Modeling with Molecular Pseudoconnectivity Descriptors. A Useful Extension of the Intrinsic I-State Concept

Lionello Pogliani†

Dipartimento di Chimica, Università della Calabria, 87030 Rende, Italy

Received: March 29, 2000; In Final Form: June 12, 2000

Intrinsic state molecular pseudoconnectivity indices, i.e., indices which are based on the intrinsic state concept and which are built on the intrinsic and the electrotopological state values, are used to model different properties of different classes of molecules: the side-chain molecular volume, the isoelectric point, the melting temperatures, the solubility, the specific rotations, and the crystal density of amino acids, the motor octane number and the melting temperatures of alkanes, the lattice enthalpy of metal halides, and the singlet excitation energies of DNA/RNA bases. A series of three activities are also modeled: the rates of hydrogen abstraction, the minimum anesthetic concentrations of chlorofluoroalkanes, and the antagonism of adrenalin by 2-bromo-2-phenethylamines. The modeling of the properties has been compared with the modeling achieved by the well-known molecular connectivity indices, while the modeling of the activities is compared with the modeling achieved by specific E-state indices. A comparison with the modeling power of the molar masses is also always stressed. Molecular pseudoconnectivity terms derived by a trial-and-error procedure are the best descriptors for the melting temperatures and crystal density of amino acids, both properties of alkanes, the lattice enthalpy of metal halides, the singlet excitation energies of DNA/RNA bases, the minimum anesthetic concentration, and the adrenalin antagonism. Further, a molecular pseudoconnectivity term of chlorofluoroalkanes, where subclasses of compounds share the same value of connectivity indices, is the best dominant descriptor for the rates of hydrogen abstraction. The advantage of these intrinsic state derived descriptors is rendered even more evident in the study of the activity of 2-bromo-2-phenethylamines, where many compounds show redundant connectivity and valence connectivity values. The modeling of the solubility of amino acids with pseudoconnectivity descriptors requires the introduction of supra-pseudoconnectivity descriptors, a fact that mimics a result already obtained with molecular connectivity indices. Sometimes a combination of molecular connectivity and pseudoconnectivity indices achieves a remarkable modeling.

Introduction

Recently, linear combinations of connectivity indices as well as molecular connectivity terms, which are derived from molecular connectivity indices by means of a trial-and-error procedure, proved to be good descriptors of a good deal of physicochemical properties.[1−5] Further, the recent introduction of semiempirical molecular connectivity terms, into which an empirical parameter is embedded, has proved that it is even possible to model different properties of a wide heterogeneous class of organic compounds.[6] One of the main characteristics of molecular connectivity terms is that they offer the advantage of being highly dominant single descriptors, a fact that normally allows the use of linear combinations of connectivity indices as well as of orthogonal descriptors to be short-circuited. A second practical advantage is that the construction procedure of these higher-level molecular connectivity descriptors allows the exponent of the molecular connectivity indices, which are their basic parameters, to be indirectly optimized. Clearly, molecular connectivity indices and terms represent only a subset, even if an important one, of the entire set of graph theoretical indices (refs 7−18 and references therein) used up to date.

Molecular connectivity indices and terms are directly based on the graph and pseudograph representations of a molecule;[19−21] this last representation, the pseudograph representation, is of no use for inorganic compounds,[1,22] and the achieved modeling of these compounds is performed by the aid of valence molecular connectivity whose δ^v values have heuristically been defined.[20] This last result brings us to consider the possibility to define new descriptors, which are not directly based on graph attributes, but which are based on some molecular feature that can impart to these new descriptors the characteristics of a molecular invariant. The finding of new invariants is, in fact, a task of paramount importance for a chemist as has indirectly been emphasized by W. Ostwald[23] at the beginning of this century. Kier and Hall in 1990 put forward a new molecular structural model, the electrotopological state (E state), based on an intrinsic I-state concept[24−26] by the aid of which it was possible from the atom's topological environment to incorporate the information about the influence of the remainder of the molecular environment. The *S* index of this E state, which incorporates the influence of the remainder of the molecular environment, is a composite index embracing both electronic and steric attributes of atoms and molecules, and it is directly comparable to molecular orbital-type indices, or to steric parameters. Further, it reflects the electronegativity of an atom, the electronegativity of proximal and distant atoms, and its topological state, and it shows some common features with the concept of free valence of an atom introduced by Coulson in 1948.[26] Since then this new descriptor has successfully been used in a wide variety of structure−activity studies, as well as in studies encoding molecular similarity (see references in ref 26). Index *S* is a function of index *I*, which defines the so-called intrinsic I state of an atom that is based on graph and pseudograph parameters. Now, while *I* and *S* values are atom-

† Fax: x39-984-492044. E-mail: lionp@unical.it.

10.1021/jp001191v CCC: $19.00
Published on Web 09/13/2000

centered values even if the last one takes into consideration the whole molecular framework, it is not at all odd to define new whole molecular indices based on them in a similar way as χ molecular connectivity indices are directly grounded on the connectivity degree of a graph (δ) and of a pseudograph (δ^v). These new indices, as they rotate around the intrinsic I concept and are indirectly based on connectivity δ and δ^v numbers, may be called intrinsic state, I_S, molecular pseudoconnectivity indices or just molecular pseudoconnectivity indices and denoted the Greek letter ψ. They will consist of two subtypes of indices: $\{\psi\} = \{\psi_I, \psi_E\}$, where ψ_I is the subset of the *I*-derived indices and ψ_E is the subset of the *S*-derived indices. Their indirect relation to topological characteristics together with their electronic features should make them interesting descriptors of physicochemical properties and activities. Thus, either as single descriptors or as linear combinations of pseudoconnectivity indices (LCpCI) or as molecular pseudoconnectivity terms, $Y = f(\psi)$, they should achieve some sort of modeling. In the present study we will attempt to model with ψ or ψ-derived descriptors different properties of amino acids, alkanes, inorganic salts, and DNA/RNA bases and to compare the achieved modeling with the modeling power of molecular connectivity descriptors and of the molar masses. The activities of chlorofluoroalkanes and 2-bromo-2-phenethylamines will also be modeled as they offer a good example of classes of compounds where a lot of molecules show the same values of connectivity indices, ${}^i\chi$, and also (the second class) of valence connectivity indices, ${}^i\chi^v$.

Method

The electronic S_i E-state index for atom *i* proposed by Kier and Hall[24−26] to describe quantitatively the mutual influence of non-hydrogen atoms in a molecule can be estimated by the following formula:

$$S_i = I_i + \sum_j \Delta I_{ij} \tag{1}$$

where $I = [(2/N)^2\delta^v + 1]/\delta$, 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. Because some *S* values are negative (in amino acids, inorganic salts, chlorofluoroalkanes, and 2-bromo-2-phenethylamines), calculation of some ψ_E values could give rise to imaginary numbers. To avoid this occurrence, *S* values have been rescaled. Buried carbon atoms bonded to highly electronegative atoms give rise to negative *S* values. It is, then, not at all odd to rescale every S_i value to the *S* value of the carbon atom in CF_4, which equals −5.5, and which is the lowest *S* value a carbon atom can assume. Inevitably, this rescaling invalidates one of the results of the electrotopological model, which states that in a molecule $\sum_i S_i = \sum_i I_i$. In alkanes, where such a rescaling has not been undertaken, as no negative *S* values are obtained, the values of the two indices, ${}^S\psi_I$ and ${}^S\psi_E$ are in fact equal (see the following equations). As already done with preceding χ-modeling the following set of eight I_S molecular pseudoconnectivity indices will be used to model the given properties and activities of organic and inorganic compounds:

$$\{\psi\} = \{{}^S\psi_I, {}^0\psi_I, {}^1\psi_I, {}^T\psi_I, {}^S\psi_E, {}^0\psi_E, {}^1\psi_E, {}^T\psi_E\} \tag{2}$$

Their definition parallels the original definition of χ indices (refs 5 and 20 and references therein)

$${}^S\psi_I = \sum_i I_i \tag{3}$$

$${}^0\psi_I = \sum_i (I_i)^{-0.5} \tag{4}$$

$${}^1\psi_I = \sum (I_i I_j)^{-0.5} \tag{5}$$

$${}^T\psi_I = (I_1 \cdot I_2 \cdot I_3 \ldots I_N)^{-0.5} \tag{6}$$

The sums in eqs 3 and 4 are taken over all atoms, while the sum in eq 5 is over all edges (σ bonds) of the molecular graph, respectively. Replacing *I* in eqs 3−6 with *S*, the subset $\{{}^S\psi_E, {}^0\psi_E, {}^1\psi_E, {}^T\psi_E\}$ of molecular pseudoconnectivity indices is obtained. Superscripts S and T stand for sum and total; the other superscripts follow the established denomination for χ indices.[20] Equations 3 and 5 deserve special attention. The index defined by eq 3 had already been proposed to describe molecular polarity, while the index of eq 5 strongly mimics the sum of the bond-E-state index, $BI_S = (I_i I_j)^{0.5}$, where the exponent is positive instead of negative.[27]

To avoid confusion, denomination ϵ for *S*-derived values (and correspondingly ι for *I*-derived values) has not been chosen as symbol as ϵ has already been used for edge-connectivity indices.[28,29] Further, the proposed name for I_S molecular pseudoconnectivity indices, especially the portion "pseudoconnectivity", has been chosen to avoid renaming the normal connectivity χ indices. In fact, naming them "indirect" molecular connectivity indices would practically have forced renaming χ indices as direct molecular connectivity indices, while to name them molecular electrotopological connectivity indices would have required redefining the meaning of connectivity. Anyway, the question of their names stays open and will surely be solved in the near future.

Assuming that the relationship between properties, *P*, and molecular pseudoconnectivity indices is linear, the modeling equation is given by the following dot product modulus: $P = |\mathbf{C}\cdot\boldsymbol{\psi}|$, where *P* is the calculated property of a compound, row vector $\mathbf{C}$ is the vector of the c_k coefficients that are determined by the least-squares procedure, and column vector ψ is the vector of the best pseudoconnectivity descriptors selected with a total combinatorial search technique and/or with a trial-and-error procedure (for terms only). The multivariate regression can be regarded as a linear combination of pseudoconnectivity indices (LCpCI) where the constant term of the regression can be considered to multiply the unitary index, $U_0 \equiv 1$, a kind of bias index. Even though *P* is not always a linear function of ψ, it is nevertheless a linear function of the c_k coefficients. If ψ is an $m \times n$ matrix (where n = number of compounds), then $\mathbf{P}$ is a property column vector of the entire class of compounds. Bars in the modeling equation stand for absolute value to get rid of calculated negative *P* values with no physical meaning and simultaneously enhance the description, provided that the corresponding experimental property is positive. The statistical performance of the different LCpCI, which can be obtained with a combinatorial procedure, is controlled by a quality, $Q = r/s$, factor, where r = correlation coefficient and s = standard deviation of estimates, by the variance *F* (Fischer ratio), $F = fr^2/[(1 - r^2)\nu]$, where f = number of freedom degrees = $n - (\nu + 1)$, ν = number of variables, and n = number of data. The parameter *Q*, which is an "internal" statistic, is apt to compare the descriptive power of different descriptors of the same property, while the parameter *F* tells us, even if *Q* improves, which additional descriptor endangers the statistical quality of the combination. For every index of a LCpCI equation

the fractional utility, $u_k = |c_k/s_k|$, as well as the average fractional utility $\langle u \rangle = \sum u_k/(\nu + 1)$ will be given. This statistical parameter will allow us to detect the paradoxical situation of a LCpCI with a good predictive power but with a poor utility at the level of its coefficients.[30] It should be noticed that Q, and r, and s values as well as $\langle u \rangle$ and **u** values, even if some of them may seem redundant, in their totality offer, a direct view of the statistical behavior of a modeling and can also be used as a check for eventual printing errors both of the author and of the journal, as two nearby printing errors is a rather rare event.

When properties of some members of a class of compounds assume negative values, the modeling equation should be used without modulus bars, $P = \mathbf{C}\cdot\boldsymbol{\psi}$. This is the case for the specific rotation (SR) of amino acids. Further, as the specific rotation (SR) of amino acids can assume negative and antithetic values for the L- and D-forms, then the correlation vectors for the L- and D-forms are related through the relation $\mathbf{C}_L = -\mathbf{C}_D$; this means that once a subset is modeled, the modeling of the other subset is straightforward.

Usually, indices used in linear combinations, LCpCI, are interrelated, a fact that has some negative effects as (i) it results in unstable estimated regression coefficients of vector **C**, (ii) it may render values predicted for compounds not in the original training data set not reliable, (iii) it may also render an analysis of the relative importance of an index in a modeling equation a useless task, and, finally, (iv) it may worsen the utilities of the regression coefficients. The construction of dominant pseudoconnectivity terms, $Y = f(\psi)$, with a trial-and-error procedure offers the possibility to reduce the modeling equation to a simple linear form, $P = |c\cdot Y + c_0U_0|$, and to short-circuit, thus, the orthogonality problem. The trial-and-error procedure to construct molecular connectivity or pseudoconnectivity terms runs in this way: (i) optimize the first index, ${}^1\psi$, (ii) given ${}^1\psi$ introduce and optimize ${}^2\psi$ (the second index), (iii) back-optimize ${}^1\psi$, (iv) check that with the new ${}^1\psi$ index, the ${}^2\psi$ index continues to be the best index, and so on. This procedure can be schematized, for the case of four parameters, by the aid of the following symbolism, where I stands for introduce, O for optimization, and C for check operations. Normally, optimization is optimal at the check point. The ratio can be built at every level, if the introduction of an index does not improve any more the description at that level:

O(1)

I(2), O(2|1), O(1|2), C(2|1)

I(3), O(3|1,2), O(2|3,1), O(1|2,3), C(3|1,2)

I(4), O(4|1,2,3), O(3|4,1,2), O(2|3,4,1), O(1|2,3,4), C(4|1,2,3)

Usually, this procedure either converges rapidly or does not work at all, and the found terms regularly have the following form: $Y = [(\psi_1)^m + a(\psi_2)^n]^q/[b(\psi_3)^o + c(\psi_1)^p]^r$, where a, b, c, m, n, o, p, q, and r are optimization parameters that can also be negative, zero, or one.

