How Far Are Molecular Connectivity Descriptors from I_S Molecular Pseudoconnectivity Descriptors?

Lionello Pogliani[†]

Dipartimento di Chimica, Università della Calabria, 87030 Rende (CS), Italy

A comparison of the characteristics of the molecular connectivity and intrinsic-state pseudoconnectivity indices in modeling different activities and properties of different classes of compounds is performed. Two different activities of chlorofluorocarbons, an activity of 2-Br-2-phenetylamines, an activity of benzimidazoles, and the boiling points of primary amines and primary alcohols are modeled. The simultaneous modeling of these last two classes of compounds is also accomplished. The comparison encompasses also the molar mass as descriptor. Further, the influence of the rescaling procedure on the modeling for the case of benzimidazoles is analyzed as well as the relationship between degeneracy of connectivity values and quality of the modeling. Molecular pseudoconnectivity terms seem to compete advantageously with molecular connectivity terms in the descriptive power of the activity and property of many classes of compounds. In many cases new mixed higher-order connectivity-pseudoconnectivity terms have been detected that consistently improve the quality of the description.

INTRODUCTION

In 1990 Kier and Hall developed the electrotopological-state concept to derive significant indices for atoms in molecules,[1] which they further refined and used, together with other authors throughout the entire following decade.[2−5] In a recent book[5] they collected all the experience and works published using this concept. This concept can practically be seen as the ultimate product of a "tour de force" started with the introduction of the concept of molecular connectivity[6] (see references therein) in connection with a mathematical algorithm proposed by Randić 25 years ago.[7]

Recently,[8] the electrotopological-state concept has been used to develop new molecular descriptors, the intrinsic-state molecular pseudoconnectivity indices, I_S-ψ, or more concisely the ψ indices. Found indices, which show a much lower degeneracy than the corresponding set of χ indices, have been used in a series of QSAR and QSPR studies of different classes of compounds.[8] Their modeling quality competes, in many cases, with the modeling power of the well-known molecular connectivity indices. Further, high predictive dominant molecular pseudoconnectivity terms have be found by the aid of a trial-and-error procedure.

Throughout the cited study[8] about ψ indices many features of these indices, due to the novelty of the subject, went unnoticed. The present study (i) underlines the unnoticed characteristics of ψ indices, (ii) enlarges the spectrum of properties and activities that can be modeled with these indices, (iii) undertakes a comparative study of the modeling capabilities of both type of indices, and (iv) discovers mixed higher-order connectivity-pseudoconnectivity terms. Two different activities of chlorofluorocarbons, an activity of 2-Br-2-phenetylamines and of benzimidazoles, and finally the boiling points of primary amines and alcohols will be modeled throughout the present study. A comparison with the molar mass as descriptors, that shows, for some classes of compounds, a high degeneracy, will be stressed throughout this study. As the molar mass degeneracy is in some cases mimicked by a corresponding degeneracy in a subset of connectivity χ indices, the relationship between degeneracy and modeling together with the influence of the rescaling procedure of ψ indices on modeling will also be inspected.

METHOD

As already done in preceding QSPR studies[8,9] a medium-sized set of eight molecular connectivity indices (eq 1) and eight molecular pseudoconnectivity indices (eq 2) will be used throughout the present study.

$$\{\chi\} = \{D, {}^0\chi, {}^1\chi, \chi_t, D^v, {}^0\chi^v, {}^1\chi^v, \chi_t^v\} \qquad (1)$$

$$\{\psi\} = \{{}^S\psi_I, {}^0\psi_I, {}^1\psi_I, {}^T\psi_I, {}^S\psi_E, {}^0\psi_E, {}^1\psi_E, {}^T\psi_E\} \qquad (2)$$

While χ indices are directly based on the δ and δ^v connectivity numbers of a graph and a pseudograph, respectively,[6,7,9] ψ indices are directly based on the I−state (ψ_I subset) and S−state (ψ_E subset) indices,[8] which have been developed by Kier and Hall within the frame of the intrinsic-state, I_S, concept. These two atom-level indices are defined in eqs 2 and 3. From these equations we notice that index S incorporates information about the influence of the remainder of the molecular environment.[1−5] Indices I and S can easily be obtained for any molecule with the software package provided in ref 5.

$$I_i = [(2/N)^2 \delta_i^{\,v} + 1]/\delta_i \qquad (3)$$

$$S_i = I_i + \Sigma_j \Delta I_{ij} \qquad (4)$$

[†] Corresponding author phone: +39-984-492105; e-mail: lionp@unical.it.

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 $d_{ij} + 1$. From the definition of ΔI_{ij} it is evident that S can also assume negative values. Equations 3 and 4 show clearly that I_i and S_i indices encode simultaneously graph (δ) and pseudograph characteristics (δ^v). Now, as a graph is a special case of a pseudograph[9] the name pseudo-indices for the intrinsic-state ψ indices seems to be well earned. To underline the similarity between indices of sets 1 and 2 formally similar indices are here shown side-by-side

$$D = \Sigma_i\delta_i; \quad {}^S\psi_I = \Sigma_i I_i \qquad (5)$$

$$ {}^0\chi = \Sigma_i(\delta_i)^{-0.5}; \quad {}^0\psi_I = \Sigma_i(I_i)^{-0.5} \qquad (6)$$

$$ {}^1\chi = \Sigma(\delta_i\delta_j)^{-0.5}; \quad {}^1\psi_I = \Sigma(I_iI_j)^{-0.5} \qquad (7)$$

$$\chi_t = (\delta_1 \cdot d_2 \cdot d_3 \cdot \dots \cdot d_N)^{-0.5}; \quad {}^T\psi_I = (I_1 \cdot I_2 \cdot I_3 \cdot \dots \cdot I_N)^{-0.5} \qquad (8)$$

Sums in eqs 5 and 6 are taken over all N atoms of a molecule, and the sum in eq 7 is over all edges (σ bonds) of the molecular graph. Replacing in eqs 5–8 δ_i with δ^v_i and I_i with S_i, the subset of four valence χ^v indices and four ψ_E indices is obtained. Apexes S and T stand for sum and total. The other apexes, instead, follow the established denomination for χ indices.[6]

To avoid negative S_i values of carbon atoms bonded to highly electronegative atoms, which could give rise to imaginary ψ_E values, those S_i values of classes of compounds with negative S_i have been rescaled to the *S* value of the carbon atom in CF_4 ($S = -5.5$). This is the lowest *S* values a carbon atom can assume. Other non-carbon atoms with more negative *S* values can be found, but the rescaling was also done with the intent to avoid (after rescaling) too large *S* values, which would render a practically useless index ${}^T\psi_I = (S_1 \cdot S_2 \cdot \dots \cdot S_N)^{-0.5}$. Inevitably, the rescaling procedure invalidates a theorem of the I_S concept ($\Sigma_i I_i = \Sigma_i S_i$) with the consequence that ${}^S\psi_I \neq {}^S\psi_E$.[8] While the rescaling procedure is mandatory for chlorofluorocarbons, CFC, and 2-bromo-2-phenethylamines, primary amines and alcohols, with no negative S_i values, do not need any rescaling procedure. Benzimidazoles, instead, whose S_i values are always positive, have been chosen for a comparative study between a rescaled and a nonrescaled case, with the intent to check the weight of the rescaling procedure.

The statistical performance of a modeling will be controlled by (i) the quality factor, $Q = r/s$, (ii) the variance *F* (Fischer ratio), (iii) the fractional utility, $u_k = |c_k/s_k|$, and (iv) the average utility $\langle u \rangle = \Sigma u_k/(\nu+1)$.[9] Here *r* is the correlation coefficient and *s* is the standard deviation of the estimate. Parameter Q is apt to compare the descriptive power of different descriptors for the same property. Ratio F tells us, even if Q improves, which additional descriptor endangers the statistical quality of the combination. The fractional utility as well as the average fractional utility will allow us to detect descriptors, which can give rise to unreliable regression coefficients (c_k) due to their large deviation (s_k). Regression coefficients, c_i, and utility values, u_i, will be collected into

C and **u** vectors, respectively. To avoid the interrelation problem between different indices and to improve the description, the construction of connectivity, $X = f(\chi)$, or pseudoconnectivity, $Y = f(\psi)$, terms, by the aid of a trial-and-error procedure, will always be attempted.[8,9]

RESULTS AND DISCUSSION

Chlorofluorocarbons (CFCs). Rates of Hydrogen Abstraction. A series of $n = 26$ chlorofluorocarbons $(C_xCl_yF_z)_n$, whose $\{\chi\}$ and $\{\psi\}$ values are collected in Tables 1 and 2, respectively, was studied for the effect of their reaction with the hydroxyl radical, in units of rates of hydrogen abstraction, log *K*, shown in Table 3. In Table 1 are also reported six CFCs that will be used for modeling the minimum anesthetic concentration (MAC) only. From Tables 1 and 2 it is possible to notice that the *M* degeneracy is rather low. Seven subclasses of compounds show degenerate nonvalence χ indices, $\{D, {}^0\chi, {}^1\chi, \chi_t\}$. Pseudoindices, instead, show no degeneracy at the level of their ψ_I and ψ_E subsets.

