CROATICA CHEMICA ACTA
CCACAA **77** (1–2) 193–201 (2004)

ISSN-0011-1643
CCA-2917
Original Scientific Paper

Modeling with Indices Obtained from Complete Graphs*

Lionello Pogliani

Dipartimento di Chimica, Università della Calabria, 87030 Rende (CS), Italy
(E-mail: lionp@unical.it)

RECEIVED JANUARY 7, 2003; REVISED JULY 15, 2003; ACCEPTED JULY 18, 2003

Key words
complete graph algorithm
core electrons
molecular modeling
QSPR
halogenated compounds

Modeling quality of two different algorithms for the δ^v number, the main parameter for many molecular connectivity indices, has been tested. The new algorithms for δ^v are able to take into account the inner-core electrons of any heteroatom with the principal quantum number $n \geq 2$. Both algorithms are based on the concept of regular complete graphs, which are an appropriate tool for completing the graph representation of a molecule, extending it to the inner-core electrons of the atoms of the molecule. The model of a theoretical and two experimental properties was performed: the electron density at the bond critical point of fluorides and chlorides, the experimental lattice enthalpy of metal halides, and the experimental partition coefficient of halogenated organic compounds. The two algorithms for the δ^v number, based on two different types of complete graphs, are tested throughout the modeling of the quantum theoretical property and of the two experimental properties. While the odd complete graphs behave better than the sequential complete graphs, both types of algorithms for δ^v used with odd complete graphs are able to model, alternatively, the different properties in a satisfactory way.

INTRODUCTION

The goal of quantitative structure-property (QSPR) or structure-activity (QSAR) studies could be synthesized into the following logical assertion for every chemical graph G: if it has a set of graph-theoretical basis indices, $\{\beta\}$, then it has also the property (or activity) P, *i.e.*, $\forall$ G ($\{\beta\} \rightarrow$ P). Clearly, this assertion has only a probabilistic character. Another goal of QSPR studies is to define a set of consistent rules that could possibly give rise to an elaborated and coherent theory that can be used for modeling purposes at any moment. The molecular connectivity theory (MC), with its variant, the electrotopological state,[1–3] is such a theory. Recently, with the aim to widen the applicability and correct the incongruity of this theory, three new concepts have been introduced, which have further improved the modeling quality of the theory: the pseudoconnectivity indices, the dual indices, and the complete graphs for the core electrons.[4–8] The MC theory has recently undergone other improvements, such as, among others, the introduction of the variable connectivity index, the edge and line-graph connectivity index, the molecular connectivity terms, together with the many interesting interpretations developed for the connectivity index.[9–20]

In this paper, we will further check the modeling quality of the molecular connectivity and pseudoconnectivity indices, whose main parameter, the δ^v number, can encode the core electrons of heteroatoms with the principal quantum number $n \geq 2$. This is done with the aid of complete graphs, which proved to be an appropriate 'graph' tool for representation of the core electrons of atoms in a molecule.[21–24] Here, two algorithms, based on

---

* Dedicated to Professor Nenad Trinajstić on the occasion of his 65th birthday.

sequential and odd complete graphs, will be tested. They will be used to derive a set of basis indices, which will model a quantum theoretical property and two experimental properties of different classes of compounds. The quantum property is the electron density at the bond critical point of a class of fluorides and chlorides. The two experimental properties are the lattice enthalpy of metal halides, and the partition coefficient of halogenated organic compounds. The choice of these halogenated compounds allows checking of the modeling quality of those indices which are based on regular complete graphs. The quantum data for the first class of compounds are taken from Refs. 25 and 26, while the sets of experimental data are taken from Refs. 27 and 28, respectively.

METHOD

The Basis Indices

The set of molecular connectivity basis indices, $\{\beta\}$, used throughout this study can be divided into three different subsets. The subset of molecular connectivity indices, $\{\chi\}$, the subset of molecular pseudoconnectivity indices, $\{\psi\}$, and the subset of the dual basis indices, $\{\beta_\mathrm{d}\}$, which can be obtained from the previous ones with a Boolean-like algorithm,

$$\{\beta\} = \{\{\chi\}\{\psi\}\{\beta_\mathrm{d}\}\} \tag{1}$$

To avoid computational problems, due to the large number of indices belonging to each of these subsets, only a small fraction of all possible indices will be used, *i.e*,

$$\{\chi\} = \{\mathrm{D}, {}^0\chi, {}^1\chi, \chi_\mathrm{t}, \mathrm{D}^\mathrm{v}, {}^0\chi^\mathrm{v}, {}^1\chi^\mathrm{v}, \chi^\mathrm{v}{}_\mathrm{t}\}$$

$$\{\psi\} = \{{}^\mathrm{S}\psi_\mathrm{I}, {}^0\psi_\mathrm{I}, {}^1\psi_\mathrm{I}, {}^\mathrm{T}\psi_\mathrm{I}, {}^\mathrm{S}\psi_\mathrm{E}, {}^0\psi_\mathrm{E}, {}^1\psi_\mathrm{E}, {}^\mathrm{T}\psi_\mathrm{E}\}$$

$$\{\beta_\mathrm{d}\} = \{{}^0\chi_\mathrm{d}, {}^1\chi_\mathrm{d}, {}^1\chi_\mathrm{s}, {}^0\chi^\mathrm{v}{}_\mathrm{d}, {}^1\chi^\mathrm{v}{}_\mathrm{d}, {}^1\chi^\mathrm{v}{}_\mathrm{s}, {}^0\psi_\mathrm{Id}, {}^1\psi_\mathrm{Id}, {}^1\psi_\mathrm{Is}, {}^0\psi_\mathrm{Ed}, {}^1\psi_\mathrm{Ed}, {}^1\psi_\mathrm{Es}\}$$

