CROATICA CHEMICA ACTA CCACAA **75** (2) 409–432 (2002)

ISSN-0011-1643
CCA-2809
Original Scientific Paper

Higher-level Descriptors in Molecular Connectivity*

Lionello Pogliani

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

Received March 12, 2001; revised June 7, 2001; accepted June 8, 2001

The optimal modeling of the lattice enthalpy of metal halides, of the boiling points of the mixed class of alcohols plus amines and of two activities of chlorofluorocarbons, the rates of hydrogen abstraction and the minimum anesthetic concentrations, has been achieved thanks to the introduction of three new types of higher-order terms. The first type of term, the mixed molecular connectivity-pseudoconnectivity term, $Z = f(X, Y)$, is function of the molecular connectivity term, X, and of the molecular pseudoconnectivity term, Y, only. Terms X and Y are function of a basis index β that can be either a molecular connectivity index, χ, or a molecular pseudoconnectivity index, ψ. The Z term can be found by trying different operational combinations of X and Y terms. The second and third type of mixed higher-order terms, $Z' = f(Z, \beta)$ and $Z' = f(X, Y, \beta)$, can be found by the aid of a trial-and-error search procedure in which Z or X and Y terms are held constant throughout the search. Modeled properties show these three types of higher-order mixed terms at work: the two different properties of chlorofluorocarbons are modeled by the aid of a $Z = f(X, Y)$ term, the lattice enthalpy of metal halides is instead modeled by a $Z' = f(Z, \beta)$ term, and the boiling point of the mixed class alcohols plus amines is modeled by a $Z' = f(X, Y, \beta)$ term.

Key words: alcohols, amines, chlorofluorocarbons, connectivity index, lattice enthalpy of metal halides, pseudoconnectivity index, QSPR models.

> *A physicist is a person who can calculate*
> *anything within an order of magnitude.*
> N. J. Giordano (Computational Physics, Prentice-Hall, 1977)

---

[*] Dedicated to Professor Milan Randić on the occasion of his 70th birthday.

INTRODUCTION

Randić in 1975 (Ref. 1) proposed an algorithm for molecular modeling around which a new theory for molecular modeling, the molecular connectivity theory, MC, based on topological concepts, was developed and successfully used in QSAR/QSPR studies.[2,3] The new theory proposed a very simple and elegant way to solve one of the main problems of material science, which is synthesized in Scheme I, and which has recently been published in a number of Science:[4]

$$\textit{Processing} \Leftrightarrow \textit{Structure} \Leftrightarrow \textit{Property or Activity} \Leftrightarrow \textit{Performance.} \qquad \text{(I)}$$

The main point of Scheme I, which QSAR/QSPR studies have tried and are trying to solve, is its central moment shown in Scheme II:

$$\textit{Structure} \Leftrightarrow \textit{Property or Activity.} \qquad \text{(II)}$$

It is this central moment that has mainly concerned, during these last twenty-five years, many scientists working in the field of mathematical chemistry.[1–10] During the last year it has been shown that with molecular connectivity indices, completed with indices based on the I_S-State concept,[11–13] it was possible to relate the topological structure of a compound, encoded by a higher-order order structural descriptor (S), with its property (P) or Activity (A), through a very efficient and simple linear equation, $P = f(S)$. The higher-order descriptor can be built either by the aid of molecular connectivity χ indices or by the aid of molecular pseudoconnectivity ψ indices, and in some cases by both type of basis indices.

METHOD

Throughout the present work we will be concerned with higher-level molecular structural descriptors whose basis indices, β, are the following eight molecular connectivity indices of subset $\{\chi\}$, and the following eight molecular pseudoconnectivity indices of subset $\{\psi\}$,

$$\{\beta\} = \{\{\chi\}\{\psi\}\} =$$
$$\{\{\mathrm{D}, {}^0\chi, {}^1\chi, \chi_\mathrm{t}, \mathrm{D}^\mathrm{v}, {}^0\chi^\mathrm{v}, {}^1\chi^\mathrm{v}, \chi_\mathrm{t}{}^\mathrm{v}\}\{{}^\mathrm{S}\psi_\mathrm{I}, {}^0\psi_\mathrm{I}, {}^1\psi_\mathrm{I}, {}^\mathrm{T}\psi_\mathrm{I,} {}^\mathrm{S}\psi_\mathrm{E}, {}^0\psi_\mathrm{E}, {}^1\psi_\mathrm{E}, {}^\mathrm{T}\psi_\mathrm{E}\}\}. \quad (1)$$

While the first eight indices can directly be derived with well-known algorithms defined within the frame of MC theory, the second type of indices are directly derived from the I_S-State concept, and through this concept they are indirectly related to the MC paradigm.[3,11] While the χ basis indices can be used to derive the molecular connectivity terms, $X = f(\chi)$, the ψ basis indices can be used to derive the molecular pseudoconnectivity term, $Y = f(\psi)$. Both terms are derived with a trial-and-error procedure. Once these terms are obtained it is possible to combine them and derive a mixed higher-order molecular connectivity-pseudoconnectivity term, $Z = f(X, Y)$. Further, by the aid of a trial-and-error procedure, where X and Y or Z terms are held constant it is possible to derive either a $Z' = f(X, Y, \beta)$ or a $Z' = f(Z, \beta)$ term.

The great advantage of using terms is that normally the relation between the higher-order structural invariant, S, and the property (or activity) is linear (Eq. 2), and only seldom, improved multiple linear regressions, $P = f(S, \beta_1,..., \beta_n)$ can be found[3,12,13]

$$P \text{ (or } A) = c_1 S + c_0 U_0 \ . \tag{2}$$

Here $S = X, Y, Z$, or Z', and $U_0 \equiv 1$. To avoid unexpected negative calculated P values (for the dimensional meaning of P see Ref. 14), with no biological or physical meaning that can further degrade the quality of the modeling it is better to use the modulus modeling equation, $P = |c_1 S + c_0 U_0|$, where bars stand for absolute value. This procedure normally enhances the description, provided that the experimental activities or properties are positive. If some experimental A or P values are negative then modulus bars should be omitted and Eq. (2) should be used. While the higher-level mixed Z or Z' terms are made up of X and Y terms, these last terms have to be constructed directly from basis indices, β, and they have different forms, which can be derived from the rational function of Eq. (3), where $S = X$ if $\beta = \chi$, and $S = Y$ if $\beta = \psi$

$$S = [a(\beta_1)^m + b(\beta_2)^n]^q \ / \ [c(\beta_3)^o + d(\beta_1)^p]^r \ . \tag{3}$$

Here $a - d$, $m - q$, and r are optimization parameters that can also be negative, zero or one. Clearly, when some of these optimization parameters are zero and/or one, relation (3) shrinks to a quite elementary expression. The use of terms in molecular connectivity modeling remind us the use of configuration interaction (CI) of molecular orbitals (MO) made up of Gaussian type orbitals (GTO) in quantum chemistry calculations. For this reason this method could loosely be called Configuration Interaction of adjustable Graph-Type Basis Indices (CI-a-GTBI).

Let us now see in detail (i) the origin of the newly introduced ψ indices (ii) and how are defined the single connectivity and pseudoconnectivity indices. While χ indices are directly based on the δ and δ^{v} connectivity numbers of a graph and a pseudograph respectively, ψ indices are indirectly based on

δ and δ^{v} numbers through the I-State (ψ_{I} subset) and S-State (ψ_{E} subset) indices,[10] which are defined in Eqs. (4) and (5) (S index here should not be confused with the structural term, S, of Eqs. (2) and (3)).

$$I_i = [(2/N)^2\delta_i^{\mathrm{v}} + 1] / \delta_i \tag{4}$$

$$S_i = I_i + \Sigma_j \Delta I_{ij} \tag{5}$$

Here, N = principal quantum number, $\Delta I_{ij} = (I_i - I_j) / r^2_{ij}$, and r_{ij} = counts of atoms in the minimum path length separating two atoms i and j, which is equal to the usual graph distance $d_{ij} + 1$. It should here be noticed that index S, through term $\Sigma_j \Delta I_{ij}$ in Eq. (5), incorporates, at the atomic level, the information about the influence of the remainder of the molecular environment, and for this reason it can even be negative. These two atom-level indices encode simultaneously graph and pseudograph characteristics, as they are directly (I) and indirectly (S) based on δ and δ^{v} numbers of a graph and a pseudograph, respectively.[10] Now, as a graph is a special case of a pseudograph[3] the name pseudoindices for the I_S-ψ indices seems to be well earned. Indices of subsets $\{\chi\}$ and $\{\psi\}$ of Eq. (1), as well as subsets of these subsets, share a formally similar definition, for this reason only ψ_{I} indices will here be defined

$$^{\mathrm{S}}\psi_{\mathrm{I}} = \sum_i I_i \tag{6}$$

$$^{0}\psi_{\mathrm{I}} = \sum_i (I_i)^{-0.5} \tag{7}$$

$$^{1}\psi_{\mathrm{I}} = \sum (I_i I_j)^{-0.5} \tag{8}$$

$$^{\mathrm{T}}\psi_{\mathrm{I}} = (I_1 \cdot I_2 \cdot I_3 \cdot \ldots\ldots\ldots \cdot I_N)^{-0.5} \ . \tag{9}$$