Let us start noticing that the molar mass is a very bad descriptor of log *K* with $Q = 0.042$, $F = 0.03$, $r = 0.033$, and $s = 0.78$. Kier and Hall[10] revealed that a combination of a χ index, a κ shape index, and a S index achieved a remarkable description for this property (two decimal figures for s are normally given only for comparison purposes)

$$\{{}^3\chi_c, {}^1\kappa_\alpha, S(1)\}: \quad Q = 3.47, \quad F = 60, \quad r = 0.943,$$
$$s = 0.27, \quad n = 26, \quad \langle u \rangle = 8.6,$$
$$\mathbf{u} = (10.4, 10.6, 9.4, 3.9)$$

The best single descriptor, $\{S(1)\}$, is, instead, a poor descriptor of log *K* with $Q = 0.86$, $F = 11$, $r = 0.56$, $s = 0.65$, and $n = 26$. While the best single-ψ index and the best linear combination of ψ indices are rather poor descriptors, the pseudoterm of eq 9 is a much better single descriptor than S(1)

$$Y_{\log K} = [({}^S\psi_E - 1.1\,{}^S\psi_I - {}^T\psi_I)/{}^0\psi_I]^{6.5} \qquad (9)$$

$$Q = 2.07, \quad F = 63, \quad r = 0.852, \quad s = 0.41, \quad n = 26,$$
$$\langle u \rangle = 24, \quad u = (8.0, 41)$$

With connectivity χ indices the modeling does not improve. Only the mixed combination $\{{}^S\psi_E, {}^1\chi\}$ achieves an interesting modeling with $Q = 2.14$, $F = 34$, and $r = 0.863$. A trial-and-error procedure discovers the following molecular connectivity term with a nice modeling power

$$X_{\log K} = (\chi_t)^4[D^v + 3.5 \cdot ({}^0\chi^v)^{1.2} + 5.8 \cdot (\chi_t^v)^{0.4} - 5.2 \cdot ({}^0\chi)]^{4.3} \qquad (10)$$

$$Q = 3.05, \quad F = 137, \quad r = 0.923, \quad s = 0.3, \quad n = 26,$$
$$\langle u \rangle = 38, \quad \mathbf{u} = (12, 65)$$

With X and Y terms at hand it is, now, possible to construct the mixed connectivity-pseudoconnectivity supraterm, $Z = f(X,Y)$ of eq 11, which mimics the *Q* value of

Table 1. Molecular Connectivity χ Values for 32 CFCs[a] (Only Heteroatoms Are Shown, See Table 2) and Their Molar Mass (M)

CFC	M	D	$^0\chi$	$^1\chi$	χ_t	D^v	$^0\chi^v$	$^1\chi^v$	χ_t^v
CCl[1]	50.5	2	2	1	1	1.78	2.13228	1.13228	1.13228
CF[1]	34	2	2	1	1	8	1.37796	0.37796	0.37796
CCl$_2$[2]	84.9	4	2.70711	1.41421	0.70711	3.56	2.97166	1.60128	0.90655
CF$_2$[2]	52	4	2.70711	1.41421	0.70711	16	1.46304	0.53452	0.10102
CClF[2]	68.5	4	2.70711	1.41421	0.70711	9.78	2.21735	1.06790	0.30261
CCF[2]	48.1	4	2.70711	1.41421	0.70711	10.0	2.08507	0.97437	0.26726
CCCl[2]	64.5	4	2.70711	1.41421	0.70711	3.78	2.83938	1.50775	0.80064
CF$_3$[3]	70	6	3.57735	1.73205	0.57735	24	1.71124	0.65465	0.03117
CCl$_2$F[3]	103	6	3.57735	1.73205	0.57735	11.56	3.21987	1.52566	0.29777
CClF$_2$[3]	86.5	6	3.57735	1.73205	0.57735	17.78	2.46556	1.09016	0.09339
CCl$_3$[3]	119	6	3.57735	1.73205	0.57735	5.34	3.97418	1.96116	0.83810
CCCl$_2$[3]	99	6	3.57735	1.73205	0.57735	5.56	3.84190	1.88479	0.74019
CCF$_2$[3]	66.1	6	3.57735	1.73205	0.57735	18.0	2.33328	1.01379	0.08248
CFCF	66.1	6	3.41421	1.91421	0.5	18.0	2.17014	1.03452	0.07143
CClCCl$_2$[4]	133	8	4.28446	2.27006	0.40825	7.34	4.68129	2.51633	0.59263
CFCF$_2$[4]	84	8	4.28446	2.27006	0.40825	26	2.41835	1.11195	0.02204
CClCF$_2$Cl[5]	135	10	5.20711	2.56066	0.35355	21.56	4.22759	2.09830	0.06475
CClCF$_3$[5]	119	10	5.20711	2.56066	0.35355	27.78	3.47328	1.72114	0.02162
CBrCF$_3$[5]	126.9	10	5.20711	2.56066	0.35355	27.26	4.30216	2.30725	0.03744
CICF$_3$[5]	209.9	10	5.20711	2.56066	0.35355	27.16	4.84100	2.68827	0.04773
CFCF$_3$[5]	102	10	5.20711	2.56066	0.35355	34	2.71896	1.18776	0.00722
CF$_2$CF$_2$	102	10	5.15470	2.64273	0.33333	34	2.66656	1.20620	0.00680
CCl$_2$CF$_3$[6]	153	12	6.07735	2.94338	0.28868	29.56	4.47580	2.16306	0.01983
CClFCF$_3$[6]	137	12	6.07735	2.94338	0.28868	35.78	3.72149	1.72756	0.00667
CF$_2$CF$_3$[6]	120	12	6.07735	2.94338	0.28868	42.0	2.96717	1.29206	0.00223
CClBrCF$_3$[6]	197.4	12	6.07735	2.94338	0.28868	29.04	5.30468	2.64162	0.03461
CFBrCF$_3$[6]	180.9	12	6.07735	2.94338	0.28868	35.26	4.55037	2.20612	0.01155
CFICF$_3$[6]	227.9	12	6.07735	2.94338	0.28868	35.16	5.08921	2.51722	0.01473
CBr$_2$CF$_3$[6]	241.8	12	6.07735	2.94338	0.28868	28.52	6.13357	3.12018	0.05995
CCF$_3$[7]	84	8	4.5	2	0.5	26.0	2.63389	1.06695	0.02700
CCF$_2$Cl[7]	101	8	4.5	2	0.5	19.78	3.38821	1.44410	0.08088
CCCl$_3$[7]	133	8	4.5	2	0.5	7.34	4.89683	2.19842	0.72582

[a] Sets of compounds with the same apex number have the same values of {D, $^0\chi$, $^1\chi$, χ_t} subset.

$\{^3\chi_c, {}^1\kappa_\alpha, S(1)\}$ combination, but with enhanced F and $\mathbf{u}$ values

$$Z_{\log K} = (X_{\log K})^{0.5} \cdot (Y_{\log K})^{0.3} \quad (11)$$

$Q = 3.41$, $F = 172$, $r = 0.937$, $s = 0.27$, $n = 26$, $\langle u \rangle = 35$, $\mathbf{u} = (13, 57)$, $\mathbf{C} = (-0.00047, 9.12978)$

The calculated log K values (Table 3) are obtained with equation **log $K = -0.00047 \cdot Z_{\log K} + 9.12978$**. Along the last column of Table 3 are the calculated log K values obtained with the leave-one-out (jack-knifing = jk) method, which together with the percent residual values, shown in the third column, $\Delta\% = |\log K - \log K\text{clc}| \cdot 100/\log K$, underlines the good level of the modeling. There is no deviation from the experimental values greater than 10%, and most of the deviations are < 5%.

Minimum Anesthetic Concentrations. The minimum anesthetic concentrations of 11 trifluoromethyl ethanes, tFMeE, expressed as logMAC is reported in Table 4. The connectivity and pseudoconnectivity values of these compounds are collected along the last rows of Tables 1 and 2. The suggested electrotopological S(F) index of the three geminal fluoro atoms[11] shows an interesting modeling power with

$\{S(F)\}$: $Q = 4.05$, $F = 74$, $r = 0.944$, $s = 0.2$, $n = 11$, $\langle u \rangle = 8.7$, $\mathbf{u} = (8.6, 8.8)$

Molar masses are less powerful, but, nevertheless, they are interesting descriptors for this property with $Q = 2.24$,

$F = 23$, $r = 0.847$, $s = 0.38$. While a valence χ index has a better modeling quality than S(F), a mixed $\chi - \psi$ combination, where D is highly degenerate (see Table 1), is the best two-index combination

$\{^1\chi^v\}$: $Q = 6.65$, $F = 200$, $r = 0.978$, $s = 0.15$, $n = 11$, $\langle u \rangle = 16$, $\mathbf{u} = (14, 17)$

$\{^0\psi_I, D\}$: $Q = 15.1$, $F = 510$, $r = 0.996$, $n = 11$, $s = 0.06_6$, $\langle u \rangle = 27$, $\mathbf{u} = (30, 24, 27)$

Things improve even further with Y term of eq 13, which is a better descriptor than the X term of eq 12

$$X_{\text{logMAC}} = ({}^1\chi^v + 1.2 \cdot \chi_t)^{0.5} \quad (12)$$

$Q = 7.84$, $F = 278$, $r = 0.981$, $s = 0.13$, $n = 11$, $\langle u \rangle = 18$, $\mathbf{u} = (17, 18)$

$$Y_{\text{logMAC}} = ({}^1\psi_I + {}^T\psi_I)/{}^S\psi_E \quad (13)$$

$Q = 15.3$, $F = 1065$, $r = 0.996$, $s = 0.06$, $n = 11$, $\langle u \rangle = 34$, $\mathbf{u} = (33, 35)$

The overall best modeling is found with the mixed connectivity-pseudoconnectivity term of eq 14. Equation **logMAC $= -20.5991 \cdot Z_{\text{logMAC}} + 27.2136$** has been used to derive the logMACclc values of Table 4, the residual values, $\Delta\%$, and the logMACclc(jk) values calculated with the leave-one-out (jack-knifing) method.

MOLECULAR CONNECTIVITY DESCRIPTORS

J. Chem. Inf. Comput. Sci., Vol. 41, No. 3, 2001 **839**

Table 2. 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$_3$Cl	6.11	1.20037	0.34879	0.34879	17.11	0.69281	0.11895	0.11895
CH$_2$Cl$_2$	9.72	1.80302	0.80550	0.19866	26.03	1.04637	0.26291	0.04122
CHCl3	13.66	2.34690	1.28314	0.10407	35.68	1.39314	0.42869	0.01386
CH$_3$F	10.00	1.06066	0.25000	0.25000	21.00	0.66645	0.10541	0.10541
CH$_2$F$_2$	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$_2$ClF	23.61	1.66331	0.69142	0.14239	30.11	1.03275	0.26350	0.03722
CHCl$_2$F	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$_3$CH$_2$Cl	7.61	2.01687	0.98010	0.28479	24.11	1.07743	0.27124	0.04552
CH$_3$CH$_2$F	11.50	1.87716	0.86603	0.20412	28.00	1.06714	0.27526	0.04163
CH$_3$CHCl$_2$	11.55	2.56074	1.46857	0.14918	33.56	1.42391	0.43028	0.01539
CH$_3$CHF$_2$	19.33	2.28132	1.22628	0.07664	41.32	1.44814	0.49327	0.01376
CH$_2$FCH$_2$F	19.00	2.34010	1.24402	0.08333	41.00	1.42984	0.44802	0.01357
CH$_2$ClCHCl$_2$	15.16	3.16340	1.96617	0.08497	42.67	1.77999	0.58378	0.00534
CH$_2$FCHF$_2$	26.83	2.74427	1.60962	0.03129	54.25	1.87126	0.74484	0.00489
CH$_2$ClCF$_2$Cl	23.97	3.86045	2.20669	0.04274	59.48	2.25853	0.99310	0.00191
CH$_2$ClCF$_3$	30.86	3.26485	2.08173	0.01592	63.85	2.42448	1.23127	0.00208
CH$_2$FCF$_3$	34.75	3.12514	1.96766	0.01141	67.74	2.62825	1.53805	0.00235
CHF$_2$CHF$_2$	34.66	3.14843	1.97816	0.01175	67.68	2.41002	1.20063	0.00196
CHCl$_2$CF$_3$	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$_2$CF$_3$	42.58	3.52930	2.33739	0.00428	81.10	4.55532	4.38982	0.00210
CH$_3$CCl$_3$	15.58	3.08133	1.95602	0.07590	43.08	1.77734	0.61915	0.00525
CH$_3$CF$_3$	27.25	2.66219	1.58114	0.02795	54.74	1.98926	0.95756	0.00542
CH$_3$CF$_2$Cl	23.36	2.80190	1.70610	0.03900	50.78	1.85862	0.77775	0.00509
CH$_2$BrCF$_3$	29.50	3.37461	2.17315	0.01941	62.14	2.52461	1.32327	0.00245
CH$_2$ICF$_3$	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$_2$CF$_3$	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).