Their definitions have been collected in pairs to underline their formal similarity,

$$D = \Sigma_i\,\delta_i \qquad {}^\mathrm{S}\psi_\mathrm{I} = \Sigma_i I_i \tag{2}$$

$${}^0\chi = \Sigma_i(\delta_i)^{-0.5} \qquad {}^0\psi_\mathrm{I} = \Sigma_i(I_i)^{-0.5} \tag{3}$$

$${}^1\chi = \Sigma(\delta_i\,\delta_j)^{-0.5} \qquad {}^1\psi_\mathrm{I} = \Sigma(I_i I_j)^{-0.5} \tag{4}$$

$$\chi_\mathrm{t} = (\Pi\delta_i)^{-0.5} \qquad {}^\mathrm{T}\psi_\mathrm{I} = (\Pi I_i)^{-0.5} \tag{5}$$

$${}^0\chi_\mathrm{d} = (-0.5)^N\Pi_i(\delta_i) \qquad {}^0\psi_\mathrm{Id} = (-0.5)^N\Pi_i(I_i) \tag{6}$$

$${}^1\chi_\mathrm{d} = (-0.5)^{(N+\mu-1)}\Pi(\delta_i+\delta_j) \qquad {}^1\psi_\mathrm{Id} = (-0.5)^{(N+\mu-1)}\Pi(I_i+I_j) \tag{7}$$

$${}^1\chi_\mathrm{s} = \Pi(\delta_i + \delta_j)^{-0.5} \qquad {}^1\psi_\mathrm{Is} = \Pi\,(I_i + I_j)^{-0.5} \tag{8}$$

Sums in Eqs. (2) and (3) as well as products (Π) in Eqs. (5) are over all the vertices of the hydrogen-suppressed chemical graph. Sums in Eqs. (4), are over all the edges of the chemical graph (σ bonds in a molecule). Replacing in Eqs. (2–8) δ with δ^v, and I_i with S_i, the corresponding valence χ^v indices and E-state ψ_E indices are obtained. Exponent μ in Eq. (7) is a cyclomatic number. This number is a molecular descriptor that indicates the number of bonds that must be broken in order to obtain a cycle-free structure. For acyclic molecules: $\mu = 0$; for monocyclic ones: $\mu = 1$ and for bicyclic ones: $\mu = 2$. While the basis χ indices and valence χ^v indices are directly related to the δ, and δ^v numbers of a hydrogen-suppressed graph and pseudograph, respectively, the basis ψ indices are indirectly related to δ and δ^v numbers through the I-state (ψ_I subset) and the S-state (ψ_E subset) atom level indices defined by Kier and Hall.[3]

$$I_i = [(2/n)^2\delta_i{}^\mathrm{v} + 1]/\delta_i \quad \text{and } S_i = I_i + \Sigma_j\Delta I_{ij} \tag{9}$$

$\Delta I_{ij} = (I_i - I_j)\ /\ r^2{}_{ij}$, where r_{ij} counts the atoms in the minimum path length separating two atoms, i and j, which equals the graph distance, $d_{ij} + 1$. As the $\delta_i{}^\mathrm{v}$ values are obtained from the hydrogen-suppressed pseudograph (ps) of a molecule, it is advantageous to redefine $\delta_i{}^\mathrm{v}$ in Eq. (9) as $\delta_i{}^\mathrm{v}(\mathrm{ps})$. In this case, the first equation can be rewritten as: $I_i = [\delta_i{}^\mathrm{v} + 1]/\delta_i$, where, $\delta_i{}^\mathrm{v} = (2/n)^2\delta_i{}^\mathrm{v}(\mathrm{ps})$. The contribution for the core electrons for $n \geq 2$ is here encoded by the $(2/n)^2$ parameter. The δ^v for the χ^v indices for atoms with $n > 2$ has, instead, been defined as $\delta^\mathrm{v} = Z^\mathrm{v}\ /\ (Z - Z^\mathrm{v} - 1)$.[2] Both definitions for δ^v introduce concepts that do not belong to graph theory. Recently,[8,21–24] to avoid this state of affairs, and render molecular connectivity more in keeping with chemical graph theory, complete graphs have been introduced to encode the core electrons of atoms with the principal quantum number $n \geq 2$ (heteroatoms). A graph G is complete if each pair of its vertices is adjacent. A complete graph of order p is denoted K_p, $(p - 1 = r)$ and is r-regular, where r denotes its regularity. A graph is regular if it has all vertices with the same degree r. Thus, a complete graph is always regular, even if a regular graph is not always complete. Figure 1 shows the hydrogen-suppressed chemical pseudograph-complete graph for the CH_3Br compound with p odd.

Two different δ^v numbers centered on parameters belonging to complete graphs can be derived from Eq. (10), as $q = 1$, and p.[23–24] Actually, these two values parallel the two 'quantum' values for δ^v previously seen. In any case, for $p = 1$, $\delta^\mathrm{v} = \delta^\mathrm{v}(ps)$. The two values for q will be used to derive the subset of valence χ^v indices and the subset of pseudo-ψ indices, dual indices included,

$$\delta^\mathrm{v} = q \cdot \delta^\mathrm{v}(\mathrm{ps})\ /\ (p \cdot r + 1) \tag{10}$$

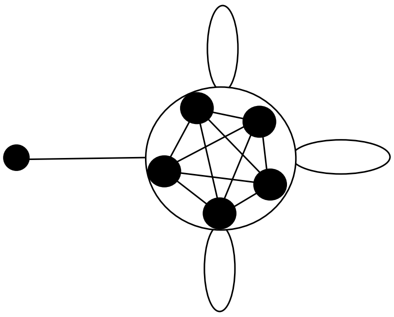

Figure 1. The hydrogen-suppressed pseudograph plus the complete graph of CH_3Br. The core electrons of the carbon atom are encoded with a K_1 complete graph, *i.e.*, a vertex. The core electrons of Br are encoded with a K_5 complete graph. The dimensions of this last complete graph have no metric meaning; it can be seen as a zoom of this vertex that allows noting its details.

Parameter $p \cdot r$ is an important parameter in graph theory since it represents the sum of the degree of the vertices of a K_p graph, which for the hand-shaking theorem equals twice the number of edges.[20] Two different types of complete graphs will be tested here, the sequential type where p can have sequential values, *i.e.*, $p = 1$, 2, 3, 4,….. for 2nd , 3rd, 4th, 5th, ... row atoms. To each value of δ^v for $q = 1$ and for $q = p$ an acronym will be assigned, *i.e.*, K_p-(p-seq), and K_p-(pp-seq), respectively. The odd type of complete graph has instead $p = 1, 3, 5, 7$,.... for 2nd , 3rd, 4th, 5th, ... row atoms. Consequently, even here we have two acronyms: K_p-(p-odd), and K_p-(pp-odd). Practically, we have four different values for the δ^v of each heteroatom.