Sum in Eqs. (6) and (7) are taken over all N atoms of a molecule, and sum in Eq. (8) is over all edges (σ bonds) of the molecular graph. Replacing in Eqs. (6) – (9) I_i with S_i, the subset of molecular pseudo-connectivity indices, $\{^{\mathrm{S}}\psi_{\mathrm{E}}, {}^{0}\psi_{\mathrm{E}}, {}^{1}\psi_{\mathrm{E}}, {}^{\mathrm{T}}\psi_{\mathrm{E}}\}$, is obtained, *e.g.*,

$$^{\mathrm{S}}\psi_{\mathrm{E}} = \sum_i S_i \ . \tag{10}$$

While apices S and T stay for sum and total, the other apices follow the established denomination for the χ indices.[2] If in Eqs. (6) – (9) I_i is replaced first with δ_i and then with δ_i^{v} then the connectivity indices of subset $\{\chi\}$ are retrieved ($^{\mathrm{S}}\psi_{\mathrm{I}} \rightarrow D$, $^{\mathrm{S}}\psi_{\mathrm{E}} \rightarrow D^{\mathrm{v}}$, $^{0}\psi_{\mathrm{I}} \rightarrow {}^{0}\chi$, $^{0}\psi_{\mathrm{E}} \rightarrow {}^{0}\chi^{\mathrm{v}}$, and so on).

One of the results of the I_S concept states that $\Sigma_i S_i = \Sigma_i I_i$,[10] with the consequence that $^{\mathrm{S}}\psi_{\mathrm{I}} = {}^{\mathrm{S}}\psi_{\mathrm{E}}$, and in this case subset $\{\psi\}$ will be made up of seven

ψ indices only. To avoid negative S values of carbon atoms bonded to highly electronegative atoms, which could give rise to imaginary ψ_E values, the S values of those classes of compounds, whose molecules have one or more negative S_i values, have been rescaled to the S value of the carbon atom in CF_4 ($S = -5.5$). This is the lowest S value that a carbon atom can assume. Clearly, if in a molecule there are atoms with $S_i < -5.5$, then the rescaling procedure should be done taking a new reference compound as zero point, but this is not our case. The rescaling procedure invalidates the cited result of the I_S concept, with the consequence that ${}^S\psi_I \neq {}^S\psi_E$.[11–13] In this work the rescaling procedure becomes mandatory for the metal halides, MeX, and for the chlorofluorocarbons, CFC. Primary amines and alcohols, instead, whose carbons show no negative S values, do not need any rescaling procedure. About the influence of the rescaling procedure on the quality of the modeling see Ref. 12.

The statistical performance of the structural invariant, $S = X, Y, Z$, and Z', is controlled (i) by a quality factor, $Q = r / s$, and (ii) by the (Fischer) ratio $F = fr^2 / [(1-r^2)\nu]$; where r and s are the correlation coefficient and the standard deviation of the estimates, respectively, f is the number of freedom degrees $= n - (\nu+1)$, ν is the number of variables, and n is the number of data. Parameter Q, which is an »internal« statistics, is apt to compare the descriptive power of different descriptors for the same property. Fischer F ratio tells us, even if Q improves, which additional descriptor endangers the statistical quality of the combination. For every invariant S and constant parameter, U_0, of Eq. (2) the fractional utility, $u_1 = |c_1/s_1|$ of S and $u_0 = |c_0/s_0|$ of U_0, as well as the average fractional utility, $<u> = (u_{k1} + u_0)/2$, will be controlled. These last statistics will allow us to detect descriptors that give rise to unreliable coefficient values (c_k), whenever they have a high deviation (s_k). Some statistics here presented seem redundant, but, like every redundant code, they are of great help in detecting printing or other type of errors.

Mixed higher-level terms allow to short-circuit the combinatorial problem caused by a search for the best combination of indices performed on a set of sixteen β basis indices, if ${}^S\psi_I \neq {}^S\psi_E$, or of fifteen β basis indices if ${}^S\psi_I = {}^S\psi_E$. It should be noticed that in the case of sixteen indices the combinations to be searched with a full combinatorial procedure[3] are 78406, while the total combinatorial space of eight indices is made up of 255 combinations only. The number of these last combinations with 8 indices are much less than the number of combinations to be searched to obtain the optimal three-index combination out of sixteen indices, which are 560. Clearly, it cannot be excluded that a linear combination derived from a full search performed on the 78406 combinations will show a better modeling quality than any kind of term. The advantage of the trial-and-error search for the optimal term resides in the fact that here the search is an open-ended search, as the combinatorial space is here theoretically infinite, even if practically a decent term can be found, if it exists, through a cyclic procedure in a matter of minutes.[3]

RESULTS AND DISCUSSION

The Lattice Enthalpy of Metal Halides

In Tables I and II are collected the χ and ψ values together with the lattice enthalpy values of twenty metal halides. The graph of these metal halides consists of two connected points, •—•. The procedure to derive δ^v and

TABLE I

Lattice enthalpies $\Delta H_L^{\varnothing}$ (kJ mol^{-1}) at 298.15 K of 20 metal halides (MeX) and their corresponding molecular connectivity values

MeX	$\Delta H_L^{\varnothing}$	D^v	$^0\chi^v$	$^1\chi^v$
LiF	1037	8	1.37796	0.37796
NaF	926	7.11111	3.37796	1.13389
KF	821	7.05882	4.50107	1.55839
RbF	789	7.02857	6.29404	2.23607
CsF	750	7.01887	7.65807	2.75162
LiCl	852	1.77778	2.13389	1.13389
NaCl	786	0.88889	4.13389	3.40168
KCl	717	0.83660	5.25700	4.67516
RbCl	695	0.80635	7.04997	6.70820
CsCl	678	0.79665	8.41400	8.25487
LiBr	815	1.25926	2.96396	1.96396
NaBr	752	0.37037	4.96396	5.89188
KBr	689	0.31808	6.08707	7.09762
RbBr	668	0.28783	7.88004	11.6190
CsBr	654	0.27813	9.24407	14.2979
LiI	761	1.15556	3.53546	2.53546
NaI	705	0.26667	5.53546	7.60639
KI	649	0.21438	6.65857	10.4540
RbI	632	0.18413	8.45154	15.0000
CsI	620	0.17442	9.81557	18.4585

TABLE II

The I_S molecular pseudo-connectivity ψ values[a] for 20 metal halides, MeX

MeX	${}^{S}\psi_I$	${}^{0}\psi_I$	${}^{1}\psi_I$	${}^{S}\psi_E$	${}^{0}\psi_E$	${}^{1}\psi_E$
LiF	10.00	1.06066	0.25000	21.00	0.66645	0.10541
NaF	9.44	1.18689	0.29463	20.45	0.69097	0.11153
KF	9.25	1.24798	0.31623	20.25	0.70113	0.11406
RbF	9.16	1.28203	0.32827	20.16	0.70588	0.11525
CsF	9.11	1.30271	0.33558	20.11	0.70854	0.11592
LiCl	6.11	1.20037	0.34879	17.11	0.69281	0.11895
NaCl	5.55	1.32660	0.41105	16.56	0.71094	0.12446
KCl	5.36	1.38769	0.44119	16.36	0.71837	0.12670
RbCl	5.27	1.42174	0.45798	16.27	0.72183	0.12775
CsCl	5.22	1.44242	0.46819	16.22	0.72378	0.12834
LiBr	4.75	1.31013	0.42640	15.75	0.71408	0.12731
NaBr	4.19	1.43636	0.50252	15.20	0.73006	0.13269
KBr	4.00	1.49745	0.53936	15.01	0.73631	0.13477
RbBr	3.91	1.53150	0.55989	14.91	0.73969	0.13590
CsBr	3.86	1.55218	0.57236	14.86	0.74142	0.13647
LiI	4.12	1.39391	0.48564	15.12	0.72743	0.13229
NaI	3.56	1.52014	0.57234	14.57	0.74234	0.13760
KI	3.37	1.58123	0.61430	14.37	0.74847	0.13976
RbI	3.28	1.61528	0.63768	14.28	0.75136	0.14077
CsI	3.23	1.63596	0.65188	14.23	0.75299	0.14135

[a] ψ_E values have been obtained after a rescaling procedure (see method section).

the corresponding χ^v indices is, instead, explained in Ref 3. For these inorganic compounds we have: ${}^{1}\chi = \chi_t$, ${}^{1}\chi_v = \chi_t^v$, ${}^{1}\psi_I \equiv {}^{T}\psi_I$, ${}^{1}\psi_E \equiv {}^{T}\psi_E$, while ${}^{S}\psi_I$ and ${}^{S}\psi_E$ are nearly coincident with a $r({}^{S}\psi_I, {}^{S}\psi_E) = 0.99998$.

