Modeling Properties with Higher-Level Molecular Connectivity Descriptors

Lionello Pogliani

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

Received April 23, 1998

The internal formal similarity of different higher-level molecular connectivity descriptors, also known as molecular connectivity terms, which can be derived by a trial-and-error procedure from a medium-sized set of eight molecular connectivity indices or a subset of it has been thoroughly analyzed. These molecular connectivity terms are interesting descriptors of a wide range of properties of biochemical but also organic and inorganic compounds: amino acids, purines, pyrimidines, alkanes, and inorganic salts. The trial-and-error procedure sometimes uncovers "dead-end" terms, that is, terms which cannot be used in multilinear combinations as they give rise to a worse description, like, e.g., the single term used to model the crystal density of amino acids and a supra term (a third-level descriptor) for the description of the solubility of the heterogeneous class of amino acids, purines, and pyrimidines. The formal similarity of many terms used to describe the different properties simplifies a lot of the otherwise rather lengthy trial-and-error procedure used to discover new meaningful higher-level descriptors. Thus, formally similar terms are capable to model, e.g., the pH at the isoelectric point, pI, the specific rotations of D- and L-amino acids in aqueous solution, the water solubility of a heterogeneous class of amino acids plus purines and pyrimidines, for which even a fourth-level descriptor can be designed, the unfrozen water content of a heterogeneous class of amino acids and inorganic salts. Further, also the motor octane number of alkanes and five different properties of the DNA−RNA bases (U, T, A, G, and C) can be modeled with formally similar second level descriptors: first and second singlet excitation energies ΔE_1 and ΔE_2, the first and second oscillator strengths, f_1 and f_2 of the first singlet excitation energy, and the molar absorption coefficient ϵ_{260}.

INTRODUCTION

A great deal of quantitative structure−property relationship (QSPR) studies are based on graph theoretical indices derived from hydrogen-suppressed chemical graphs, the molecular connectivity χ indices, defined in 1975[1−5] and further refined over the last 22 years by several authors[6−40] into a self-consistent theoretical frame known as molecular connectivity, MC, theory (or RKH theory from Randić, Kier, and Hall). It is a theory, which allows the modeling of properties and activities of compounds with quick and easy computational methods. In fact, the few mathematical tools used for this purpose are elementary and easy to grasp. Further, this method shows the characteristic of completeness, as with the recent introduction of an orthogonalization procedure, which allows for construction of highly stable and useful relations[19−21] and of a procedure, which allows the molecular connectivity relationships to revert to molecular graphs;[30,31] this theory can start from a molecule, illustrate its property or activity, and from a desired property or activity revert to the corresponding molecule. Recently, with the methodical utilization of linear combinations of molecular connectivity indices (LCCI)[41−47] it has been possible to model physicochemical properties of heterogeneous classes of compounds, to detect association phenomena in solution, and even to model properties of inorganic compounds. The very recent introduction[48−50] of second-level descriptors or molecular connectivity (composite) terms, $X = f(\chi)$, derived by the aid of a trial-and-error composition procedure based on a medium-sized set of eight molecular connectivity χ indices, which can sometimes be restricted to the subset of best χ descriptors, widened the possibility to construct new and powerful terms to be used as single descriptors or in combination with other terms or χ indices (LCXCT: linear combination of X molecular connectivity terms). It has been checked that, with a Randić's orthogonalization procedure,[19−21] the utility of these terms can be enhanced with the introduction of the corresponding orthogonal molecular connectivity terms.[48−50] Composite molecular connectivity indices had already been suggested by Kier and Hall[10] many years ago, but till now no straightforward and systematic study had been undertaken to unveil their construction and characteristics. Aim of the present paper is to deepen the study and use of second- and/or higher level molecular connectivity descriptors or terms and especially to detect if there is a way (other than the researcher's skills and his "connectivity" feelings) which can reduce the huge number of possible compositions a trial-and-error procedure with eight or even less χ indices can give rise. During the recent studies on X terms[48−50] the possibility has been put forward that these terms could share a rather similar form and also the possibility that second-level descriptors can give rise to third (supraterms) and even fourth-level descriptors (supersupraterms). Such facts, if confirmed, would drastically enhance the possibility to derive higher-level molecular connectivity descriptors.

METHOD

The trial-and-error procedure for the composition of molecular X connectivity terms is based on a $\{\chi\}$ set of eight already tested optimal[41−43,47,48] molecular connectivity indices to avoid both a severe combinatorial problem and to

10.1021/ci980054g CCC: $18.00
Published on Web 12/05/1998

Table 1. Molecular Connectivity Indices for 21 Amino Acids and Their Molar (MW) Masses

AA (MW)	D	D^v	${}^0\chi$	${}^0\chi^v$	${}^1\chi$	${}^1\chi^v$	χ_t	χ_t^v
Gly (75)	8	20	4.28446	2.63992	2.27006	1.18953	0.40825	0.03727
Ala (89)	10	22	5.15470	3.51016	2.64273	1.62709	0.33333	0.03043
Cys (121)	12	23.56	5.86181	4.55358	3.18074	2.40290	0.23570	0.02875
Ser (105)	12	28	5.86181	3.66448	3.18074	1.77422	0.23570	0.00962
Val (117)	14	26	6.73205	5.08751	3.55342	2.53777	0.19245	0.01757
Thr (119)	14	30	6.73205	4.53473	3.55342	2.21862	0.19245	0.00786
Met (149)	16	26.67	7.27602	6.14607	4.18074	4.04355	0.11785	0.01859
Pro (115)	16	28	5.98313	4.55413	3.80453	2.76688	0.08333	0.00932
Leu (131)	16	28	7.43916	5.79462	4.03658	3.02094	0.13608	0.01242
Ile (131)	16	28	7.43916	5.79462	4.09142	3.07578	0.13608	0.01242
Asn (132)	16	36	7.43916	4.70278	4.03658	2.30434	0.13608	0.00254
Asp (133)	16	38	7.43916	4.57273	4.03658	2.23927	0.13608	0.00196
Lys (146)	18	32	7.98313	5.91594	4.68074	3.36624	0.08333	0.00439
Hyp (132)	18	34	6.85337	4.87159	4.19838	2.84158	0.06804	0.00340
Gln (146)	18	38	8.14627	5.40997	4.53658	2.80434	0.09623	0.00179
Glu (147)	18	40	8.14627	5.27984	4.53658	2.73927	0.09623	0.00139
His (155)	22	42	8.26758	5.81918	5.19838	3.15529	0.03402	0.00080
Arg(174)	22	42	9.56048	6.70883	5.53658	3.60022	0.04811	0.00078
Phe (165)	24	42	8.97469	6.60402	5.69838	3.72222	0.02406	0.00069
Tyr (181)	26	48	9.84493	6.97388	6.09222	3.85651	0.01964	0.00027
Trp (204)	32	54	10.83650	8.10402	7.18154	4.71624	0.00567	0.00009

circumscribe the onerous trial-and-error procedure:

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

These indices are based on the degree δ_i of each i vertex of a molecular graph and can be computed as follows: the sum-delta D index is given[27] by

$$D = \sum \delta_i \tag{1}$$

The zeroth- and first-order indices are defined as[10] follows

$${}^0\chi = \sum (\delta_i)^{-0.5} \tag{2}$$

$${}^1\chi = \sum (\delta_i \delta_j)^{-0.5} \tag{3}$$

while the total structure χ index, over the N heteroatoms of the molecule[17] is

$$\chi_t = (\delta_1 \delta_2 \ldots \delta_N)^{-0.5} \tag{4}$$

Sums in eqs 1, 2, and 3 run over the total number of vertices (eqs 1 and 2) and edges (eq 3) of the hydrogen-suppressed molecular graph respectively (corresponding to heteroatoms and hetero-σ bonds, respectively). Replacing in eqs 1−4 δ with valence δ^v, which represent the degree of a vertex in a pseudograph with multiple edges and loops (practically, the count of all non-hydrogen electrons contributed by atom i) the corresponding four valence molecular connectivity χ^v indices are obtained. The $\delta^v(S)$ values in amino acids Cys and Met (0.56 and 0.67, respectively) as well as the δ^v definition, $\delta^v = Z^v/(Z - Z^v - 1)$, for the inorganic salts (here, metal chlorides) have been taken from Kier and Hall,[10] where Z^v is the number of valence electrons and Z is the atomic number of the corresponding atom. Recently, supraconnectivity indices have been introduced to improve the modeling of some experimental properties.[32,43] These supra indices are obtained multiplying the normal molecular connectivity indices by an association constant a, that can be derived from experimental evidence of associative phenomena or inferred from anomalous values of the physicochemical properties.

The linear estimation problem of a physicochemical property P with χ indices or $X = f(\chi)$ terms is the estimation of the experimental property P_{exp} in terms of a linear function of χ or X, which in matrix notation is given by the following dot product

$$\mathbf{P} = |\mathbf{C} \cdot \mathbf{X}| \tag{5}$$

where $\mathbf{X} = \chi$ depicts a special case and $\mathbf{P}$ is the calculated property of a compound. Row vector $\mathbf{C}$ is the vector of coefficients c_k which are determined by the linear least-square procedure, and column vector $\mathbf{X}$ is the vector of the best molecular connectivity $\mathbf{X}$ terms, derived by the aid of a combinatorial procedure. The descriptor corresponding to the constant c_0 term is the unitary index $\mathbf{X}^0 \equiv 1$. If $\mathbf{X}$ is a $m \cdot n$ matrix, with n = number of compounds, then $\mathbf{P}$ is the column vector of the properties of the entire class of compounds. Bars in eq 5 stand for absolute value to get rid of negative $\mathbf{P}$ values with no physical meaning and simultaneously enhance the description of the property.[46,48] For every term of a LCXCT equation the utility, $u_k = |c_k/s_k|$ of every index as well as the average utility $\langle u \rangle = \sum u_k/m$ of the indices of the found linear combination will be estimated.[48] Molecular connectivity terms can also be orthogonalized generating orthogonal molecular connectivity terms that (i) short-circuit the collinearity problem due to the mutual interrelation among the X terms, (ii) improve $\langle u \rangle$, (iii) generate coefficients that are stable upon introduction of a new orthogonal index, and (iv) detect dominant descriptors whenever $\mathbf{X}$ indices are poor descriptors.[19−21,34] When properties with negative and antithetical values are modeled, like the specific SR rotations of L- and D-amino acids, eq 5 has to be recast into the following form ($\cup$ stands for the logical sign "or")

$$\mathbf{P}_{L \cup D} = \mathbf{C}_{L \cup D} \cdot \mathbf{X} \quad \text{with} \quad \mathbf{C}_D = -\mathbf{C}_L \tag{6}$$

This algorithm allows to model with n SR values $2n$ compounds, here L- and D-AA.

Composition of χ indices with a trial-and-error procedure is stopped as soon as a higher-level descriptor with satisfactory values for the following statistics is found: $\langle u \rangle$, $Q = r/s$, $F = fr^2/[(1 - r^2)v]$, where r = correlation coefficient, s

Table 2. Calculated χ Values for 23 Purine and Pyrimidine Bases[a] and Their Molar Masses

PP (MW)	D	D^v	$^0\chi$	$^0\chi^v$	$^1\chi$	$^1\chi^v$	χ_t	χ_t^v
7I8MTp (250.3)	38	62	13.61036	11.38981	8.34111	5.97071	0.003564	8.51E-05
7B8MTp (250.3)	38	62	13.44723	11.22667	8.48527	6.11486	0.003086	7.37E-05
7ITp (236.3)	36	60	12.74012	10.46716	7.93043	5.53989	0.004365	9.82E-05
7BTp (236.3)	36	60	12.57699	10.30402	8.07459	5.68405	0.00378	8.51E-05
1BTb (236.3)	36	60	12.57699	10.30402	8.07459	5.68405	0.00378	8.51E-05
7PTp (222.2)	34	58	11.86988	9.59691	7.57459	5.18405	0.005346	0.00012
1PTb (222.2)	34	58	11.86988	9.59692	7.57459	5.18405	0.005346	0.00012
7ETp (208.2)	32	56	11.16277	8.88981	7.07459	4.68405	0.00756	0.00017
lETb (208.2)	32	56	11.16277	8.88981	7.07459	4.68405	0.00756	0.00017
Cf (194.2)	30	54	10.45567	8.1827	6.53658	4.10793	0.01069	0.00024
Tp (180.2)	28	52	9.58542	7.23549	6.1259	3.71758	0.013095	0.000269
Tb (180.2)	28	52	9.58542	7.23549	6.10906	3.7135	0.013095	0.000269
UA (168.1)	26	54	8.71518	5.72474	5.6647	3.11237	0.01604	0.00013
OA (156.1)	22	50	8.43072	5.24931	5.09222	2.66333	0.03928	0.00027
X (152.1)	24	48	7.84493	5.34106	5.27086	2.92873	0.01964	0.00034
IsoG (151.1)	24	46	7.84493	5.45738	5.27086	2.96049	0.01964	0.00043
G (151.1)	24	46	7.84493	5.45738	5.27086	2.96049	0.01964	0.00043
HypoX (136.1)	22	42	6.97469	4.95738	4.87701	2.74509	0.02406	0.00085
A(135.1)	22	40	6.97469	5.07369	4.87701	2.77277	0.02406	0.00108
T (126.1)	18	36	6.85337	4.89385	4.19838	2.4856	0.06804	0.00301
5MC (125.1)	18	34	6.85337	5.01016	4.19838	2.51736	0.06804	0.0038
U (112.1)	16	34	5.98313	3.9712	3.78769	2.06893	0.08333	0.00347
C (111.1)	16	32	5.98313	4.08751	3.78769	2.1007	0.08333	0.00439