$$Z_{logMAC} = [(X_{logMAC})^{0.01} + (Y_{logMAC})^{0.9}]^5 \quad (14)$$

$$Q = 17.7, \quad F = 1414, \quad r = 0.997, \quad s = 0.06, \quad n = 11,$$
$$\langle u \rangle = 38, \quad \mathbf{u} = (38, 39), \quad \mathbf{C} = (-20.5991, 27.2136)$$

2-Br-2-Phenetylamines. In Tables 5 and 6 are collected the connectivity and pseudoconnectivity values of 2-Br-2-phenetylamines, together with their molar mass M (see Table 5). Their hydrogen-suppressed formula is **NCCBrΦY$_p$Z$_m$**, where Y$_p$ and Z$_m$ are two substituents in the para and meta positions on the phenyl ring, Φ, and Br is bound to the second carbon atom from the left. Table 5 shows that this class of compounds has seven pairs of degenerate M values and that it can be divided into different subclasses. The first type of subclass, which is characterized by the same number as apex, has degenerate values for subset {D, $^0\chi$, $^1\chi$, χ_t}, and the second type of subclass, instead, which is characterized by the same number and letter as apexes, has degenerate values for the entire set of eight χ values. No redundancy can, instead, be detected with ψ indices (see Table 6). The modeling of the antagonism of adrenaline of these compounds measured in units of pED$_{50}$ (Table 7, first−fourth columns) shows that molar mass is a bad descriptor with $Q = 0.994$, $F = 6.7$, $r = 0.501$, and $s = 0.50$. A previous modeling[12] detected the following single-E-state-index and three-E-state-index descriptions

$$\{S11\_S12\}: \quad Q = 0.75, \quad F = 4, \quad r = 0.40, \quad s = 0.5$$
$$\{S11\_S12, S(10), S(9)\}: \quad Q = 4.75, \quad F = 49,$$
$$r = 0.95, \quad s = 0.2, \quad n = 22, \quad \langle u \rangle = 5.8,$$
$$\mathbf{u} = (1.0, 11, 11, 0.3)$$

While χ indices disclose no interesting modeling, in keeping with their high degeneracy, a ψ modeling discovers the following optimal descriptors

$$\{^0\psi_I\}: \quad Q = 2.17, \quad F = 32, \quad r = 0.785,$$
$$s = 0.4, \quad n = 22$$

$$\{^T\psi_I, {}^0\psi_E, {}^1\psi_E\}: \quad Q = 3.42, \quad F = 26, \quad r = 0.902,$$
$$s = 0.3, \quad n = 22, \quad \langle u \rangle = 4.4, \quad \mathbf{u} = (2.3, 7.6, 6.1, 1.4)$$

The last description with three indices is far from being optimal, especially for what concerns the last and first utility values. The description does not improve with the term X = f(χ) of eq 15, while it improves consistently with the term Y = f(ψ) of eq 16

$$X_{pED} = \chi_t \cdot [(^0\chi^v)^{-1} + (\chi_t^v)^{1.3}]^{3.5} \quad (15)$$

$$Q = 1.92, \quad F = 25, \quad r = 0.746, \quad s = 0.4, \quad n = 22$$

$$Y_{pED} = [^0\psi_I + (1.1 \cdot {}^T\psi_I)^{0.9} - {}^0\psi_E - 3.3]^{-0.01} \quad (16)$$

$$Q = 4.00, \quad F = 108, \quad r = 0.919, \quad s = 0.2, \quad n = 22,$$
$$\langle u \rangle = 11, \quad \mathbf{u} = (10, 11)$$

The overall best description is again achieved by the higher-level mixed connectivity-pseudoconnectivity term, Z = f(X, Y), of eq 17. It is by the aid of correlation vector **C** = (−15450.5, 104.947) and of the corresponding modeling equation **pED = −15450.5·Z$_{pED}$ + 104.947** that pED$_{50}$clc

Table 3. Rates of Hydrogen Abstraction in $n = 26$ CFCs in log K (Pex) Units, Calculated log Kclc (P_{Calc}), Calculated log K(jk) with the Leave-One-Out or Jackknifing Method, and Percent Residual Relatively to PK Values, $\Delta\% = |(P_{ex} - P_{calc})|\cdot100/P_{exp}$

CFC	log K	log Kclc	$\Delta\%$	log K(jk)
CH_3Cl	7.36	7.66	4.0	7.69
CH_2Cl_2	8.00	7.83	2.1	7.83
$CHCl_3$	7.80	7.45	4.5	7.46
CH_3F	6.95	7.03	1.1	7.03
CH_2F_2	6.81	6.52	4.2	6.54
CHF_3	5.10	5.30	3.9	5.38
CH_2ClF	7.46	7.47	0.11	7.47
$CHCl_2F$	7.30	7.23	1.0	7.22
$CHClF_2$	6.45	6.56	1.7	6.61
CH_3CH_2Cl	8.37	8.25	1.5	8.24
CH_3CH_2F	8.14	7.99	1.9	7.99
CH_3CHCl_2	8.20	7.91	3.5	7.89
CH_3CHF_2	7.48	7.18	4.0	7.17
CH_2FCH_2F	7.83	7.76	1.0	7.75
$CH_2ClCHCl_2$	8.28	8.13	1.8	8.12
CH_2FCHF_2	7.47	7.19	3.8	7.17
CH_2ClCF_2Cl	7.20	7.59	5.5	7.64
CH_2ClCF_3	6.95	7.04	1.4	7.05
CH_2FCF_3	6.70	6.40	4.5	6.36
CHF_2CHF_2	6.50	6.73	3.6	6.78
$CHCl_2CF_3$	7.40	7.87	6.4	7.96
$CHClFCF_3$	6.87	6.67	2.9	6.68
CHF_2CF_3	6.48	6.13	5.5	6.11
CH_3CCl_3	6.80	7.40	8.9	7.45
CH_3CF_3	5.95	6.00	0.8	6.05
CH_3CF_2Cl	6.60	6.81	3.2	6.86

Table 4. Logarithm of the Minimum Anesthetic Concentrations, logMAC (P_{ex}), of 11 Trifluoromethylethanes, TFMeE, Calculated logMAC (P_{calc}), Calculated logMAC(jk) with the Leave-One-Out or Jackknifing Method, and Percent Residual, $\Delta\%$

tFMeE	logMAC	logMACclc	$\Delta\%$	logmac(jk)
CH_3CF_3	1.60	1.62	1.2	1.63
CH_2ClCF_3	0.90	0.92	1.9	0.92
CH_2BrCF_3	0.45	0.39	15	0.37
CH_2ICF_3	0.10	0.09	11	0.09
CHF_2CF_3	1.70	1.71	0.7	1.72
$CHFClCF_3$	1.18	1.12	4.8	1.11
$CHFBrCF_3$	0.70	0.66	5.5	0.66
$CHFICF_3$	0.30	0.39	29	0.40
$CHCl_2CF_3$	0.43	0.43	1.1	0.42
$CHClBrCF_3$	- 0.10	- 0.01	92	0.01
$CHBr_2CF_3$	- 0.40	- 0.45	11	- 0.47

values and $\Delta\%$ values of Table 7 have been derived

$$Z_{pED} = (0.02\cdot Y_{pED} - X_{pED})^{1.3} \qquad (17)$$

$Q = 4.03, \quad F = 110, \quad r = 0.920, \quad s = 0.2, \quad n = 22,$
$$\langle u \rangle = 11, \quad \mathbf{u} = (11, 11)$$

Benzimidazoles I. In Tables 8 and 9 are collected the χ and ψ indices of 15 benzimidazoles, whose standard hydrogen-suppressed formula is $\mathbf{R}_n\boldsymbol{\Phi}\mathbf{ImR_2}$, where R_2 is a substituent at C2 in the imidazole portion and R_n is a substituent (one or more than one) at the phenyl ring, Φ. In this section pseudoconnectivity ψ_E values have been rescaled, for comparison purposes, even if every S_i value in these compounds is positive. The molar mass shows four pairs of compounds and a triplet of compounds with degenerate M values. Further, these compounds show three subclasses with degenerate $\{D, {}^0\chi, {}^1\chi, \chi_t\}$ values and a subclass of two compounds (class 2a) with degenerate values for the entire set of eight molecular connectivity indices. This class shows also a subclass (Table 9, with apex 1) of two compounds

with degenerate values for $\{{}^S\psi_I, {}^0\psi_I, {}^1\psi_I, {}^T\psi_I\}$ subset. It is, here, noticeable that D and D^v are collinear, while indices ${}^1\chi$ and ${}^1\chi^v$ have r(${}^1\chi, {}^1\chi^v$) = 0.9998.

Benzimidazoles are known for their effect in inhibiting influenza viruses, measured in units of pK[13] (see Table 7, sixth column). The description of pK offered by the molar mass, M, is rather remarkable with $Q = 4.70$, $F = 81$, $r = 0.928$, $s = 0.2$, $\langle u \rangle = 5.0$, and $\mathbf{u} = (8.9, 1.1)$. The description by following two E-state indices[13] by Kier and Hall is even more interesting

$\{S(N-1,3), S(C-2)\}$: $\quad Q = 5.89, \quad F = 63, \quad r = 0.956,$
$$n = 15, \quad s = 0.16, \quad \langle u \rangle = 7.2, \quad \mathbf{u} = (9.8, 4.1, 7.8)$$

Only $\{\chi_t\}$ offers a better description than M with $Q = 5.47$, $F = 110$, $r = 0.945$, $s = 0.17$, $n = 15$, $\langle u \rangle = 22$, $\mathbf{u} = (11, 34)$. No two-χ indices or two-ψ indices as well as a no mixed $\{\chi, \psi\}$ combinations achieve a better modeling than $\{S(N-1,3), S(C-2)\}$. The trial-and-error procedure discovers the Y term of eq 19, which is better than the X term of eq 18

$$X_{pK} = [(\chi_t)^{1.7} + 6\cdot(\chi_t{}^v)^{1.3}]^{1.1} \qquad (18)$$

$Q = 5.72, \quad F = 120, \quad r = 0.950, \quad s = 0.17, \quad n = 15,$
$$\langle u \rangle = 27, \quad \mathbf{u} = (11, 43)$$

$$Y_{pK} = [({}^S\psi_I)^{1.1} + 0.9\cdot({}^S\psi_E)]/{}^1\psi_E \qquad (19)$$

$Q = 6.07, \quad F = 135, \quad r = 0.955, \quad s = 0.16, \quad n = 15,$
$$\langle u \rangle = 12, \quad \mathbf{u} = (12, 12)$$

Now, X and Y terms can be employed to build the optimal mixed term of eq 20

$$Z_{pK} = (X_{pK} + 0.004\cdot Y_{pK})^{0.1} \qquad (20)$$

$Q = 6.22, \quad F = 141, \quad r = 0.957, \quad s = 0.15, \quad n = 15,$
$$\langle u \rangle = 12, \quad \mathbf{u} = (12, 12), \quad \mathbf{C} = (-497.069, 436.819)$$

Finally, equation $\mathbf{p}K = -\ \mathbf{497.069}\cdot\mathbf{Z_{pK}} + \mathbf{436.819}$ can be used to derive the calculated pKclc and $\Delta\%$ values shown in the seventh and eighth columns of Table 7.