A trial-and-error procedure discovers the following very good Y pseudoterm of Eq. (11) and the less good X term of Eq. (12)

$$Y_{\Delta H} = \frac{\left(1.5 \cdot {}^{0}\psi_E + {}^{1}\psi_E\right)^8}{\left({}^{0}\psi_E\right)^5} \tag{11}$$

$Q = 0.053$, $F = 584$, $r = 0.985$, $s = 19$, $n = 20$, $\langle u \rangle = 26$, $\boldsymbol{u} = (24, 28)$

$$X_{\Delta H} = \frac{(D^{v})^{0.5} + 0.2}{D^{v} + 4.2 \cdot {}^{0}\chi^{v}} \ . \tag{12}$$

Q = 0.037, F = 281, r = 0.969, s = 24, n = 20, $<u>$ = 41, $\boldsymbol{u}$ = (17, 65)

The u_0 utility of the $X_{\Delta H}$ terms improves over the u_0 utility of the $Y_{\Delta H}$ term but at some expenses of the u_1 utility. The search for a mixed higher-order term discovers the following interesting Z term

$$Z_{\Delta H} = [0.7 \cdot (Y_{\Delta H})^{1.1} - 1.1 \cdot (X_{\Delta H})^{0.8}] \ . \tag{13}$$

Q = 0.056, F = 647, r = 0.986, s = 18, n = 20, $<u>$ = 24, $\boldsymbol{u}$ = (25, 23)

But even more interesting for its modeling power is the following trail-and-error higher-level mixed term, Z', which is the result of a linear combination of the previous mixed Z term with a pseudoconnectivity index

$$Z'_{\Delta H} = [Z - 1.4 \cdot {}^{0}\psi_{\mathrm{E}}] \ . \tag{14}$$

Q = 0.063, F = 819, r = 0.989, s = 16, n = 20, $<u>$ = 21,
$\boldsymbol{u}$ = (29, 14), $\boldsymbol{C}$ = (973.409, – 739.058)

Figure 1 has been obtained with the correlation $\boldsymbol{C}$ vector of this last Z' mixed descriptor.

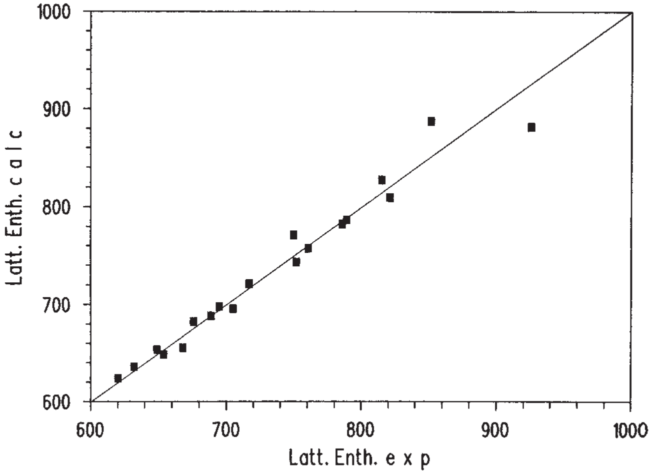

Figure 1. Plot of the calculated (calc) *versus* the experimental (exp) lattice enthalpy, Latt. Ent., for twenty metal halides.

The treatment that allows to describe inorganic salts by the aid of graph-theoretical indices can give rise to some doubts which have been treated in Ref. 3. Recently, some studies have also focused their attention on the problem ionic-covalent character of an inorganic compound.[15,16] In these studies it has been established that when discussing structure-related properties it is unnecessary and unhelpful to worry about whether a bond is ionic or covalent,[15] and that compounds like BF_3 and SiF_4 can be described in terms of ionic or polar covalent bonding and that F^- anions are strongly polarized in a manner indicating significant covalent contributions.[16] Further, it should be noticed that the *Kapustinskii equation*,[17] which is normally used to derive an approximate value for the lattice energy, uses the chemical formula of a compound and its number of ions. Now, it is the chemical formula of a compound that makes up the basic tool of a chemical graph.

The Mixed Class of Primary Amines Plus Alcohols

In Tables III–VII are collected the boiling points, T_b, of $n = 27$ alcohols and $n = 21$ amines and their corresponding β values. Up to now failed every attempt to derive a higher-level Z or Z' term of this property for the single classes of primary amines and alcohols.[12,13] For the boiling points of the mixed class made up of $n = 48$ [Am + Al], instead, it is possible to derive a higher-level mixed connectivity-pseudoconnectivity term, Z'. First let us notice that, due to collinearity,[12,13] the set of fifteen basis indices is reduced to a subset of five χ indices and a subset of seven ψ indices respectively, that is, $\{D, {}^0\chi, {}^1\chi, \chi_t, {}^1\chi^v\}$ and $\{{}^S\psi_I, {}^0\psi_I, {}^1\psi_I, {}^T\psi_I, {}^0\psi_E, {}^1\psi_E, {}^T\psi_E\}$.

The Y_T trial-and-error term of Eq. (15) offers the best overall description for the boiling points of this mixed class, while the X_T term is a less efficient descriptor for this property. Found X_T and Y_T terms are formally quite similar to the terms used to simulate the boiling points of primary amines only, X_T(Am) and Y_T(Am).[12]

$$Y_T = [({}^0\psi_I)^{0.9} - 2.6 \cdot ({}^T\psi_E)^{0.7}] \tag{15}$$

$Q = 0.084$, $F = 718$, $r = 0.969$, $s = 12$, $n = 48$, $\langle u \rangle = 44$, $\boldsymbol{u} = (27, 62)$

$$X_T = [{}^1\chi - 2.5 \cdot (\chi_t)^{1.5}] \tag{16}$$

$Q = 0.056$, $F = 325$, $r = 0.936$, $s = 17$, $n = 48$, $\langle u \rangle = 37$, $\boldsymbol{u} = (18, 55)$

While the X_T term is a strict »dead-end« term, Y_T term together with a pseudoindex offers an improved Q description for the $n = 48$ boiling points of this mixed class of [Am + Al],

TABLE III

Boiling points (T_b / K) of alcohols (R–OH) and primary amines (R–NH_2)

Alcohol – R	T_b / K	Amine – R	T_b / K
$(CH_3)_2CH-$	355.5_5	CH_3-	256.6_5
$CH_3CH_2CH_2-$	370.2_5	CH_3CH_2-	290.1_5
$CH_3CH_2C(CH_3)_2-$	375.4_5	$(CH_3)_2CH-$	307.1_5
$CH_3CH(CH_3)CH_2-$	381.2_5	$CH_3CH_2CH_2-$	322.1_5
$CH_3(CH_2)_3-$	390.7_5	$CH_3CH_2CH(CH_3)-$	336.1_5
$CH_3CH_2CH_2CH(CH_3)-$	392.0_5	$CH_3CH(CH_3)CH_2-$	341.1_5
$CH_3C(CH_3)_2CH(CH_3)-$	393.5_5	$CH_3CH_2CH_2CH_2-$	350.9_5
$CH_3(CH_2)_2C(CH_3)_2-$	396.1_5	$CH_3CH_2C(CH_3)_2-$	351.1_5
$CH_3CH(CH_3)CH(CH_3CH_2)-$	400.6_5	$(CH_3CH_2)_2CH-$	364.1_5
$CH_3CH_2CH(CH_3)CH_2-$	402.0_5	$CH_3CH_2CH_2CH(CH_3)-$	365.1_5
$CH_3CH(CH_3)CH_2CH_2-$	405.1_5	$CH_3CH(CH_3)CH_2CH_2-$	368.1_5
$CH_3CH(CH_3)CH_2CH(CH_3)-$	406.1_5	$CH_3C(CH_3)_2CH(CH_3)-$	375.1_5
$(CH_3CH_2)_2C(CH_3)-$	409.1_5	$CH_3(CH_2)_4-$	377.5_5
$CH_3CH_2C(CH_3)_2CH_2-$	409.8_5	$CH_3(CH_2)_3CH(CH_3)-$	390.6_5
$CH_3(CH_2)_4-$	411.1_5	$CH_3(CH_2)_5-$	403.1_5
$(CH_3CH(CH_3))_2CH-$	413.1_5	$CH_3CH_2CH(CH_3)CH_2CH(CH_3)-$	406.6_5
$(CH_3CH_2)_3C-$	415.1_5	$CH_3(CH_2)_4CH(CH_3)-$	415.1_5
$CH_3CH(CH_3)CH(CH_3)CH_2-$	418.1_5	$CH_3(CH_2)_6-$	430.0_5
$CH_3CH_2CH_2CH(CH_3)CH_2-$	421.1_5	$CH_3(CH_2)_7-$	449.1_5
$CH_3CH(CH_3)CH_2CH(CH_3)-$	432.9_5	$CH_3(CH_2)_8-$	465.1_5
$(CH_3CH_2)_2(CH_3)C-$	434.1_5	$CH_3(CH_2)_9-$	490.1_5
$(CH_3(CH_2)_3)(CH_3CH_2)(CH_3)C-$	436.1_5		
$CH_3(CH_2)_6-$	449.9_5		
$CH_3(CH_2)_5C(CH_3)_2-$	451.1_5		
$(CH_3CH_2CH_2)_2(CH_3CH_2)C-$	455.1_5		
$CH_3CH(CH_3)CH_2(CH_2)_4-$	461.1_5		
$CH_3(CH_2)_7-$	467.5_5		

$$\{Y_T,\ {}^S\psi_I\}:\ Q = 0.091,\ F = 423,\ r = 0.974,\ s = 11,\ n = 48,$$
$$<u> = 11,\ \boldsymbol{u} = (8.9,\ 2.9,\ 21)\ .$$