[a] A = adenine, G = guanine, U = uracil, T = thymine, C = cytosine, OA = orotic acid, UA = uric acid, X = xanthine, M = methyl, P = propyl, B = butyl, I = isobutyl, Cf = caffein = 137MMMX = 7MTp, Tb = theobromine = 37MMX, Tp = theophylline = 13MMX.

Table 3. Experimental Solubility, S (at 25 °C in Units of Grams per kg of Water), Experimental pH at the Isoelectric Point (pI), Experimental Crystal CD Densities for 10 Amino Acids (AA), Experimental Specific Rotations in Angular Degrees, SR_L for L-AA and SR_D for D-AA in Water (in Parentheses, ±1 °C, When $T \neq 20$ °C), Experimental Unfrozen Water Content UWC (g H_2O/g AA) for Eight AA and Experimental Solubility, S (at the Indicated T °C, in Units of Grams per 100 mL of Water) for 23 Purines and Pyrimidines (PP)

AA	S	pI	CD	SR_L	SR_D	UWC	PP	S (T °C)
Gly	251	5.97	1.601				7I8MTp	0.63 (20)
Ala	167	6	1.401	2.7 (22)			7B8MT	0.45 (20)
Cys		5.07					7ITp	2.7 (20)
Ser	422	5.68	1.537	−6.83	6.87	0.48	7BTp	0.37 (30)
Val	58	5.96	1.230	6.42	−6.06		1BTb	0.56 (3.0)
Thr	97	5.60		28.4 (26)	−28.3	0.72	7PTp	23.11
Met	56	5.74	1.340	−8.11	8.12 (25)		1PTb	1.38 (30)
Pro	1622	6.30		−85 (23)	81.5	1.07	7ETp	3.66 (30)
Leu	23	5.98	1.165	−10.8	10.34		lETb	3.98 (30)
Ile	34	6.02		11.29	−10.55		Cf	2.58 (30)
Asn	25	5.41					Tp	0.81 (30)
Asp	5	2.77	1.660	4.7 (18)			Tb	0.054
Lys	6	9.74		14.6		0.93	UA	0.002
Hyp	361	5.8		−75.2	75.2 (21)	0.70	OA	0.18 (18)
Gln	42	5.65					X	0.05 (20)
Glu	8.6	3.22	1.538	11.5 (18)	0.97		IsoG	0.006
His	43	7.59		−39.01	39.8 (23)	0.66	G	0.004
Arg	181	10.76	1.100	12.5		0.46	HypoX	0.07 (19)
Phe	29	5.48		−35.14	35.0		A	0.09 (25)
Tyr	0.5	5.66	1.456				T	0.40 (25)
Trp	12	5.89		−31.5	32.45		5MC	0.45 (25)
							U	0.36 (25)
							C	0.77 (25)

= standard deviation of the estimate, f = degrees of freedom, v = number of χ or **X** indices. Molecular connectivity terms show in many occasions, the interesting feature to be leading terms, that is, combinations of descriptors with no X term are poor combinations, a fact which helps to restrict the combinatorial search for a multilinear description.[48−50] In fact, with X terms the case is not rare where the complete combinatorial search changes over to a forward selection search, where the search is restricted to the next best index only.

RESULTS AND DISCUSSION

The molecular connectivity indices for the different classes of compounds as well as the experimental data for the different modeled properties have already been published elsewhere[43,46,48−50] but, for the sake of clarity, are here again collected in Tables 1−6. The experimental values have been taken from the available literature.[51−55]

The Second-Level Descriptor for the pH at the Isoelectric Point, pI, of Amino Acids. The most appropriate term for this property of 21 amino acids which can be derived by the trial-and-error procedure is the one given by eq 7 (where, e.g., for $\chi = D^v \rightarrow X_{pI} \equiv {}^DX^v_{pI}$, and so on). The key to derive this term is based on considerations about the importance for pI values of the number of functional basic and/or acidic groups. In fact, the expression in parentheses with $\Delta n = n_A - n_B$, ($n_A = n°$ of acidic groups, 1 for Asp and Glu; $n_B = n°$ of basic groups, 1 for Lys and His and 2

Table 4. Unfrozen Water Content UWC (g H_2O/g MeCl) for Five Metal Chlorides and Their Corresponding Molecular Connectivity Index Values

MeCl	UWC	D	$^0\chi$	$^1\chi$	D^v	$^0\chi^v$	$^1\chi^v$	χ_t^v
LiCl	6.5	2	2	1	1.7778	2.1339	1.1339	1.1339
NaCl	3.0	2	2	1	0.8889	4.1339	3.4016	3.4017
KCl	1.8	2	2	1	0.8366	5.2570	4.6752	4.6752
$CaCl_2$	4.0	4	2.7071	1.4142	1.6732	5.1832	6.6117	3.7485
$CuCl_2$	4.0	4	2.7071	1.4142	1.6325	5.8733	8.1777	4.6357

Table 5. Experimental (exp) Molar Absorption Coefficient $\epsilon_{260,exp}$ at 260 nm and pH = 7.0, first ΔE_1 and Second ΔE_2 Singlet Excitation Energies in eV and First f_1 and Second f_2 Oscillator Strength Values (of the First Singlet Excitation Energies) of the Nucleotide DNA−RNA Bases[a]

bases	ϵ_{260}/1000	ΔE_1	ΔE_2	f_1	f_2
A	15.4	4.75	5.99	0.28	0.54
G	11.7	4.49	5.03	0.20	0.27
U	9.9	4.81	6.11	0.18	0.30
T	9.2	4.67	5.94	0.18	0.37
C	7.5	4.61	6.26	0.13	0.72

[a] A = adenine, G = guanine, U = uracil, T = thymine, C = cytosine.