Benzimidazoles II. Along columns 2−5 of Table 10 are collected the nonrescaled values of $\{\psi'_E\}$ subset only. The full set modeling set now is $\{\psi'\} = \{{}^S\psi_I, {}^0\psi_I, {}^1\psi_I, {}^T\psi_I, {}^0\psi'_E, {}^1\psi'_E, {}^T\psi'_E\}$, where ψ'_E values are the nonrescaled ψ_E values. Due to the fact that, now (see method section), ${}^S\psi_I = {}^S\psi'_E$, the size of the set is reduced to seven indices only.

The best single ψ'_E index is nearly as good as the corresponding best single $\{\chi_t\}$ index found in *the* benzimidazole I section; only the utility of the unitary index U_0 (index of the constant parameter of the regression) is consistently worse

$\{{}^1\psi'_E\}$: $\quad Q = 5.23, \quad F = 100, \quad r = 0.941, \quad s = 0.18,$
$$n = 15, \quad \langle u \rangle = 5.2, \quad \mathbf{u} = (10, 0.3)$$

Like with benzimidazoles I no linear combination of ψ indices shows an improved modeling quality. Here, the low utility of the unitary index prevents the consideration of the ${}^1\psi'_E$ index as an interesting descriptor for this property. A mixed combination of a χ-type index and a ψ'_E-type index

Molecular Connectivity Descriptors

J. Chem. Inf. Comput. Sci., Vol. 41, No. 3, 2001 **841**

Table 5. Molecular Connectivity Values for $n = 22$ 2-Br-2-Phenethylamines Whose Heteroatomic Formula Is $NCCBr\Phi Y_pZ_m{}^a$

Y_p/Z_m	M	O.D	$P.^0\chi$	$Q.^1\chi$	$R.\chi_t$	D^v	$S.^0\chi v$	$T.^1\chi v$	$U.\chi_t{}^v$
H/H	200.1	20	7.39734	4.84254	0.04167	27.26	7.20972	4.14813	0.01483
F/H[1a]	218.1	22	8.26758	5.23638	0.03402	35.26	7.51033	4.24780	0.00485
H/F[1a]	218.1	22	8.26758	5.23638	0.03402	35.26	7.51033	4.24780	0.00485
Cl/H[1b]	234.5	22	8.26758	5.23638	0.03402	29.04	8.26465	4.62495	0.01454
H/Cl[1b]	234.5	22	8.26758	5.23638	0.03402	29.04	8.26465	4.62495	0.01454
Br/H[1c]	279	22	8.26758	5.23638	0.03402	28.52	9.09353	5.03940	0.02518
H/Br[1c]	279	22	8.26758	5.23638	0.03402	28.52	9.09353	5.03940	0.02518
I/H[1d]	326	22	8.26758	5.23638	0.03402	28.42	9.63237	5.30882	0.03210
H/I[1d]	326	22	8.26758	5.23638	0.03402	28.42	9.63237	5.30882	0.03210
Me/H[1e]	214	22	8.26758	5.23638	0.03402	29.26	8.13237	4.55882	0.01284
H/Me[1e]	214	22	8.26758	5.23638	0.03402	29.26	8.13237	4.55882	0.01284
F/Cl[2]	252.5	24	9.13783	5.64707	0.02778	37.04	8.56526	4.73060	0.00476
F/Br[2]	297	24	9.13783	5.64707	0.02778	36.52	9.39415	5.14505	0.00824
F/Me[2]	232	24	9.13783	5.64707	0.02778	37.26	8.43298	4.66446	0.00420
Cl/Cl[2]	269	24	9.13783	5.64707	0.02778	30.82	9.31957	5.10776	0.01426
Cl/Br[2a]	313.4	24	9.13783	5.64707	0.02778	30.30	10.1485	5.52220	0.02469
Br/Cl[2a]	313.4	24	9.13783	5.64707	0.02778	30.30	10.1485	5.52220	0.02469
Cl/Me[2]	248.6	24	9.13783	5.64707	0.02778	31.04	9.18730	5.04162	0.01259
Br/Br[2]	357.9	24	9.13783	5.64707	0.02778	29.78	10.9773	5.93664	0.04277
Br/Me[2b]	293	24	9.13783	5.64707	0.02778	30.52	10.0162	5.45606	0.02181
Me/Br[2b]	293	24	9.13783	5.64707	0.02778	30.52	10.0162	5.45606	0.02181
Me/Me[2]	228.1	24	9.13783	5.64707	0.02778	31.26	9.05502	4.97548	0.01112

[a] Along column *M* are the molar mass values. Compounds with the same apex number share the same $\{D, {}^0\chi, {}^1\chi, \chi_t\}$ values; compounds with the same apex number and letter share the same $\{D, D^v, {}^0\chi, {}^0\chi^v, {}^1\chi, {}^1\chi^v, \chi_t, \chi_t{}^v\}$ values.

Table 6. I_S Molecular Pseudoconnectivity Values for $n = 22$ 2-Br-2-Phenethylamines[a] Whose Heteroatomic Formula Is $NCCBr\Phi Y_pZ_m$

Y_p/Z_m	${}^S\psi_I$	${}^0\psi_I$	${}^1\psi_I$	${}^T\psi_I$	${}^S\psi_E$	${}^0\psi_E$	${}^1\psi_E$	${}^T\psi_E$
H/H	21.25	7.09599	5.40447	0.029201	76.24	3.66247	1.40023	0.000042
F/H	28.92	7.51626	5.77241	0.11298	89.41	4.03535	1.60992	0.000014
Cl/H	25.03	7.65597	5.88052	0.015763	85.50	4.01417	1.55360	0.000014
Br/H	23.67	7.76573	5.96545	0.019270	84.16	4.02434	1.55079	0.000015
I/H	23.04	7.84951	6.03028	0.21948	83.52	4.03527	1.54903	0.000016
Me/H	22.92	7.86981	6.04600	0.022597	83.43	4.03725	1.54845	0.000016
H/F	28.92	7.51626	5.77241	0.011298	89.42	4.04020	1.61651	0.000014
H/Cl	25.03	7.65597	5.88052	0.015763	85.53	4.01514	1.56224	0.000014
H/Br	23.67	7.76573	5.96545	0.019270	84.15	4.02515	1.55212	0.000015
H/I	23.04	7.84951	6.03028	0.021948	83.53	4.03515	1.54889	0.000016
H/Me	22.92	7.86981	6.04600	0.022597	83.41	4.03785	1.54876	0.000016
F/Cl	32.70	8.07624	6.25291	0.006099	98.69	4.40118	1.79407	0.000005
F/Br	31.34	8.18600	6.33784	0.007456	97.34	4.40714	1.77686	0.000005
F/Me	30.59	8.29008	6.41839	0.008743	96.59	4.41980	1.77130	0.000005
Cl/Cl	28.81	8.21595	6.36102	0.008509	94.82	4.37065	1.73098	0.000005
Cl/Br	27.45	8.32571	6.44595	0.010402	93.44	4.37909	1.71746	0.000005
Cl/Me	26.70	8.42979	6.52650	0.012198	92.70	4.39135	1.71258	0.000005
Br/Cl	27.45	8.32571	6.44595	0.010402	93.44	4.37996	1.71867	0.000005
Br/Br	26.09	8.43546	6.53089	0.012717	92.09	4.38741	1.70467	0.000005
Br/Me	25.34	8.53955	6.61143	0.014912	91.32	4.40027	1.70062	0.000006
Me/Me	24.59	8.64363	6.69197	0.017486	90.58	4.41277	1.69715	0.000006
Me/Br	25.34	8.53955	6.61143	0.014912	91.35	4.39991	1.70071	0.000006

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

offers an interesting modeling but with deceiving utility values

$$\{\chi_t, {}^1\psi'_E\}: \quad Q = 5.76, \quad F = 61, \quad r = 0.954, \quad s = 0.17,$$
$$n = 15, \quad \langle u \rangle = 1.7, \quad \mathbf{u} = (1.8, 1.5, 1.9)$$

A trial-and-error procedure discovers the following optimal term, with an extraordinary improvement in u(U₀) relative to ${}^1\psi'_E$ (from 0.3 to 14). This term is even better than the mixed connectivity-pseudoconnectivity term, Z = f(X, Y), of benzimidazoles I

$$Y'_{pK} = [{}^S\psi_I + 2\cdot{}^1\psi_I]/[{}^1\psi'_E + 3\cdot{}^0\psi_I] \qquad (21)$$

$$Q = 6.41, \quad F = 150, \quad r = 0.959, \quad s = 0.15, \quad n = 15,$$
$$\langle u \rangle = 13, \quad \mathbf{u} = (12, 14)$$

But the Y'_{pK} term is not the best overall term. The mixed connectivity-pseudoconnectivity term, $Z' = f(X, Y)$, of eq 22 is the best overall descriptor for pK (for X see eq 18). Calculated, pK_{clc}, and percent residual $\Delta\%$ values can be obtained with equation $\mathbf{pK = -7.78417 \cdot Z' pK + 12.4389}$.

$$Z'_{pK} = [Y'_{pK} + 15\cdot(X_{pK})^{1.1}]^{1.5} \qquad (22)$$

$$Q = 6.50, \quad F = 155, \quad r = 0.960, \quad s = 0.15, \quad n = 15,$$
$$\langle u \rangle = 15, \quad \mathbf{u} = (12, 17), \quad \mathbf{C} = (-7.78417, 12.4389)$$