Let us try to improve *F*, and, hopefully, the other statistics also, with an algebraic sum of the two descriptors of this combination, which gives rise to a single descriptor,

TABLE IV

Molecular connectivity χ indices of primary amines ($R–NH_2$)

R	D	$^0\chi$	$^1\chi$	χ_t	$^1\chi^v$
CH_3–	2	2	1	1	0.57735
CH_3CH_2–	4	2.70711	1.41421	0.70711	1.11536
$(CH_3)_2CH$–	6	3.57735	1.73205	0.57735	1.48803
$CH_3CH_2CH_2$–	6	3.41521	1.91421	0.50000	1.61536
$CH_3CH_2CH(CH_3)$–	8	4.28446	2.27006	0.40825	2.02604
$CH_3CH(CH_3)CH_2$–	8	4.28446	2.27006	0.40825	1.97120
$CH_3CH_2CH_2CH_2$–	8	4.12132	2.41421	0.35355	2.11536
$CH_3CH_2C(CH_3)_2$–	10	5.20711	2.56066	0.35355	2.34934
$(CH_3CH_2)_2CH$–	10	4.99156	2.80806	0.28868	2.56404
$CH_3CH_2CH_2CH(CH_3)$–	10	4.99156	2.77006	0.28868	2.52604
$CH_3CH(CH_3)CH_2CH_2$–	10	4.99156	2.77006	0.28868	2.47120
$CH_3C(CH_3)_2CH(CH_3)$–	12	6.07735	2.94338	0.28868	2.69936
$CH_3(CH_2)_4$–	10	4.82843	2.91421	0.25000	2.61536
$CH_3(CH_2)_3CH(CH_3)$–	12	5.69867	3.27006	0.20412	3.02604
$CH_3(CH_2)_5$–	12	5.53553	3.41421	0.17678	3.11536
$CH_3CH_2CH(CH_3)CH_2CH(CH_3)$–	14	6.56891	3.66390	0.16667	3.41989
$CH_3(CH_2)_4CH(CH_3)$–	14	6.40578	3.77006	0.14434	3.52604
$CH_3(CH_2)_6$–	14	6.24264	3.91421	0.12500	3.61536
$CH_3(CH_2)_7$–	16	6.94975	4.41421	0.08839	4.11536
$CH_3(CH_2)_8$–	18	7.65685	4.91421	0.06250	4.61536
$CH_3(CH_2)_9$–	20	8.36396	5.41421	0.04419	5.11536

$$\{(Y_T + {}^S\psi_I)\}: Q = 0.060,\ F = 382,\ r = 0.945,\ s = 15,\ n = 48,$$
$$\langle u \rangle = 22,\ \boldsymbol{u} = (20,\ 25)\ .$$

As already seen elsewhere[13] this single »synthetic« descriptor shows improvement, neither in F nor in Q values, only the utility values improve. The statistics of the difference descriptor, $\{(Y_T - {}^S\psi_I)\}$, instead, are even worse with: $Q = 0.037$, $F = 138$, $r = 0.866$.

The following trial-and-error mixed higher-level term shows the highest modeling quality up to now, with a consistent improvement relative to Y_T term and $\{Y_T, {}^S\psi_I\}$ combination

TABLE V

I_S molecular pseudo-connectivity ψ indices of primary amines (R–NH_2)

R[a]	$^S\psi_I$	$^0\psi_I$	$^1\psi_I$	$^T\psi_I$	$^0\psi_E$	$^1\psi_E$	$^T\psi_E$
Me–	6.00	1.20711	0.35355	0.35355	1.28790	0.38490	0.38490
Et–	7.50	2.02360	0.98560	0.28868	2.33425	1.36203	0.38038
Me_2CH–	9.33	2.78132	1.65983	0.21678	3.61907	3.26968	0.39695
$EtCH_2$–	9.00	2.84010	1.65227	0.23570	3.20039	2.20962	0.32710
EtCH(Me)–	10.83	3.59782	2.33204	0.17700	4.42140	4.08218	0.33356
$MeCH(Me)CH_2$–	10.83	3.59782	2.34252	0.17700	4.16196	3.55517	0.28643
$Me(CH_2)_3$–	10.50	3.65660	2.31893	0.19245	4.04722	2.94732	0.27600
$EtC(Me)_2$–	12.75	4.33224	3.01977	0.12910	8.50171	14.7184	0.71225
$(Et)_2CH$–	12.33	4.41432	3.00424	0.14452	5.21030	4.81326	0.27522
$EtCH_2CH(Me)$–	12.33	4.41432	2.99870	0.14452	5.23851	4.73375	0.27674
$MeCH(Me)(CH_2)_2$–	12.33	4.41432	3.00919	0.14452	4.96471	4.11344	0.23505
$MeC(Me)_2CH(Me)$–	14.58	5.08996	3.71963	0.09695	6.89190	9.38091	0.32777
$CH_3(CH_2)_4$–	12.00	4.47309	3.15470	0.15713	4.87475	3.64489	0.22826
$Me(CH_2)_3CH(Me)$–	13.83	5.23081	3.66537	0.11800	6.07585	5.43429	0.23069
$CH_3(CH_2)_5$–	13.50	5.28959	3.65227	0.12830	5.71228	4.35026	0.19064
$EtCH(Me)CH_2CH(Me)$–	15.66	5.98853	4.36116	0.08861	7.02595	6.71190	0.20030
$Me(CH_2)_4CH(Me)$–	15.33	6.04731	4.33204	0.09635	6.91203	6.13575	0.19221
$CH_3(CH_2)_6$–	15.00	6.10609	4.31893	0.10476	6.53893	5.03550	0.15731
$CH_3(CH_2)_7$–	16.50	6.92258	4.98560	0.08533	7.35768	5.71122	0.12881
$CH_3(CH_2)_8$–	18.00	7.73908	5.65227	0.06984	8.18205	6.39221	0.10603
$CH_3(CH_2)_9$–	19.50	8.55558	6.31893	0.05702	8.99993	7.06167	0.08668

[a] Me = CH_3; Et = CH_3CH_2.

$$Z'_T = [0.7 \cdot Y_T + X_T + 0.7 \cdot (\chi_t)^3 - 1.1 \cdot (^1\psi_I)^{0.9}] \tag{17}$$

$$Q = 0.108,\ F = 1208,\ r = 0.981,\ s = 9.1,\ n = 48,\ <u> = 77,$$
$$\boldsymbol{u} = (35,\ 120),\ \boldsymbol{C} = (39.6549,\ 316.733)\ .$$

The excellent Figure 2 has been obtained by the aid of the ***C*** vector of this Z'_T descriptor.

Let us see how the three terms, X, Y, and Z' describe the single classes of compounds, to check which of them is a robust term, *i.e.*, which term is a good descriptor of the single classes of compounds

$$\text{Al)}\ \{Y_T\}\text{: } n = 27,\ Q = 0.069,\ F = 100,\ r = 0.895,\ s = 13$$

$$\text{Am)}\ \{Y_T\}\text{: } n = 21,\ Q = 0.156,\ F = 1570,\ r = 0.994,\ s = 6.4$$

$$\text{Al)}\ \{X_T\}\text{: } n = 27,\ Q = 0.086,\ F = 154,\ r = 0.928,\ s = 11$$

$$\text{Am)}\ \{X_T\}\text{: } n = 21,\ Q = 0.107,\ F = 728,\ r = 0.987,\ s = 9.3$$

$$\text{Al)}\ \{Z'_T\}\text{: } n = 27,\ Q = 0.099,\ F = 204,\ r = 0.944,\ s = 9.6$$

$$\text{Am)}\ \{Z'_T\}\text{: } n = 21,\ Q = 0.124,\ F = 981,\ r = 0.990,\ s = 8.0\ .$$

From these results there is no doubt that Z'_T mixed term is also a valid term for both classes of compounds separately. In fact, if we take as a sign of robustness the average value for r, $<r>$, and the average value for s, $<s>$, then Z'_T term is the most robust term. Thus, a unique graph-structural term is responsible for the simulation of a common property of two different classes of compounds, as it should be expected in an ideal case.

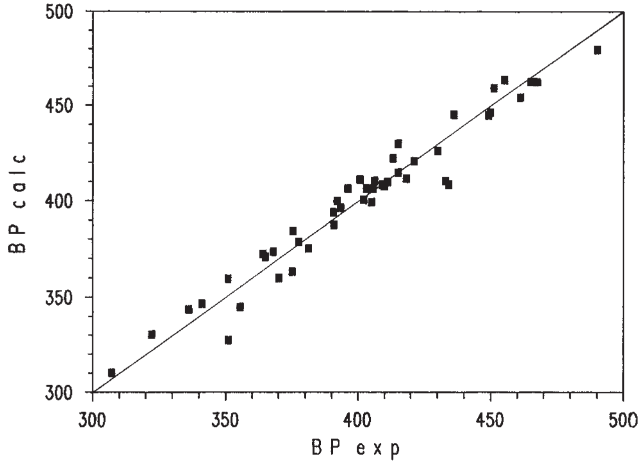

Figure 2. Plot of the calculated (calc) *versus* the experimental (exp) boiling points, BP, for 21 primary amines plus 27 primary alcohols.