Table 6. Experimental Motor Octane Numbers (MON) and Calculated Molecular Connectivity χ Indices for 30 Alkanes[a]

names	MON	D	$^0\chi$	$^1\chi$	χ_t
4	90.1	6	3.414 21	1.914 21	0.5000
2M3	97.6	8	4.284 45	2.270 05	0.4082
2M4	90.3	6	3.577 35	1.732 05	0.5774
2M5	73.5	10	4.991 56	2.770 05	0.2887
24MM6	69.9	14	6.569 81	3.663 90	0.1667
33MM5	86.6	12	5.914 21	3.121 32	0.2500
5	61.9	8	4.121 32	2.414 21	0.3536
23MM4	94.4	10	5.154 70	2.642 73	0.3333
33MM6	83.4	14	6.621 32	3.621 32	0.1768
22MM5	95.6	12	5.914 21	3.060 66	0.2500
22MM6	77.4	14	6.621 32	3.560 66	0.1768
4M7	39	14	6.405 77	3.808 06	0.1443
3M7	35	14	6.405 77	3.808 06	0.1443
3M6	55.0	12	5.698 67	3.308 06	0.2041
24MM5	83.5	12	5.861 80	3.125 89	0.2357
23MM5	88.5	12	5.861 80	3.180 73	0.2357
3E5	65.0	12	6.698 67	3.346 06	0.2041
2M6	46.4	12	5.698 67	3.270 05	0.2041
3M5	74.3	10	4.991 56	2.808 06	0.2887
23ME5	88.1	14	6.568 91	3.718 74	0.1667
223MMM5	99.9	14	6.784 45	3.481 38	0.2041
234MMM5	95.9	14	6.732 05	3.553 41	0.1925
2M7	23.8	14	6.405 77	3.770 05	0.1443
224MMM5	100.0	14	6.784 45	3.416 50	0.2041
233MMM5	99.4	14	6.784 45	3.504 03	0.2041
22MM4	93.4	10	5.207 10	2.560 66	0.3536
6	26.0	10	4.828 42	2.914 21	0.2500
25MM6	55.7	14	6.569 81	3.625 89	0.1667
7	0.0	12	5.535 53	3.414 21	0.1768
22MM3	80.2	8	5.500 00	2.000 00	0.5000

[a] Abbreviations: 2 = ethane, 3 = propane, etc.; M = methyl, E = ethyl; e.g., 34ME6 = 3-methyl-4-ethylhexane.

for Arg) and $n_T = 3$ (total number of functional groups, for $n_T = 2$, $\Delta n = 0$)[48] is by itself a fine and leading descriptor ($X_{pI} \equiv {}^0X^v$)

$$X_{pI} = \frac{\chi}{{}^0\chi^v}\left(1 + \frac{\Delta n}{n_T}\right) \tag{7}$$

$\{{}^0X^v\}_{pI}$: $Q = 2.12$, $F = 267$, $r = 0.966$, $s = 0.46$, $\langle u \rangle = 22.4$

A better Q description can be obtained with the following combination

$\{{}^0X^v, {}^0X, {}^DX^v, {}^1X\}_{pI}$: $Q = 2.53$, $F = 95$, $r = 0.980$, $s = 0.39$, $\langle u \rangle = 7.9$

The very good F value of the 1-X-LCXCT together with its excellent $\mathbf{u} = (16.3, 28.4)$ vector clearly indicates the good quality of this elementary linear combination, while analysis of the utility of each term of the 4 X combination reveals some deceiving results, with the exception of the last unitary term

$$\mathbf{X}_{pI} = ({}^DX^v, {}^0X, {}^0X^v, {}^1X, X^0),$$
$$\mathbf{u} = (3.1, 2.8, 4.7, 2.8, 26.3)$$

The modeling **C** vector being **C** = (−1.79023, 8.20212, −18.4178, 13.7802, 12.9388). To improve these utilities and detect possibly dominant descriptors use is done of the following vector of orthogonalized terms, $\Omega = ({}^1\Omega, {}^2\Omega, {}^3\Omega, {}^4\Omega, \Omega^0)$, where ${}^1\Omega \leftarrow {}^0X^v$, ${}^2\Omega \leftarrow {}^DX^v$, ${}^3\Omega \leftarrow {}^1X$, ${}^4\Omega \leftarrow {}^0X$, and $\Omega^0 \equiv 1$. The orthogonalized vector shows the following utilities: $\mathbf{u} = (19.2, 1.3, 1.0, 2.8, 33.4)$. This utility vector tells us that only the first ${}^1\Omega = {}^0X^v$ and last $\Omega^0 \equiv 1$ parameters are important descriptors, thus, we are back to the single term description but with an enhanced utility of the first and last term: 19 and 33 instead of 16 and 28. For comparison it should be noticed that the statistical Q/F score of the molar masses for this property is $Q = 0.002$ and $F = 0.14$.

The Second-Level Descriptor for the Specific SR Rotations of D- and L-Amino Acids. The second-level descriptor for the modeling of the specific rotation SR_L of $n = 16$ L-AA in aqueous solution can be derived with a trial-and-error composition procedure with the following best molecular connectivity index combination where no valence molecular connectivity indices play a significant role

$\{D, {}^0\chi, \chi_t\}$: $n = 16$, $Q = 0.088$, $F = 41.2$; $r = 0.955$, $s = 10.8$, $\langle u \rangle = 7.2$

This last combination, with $\mathbf{C}_L = -\mathbf{C}_D$, is also a relatively good descriptor for the experimental specific rotations SR_D of D-AA ($n = 11$) with $Q = 0.087$ and $F = 30.0$, while molar masses are very inadequate descriptors for this property with $Q = 0.0003$, $F = 0.12$ for L-AA and $Q = 0.005$, $F = 0.34$ for D-AA. The single best $\{\chi_t\}$ index rates, instead, rather badly with $Q = 0.014$, $F = 3.2$, $r = 0.43$, $s = 30.4$, $\langle u \rangle = 2.1$.