Even regular complete graphs with $p = 2, 4, 6$,…. have been avoided due to conceptual difficulties. This choice would oblige one to encode the inner-core electrons for heteroatoms with $n = 2$, either with a K_0 null graph (with no vertices and edges),[29] or with a K_2 complete graph. This last possibility would oblige one to redefine most of the chemical graphs used to date, made up of K_1 vertices. Formally, the K_0 null graph could be used to encode the depleted hydrogens of a hydrogen-suppressed graph.

To obtain the δ^v values for $q = 1$, the adjacency matrix of the hydrogen-suppressed K_p-(p-odd or p-seq) chemical pseudograph of a molecule is needed. In Eq. (11), a seemingly symmetric adjacency pseudograph-K_p matrix for a tri-atomic system is shown,

$$\boldsymbol{A} = (p \cdot r + 1)^{-1}_{K_p} \begin{pmatrix} ps_{1,1} & g_{2,1} & g_{1,2} \\ g_{2,1} & ps_{2,2} & g_{2,3} \\ g_{3,1} & g_{3,2} & ps_{3,3} \end{pmatrix} \qquad (11)$$

Here, $g_{i,j}$ can be either 0 or 1. It is one only if vertices i and j are connected, otherwise it is zero; $ps_{i,i}$ is the sum of the self-connections (they count twice) and multiple connections of vertex i (pseudograph characteristics only, ps). Factor $(p{\cdot}r + 1)^{-1}{}_{Kp}$ encodes the complete-graph characteristics for the case $q = 1$. It depends on the p value of the complete graph used to model the core electrons of each atom. The final form of the matrix is normally asymmetric, as can be seen in the following example for the K_p-(p-odd) pseudograph adjacency matrix of MgI_2 (K_7 for I, and K_3 for Mg; the atom numbering indicates the row):

$$\boldsymbol{A}(I_1\, Mg_2\, I_3) = \begin{pmatrix} 6/43 & 1/43 & 0 \\ 1/7 & 2 & 1/7 \\ 0 & 1/43 & 6/43 \end{pmatrix}$$

Higher-order Descriptors

The basis indices can either be used in simple linear relationships, $P = c_1 \beta + c_0 U_0$ ($U_0 \equiv 1$), or in multilinear relationships, $P = \Sigma_i\, c_i \beta_i$ ($\beta_0 = U_0$), or they can be used, with the aid of a trial-and-error procedure, to build a series of powerful higher-order single descriptors (S). Three different types of higher-order descriptors are normally used, $S = X, Y$, and the mixed term Z: $X = f(\chi)$, $Y = f(\psi)$, and $Z = f(X,Y)$ or $f(X, Y, \beta)$, where $\beta = \chi$ or ψ. Dual indices are normally introduced at the Z level, giving rise to $Z' = f(X,Y, \beta_d)$ or $f(X, Y, \beta, \beta_d)$ terms. This procedure is used to short-circuit the huge combinatorial problem generated by twenty-eight indices. Clearly, it is also possible to construct terms like: $X' = f(\chi, \beta_d)$, $Y' = f(\psi, \beta_d)$.

The general form of terms X and Y is $S = [a(\beta_1)^m + b(\beta_2)^n]^q / [c(\beta_3)^o + d(\beta_1)^p]^r$, where β is a basis index, and $S = X$ or Y for $\beta = \chi$ or $\beta = \psi$, respectively. Parameters a–d, m–q, and r are optimization parameters. If some of these parameters are zero or one, the rational function assumes a simplified form.

Statistical performance of the graph-structural invariants, be it a basis index or S, is controlled by a quality factor, $Q = r / s$, and by the (Fischer) ratio $F = n_{fd}\, r^2/[(1-r^2)\nu]$, where r and s are the correlation coefficient and the standard deviation of the estimates, respectively, n_{fd} is the number of freedom degrees $= n - (\nu + 1)$, ν is the number of variables, and n is the number of data. Parameter Q has no absolute meaning, it can only compare the descriptive power of different descriptors within the same property. The F ratio, which has the character of an 'inter' statistical parameter, tells us, even if Q improves, which additional descriptor endangers the statistical quality of the combination. For every structural descriptor, be it a basis index of a linear combination, a term or the unitary descriptor, U_0, the fractional utility, $u_i = |c_i / s_i|$, has been defined; here s_i is the confidence interval of c_i. The average fractional utility $<u> = \Sigma u_i / (\nu + 1)$ will also be given. Utility statistics allows detecting descriptors that give rise to unreliable coefficient values (c_i) whenever they give rise to a high deviation interval (s_i). The standard deviation of the estimates, s, has a critical importance throughout these modeling studies. For this reason, the ratio $s_R = s_0/s_i$ has recently been introduced,[8,22] where s_0 is the s value of the best single basis index description and s_i refers to the s values of improved descriptions. This parameter gives a direct idea about how much s improves throughout a series of descriptions. The reader should keep in mind that (i) all statistical parameters will now grow with improving modeling, that (ii) every modeling will be under the control of all of these statistics, and that (iii) nothing justifies that an improved Q is a good receipt for good modeling. The abundance of statistical parameters can also be used to detect possible printing errors, since redundancy is very useful in self-correcting codes. To avoid bothering the reader with dimensional problems, every property P should be read as P/P° where P° is the unitary value of the property; this will allow to read P as a pure number.

