–out values, $V(Clc)$–8–l.o. (*)

A A	V	V(Clc)–8–l.o.	pI
Gly	36.3	*36.1	5.97
Ala	52.6	51.1	6.00
Ser	54.9	*54.5	5.68
Val	85.1	84.8	5.96
Thr	71.2	*69.3	5.60
Pro	73.6	74.9	6.30
Leu	102	*102.8	5.98
Ile	102	103	6.02
Asn	72.4	*72.0	5.41
Asp	68.4	69.5	2.77
Lys	105.1	*105.0	9.74
Gln	92.7	92.9	5.65
Glu	84.7	*82.1	3.22
His	91.1	92.6	7.59
Arg	109.1	*112.8	10.76
Phe	113.9	111.1	5.48
Tyr	116.2	114.4	5.66
Trp	135.4	136.8	5.89

During the leave–one–out method we assumed that the X term does not change. Actually, there are minor changes (± 0.1) in the power of ($^0\chi^v/D$), if model maximization is achieved for every one–out compound. The improvement over the $f_\delta = 0$ case is achieved for $n = 2$, *i.e.*, a drastic hydrogen perturbation achieves a small improvement. The side–chain V case, instead, showed a consistent improvement with a minor hydrogen perturbation ($n = 6$). The model of pI and V tell us that the hydrogen perturbation is property–dependent.

3.2 Boiling Points of Amines and Alcohols

The experimental values are collected in Table 3 and are taken from Ref. [2]. The two classes of compounds show, for $f_\delta = 0$, different model quality, the amines showing a high quality model, while the alcohols a less satisfactory model. Results for $f_\delta^n \neq 0$ are rather unexpected. Practically the boiling points of the single classes of amines and alcohols show no interesting improvement when the hydrogen perturbation is considered. Nevertheless, as soon as we consider the boiling points of a class made up of the amines and alcohols [$R–NH_2 + ROH$, $n_H/n_O = 2.1$] things change significantly. In fact for $f_\delta = 0$ we obtain the following best single basis index and term descriptors:

$$\{^1\chi\}: F = 284, r^2 = 0.861, s = 17.9, N = 48$$
$$X(Am+Al) = [(\chi_t)^{0.3} + 0.05\,^1\chi^v]: F = 383, r^2 = 0.895, s = 15.7, N = 48$$

Table 3. Primary amines and alcohols, their boiling points T_b (in K)

Amines	T_b	Alcohols	T_b
CH_3-	256.6₅	$(CH_3)_2CH-$	355.5₅
CH_3CH_2-	290.1₅	$CH_3CH_2CH_2-$	370.2₅
$(CH_3)_2CH-$	307.1₅	$CH_3CH_2C(CH_3)_2-$	375.4₅
$CH_3CH_2CH_2-$	322.1₅	$CH_3CH(CH_3)CH_2-$	381.2₅
$CH_3CH_2CH(CH_3)-$	336.1₅	$CH_3(CH_2)_3-$	390.7₅
$CH_3CH(CH_3)CH_2-$	341.1₅	$CH_3CH_2CH_2CH(CH_3)-$	392.0₅
$CH_3CH_2CH_2CH_2-$	350.9₅	$CH_3C(CH_3)_2CH(CH_3)-$	393.5₅
$CH_3CH_2C(CH_3)_2-$	351.1₅	$CH_3(CH_2)_2C(CH_3)_2-$	396.1₅
$(CH_3CH_2)_2CH-$	364.1₅	$CH_3CH(CH_3)CH(CH_3CH_2)-$	400.6₅
$CH_3CH_2CH_2CH(CH_3)-$	365.1₅	$CH_3CH_2CH(CH_3)CH_2-$	402.0₅
$CH_3CH(CH_3)CH_2CH_2-$	368.1₅	$CH_3CH(CH_3)CH_2CH_2-$	405.1₅
$CH_3C(CH_3)_2CH(CH_3)-$	375.1₅	$CH_3CH(CH_3)CH_2CH(CH_3)-$	406.1₅
$CH_3(CH_2)_4-$	377.5₅	$(CH_3CH_2)_2C(CH_3)-$	409.1₅
$CH_3(CH_2)_3CH(CH_3)-$	390.6₅	$CH_3CH_2C(CH_3)_2CH_2-$	409.8₅
$CH_3(CH_2)_5-$	403.1₅	$CH_3(CH_2)_4-$	411.1₅
$CH_3CH_2CH(CH_3)CH_2CH(CH_3)-$	406.6₅	$(CH_3CH(CH_3))_2CH-$	413.1₅
$CH_3(CH_2)_4CH(CH_3)-$	415.1₅	$(CH_3CH_2)_3C-$	415.1₅
$CH_3(CH_2)_6-$	430.0₅	$CH_3CH(CH_3)CH(CH_3)CH_2-$	418.1₅
$CH_3(CH_2)_7-$	449.1₅	$CH_3CH_2CH_2CH(CH_3)CH_2-$	421.1₅
$CH_3(CH_2)_8-$	465.1₅	$CH_3CH_2CH(CH_3)CH_2CH(CH_3)-$	432.9₅
$CH_3(CH_2)_9-$	490.1₅	$(CH_3CH_2)_2(CH_3)C-$	434.1₅
		$(CH_3(CH_2)_3)(CH_3CH_2)(CH_3)C-$	436.1₅
		$CH_3(CH_2)_6-$	449.9₅
		$CH_3(CH_2)_5C(CH_3)_2-$	451.1₅
		$(CH_3CH_2CH_2)_2(CH_3CH_2)C-$	455.1₅
		$CH_3CH(CH_3)CH_2(CH_2)_4-$	461.1₅
		$CH_3(CH_2)_7-$	467.5₅