Chlorofluorocarbons (CFCs) – Rates of Hydrogen Abstraction

A series of n = 26 and 11 chlorofluorocarbons $(C_xCl_yF_z)_n$, whose $\{\beta\}$ values are collected in Tables VIII and IX, is here studied for the effect of their reaction with the hydroxyl radical, in units of rates of hydrogen abstraction, log*K*, shown in Table X. In Table X are also collected eleven CFC values, which will be used for modeling the minimum anesthetic concentration (MAC) only.

A trial-and-error search procedure discovers the following molecular connectivity term with a very good modeling power,

TABLE VI

Molecular connectivity χ indices of alcohols (R–OH)

R	D	${}^0\chi$	${}^1\chi$	χ_t	${}^1\chi^v$
$(CH_3)_2CH–$	6	3.57735	1.73205	0.57735	1.41290
$CH_3CH_2CH_2–$	6	3.41421	1.91421	0.50000	1.52333
$CH_3CH_2C(CH_3)_2–$	10	5.20711	2.56066	0.35355	2.28427
$CH_3CH(CH_3)CH_2–$	8	4.28446	2.27006	0.40825	1.87918
$CH_3(CH_2)_3–$	8	4.12132	2.41421	0.35355	2.02333
$CH_3CH_2CH_2CH(CH_3)–$	10	4.99156	2.77006	0.28868	2.45090
$CH_3C(CH_3)_2CH(CH_3)–$	12	6.07735	2.94338	0.28868	2.62422
$CH_3(CH_2)_2C(CH_3)_2–$	12	5.91421	3.06066	0.25000	2.78427
$CH_3CH(CH_3)CH(CH_3CH_2)–$	12	5.86181	3.18074	0.23570	2.86159
$CH_3CH_2CH(CH_3)CH_2–$	10	4.99156	2.80806	0.28868	2.41718
$CH_3CH(CH_3)CH_2CH_2–$	10	4.99156	2.77006	0.28868	2.37918
$CH_3CH(CH_3)CH_2CH(CH_3)–$	12	5.86181	3.12590	0.23570	2.80675
$(CH_3CH_2)_2C(CH_3)–$	12	5.91421	3.12132	0.25000	2.84493
$CH_3CH_2C(CH_3)_2CH_2–$	12	5.91421	3.12132	0.25000	2.73044
$CH_3(CH_2)_4–$	10	4.82843	2.91421	0.25000	2.52333
$(CH_3CH(CH_3))_2CH–$	14	6.73205	3.55342	0.19245	3.23427
$(CH_3CH_2)_3C–$	14	6.62132	3.68198	0.17678	3.40559
$CH_3CH(CH_3)CH(CH_3)CH_2–$	12	5.86181	3.18074	0.23570	2.78986
$CH_3CH_2CH_2CH(CH_3)CH_2–$	12	5.69867	3.30806	0.20412	2.91718
$CH_3CH(CH_3)CH_2CH(CH_3)–$	12	5.86181	3.12590	0.23570	2.80675
$(CH_3CH_2)_2(CH_3)C–$	12	5.91421	3.12132	0.25000	2.84493
$(CH_3(CH_2)_3)(CH_3CH_2)(CH_3)C–$	16	7.32843	4.12132	0.12500	4.12132
$CH_3(CH_2)_6–$	14	6.24264	3.91421	0.12500	3.52333
$CH_3(CH_2)_5C(CH_3)_2–$	18	8.03553	4.56066	0.08839	4.28427
$(CH_3CH_2CH_2)_2(CH_3CH_2)C–$	18	8.03553	4.68198	0.08839	4.40559
$CH_3CH(CH_3)CH_2(CH_2)_4–$	16	7.11288	4.27006	0.10206	3.87918
$CH_3(CH_2)_7–$	16	6.94975	4.41421	0.08839	4.02333

$$X_{\log K} = (\chi_\mathrm{t})^4[D^\mathrm{v} + 3.5 \cdot ({}^0\chi^\mathrm{v})^{1.2} + 5.8 \cdot (\chi_\mathrm{t}{}^\mathrm{v})^{0.4} - 5.2 \cdot ({}^0\chi)]^{4.3} \qquad (18)$$

$Q = 3.047$, $F = 137$, $r = 0.923$, $s = 0.3$, $n = 26$, $<u> = 38$, $\boldsymbol{u} = (12, 65)$.

TABLE VII

I_S molecular pseudo-connectivity ψ indices of alcohols (R–OH)

R[a]	${}^S\psi_I$	${}^0\psi_I$	${}^1\psi_I$	${}^T\psi_I$	${}^0\psi_E$	${}^1\psi_E$	${}^T\psi_E$
$(Me)_2CH$–	11.33	2.68957	1.58027	0.17700	1.44903	0.44003	0.01629
$EtCH_2$–	11.00	2.74835	1.57735	0.19245	1.45066	0.42267	0.01646
$EtC(Me)_2$–	14.75	4.24049	2.93771	0.10541	2.21342	0.76996	0.00235
$MeCH(Me)CH_2$–	12.83	3.50607	2.26760	0.14452	1.82558	0.58135	0.00614
$Me(CH_2)_3$–	12.50	3.56484	2.24402	0.15713	1.83280	0.74179	0.00628
$EtCH_2CH(Me)$–	14.33	4.32256	2.91914	0.11800	2.21488	0.73920	0.00237
$MeC(Me)_2CH(Me)$–	16.58	4.99821	3.64007	0.07916	2.58666	0.92354	0.00087
$Me(CH_2)_2C(Me)_2$–	16.25	5.05699	3.60437	0.08607	2.59532	0.91764	0.00089
$MeCH(Me)CH(Et)$–	16.16	5.08028	3.65128	0.08861	2.59542	0.90514	0.00090
$EtCH(Me)CH_2$–	14.33	4.32256	2.93981	0.11800	2.20936	0.73046	0.00235
$MeCH(Me)CH_2CH_2$–	14.33	4.32256	2.93427	0.11800	2.20749	0.72681	0.00234
$MeCH(Me)CH_2CH(Me)$–	16.16	5.08028	3.60940	0.08861	2.59037	0.89667	0.00089
$(Et)_2C(Me)$–	16.25	5.05699	3.61290	0.08607	2.59844	0.92179	0.00090
$EtC(Me)_2CH_2$–	16.25	5.05699	3.63619	0.08607	2.58453	0.89840	0.00088
$Me(CH_2)_4$–	14.00	4.38134	2.91068	0.12830	2.21250	0.71601	0.00238
$(MeCH(Me))_2CH$–	17.99	5.83801	4.31031	0.06654	2.97342	1.06799	0.00034
$(Et)_3C$–	17.75	5.87349	4.28809	0.07027	2.98236	1.07235	0.00034
$MeCH(Me)CH(Me)CH_2$–	16.16	5.08028	3.63262	0.08861	2.58446	0.88825	0.00088
$EtCH_2CH(Me)CH_2$–	15.83	5.13906	3.60647	0.09635	2.59002	0.87655	0.00089
$MeCH(Me)CH_2CH(Me)$–	16.16	5.08028	3.60940	0.08861	2.59037	0.89667	0.00089
$(Et)_2(Me)C$–	16.25	5.05699	3.61290	0.08607	2.59844	0.92179	0.00090
$(Me(CH_2)_3)(Et)(Me)C$–	19.25	6.68998	4.94623	0.05738	3.36119	1.21539	0.00013
$CH_3(CH_2)_6$–	17.00	6.01433	4.24402	0.08553	2.97182	1.00549	0.00034
$Me(CH_2)_5C(Me)_2$–	20.75	7.50648	5.60437	0.04685	3.73582	1.35389	0.00005
$(Pr)_2(Et_2)C$–	20.75	7.50648	5.62142	0.04685	3.74605	1.36665	0.00005
$MeCH(Me)CH_2(CH_2)_4$–	18.83	6.77205	4.93427	6.06423	3.34440	1.15888	0.00013
$CH_3(CH_2)_7$-	18.50	6.83083	4.91068	0.06984	3.35119	1.14979	0.00013

[a] Me = CH_3; Et = CH_3CH_2; Pr = $CH_3CH_2CH_2$.