Results

The experimental values of the physicochemical properties of amino acids and alkanes are collected in Tables 1 and 2, respectively. The lattice enthalpy values of metal halides are collected in the second column of Table 8, and the first and second singlet excitation energies of DNA/RNA bases are collected in Table 10. The rates of hydrogen abstraction and minimum anesthetic concentrations of chlorofluoroalkanes, in log units, are collected in Tables 12 and 14, respectively. The measure of biological response of 2-bromo-2-phenethylamines in pED_{50} units, i.e., the negative logarithm of the estimated dose for 50% response, is given in Table 17. While the experimental values of Tables 12, 14, and 17 have been taken from ref 26, the other experimental values have been taken from the cited papers of the author. I_i and S_i values of $\mathbf{I_S}$ and $\mathbf{E_S}$ matrices and vectors of Tables 3, 5, 7, 9, 11, and 15 have been obtained by the aid of the E-calc software provided in a CD of a cited book.[26] The indices ψ_I and ψ_E of Tables 4, 6, 8, 10, 12, 13, and 16 have, instead, been obtained by the aid of a Turbo Basic software program written by the author. The meaning of columns P_{calc} and P_{jk} of Tables 1, 2, 8, and 14 will be explained in the following sections. Amino acids, as well as DNA/RNA bases, have been represented by two-row $\mathbf{I_S}$ and $\mathbf{E_S}$ matrices, a representation already used in a preceding paper where two-row χ matrices have also been used[1,31] for amino acids and purine and pyrimidine bases. In these matrices heteroatoms are substituted by their corresponding I values, in $\mathbf{I_S}$ matrices, and S values, in $\mathbf{E_S}$ matrices. This type of representation is also used for alkanes, which give rise in some cases to three-row matrices, chlorofluoroalkanes, 2-Br-2-phenethylamines, and inorganic salts, which give rise to one-row matrices or vectors. In cyclic amino acids of Table 3 (Pro, Hyp, His, Phe, Tyr, and Trp) the ring closure point is represented by an underlined value which connects to the corresponding values denoted by a left broken arrow. Notice that in Trp there are two ring closure points. The same "strategy" was used in Table 9 for the bases A and G, while in bases U, T, and C the ring closure is between the value on top of the "→" sign and the value on top of the left broken arrow sign. This last strategy has also been used with 2-bromo-2-phenethylamines. In alkanes, and chlorofluoroalkanes (CFC), instead, the problem of double methyl (M) or double halogen substitution at the same carbon was solved, underlining the corresponding I or S value, which means that the value should be reported in a virtual third row on the top of the first one. Single ethyl (E) substitution was here solved by adding a third line, while ME or EE substitution at the same carbon atom was solved, in the first case, by placing the corresponding value of M in parentheses, and, in the second case, by underlining both ethyl I and S values. A rapid glance at these $\mathbf{I_S}$ and $\mathbf{E_S}$ matrices or vectors offers some interesting insights into the meaning of the electrotopological state: (i) while some compounds show similar $\mathbf{I_S}$ matrices, no compound has the same $\mathbf{E_S}$ matrix, (ii) single values in $\mathbf{E_S}$ matrices seem to reflect electronegativity considerations, with the most negatively charged atom showing the highest S value, and (iii) the trend of carbon S values in amino acid and alkane $\mathbf{E_S}$ matrices seems to reflect the NMR chemical shift trend of the corresponding carbon atoms,[32] a fact already noticed.[26] It would be interesting to notice if different S_i values for the same type of carbon in different amino acids or alkanes are also mapped by different chemical shifts and in the same order.

Discussion

Amino Acids. The average interrelation value of the $\{\psi\}$ indices for amino acids is $\langle r \rangle = 0.818$, while the strongest and weakest interrelations are $r({}^0\psi_I, {}^1\psi_I) = 0.992 = r({}^0\psi_I, {}^0\psi_E)$ and $r({}^S\psi_I, {}^T\psi_E) = 0.568$, respectively. Before entering the modeling of the different properties of amino acids let us notice that the $\{{}^0\psi_E\}$ index is the best descriptor for the molar masses, M, of amino acids with $Q = 0.159$, $F = 459$, $r = 0.980$, and $s = 6.2$, while $\{{}^0\chi\}$ had $Q = 0.156$, $F = 441$, $r = 0.979$, and $s = 6.3$.

TABLE 1: Experimental Properties of Amino Acids, AAs (M = Molar Mass): Solubility, S (at 25 °C in Units of g/kg of water), the pH at the Isoelectric Point, pI, Melting Temperature, T_m (°C), Crystal Density, CD, Side-Chain Molecular Volume, V (Å^3), and Specific Rotations in Angular Degrees, SR_L, in Water[a]

AA (M)	S	P_{calc}	P_{jk}	pI	T_m	CD	V	SR_L
Gly (75)	251	244	243	5.97	290	1.601	36.3	
Ala (89)	167	149	148	6	297	1.401	52.6	2.7 (22)
Cys (121)				5.07	178			
Ser (105)	422	412	412	5.68	228	1.537	54.9	−6.83
Val (117)	58	59	59	5.96	293.5	1.230	85.1	6.42
Thr (119)	97	84	83	5.60	253		71.2	28.4 (26)
Met (149)	56	25	23	5.74	283	1.340		−8.11 (25)
Pro (115)	1622	1636	1793	6.30	222		73.6	−85 (23)
Leu (131)	23	35	36	5.98	337	1.165	102	−10.8 (25)
Ile (131)	34	35	35	6.02	284		102	11.29
Asn (132)	25	75	78	5.41	236		72.4	
Asp (133)	5	82	86	2.77	270	1.660	68.4	4.7 (18)
Lys (146)	6	31	33	9.74	224.5		105.1	14.6
Hyp (132)	361	276	271	5.8				−75.2 (23)
Gln(146)	42	47	48	5.65	185		92.7	
Glu (147)	8.6	52	55	3.22	249	1.538	84.7	11.5 (18)
His (155)	43	34	34	7.59	277		91.1	−39.01 (25)
Arg (174)	181	162	161	10.76	238	1.100	109.1	12.5
Phe (165)	29	14	13	5.48	284		113.9	−35.14
Tyr (181)	0.5	5.6	5.9	5.66	344	1.456	116.2	
Trp (204)	12	15	17	5.89	282		135.4	−31.5 (23)

[a] In parentheses are reported $T \neq 20 \pm 1$ °C. In the third and fourth columns are the calculated solubility P_{calc} values and calculated jackknifing solubility P_{jk} values.

TABLE 2: Experimental Melting Points, MPs (K), for 17 Alkanes, Motor Octane Numbers, MONs, Calculated MON Values (P_{Calc}), and Calculated Jackknifing MON Values (P_{jk}) for 30 Alkanes[a]

alkane	MP	MON	P_{calc}	P_{jk}	alkane	MP	MON	P_{calc}	P_{jk}
4		90.1	83.4	83.2					
2M3		97.6	99.0	99.1	3M5		74.3	76.7	76.8
2M4		90.3	89.3	89.2	23ME5	158.2	88.1	72.3	71.7
2M5		73.5	76.7	76.8	223MMM5		99.9	91.3	90.8
24MM6	135.7	69.9	72.3	72.4	234MMM5		95.9	87.3	86.9
33MM5	138.7	86.6	88.0	88.0	2M7		23.8	36.8	38.3
5		61.9	63.9	63.9	224MMM5		100.0	91.3	90.8
23MM4	144.6	94.4	92.8	92.7	233MMM5		99.4	91.3	90.8
33MM6	147.1	83.4	78.3	79.1	22MM4	173.3	93.4	95.8	96.0
22MM5	149.3	95.6	88.0	87.6	6		26.0	34.6	35.7
22MM6	152.0	77.4	78.3	78.3	25MM6	182.0	55.7	72.3	72.9
4M7		39	36.8	36.6	7		0.0	12.7	22.5
3M7		35	36.8	37.0	23MM7	157.2			
3M6		55.0	59.9	60.1	22MM7	160.2			
24MM5	153.9	83.5	83.7	83.7	26MM7	170.3			
23MM5	154.1	88.5	83.7	83.5	33ME5	182.3			
3E5		65.0	59.9	59.7	33EE5	240.0			
2M6		46.4	59.9	60.5	22MM3	256.6	80.2	102.4	104.4

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

This result indicates that ψ indices should at least be decent descriptors of M-dependent properties.

Side-Chain Molecular Volume. Let us start modeling the side-chain molecular volume, V, of $n = 18$ amino acids, shown in Table 2 (no Met, Cys, and Hyp). The best χ index for this property[1,31] is ${}^0\chi^v$ with a remarkable statistical score

$$\{{}^0\chi^v\}: Q = 0.25,\ F = 691,\ r = 0.989,\ s = 3.9_5,\ \langle u\rangle = 15,\ \mathbf{u} = (26, 3.3)$$

For comparison purposes, the molar masses, M, describe this property with $Q = 0.069$ and $F = 83$. The best multi-χ index LCCI is given by the following two-χ index combination:[31]

$$\{D^v, {}^0\chi^v\}: Q = 0.40,\ F = 887,\ r = 0.996,\ s = 2.5,\ \langle u\rangle = 12,\ \mathbf{u} = (5.1, 28, 4.0)$$

where the last value is the utility of the bias index, $U_0 \equiv 1$. Molecular pseudoconnectivity indices of the set of eq 2 do not achieve the same good statistical score

$$\{{}^0\psi_I\}: Q = 0.123,\ F = 140,\ r = 0.947,\ s = 8.4,\ \langle u\rangle = 6.2,\ \mathbf{u} = (12, 0.7)$$

$$\{{}^0\psi_I,\ {}^1\psi_E\}: Q = 0.140,\ F = 106,\ r = 0.966,\ s = 6.6,\ \langle u\rangle = 3.1,\ \mathbf{u} = (5.4, 2.9, 1.0)$$

Further, the utilities are not at all impressive, and especially, the utility of the unitary index U_0 shows that both regressions, the single-index and the two-index regressions, are useless. A trial-and-error search for the best molecular pseudoconnectivity term, $Y_V = f(\psi)$, for this property discovers the following interesting term:

$$Y_V = \frac{({}^0\psi_I)^{0.35} - 0.8({}^1\psi_E)^{0.35}}{(1\psi_I)^{0.06}} \qquad (7)$$

This term rates $Q = 0.176$, $F = 341$, $r = 0.977$, $s = 5.6$, $\langle u\rangle =$

TABLE 3: Intrinsic State Value Matrices, I_S, and the Electrotopological E_S Matrices of 21 Amino Acids[a]