For $f_\delta^2 \neq 0$ (*i.e., n* = 2) an improved single–index descriptor is obtained, while with f_δ^8 (*i.e., n* = 8) an improved term is obtained,

$$^{f2}\{D^v\}: F = 484, r^2 = 0.914, s = 14.1, N = 48$$
$$^{f8}X(Am+Al) = [(\chi_t^v)^{0.2} + 0.03 {}^1\chi^v] : F = 1150, r^2 = 0.962, s = 9.4, N = 48$$

The achieved improvement is substantial, especially the one obtained with the $^{f8}X(Am+Al)$ term. Now, every second compound starting with the first one in both classes of amines and alcohols of Table 3 (*i.e.,* 1°, 3°, 5° , …) will be excluded from the model but keeping n_H/n_O constant. Thus, the number of training amines is ten and the number of training alcohols is thirteen totaling $N = 23$ training points. The best descriptors for $f_\delta \neq 0$, and $f_\delta = 0$ are:

$$\{\chi_t\}: F = 131, r^2 = 0.861, s = 16.5$$
$$^{f2}\{D^v\}: F = 212, r^2 = 0.910, s = 13.3$$

$$X'(Am+Al) = [(\chi_t)^{0.4} + 0.08 {}^1\chi]^{0.6} : F = 192, r^2 = 0.901, s = 13.9$$

$$^{f8}X'(Am+Al) = [(\chi_t^v)^{0.2} + 0.03 {}^1\chi^v]^{1.4} : F = 693, r^2 = 0.970, s = 7.6;$$

Practically, the importance of the hydrogen perturbation for these two types of descriptors with a smaller training set does not change, while the optimal term is only slightly different from the previous one with $N = 48$. At the level of the composite class [amines + alcohols] N changes seem to affect only slightly the importance and type of the hydrogen perturbation. In Figure 3 it is

displayed the model of the twenty–three training points plus the twenty–five evaluating points with equation (utilities are in square parenthesis):

$$BP(Am + Al) = -599.922 \,^{f8}X'(Am+Al)[26] + 777.008 \,[53] \tag{15}$$
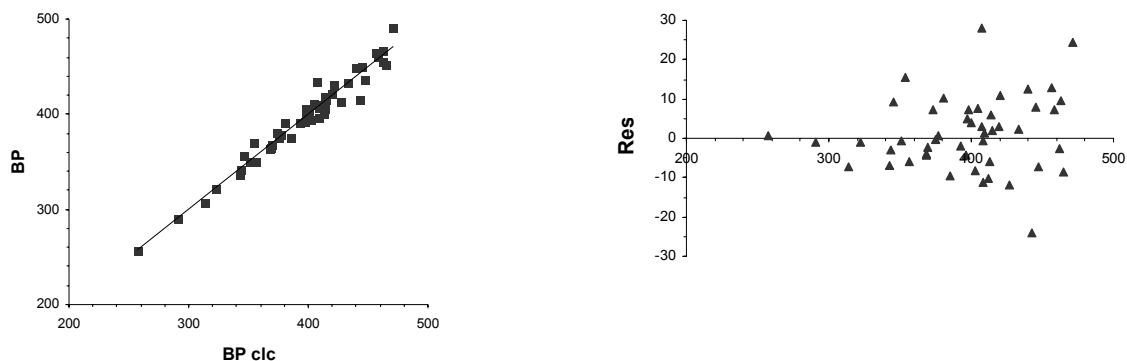
Figure 3. *Left*: Model of the boiling points of amines plus alcohols, where twenty–three points are training points and twenty–five points are evaluating points. *Right*: plot of the *BP* residuals *vs.* the calculated boiling point values.