The derived X_{SR} has the following form, where two χ indices of the preceding best set are present

$$X_{SR}(a) = \frac{{}^0\chi}{(D + a\chi_t)} \tag{8}$$

This term, due to the presence of the flexible parameter a, shows the following description for the SR of 16 L-amino acids, with $\{X_{SR}(7)\}$ showing a better quality than $\{X_{SR}(1)\}$

$$\{X_{SR}(1)\}: \quad Q = 0.035, \quad F = 19.8, \quad r = 0.77, \quad s = 21.7, \quad \langle u \rangle = 4.8$$

$$\{X_{SR}(7)\}: \quad Q = 0.044, \quad F = 30.0, \quad r = 0.83, \quad s = 19.0, \quad \langle u \rangle = 5.6$$

Both are leading terms[49,50] which choose the same $^1\chi$ index as second best index, a fact that simplifies greatly the combinatorial search, and, actually, achieving a combinatorial search with the following two sets is possible to enhance the modeling of this property

$$\{D, D^v, {}^0\chi, {}^0\chi^v, {}^1\chi, {}^1\chi^v, \chi_t, \chi_t^v, X_{SR}(7)\} \quad \text{and} \quad \{D, D^v, {}^0\chi, {}^0\chi^v, {}^1\chi, {}^1\chi^v, \chi_t, \chi_t^v, X_{SR}(1)\}$$

The search ends up with the following optimal 2- and 3-index description for SR_L

$$\{{}^1\chi, \chi_t, X_{SR}(7)\}: \quad Q = 0.097, \quad F = 50.1, \quad r = 0.962, \quad s = 9.9, \quad \langle u \rangle = 8.1$$

$$\{{}^1\chi, X_{SR}(1)\}: \quad Q = 0.100, \quad F = 79.5, \quad r = 0.961, \quad s = 9.6, \quad \langle u \rangle = 10.6$$

Notice that the worse $X_{SR}(1)$ term achieves, when combined with χ indices, a better description than $X_{SR}(7)$ term. The gain of $\{^1\chi, X_{SR}(1)\}$ over $\{^1\chi, \chi_t, X_{SR}(7)\}$ combination is remarkable, mainly in $\langle u \rangle$ and F. Combination, $\{^1\chi, X_{SR}(1)\}$, describes the 11 experimental SR_D (for D-AA) values with $Q = 0.085$ and $F = 43.9$. To model the 16 SR_L plus 11 SR_D experimental values (see eq 6, and Table 3) the following vectors have been used, where $X^0 \equiv 1$ is the unitary vector for the constant parameter of the regression

$$\mathbf{X} = ({}^1\chi, X_{SR}, X^0), \quad \mathbf{C}_L = -\mathbf{C}_D = (26.28495, 965.8255, -545.67), \quad \mathbf{u} = (7.63, 12.34, 11.8)$$

This optimal description of 27 SR values by two nonvalence descriptors, $^1\chi$ and X_{SR}, could easily be extended to 32 SR values, as, theoretically, SR_D values $= -SR_L$ values (see eq 6).

Second-, Third-, and Fourth-Level Descriptors for the Solubility of Amino Acids, Purines, and Pyrimidines. The solubility (g per Kg of H_2O) of $n = 43$ amino acids ($n = 20$ AA) plus purines and pyrimidines ($n = 23$ PP) was the first heterogeneous class of compounds to be simulated by the aid of second level descriptors with supraconnectivity indices.[48] This former modeling was achieved with the following combination of molecular connectivity terms, where ^DX is the dominant term

$$\{{}^DX \equiv c\mathrm{D}\chi_t^v, {}^DX^v \equiv c\mathrm{D}^v\chi_t^v, X_t^v \equiv b\chi_t^v\}: \quad Q = 0.019, \quad F = 326, \quad r = 0.981, \quad s = 52.4, \quad \langle u \rangle = 6.0$$

where $c = a \cdot b$, with $b = 1$ for every amino acid and $b = a$ for purines and pyrimidines. The single term description, instead, rates

$$\{{}^DX\}: \quad Q = 0.008, \quad F = 196, \quad r = 0.909, \quad s = 108.5, \quad \langle u \rangle = 7.6$$

Molar masses are even here bad descriptors with $Q = 0.001$, $F = 2.9$. Recently,[49,50] the following family of third-level descriptor or supraterms derived with a trial-and-error composition procedure with the $\{{}^DX, {}^DX^v, X_t^v\}$ best set was capable of describing this property in a very satisfactory way

$$X_s(d) = \frac{{}^DX^v}{[1 - \exp(-dX_t^v/{}^DX^v)]} \tag{9}$$

for $d = 1$ and for $d = 70$ we have (compare with the single ^DX term description)

$$\{X_S(70)\}: \quad Q = 0.013, \quad F = 503, \quad r = 0.962, \quad s = 72, \quad \langle u \rangle = 12$$

$$\{X_S(1)\}: \quad Q = 0.012, \quad F = 415, \quad r = 0.954, \quad s = 78.3, \quad \langle u \rangle = 10.8$$

The search for an improved modeling with a combination of supraterms shows that the $X_S(70)$ is a dead-end supraterm, giving rise to no satisfactory combinations, while the following set, where $X_S(1)$ is the dominant supraterm, can further improve the modeling at the level of Q, r, and s statistics

$$\{({}^DX)^2, ({}^DX^v)^2, ({}^0X)^2, ({}^0X^v)^2, ({}^1X)^2, ({}^1X^v)^2, X_t, X_t^v, X_S(1)\}$$

with the following optimal combination of three supraterms

$$\{({}^DX)^2, ({}^0X)^2, X_S(1)\}: \quad Q = 0.021, \quad F = 402, \quad r = 0.984, \quad s = 47.4, \quad \langle u \rangle = 6.8$$

A short glance at the single utilities of this combination reveals that the utility of the unitary term is very bad

$$\mathbf{X} = (({}^DX)^2, ({}^0X)^2, X_S(1), X^0), \quad \mathbf{u} = (8.5, 8.2, 9.9, 0.7)$$

The correlation vector being $\mathbf{C} = (-7008.32, 24748.7, 14.2480, -6.03403)$. We can try to improve the utility of the unitary supraterm with the following orthogonalization[19-21] procedure $\{^1\Omega \equiv X_S(1)\ {}^2\Omega \leftarrow ({}^DX)^2, {}^3\Omega \leftarrow ({}^0X)^2, \Omega^0 \equiv 1\}$). The orthogonal supraterms show now the following utility vector $\mathbf{u} = (34, 2.3, 8.2, 4.6)$, where the main supraterm shows an outstanding utility (from 9.9 to 34), the unitary orthogonal supraterm has improved its utility more than five times (from 0.7 to 4.6), the third orthogonal supraterm (corresponding to the second supraterm) shows an unchanged good utility, and the low utility value of $^2\Omega$ indicates the minor importance of this descriptor for the modeling.