AA	I_S matrices	E_S matrices
Gly	$\begin{pmatrix} 1.5 & 1.67 & 7 \\ 4 & 6 & 0 \end{pmatrix}$	$\begin{pmatrix} -0.28 & -0.97 & 9.24 \\ 4.57 & 7.60 & 0 \end{pmatrix}$
Ala	$\begin{pmatrix} 2 & 1.33 & 1.67 & 7 \\ 0 & 4 & 6 & 0 \end{pmatrix}$	$\begin{pmatrix} 1.42 & -0.73 & -0.96 & 9.57 \\ 0 & 4.84 & 7.87 & 0 \end{pmatrix}$
Cys	$\begin{pmatrix} 3.22 & 1.5 & 1.33 & 1.67 & 7 \\ 0 & 0 & 4 & 6 & 0 \end{pmatrix}$	$\begin{pmatrix} 3.65 & 0.19 & -0.82 & -1.00 & 9.76 \\ 0 & 0 & 4.94 & 8.01 & 0 \end{pmatrix}$
Ser	$\begin{pmatrix} 6 & 1.5 & 1.33 & 1.67 & 7 \\ 0 & 0 & 4 & 6 & 0 \end{pmatrix}$	$\begin{pmatrix} 8.00 & -0.50 & -1.13 & -1.18 & 9.65 \\ 0 & 0 & 4.77 & 7.90 & 0 \end{pmatrix}$
Val	$\begin{pmatrix} 2 & 1.33 & 1.33 & 1.67 & 7 \\ 0 & 2 & 4 & 6 & 0 \end{pmatrix}$	$\begin{pmatrix} 1.78 & 0.02 & -0.71 & -0.93 & 10.02 \\ 0 & 1.78 & 5.16 & 8.23 & 0 \end{pmatrix}$
Thr	$\begin{pmatrix} 6 & 1.33 & 1.33 & 1.67 & 7 \\ 0 & 2 & 4 & 6 & 0 \end{pmatrix}$	$\begin{pmatrix} 8.49 & -0.98 & -1.16 & -1.18 & 9.86 \\ 0 & 1.33 & 4.91 & 8.07 & 0 \end{pmatrix}$
Met*	$\begin{pmatrix} 2 & 1.83 & 1.5 & 1.5_H \\ 0 & 0 & 0 & 0_H \end{pmatrix}$	$\begin{pmatrix} 1.93 & 1.60 & 0.81 & 0.55 & -0.68 & -0.91 & 10.07 \\ 0 & 0 & 0 & 0 & 5.19 & 8.27 & 0 \end{pmatrix}$
Leu	$\begin{pmatrix} 2 & 1.33 & 1.5_H \\ 0 & 2 & 0_H \end{pmatrix}$	$\begin{pmatrix} 1.95 & 0.36 & 0.55 & -0.69 & -0.91 & 10.11 \\ 0 & 1.95 & 0 & 5.22 & 8.31 & 0 \end{pmatrix}$
Ile	$\begin{pmatrix} 2 & 1.5 & 1.33_H \\ 0 & 0 & 2_H \end{pmatrix}$	$\begin{pmatrix} 1.92 & 0.81 & 0.072 & -0.70 & -0.91 & 10.17 \\ 0 & 0 & 1.83 & 5.27 & 8.36 & 0 \end{pmatrix}$
Asn	$\begin{pmatrix} 4 & 1.67 & 1.5_H \\ 0 & 7 & 0_H \end{pmatrix}$	$\begin{pmatrix} 4.65 & -0.70 & -0.31 & -1.16 & -1.21 & 9.90 \\ 0 & 9.99 & 0 & 4.92 & 8.10 & 0 \end{pmatrix}$
Asp	$\begin{pmatrix} 7 & 1.67 & 1.5_H \\ 0 & 6 & 0_H \end{pmatrix}$	$\begin{pmatrix} 9.77 & -1.20 & -0.53 & -1.29 & -1.29 & 9.85 \\ 0 & 7.99 & 0 & 4.84 & 8.04 & 0 \end{pmatrix}$
Gln	$\begin{pmatrix} 4 & 1.67 & 1.5 & 1.5_H \\ 0 & 7 & 0 & 0_H \end{pmatrix}$	$\begin{pmatrix} 4.75 & -0.53 & 0.021 & 0.10 & -0.98 & -1.11 & 10.03 \\ 0 & 10.10 & 0 & 0 & 5.06 & 8.22 & 0 \end{pmatrix}$
Glu	$\begin{pmatrix} 7 & 1.67 & 1.5 & 1.5_H \\ 0 & 6 & 0 & 0_H \end{pmatrix}$	$\begin{pmatrix} 9.88 & -1.03 & -0.20 & -0.02 & -1.06 & -1.17 & 9.99 \\ 0 & 8.10 & 0 & 0 & 5.00 & 8.18 & 0 \end{pmatrix}$
Lys	$\begin{pmatrix} 4 & 1.5 & 1.5 & 1.5 & 1.5_H \\ 0 & 0 & 0 & 0 & 0_H \end{pmatrix}$	$\begin{pmatrix} 5.20 & 0.60 & 0.85 & 0.80 & 0.52 & -0.72 & -0.93 & 10.14 \\ 0 & 0 & 0 & 0 & 0 & 5.22 & 8.33 & 0 \end{pmatrix}$
Arg	$\begin{pmatrix} 4 & 1.67 & 2.5 & 1.5 & 1.5 & 1.5_H \\ 0 & 4 & 0 & 0 & 0 & 0_H \end{pmatrix}$	$\begin{pmatrix} 5.10 & 0.14 & 2.66 & 0.55 & 0.63 & 0.41 & -0.80 & -0.99 & 10.23 \\ 0 & 5.10 & 0 & 0 & 0 & 0 & 5.24 & 8.39 & 0 \end{pmatrix}$
Pro	$\begin{pmatrix} 1.5 & 1.5 & 1.5 & 1.33 & 1.67 & 7 \\ \underline{2.5} & 0 & 0 & \lrcorner & 6 & 0 \end{pmatrix}$	$\begin{pmatrix} 0.86 & 1.00 & 0.79 & -0.27 & -0.72 & 10.14 \\ \underline{2.86} & 0 & 0 & \lrcorner & 8.35 & 0 \end{pmatrix}$
Hyp	$\begin{pmatrix} 1.5 & 1.33 & 1.5 & 1.33 & 1.67 & 7 \\ \underline{2.5} & 6 & 0 & \lrcorner & 6 & 0 \end{pmatrix}$	$\begin{pmatrix} 0.40 & -0.48 & 0.33 & -0.54 & -0.88 & 10.17 \\ \underline{2.66} & 8.82 & 0 & \lrcorner & 8.36 & 0 \end{pmatrix}$
His	$\begin{pmatrix} 2.5 & 2 & 2.5 & 1.67 & 1.5_H \\ \underline{2} & 0 & 0 & \lrcorner & 0_H \end{pmatrix}$	$\begin{pmatrix} 2.77 & 1.62 & 2.82 & 0.79 & 0.32 & -0.83 & -0.99 & 10.28 \\ \underline{1.68} & 0 & 0 & \lrcorner & 0 & 5.27 & 8.43 & 0 \end{pmatrix}$
Phe	$\begin{pmatrix} 2 & 2 & 2 & 2 & 1.67 & 1.5_H \\ \underline{2} & 0 & 0 & 0 & \lrcorner & 0_H \end{pmatrix}$	$\begin{pmatrix} 1.86 & 1.88 & 1.87 & 1.86 & 0.95 & 0.39 & -0.80 & -0.96 & 10.38 \\ \underline{1.86} & 0 & 0 & 0 & \lrcorner & 0 & 5.35 & 8.52 & 0 \end{pmatrix}$
Tyr	$\begin{pmatrix} 2 & 1.67 & 2 & 2 & 1.67 & 1.5_H \\ \underline{2} & 6 & 0 & 0 & \lrcorner & 0_H \end{pmatrix}$	$\begin{pmatrix} 1.50 & 0.16 & 1.50 & 1.65 & 0.80 & 0.27 & -0.88 & -1.02 & 10.40 \\ \underline{1.65} & 8.95 & 0 & 0 & \lrcorner & 0 & 5.32 & 8.52 & 0 \end{pmatrix}$
Trp	$\mathbf{I_S} = \begin{pmatrix} 2 & 2 & 2 & 2 & 1.67 & 2.5 & 2 & 1.67 & 1.5_H \\ \underline{1.67} & 0 & 0 & 0 & \lrcorner & 0 & 0 & \lrcorner & 0_H \end{pmatrix}$	
Trp	$\mathbf{E_S} = \begin{pmatrix} 1.95 & 1.92 & 1.93 & 1.95 & 1.00 & 3.08 & 1.81 & 0.94 & 0.35 & -0.84 & -0.97 & 10.63 \\ \underline{1.04} & 0 & 0 & 0 & \lrcorner & 0 & 0 & \lrcorner & 0 & 5.48 & 8.72 & 0 \end{pmatrix}$	

[a] For the sake of being brief and for space the letter *H* stands for "head" of the amino acid, and it corresponds to the last three columns (from the left) in the amino acid Ala. For a general explanation of these matrices see the Results section.

16, and $\mathbf{u} = (19, 13)$. Not only the good score of the single ${}^0\chi^v$ continues to be unsurpassed, but if the score of the best molecular connectivity term of eq 8 is considered, $Q = 0.438$,

$$X_V = \frac{(D^v)^{1.3} + ({}^0\chi)^{2.1}}{D^v - 0.7D} \tag{8}$$

$F = 2109$, $r = 0.996$, $s = 2.3$, and $\mathbf{u} = (46, 17)$, then the Y_V term has no chance to compete with χ-derived descriptors at any statistical level. Both terms are "dead-end" terms; i.e., it is not possible to improve the description with any linear combination of terms or of terms and indices. In this respect, from now on, it will be assumed that every term is a dead-end term, unless it is used in an improved linear combination with other descriptors. If a mixed set composed of eight molecular connectivity indices, $\{D, D^v, {}^0\chi, {}^0\chi^v, {}^1\chi, {}^1\chi^v, \chi_t, \chi_t{}^v\}$, and of eight molecular pseudoconnectivity indices (see eq 2) is considered, then a very interesting modeling is achieved, quite

TABLE 4: I_S Molecular Pseudoconnectivity Indices for 21 Amino Acids[a]

AA	${}^{S}\psi_I$	${}^{0}\psi_I$	${}^{1}\psi_I$	${}^{T}\psi_I$	${}^{S}\psi_E$	${}^{0}\psi_E$	${}^{1}\psi_E$	${}^{T}\psi_E$
Gly	20.17	2.87653	1.64846	0.04876	55.59	1.79888	0.61430	0.00533
Ala	22.00	3.63425	2.32607	0.03661	55.01	2.14940	0.78058	0.00179
Cys	24.72	4.30090	2.87594	0.02356	63.23	2.52101	0.94227	0.00065
Ser	27.50	4.15189	2.75426	0.01726	66.02	2.52046	0.96944	0.00061
Val	25.33	5.20847	3.69109	0.02244	69.35	2.92156	1.10863	0.00026
Thr	29.33	4.90961	3.43195	0.01296	73.34	2.91805	1.16460	0.00024
Met	26.83	6.00647	4.23467	0.01804	76.37	3.29648	1.22896	0.00001
Pro	23.00	5.50909	4.38551	0.03564	76.01	2.94957	1.21321	0.00029
Leu	26.83	6.02497	4.35520	0.01833	76.36	3.30240	1.25155	0.00010
Ile	26.83	6.02497	4.36330	0.01833	73.33	3.41587	1.26010	0.00013
Asn	34.17	5.39543	3.73214	0.00618	83.68	3.26195	1.30864	0.00008
Asp	36.17	5.30368	3.66114	0.00505	85.68	3.27170	1.33558	0.00008
Lys	30.00	6.69313	4.82917	0.01150	85.01	3.66577	1.38908	0.000035
Hyp	28.83	5.96795	4.82216	0.01545	78.34	3.32198	1.42521	0.00010
Gln	35.67	6.21193	4.39881	0.00505	90.66	3.64139	1.44242	0.00003
Glu	37.67	6.12018	4.32781	0.00412	92.67	3.64889	1.46916	0.00003
His	32.17	7.19659	5.43098	0.00654	92.66	4.01363	1.63775	0.000012
Arg	36.67	7.78291	5.53389	0.00345	102.66	4.36498	1.67151	0.000004
Phe	33.17	8.05300	6.14710	0.00578	99.17	4.40001	1.78495	0.000005
Tyr	38.84	8.52796	6.55737	0.00258	110.32	4.76156	1.97446	0.000002
Trp	39.01	10.2331	8.32851	0.00219	121.50	5.51765	2.35735	0.0000002

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

similar to the modeling of the combination made up of two χ indices alone seen at the beginning of this paragraph

$$\{{}^{0}\chi^{v}, {}^{1}\psi_E\}: Q = 0.373, F = 764, r = 0.995, s = 2.7, \langle u\rangle = 9.1, \mathbf{u} = (17, 4.5, 5.9)$$

We will now try to check if a point in the data set is inflating or deteriorating the modeling by leaving it out of the modeling and detecting the new Q value for the given X_V and Y_V terms to check which of these terms is too much influenced by a specific point of the data set. This procedure will be applied for each data point, and the average $\langle Q(X)\rangle$ and $\langle Q(Y)\rangle$ values as well as the Q_{MAX} and Q_{MIN} values for an excluded point will be given. This kind of leave-one-out method will be called the "*Q*-1-out" method. For V we have $\langle Q(X_V)\rangle = 0.442$, $Q(X_V)_{MIN} = 0.593$ (Glu-out), $Q(X_V)_{MIN} = 0.423$ (Trp-out), and $\langle Q(Y_V)\rangle = 0.176$, $Q(Y_V)_{MIN} = 0.201$ (His-out), $Q(Y_V)_{MIN} = 0.170$ (many AA-out). No point is deteriorating or inflating, in a consistent way, the modeling.

Isoelectric Point. As this property is highly dependent on the type of functional groups an amino acid has, its modeling affords the construction of special Y_{pI}, as already done with normal χ indices.[2−4] In fact the best single pseudoindex rates quite poorly, $\{{}^{0}\psi_I\}$ with $Q = 0.153$, $F = 1.4$, $r = 0.3$, and $s = 1.70$, and no linear combination of pseudoindices consistently improves the modeling. The statistical Q/F score of the molar masses is, instead, much worse with $Q = 0.002$ and $F = 0.14$. The family of eight terms that can be derived from the following general type of term can achieve an improved description:

$$Y_{pI} = \frac{\psi}{S_{\psi_E}}\left(1 + \frac{\Delta n}{n_T}\right) \tag{9}$$

Here $\Delta n = n_A - n_B$, n_A = number of acidic groups (2 for Asp and Glu, 1 for all others), n_B = number of basic groups (2 for Lys and His, 3 for Arg, and 1 for all others), and $n_T = 3$ (total number of functional groups), and if $n_T = 2$, then $\Delta n = 0$. Clearly, there are eight such terms following the type of ψ index in expression 9. The nomenclature for such terms can be defined in the following way: for $\psi \equiv {}^{S}\psi_E \rightarrow Y_{pI} \equiv {}^{S}Y_E$, and so on. The best single descriptor for pI is the trivial, ${}^{S}Y_E$, term, which is just equal to $1 + \Delta n/n_T$. This descriptor scores $Q = 2.12$, $F = 270$, $r = 0.966$, $s = 0.5$, $\langle u\rangle = 22$, and $\mathbf{u} = (16, 28)$. This $1 + \Delta n/n_T$ descriptor is also the best descriptor in the χ description. The following linear combination of two pseudoterms (LCpCI) improves the modeling only in a minor way:

$$\{{}^{0}Y_I, {}^{S}Y_E\}: Q = 2.16, F = 139, r = 0.969, s = 0.4_5, \langle u\rangle = 12.5, \mathbf{u} = (1.3, 9.4, 27)$$

Even the following convoluted higher-level I_S pseudoconnectivity term, with its score of $Q = 2.20$, $F = 288$, $r = 0.969$, $s = 0.4$, $\langle u\rangle = 15$, and $\mathbf{u} = (17, 30)$, does not improve the modeling:

$$Y_{pI} = \frac{({}^{S}Y_E - {}^{T}\psi_I)({}^{S}\psi_E - 0.8\,{}^{S}\psi_I)}{({}^{0}\psi_I - {}^{T}\psi_I)} \tag{10}$$

A conclusive word about pI simulation is given by a linear combination of the two-X_{pI} index (LCCT), and by a higher-level molecular connectivity X term. The two-X LCCT shows an unreliable u_1 value, a fact which endangers the utility of this LCCT

$$\{{}^{D}X^{v}, {}^{0}X^{v}\}: Q = 2.14, F = 136, r = 0.966, s = 0.5, \langle u\rangle = 12, \mathbf{u} = (1.2, 5.6, 28)$$

The following higher-level molecular connectivity X_{pI} term, instead, shows an improved modeling power, with $Q = 3.41$, $F = 693$, $r = 0.987$, $s = 0.3$, $\langle u\rangle = 58$, and $\mathbf{u} = (26, 90)$. This term is in every insight better than the corresponding higher-level I_S molecular pseudoconnectivity term Y_{pI}.

$$X_{pI} = \frac{[(1\chi^{v})^{0.5} + 180\chi_t^{v}]}{D}\left(0.04\chi_t^{v} - \frac{\Delta n}{n_T}\right) \tag{11}$$

No interesting mixed $\{\chi, \psi\}$ descriptions can be obtained for pI.