This figure is hardly different from the model displayed in a plot (not shown) where all points have been considered training points. The residual plot shows that there are three outliers in the high *BP* region, *i.e.*, these residuals are well within a 10% error.

4 CONCLUSIONS

The recently introduced hydrogen perturbation [20,21], f_δ^n, into the definition for δ^v is able to encode, by the aid of graph concepts only, all molecular features, from the core electrons, to non–bonding electrons, inclusive multiple bonding, and, finally, is able to solve the degeneracy of many δ^v values which stem from the number of bonded hydrogen atoms. Molecular connectivity is now able, without making use of concepts external to graph theory, to obtain a specific index value for every type of compound. Present study on the hydrogen perturbation underlines that this perturbation is property–dependent, and *N*–dependent, as it is nicely underlined by the model of the boiling points of the composite class [amines+alcohols] relatively to the boiling points of amines and alcohols. The percent of hydrogen content of a molecule is not the only feature responsible for the f_δ^n values. Quite probably the recently suggested molecular variable connectivity index [28,29] embeds among other features also the hydrogen perturbation.

The fact that studied properties are mainly modeled by the aid of valence connectivity indices, *i.e.*, by indices which are δ^v–dependent underline the importance of a graph encoding of the overall electronic structure of a compound, *i.e.*, of electronic features that can only be encoded with general and complete graphs. The side–chain volume is mainly modeled by two vertex–based indices, D^v,

and $^0\chi^{\mathrm{v}}$, *i.e.*, the atoms contribute mainly the overall volume of the compound. The importance of the overall index, χ_t^{v}, as well as of the vertex–based index, $^0\chi^{\mathrm{v}}$, throughout the model of the isoelectric points, *pI*, tells that a delicate interplay between molecular and atomic characteristic is here at work. For the boiling points of the mixed class of amines and alcohols the contribution of the overall valence index χ_t^{v}, as well as of the edge valence index, $^1\chi^{\mathrm{v}}$, shows that molecular and bond characteristics are here playing an important role.

5 REFERENCES

[1]   L. B. Kier and L. H. Hall, *Molecular Connectivity in Structure–Activity Analysis*, Wiley & Sons, New York, 1986.

[2]   L. B. Kier and L. H. Hall, *Molecular Structure Description. The Electrotopological State*, Academic Press, New York, 1999.

[3]   M. Randić, On characterization of molecular branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609–6615.

[4]   S. C. Basak, A. T. Balaban, G. D. Grunwald, and B. D. Gute, Topological indices: their nature and mutual relatedness. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 891–898.

[5]   E. Estrada and L. Rodriguez, Edge–connectivity indices in QSPR/QSAR studies. 1. Comparison to other topological indices in QSPR studies. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1037–1041.

[6]   E. Estrada, Edge–connectivity indices in QSPR/QSAR studies. 2. Accounting for long–range bond contributions. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1042–1048.

[7]   E. Estrada and E. Molina, Novel local (fragment–based) topological molecular descriptors for QSPR/QSAR and molecular design. *J. Mol. Graph. Mod.* **2001**, *20*, 54–64.

[8]   E. Estrada, Physicochemical interpretation of molecular connectivity indices. *J. Phys. Chem. A* **2002**, *106*, 9085–9091.

[9]   J. Galvez, R. Garcia–Domenech, J. V. de Julian–Ortiz, and R. Soler, Topological approach to drug design. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 272–284.

[10] J. Galvez, On a topological interpretation of electronic and vibrational molecular energies. *J. Mol. Struct.* (*THEOCHEM*) **1998**, *429*, 255–264.