TABLE VIII

Molecular connectivity χ values for 32 CFCs (hydrogens are not shown)

CFC	D	$^0\chi$	$^1\chi$	χ_t	D^v	$^0\chi^v$	$^1\chi^v$	χ_t^v
CCl	2	2.00000	1.00000	1.00000	1.78	2.13228	1.13228	1.13228
CF	2	2.00000	1.00000	1.00000	8.00	1.37796	0.37796	0.37796
$\mathrm{CCl_2}$	4	2.70711	1.41421	0.70711	3.56	2.97166	1.60128	0.90655
$\mathrm{CF_2}$	4	2.70711	1.41421	0.70711	16.00	1.46304	0.53452	0.10102
CClF	4	2.70711	1.41421	0.70711	9.78	2.21735	1.06790	0.30261
CCF	4	2.70711	1.41421	0.70711	10.00	2.08507	0.97437	0.26726
CCCl	4	2.70711	1.41421	0.70711	3.78	2.83938	1.50775	0.80064
$\mathrm{CF_3}$	6	3.57735	1.73205	0.57735	24.00	1.71124	0.65465	0.03117
$\mathrm{CCl_2F}$	6	3.57735	1.73205	0.57735	11.56	3.21987	1.52566	0.29777
$\mathrm{CClF_2}$	6	3.57735	1.73205	0.57735	17.78	2.46556	1.09016	0.09339
$\mathrm{CCl_3}$	6	3.57735	1.73205	0.57735	5.34	3.97418	1.96116	0.83810
$\mathrm{CCCl_2}$	6	3.57735	1.73205	0.57735	5.56	3.84190	1.88479	0.74019
$\mathrm{CCF_2}$	6	3.57735	1.73205	0.57735	18.00	2.33328	1.01379	0.08248
CFCF	6	3.41421	1.91421	0.50000	18.00	2.17014	1.03452	0.07143
$\mathrm{CClCCl_2}$	8	4.28446	2.27006	0.40825	7.34	4.68129	2.51633	0.59263
$\mathrm{CFCF_2}$	8	4.28446	2.27006	0.40825	26.00	2.41835	1.11195	0.02204
$\mathrm{CClCF_2Cl}$	10	5.20711	2.56066	0.35355	21.56	4.22759	2.09830	0.06475
$\mathrm{CClCF_3}$	10	5.20711	2.56066	0.35355	27.78	3.47328	1.72114	0.02162
$\mathrm{CBrCF_3}$	10	5.20711	2.56066	0.35355	27.26	4.30216	2.30725	0.03744
$\mathrm{CICF_3}$	10	5.20711	2.56066	0.35355	27.16	4.84100	2.68827	0.04773
$\mathrm{CFCF_3}$	10	5.20711	2.56066	0.35355	34.00	2.71896	1.18776	0.00722
$\mathrm{CF_2CF_2}$	10	5.15470	2.64273	0.33333	34.00	2.66656	1.20620	0.00680
$\mathrm{CCl_2CF_3}$	12	6.07735	2.94338	0.28868	29.56	4.47580	2.16306	0.01983
$\mathrm{CClFCF_3}$	12	6.07735	2.94338	0.28868	35.78	3.72149	1.72756	0.00667
$\mathrm{CF_2CF_3}$	12	6.07735	2.94338	0.28868	42.00	2.96717	1.29206	0.00223
$\mathrm{CClBrCF_3}$	12	6.07735	2.94338	0.28868	29.04	5.30468	2.64162	0.03461
$\mathrm{CFBrCF_3}$	12	6.07735	2.94338	0.28868	35.26	4.55037	2.20612	0.01155
$\mathrm{CFICF_3}$	12	6.07735	2.94338	0.28868	35.16	5.08921	2.51722	0.01473
$\mathrm{CBr_2CF_3}$	12	6.07735	2.94338	0.28868	28.52	6.13357	3.12018	0.05995
$\mathrm{CCF_3}$	8	4.50000	2.00000	0.50000	26.00	2.63389	1.06695	0.02700
$\mathrm{CCF_2Cl}$	8	4.50000	2.00000	0.50000	19.78	3.38821	1.44410	0.08088
$\mathrm{CCCl_3}$	8	4.50000	2.00000	0.50000	7.34	4.89683	2.19842	0.72582

TABLE IX

I_S-ψ values for 32 CFCs[a]

CFC	${}^{S}\psi_{I}$	${}^{0}\psi_{I}$	${}^{1}\psi_{I}$	${}^{T}\psi_{I}$	${}^{S}\psi_{E}$	${}^{0}\psi_{E}$	${}^{1}\psi_{E}$	${}^{T}\psi_{E}$
CH_3Cl	6.11	1.20037	0.34879	0.34879	17.11	0.69281	0.11895	0.11895
CH_2Cl_2	9.72	1.80302	0.80550	0.19866	26.03	1.04637	0.26291	0.04122
$CHCl_3$	13.66	2.34690	1.28314	0.10407	35.68	1.39314	0.42869	0.01386
CH_3F	10.00	1.06066	0.25000	0.25000	21.00	0.66645	0.10541	0.10541
CH_2F_2	17.50	1.52360	0.57735	0.10206	34.01	1.03057	0.26552	0.03413
CHF_3	25.33	1.92777	0.91971	0.03832	47.34	1.50947	0.56938	0.01251
CH_2ClF	23.61	1.66331	0.69142	0.14239	30.11	1.03275	0.26350	0.03722
$CHCl_2F$	17.55	2.20719	1.16200	0.07559	39.35	1.40229	0.45671	0.01310
$CHClF_2$	21.44	2.06748	1.04085	0.05346	43.45	1.42839	0.49624	0.01245
CH_3CH_2Cl	7.61	2.01687	0.98010	0.28479	24.11	1.07743	0.27124	0.04552
CH_3CH_2F	11.50	1.87716	0.86603	0.20412	28.00	1.06714	0.27526	0.04163
CH_3CHCl_2	11.55	2.56074	1.46857	0.14918	33.56	1.42391	0.43028	0.01539
CH_3CHF_2	19.33	2.28132	1.22628	0.07664	41.32	1.44814	0.49327	0.01376
CH_2FCH_2F	19.00	2.34010	1.24402	0.08333	41.00	1.42984	0.44802	0.01357
$CH_2ClCHCl_2$	15.16	3.16340	1.96617	0.08497	42.67	1.77999	0.58378	0.00534
CH_2FCHF_2	26.83	2.74427	1.60981	0.03129	54.25	1.87126	0.74484	0.00489
CH_2ClCF_2Cl	23.97	3.86045	2.20669	0.04274	59.48	2.25853	0.99310	0.00191
CH_2ClCF_3	30.86	3.26485	2.08173	0.01592	63.85	2.42448	1.23127	0.00208
CH_2FCF_3	34.75	3.12514	1.96766	0.01141	67.74	2.62825	1.53805	0.00235
CHF_2CHF_2	34.66	3.14843	1.97816	0.01175	67.68	2.41002	1.20063	0.00196
$CHCl_2CF_3$	34.80	3.80872	2.57968	0.00834	73.28	2.91742	1.63042	0.00083
$CHClFCF_3$	38.69	3.66901	2.45853	0.00598	77.79	3.26865	2.19995	0.00104
CHF_2CF_3	42.58	3.52930	2.33739	0.00428	81.10	4.55532	4.38982	0.00210
CH_3CCl_3	15.58	3.08133	1.95602	0.07590	43.08	1.77734	0.61915	0.00525
CH_3CF_3	27.25	2.66219	1.58114	0.02795	54.74	1.98926	0.95756	0.00542
CH_3CF_2Cl	23.36	2.80190	1.70610	0.03900	50.78	1.85862	0.77775	0.00509
CH_2BrCF_3	29.50	3.37461	2.17315	0.01941	62.14	2.52461	1.32327	0.00245
CH_2ICF_3	28.87	3.45839	2.39975	0.02217	61.88	2.39562	1.14897	0.00217
$CHFBrCF_3$	37.33	3.77877	2.55371	0.00731	75.84	3.13653	1.95688	0.00100
$CHFICF_3$	36.70	3.86255	2.62635	0.00832	75.22	3.09680	1.87379	0.00100
$CHClBrCF_3$	33.44	3.91848	2.67485	0.01020	71.95	2.86487	1.52435	0.00083
$CHBr_2CF_3$	32.08	4.02824	2.77002	0.01246	70.57	2.82553	1.43874	0.00083

[a] ψ_E values have been obtained after a rescaling procedure (see method section).