The modeling with both rescaled and nonrescaled ψ values tells us (Table 7, last columns and Table 10, last columns) that compound Et/5-Me is a strong outlier, with $\Delta\% \approx 10$, while for the other compounds the modeling is rather

Table 7. Antagonism of Adrenalin by 2-Br-2-Phenethylamines (A: Y_p/Z_m) in PED_{50} Units, Calculated $pED_{50}clc$ Values, and Percent Residual Relative to PED_{50} and Inhibition of Influenza Virus by Benzimidazoles (B: R_2/R_n) in pK Units, Calculated pKclc Values, and Percent Residual, $\Delta\%$

A: Y_p/Z_m	pED_{50}	$pED_{50}clc$	$\Delta\%$	B: R_2/R_n	pK	$pKclc$	$\Delta\%$
H/H	7.46	7.51	0.7	H/H	2.14	1.12	0.7_5
F/H	8.16	8.20	0.5	Me/H	2.51	2.63	4.8
Cl/H	8.68	8.28	4.6	H/5-Me	2.72	2.55	6.1
Br/H	8.89	8.57	3.6	H/5,6-dMe	2.72	2.88	5.7
I/H	9.25	9.21	0.5	H/4,6-dMe	2.82	2.90	3.0
Me/H	9.30	8.83	5.0	Me/5-Me	2.89	2.93	1.4
H/F	7.52	7.52	0.05	H/4,5-dMee	2.87	2.93	3.1
H/Cl	8.16	8.28	1.5	Me/5,6-dMe	3.05	3.20	5.1
H/Br	8.30	8.57	3.2	Me/4,5-dMe	3.20	3.21	0.3
H/I	8.40	8.75	4.2	H/5,6-dEt	3.39	3.52	3.9
H/Me	8.46	8.83	4.4	2-Pr/5-Me	3.60	3.64	1.1
F/Cl	8.19	8.29	1.3	Me/4,5,6,7-ttMe	3.66	3.60	1.7
F/Br	8.57	8.57	0.04	Et/5-Me	3.74	3.35	10.5
F/Me	8.82	8.82	0.01	Bu/5-Me	3.77	3.88	2.8
Cl/Cl	8.89	8.74	1.7	iPr/5-Me	3.77	3.66	2.8
Cl/Br	8.92	8.94	0.3				
Cl/Me	8.96	9.12	1.8	Et = ethyl		Me = methyl	
Br/Cl	9.00	9.16	1.8	Pr = propyl		Bu = butyl	
Br/Br	9.35	9.13	2.3				
Br/Me	9.22	9.27	0.6	d = di		t = tri	
Me/Me	9.30	9.42	1.3	tt = tetra		i = iso	
Me/Br	9.52	9.27	2.6				

wealthy. The mean $<\Delta\%>$ value is 3.5 for benzimidazoles I (Table 7) and 3.2 for benzimidazoles II (Table 10), attesting no decisive preference either for rescaled or nonrescaled ψ values.

<div align="center">PRIMARY AMINES</div>

The boiling points of primary amines, $T_b(Am)$, examined by Kier and Hall[14] with E-state indices within the frame of a bimolecular encounter model offer an interesting benchmark to evaluate the different descriptive capabilities of χ and ψ indices, as for this property even the molar mass M is a good descriptor. The boiling points of $n = 21$ primary amines are shown in Table 11, while χ and ψ indices are collected in Tables 12 and 13. Kier and Hall achieved an optimal description of this property with a mixed set of four indices, i.e, three bimolecular encounter indices and an E-state index of the primary amine group, $S(-NH_2)$: $Q = 0.32$, $F = 1647$, $r = 0.999$, $s = 3.1$, $n = 21$, $\langle u \rangle = 7.0$, $\mathbf{u} = (9.2, 5.9, 15, 3.6, 1.1)$. The only negative point of this interesting description is the rather low utility value of the constant parameter (1.1).

The medium-sized set of χ indices of these amines (see Table 12) is reduced, due to collinearity, to five indices only, i.e., $\{D, {}^0\chi, {}^1\chi, \chi_t, {}^1\chi^v\}$. In fact $r(D, D^v) = r({}^0\chi, {}^0\chi^v) = r(\chi_t, \chi_t^v) = 1$. As ψ_E indices for amines are not rescaled, the medium-sized set of ψ indices is made up of seven indices only: $\{{}^S\psi_I, {}^0\psi_I, {}^1\psi_I, {}^T\psi_I, {}^0\psi_E, {}^1\psi_E, {}^T\psi_E\}$. Two pairs of amines show degenerate nonvalence connectivity indices, $\{D, {}^0\chi, {}^1\chi, \chi_t\}$. No amines, thanks to the ${}^1\psi_I$ index, show any degeneracy for the entire subset $\{{}^S\psi_I, {}^0\psi_I, {}^1\psi_I, {}^T\psi_I\}$ (see Tables 12 and 13). Table 11 tells us that there is a consistent degeneracy in the molar mass, with only 11 different values for M over 21 compounds. As already said, the molar mass, notwithstanding its redundancy, is a rather good descriptor of $T_b(Am)$, with $Q = 0.12$, $F = 869$, $r = 0.989$, $s = 8.5$, $n = 21$, $\mathbf{u} = (30, 36)$, $\langle u \rangle = 33$. The best single- and two- and

three-index descriptions are achieved by χ-type indices with

$$\{{}^1\chi\}: \quad Q = 0.20, \quad F = 2554, \quad r = 0.996, \quad s = 5,$$
$$n = 21, \quad \langle u \rangle = 60, \quad \mathbf{u} = (51, 70)$$

$$\{{}^1\chi, \chi_t\}: \quad Q = 0.29, \quad F = 2717, \quad r = 0.998, \quad s = 3.4,$$
$$n = 21, \quad \langle u \rangle = 22, \quad \mathbf{u} = (27, 4.7, 36)$$

$$\{D, {}^1\chi, \chi_t\}: \quad Q = 0.32, \quad F = 2210, \quad r = 0.998,$$
$$s = 3.1, \quad n = 21, \quad \langle u \rangle = 14, \quad \mathbf{u} = (2.2, 11, 5.7, 39)$$

With more than three indices the description starts to worsen. The last description is somewhat better than the cited description achieved in ref 15. Notice that connectivity combinations are composed of nonvalence indices only, while the nondegenerate ${}^1\chi^v$ index has no role in the description of this property. A trial-and-error search for the best term in both χ- and ψ-representations discovers the two terms of eqs 23 and 24, where the χ-based term of eq 23, centered on two-nonvalence indices, shows an exceptional good quality at every statistical level

$$X_T(Am) = [{}^1\chi - 1.1 \cdot (\chi_t)^{0.7}] \qquad (23)$$

$$Q = 0.31, \quad F = 6184, \quad r = 0.998, \quad s = 3.2, \quad n = 21,$$
$$\langle u \rangle = 128, \quad \mathbf{u} = (79, 177), \quad \mathbf{C} = (42.2300, 268.602)$$

$$Y_T(Am) = [({}^0\psi_I)^{0.9} - 2({}^T\psi_E)^{0.4}] \qquad (24)$$

$$Q = 0.20, \quad F = 2446, \quad r = 0.996, \quad s = 5.1, \quad n = 21,$$
$$\langle u \rangle = 83, \quad \mathbf{u} = (49, 117)$$

Interesting here is the very good improvement in Q value achieved by X_T when combined with the nondegenerate ${}^1\chi^v$ index

$$\{X_T, {}^1\chi^v\}: Q = 0.42, \quad F = 5545, \quad r = 0.9992, \quad s = 2.4,$$
$$n = 21, \quad \langle u \rangle = 6.5, \quad \mathbf{u} = (8.5, 4.0, 6.9)$$

This $\{X_T, {}^1\chi^v\}$ combination overrates at the Q, F, and u_i level the original Kier and Hall combination. Any attempt to detect a higher level optimal $Z = f(X, Y)$ term has, up to now, failed. Calculated T_{bclc} values of Table 11 as well as the corresponding percent residual values, $\Delta\%$, have been obtained with equation $\mathbf{T_b = 42.2300 \cdot X_T(Am) + 268.602}$. From $\Delta\%$ values we notice that the simulation is quite good, as no $\Delta\%$ exceeds 1.1%.

<div align="center">PRIMARY ALCOHOLS</div>

The boiling points of $n = 27$ alcohols, $T_b(Al)$, are collected in Table 14, and their connectivity and pseudoconnectivity values are collected in Tables 15 and 16. This modeling has also been studied by Kier and Hall within the frame of the E-state and bimolecular encounter models.[15] The achieved modeling with three indices, made up of two bimolecular encounter indices and an $S(-OH)$ index, was very satisfactory with $Q = 0.166$, $F = 104$, $r = 0.962$, $s = 5.8$, $n = 27$, $\langle u \rangle = 6.4$, $\mathbf{u} = (2.6, 9.4, 5.8, 7.8)$. From this result and the corresponding results in amines we notice that the boiling points of amines are better described than the boiling points of alcohols. Primary alcohols show a high M degeneracy with seven different molar mass values only out of 21

Table 8. Molar Mass and Molecular Connectivity Values for $n = 15$ Benzimidazoles Whose Heteroatomic Formula Is $R_n\Phi ImR_2$[a,b]

R_2/R_n[c]	M	D	$^0\chi$	$^1\chi$	χ_t	D^v	$^0\chi^v$	$^1\chi^v$	χ_t^v
H/H	118.1	20	6.10445	4.46633	0.02946	32	4.83396	2.84783	0.00359
Me/H[1]	132.2	22	6.97469	4.86017	0.02406	34	5.75661	3.27456	0.00311
H/5-Me[1]	132.2	22	6.97469	4.86017	0.02406	34	5.75661	3.25851	0.00311
H/5,6-dMe[2a]	149.2	24	7.84493	5.27086	0.01964	36	6.67926	3.67518	0.00269
H/4,6-dMe[2a]	146.2	24	7.84493	5.27086	0.01964	36	6.67926	3.67518	0.00269
Me/5-Me[2]	146.2	24	7.84493	5.25402	0.01964	36	6.67926	3.68525	0.00269
H/4,5-dMe[2]	146.2	24	7.84493	5.25402	0.01964	36	6.67926	3.68116	0.00269
Me/5,6-dMe[3]	160.2	26	8.71518	5.66470	0.01604	38	7.60191	4.10191	0.00233
Me/4,5-dMe[3]	160.2	26	8.71518	5.66470	0.01604	38	7.60191	4.10790	0.00233
H/5,6-dEt	174.3	28	9.25915	6.34687	0.00982	40	8.09348	4.79650	0.00134
Pr/5-Me	174.3	28	9.25915	6.29203	0.00982	40	8.09348	4.74591	0.00134
Me/4−7-ttMe	188.3	30	10.4557	6.51974	0.01069	42	9.44721	4.94721	0.00175
Et/5-Me	160.2	26	8.55204	5.79203	0.01389	38	7.38637	4.24591	0.00190
Bu/5-Me	188.3	30	9.96626	6.79203	0.00694	42	8.80058	5.24591	0.00095
*i*Pr/5-Me	174.3	28	9.42229	6.16470	0.01134	40	8.25661	4.62862	0.00155

[a] R_2 is at C2 position on the imidazole portion, Im, and R_n is on the benzene portion, Φ, and can be more than one substituent. [b] Me = methyl, Et = ethyl, Pr = propyl, Bu = butyl, d = di, tt = tetra, i = iso, 4−7 = 4,5,6,7. [c] Sets of compounds with the same apex number share the same $\{D, {}^0\chi, {}^1\chi, \chi_t\}$ values; compounds with the same apex number and letter share the same $\{D, D^v, {}^0\chi, {}^0\chi^v, {}^1\chi, {}^1\chi^v, \chi_t, \chi_t^v\}$ values.