With supraterms $X_S(70)$ and $X_S(1)$ it is even possible to derive the following two different fourth-level descriptors (supersupraterms), which, both, are dead-end descriptors with a very high utility and a noticeable enhanced F value for

the second fourth-level descriptor (relatively to the corresponding supraterms)

$(X_S(70) + 0.05\cdot\ln X_S(70)\}$: $Q = 0.014$, $F = 538$, $r = 0.964$, $s = 69.5$, $\langle u\rangle = 14.2$

$(X_S(1) + 10\cdot\ln X_S(1)\}$: $Q = 0.015$, $F = 611$, $r = 0.968$, $s = 65.5$, $\langle u\rangle = 12.5$

To shed some more light on dead-end descriptors let us model the crystal densities, CD, of amino acids. Lately,[33,41] the modeling of the crystal CD densities of $n = 10$ amino acids has been satisfactorily achieved with a combination of 3, 4 χ indices, but the single index description of this property resulted to be a rather poor descriptor

$\{{}^0\chi^v\}$: $Q = 3.43$, $F = 3.9$, $r = 0.57$, $s = 0.2$, $\langle u\rangle = 5.4$

Now, a trial-and-error search for a better descriptor on the optimal set $\{D, D^v, {}^0\chi, {}^0\chi^v\}$ finds the following satisfactory molecular connectivity terms centered around the ${}^0\chi^v$ index

$$\left\{{}^1X_{CD} = \frac{{}^0\chi^v}{({}^1\chi + 1)}\right\}: \quad Q = 5.86, \quad F = 11.2, \quad r = 0.76, \quad s = 0.13, \quad \langle u\rangle = 5.3 \tag{10}$$

$$\{{}^2\chi_{CD} = {}^0\chi^v \exp(-D/{}^0\chi)\}: \quad Q = 6.38, \quad F = 13.3, \quad r = 0.79, \quad s = 0.12, \quad \langle u\rangle = 6.5 \tag{11}$$

Interestingly enough is not only the fact that they are both centered on the best single ${}^0\chi^v$ index descriptor but also the fact that none of these X_{CD} single descriptors can be used in combination with other indices or terms for an improved description. They are, thus, dead-end descriptors.

The Second-Level Descriptor for the Unfrozen Water Content UWC of Amino Acids and Inorganic Salts. Molar masses are very poor descriptors of this property with $Q = 0.44$, $F = 8.1$. To derive the second-level descriptor for this property we just start the trial-and-error search with a term whose form reminds us of the X_{SR} term. And, in fact, the search derives the following interesting family of terms

$$X_{UWC}(\chi) = \frac{{}^1\chi^v}{(\chi + a\,{}^0\chi^v)} \tag{12}$$

The optimal description is obtained for $a = -1$, and $\chi = D^v$

$\{X_{UWC}(-1, D^v)\}$: $Q = 2.79$, $F = 328$, $r = 0.984$, $s = 0.35$, $\langle u\rangle = 12.8$

The very good description achieved by this term is underlined by its good utility vector, $\mathbf{u} = (18.1, 7.8)$. The Q, r, and s statistics of this last term can further be improved, with a minimal loss in F and $\langle u\rangle$ statistics, with the following combination

$\{X_{UWC}(-1, D^v), \chi_t^v\}$: $Q = 3.11$, $F = 203$, $r = 0.988$, $s = 0.32$, $\langle u\rangle = 9.0$

Second-Level Descriptors for Five Properties of DNA/RNA Bases. Newly, molecular connectivity terms formally similar to X_{SR} or X_{UWC} have been found for three different properties of U, T, A, G, and C DNA/RNA bases: the first and second f_1, and f_2 oscillator strengths of the first singlet excitation energy and the molar absorption coefficient ϵ_{260} at 260 nm and pH = 7[49,50]

$$X_{f1}(a) = \frac{{}^1\chi^v}{({}^0\chi + a\,{}^0\chi^v)} \tag{13}$$

$$X_{f2}(b) = \frac{\chi_t^v}{(\chi_t + b\chi_t^v)} \tag{14}$$

$$X_{\epsilon}(c) = \frac{{}^1\chi^v}{({}^0\chi + c\,{}^0\chi^v)} \tag{15}$$

The resulting description for $a = 0.6$, $b = -16$, and $c = 2$, is more than decent, especially when compared to the description resulting from the molar masses (M)

f_1: $\{X_{f1}(0.6)\}$: $Q = 31$, $F = 12$, $r = 0.89$, $s = 0.03$, $\langle u\rangle = 3.1$; $Q(M) = 11$, $F(M) = 1.4$

f_2: $\{X_{f2}(-16)\}$: $Q = 11$, $F = 17$, $r = 0.92$, $s = 0.08$, $\langle u\rangle = 3.3$; $Q = 2.3$, $F(M) = 0.7$

ϵ_{260}: $\{X_{\epsilon}(2)\}$: $Q = 0.8$, $F = 21$, $r = 0.936$, $s = 1.2$, $\langle u\rangle = 4.2$; $Q(M) = 0.2$, $F(M) = 1.9$

Actually, it is possible to model even the first and second, ΔE_1, and ΔE_2, singlet excitation energies with a similar second-level descriptor. The following term, with $d = 10^3$, is in fact, a rather efficient descriptor for these two properties, especially for the second (Q and F values have been obtained with three figures for r and s)

$$X_E(d) = \left(\frac{{}^0\chi}{\chi_t + d\chi_t^v}\right)^5 \tag{16}$$

ΔE_1: $\{X_E\}$: $Q = 8.9$, $F = 5.0$, $r = 0.79$, $s = 0.1$, $\langle u\rangle = 55$; $Q(M) = 4.9$, $F(M) = 1.5$

ΔE_2: $\{X_E\}$: $Q = 6.9$, $F = 44$, $r = 0.97$, $s = 0.1$, $\langle u\rangle = 46$; $Q(M) = 3.7$, $F(M) = 12$

Practically the five properties are described by a formally similar second-level molecular connectivity descriptor; further the two energies are modeled by the same term, every term being a mixing of χ and χ^v type of indices. Every property is thus described by terms, which are δ^v dependent.