TABLE X

Rates of hydrogen abstraction in n = 26 CFCs in log$K$ units, and logarithm of the Minimum Anesthetic Concentrations, log MAC, of 11 trifluoromethylethanes, tFMeE

CFC	log$K$	CFC	log$K$	tFMeE	log MAC
CH_3Cl	7.36	CH_2FCH_2F	7.83	CH_3CF_3	1.60
CH_2Cl_2	8.00	$CH_2ClCHCl_2$	8.28	CH_2ClCF_3	0.90
$CHCl_3$	7.80	CH_2FCHF_2	7.47	CH_2BrCF_3	0.45
CH_3F	6.95	CH_2ClCF_2Cl	7.20	CH_2ICF_3	0.10
CH_2F_2	6.81	CH_2ClCF_3	6.95	CHF_2CF_3	1.70
CHF_3	5.10	CH_2FCF_3	6.70	$CHFClCF_3$	1.18
CH_2ClF	7.46	CHF_2CHF_2	6.50	$CHFBrCF_3$	0.70
$CHCl_2F$	7.30	$CHCl_2CF_3$	7.40	$CHFICF_3$	0.30
$CHClF_2$	6.45	$CHClFCF_3$	6.87	$CHCl_2CF_3$	0.43
CH_3CH_2Cl	8.37	CHF_2CF_3	6.48	$CHClBrCF_3$	–0.10
CH_3CH_2F	8.14	CH_3CCl_3	6.80	$CHBr_2CF_3$	–0.40
CH_3CHCl_2	8.20	CH_3CF_3	5.95		
CH_3CHF_2	7.48	CH_3CF_2Cl	6.60		

The best trial-and-error Y term is, instead, a poorer descriptor

$$Y_{\log K} = [({}^{S}\psi_E - 1.1\,{}^{S}\psi_I - {}^{T}\psi_I) / {}^{0}\psi]^{6.5} \tag{19}$$

$$Q = 2.07,\ F = 63,\ r = 0.852,\ s = 0.4,\ n = 26,\ <u> = 24,\ \boldsymbol{u} = (8.0,\ 41)\ .$$

Term $X_{\log K}$ together with ${}^0\chi$ index gives rise to combination $\{X_{\log K}, {}^0\chi\}$, with an improved $Q : Q = 3.089$, $F = 71$, $r = 0.927$, $s = 0.3$. With these two terms, $X_{\log K}$ and $Y_{\log K}$, it is, now, possible to construct by the aid of a trial-and-error procedure the following mixed connectivity-pseudoconnectivity term, $Z'_{\log K} = f(X, Y, \beta)$,

$$Z'_{\log K} = [(X_{\log K})^2 + 9 \cdot Y_{\log K} + 5 \cdot 10^6 \cdot {}^0\chi^{v}]^{0.5} \tag{20}$$

$$Q = 3.64,\ F = 196,\ r = 0.944,\ s = 0.3,\ n = 26,\ <u> = 34,$$
$$\boldsymbol{u} = (14,\ 55),\ \boldsymbol{C} = (-0.00043,\ 9.46964)\ .$$

The mixed term $Z_{CFC} = f(X, Y)$ of Eq. (21) shows, instead, a lower but always quite interesting quality. This Z_{CFC} term is here included as its importance as well its name will become evident with the modeling of the next activity, *i.e.*, log MAC of CFCs.

$$Z_{\mathrm{CFC}} = (X_{\log K})^{0.5} \cdot (Y_{\log K})^{0.3} \tag{21}$$

$$Q = 3.41,\ F = 172,\ r = 0.937,\ s = 0.3,\ n = 26,\ <u> = 35,$$
$$\boldsymbol{u} = (13,\ 57),\ \boldsymbol{C} = (-0.00047,\ 9.12978)\ .$$

Found $Z'_{\log K}$ and Z_{CFC} terms show how far from ψ- and χ-based descriptors it is possible to ascend to derive dominant terms, and also how the search for an improved descriptor can never be considered concluded. The last Z_{CFC} term combined with ${}^0\chi^{\mathrm{v}}$ index shows improved Q, and r statistics: $\{Z_{\mathrm{CFC}},\ {}^0\chi^{\mathrm{v}}\}$: $Q = 3.50$, $F = 91$, $r = 0.942$, s = 0.3, $n = 26$, $<u> = 16$, $\boldsymbol{u} = (13, 1.4, 34)$. Here, the less satisfactory F value as well as the low utility value of the ${}^0\chi^{\mathrm{v}}$ basis index arises some doubts about the reliability of this combination.

Chlorofluorocarbons (CFCs) – Minimum Anesthetic Concentrations

The minimum anesthetic concentrations of eleven trifluoromethylethanes, tFMeE, expressed as log MAC is reported in Table X. While neither $X_{\log K}$ nor $Y_{\log K}$ can be used to simulate log MAC, as they show a bad quality, an »*ad hoc*« trial-and-error procedure discovers the following two terms of Eqs. (22) and (23), $X_{\log \mathrm{MAC}}$ and $Y_{\log \mathrm{MAC}}$, where, $Y_{\log \mathrm{MAC},}$ is the best homogeneous descriptor with very good statistical values

$$X_{\log \mathrm{MAC}} = ({}^1\chi^{\mathrm{v}} + 1.2 \cdot \chi_{\mathrm{t}})^{0.5} \tag{22}$$

$$Q = 7.84,\ F = 278,\ r = 0.981,\ s = 0.1\ (0.13),\ n = 11,\ <u> = 18,\ \boldsymbol{u} = (17,\ 18)$$

$$Y_{\log \mathrm{MAC}} = ({}^1\psi_{\mathrm{I}} + {}^{\mathrm{T}}\psi_{\mathrm{I}})\ /\ {}^{\mathrm{S}}\psi_{\mathrm{E}} \tag{23}$$

$$Q = 15.3,\ F = 1065,\ r = 0.996,\ s = 0.1\ (0.06),\ n = 11,\ <u> = 34,\ \boldsymbol{u} = (33,\ 35).$$

Both $X_{\log \mathrm{MAC}}$ and $Y_{\log \mathrm{MAC}}$ are bad descriptors of logK of CFCs. While no better combinations, $\{Y_{\log \mathrm{MAC}},\ \psi\}$, could be found, the overall best modeling quality can be ascribed to the higher-level mixed connectivity-pseudoconnectivity term of Eq. (24),

$$Z_{\log \mathrm{MAC}} = [(X_{\log \mathrm{MAC}})^{0.01} + (Y_{\log \mathrm{MAC}})^{0.9}]^5 \tag{24}$$

$$Q = 17.7,\ F = 1414,\ r = 0.997,\ s = 0.06,\ n = 11,\ <u> = 38,$$
$$\boldsymbol{u} = (38,\ 39),\ \boldsymbol{C} = (-20.5991,\ 27.2136)\ .$$

Up to now no higher-level mixed term, $Z'_{\log \mathrm{MAC}} = f(X,\ Y,\ \beta)$, could be found.

Let us, now, retrieve the Z_{CFC} term (see Eq. (21)) found in the preceding paragraph. If we try to simulate log MAC by the aid of this Z_{CFC} term we notice that this term is also a good descriptor for log MAC,

$$Z_{\mathrm{CFC}} = (X_{\log K})^{0.5} \cdot (Y_{\log K})^{0.3} \tag{25}$$

$$Q = 11.4,\ F = 586,\ r = 0.982,\ s = 0.1,\ n = 11,\ \langle u \rangle = 25,$$
$$\boldsymbol{u} = (24, 26),\ \boldsymbol{C} = (-16.5121, 8.14628)\ .$$

In fact, Figure 3, where the thirty-seven points of the calculated *versus* the experimental [log$K$ + log MAC] have been plotted, is very impressive. Here the calculated points of log$K$ and log MAC have been obtained with Eqs. (26) and (27).

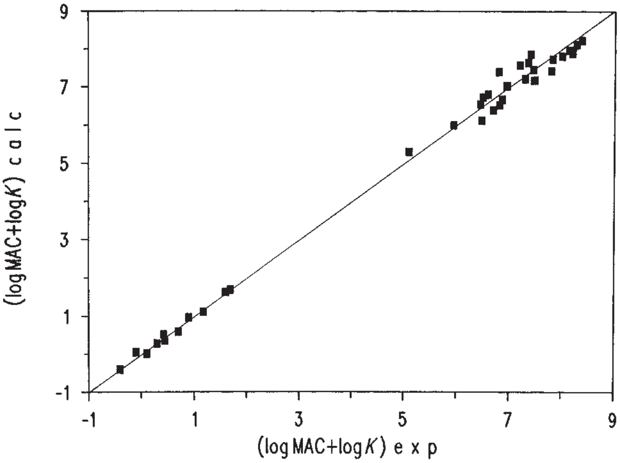

Figure 3. Plot of the calculated (calc) *versus* the experimental (exp) log MAC+log$K$ for thirty-seven CFCs.

$$\log K = -0.0005 \cdot Z_{\mathrm{CFC}} + 9.13 \tag{26}$$

$$\log \mathrm{MAC} = -16.51 \cdot Z_{\mathrm{CFC}} + 8.15 \tag{27}$$

That is, a unique mixed higher-level graph-structural term is required to describe both properties of CFCs, log$K$ and log MAC. Clearly, each time with its own $\boldsymbol{C}$ vector (see $\boldsymbol{C}$ vectors of Eqs. (21) and (25)), as it should be expected for two different properties of the same class of compounds. The fact that a single S descriptor can be used for two different properties can be seen as a kind of external validation text for the found descriptor.