Table 9. I_s Molecular Pseudoconnectivity Values for $n = 15$ Benzimidazoles[a]

R_2/R_n[b]	$^S\psi_I$	$^0\psi_I$	$^1\psi_I$	$^T\psi_I$	$^S\psi_E$	$^0\psi_E$	$^1\psi_E$	$^T\psi_E$
H/H	18.84	6.29299	4.98479	0.03865	68.34	3.28158	1.34686	0.00011
Me/H	20.51	7.06681	5.61268	0.02991	75.50	3.65971	1.49758	0.000042
H/5-Me	20.51	7.06681	5.62632	0.02991	75.50	3.65685	1.49489	0.000042
H/5,6-dMe[1]	22.18	7.84063	6.27230	0.02315	82.67	4.03159	1.64248	0.000016
H/4,6-dMe[1]	22.18	7.84063	6.27230	0.02315	82.65	4.03244	1.64325	0.000016
Me/5-Me	22.18	7.84063	6.25421	0.02315	82.69	4.03416	1.64483	0.000016
H/4,5-dMe	22.18	7.84063	6.27675	0.02315	82.69	4.03119	1.64249	0.000016
Me/5,6-dMe	23.85	8.61446	6.90019	0.01791	89.84	4.40958	1.79323	0.000006
Me/4,5-dMe	23.85	8.61446	6.90464	0.01791	89.84	4.40955	1.79342	0.000006
H/5,6-dEt	25.18	9.47363	7.59629	0.01543	96.68	4.79045	1.92871	0.000002
Pr/5-Me	25.18	9.47363	7.58287	0.01543	96.70	4.79445	1.93431	0.000002
Me/4−7-ttMe	27.19	10.1621	8.20104	0.01072	104.15	5.15994	2.08987	0.000001
Et/5-Me	23.68	8.65713	6.91621	0.01890	89.70	4.41493	1.78998	0.000006
Bu/5-Me	26.68	10.2901	8.24954	0.01260	103.66	5.17445	2.07883	0.000001
*i*Pr/5-Me	25.51	9.41485	7.60430	0.01419	96.98	4.79045	1.94886	0.000002

[a] ψ_E values have been obtained after a rescaling procedure (see Method section). [b] Sets of compounds with the same apex values share the same values for subset $\{\psi_I\}$.

compounds (see Table 14). Further, the $\{\chi\}$ set has only five indices due to collinearity: $r(D, D^v) = r({}^0\chi, {}^0\chi^v) = r(\chi_t, \chi_t^v) = 1$. While no subset of ψ indices shows the same set of values for any alcohol (Table 16), subset $\{D, {}^0\chi, {}^1\chi, \chi_t\}$ shows a couple and a triplet of alcohols with degenerate values (Table 15). The molar mass, with a high degeneracy, is no interesting descriptor for this property of alcohols with $Q = 0.064$, $F = 87$, $r = 0.881$, $s = 14$, and $n = 27$.

The best single-, two- and three-(χ or ψ)-index descriptions are

$$\{\chi_t\}: \quad Q = 0.079, \quad F = 130, \quad r = 0.916, \quad s = 11.6,$$
$$\langle u \rangle = 50, \quad \mathbf{u} = (11, 88)$$

$$\{{}^0\psi_I, {}^0\psi_E\}: \quad Q = 0.100, \quad F = 105, \quad r = 0.947,$$
$$s = 9.5, \quad \langle u \rangle = 12, \quad \mathbf{u} = (5.4, 4.7, 26)$$

$$\{{}^0\psi_I, {}^0\psi_E, {}^T\psi_E\}: \quad Q = 0.102, \quad F = 73, \quad r = 0.951,$$
$$s = 9.3, \quad \langle u \rangle = 8.1, \quad \mathbf{u} = (5.6, 5.0, 1.4, 20)$$

No improved mixed $\psi - \chi$ description can be detected. It is at the level of the connectivity and pseudoconnectivity terms that the best description is to be found. The best description is achieved by the pseudoconnectivity term of

eq 26, while no Z term could be detected

$$X_T(Al) = \chi_t \cdot ({}^1\chi + 0.8) \tag{25}$$

$$Q = 0.101, \quad F = 214, \quad r = 0.946, \quad s = 9.4, \quad \langle u \rangle = 45,$$
$$\mathbf{u} = (15, 76)$$

$$Y_T(Al) = \left[\frac{{}^0\psi_I - {}^0\psi_E}{{}^S\psi_I + {}^1\psi_E} \right]^{2.4} \tag{26}$$

$$Q = 0.107, \quad F = 238, \quad r = 0.951, \quad s = 8.9, \quad \langle u \rangle = 26,$$
$$\mathbf{u} = (15, 37), \quad \mathbf{C} = (16947.9, 293.506)$$

Calculated T_{bclc} values of Table 14 have been obtained with equation $T_b = \mathbf{16947.9 \cdot Y_T + 293.506}$. The percent residual $\Delta\%$ values show that only a compound exceeds $\Delta\% = 5\%$, and four have $\Delta\%$ between 2% and 3%.

THE MIXED CLASS AMINES PLUS ALCOHOLS

The boiling points, T_b of this mixed class of $n = 48$ [Am + Al], simulated by the aid of the connectivity and pseudoconnectivity sets of five χ and seven ψ indices, $\{D, {}^0\chi, {}^1\chi, \chi_t, {}^1\chi^v\}$ and $\{{}^S\psi_I, {}^0\psi_I, {}^1\psi_I, {}^T\psi_I, {}^0\psi_E, {}^1\psi_E, {}^T\psi_E\}$, respectively, show a high M degeneracy. This degeneracy

844 *J. Chem. Inf. Comput. Sci., Vol. 41, No. 3, 2001*

POGLIANI

Table 10. I_s Molecular Pseudoconnectivity $\{\psi'_E\}$ Values for $n = 15$ Benzimidazoles, Where $\{\psi'_E\}$ Have Not Been Rescaled[a]

R_2/R_n	$^S\psi_I{}^b$	$^0\psi'_E$	$^1\psi'_E$	$^T\psi'_E$	pK	pKclc	$\Delta\%$
H/H	18.84	6.62013	5.69519	0.05242	2.14	2.12	0.5
Me/H	20.51	7.53400	6.59110	0.04595	2.51	2.63	4.6
H/5-Me	20.51	7.44363	6.47705	0.04246	2.72	2.55	6.1
H/5,6-dMe	22.18	8.24859	7.24784	0.03386	2.72	2.87	5.6
H/4,6-dMe	22.18	8.25984	7.27783	0.03430	2.82	2.89	2.4
Me/5-Me	22.18	8.34573	7.34822	0.03677	2.89	2.93	1.5
H/4,5-dMe	22.18	8.24172	7.27605	0.03364	2.96	2.88	2.6
Me/5,6-dMe	23.85	9.16085	8.13803	0.02964	3.05	3.19	4.5
Me/4,5-dMe	23.85	9.15435	8.16414	0.02951	3.20	3.19	0.2
H/5,6-dEt	25.18	9.98759	8.68659	0.02474	3.39	3.59	4.3
Pr/5-Me	25.18	10.1028	8.92516	0.02747	3.60	3.63	0.8
Me/4−7-ttMe	27.19	10.7625	9.73167	0.01883	3.66	3.56	2.9
Et/5-Me	23.68	9.25191	8.18209	0.03251	3.74	3.35	10.4
Bu/5-Me	26.68	10.9498	9.64630	0.02317	3.77	3.84	1.8
iPr/5-Me	25.51	10.3734	9.90823	0.03143	3.77	3.77	0.05

[a] See text. For the meaning of the last three columns see Table 7 and text. [b] Nonrescaled $^S\psi'_E$ values equal $^S\psi_I$ values (see text).

Table 11. Primary Amines, Their Molar Mass, M, Boiling Points (T_b/K), Calculated Boiling Points (T_{bclc}/K), and Percent Residual Relatively to T_b Values, $\Delta\%$

amine − R[a]	M	T_b/K	T_{bclc}/K	$\Delta\%$
CH$_3$−	31.1	256.6$_5$	264.4	0.5
CH$_3$CH$_2$−	45.1	290.1$_5$	291.9	0.6
(CH$_3$)$_2$CH−	50.1	307.1$_5$	310.1	1.0
CH$_3$CH$_2$CH$_2$−	59.1	322.1$_5$	320.8	0.4
CH$_3$CH$_2$CH(CH$_3$)−[1]	73.1	336.1$_5$	339.7	1.0
CH$_3$CH(CH$_3$)CH$_2$−[1]	73.1	341.1$_5$	339.7	0.4
CH$_3$CH$_2$CH$_2$CH$_2$−[1]	73.1	350.9$_5$	348.1	0.8
CH$_3$CH$_2$C(CH$_3$)$_2$−[2]	87.2	351.1$_5$	354.3	0.9
(CH$_3$CH$_2$)$_2$CH−[2]	87.2	364.1$_5$	367.7	1.0
CH$_3$CH$_2$CH$_2$CH(CH$_3$)−[2]	87.2	365.1$_5$	366.1	0.3
CH$_3$CH(CH$_3$)CH$_2$CH$_2$−[2]	87.2	368.1$_5$	366.1	0.6
CH$_3$C(CH$_3$)$_2$CH(CH$_3$)−[3]	101.2	375.1$_5$	373.4	0.5
CH$_3$(CH$_2$)$_4$−[2]	87.2	377.5$_5$	374.1	0.9
CH$_3$(CH$_2$)$_3$CH(CH$_3$)−[3]	101.2	390.6$_5$	391.4	0.2
CH$_3$(CH$_2$)$_5$−[3]	101.2	403.1$_5$	399.0	1.0
CH$_3$CH$_2$CH(CH$_3$)CH$_2$CH(CH$_3$)−[4]	115.2	406.6$_5$	410.1	1.1
CH$_3$(CH$_2$)$_4$CH(CH$_3$)−[4]	115.2	415.1$_5$	415.8	0.2
CH$_3$(CH$_2$)$_6$−[4]	115.2	430.0$_5$	423.1	1.6
CH$_3$(CH$_2$)$_7$−	129.2	449.1$_5$	446.5	0.6
CH$_3$(CH$_2$)$_8$−	143.2	465.1$_5$	469.5	0.9
CH$_3$(CH$_2$)$_9$−	157.3	490.1$_5$	492.0	0.4

[a] Compounds with the same apex number share the same molar mass value.

is in a minor way reflected in the high similarity of some T_b values. Notwithstanding its high degeneracy the M description seems rather good with $Q = 0.051$, $F = 263$, $r = 0.923$, $s = 18.2$, $n = 48$, $\langle u \rangle = 20$, $\mathbf{u} = (16, 24)$.