The Second-Level Descriptor for the Motor MON Octane Number of Alkanes. Molecular connectivity terms for this property of 30 alkanes have already been found,[48,50] but some of their forms are far away from the general form discussed in this paper. A search for a similar term with the $\{D, {}^0\chi, {}^1\chi, \chi_t\}$ set (alkanes do not have valence molecular connectivity indices) ends with the following term and

description

$$X_{\mathrm{MON}}(a) = \frac{{}^0\chi}{({}^0\chi + a\,{}^1\chi)} \qquad (17)$$

$\{X_{\mathrm{MON}}(-1.5)\}$: $Q = 0.079$, $F = 125$, $r = 0.904$, $s = 11.5$, $\langle u \rangle = 16.7$

The utility vector of the $\mathbf{X_{MON}} = (X_{\mathrm{MON}}(-1.5), X^0)$ vector is quite good with $\mathbf{u} = (11.2, 22.3)$. The correlation vector being $\mathbf{C} = (-10.3022, 136.687)$. A combinatorial search for a better combination finds the following rather good combination

$\{X_{\mathrm{MON}}(-1.5), D, {}^0\chi, {}^1\chi\}$: $Q = 0.121$, $F = 74$, $r = 0.960$, $s = 7.9$, $\langle u \rangle = 5.0$

CONCLUSION

A striking feature of the present study is clearly the strange similarity between many second-level descriptors used to model many different properties of different classes of compounds, either homogeneous or heterogeneous. It seems, in fact, that many molecular connectivity terms are clustered around some well defined χ compositions, a fact of no secondary importance, as the only method to derive these kinds of descriptors is the trial-and-error procedure, which formally would require a huge search all over the myriad of possible compositions a given set of basis χ indices can give rise. Thus, a closer look at terms 8, 10, 12−17 shows that they can be derived from the following common general molecular connectivity term

$$X(a, \chi_1, \chi_2, \chi_3) = \frac{\chi_1}{(\chi_2 + a\chi_3)} \qquad (18)$$

where for ${}^1X_{\mathrm{CD}}$ (eq 10) we have $a = 1$ and $\chi_3 = \chi^0 \equiv 1$. Terms for pI, X_{pI} (eq 7), for CD, ${}^2X_{\mathrm{CD}}$ (eq 11), supraterm for S, $X_{\mathrm{S}}(d)$ (eq 9), seems to fall outside this category. But, the rather similar descriptive power of ${}^1X_{\mathrm{CD}}$ (eq 10) and ${}^2X_{\mathrm{CD}}$ renders this second ${}^2X_{\mathrm{CD}}$ term unimportant, while the form of X_{pI} is not too far away from the general $X(a, \chi_1, \chi_2, \chi_3)$ form if allowance is done for $a = 0$. Even the $X_{\mathrm{S}}(d)$ supraterm can rather easily be rewritten in the given general form if the following definitions are introduced $\chi_1 = {}^{\mathrm{D}}X^{\mathrm{V}}$, $\chi_2 = \chi^0$, $a = -1$, and $\chi_3 = \exp(-\mathrm{d}X_{\mathrm{t}}^{\mathrm{v}}/\chi_1)$.

Found similar molecular connectivity X terms, derived by a trial-and-error composition procedure from a medium-sized set of molecular connectivity indices or a subset of it, are rather powerful descriptors of a wide number of properties of different homogeneous or heterogeneous classes of compounds, like amino acids, purines, pyrimidines, inorganic salts, and alkanes, and their modeling power can be further enhanced, with the construction of even higher-level descriptors, as it has been successfully attempted for the solubility S of the heterogeneous class of amino acids, purines, and pyrimidines, where even a fourth-level descriptor could be derived and/or with the introduction of the corresponding orthogonal molecular connectivity terms or supraterms, which can be considered a quite general form of higher-level descriptors. Interesting is also the fact that some higher-level descriptors seem to be dead-end descriptors, in that they do not allow for derivation of better modeling with a multilinear equation. These striking similarities among second-level descriptors of so many properties reminds us in a way of a stimulating observation from Callen's[56] masterpiece, "subtle relationships exist among apparently unconnected properties".

ACKNOWLEDGMENT

I am particularly indebted to Professor Lowell H. Hall (Eastern Nazarene College, Quincy, MA) for support and help. Thanks are also due to Dr. Subhash C. Basak and Dr. Brian D. Gute (Natural Research Resource Institute, Duluth, MN), for much advice and help before and during the first Indo-U.S. workshop. To Prof. Milan Randić (Drake University, Iowa) I owe an interesting overhead.

REFERENCES AND NOTES

(1) Randić, M. On characterization of molecular branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609−6615.
(2) Kier, L. B.; Hall, L. H.; Murray, W. J.; Randić, M. Molecular Connectivity I: relationship to nonspecific local anesthesia. *J. Pharm. Sci.* **1975**, *64*, 1971−1974.
(3) Hall, L. H.; Kier, L. B.; Murray, W. J. Molecular Connectivity II: relationship to water solubility and boilng point. *J. Pharm. Sci.* **1975**, *64*, 1974−1977.
(4) Murray, J. M.; Hall, L. H.; Kier, L. B. Molecular Connectivity III: relationship to partition coefficients. *J. Pharm. Sci.* **1975**, *64*, 1978−1981.
(5) Kier, L. B.; Hall. L. H. Molecular Connectivity VII: specific treatment of heteroatoms. *J. Pharm. Sci.* **1976**, *65*, 1806−1809.
(6) *Chemical applications of Graph theory*; Balaban, A. T., Ed.; Academic Press: London, 1976.
(7) Hall, L. H.; Kier, L. B. A molecular connectivity study of electron density in alkanes. *Tetrahedron* **1977**, *33*, 1953−1957.
(8) Kier, L. B.; Hall, L. H. Derivation and significance oof valence molecular connectivity indices. *J. Pharm. Sci.* **1981**, *70*, 583−589.
(9) Trinajstić, N. *Chemical graph theory*; CRC Press: Boca Raton, Fl, 1983; 2nd ed. 1992.
(10) Kier, L. B.; Hall, L. H. *Molecular connectivity in structure-activity analysis*; Wiley: New York, 1986; and references therein.
(11) Turro, N. J. Geometric and topological thinking in organic chemistry. *Angew. Chem., Int. Ed. Engl.* **1986**, *25*, 882−901.
(12) Seybold, P. G.; May, M.; Bagal, A. U. Molecular structure−property relationships. *J. Chem. Educ.* **1987**, *64*, 575−581.
(13) Randić, M. On characterization of three-dimensional structures. *Int. J. Quant. Chem: Quant. Biol. Symp.* **1988**, *15*, 201−208.
(14) Randić, M.; Hansen, P. J.; Jurs, P. C. Search for useful graph theoretical invariants of molecular structure. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 60−68.
(15) Basak, S. C.; Magnuson, V. R.; Niemi, G. J.; Regal, R. R. Determining structural similarity of chemicals using graph-theoretical indices. *Discr. Appl. Math.* **1988**, *19*, 17−44.
(16) Hansen, P. J.; Jurs, P. C. Chemical applications of graph theory. *J. Chem. Educ.* **1988**, *65*, 574−580.
(17) Needham, D. E.; Wei, I.-C.; Seybold, P. G. Molecular modeling of the physical properties of the alkanes. *J. Am. Chem. Soc.* **1988**, *110*, 4186−4194.
(18) Rouvray, D. H. The limits of applicability of topological indices. *J. Mol. Struct. (Theochem)* **1989**, *185*, 187−201.
(19) Randić, M. Orthogonal molecular descriptors. *N. J.Chem.* **1991**, *15*, 517−525.
(20) Randić, M. Resolution of ambiguities in structure−property studies by use of orthogonal descriptors. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 311−320.
(21) Randić, M. Search for optimal molecular descriptors. *Croat. Chim. Acta* **1991**, *64*, 43−54.
(22) Basak, S. C.; Niemi, G. J.; Veith, G. D. Predicting properties of molecules using graph invariants. *J. Math. Chem.* **1991**, *7*, 243−272.
(23) Balaban, A. T. Using real numbers as vertex invariants for third-generation topological indices. *J. Chem. Inf: Comput. Sci.* **1992**, *32*, 23−28.
(24) Maier, B. J. Wiener and Randić topological indices for graphs. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 87−90.
(25) Mihalić, Z.; Trinajstić, N. A graph-theoretical approach to structure−property relationships. *J. Chem. Educ.* **1992**, *69*, 701−712.
(26) Mihalić, Z.; Nikolić, S.; Trinajstić, N. Comparative study of molecular descriptors derived from the distance matrix. *J. Chem. Inf: Comput. Sci.* **1992**, *32*, 28−37.