The *Q*-1-out method for this property has $\langle Q(X_{pI})\rangle = 3.43$, $Q(X_{pI})_{MAX} = 4.16$ (Cys-out), $Q(X_{pI})_{MIN} = 3.31$ (Asp-out), and $\langle Q(Y_{pI})\rangle = 2.23$, $Q(Y_{pI})_{MAX} = 3.43$ (Lys-out), $Q(Y_{pI})_{MIN} = 2.10$ (Arg-out). The amino acid Cys is deteriorating the descriptive power of X_{pI}, as without Cys the correlation coefficient enhances

TABLE 5: Intrinsic State Value Matrices, I_S, and the Electrotopological E_S Matrices of 35 Alkanes[a]

Alk	I_S matrices	E_S matrices	Alk	I_S matrices	E_S matrices
4	(2 1.5 1.5 2)	(2.18 1.32 1.32 2.18)	2·3	$\begin{pmatrix} 2 & 1.33 & 2 \\ 0 & 2 & 0 \end{pmatrix}$	$\begin{pmatrix} 2.17 & 0.83 & 2.17 \\ 0 & 2.17 & 0 \end{pmatrix}$
2·4	$\begin{pmatrix} 2 & 1.33 & 1.5 & 2 \\ 0 & 2 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2.22 & 0.88 & 1.31 & 2.2 \\ 0 & 2.22 & 0 & 0 \end{pmatrix}$	22:3	$\begin{pmatrix} 2 & 1.25 & 2 \\ 0 & \underline{2} & 0 \end{pmatrix}$	$\begin{pmatrix} 2.19 & 0.5 & 2.19 \\ 0 & \underline{2.19} & 0 \end{pmatrix}$
22:4	$\begin{pmatrix} 2 & 1.25 & 1.5 & 2 \\ 0 & \underline{2} & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2.24 & 0.51 & 1.27 & 2.21 \\ 0 & \underline{2.24} & 0 & 0 \end{pmatrix}$	5	(2 1.5 1.5 1.5 2)	(2.21 1.34 1.39 1.34 2.21)
23:4	$\begin{pmatrix} 2 & 1.33 & 1.33 & 2 \\ 0 & 2 & 2 & 0 \end{pmatrix}$	$\begin{pmatrix} 2.24 & 0.85 & 0.85 & 2.24 \\ 0 & 2.24 & 2.24 & 0 \end{pmatrix}$	2·5	$\begin{pmatrix} 2 & 1.33 & 1.5 & 1.5 & 2 \\ 0 & 2 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2.25 & 0.90 & 1.38 & 1.33 & 2.22 \\ 0 & 2.25 & 0 & 0 & 0 \end{pmatrix}$
6	(2 1.5 1.5 1.5 1.5 2)	(2.23 1.36 1.41 1.41 1.36 2.23)	3·5	$\begin{pmatrix} 2 & 1.5 & 1.33 & 1.5 & 2 \\ 0 & 0 & 2 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2.23 & 1.33 & 0.94 & 1.33 & 2.23 \\ 0 & 0 & 2.28 & 0 & 0 \end{pmatrix}$
2·6	$\begin{pmatrix} 2 & 1.33 & 1.5 & 1.5 & 1.5 & 2 \\ 0 & 2 & 0 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2.27 & 0.90 & 1.40 & 1.40 & 1.35 & 2.24 \\ 0 & 2.27 & 0 & 0 & 0 & 0 \end{pmatrix}$	22:5	$\begin{pmatrix} 2 & 1.25 & 1.5 & 1.5 & 2 \\ 0 & \underline{2} & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2.27 & 0.55 & 1.34 & 1.31 & 2.23 \\ 0 & \underline{2.27} & 0 & 0 & 0 \end{pmatrix}$
3·6	$\begin{pmatrix} 2 & 1.5 & 1.33 & 1.5 & 1.5 & 2 \\ 0 & 0 & 2 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2.25 & 1.34 & 0.95 & 1.40 & 1.34 & 2.24 \\ 0 & 0 & 2.31 & 0 & 0 & 0 \end{pmatrix}$	23:5	$\begin{pmatrix} 2 & 1.33 & 1.33 & 1.5 & 2 \\ 0 & 2 & 2 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2.27 & 0.86 & 0.90 & 1.32 & 2.24 \\ 0 & 2.27 & 2.30 & 0 & 0 \end{pmatrix}$
25:6	$\begin{pmatrix} 2 & 1.33 & 1.5 & 1.5 & 1.33 & 2 \\ 0 & 2 & 0 & 0 & 2 & 0 \end{pmatrix}$	$\begin{pmatrix} 2.28 & 0.89 & 1.39 & 1.39 & 0.89 & 2.28 \\ 0 & 2.28 & 0 & 0 & 2.28 & 0 \end{pmatrix}$	24:5	$\begin{pmatrix} 2 & 1.33 & 15 & 1.33 & 2 \\ 0 & 2 & 0 & 2 & 0 \end{pmatrix}$	$\begin{pmatrix} 2.26 & 0.88 & 1.36 & 0.88 & 2.26 \\ 0 & 2.26 & 0 & 2.26 & 0 \end{pmatrix}$
24:6	$\begin{pmatrix} 2 & 1.33 & 1.5 & 1.33 & 1.5 & 2 \\ 0 & 2 & 0 & 2 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2.28 & 0.88 & 1.39 & 0.93 & 1.33 & 2.26 \\ 0 & 2.28 & 0 & 2.32 & 0 & 0 \end{pmatrix}$	33:5	$\begin{pmatrix} 2 & 1.5 & 1.25 & 1.5 & 2 \\ 0 & 0 & \underline{2} & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2.24 & 1.30 & 0.58 & 1.30 & 2.24 \\ 0 & 0 & \underline{2.30} & 0 & 0 \end{pmatrix}$
22:6	$\begin{pmatrix} 2 & 1.25 & 1.5 & 1.5 & 1.5 & 2 \\ 0 & \underline{2} & 0 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2.29 & 0.55 & 1.36 & 1.37 & 1.33 & 2.24 \\ 0 & \underline{2.29} & 0 & 0 & 0 & 0 \end{pmatrix}$	223∴5	$\begin{pmatrix} 2 & 1.25 & 133 & 1.5 & 2 \\ 0 & \underline{2} & 2 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2.29 & 0.51 & 0.85 & 1.29 & 2.25 \\ 0 & \underline{2.29} & 2.30 & 0 & 0 \end{pmatrix}$
33:6	$\begin{pmatrix} 2 & 1.5 & 1.25 & 1.5 & 1.5 & 2 \\ 0 & 0 & \underline{2} & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2.26 & 1.31 & 0.59 & 1.36 & 1.32 & 2.25 \\ 0 & 0 & \underline{2.33} & 0 & 0 & 0 \end{pmatrix}$	224∴5	$\begin{pmatrix} 2 & 1.25 & 15 & 1.33 & 2 \\ 0 & \underline{2} & 0 & 2 & 0 \end{pmatrix}$	$\begin{pmatrix} 2.28 & 0.52 & 1.33 & 0.84 & 2.27 \\ 0 & \underline{2.28} & 0 & 2.27 & 0 \end{pmatrix}$
7	(2 1.5 1.5 1.5 1.5 1.5 2)	(2.25 1.36 1.42 1.44 1.42 1.36 2.25)	233∴5	$\begin{pmatrix} 2 & 1.33 & 125 & 1.5 & 2 \\ 0 & 2 & \underline{2} & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2.28 & 0.81 & 0.54 & 1.28 & 2.25 \\ 0 & 2.28 & \underline{2.31} & 0 & 0 \end{pmatrix}$
2·7	$\begin{pmatrix} 2 & 1.3 & 1.5 & 1.5 & 1.5 & 1.5 & 2 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2.28 & 0.90 & 1.41 & 1.42 & 1.42 & 1.35 & 2.25 \\ 0 & 2.28 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$	234∴5	$\begin{pmatrix} 2 & 1.33 & 133 & 1.33 & 2 \\ 0 & 2 & 2 & 2 & 0 \end{pmatrix}$	$\begin{pmatrix} 2.28 & 0.84 & 0.87 & 0.84 & 2.28 \\ 0 & 2.28 & 2.31 & 2.28 & 0 \end{pmatrix}$
3·7	$\begin{pmatrix} 2 & 1.5 & 1.33 & 1.5 & 1.5 & 1.5 & 2 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2.26 & 1.35 & 0.95 & 1.42 & 1.41 & 1.35 & 2.25 \\ 0 & 0 & 2.33 & 0 & 0 & 0 & 0 \end{pmatrix}$	3-5	$\begin{pmatrix} 2 & 1.5 & 1.33 & 1.5 & 2 \\ 0 & 0 & 1.5 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2.26 & 1.35 & 0.99 & 1.35 & 2.26 \\ 0 & 0 & 1.35 & 0 & 0 \\ 0 & 0 & 2.26 & 0 & 0 \end{pmatrix}$
4·7	$\begin{pmatrix} 2 & 1.5 & 1.5 & 1.33 & 1.5 & 1.5 & 2 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2.26 & 1.35 & 1.41 & 0.96 & 1.41 & 1.35 & 2.26 \\ 0 & 0 & 0 & 2.34 & 0 & 0 & 0 \end{pmatrix}$	23≠5	$\begin{pmatrix} 2 & 1.33 & 1.33 & 1.5 & 2 \\ 0 & 2 & 1.5 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2.30 & 0.88 & 0.95 & 1.34 & 2.27 \\ 0 & 2.30 & 1.34 & 0 & 0 \\ 0 & 0 & 2.27 & 0 & 0 \end{pmatrix}$
23:7	$\begin{pmatrix} 2 & 1.33 & 1.33 & 1.5 & 1.5 & 1.5 & 2 \\ 0 & 2 & 2 & 0 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2.31 & 0.87 & 0.92 & 1.41 & 1.40 & 1.34 & 2.26 \\ 0 & 2.31 & 2.35 & 0 & 0 & 0 & 0 \end{pmatrix}$	33≠5*	$\begin{pmatrix} 2 & 1.5 & 1.25(2) & 1.5 & 2 \\ 0 & 0 & 1.5 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2.27 & 1.32 & 0.63(2.35) & 1.32 & 2.27 \\ 0 & 0 & 1.32 & 0 & 0 \\ 0 & 0 & 2.27 & 0 & 0 \end{pmatrix}$
26:7	$\begin{pmatrix} 2 & 1.33 & 1.5 & 1.5 & 1.5 & 1.33 & 2 \\ 0 & 2 & 0 & 0 & 0 & 2 & 0 \end{pmatrix}$	$\begin{pmatrix} 2.29 & 0.89 & 1.40 & 1.41 & 1.40 & 0.89 & 2.29 \\ 0 & 2.29 & 0 & 0 & 0 & 2.29 & 0 \end{pmatrix}$	33=5	$\begin{pmatrix} 2 & 1.5 & 1.25 & 1.5 & 2 \\ 0 & 0 & \underline{1.5} & 0 & 0 \\ 0 & 0 & \underline{2} & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2.30 & 1.34 & 0.67 & 1.34 & 2.27 \\ 0 & 0 & \underline{1.34} & 0 & 0 \\ 0 & 0 & \underline{2.30} & 0 & 0 \end{pmatrix}$
22:7	$\begin{pmatrix} 2 & 1.25 & 1.5 & 1.5 & 1.5 & 1.5 & 2 \\ 0 & \underline{2} & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2.31 & 0.55 & 1.38 & 1.40 & 1.40 & 1.34 & 2.25 \\ 0 & \underline{2.31} & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$			

[a] A number underlined in the second row has a symmetric companion on the top of the first row; "." means methyl; ":" means dimethyl; "∴" means trimethyl; "−" means E, "≠" means ME, and "=" means EE. For example: 223∴5 means 2,2,3-trimethylpentane, and 33=5 means 33-diethylpentane. An asterisk indicates that the value in parentheses in 33ME5 should be placed on top of the value at its left.

up to $r(X_{pI}) = 0.991$, and the amino acid Lys is deteriorating the corresponding modeling power of Y_{pI}, as with Lys-out $r(Y_{pI}) = 0.983$.