The best single-χ and single-ψ descriptors are

$\{^1\chi\}$: $Q = 0.051$, $F = 263$, $r = 0.923$, $s = 18.2$,
$n = 48$, $\langle u \rangle = 22$, $\mathbf{u} = (16, 27)$

$\{^S\psi_I\}$: $Q = 0.052$, $F = 282$, $r = 0.927$, $s = 17.7$,
$n = 48$, $\langle u \rangle = 17$, $\mathbf{u} = (17, 18)$

Only the following $\chi - \psi$ mixed description shows a consistent improvement relative to the M description

$\{^1\chi, {}^T\psi_E\}$: $Q = 0.087$, $F = 387$, $r = 0.972$,
$s = 11.2$, $n = 48$, $\langle u \rangle = 24$, $\mathbf{u} = (20, 8.8, 41)$

Table 12. Molecular Connectivity Indices of Primary Amines of Table 11

amine − R[a]	D	$^0\chi$	$^1\chi$	χ_t	$^1\chi^v$
CH$_3$−	2	2	1	1	0.57735
CH$_3$CH$_2$−	4	2.70711	1.41421	0.70711	1.11536
(CH$_3$)$_2$CH−	6	3.57735	1.73205	0.57735	1.48803
CH$_3$CH$_2$CH$_2$−	6	3.41521	1.91421	0.50000	1.61536
CH$_3$CH$_2$CH(CH$_3$)−[1]	8	4.28446	2.27006	0.40825	2.02604
CH$_3$CH(CH$_3$)CH$_2$−[1]	8	4.28446	2.27006	0.40825	1.97120
CH$_3$CH$_2$CH$_2$CH$_2$−	8	4.12132	2.41421	0.35355	2.11536
CH$_3$CH$_2$C(CH$_3$)$_2$−	10	5.20711	2.56066	0.35355	2.34934
(CH$_3$CH$_2$)$_2$CH−	10	4.99156	2.80806	0.28868	2.56404
CH$_3$CH$_2$CH$_2$CH(CH$_3$)−[2]	10	4.99156	2.77006	0.28868	2.52604
CH$_3$CH(CH$_3$)CH$_2$CH$_2$−[2]	10	4.99156	2.77006	0.28868	2.47120
CH$_3$C(CH$_3$)$_2$CH(CH$_3$)−	12	6.07735	2.94338	0.28868	2.69936
CH$_3$(CH$_2$)$_4$−	10	4.82843	2.91421	0.25	2.61536
CH$_3$(CH$_2$)$_3$CH(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$_3$CH$_2$CH(CH$_3$)CH$_2$CH(CH$_3$)−	14	6.56891	3.66390	0.16667	3.41989
CH$_3$(CH$_2$)$_4$CH(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

[a] Compounds with the same apex value share the same $\{D, {}^0\chi, {}^1\chi, \chi_t\}$ values.

The Y_T trial-and-error term of eq 27 offers the best overall description of the boiling points of this mixed class. Notice that X_T (eq 26) and Y_T terms are formally quite similar to terms used to simulate the boiling points of amines, $X_T(Am)$ and $Y_T(Am)$

$$X_T = [^1\chi - 2.5 \cdot (\chi_t)^{1.5}] \tag{26}$$

$Q = 0.056$, $F = 325$, $r = 0.936$, $s = 16.6$, $n = 48$,
$\langle u \rangle = 37$, $\mathbf{u} = (18, 55)$

$$Y_T = [(^0\psi_I)^{0.9} - 2.6 \cdot (^T\psi_E)^{0.7}] \tag{27}$$

$Q = 0.084$, $F = 718$, $r = 0.969$, $s = 11.6$, $n = 48$,
$\langle u \rangle = 44$, $\mathbf{u} = (27, 62)$

While the X_T term is a strict "dead-end" term,[9] term Y_T in combination with a pseudoindex offers an improved Q description, which will here be used to simulate the boiling points of this mixed class of compounds [Am + Al]

$\{Y_T, {}^S\psi_I\}$: $Q = 0.091$, $F = 423$, $r = 0.974$,
$s = 10.7$, $n = 48$, $\langle u \rangle = 11$, $\mathbf{u} = (8.9, 2.9, 21)$

$\mathbf{C} = (45.3962, -6.677, 324.882)$

It is by the aid of equation $\mathbf{T_b = 45.3962 \cdot Y_T - 6.677 \cdot {}^S\psi_I + 324.882}$ that the plot of Figure 1 has been obtained. This figure underlines the good quality of the simulation.

Let us now see how these two terms describe the single classes of compounds, i.e., which term is also a good descriptor of both amines and alcohols separately

Al) $\{Y_T\}$: $n = 27$, $Q = 0.069$, $F = 100$, $r = 0.895$,
$s = 12.9$, $n = 48$

Am) $\{Y_T\}$: $n = 21$, $Q = 0.156$, $F = 1570$,
$r = 0.994$, $s = 6.4$, $n = 48$

Table 13. I_S Molecular Pseudoconnectivity Indices of Primary Amines of Table 11[a]

amine − R	$^S\psi_I$	$^0\psi_I$	$^1\psi_I$	$^T\psi_I$	$^0\psi_E$	$^1\psi_E$	$^T\psi_E$
Me—	6	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$_2$CH—	9.33	2.78132	1.65983	0.21678	3.61907	3.26968	0.39695
EtCH$_2$—	9	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)$_2$CH—	12.33	4.41432	3.00424	0.14452	5.21030	4.81326	0.27522
EtCH$_2$CH(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)$_2$CH(Me)—	14.58	5.08996	3.71963	0.09695	6.89190	9.38091	0.32777
CH$_3$(CH$_2$)$_4$—	12	4.47309	3.15470	0.15713	4.87475	3.64489	0.22826
Me(CH$_2$)$_3$CH(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$_2$CH(Me)—	15.66	5.98853	4.36116	0.08861	7.02595	6.71190	0.20030
Me(CH$_2$)$_4$CH(Me)—	15.33	6.04731	4.33204	0.09635	6.91203	6.13575	0.19221
CH$_3$(CH$_2$)$_6$—	15	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	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$_3$CH$_2$.
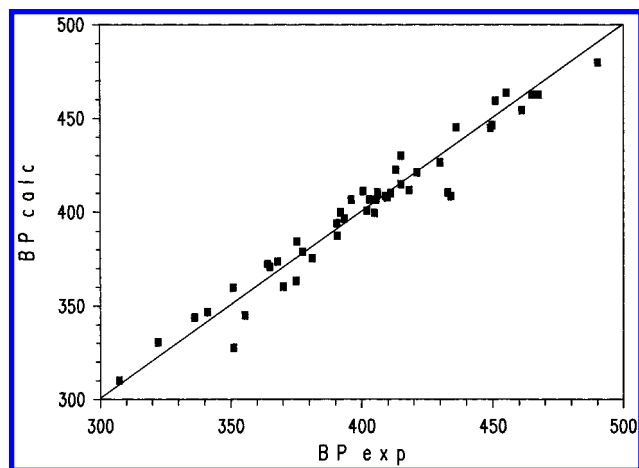
Figure 1. Plot of the calculated versus the experimental boiling points (BPcalc vs BP exp) of primary amines ($n = 21$) and alcohols ($n = 27$).

Al) $\{X_T\}$: $n = 27$, $Q = 0.086$, $F = 154$, $r = 0.928$,
$$s = 10.8, \quad n = 48$$

Am) $\{X_T\}$: $n = 21$, $Q = 0.107$, $F = 728$,
$$r = 0.987, \quad s = 9.3, \quad n = 48$$

Taking average $\langle r \rangle$, as a sign of "robustness" of the term, then the X_T term is the most robust term, if, instead, average $\langle Q \rangle$ is taken as a sign of "robustness", the most robust term is Y_T term.

CONCLUSION

Let us now try to find an answer to the title of this paper. Between the two types of indices, χ and ψ, the score is just flat, and there is no way to tell with "a priori" considerations which of the two types of indices is to be preferred, as much as the degeneracy of a particular set of indices is not[16] excessive. The discovery that mixed connectivity-pseudoconnectivity mixed terms, $Z = f(X,Y)$, can be good descriptors of properties and activities renders the mixed basis set, $\{\chi,$