(27) Pogliani, L. Molecular connectivity model for determination of isoelectric points of amino acids. *J. Pharm. Sci.* **1992**, *81*, 334−336.
(28) Pogliani, L. Molecular Connectivity: treatment of the electronic structure of amino acids. *J. Pharm. Sci.* **1992**, *81*, 967−969.
(29) Stanton, D. T.; Jurs, P. C. Computer-assisted study of the relationship between molecular structure and surface tension of organic compounds. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 109−115.
(30) Kier, L. B.; Hall, L. H.; Frazer, J. W. Design of molecules from quantitative structure−activity relationship models. 1. Information transfer between path and vertex degree counts. *J. Chem. Inf: Comput. Sci.* **1993**, *33*, 143−147.
(31) Hall, L. H.; Kier, L. B.; Frazer, J. W. Design of molecules from quantitative structure−activity realtionship models. 2. Derivation and proof of information transfer relating equations. *J. Chem. Inf: Comput. Sci.* **1993**, *33*, 148−152.
(32) Pogliani, L. Molecular connectivity model for determination of T_1 relaxation times of α-carbons of amino acids and cyclic dipeptides. *Comput. Chem.* **1993**, *17*, 283−286.
(33) Pogliani, L. Molecular connectivity model for determination of physicochemical properties of α-amino acids. *J. Phys. Chem.* **1993**, *97*, 6731−6736.
(34) Randić, M. Curve-Fitting paradox. *Int. J. Quant. Chem.: Quant. Biol. Symp.* **1994**, *21*, 215−225.
(35) Balaban, A. T.; Bertelsen, S. New centric topological indexes for acyclic molecules (trees) and substituents (rooted trees) and coding oofrooted trees. *Match* **1994**, *30*, 55−72.
(36) Basak, S. C.; Grunwald, G. D. Molecular similarity and risk assessment: analog selection and property estimation using graph invariants. *SAR & QSAR Environ. Res.* **1994**, *2*, 289−307.
(37) Pogliani, L. On a graph theoretical characterization of cis/trans isomers. *J. Chem. Inf: Comput. Sci.* **1994**, *34*, 801−804.
(38) Galvez, J.; Garcia-Domenech, R.; Julian-Ortiz, J. V.; Soler, R. Topological approach to drug design. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 272−284.
(39) Kuanar, M.; Mishra, B. K. Optimization of a regression model for a quantitative structure mutagenicity relationship of some natural amino acids. *Bull. Chem. Soc. Jpn.* **1998**, *71*, 191−198.
(40) Balaban, T. A.; Kier, L. B.; Joshi, N. Structure-Property analysis of octane numbers for hydrocarbons (Alkanes, Cycloalkanes, Alkenes). *Match (Comm. Math. Chem.)* **1992**, *28*, 13−27.
(41) Pogliani, L. Molecular connectivity descriptors of the physicochemical properties of the α-amino acids. *J. Phys. Chem.* **1994**, *98*, 1494−1499.
(42) Pogliani, L. Modeling the solubility and activity of amino acids with the LCCI method. *Amino Acids* **1995**, *9*, 217−228.
(43) Pogliani, L. Molecular modeling by linear combinations of connectivity indices. *J. Phys. Chem.* **1995**, *99*, 925−937.
(44) Lucić, B.; Nikolić, S.; Trinajstić, N.; Juretić, D.; Jurić, A. A novel QSPR approach to physicochemical properties of the α-amino acids. *Croat. Chim. Acta* **1995**, *68*, 435−450.
(45) Pogliani, L. A strategy for molecular modeling of a physicochemical property using a linear combination of connectivity indices. *Croat. Chim. Acta* **1996**, *69*, 95−109.
(46) Pogliani, L. Modeling purines and pyrimidines with the linear combination of connectivity indices - molecular connectivity "LCCI-MC" method. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1082−1091.
(47) Pogliani, L. Modeling enthalpy and hydration processes of inorganic compounds. *Croat. Chem. Acta* **1997**, *70*, 803−817.
(48) Pogliani, L. Modeling with special descriptors derived from a medium-sized set of connectivity indices. *J. Phys. Chem.* **1996**, *100*, 18065−18077.
(49) Pogliani, L. Modeling biochemicals with leading molecular connectivity terms. *Med. Chem. Res.* **1997**, *7*, 180−393.
(50) Pogliani, L. Properties of molecular connectivity terms and physicochemical properties. *J. Mol. Struct. (Theochem.)*, in press.
(51) *CRC Handbook of Chemistry and Physics*, 65th ed.; Weast, R. C., Editor-in-chief; CRC Press: Boca Raton, FL, 1984−1985.
(52) *CRC Handbook of Chemistry and Physics*, 72 ed.; Lide, D. R., Ed.; CRC Press: Boca Raton, FL, 1984−1985.
(53) Ladik, J.; Appel, K. Pariser-Parr-Pople calculations on different DNA constituents. *Theor. Chim. Acta* **1966**, *4*, 132−144.
(54) Lehninger, A. *Biochemistry*; Worth: New York, 1977.
(55) Nagashima, N.; Suzuki, E. Studies of hydration by Broad-Line pulsed NMR. *Appl. Spectr. Revs.* **1984**, *20*, 1−53.
(56) Callen, H. B. *Thermodynamics and Introduction to Thermostatics*; Wiley: New York, 1985.

CI980054G