Melting Temperature. Molar masses are very bad descriptors for the melting points, T_m, of 20 amino acids (no Hyp) with $Q = 0.003$ and $F = 0.15$. Normally the modeling of this property is rather problematic, but it can be interesting to detect how pseudoindices rate relative to connectivity indices. The χ-modeling affords a molecular connectivity term, X_{T_m}, quite similar to the X_{pI} term (see eq 9, where $^0\chi^v$ replaces $^S\psi_E$), used to model

TABLE 6: $I_S-\psi$ Indices and the Corresponding *M* Values for 35 Alkanes

alkane	*M*	${}^{S}\psi_I$	${}^{0}\psi_I$	${}^{1}\psi_I$	${}^{T}\psi_I$	${}^{0}\psi_E$	${}^{1}\psi_E$	${}^{T}\psi_E$
4	58	7	3.04721	1.82137	0.33333	3.09535	1.93658	0.34752
2M3	58	7.33	2.98843	1.83942	0.30657	3.13418	2.23538	0.34338
2M4	72	8.83	3.80493	2.51162	0.25031	3.95622	2.95133	0.28285
2M5	86	10.33	4.62142	3.17829	0.20438	4.77695	3.62286	0.23209
24MM6	114	13.66	6.19564	4.54075	0.12531	6.46451	5.35240	0.15572
33MM5	100	12.25	5.35585	3.88021	0.14907	5.72225	5.20690	0.19605
5	72	8.5	3.86370	2.48803	0.27217	3.92127	2.62765	0.28641
23MM4	86	10.66	4.56265	3.20444	0.18797	4.84192	4.07533	0.23447
33MM6	114	13.75	6.17234	4.54687	0.12172	6.54557	5.86741	0.16158
22MM5	100	12.25	5.35585	3.87168	0.14907	5.74679	5.18957	0.19927
22MM6	114	13.75	6.17234	4.53835	0.12172	6.57797	5.88218	0.16515
4M7	114	13.33	6.25442	4.51716	0.13625	6.41035	4.98086	0.15509
3M7	114	13.33	6.25442	4.51716	0.13625	6.41562	4.99395	0.15604
3M6	100	11.83	5.43792	3.85049	0.16688	5.59164	4.31168	0.18965
24MM5	100	12.16	5.37914	3.86854	0.15348	5.65026	4.66456	0.19078
23MM5	114	12.16	5.37914	3.87664	0.15348	5.65779	4.76215	0.19201
3E5	100	11.83	5.43792	3.85603	0.16688	5.58260	4.31251	0.18859
2M6	100	11.83	5.43792	3.84495	0.16688	5.60066	4.30686	0.19074
3M5	86	10.33	4.62142	3.18382	0.20438	4.76721	3.63311	0.23031
23ME5	114	13.66	6.19564	4.54884	0.12531	6.46593	5.41886	0.15633
223MMM5	114	14.08	6.11357	4.57141	0.11194	6.67389	6.55196	0.16963
234MMM5	114	13.95	6.13686	4.56946	0.11525	6.58029	5.98920	0.16440
2M7	114	13.33	6.25442	4.51162	0.13625	6.42647	4.99086	0.15732
224MMM5	114	14.08	6.11357	4.56193	0.11194	6.65920	6.35212	0.16788
233MMM5	114	14.08	6.11357	4.57440	0.11194	6.66292	6.56652	0.16917
22MM4	86	10.75	4.53935	3.20501	0.18257	4.96477	4.64625	0.24931
6	86	10	4.68020	3.15470	0.22222	4.73859	3.30194	0.23385
25MM6	114	13.66	6.19564	4.53521	0.12531	6.46544	5.32558	0.15550
7	100	11.50	5.49670	3.84373	0.18144	5.56001	3.98114	0.19178
23MM7	128	15.16	7.01214	5.20997	0.10232	7.29928	6.10314	0.12910
22MM7	128	15.25	6.98884	5.20501	0.09938	7.39435	6.54912	0.13449
26MM7	128	15.16	7.01214	5.20188	0.10232	7.29573	6.01709	0.12885
33ME5	114	13.75	6.17234	4.55540	0.12172	6.51454	5.84471	0.15845
33EE5	128	15.25	6.98884	5.23059	0.09938	7.31469	6.50000	0.12862
22MM3	72	9.25	3.72285	2.52982	0.22361	4.11716	3.82255	0.29487

pI. The main difference between the two properties resides in the definition of Δn. Here, amino acids Leu and Tyr have $\Delta n = 1$, amino acids Pro, Ser, Thr, Cys, Asn, Asp, Gln, Glu, Lys, His, and Arg instead have $\Delta n = -1$, and the rest have $\Delta n = 0$.[2,3] Practically the best molecular connectivity description is given by the following term:

$$\{{}^{0}X^{v}\}: Q = 0.037, F = 48, r = 0.854, s = 23, \langle u \rangle = 6, \mathbf{u} = (6.9, 6.1)$$

Practically, because the ${}^{0}X$ term shows a very similar modeling power, with a somewhat better $F = 50$, and $r = 0.857$, we prefer to stress the importance of the trivial index $1 + \Delta n/n_T \equiv {}^{0}X^{v}$. The best molecular pseudoconnectivity description, instead, is given by a formally similar I_S pseudoterm, Y_{Tm}, where ${}^{0}\psi_I$ replaces ${}^{S}\psi_E$ in eq 9 and where Δn fits the new definition used for X_{T_m}

$$Y_{T_m} = \frac{\psi}{{}^{0}\psi_I}\left(1 + \frac{\Delta n}{n_T}\right) \tag{12}$$

The term $\{{}^{1}Y_E\}$, where ${}^{1}\psi_E$ replaces ψ in the numerator, achieves here the best description with $Q = 0.039$, $F = 54$, $r = 0.867$, $s = 22$, $\langle u \rangle = 6.4$, and $\mathbf{u} = (7.4, 5.4)$. Like the isoelectric point, no positive modeling for T_m can be achieved with a mixed set of $\{\chi, \psi\}$ indices. Because the modeling of this property is far from being optimal and the approximations done are too drastic, the *Q*-1-out method will not be considered here.

Solubility. The modeling of the solubility, *S*, of 20 amino acids (no Cys), whose experimental values are in Table 2, was performed with LCCI and with molecular connectivity terms.[2-5] It was the modeling of this property, which includes such strong outliers as Arg, Ser, Hyp, and Pro, that obliged introduction of the following set of reciprocal connectivity indices ($R = 1/\chi$) weighted by an association parameter, *a*:

$$\{aR\} = \{a^{D}R, a^{D}R^{v}, a^{0}R, a^{0}R^{v}, a^{1}R, a^{1}R^{v}, R_t/a, R_t v/a\} \tag{13}$$

where, a(Pro) = 8, a(Ser, Hyp, Arg) = 2, and a(others) = 1. This parameter, *a*, which can be seen as a statistical weighting factor, can, in some cases, assume the physical meaning of an association parameter, due to association or self-association phenomena in solution.[2] The fact that total connectivity indices have to be divided, instead of multiplied by the association parameter *a*, resides in their definition: their values decrease with increasing complexity of the chemical graph.

While the modeling of the molar masses is quite insufficient with $Q = 0.0009$, and $F = 2.1$, a single suprareciprocal molecular connectivity index achieves a more than optimal modeling

$$\{a^{0}R^{v}\}: Q = 0.029, F = 2052, r = 0.996, s = 35, \langle u \rangle = 29, \mathbf{u} = (45, 13)$$

This modeling shows next to the exceptionally good *r* and utility values a rather unsatisfactory *s* value. The correlation **C** vector for this description is $\mathbf{C} = (1010.8, -139.39)$.

A quite similar set of reciprocal pseudoconnectivity indices has to be introduced to model the same property, with the same values for *a* ($R = 1/\psi$, i.e., ${}^{1}R_E = 1/{}^{1}\psi_E$)

$$\{a(1/\psi)\} = \{a^{S}R_I, a^{0}R_I, a^{1}R_I, {}^{T}R_I/a, a^{S}R_E, a^{0}R_E, a^{1}R_E, {}^{T}R_E/a\} \tag{14}$$

The best description is given by the following I_S pseudo-suprareciprocal index:

$$\{a^0R_I\}: Q = 0.013, F = 418, r = 0.979, s = 77, \langle u\rangle = 15, \mathbf{u} = (23, 6.2)$$

This last description is a much worse description than the corresponding $1/\chi$ description.

Now if the modeling is done by the aid of the combined set of eqs 13 and 14, then it is possible to find the following interesting combination

$$\{a^0R^v, {}^TR_I/a\}: Q = 0.028, F = 1009, r = 0.996, s = 35, \langle u\rangle = 17, \mathbf{u} = (41, 0.83, 8.1)$$

whose negative point is the low u_2 utility value of the second index and also a poor s value. Also the combination $\{a^0R^v, {}^TR_E/a\}$ shows a similar descriptive power with $Q = 0.028$, $F = 1005$, and similar poor u_2 and s values.

The Q-1-out method for the solubility of amino acids has $\langle Q(a^0R^v)\rangle = 0.029$, $Q(a^0R^v)_{MAX} = 0.035$ (Hyp-out), $Q(a^0R^v)_{MIN} = 0.028$ (many AA-out), and $\langle Q(a^0R_I)\rangle = 0.014$, $Q(a^0R_I)_{MAX} = 0.025$ (Ser-out), $Q(a^0R_I)_{MIN} = 0.011$ (Pro-out). While the amino acid Hyp has a minor influence on r, which, with Hyp-out, equals $r(a^0R^v) = 0.997$, the amino acid Ser has a greater influence on r of $\{a^0R_I\}$, which, with Ser-out, equals $r(a^0R_I) = 0.994$. To further determine the individual influence of a given experimental point on the modeling characteristics of $\{a^0R^v\}$, we remove sequentially each solubility value from the data set and determine again the **C** vector of the model, and the new model is used to predict the observation left out. In the third column of Table 1 are reported the calculated P_{calc} solubility values with the normal $\mathbf{C} = (1010.8, -139.39)$ vector, and in the fourth column of the same table are reported the P_{jk} calculated solubility values calculated with this method, where the subscript jk stands for *jackknifing*, the vulgar name for this sort of external validation method, which has in the Q-1-out method another example. It should be noticed that calculated S values, to avoid negative values (here Trp), have been obtained with the modulus equation $S = |c_1(a^0R^v) + c_2U_0|$. The figures in columns P_{calc} and P_{jk} are of interest. In spite of some minor variations there is hardly any doubt that there is a similar trend all along the two sets of values; further, the coefficients of the correlation vector all along the jackknifing procedure show only minor variations (~2%). Closer analysis of both P_{calc} and P_{jk} shows that, while prediction of the solubility of some amino acids is insufficient, prediction of S(Asp) fails totally. The Q-1-out method for Asp-out has $Q = 0.033$, quite close to Q_{MAX}, and this shows that the two methods even if complementary do not superpose. The negative results are to be ascribed to the poor s value. The physical meaning of this failure might be due to the neglect of other association phenomena in solution.

Specific Rotation. The specific rotation of 16 L-amino acids, SR_L (in angular degrees, normally given as $[\alpha]^{25}_D$) is an interesting property, as not only some of its values are negative, but also because the modeling of SR_L leads directly to the corresponding modeling of SR_D for the D-amino acids. In fact, the SR_D modeling can be achieved with just a change in sign of the correlation vector, i.e, $\mathbf{C}_L = -\mathbf{C}_D$.[4,5,31] Practically, a modeling done on n amino acids is in this way automatically extended to $2n$ amino acids. Molar mass once again is a quite deceiving descriptor, with $Q = 0.0003$ and $F = 0.34$. The overall best molecular connectivity description is given by the following combination of two reciprocal connectivity indices ($R = 1/\chi$)

$$\{{}^DR, {}^0R\}: Q = 0.089, F = 62, r = 0.952, s = 11, \langle u\rangle = 9.2, \mathbf{u} = (11, 11, 5.8)$$

while the overall best single index description is quite poor and is achieved by $\{\chi_t\}$ with $Q = 0.014$ and $F = 3.2$. The trial-and-error procedure discovers the following optimal term for SR

$$X_{SR} = \frac{{}^0\chi - (\chi_t^v)^{0.3}}{D^{0.8} + 0.2(\chi_t)^{0.02}} \quad (15)$$

with the following statistics: $Q = 0.084$, $F = 112$, $r = 0.943$, $s = 11$, $\langle u\rangle = 11$, $\mathbf{u} = (11, 11)$.

Molecular pseudoconnectivity indices do not achieve the same kind of description; further there are no reciprocal pseudoindices with improved modeling power. While the best single pseudoindex description is given by $\{{}^1\psi_I\}$ with $Q = 0.011$ and $F = 1.9$, the best two pseudoindex descriptions do not achieve the same optimal description of the reciprocal $1/\chi$ combination

$$\{{}^1\psi_I, {}^0\psi_I\}: Q = 0.044, F = 15, r = 0.838, s = 19, \langle u\rangle = 4.5, \mathbf{u} = (5.1, 5.3, 3.2)$$

The I_S molecular pseudoconnectivity term Y_{SR} of eq 16 with $Q = 0.059$, $F = 56$, $r = 0.894$, $s = 15$, $\langle u\rangle = 10$, and $\mathbf{u} = (7.5, 12)$ cannot be compared with X_{SR}

$$Y_{SR} = \frac{{}^1\psi_I + 15^T\psi_I}{({}^0\psi_I - {}^1\psi_I)^{1.5}} \quad (16)$$

This last term shows a rather homogeneous structure as it is made up of ψ_I-type indices only. Here, use of the mixed set $\{\chi, \psi\}$ of eight molecular connectivity indices and eight I_S molecular pseudoconnectivity indices allows the following combination that achieves a better modeling than the set of two I_S molecular pseudoconnectivity indices to be found:

$$\{{}^0\chi, {}^0\psi_E\}: Q = 0.056, F = 25, r = 0.892, s = 16, \langle u\rangle = 6.4, \mathbf{u} = (7.0, 7.1, 5.1)$$

The Q-1-out method for this property has $\langle Q(X_{SR})\rangle = 0.084$, $Q(X_{SR})_{MAX} = 0.092$ (Leu- and Arg-out), $Q(X_{SR})_{MIN} = 0.079$ (Pro-out), and $\langle Q(Y_{SR})\rangle = 0.060$, $Q(Y_{SR})_{MAX} = 0.069$ (Thr-out), $Q(Y_{SR})_{MIN} = 0.054$ (Pro-out). Practically, no amino acid is deteriorating or inflating the modeling power of both X_{SR} and Y_{SR} terms.