Before leaving this section, it should be underlined that the factor $\Sigma_j \Delta I_{ij}$ in $S_i = I_i + \Sigma_j \Delta I_{ij}$ (see Eqs. 9) incorporates information about the influence of the remainder of the molecular environment; thus, highly electropositive atoms have $S < 0$.[3] To avoid this possibility, which could give rise to imaginary ψ_E values, a rescaling procedure is mandatory.[4,5] In this study, the S value for metal halides, MeX, has been rescaled to the S value of Ba in BaF_2, where, $S[Ba(BaF_2)] = -3.083$. The S value for the fluorides and chlorides has, instead, been rescaled to a hypothetical $S = -7.711 = (S[Si(SiF_4)] - 1.1)$, to avoid gigantic values for the ψ_E indices. For the halocompounds used in the partition coefficient section, the S values have been rescaled to $S[C(CF_4)] = -5.5$. The rescaling procedure, which is also done with the intent to avoid either too small or too large S_i values, has a minor influence on the quality of the modeling.[5]

RESULTS AND DISCUSSION

Electron Density at the Bond Critical Point, ρ_b, of Fluorides and Chlorides

The best algorithm for δ^v for this property of this class of compounds is the K_p-(*pp*-odd) algorithm, *i.e.*, $\delta^v = p \cdot \delta^v(ps) / (p \cdot r + 1)$ with p odd. The best single- and two-basis indices are:

$$\{^T\psi_I\}: Q = 11.03, F = 16, r = 0.682, s_0 = 0.06$$

$$\{D, {}^1\chi^v\}: Q = 30.36, F = 59, r = 0.935, s_R = 2.0, n = 20, \langle u \rangle = 9.7, \boldsymbol{u} = (10, 10, 8.8)$$

The best terms are:

$$X = [(D^v)^{0.7}({}^1\chi^v)^{1.1}\chi_t^v]^{-0.3}: Q = 29.56, F = 112, r = 0.928, s_R = 2.0, n = 20, \langle u \rangle = 7.8, \boldsymbol{u} = (11, 4.9)$$

$$Y = [({}^T\psi_I \cdot {}^1\psi_E)^{-1} + 0.6 \cdot {}^S\psi_E]: Q = 22.64, F = 66, r = 0.886, s_R = 1.54, n = 20$$

$$Z = [X - 0.003Y]^{1.8}: Q = 30.21, F = 117, r = 0.931, s_R = 2.0, n = 20, \langle u \rangle = 6.4, \boldsymbol{u} = (11, 1.9)$$

This last term can be improved by insertion of a dual index, giving rise to the following mixed dual term:

$$Z' = [Z - 0.8 \cdot ({}^1\chi^v{}_s)]^{1.1}: Q = 34.97, F = 157, r = 0.947, s_R = 2.22, \langle u \rangle = 6.5, \boldsymbol{u} = (13, 0.6), \boldsymbol{C} = (0.12552, 0.01250)$$

We will not let us be excessively bothered by the low utility values of constant term of the regression, $u_0 = 0.6$, since the value of the corresponding regression parameter, c_0, is nearly zero (0.0125) and even small deviations around zero can cause high deviations in $u_0 = c_0 / s_0$. Table I shows the original quantum mechanical values, the residuals relative to the present calculated values (with Z' term), and the residuals relative to the present calculated values with the leave-one-out method. In Figure 2, instead, quantum values *vs.* our own values and the corresponding residuals are plotted. As the original values are quantum values, it can be concluded that the two methods are not divergent, and that the molecular connectivity results are as good as the quantum results.

TABLE I. Electron density at the bond critical point of fluorides and chlorides, ρ_b/au, residuals of the modeled electron density at the bond critical point, $\Delta\rho$/au, and residuals of the modeled electron density at the bond critical point with the leave-one-out method, $\Delta\rho$(loo)/au

Compound[(a)]	ρ_b/au	$\Delta\rho$/au	$\Delta\rho$(loo)/au
LiF	0.075	–0.019	–0.021
BeF_2	0.145	–0.011	–0.011
BF_3	0.217	–0.010	–0.013
CF_4	0.309	–0.023	–0.032
NF_3	0.314	0.026	0.032
OF_2	0.295	0.025	0.029
F_2	0.288	0.028	–0.032
LiCl	0.047	–0.015	0.018
$BeCl_2$	0.097	0.003	0.003
BCl_3	0.157	0.038	0.041
CCl_4	0.182	0.041	0.044
NCl_3	0.176	0.022	0.023
OCl_2	0.184	0.014	0.015
FCl	0.187	0.002	–0.002
NaF	0.051	0.007	0.009
MgF_2	0.080	–0.010	–0.011
AlF_3	0.115	–0.025	–0.026
SiF_4	0.154	–0.069	–0.074
PF_3	0.168	–0.021	–0.023
SF_2	0.182	0.002	0.002

(a) The first fourteen data are taken from Ref. 25, the others are from Ref. 26.

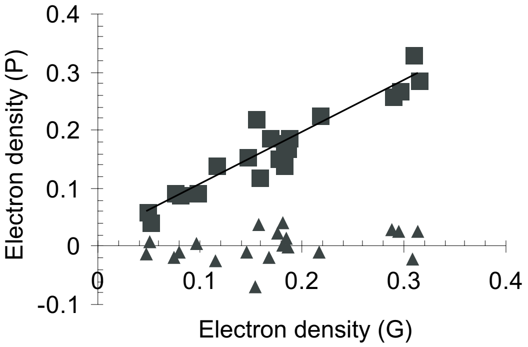

Figure 2. Plot of the electron density at the bond critical point, ρ_b/au of Gillespies' calculations (G) *vs.* present calculations (P), and plot of the corresponding residuals, $\Delta\rho_b$/au (▲).

Lattice Enthalpy of Metal Halides (MeX)

The lattice enthalpy, ΔH_L^o (kJ mol^{-1}), of twenty metal halides, MeX, at 298.15 K are collected in Table II, together with the corresponding calculated values with two different odd complete graph algorithms, the K_p-(p-odd), and the K_p-(pp-odd) algorithms, *i.e.*, $\Delta H_L^o(K_p)$, and $\Delta H_L^o(K_{pp})$, respectively. Table II also gives the corresponding modulus percent residual obtained with the two different sets of calculated values, $|\Delta\%| = |(\Delta H_L^o - \Delta H_L^o(K_i))100/\Delta H_L^o|$, with $i = p$ or pp. As in the preceding case, also here four complete graph algorithms have been considered for δ^v, two based on sequential complete graphs and two on odd complete graphs. About the possibility to model salts with graph concepts see Refs. 6, 20 and 25–27. Here, the vague and ill-defined concept of purely ionic bonds and purely covalent bonds is reconsidered.

TABLE II. Lattice enthalpies for twenty metal halides, ΔH_L^o (kJ mol^{-1}), at 298.15 K, the calculated K_p-(p-odd) values, ΔH_L^o(po), the corresponding modulus percent residual, |Δ%(po)|, the calculated K_p-(pp-odd) values, ΔH_L^o(ppo), and the corresponding residual, |Δ%(*ppo*)|.