Table 14. Alcohols, Their Molar Mass, M, Boiling Points (T_b/K), Calculated Boiling Points (T_{bclc}/K), and Percent Residual Relatively to T_b Values, $\Delta\%$

alcohol − R[a]	M	T_b/K	T_{bclc}/K	$\Delta\%$
(CH$_3$)$_2$CH—[1]	60.1	355.5$_5$	354.8	0.2
CH$_3$CH$_2$CH$_2$—[1]	60.1	370.2$_5$	366.2	1.1
CH$_3$CH$_2$C(CH$_3$)$_2$—[2]	88.2	375.4$_5$	387.1	3.1
CH$_3$CH(CH$_3$)CH$_2$—[3]	74.1	381.2$_5$	380.8	0.1
CH$_3$(CH$_2$)$_3$—[3]	74.1	390.7$_5$	392.8	0.5
CH$_3$CH$_2$CH$_2$CH(CH$_3$)—[2]	88.2	392.0$_5$	402.7	2.7
CH$_3$C(CH$_3$)$_2$CH(CH$_3$)—[4]	102.2	393.5$_5$	396.5	0.7
CH$_3$(CH$_2$)$_2$C(CH$_3$)$_2$—[4]	102.2	396.1$_5$	406.5	2.6
CH$_3$CH(CH$_3$)CH(CH$_3$CH$_2$)—[4]	102.2	400.6$_5$	409.7	2.3
CH$_3$CH$_2$CH(CH$_3$)CH$_2$—[2]	88.2	402.0$_5$	403.0	0.2
CH$_3$CH(CH$_3$)CH$_2$CH$_2$—[2]	88.2	405.1$_5$	403.3	0.4
CH$_3$CH(CH$_3$)CH$_2$CH(CH$_3$)—[4]	102.2	406.1$_5$	410.9	1.2
(CH$_3$CH$_2$)$_2$C(CH$_3$)—[4]	102.2	409.1$_5$	406.1	0.7
CH$_3$CH$_2$C(CH$_3$)$_2$CH$_2$—[4]	102.2	409.8$_5$	407.3	0.6
CH$_3$(CH$_2$)$_4$—[2]	88.2	411.1$_5$	416.1	1.2
(CH$_3$CH(CH$_3$))$_2$CH—[5]	116.2	413.1$_5$	416.6	0.8
(CH$_3$CH$_2$)$_3$C—[5]	116.2	415.1$_5$	422.9	1.9
CH$_3$CH(CH$_3$)CH(CH$_3$)CH$_2$—[4]	102.2	418.1$_5$	411.2	1.7
CH$_3$CH$_2$CH$_2$CH(CH$_3$)CH$_2$—[4]	102.2	421.1$_5$	422.8	0.4
CH$_3$CH(CH$_3$)CH$_2$CH(CH$_3$)—[4]	102.2	432.9$_5$	410.9	5.1
CH$_3$(CH$_2$)$_2$(CH$_3$)C—[4]	102.2	434.1$_5$	406.1	6.5
(CH$_3$(CH$_2$)$_3$)(CH$_3$CH$_2$)(CH$_3$)C—[6]	130.2	436.1$_5$	438.6	0.6
CH$_3$(CH$_2$)$_6$—[5]	116.2	449.9$_5$	453.3	0.7
CH$_3$(CH$_2$)$_5$C(CH$_3$)$_2$—[7]	144.3	451.1$_5$	452.9	0.4
(CH$_3$CH$_2$CH$_2$)$_2$(CH$_3$CH$_2$)C—[7]	144.3	455.1$_5$	451.6	0.8
CH$_3$CH(CH$_3$)CH$_2$(CH$_2$)$_4$—[6]	130.2	461.1$_5$	456.0	1.1
CH$_3$(CH$_2$)$_7$—[6]	130.2	467.5$_5$	468.2	0.1

[a] Compounds with the same apex value share the same M values.

$\psi\}$, a high recommendable basis set for molecular modeling. The advantage of Z terms is that they allow an easy mixing of the modeling characteristics of χ and ψ indices that, if achieved at the level of a linear combination of χ and ψ indices with a total combinatorial technique, would oblige one to search over a huge number of combinations. The mixing achieved among basis indices, χ and/or ψ, by X, Y, and especially Z terms, reminds one vaguely of the mixing of configurations in CI (configuration interaction) methodology used in many quantum chemistry calculations. In both cases a direct physical picture gets lost at the advantage of

Table 15. Molecular Connectivity Indices of Alcohols of Table 14

alcohol − R[a]	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
$CH_3CH_2CH_2CH(CH_3)-$[1]	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	0.25	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
$CH_3CH(CH_3)CH_2CH_2-$[1]	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
$(CH_3CH_2)_2C(CH_3)-$[2]	12	5.91421	3.12132	0.25	2.84493
$CH_3CH_2C(CH_3)_2CH_2-$[2]	12	5.91421	3.12132	0.25	2.73044
$CH_3(CH_2)_4-$	10	4.82843	2.91421	0.25	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
$(CH_3CH_2)_2(CH_3)C-$[2]	12	5.91421	3.12132	0.25	2.84493
$(CH_3(CH_2)_3)(CH_3CH_2)(CH_3)C-$	16	7.32843	4.12132	0.125	4.12132
$CH_3(CH_2)_6-$	14	6.24264	3.91421	0.125	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

[a] Compounds with the same apex value share the same $\{D, {}^0\chi, {}^1\chi, \chi_t\}$ values.

Table 16. I_S Molecular Pseudoconnectivity Indices of Alcohols of Table 14[a]

alcohol − R	$^S\psi_I$	$^0\psi_I$	$^1\psi_I$	$^T\psi_I$	$^0\psi_E$	$^1\psi_E$	$^T\psi_E$
$(Me)_2CH-$	11.33	2.68957	1.58027	0.17700	1.44903	0.44003	0.01629
$EtCH_2-$	11	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	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	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
$(Pr)_2(Et)C-$	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] $Me = CH_3$; $Et = CH_3CH_2$; $Pr = CH_3CH_2CH_2$.

more precise values. Further it seems that index degeneracy seems to play an important role in modeling, as less degenerate indices or indices that remove the degeneracy are normally preferred in high-quality modeling.

Let us now summarize some achieved results. For the rates of hydrogen abstraction of chlorofluorocarbons the term X = f(χ) is a better descriptor than the term Y = f(ψ). As the X term is, here, mainly composed of valence connectivity indices, X = f(D^v, $^0\chi^v$, χ_t^v, $^0\chi$), the high degeneracy of the nonvalence indices, $\{D, {}^0\chi, {}^1\chi, \chi_t\}$, is avoided. The best overall descriptor for this property is a mixed higher-order term, Z = f(X, Y). Here, pseudoindices contribute to remove the degeneracy due to $^0\chi$ index. For the minimum anesthetic concentrations of chlorofluorocarbons the Y = f(ψ) term is a better descriptor than the term X = f($^1\chi^v$, χ_t). This last term, thanks to the $^1\chi^v$ index, has the degeneracy of the χ_t

index removed. Even for this property the best descriptor is a mixed term, $Z = f(X, Y)$.

For the activity of 2-Br-2-phentylamines χ indices shows a wide degeneracy that encompass, in some cases, even the valence χ^v indices. For the activity of this class of compounds the $Y = f(\psi)$ term is a better descriptor than the term $X = f(\chi)$, a result endorsed also by considerations on index degeneracy. Also for this activity the best overall descriptor is a $Z = f(X,Y)$ higher-order term.

With the activity of the benzimidazoles the rather degenerate molar mass, instead, turns out to be an interesting descriptor. Molecular connectivity indices show again a consistent degeneracy, which includes even a subclass of two compounds with the same values for the entire set of eight χ indices. The best overall descriptor for pK of benzimidazoles is a higher-order mixed $Z = f(X, Y')$ term, where Y' is made up of nonrescaled ψ indices. Comparison between the rescaled (benzimidazoles I) and the nonrescaled case (benzimidazoles II) tells us that the rescaling procedure does not cause dramatic changes in the modeling quality.

Primary amines show (i) a consistent degeneracy in M, (ii) which is nevertheless a good descriptor for the boiling points, (iii) an extensive collinearity among χ indices, which reduces the full $\{\chi\}$ set to $\{D, {}^0\chi, {}^1\chi, \chi_t, {}^1\chi^v\}$ set, and (iv) a small degeneracy in subset $\{D, {}^0\chi, {}^1\chi, \chi_t\}$. Their boiling points have in the term $X = f({}^1\chi, \chi_t)$ the best single descriptor and in combination $\{X_T, {}^1\chi^v\}$ the best overall descriptor, where ${}^1\chi^v$ index helps to remove the degeneracy of the two indices, ${}^1\chi$ and χ_t, of term X_T. The modeling of the boiling points of alcohols parallels, at an inferior quality level, the modeling of the boiling points of amines. The best overall descriptor is here a nondegenerate $Y = f(\psi_I, \psi_E)$ term. For the boiling points of the mixed class of amines plus alcohols the best descriptor is again a $Y = f(\psi_I, \psi_E)$ term, which can also be considered a robust term.

We would like to close this section with the recommendation that the smallest set of powerful and meaningful theoretical descriptors is $\{S_S, Z\}$, where S_S is the atom-level descriptor and $Z = f(X,Y)$ is the molecular-level descriptor. For the moment being this minimum requirement needs further investigation. In defining a minimum it should not be forgotten that "*the minimum could be defined as the perfection that an artifact achieves when it is no longer possible to improve it by subtraction. This is the quality that an object has when every component, every detail, and every junction has been reduced or condensed to the essentials. It is the result of the omission of the inessentials*".[16]

REFERENCES AND NOTES

(1) Kier, L. B.; Hall, L. H. An Electrotopological-State Index for Atoms in Molecules. *Pharm. Res.* **1990**, *7*, 801−807.
(2) Hall, L. H.; Mohney, B. K.; Kier, L. B. The Electrotopological State: Structure Information at the Atomic Level for Molecular Graphs. *J. Chem. Inf.Comput. Sci.* **1991**, *31*, 76−82.
(3) Hall, L. H.; Mohney, B. K.; Kier, L. B. Comparison of Electrotopological State Indexes with Molecular Orbital Parameters: Inhibition of MAO by Hydrazides. *Quant. Struct.-Act. Relat.* **1993**, *12*, 44−48.
(4) Kier, L. B. The Electro-topological state in QSAR and similarity studies. Lecture held by Kier, L. B. at the École de Pharmacie, BEP, UNIL, CH-1015 Lausanne-Dorigny (Suisse) on 4 Juin 1998.
(5) Kier, L. B.; Hall, L. H. *Molecular Structure Description. The Electrotopological State*; Academic Press: New York, 1999.
(6) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure−Activity Analysis*; John Wiley & Sons: New York, 1986.
(7) Randić, M. On characterization of molecular branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609−6615.
(8) Pogliani, L. Modeling with molecular pseudoconnectivity descriptors. An useful extension of the intrinsic I−Sate concept. *J. Phys. Chem.* **2000**, *104*, 9029−9045.
(9) Pogliani, L. From molecular connectivity indices to semiempirical connectivity terms. Recent trends in graph theoretical descriptors. *Chem. Rev.* **2000**, *100*, 3827−3858 (and references therein).
(10) Kier, L. B.; Hall, L. H. *Molecular Structure Description. The Electrotopological State*; Academic Press: New York, 1999; pp 170−172.
(11) Kier, L. B.; Hall, L. H. *Molecular Structure Description. The Electrotopological State*; Academic Press: New York, 1999; pp 163−164.
(12) Kier, L. B.; Hall, L. H. *Molecular Structure Description. The Electrotopological State*; Academic Press: New York, 1999; pp 90−93.
(13) Kier, L. B.; Hall, L. H. *Molecular Structure Description. The Electrotopological State*; Academic Press: New York, 1999; pp 151−152.
(14) Kier, L. B.; Hall, L. H. *Molecular Structure Description. The Electrotopological State*; Academic Press: New York, 1999; pp 176−177.
(15) Kier, L. B.; Hall, L. H. *Molecular Structure Description. The Electrotopological State*; Academic Press: New York, 1999; pp 177−178.
(16) Pawson, J. *Minimum*; Phaidon Press: London, 1996.

CI000142C