Crystal Density. Let us start noticing that the descriptive power of the molar mass for this property, for $n = 10$ amino acids, is $Q = 1.9$ and $F = 1.1$. The best molecular I_S pseudoconnectivity index and the best two-index LCpCI for the 10 points of this property are

$$\{{}^0\psi_I\}: Q = 2.38, F = 1.9, r = 0.43, s = 0.2, \langle u\rangle = 4.8, \mathbf{u} = (1.4, 8.3)$$

$$\{{}^0\psi_I, {}^0\psi_E\}: Q = 6.81, F = 7.6, r = 0.827, s = 0.1, \langle u\rangle = 4.9, \mathbf{u} = (3.5, 3.3, 7.7)$$

These last values should be compared with the modeling power of the corresponding molecular connectivity indices, which show only a small but evident improvement relative to the aforementioned indices[4,31]

$\{{}^0\chi^v\}$: $Q = 3.44$, $F = 3.9$, $r = 0.570$, $s = 0.2$,
$\langle u\rangle = 5.4$, $\mathbf{u} = (2.0, 8.9)$

$\{D^v, {}^0\chi\}$: $Q = 9.05$, $F = 13$, $r = 0.890$, $s = 0.10$,
$\langle u\rangle = 11.5$, $\mathbf{u} = (4.7, 5.2, 13)$

But a more accurate descriptor for this property is surely the following I_S molecular pseudoconnectivity term, which shows a nice improvement in every statistics, with $Q = 9.92$, $F = 32$, $r = 0.895$, $s = 0.1$, $\langle u\rangle = 4.1$, and $\mathbf{u} = (5.7, 2.6)$

$$Y_{CD} = \frac{({}^S\psi_I)^{0.7} + 2.1({}^1\psi_I)^{1.2}}{{}^0\psi_I - 0.8({}^1\psi_E)^{1.1} - 0.8({}^T\psi_I)^{0.7}} \qquad (17)$$

This pseudoterm even overrates the corresponding molecular connectivity term, whose statistical score is $Q = 7.91$, $F = 20.4$, $r = 0.848$, $s = 0.1$, $\langle u\rangle = 5.4$, and $\mathbf{u} = (4.5, 6.4)$.

$$X_{CD} = \frac{({}^0\chi^v)^{1.2} + 1.8{}^0\chi}{({}^1\chi)^{0.8} - 1.3(\chi_t)^{2.1}} \qquad (18)$$

The only negative point of Y_{CD} is the low utility value for U_0. But the last word about the modeling of the crystal density is left to a combination of a molecular connectivity index and a molecular pseudoconnectivity index, which by far overrates even the found trial-and-error terms

$\{{}^0\chi, {}^S\psi_E\}$: $Q = 12.2$, $F = 24$, $r = 0.935$, $s = 0.1$,
$\langle u\rangle = 10$, $\mathbf{u} = (6.7, 6.4, 17)$

It is interesting to notice that these two indices, a normal and a I_S molecular pseudoconnectivity index, are quite correlated with $r({}^0\chi, {}^S\psi_E) = 0.987$.

The Q-1-out method for CD has $\langle Q(X_{CD})\rangle = 8.09$, $Q(X_{CD})_{MAX} = 12.0$ (Arg-out), $Q(X_{CD})_{MIN} = 7.07$ (Leu-out), and $\langle Q(Y_{CD})\rangle = 10.5$, $Q(Y_{CD})_{MAX} = 18.7$ (Met-out), $Q(Y_{CD})_{MIN} = 8.98$ (Asp-out). While the amino acid Arg is deteriorating the descriptive power of X_{CD}, as with Arg-out $r(X_{CD})$ is enhanced to 0.907, the amino acid Met is deteriorating in a consistent way the descriptive power of Y_{CD}, as with Met-out $r(Y_{pI})$ is enhanced to 0.970.

Alkanes. Alkanes do not have multiple bonds and lone-pair electrons; for this reason the subset $\{\chi^v\}$ of valence molecular connectivity indices coincides with the subset of nonvalence $\{\chi\}$ indices. For pseudoconnectivity indices, instead, the only reduction in number of I_S pseudoindices is due to the E-state relation, ${}^S\psi_I \equiv {}^S\psi_E$, which reduces the number of chosen pseudoindices from eight to seven only, i.e., $\{{}^S\psi_I, {}^0\psi_I, {}^1\psi_I, {}^T\psi_I, {}^0\psi_E, {}^1\psi_E, {}^T\psi_E\}$. That for alkanes the relation ${}^S\psi_I \equiv {}^S\psi_E$ holds is due to the fact that ψ_E values need not be rescaled, as every S value is here positive.

Motor Octane Number. While the description of this property for 30 alkanes (see Table 2) by the aid of the molar masses is very bad with $Q = 0.006$, $F = 0.79$, and $r = 0.17$, an optimal description of M is, instead, achieved by the $\{{}^1\psi_I\}$ pseudoindex with $Q = 2.60$, $F = 61496$, $r = 0.9998$, and $s = 0.4$. The best connectivity index for M, $\{D\}$, achieves only $Q = 0.262$, $F = 625$, $r = 0.978$, and $s = 3.7$. The modeling of the motor octane number with indices of the set $\{D, {}^0\chi, {}^1\chi, \chi_t\}$[2,4,5] shows an unsatisfactory description at the level of the best single-index $\{\chi_t\}$ which fares only $Q = 0.015$ and $F = 4.7$, while the best three-index LCCI shows an adequate modeling with

$\{D, {}^0\chi, {}^1\chi\}$: $Q = 0.092$, $F = 57$, $r = 0.932$, $s = 10$,
$\mathbf{u} = (6.1, 7.1, 4.3, 5.3)$, $\langle u\rangle = 5.7$

The trial-and-error search finds a very effective X_{MON} term, whose ratings are

$Q = 0.085$, $F = 146$, $r = 0.916$, $s = 11$,
$\langle u\rangle = 19.5$, $\mathbf{u} = (12, 27)$

$$X_{MON} = \frac{({}^0\chi{\cdot}\chi_t)^{0.1} + (D)^{1.3}}{({}^0\chi - 1.5{}^1\chi)^{1.2}} \qquad (19)$$

This dominant term offers the possibility, through a forward combinatorial search, to find a mixed linear combination made up of a molecular connectivity term and three molecular connectivity indices, with enhanced Q, r, and s values

$\{X_{MON}, D, {}^0\chi, {}^1\chi\}$: $Q = 0.129$, $F = 85$, $r = 0.965$, $s = 7.9$,
$\langle u\rangle = 7.5$, $\mathbf{u} = (4.8, 5.6, 5.4, 5.4, 4.1)$

Molecular pseudoconnectivity indices, which have a mean interrelation of $\langle r\rangle = 0.964$, $r_{max}({}^S\psi_I, {}^0\psi_E) = 0.998$, and $r_{min}({}^1\psi_E, {}^T\psi_E) = 0.86$, show a satisfactory modeling only with a LCpCI of four indices

$\{{}^S\psi_I, {}^0\psi_I, {}^T\psi_I, {}^T\psi_E\}$: $Q = 0.107$, $F = 59$, $r = 0.951$,
$s = 8.9$, $\mathbf{u} = (8.9, 8.3, 5.5, 4.7, 0.5)$, $\langle u\rangle = 5.6$

Combinations with more indices do not improve any further the descriptive power of the modeling. The following molecular pseudoconnectivity term, Y_{MON}, found by a trial-and-error procedure, instead, shows a more than impressive modeling power, which overrates even the descriptive power of X_{MON}

$Q = 0.108$, $F = 237$, $r = 0.946$, $s = 8.7$,
$\mathbf{u} = (15, 32)$, $\langle u\rangle = 24$

$$Y_{MON} = \frac{{}^S\psi_I + {}^T\psi_E}{({}^S\psi_I - 2{}^0\psi_I)^{1.2}} \qquad (20)$$

This term in a linear combination with the ${}^1\psi_E$ descriptor shows a small improvement in $Q = 0.110$, due to an improved $r = 0.950$, but F worsens to $F = 124$. The correlation vector of this pseudoterm is $\mathbf{C} = (-5.29525, 127.268)$. No mixed set $\{\chi, \psi\}$ LCCI shows an improved modeling.

The Q-1-out method for this property of alkanes has $\langle Q(X_{MON})\rangle = 0.085$, $Q(X_{MON})_{MAX} = 0.095$ (22MM3-out), $Q(X_{MON})_{MIN} = 0.081$ (7-out), and $\langle Q(Y_{MON})\rangle = 0.108$, $Q(Y_{MON})_{MAX} = 0.124$ (22MM3-out), $Q(Y_{MON})_{MIN} = 0.105$ (4M7-out). Alkane 22MM3 is not consistently deteriorating the descriptive power of both X_{MON} and Y_{MON}, and also, no alkane is excessively inflating the modeling of MON in both X and Y representations. In Table 2, the fourth, fifth, ninth, and last columns report the calculated MON values and the calculated MON jackknifing values, P_{calc} and P_{jk}, respectively. The two sets of values, P_{calc} and P_{jk}, are quite similar, more similar than the corresponding values for the solubility of the amino acids. The only consistent deviation between P_{calc} and P_{jk} can be detected in compound 22MM3. Even here calculated MON values have been obtained with the modulus equation MON $= |c_1 Y_{MON} + c_2 U_0|$, as alkane 7 has $P_{calc} < 0$.

Melting Temperature. The modeling of the subclass of 17 melting points of Table 2, which is made up of similar alkanes, [MMi + MEi + EEi], where i is the main chain, and M and E

TABLE 7: Intrinsic State Value Vectors, I_S, and the Electrotopological E_S Vectors of 20 Metal Halids, MeX[a]

MeX (*M*)	$\mathbf{I}_S$ vector	$\mathbf{E}_S$ vector	MeX (*M*)	$\mathbf{I}_S$ vector	$\mathbf{E}_S$ vector
LiF (25.9)	(2 8)	(0.5 9.5)	LiCl (42.4)	(2 4.11)	(1.47 4.64)
NaF (42)	(1.44 8)	(−0.19 9.64)	NaCl (58.4)	(1.44 4.11)	(0.78 4.78)
KF (50.1)	(1.25 8)	(−0.44 9.69)	KCl (74.6)	(1.25 4.11)	(0.53 4.83)
RbF (10.5)	(1.16 8)	(−0.55 9.71)	RbCl (120.9)	(1.16 4.11)	(0.42 4.85)
CsF (151.9)	(1.11 8)	(−0.61 9.72)	CsCl (168.4)	(1.11 4.11)	(0.36 4.86)
LiBr (86.8)	(2 2.75)	(1.81 2.94)	LiI (133.8)	(2 2.12)	(1.97 2.15)
NaBr (102.9)	(1.44 2.75)	(1.12 3.08)	NaI (149.9)	(1.44 2.12)	(1.28 2.29)
KBr (119)	(1.25 2.75)	(0.88 3.13)	KI (166)	(1.25 2.12)	(1.03 2.34)
RbBr (165.4)	(1.16 2.75)	(0.76 3.15)	RbI (212.4)	(1.16 2.12)	(0.92 2.36)
CsBr (212.8)	(1.11 2.75)	(0.70 3.16)	CsI (259.8)	(1.11 2.12)	(0.86 2.37)

[a] M = molar mass.

stand for methyl and ethyl substitutents along the main chain, has never been satisfactory.[2] The best single descriptor for this set of compounds is the following X_{MP} term, whose statistical ratings are

$$Q = 0.033,\ F = 19,\ r = 0.749,\ s = 23,\ \langle u\rangle = 5.7,\ \mathbf{u} = (4.4, 7.0)$$

$$X_{MP} = \frac{(D - {}^0\chi)^2}{({}^1\chi - 3.9\chi_t)^{0.6}} \tag{21}$$

The subscript MP stands for melting points to differentiate it from the subscript T_m used for amino acids. This term is far from being an optimal descriptor, and a better but not optimal description for this property can be achieved by a normal LCCI composed by the following two indices:

$$\{{}^1\chi, \chi_t\}:\ Q = 0.043,\ F = 16,\ r = 0.834,\ s = 20,\ \langle u\rangle = 5.0,\ \mathbf{u} = (5.2, 5.6, 4.3)$$

The given description is nevertheless much better than the description achieved by molar masses, which shows the following statistical values: $Q = 0.005$, $F = 0.38$, and $r = 0.16$.

The best I_S molecular pseudoconnectivity descriptor for this property is the following Y_{MP} term, highly influenced by the total type of pseudoindices, ${}^T\psi$. This $I_S-\psi$ term rates

$$Q = 0.066,\ F = 77,\ r = 0.914,\ s = 14,\ \langle u\rangle = 26,\ \mathbf{u} = (8.8, 43)$$

$$Y_{MP} = \left(\frac{{}^T\psi_I + {}^T\psi_E}{{}^1\psi_I - 0.8\,{}^T\psi_E}\right)^6 \tag{22}$$

This term overrates the descriptive power of X_{MP}. It is, further, important to notice that single-ψ or LCpCI fares better than single-χ or LCCI descriptions. In fact, for what concerns the single descriptors we have in both graph−χ and $I_S-\psi$ representations $\{\chi_t\}$, $Q = 0.011$, $F = 2$, $r = 0.339$, $s = 32$, and $\{{}^T\psi_I\}$, $Q = 0.023$, $F = 9$, $r = 0.617$, $s = 27$. Noteworthy is also the descriptions achieved by a mixed combination of molecular connectivity and pseudoconnectivity indices, whose modeling power is not unimportant: $\{\chi_t, {}^T\psi_E\}$, $Q = 0.040$, $F = 14$, $r = 0.817$, $s = 21$, $\mathbf{u} = (5.0, 4.8, 6.6)$, $\langle u\rangle = 5.5$.

The Q-1-out method for MP has $\langle Q(X_{MP})\rangle = 0.033$, $Q(X_{MP})_{MAX} = 0.056$ (33EE5-out), $Q(X_{MP})_{MIN} = 0.017$ (22MM3-out) and $\langle Q(Y_{MP})\rangle = 0.066$, $Q(Y_{MP})_{MAX} = 0.075$ (33ME5- and 22MM5-out), $Q(Y_{MP})_{MIN} = 0.058$ (22MM3-out). For what concerns the descriptive power of X_{MP}, while alkane 33EE5 is deteriorating the description, in fact, with 33EE5-out we have $r(X_{MP}) = 0.856$, alkane 22MM3 is, instead, inflating dramatically the description, in fact, with 22MM3-out $r(X_{MP})$ worsens to 0.405. The descriptive power of Y_{MP} instead is not deteriorated by 33ME5 and 22MM5, as with 33ME5-out and 22MM5-out we have $r(Y_{MP}) = 0.935$.