MeX	ΔH_L^o	$\Delta H_L^o(po)$	$\lvert\Delta\%(po)\rvert$	$\Delta H_L^o(ppo)$	$\lvert\Delta\%(ppo)\rvert$
LiF	1037	1036	0.1	1034	0.3
NaF	926	910	1.7	902	2.6
KF	821	837	2.0	825	0.5
RbF	789	787	0.3	781	1.0
CsF	750	754	0.6	752	0.2
LiCl	852	872	2.3	886	4.0
NaCl	786	771	1.9	787	0.2
KCl	717	719	0.3	732	2.1
RbCl	695	687	1.2	702	1.0
CsCl	678	666	1.7	681	0.5
LiBr	815	813	0.2	805	1.2
NaBr	752	724	3.7	749	0.4
KBr	689	686	0.5	683	0.9
RbBr	668	664	0.7	660	1.3
CsBr	654	650	0.6	644	1.5
LiI	761	777	2.1	763	0.3
NaI	705	697	1.1	692	1.7
KI	649	664	2.3	656	1.2
RbI	632	646	2.2	637	0.8
CsI	620	635	2.5	624	0.7

The graphs of metal halides can be represented by two connected points: •—•. Thus, the only meaningful indices here are: $\{\chi\} = \{D^v, {}^0\chi^v, {}^1\chi^v\}$, and $\{\psi\}=\{{}^S\psi_I, {}^0\psi_I, {}^1\psi_I, {}^S\psi_E, {}^0\psi_E, {}^1\psi_E\}$. Two descriptions compete in modeling this property: the K_p-(p-odd) and K_p-(pp-odd) descriptions; the K_p-(p-seq) and K_p-(pp-seq) descriptions, instead, lead to poorer modeling.

K_p-(p-odd) Description. – The K_p-(p-odd) description starts with a deceptive single-ψ descriptor:

$$\{{}^0\psi_E\}\text{: } Q = 0.023,\ F = 108,\ r = 0.926,\ s_0 = 40.6$$

but improves consistently with the following combination:

$$\{\,{}^0\chi^v, {}^1\psi_E\}\text{: } Q = 0.042,\ F = 183,\ r = 0.978,\ s_R = 1.8,$$
$$\boldsymbol{u} = (8.1,\ 11,\ 31),\ \langle u\rangle = 16,\ n = 20$$

It improves even more with the following X term:

$${}^{\text{odd}}X_p = [(D^v)^{0.5}/(D^v + 13.3\cdot({}^0\chi^v)^{0.9}]^{0.6}$$

$$Q = 0.056,\ F = 654,\ r = 0.987,\ s_R = 2.3,\ n = 20,$$
$$\boldsymbol{u} = (26,\ 74),\ \langle u\rangle = 50,\ n = 20$$

Thanks to a less satisfactory Y term, ${}^{\text{odd}}Y_p = ({}^0\psi_E + {}^1\psi_E)$: $Q = 0.025$, $F = 133$, $r = 0.938$, $s_R = 1.1$, a quite interesting ${}^{\text{odd}}Z_p$ term could be detected:

$${}^{\text{odd}}Z_p = [{}^{\text{odd}}X_p + 0.3\ {}^{\text{odd}}Y_p]$$

$$Q = 0.068,\ F = 953,\ r = 0.991,\ s_R = 2.8,$$
$$\boldsymbol{u} = (31,\ 10),\ \langle u\rangle = 21,\ n = 20$$

Further, insertion of a zeroth-order dual basis index improves the quality of the description with the following mixed dual term:

$${}^{\text{odd}}Z'_p = [1.2\cdot Z^{0.3} + 0.004({}^0\chi^v{}_d)^2]^{0.1}$$

$$Q = 0.074,\ F = 1145,\ r = 0.992,\ s_R = 3.0,$$
$$\boldsymbol{u} = (34,\ 33),\ n = 20,\ \boldsymbol{C} = (36789.0,\ -35815.7)$$

The ΔH_L^o values, calculated with the aid of this ${}^{\text{odd}}Z'_p$ term, $\Delta H_L^o(po)$, and the corresponding modulus percent residual, $|\Delta\%(po)| = |\Delta H_L^o - \Delta H_L^o(K_p)|\,100/\,\Delta H_L^o$, are shown in Table II. In Figure 3, the calculated *vs.* the experimental values (top) and the corresponding residuals, $\Delta H_L^o - \Delta H_L^o(po)$ (bottom) are plotted. The agreement among the data seems satisfactory.

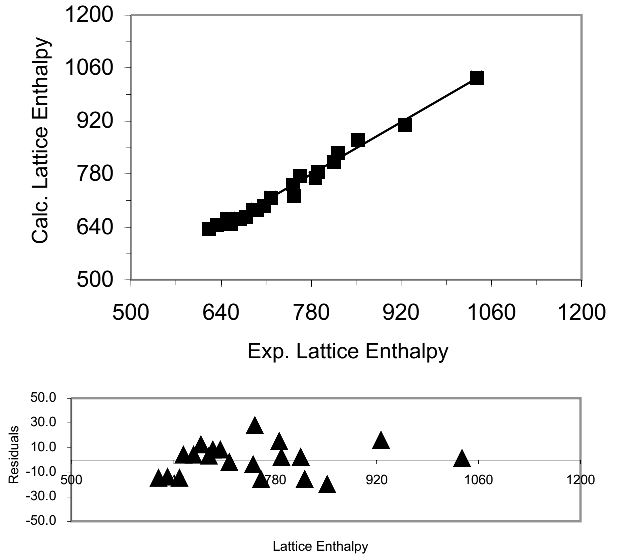

Figure 3. Top: plot of the calculated (with K_p-(p-odd) algorithm) *vs.* the experimental lattice enthalpies, $\Delta H_L{}^o$, for twenty metal halides, MeX. Bottom: the corresponding residuals.