[11] J. Galvez, R. Garcia–Domenech, and C. Gregorio–Alapont, Indices of differences of path lengths: novel topological descriptors derived from electronic interferences in graphs. *J. Comput. –Aid. Molec. Des.* **2000**, *14*, 679–687.

[12] G. Grassy, B. Calas, A. Yasri, R. Lahana, J. Woo, S. Iyer, M. Kaczorek, R. Floc'h, and R. Buelow, Computer–assisted rational design of immunosuppressive compounds. *Nature Biotechn.* **1998**, *16*, 746–752.

[13] I. Gutman and E. Estrada, Topological Indices based on the line graph of the molecular graph. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 541–543.

[14] I. Gutman and Z. Tomović, More on the line graph model for predicting physico–chemical properties of alkanes. *Models in Chemistry* **2000**, *137*, 439–445.

[15] L. Pogliani, From molecular connectivity indices to semiempirical connectivity terms: recent trends in graph theoretical descriptors, *Chem. Rev.* **2000**, *100*, 3827–3858.

[16] L. Pogliani, Modeling Molecular Polarizabilities with Graph–Theoretical concepts, in *Computational Aspects of Electric Polarizability Calculations: Atom, Molecules, and Clusters* (G. Maroulis, Ed.), Part II, 737–751, IOS Press, Amsterdam, 2005, ISBN 1586034952; see also, L. Pogliani, *J. Comput. Meth. Sci .& Eng.* **2004**, *4*, 737–751.

[17] L. Pogliani, Encoding the core electrons with graph concepts, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 42–49.

[18] L. Pogliani, Model of Physical Properties of Halides with Complete Graph based Indices, *Int. J. Quant. Chem.* **2005**, *102*, 38–52.

[19] L. Pogliani, The complete graph conjecture for the inner–core electrons. The homogeneous index case, *J. Comput. Chem.* **2003**, *24*, 1097–1109.

[20] L. Pogliani, A Natural Graph–Theoretical Model for the Partition and Kinetic Coefficients, *New J. Chem.* **2005**,

29,1082–1088.

[21] L. Pogliani, The hydrogen perturbation in molecular connectivity computations. *J. Comput. Chem.* **2006**, *27*, 868–882.

[22] E. Besalu, J. V. de Julian–Ortiz, M. Iglesias, and L. Pogliani, An Overlooked Property of Plot methods. *J. Math. Chem.* (2006) to be published (accessible on line: 06 Jan 2006).

[23] E. Besalu, J. V. de Julian–Ortiz, and L. Pogliani, Some Plots Are not that Equivalent. *MATCH – Commun. Math. Comput. Chem.* **2006**, *55*, 281–286.

[24] M. Randić and S. C. Basak, On use of the variable connectivity index $^1\chi^f$ in QSAR: toxicity of aliphatic ethers, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 614–618.

[25] M. Randić and M. Pompe, The variable connectivity index $^1\chi^f$ versus the traditional molecular descriptors: A comparative study of $^1\chi^f$ against descriptors of CODESSA. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 631–638.

[26] R. Todeschini and V. Consonni, *Handbook of Molecular Descriptors*, Wiley–VCH, New York, 2000.

[27] M. Diudea (Ed.), *QSPR/QSAR studies by molecular descriptors*, Nova Science, New York, 2000.

[28] R. Carbó–Dorca, D. Robert, Ll. Amat, X. Girones, and E. Besalu, *Molecular Quantum Similarity in QSAR and Drug Design*, Springer, Berlin, 2000.

[29] A. M. Doweyko, 3D–QSAR illusions. *J. Comput.–Aid. Mol. Des.* **2004**, *18*, 587–596.

Biographies

Lionello Pogliani is professor of physical chemistry at the University of Calabria, Cosenza, Italy. After obtaining a Ph.D. degree in chemistry from the University of Firenze, Italy, the author undertook postdoctoral research with Dr. Eng. M. Ellenberger of the C.E.A. of Saclay, France, Prof. D. Ziessow of the University of Berlin (TU), and Prof. T. James of the University of California, San Francisco. More recently, the author has collaborated on projects with Professor M. N. Berberan–Santos of the University of Lisbon, and J. Galvez and R. Garcia–Domenech of the University of Valencia.