Lattice Enthalpy of Inorganic Salts. The lattice enthalpies of $n = 20$ metal halides, MeX, of Table 7 can be modeled in an adequate way by the molar mass M, with $Q = 0.015$, $F = 45$, and $r = 0.846$. The best ψ descriptor for M is $\{{}^0\psi_I\}$ with $Q = 0.028$, $F = 62$, $r = 0.880$, and $s = 31$, while the best χ index for M is $\{{}^1\chi^v\}$ with $Q = 0.030$, $F = 70$, $r = 0.895$, and $s = 29$. For these metal halides, whose graph may approximately be represented as two connected points, ●—●, we have ${}^1\psi_I \equiv {}^T\psi_I$ and ${}^1\psi_E \equiv {}^T\psi_E$, while ${}^S\psi_I$ and ${}^S\psi_E$ are nearly coincident with $r = 0.99998$. The overall mean $\langle r\rangle$ equals 0.941, and the weakest interrelation is $r({}^0\psi_I, {}^S\psi_E) = 0.863$. The best single-, two-, and three-index LCpCI descriptions are

$$\{{}^0\psi_I\}:\ Q = 0.027,\ F = 157,\ r = 0.947,\ s = 35,\ \mathbf{u} = (12, 23),\ \langle u\rangle = 18$$

$$\{{}^0\psi_I, {}^1\psi_I\}:\ Q = 0.040,\ F = 167,\ r = 0.975,\ s = 24,\ \mathbf{u} = (8.1, 4.4, 16),\ \langle u\rangle = 9.6$$

$$\{{}^0\psi_I, {}^1\psi_I, {}^S\psi_E\}:\ Q = 0.060,\ F = 250,\ r = 0.989,\ s = 17,\ \mathbf{u} = (13, 7.6, 4.6, 16),\ \langle u\rangle = 10$$

From these ratings we notice the overall good quality of the three-index LCpCI. Only the average utility decreases along this series but not in a way to endanger the utility of the LCpCI. Now, three indices to model twenty properties could be judged a rather risky choice. A trial-and-error procedure discovers, in fact, the following very good I_S pseudo-term

$$Y_{\Delta H} = \frac{(1.5\,{}^0\psi_E + {}^1\psi_E)^8}{({}^0\psi_I)^5} \tag{23}$$

whose statistical values are $Q = 0.053$, $F = 584$, $r = 0.985$, $s = 19$, $\mathbf{u} = (24, 28)$, and $\langle u\rangle = 26$. The correlation vector is $\mathbf{C} = (373.966, 406.508)$, and no mixed set $\{\chi, \psi\}$ combination offers a LCpCI with better modeling. The connectivity descriptors show, instead, the following modeling power:

$$\{{}^0\chi^v\}:\ Q = 0.015,\ F = 45,\ r = 0.846,\ s = 57,\ \langle u\rangle = 17$$

$$\{D^v, {}^0\chi^v\}:\ Q = 0.033,\ F = 115,\ r = 0.965,\ s = 29,\ \langle u\rangle = 19$$

$$\{{}^0\chi^v, {}^1\chi^v, D^z\}:\ Q = 0.043,\ F = 131,\ r = 0.980,\ s = 22,\ \langle u\rangle = 6.3$$

$$X_{\Delta H} = \frac{(D^v)^{0.5} + 0.2}{D^v + 4.2\,{}^0\chi^v} \tag{24}$$

TABLE 8: Experimental Lattice Enthalpies $\Delta H_L^{\varnothing}$ at 298.15 K (kJ mol^{-1}) for 20 Metal Halides, MeX, Their Corresponding Calculated (P_{calc}) and Calculated Jackknifing (P_{jk}) Values, and the I_S Molecular Pseudoconnectivity Values[a]

MeX	$\Delta H_L^{\varnothing}$	P_{calc}	P_{jk}	$^S\psi_I$	$^0\psi_I$	$^1\psi_I$	$^S\psi_E$	$^0\psi_E$	$^1\psi_E$
LiF	1037	1026	1018	10	1.06066	0.25	21	0.66645	0.10541
NaF	926	885	879	9.44	1.18689	0.29463	20.45	0.69097	0.11153
KF	821	828	828	9.25	1.24798	0.31623	20.25	0.70113	0.11406
RbF	789	796	797	9.16	1.28203	0.32827	20.16	0.70588	0.11525
CsF	750	778	780	9.11	1.30271	0.33558	20.11	0.70854	0.11592
LiCl	852	892	899	6.11	1.20037	0.34879	17.11	0.69281	0.11895
NaCl	786	775	774	5.55	1.32660	0.41105	16.56	0.71094	0.12446
KCl	717	728	729	5.36	1.38769	0.44119	16.36	0.71837	0.12670
RbCl	695	703	704	5.27	1.42174	0.45798	16.27	0.72183	0.12775
CsCl	678	689	689	5.22	1.44242	0.46819	16.22	0.72378	0.12834
LiBr	815	819	819	4.75	1.31013	0.42640	15.75	0.71408	0.12731
NaBr	752	722	721	4.19	1.43636	0.50252	15.20	0.73006	0.13269
KBr	689	683	682	4	1.49745	0.53936	15.01	0.73631	0.13477
RbBr	668	663	663	3.91	1.53150	0.55989	14.91	0.73969	0.13590
CsBr	654	652	651	3.86	1.55218	0.57236	14.86	0.74142	0.13647
LiI	761	763	763	4.12	1.39391	0.48564	15.12	0.72743	0.13229
NaI	705	683	681	3.56	1.52014	0.57234	14.57	0.74234	0.13760
KI	649	651	651	3.37	1.58123	0.61430	14.37	0.74847	0.13976
RbI	632	634	634	3.28	1.61528	0.63768	14.28	0.75136	0.14077
CsI	620	624	624	3.23	1.63596	0.65188	14.23	0.75299	0.14135

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

TABLE 9: Intrinsic State Value Matrices, I_S, and the Electrotopological E_S Matrices of the Five DNA/RNA Bases[a]

Bases (M)	I_S matrices	E_S matrices
A (135)	(3.0 2.0 3.0 1.67 1.67 3.0 2.0 2.5 / 1.67 0.0 0.0 4.0 ↵ 0.0 0.0 ↵)	(3.89 1.40 3.76 0.41 0.63 3.91 1.54 2.82 / 0.67 0.0 0.0 5.47 ↵ 0.0 0.0 ↵)
G (151)	(3.0 1.67 2.5 1.67 1.67 3.0 2.0 2.5 / 1.67 4.0 0.0 7.0 ↵ 0.0 0.0 ↵)	(3.80 0.09 2.33 −0.33 0.28 3.74 1.40 2.68 / 0.41 5.27 0.0 11.0 ↵ 0.0 0.0 ↵)
U (112)	(2.5 1.67 2.5 1.67 2.0 2.0 / ↦ 7.0 0.0 7.0 0.0 ↵)	(2.26 −0.47 2.00 −0.38 1.24 1.29 / ↦ 10.18 0.0 10.21 0.0 ↵)
T (126)	(2.5 1.67 2.5 1.67 1.67 2.0 / ↦ 7.0 0.0 7.0 2.0 ↵)	(2.33 −0.47 2.07 −0.33 0.51 1.38 / ↦ 10.33 0.0 10.56 1.62 ↵)
C (111)	(2.5 1.67 3.0 1.67 2.0 2.0 / ↦ 7.0 0.0 4.0 2.0 ↵)	(2.33 −0.41 3.43 0.24 1.51 1.45 / ↦ 10.24 0.0 5.13 0.0 ↵)

[a] Their molar masses, M, are given in parentheses.

The modeling power of this term is $Q = 0.037$, $F = 281$, $r = 0.969$, $s = 24$, $\mathbf{u} = (17, 65)$, and $\langle u \rangle = 41$. The u_2 utility of the $X_{\Delta H}$ terms improves over the u_2 utility of the $Y_{\Delta H}$ term but at some expense of the u_1 utility. No mixed $\{\chi, \psi\}$ index description is worth being cited for this property.

The Q-1-out method for this property of metal halides shows $\langle Q(X_{\Delta H}) \rangle = 0.037$, $Q(X_{\Delta H})_{MAX} = 0.042$ (NaBr-out), $Q(X_{\Delta H})_{MIN} = 0.035$ (LiF-out), and $\langle Q(Y_{\Delta H}) \rangle = 0.053$, $Q(Y_{\Delta H})_{MAX} = 0.062$ (LiCl- or KF-out), $Q(Y_{\Delta H})_{MIN} = 0.051$ (for many consecutively MeX-out). For what concerns the descriptive power of both terms $X_{\Delta H}$ and $Y_{\Delta H}$ neither of them seems to be consistently influenced by the exclusion of any metal halide. In the third column of Table 8 the calculated lattice enthalpy values, P_{calc}, with vector $\mathbf{C} = (373.966, 406.508)$ are shown, and along the fourth column of the same table the corresponding jackknifing values, P_{jk}, are shown. The two sets of values are quite similar, even more similar than the MON case.

Singlet Excitation Energies of DNA/RNA Bases. The first and second singlet excitation energies, ΔE_1 and ΔE_2, of DNA/RNA bases A, G, U, T, and C offer an interesting benchmark for the newly defined I_S molecular pseudoconnectivity indices. In fact these two dimensionally similar properties have recently[4] been modeled by a single and the same molecular connectivity term, whose form and statistical values are

$$X_{\Delta E} = \left(\frac{^0\chi}{\chi_t + 1000\chi_t^v} \right)^5 \tag{25}$$

ΔE_1: $Q = 8.96$, $F = 4.5$, $r = 0.790$, $s = 0.1$, $\mathbf{u} = (2.3, 107)$, $\langle u \rangle = 55$

ΔE_2: $Q = 6.85$, $F = 44$, $r = 0.967$, $s = 0.1_4$, $\mathbf{u} = (6.6, 86)$, $\langle u \rangle = 46$

From these results it is evident that $X_{\Delta E}$ is a good ΔE_2 descriptor, while the description of ΔE_1 is less satisfactory. The discrepancy in Q values (the best Q value for the worse r value) is due to the larger s value for ΔE_2, as the original s value for ΔE_1 equals 0.088. The interesting side of this modeling resides in the "broad" validity of this term, which is an optimal descriptor for both properties. Now, if pseudoconnectivity terms are used, not only the "broad" validity is maintained but is even en-

TABLE 10: Experimental First ΔE_1 and Second ΔE_2 Singlet Excitation Energies (eV) of the DNA/RNA Bases and Their Corresponding I_S Molecular Pseudoconnectivity Values[a]

base	ΔE_1	ΔE_2	${}^{S}\psi_I$	${}^{0}\psi_I$	${}^{1}\psi_I$	${}^{T}\psi_I$	${}^{S}\psi_E$	${}^{0}\psi_E$	${}^{1}\psi_E$	${}^{T}\psi_E$
A	4.75	5.99	24.51	6.60019	5.08618	0.01410	79.50	3.60497	1.49359	0.00004
G	4.49	5.03	30.68	7.09998	5.54098	0.00639	91.27	3.98554	1.71209	0.00001
U	4.81	6.11	26.34	4.98270	3.54757	0.01711	70.33	2.88573	1.16416	0.00024
T	4.67	5.94	28.01	5.75652	4.19355	0.01324	77.50	3.27130	1.32709	0.00009
C	4.61	6.26	23.84	5.04963	3.55672	0.02066	67.92	2.86843	1.12262	0.00025

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

TABLE 11: Intrinsic State Value Matrices, I_S, and the Electrotopological E_S Matrices of 32 Chlorofluorocarbons (CFCs)[a]