K_p-(pp-odd) Description. – This description is slightly better than the preceding one, but the difference is not dramatic. The best single- and two-basis index descriptions are ($s_0 = 40.6$):

$$\{{}^0\psi_I\}\text{: } Q = 0.028,\ F = 159,\ r = 0.948,\ s_R = 1.2,$$
$$\boldsymbol{u} = (13, 23),\ \langle u\rangle = 18,\ n = 20$$

$$\{{}^0\psi_I,\ {}^1\psi_I\}\text{: } Q = 0.041,\ F = 173,\ r = 0.976,\ s_R = 1.7,$$
$$\boldsymbol{u} = (8.2, 4.4, 16),\ \langle u\rangle = 9.7,\ n = 20$$

This last description, based on ψ_I indices alone, is quite similar to the description achieved by the ψ_I indices calculated with the $\delta^v=(2/n)^2\,\delta^v(\mathrm{ps})$ algorithm. Actually, the two algorithms give rise to similar, but not equal δ^v values; in fact, the small difference plays a major role in determining the better quality of the higher-order terms based on the K_p-(*pp*-odd) algorithm. The following *X, Y* (a deceptively simple term), *Z*, and Z′ terms can be detected:

$${}^{\mathrm{odd}}X_{pp} = (D^v)^{0.4}/[0.9D^v + 50({}^0\chi^v)^{1.1}]^{0.9}$$

$$Q = 0.071,\ F = 1035,\ r = 0.991,$$
$$s_R = 2.9,\ \boldsymbol{u} = (32, 71),\ \langle u\rangle = 52,\ n = 20$$

The best *Y* term is of inferior quality and gives rise, together with the *X* term, to a *Z* term, which is hardly better than the *X* term:

$${}^{\mathrm{odd}}Y_{pp} = ({}^0\psi_I)^{0.01}$$

$$Q = 0.031,\ F = 206,\ r = 0.959,\ s_R = 1.3,\ n = 20$$

$${}^{\mathrm{odd}}Z_{pp} = (0.6X + 0.1Y^3)^{0.01}$$

$$Q = 0.072,\ F = 1087,\ r = 0.992,\ s_R = 3.0,\ n = 20$$

Anyway, with this last term and with the help of a dual basis index, it is possible to derive a higher-order dual term with an overall improved quality:

$${}^{\mathrm{odd}}Z'_{pp} = [Z^{0.9} + 0.0001({}^1\chi^v{}_s)]^{0.2}$$

$$Q = 0.081,\ F = 1374,\ r = 0.994,\ s_R = 3.3,$$
$$u = (37, 37),\ n = 20,\ \boldsymbol{C} = (258391, -252589)$$

The ΔH_L values, calculated with the aid of this $Z'[K_p$-(*pp*-odd)] term, $\Delta H_L{}^o(ppo)$, and the corresponding modulus percent residual, $|\Delta\%(ppo)| = |\Delta H_L{}^o - \Delta H_L{}^o(K_p)|\,100/\Delta H_L{}^o$, are shown in the last two columns of Table II. In Figure 4, the calculated *vs.* the experimental values (top) and the corresponding residuals, $\Delta H_L{}^o - \Delta H_L{}^o(ppo)$ (bottom) are plotted. Even here the agreement between experimental and calculated data is quite satisfactory. We leave to the reader the choice between the K_p-(*p*-odd) and K_p-(*pp*-odd) representation. Table II shows two anomalous residuals, a 4 % one for LiCl-(*pp*o), and a 3.7 % one for NaBr-(*p*o). The origin of this relative anomaly is far from obvious, considering that the two kinds of atoms involved give rise, with other atoms, to decent residuals. The reason could be that a small discrepancy in the δ^v value of each atom adds up, giving rise in each case, K_p-(*p*-odd) for NaBr and K_p-(*pp*-odd) for LiCl, to a large discrepancy, *i.e.*, to a large residual.

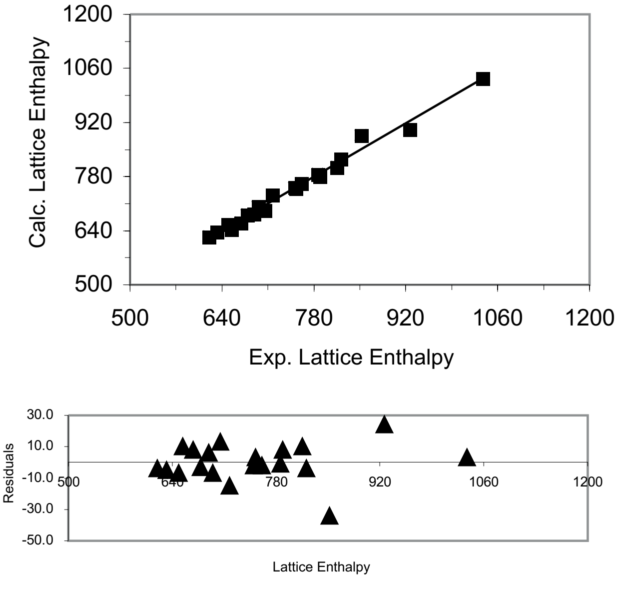

Figure 4. Top: plot of the calculated (with K_p-(pp-odd) algorithm) *vs.* the experimental lattice enthalpies, $\Delta H_L{}^o$, for twenty metal halides, MeX. Bottom: the corresponding residuals.

Partition Coefficient (PC) of Halogenated Organic Compounds

The air partition coefficient, at 37 °C, of the rat fat tissue of twenty-five halogenated methanes, ethanes, and ethylenes and their Log values are collected in Table III, together with the corresponding calculated values, $\mathrm{Log(PC)_C}$, whose χ^v, and $\psi_{I,E}$ values have been obtained with the odd complete K_p-(p-odd) graph algorithm. Table III also shows the corresponding modulus percent residual obtained, $|\Delta\%\mathrm{Log(PC)}| = |\mathrm{Log(PC)} - \mathrm{Log(PC)_C}|100/\mathrm{Log(PC)}$. The experimental data were taken from Ref. 24, where all values are mean values with a maximal error of *ca.* 10 %.