In Table XI are summarized the correlation found for the different properties of the different classes of compounds, with their statistics. Q statistics has here been excluded as it is an internal statistics of each property (see method section). This Table gives an overall impression of the validity of the achieved descriptions allowing a rapid comparison with the best correlations of lower quality (some of them are multiple linear regressions) to be found in the literature, concerning, especially, the lattice enthalpy of metal halides,[11,18] and the $\log K$ of CFCs.[12,19]

TABLE XI

Summary of the best correlations for the lattice enthalpy, $\Delta H_{\mathrm{L}}^{\varnothing}$, of metal halides (MeX), for the boiling points, T_{b}, of primary amines & alcohols [Am + Al], for the $\log K$ of chlorofluorocarbons (CFC), for the $\log \mathrm{MAC}$ of CFC, and for both activities [$\log K$ + $\log \mathrm{MAC}$] of the entire class of CFC simulated by a single descriptor, Z_{CFC}

Class	Correlations	F	r	s	n	$<u>$
MeX	$\Delta H_{\mathrm{L}}^{\varnothing} = 973.409 \cdot Z'_{\Delta H} - 739.1058$	819	0.989	16	20	21
[Am + Al]	$T_{\mathrm{b}} = 39.6549 \cdot Z'_{\mathrm{T}} + 316.733$	1208	0.981	9.1	48	77
CFC	$\log K = -0.00043 \cdot Z'_{\log K} + 9.46964$	196	0.944	0.3	26	34
CFC	$\log \mathrm{MAC} = -20.599 \cdot Z_{\log \mathrm{MAC}} + 27.214$	1414	0.997	0.06	11	38
CFC	$\log K = -0.00047 \cdot Z_{\mathrm{CFC}} + 9.12978$	172	0.937	0.3	26	35
CFC	$\log \mathrm{MAC} = -16.5121 \cdot Z_{\mathrm{CFC}} + 8.14628$	586	0.982	0.1	11	25

CONCLUSIONS

Recently it has been said »*or is it that there are better descriptors but we have simply been unsuccessful in finding them?*«.[20] Found higher-levels terms made up of homogeneous connectivity and pseudoconnectivity terms, $Z = f(X, Y)$ and $Z' = f(X, Y, \beta)$, seem to give an affirmative answer to the need of finding new powerful descriptors using old tools, that is, those tools based on graph-theoretical molecular connectivity (MC) concepts, which came into being in 1975.[1,2] Throughout this paper it has been shown how the Configuration Interaction of adjustable Graph-Type Basis Indices (CIaGTBI) is able to derive optimal descriptors with a set of basis indices defined within the frame of MC theory. Stimulating is also the fact that the trial-and-error search procedure used to derive a good term is formally an open-ended procedure but practically a rapidly converging search procedure.

Often the claim has arisen that connectivity indices, first, and, later, connectivity terms, do nothing else but mimic random numbers. Some years ago Kier and Hall[21] tackled this problem and showed that random numbers were unable to model in a satisfactory way any property or activity whatsoever. It should be noticed that if some »good« random numbers were able to model properties or activities of classes of compounds they could not be considered random anymore. Further, the detection of these rare objects, »good« random numbers, if they existed, would be of primary importance in any QSAR/QSPR study.

The other claim that terms have no physical meaning, could actually be extended to the basis indices, as we will shortly see. Kline in a very enjoyable book[22] asserts that most of physics has no physical meaning, but that physics turns around mathematical objects that are extremely useful in physics. We would not go that far, but it is not at all impolite to question ourselves about the physical meaning of a Gaussian-Type Orbital (GTO)? And what about the physical meaning of a linear combination of GTOs with adjustable parameters followed by CI (Configuration Interaction) calculations? In addition, could not these heavy and complex calculations be mimicked by random numbers?

But things are here even subtler. Molecular connectivity and pseudoconnectivity indices and terms are derived from graphs and pseudographs, which are mathematical (topological) objects, and insofar they have no physical meaning, they are just mathematical algorithms, which are used to solve physical, chemical or biological problems. Let us here read the stimulating words by Ugi on the argument:[23] »*the goal of mathematical chemistry is the mathematization of chemistry without the intermediary of physics and the direct solution of chemical problems by qualitative mathematical methods*«. Chemistry, like physics and also biology, is looking for extremely useful mathematical objects.

With the introduction of graph-pseudograph theoretical higher-level invariants we are not only getting nearer and nearer to one of the main goals of material science exemplified in Scheme I and II in the introduction section, but also to the overall goal of science, that is, the reduction of databases into algorithms. This last goal was phrased by E. Mach at the beginning of 20th century with very impressive words: »*because our memory is limited, data must be reduced. If, for example, to every time of falling, we knew the corresponding space fallen through, we could be satisfied with that. Only, what a gigantic memory would be needed to contain the table of the correspondences of* s *and* t*. Instead of this we remember the formula* $s = gt^2/2$*, that is to say, the rule of derivation by means of which we find from a given* t*, the corresponding* s*, and this replaces the table just mentioned in a very complete, convenient and compendious manner*«.[24]

Acknowledgement. – It is my duty and pleasure to underline the help, support, and advise I received from Professor Milan Randić during these last years.

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 (and references therein).
3. L. Pogliani, *Chem. Rev.* **100** (2000) 3827–3858.
4. G. B. Olson, *Science* **288** (2000) 993–998.
5. N. Trinajstić, *Chemical Graph Theory*, 2nd Ed., CRC Press, Boca Raton, 1992.
6. J. Devillers and A. T. Balaban, *Topological Indices and Related Descriptors in QSAR and QSPR*, Gordon and Breach, UK, 1999.
7. M. Reinhard and A. Drefahl, *Handbook for Estimating Physicochemical Properties of Organic Compounds*, Wiley, New York, 1999.
8. M. V. Diudea, *QSPR/QSAR Studies by Molecular Descriptors*, Nova Science Pubs., Inc., New York, 2001.
9. M. Randić, *Topological Indices*, in: P. V. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer III, and P. R. Schreiner, (Eds.) *The Encyclopedia of Computational Chemistry*, Wiley, Chichester, 1998, pp. 3018–3032.
10. L. B. Kier and L. H. Hall, *Molecular Structure Description. The Electrotopological State*, Academic Press, New York, 1999.
11. L. Pogliani, *J. Phys. Chem.* A **104** (2000) 9029–9045.
12. L. Pogliani, *J. Chem. Inf. Comput. Sci.* **41** (2001) 836–841.
13. L. Pogliani, *Mixed Higher-Order Connectivity-Pseudoconnectivity Terms,* in: *Volume on Harry Wiener International Memorial Conference on the Role of Topology in Chemistry*, 2001.
14. M. N. Berberan-Santos and L. Pogliani, *J. Math. Chem.* **26** (1999) 255–261.
15. I. D. Brown, *J. Chem. Ed.* **77** (2000) 1070–1075.
16. A. Haaland, T. U. Helgaker, K. Ruud, and D. J. Shorokhov, *J. Chem. Ed.* **77** (2000) 1076–1080.
17. S. E. Dann, *Reactions and Characterization of Solids*, The Royal Society of Chemistry, Cambridge, UK, 2000, pp. 34–35.
18. L. Pogliani, *Croat. Chem. Acta* **70** (1997) 803–817, p. 811.
19. L. B. Kier and L. H. Hall, *Molecular Structure Description. The Electrotopological State*, Academic Press, New York, 1999, pp. 170–172.
20. M. Randić and S. C. Basak, *Variable Molecular Descriptors*, in: D. K. Sinha, S. C. Basak, R. K. Mohanty, and I. N. Basumallik (Eds.), *Some Aspects of Mathematical Chemistry*, in press.
21. L. B. Kier and L. H. Hall, *Molecular Connectivity in Structure-Activity Analysis*, Wiley, New York, 1986, Chs. 5 and 7.
22. M. Kline, *Mathematics and the Physical World*, Dover, New York, 1981.
23. I. Ugi, *Angew. Chem., Int. Ed. Eng.* **32** (1993) 201–227.
24. M. Jo Nye, *A History of Modern Physics 1800–1954*. Vol IV: *The Question of the Atom*, Tomash Pub., Los Angeles, San Francisco, 2nd Ed, 1986, p. 176.

SAŽETAK

Deskriptori višeg stupnja u molekulskoj povezanosti

Lionello Pogliani

Optimalno modeliranje niza observabli (entalpijâ rešetke metalnih halogenida, vrelišta mješovite klase alkohola i amina te dvije aktivnosti kloroflorougljika, brzinâ odcjepljivanja vodika te minimalne anestetičke koncentracije) postignuto je zahvaljujući uvođenju tri nove vrste članova višega reda u molekulsku povezanost. Prva vrsta člana, tj. mješoviti član molekulske povezanosti i pseudopovezanosti, $Z = f(X, Y)$, funkcija je jedino člana molekulske povezanosti, X, i člana molekulske pseudopovezanosti, Y. Članovi X i Y funkcije su osnovnog indeksa β koji može biti bilo indeks molekulske povezanosti, χ, ili molekulske pseudopovezanosti, Ψ. Član Z može se dobiti različitim kombinacijama članova X i Y. Drugi i treći tip mješovitih članova višega reda, $Z' = f(Z, \beta)$ i $Z' = f(X, Y, \beta)$, mogu se dobiti metodom pokušaja i pogreške u kojoj Z ili X i Y su konstante tijekom cijeloga traženja. Tri tipa mješovitih članova višega reda modeliraju svojstva na sljedeći način: dva različita svojstva klorofluorougljika uspješno su modelirana s pomoću $Z = f(X, Y)$, međutim entalpije rešetke metalnih halida modelirane su s pomoću $Z' = f(Z, \beta)$, a vrelišta mješovite klase alkohola i amina modelirana su s pomoću $Z' = f(X, Y, \beta)$.