TABLE III. Air partition coefficients (PC), at 37 °C, of rat fat tissue, Log(PC), calculated $\mathrm{Log(PC)_C}$, and modulus percent residual $|\Delta\%\mathrm{Log(PC)}|$

Molecule	Log(PC)	$\mathrm{Log(PC)_C}$	$\lvert\Delta\%\mathrm{Log(PC)}\rvert$
MeCl	1.130	1.118	1.1
$MeCl_2$	2.079	1.869	10
$CHCl_3$	2.307	2.305	0.1
CCl_4	2.555	2.628	2.9
CH_2=CHCl	1.301	1.296	0.4
CCl_2=CH_2	1.836	1.781	3.0
CHCl=CHCl(*cis*)	2.356	2.117	10
CHCl=CHCl(*trans*)	2.170	2.117	2.5
CCl_2=CHCl	2.744	2.608	4.9
CCl_2=CCl_2	3.214	3.101	3.5
CH_3-CH_2Cl	1.587	1.869	18
$CHCl_2$–CH_3	2.215	2.305	4.1
CH_2Cl–CH_2Cl	2.537	2.658	4.8
CCl_3–CH_3	2.420	2.628	8.6
$CHCl_2$–CH_2Cl	3.158	3.110	1.5
$CHCl_2$–$CHCl_2$	3.576	3.569	0.2
CCl_3–CH_2Cl	3.332	3.443	3.3
CH_2F_2	0.155	0.333	115
CH_2FCl	1.188	1.101	7.3
CH_2BrCl	2.512	2.352	6.4
CH_2Br_2	2.899	2.834	2.2
CF_3–CHClBr	2.260	2.209	2.2
CH_2=CHBr	1.692	1.735	2.6
CH_2Br–CH_2Cl	2.982	3.140	5.3
CF_3–CH_2Cl	1.326	1.304	1.7

Even here the four-valued complete graph algorithm for δ^v has been considered. Only the δ^v-K_p-(p-odd) case leads to satisfactory results. The other cases give rise to a deceiving model. Nevertheless, for comparison purposes, we will also show some results of the δ^v-K_p-(pp-odd) case. Let us start telling the reader that the previous model by Cargas *et al.*[28] was based on a mixed set of five different parameters: two $\chi^v = f[\delta^v = Z^v/(Z - Z^v - 1)]$ indices, a Cl index, an F index, and a polar hydrogen factor index. All together they generated an interesting model:

$$\{^1\chi^v, {}^3\chi^v_c, N_F, N_{Cl}, Q_H\}: Q = 9.04, F = 256,$$
$$r = 0.993, s = 0.11, n = 25$$

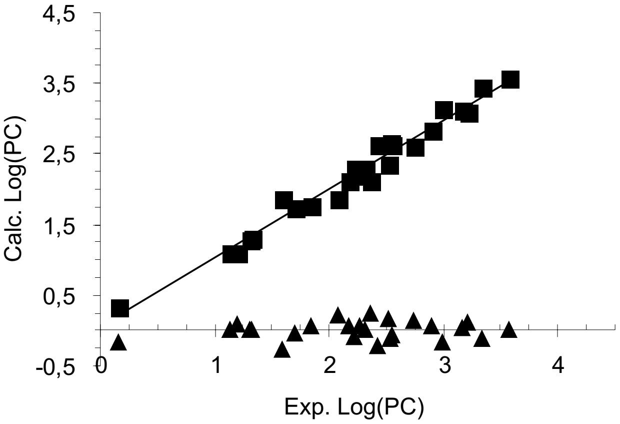

Figure 5. Top: plot of the calculated (with K_p-(p-odd) algorithm) *vs.* the experimental lattice partition coefficient, Log(PC), for twenty five halogenated organic compounds, and plot of the corresponding residuals (▲).

If we restrict the number of parameters used for this model, *e.g.*, to two indices alone, we have:

$$\{^1\chi^v, {}^3\chi^v_c\}: Q = 2.69, F = 57, r = 0.915,$$
$$s = 0.34, n = 25$$

Here, the second index, $^3\chi^v_c$, is rather useless; in fact, for the single best index we have:

$$\{^1\chi^v\}: Q = 2.67, F = 111, r = 0.910,$$
$$s = 0.34, n = 25$$

Consistent deterioration of the F value produced by the $^3\chi^v_c$ index is evident.

K_p-(p-odd) Description. – In the following lines, descriptions with no more than three indices will be considered, a choice that seems reasonable for $n = 25$ points.

$$\{^1\chi^v\}: Q = 2.815, F = 124, r = 0.918,$$
$$s_0 = 0.34, n = 25$$

$$\{^0\chi, {}^0\psi_I\}: Q = 5.813, F = 264, r = 0.980,$$
$$s_R = 2.0, n = 25$$

$$\{^0\chi, {}^0\psi_I, {}^1\chi^v\}: Q = 7.116, F = 264, r = 0.987,$$
$$s_R = 2.4, n = 25, \langle u \rangle = 7.0$$
$$\boldsymbol{u} = (10, 9.8, 3.5, 4.5);$$
$$\boldsymbol{C} = (-1.08143, 1.64032, 0.42748, 0.53332)$$

Four points can be highlighted: (i) the good quality of the present combinations, (ii) the positive character of every new basis index, whose introduction does not deteriorate the F statistics. The third and fourth point are: (iii) the impressive statistics of the two-basis index combination, and (iv) the good quality of the three-basis index combination achieved with the help of no 'external' index. The r_{loo} obtained, by the leave-one-out method, with the {$^0\chi$, $^0\psi_I$, $^1\chi^v$} combination, is just equal to the obtained r up to the third decimal figure. The calculated Log(PC)$_C$ values of Table III and the modulus percent residual, |Δ%Log(PC)|, have been obtained with {$^0\chi$, $^0\psi_I$, $^1\chi^v$} and with vector ***C***. Figure 5 has been obtained with this combination and with its correlation vector. Here, the calculated values are plotted *vs*. the experimental ones, together with their residuals.

K_p-(pp-odd) Description. – For comparison, we will give the quality of the (second) best descriptions obtained with the K$_p$-(*pp*-odd) algorithm for δ^v (s_0 = 0.34), which are self-explanative:

$$\{^1\chi^v\}: Q = 1.534,\ F = 37,\ r = 0.784,\ s_R = 0.67,\ n = 25$$

$$\{D^v, {}^1\chi\}: Q = 2.957,\ F = 68,\ r = 0.928,\ s_R = 1.1,\ n = 25$$

$$\{D^v, {}^1\chi, {}^S\psi_E\}: Q = 3.214,\ F = 54,\ r = 0.941,\ s_R = 1.2,\ n = 25$$

CONCLUSIONS

The core electrons of atoms that have so far been ignored by chemical graph theory can be encoded advantageously, for what concerns the molecular connectivity theory, with odd complete graphs. Thus far, the δ^v valence number used to be two-valued, following the two values of q parameter of Eq. (10), *i.e.*, q = 1, and p. The two-valued δ^v is based on the concept of odd complete graphs, and gives rise to the K$_p$-(p-odd) and K$_p$-(pp-odd) cases, where p = odd. These two δ^v values can be used to derive the basis χ^v and ψ indices. The modeled properties of the three different halogenated classes of compounds show that it is not possible to predict *'a priori'* which δ^v works better for a class of compounds. The choice can be made only with considerations based on *'a posteriori'* arguments. Nevertheless, these results together with the results from other studies [8, 23-24] suggest that to 'graph' encode the core electrons of atoms in molecules, odd complete regular graphs are more useful than other types of complete graphs. The model achieved for the first two properties underlines also the positive role of the recently introduced dual indices. The quality of the modeling achieved for the quantum-theoretical property indicates, instead, the parallelism of the two computing methods.

Recently, Basak *et al*. published a work on forty-six partition coefficients,[30] where the aim was to optimize the model, a somewhat different task from ours. Here, our intent was to check different molecular connectivity complete graph conjectures with different sets of properties of different classes of compounds.

Acknowledgements. – The author thanks the three referees for their support and suggestions.

REFERENCES

1. M. Randić, *J. Am. Chem. Soc.* **97** (1975) 6609–6615.
2. L. B. Kier and L. H. Hall, *Molecular Connectivity in Structure-Activity Analysis*, Wiley, New York, 1986.
3. L. B. Kier and L. H. Hall, *Molecular Structure Description. The Electrotopological State*, Academic Press, New York, 1999.
4. L. Pogliani, *J. Phys. Chem. A* **104** (2000) 9029–9045.
5. L. Pogliani, *J. Chem. Inf. Comput. Sci.* **41** (2001) 836–847.
6. L. Pogliani, *Croat. Chem. Acta* **75** (2002) 409–432.
7. L. Pogliani, *J. Mol. Struct. (THEOCHEM)* **581** (2002) 87–109.
8. L. Pogliani, *J. Chem. Inf. Comput. Sci.* **42** (2002) 1028–1042.
9. B. Lučić, S. Nikolić, N. Trinajstić, D. Juretić and A. Jurić, *Croat. Chem. Acta* **68** (1995) 435–450.
10. M. Randić and S. C. Basak, *J. Chem. Inf. Comput. Sci.* **41** (2001) 614–618.
11. M. Randić and M. Pompe, *J. Chem. Inf. Comput. Sci.* **41** (2001) 575–581.
12. N. Kezele, L. Klasinc, J. von Knop, S. Ivaniš, and S. Nikolić, *Croat. Chem. Acta* **75** (2002) 651–665.
13. E. Estrada and L. Rodriguez, *J. Chem. Inf. Comput. Sci.* **39** (1999) 1037–1041.
14. E. Estrada, *J. Chem. Inf. Comput. Sci.* **39** (1999) 1042–1048.
15. S. C. Basak, S. Nikolić, and N. Trinajstić, *J. Chem. Inf. Comput. Sci.* **40** (2000) 927–933.
16. S. Nikolić, N. Trinajstić, and S. Ivaniš, *Croat. Chem. Acta* **72** (1999) 875–883.
17. J. Galvez, *J. Mol. Struct. (THEOCHEM)* **429** (198) 255–264.
18. L. B. Kier and L. H. Hall, *J. Chem. Inf. Comput. Sci.* **40** (2000) 792–795.
19. E. Estrada, *J. Phys. Chem. A* **106** (2002) 9085–8091.
20. L. Pogliani, *Chem. Rev.* **100** (2000) 3827–3858.
21. L. Pogliani, *Data Sci. J.* **2** (2003) 1–11.
22. L. Pogliani, *New J. Chem.* **27** (2003) 919–927.
23. L. Pogliani, *J. Comput. Chem.* **24** (2003) 1097–1109.
24. L. Pogliani, *Indian J. Chem.* **42** (2003) 1347–1353.
25. C. F. Matta and R. J. Gillespie, *J. Chem. Educ.* **79** (2002) 1141–1152.
26. R. J. Gillepsie, *J. Chem. Educ.* **78** (2001) 1688–1692.
27. L. Pogliani, *Croat. Chem. Acta* **70** (1997) 803–817.
28. M. L. Cargas, P. G. Seybold, and M. E. Andersen, *Toxicol. Lett.* **43** (1988) 235–256.
29. F. Harary and R. Read, *Proc. Graphs and Combinatorics Conference*, George Washington University, Springer, New York, 1977.
30. S. C. Basak, D. Mills, D. M. Hawking, and H. A. El-Masri, *SAR QSAR Environ. Res.* **13** (2002) 649–665.

SAŽETAK

Modeliranje s indeksima dobivenim iz potpunih grafova

Lionello Poglianni

Uspoređena je kvaliteta modeliranja dvaju algoritama za računanje δ^v, koji je glavni parametar na kojemu se temelje mnogi indeksi povezanosti. Ti algoritmi uzimaju u obzir i doprinose unutarnjih, a ne samo valentnih elektrona za bilo koji heteroatom s glavnim kvantnim brojem $n \geq 2$. Oba se algoritma temelje na regularnim potpunim grafovima, koji su pogodni za prikaz molekula, kada se uz valentne elektrone uzimaju u obzir i elektroni unutarnjih ljusaka. Jedan se algoritam temelji na neparnim potpunim grafovima, a drugi na sekvenci potpunih grafova. Prikazano je modeliranje triju molekularnih svojstava: elektronske gustoće kritične veze kod fluorida i klorida, entalpije rešetke metalnih halogenida i particijskih koeficijenata halogeniranih organskih spojeva. Krajnji je zaključak ovoga rada kako se ne može *a priori* predvidjeti koji je algoritam bolji za predviđanje svojstava neke klase molekula.