CFC	I_S matrices	E_S matrices	CFC	I_S matrices	E_S matrices
CH_3Cl	(2 4.11)	(1.47 4.64)	CH_2Cl_2	(4.11 1.5 4.11)	(4.76 0.19 4.76)
$CHCl_3$	$\begin{pmatrix}4.11 & 1.33 & 4.11\\ 0 & 4.11 & 0\end{pmatrix}$	$\begin{pmatrix}4.81 & -0.75 & 4.81\\ 0 & 4.81 & 0\end{pmatrix}$	CH_3F	(2 8)	(0.5 9.5)
CH_2F_2	(8 1.5 8)	(9.63 -1.75 9.63)	CHF_3	$\begin{pmatrix}8 & 1.33 & 8\\ 0 & 8 & 0\end{pmatrix}$	$\begin{pmatrix}9.67 & -3.67 & 9.67\\ 0 & 9.67 & 0\end{pmatrix}$
CH_2ClF	(4.11 1.5 8)	(4.33 -0.78 10.06)	$CHCl_2F$	$\begin{pmatrix}4.11 & 1.33 & 4.11\\ 0 & 8 & 0\end{pmatrix}$	$\begin{pmatrix}4.37 & -1.72 & 4.37\\ 0 & 10.53 & 0\end{pmatrix}$
$CHClF_2$	$\begin{pmatrix}8 & 1.33 & 8\\ 0 & 4.11 & 0\end{pmatrix}$	$\begin{pmatrix}10.1 & -2.69 & 10.1\\ 0 & 3.94 & 0\end{pmatrix}$	CH_3CH_2Cl	(2 1.5 4.11)	(1.89 0.72 5)
CH_3CH_2F	(2 1.5 8)	(1.46 -0.25 10.29)	CH_3CHCl_2	$\begin{pmatrix}2 & 1.33 & 4.11\\ 0 & 4.11 & 0\end{pmatrix}$	$\begin{pmatrix}1.7 & -0.22 & 5.04\\ 0 & 5.04 & 0\end{pmatrix}$
CH_3CHF_2	$\begin{pmatrix}2 & 1.33 & 8\\ 0 & 8 & 0\end{pmatrix}$	$\begin{pmatrix}0.83 & -2.17 & 10.33\\ 0 & 10.33 & 0\end{pmatrix}$	CH_2FCH_2F	(8 1.5 1.5 8)	(10.35 -0.85 -0.85 10.35)
$CH_2ClCHCl_2$	$\begin{pmatrix}4.11 & 1.5 & 1.33 & 4.11\\ 0 & 0 & 4.11 & 0\end{pmatrix}$	$\begin{pmatrix}5.07 & 0.31 & -0.41 & 5.1\\ 0 & 0 & 5.1 & 0\end{pmatrix}$	CH_2FCHF_2	$\begin{pmatrix}8 & 1.5 & 1.33 & 8\\ 0 & 0 & 8 & 0\end{pmatrix}$	$\begin{pmatrix}10.37 & -1.53 & -2.78 & 10.39\\ 0 & 0 & 10.39 & 0\end{pmatrix}$
CH_2ClCF_2Cl	$\begin{pmatrix}4.11 & 1.5 & 1.25 & 4.11\\ 0 & 0 & \underline{8} & 0\end{pmatrix}$	$\begin{pmatrix}4.25 & -0.82 & -3.22 & 4.25\\ 0 & 0 & \underline{11.08} & 0\end{pmatrix}$	CH_2ClCF_3	$\begin{pmatrix}4.11 & 1.5 & 1.25 & 8\\ 0 & 0 & \underline{8} & 0\end{pmatrix}$	$\begin{pmatrix}4.35 & -1.26 & -4.19 & 10.65\\ 0 & 0 & \underline{10.65} & 0\end{pmatrix}$
$CHFCF_3$	$\begin{pmatrix}8 & 1.5 & 1.25 & 8\\ 0 & 0 & \underline{8} & 0\end{pmatrix}$	$\begin{pmatrix}10.37 & -2.23 & -4.63 & 10.41\\ 0 & 0 & \underline{10.41} & 0\end{pmatrix}$	CHF_2CHF_2	$\begin{pmatrix}8 & 1.33 & 1.33 & 8\\ 0 & 8 & 8 & 0\end{pmatrix}$	$\begin{pmatrix}10.41 & -3.48 & -3.48 & 10.41\\ 0 & 10.41 & 10.41 & 0\end{pmatrix}$
$CHCl_2CF_3$	$\begin{pmatrix}4.11 & 1.33 & 1.25 & 8\\ 0 & 4.11 & \underline{8} & 0\end{pmatrix}$	$\begin{pmatrix}4.39 & -2.26 & -4.47 & 10.91\\ 0 & 4.39 & \underline{10.91} & 0\end{pmatrix}$	$CHClFCF_3$	$\begin{pmatrix}4.11 & 1.33 & 1.25 & 8\\ 0 & 8 & \underline{8} & 0\end{pmatrix}$	$\begin{pmatrix}3.96 & -3.23 & -4.90 & 10.67\\ 0 & 10.85 & \underline{10.67} & 0\end{pmatrix}$
CHF_2CF_3	$\begin{pmatrix}8 & 1.33 & 1.25 & 8\\ 0 & 8 & \underline{8} & 0\end{pmatrix}$	$\begin{pmatrix}10.42 & -4.20 & -5.33 & 10.43\\ 0 & 10.42 & \underline{10.43} & 0\end{pmatrix}$	CH_3CCl_3	$\begin{pmatrix}2 & 1.25 & 4.11\\ 0 & \underline{4.11} & 0\end{pmatrix}$	$\begin{pmatrix}1.48 & -1.08 & 5.06\\ 0 & \underline{5.06} & 0\end{pmatrix}$
CH_3CF_3	$\begin{pmatrix}2 & 1.25 & 8\\ 0 & \underline{8} & 0\end{pmatrix}$	$\begin{pmatrix}0.19 & -4.0 & 10.35\\ 0 & \underline{10.35} & 0\end{pmatrix}$	CH_3CF_2Cl	$\begin{pmatrix}2 & 1.25 & 4.11\\ 0 & \underline{8} & 0\end{pmatrix}$	$\begin{pmatrix}0.62 & -3.03 & 4.2\\ 0 & \underline{10.79} & 0\end{pmatrix}$
CH_2BrCF_3	$\begin{pmatrix}2.75 & 1.5 & 1.25 & 8\\ 0 & 0 & \underline{8} & 0\end{pmatrix}$	$\begin{pmatrix}2.24 & -0.92 & -4.04 & 10.74\\ 0 & 0 & \underline{10.74} & 0\end{pmatrix}$	$CHICF_3$	$\begin{pmatrix}2.12 & 1.5 & 1.25 & 8\\ 0 & 0 & \underline{8} & 0\end{pmatrix}$	$\begin{pmatrix}1.27 & -0.76 & -3.97 & 10.78\\ 0 & 0 & \underline{10.78} & 0\end{pmatrix}$
$CHFBrCF_3$	$\begin{pmatrix}2.75 & 1.33 & 1.25 & 8\\ 0 & 8 & \underline{8} & 0\end{pmatrix}$	$\begin{pmatrix}1.70 & -2.89 & -4.75 & 10.76\\ 0 & 11 & \underline{10.76} & 0\end{pmatrix}$	$CHFICF_3$	$\begin{pmatrix}2.12 & 1.33 & 1.25 & 8\\ 0 & 8 & \underline{8} & 0\end{pmatrix}$	$\begin{pmatrix}0.66 & -2.73 & -4.68 & 10.80\\ 0 & 11.07 & \underline{10.80} & 0\end{pmatrix}$
$CHClBrCF_3$	$\begin{pmatrix}2.75 & 1.33 & 1.25 & 8\\ 0 & 4.11 & \underline{8} & 0\end{pmatrix}$	$\begin{pmatrix}2.14 & -1.92 & -4.32 & 11\\ 0 & 4.55 & \underline{11} & 0\end{pmatrix}$	$CHBr_2CF_3$	$\begin{pmatrix}2.75 & 1.33 & 1.25 & 8\\ 0 & 2.75 & \underline{8} & 0\end{pmatrix}$	$\begin{pmatrix}2.29 & -1.58 & -4.17 & 11.08\\ 0 & 2.29 & \underline{11.08} & 0\end{pmatrix}$

[a] A number underlined in the second row has a symmetric companion on the top of the first row.

hanced; in fact following term shows enhanced statistics for both energies:

$$Y_{\Delta E} = \left(\frac{{}^{S}\psi_I - 4.8\,{}^{1}\psi_I}{{}^{T}\psi_I + 120\,{}^{T}\psi_E}\right)^4 \qquad (26)$$

ΔE_1: $Q = 9.34$, $F = 5.4$, $r = 0.802$, $s = 0.09$,

$$\mathbf{u} = (2.4, 105), \langle u \rangle = 54$$

ΔE_2: $Q = 8.62$, $F = 70$, $r = 0.979$, $s = 0.1$,

$$\mathbf{u} = (8.3, 103), \langle u \rangle = 56$$

Even here the discrepancy in Q values is due to the different s vales ($s = 0.085$ and 0.11, respectively). For comparison purposes it is to be noticed that the molar masses as descriptors rate, for ΔE_1, $Q = 4.9$, $F = 1.5$, and $r = 0.578$, and, for ΔE_2, $Q = 3.7$, $F = 13$, and $r = 0.899$. The last result for ΔE_2 is the only interesting modeling achieved by M in this study.

QSAR Studies of Chlorofluoroalkanes (CFCs) and of 2-Bromo-2-Phenethylamines (BrPhAm). *Rates of Hydrogen Abstraction.* Recently, it has been possible to simulate by the aid of the E-state index of the carbon atom bonding the largest number of hydrogen atoms, $S(1)$, and of two non-E-state descriptors, the ${}^{3}\chi_c$ molecular connectivity index, and the ${}^{1}\kappa_\alpha$ shape index[33] the rates of hydrogen abstraction in units of log K of 26 CFCs, that is, the effect of their reaction with the

TABLE 12: Rates of Hydrogen Abstraction (log *K*), $I_S-\psi$ Values for 26 CFCs[a] and Their Molar Mass (*M*)

CFC[b] (*M*)	log *K*	${}^{S}\psi_I$	${}^{0}\psi_I$	${}^{1}\psi_I$	${}^{T}\psi_I$	${}^{S}\psi_E$	${}^{0}\psi_E$	${}^{1}\psi_E$	${}^{T}\psi_E$
CCl (50.5)	7.36	6.11	1.20037	0.34879	0.34879	17.11	0.69281	0.11895	0.11895
CCl_2 (84.9)	8.00	9.72	1.80302	0.80550	0.19866	26.03	1.04637	0.26291	0.04122
CCl3 (119)	7.80	13.66	2.34690	1.28314	0.10407	35.68	1.39314	0.42869	0.01386
CF (34)	6.95	10.00	1.06066	0.25000	0.25000	21.00	0.66645	0.10541	0.10541
CF_2 (52)	6.81	17.50	1.52360	0.57735	0.10206	34.01	1.03057	0.26552	0.03413
CF_3 (70)	5.10	25.33	1.92777	0.91971	0.03832	47.34	1.50947	0.56938	0.01251
CClF (68.5)	7.46	23.61	1.66331	0.69142	0.14239	30.11	1.03275	0.26350	0.03722
CCl_2F (103)	7.30	17.55	2.20719	1.16200	0.07559	39.35	1.40229	0.45671	0.01310
$CClF_2$ (86.5)	6.45	21.44	2.06748	1.04085	0.05346	43.45	1.42839	0.49624	0.01245
CCCl (64.5)	8.37	7.61	2.01687	0.98010	0.28479	24.11	1.07743	0.27124	0.04552
CCF (48.1)	8.14	11.50	1.87716	0.86603	0.20412	28.00	1.06714	0.27526	0.04163
$CCCl_2$ (99)	8.20	11.55	2.56074	1.46857	0.14918	33.56	1.42391	0.43028	0.01539
CCF_2 (66.1)	7.48	19.33	2.28132	1.22628	0.07664	41.32	1.44814	0.49327	0.01376
CFCF (66.1)	7.83	19.00	2.34010	1.24402	0.08333	41.00	1.42984	0.44802	0.01357
$CClCCl_2$ (133)	8.28	15.16	3.16340	1.96617	0.08497	42.67	1.77999	0.58378	0.00534
$CFCF_2$ (84)	7.47	26.83	2.74427	1.60981	0.03129	54.25	1.87126	0.74484	0.00489
$CClCF_2Cl$ (135)	7.20	23.97	3.86045	2.20669	0.04274	59.48	2.25853	0.99310	0.00191
$CClCF_3$ (119)	6.95	30.86	3.26485	2.08173	0.01592	63.85	2.42448	1.23127	0.00208
$CFCF_3$ (102)	6.70	34.75	3.12514	1.96766	0.01141	67.74	2.62825	1.53805	0.00235
CF_2CF_2 (102)	6.50	34.66	3.14843	1.97816	0.01175	67.68	2.41002	1.20063	0.00196
CCl_2CF_3 (153)	7.40	34.80	3.80872	2.57968	0.00834	73.28	2.91742	1.63042	0.00083
$CClFCF_3$ (137)	6.87	38.69	3.66901	2.45853	0.00598	77.79	3.26865	2.19995	0.00104
CF_2CF_3 (120)	6.48	42.58	3.52930	2.33739	0.00428	81.10	4.55532	4.38982	0.00210
$CCCl_3$ (133)	6.80	15.58	3.08133	1.95602	0.07590	43.08	1.77734	0.61915	0.00525
CCF_3 (84)	5.95	27.25	2.66219	1.58114	0.02795	54.74	1.98926	0.95756	0.00542
CCF_2Cl (101)	6.60	23.36	2.80190	1.70610	0.03900	50.78	1.85862	0.77775	0.00509

[a] ψ_E values have been obtained after a rescaling procedure (see the Method section). [b] Only heteroatoms are written (see the first column of Table 11).

TABLE 13: $I_S-\psi$ (and *M*) Values for Six CFCs Not in Table 12 and Used for the MAC Simulation

CFC[a] (*M*)	${}^{S}\psi_I$	${}^{0}\psi_I$	${}^{1}\psi_I$	${}^{T}\psi_I$	${}^{S}\psi_E$	${}^{0}\psi_E$	${}^{1}\psi_E$	${}^{T}\psi_E$
$CBrCF_3$ (126.9)	29.50	3.37461	2.17315	0.01941	62.14	2.52461	1.32327	0.00245
$CICF_3$ (209.9)	28.87	3.45839	2.39975	0.02217	61.88	2.39562	1.14897	0.00217
$CFBrCF_3$ (180.9)	37.33	3.77877	2.55371	0.00731	75.84	3.13653	1.95688	0.00100
$CFICF_3$ (227.9)	36.70	3.86255	2.62635	0.00832	75.22	3.09680	1.87379	0.00100
$CClBrCF_3$ (197.4)	33.44	3.91848	2.67485	0.01020	71.95	2.86487	1.52435	0.00083
CBr_2CF_3 (241.8)	32.08	4.02824	2.77002	0.01246	70.57	2.82553	1.43874	0.00083

[a] Only heteroatoms are written (see the first column of Table 11).

hydroxyl radical. In Table 11 are collected the two-row matrices or vectors of 32 CFCs, where the last 6 CFCs are used to model the next activity. From these matrices *S*(1) values can easily be read, but they are also reported in ref 32. In Table 12 are, instead, collected the 26 $\{\psi\}$ values together with the log *K* values. It is here interesting to stress that the subsets of monosubstituted, bisubstituted, and trisubstituted CFC compounds at the same carbon atom and at two different carbon atoms share the same value for each of the four molecular connectivity indices of subset $\{D, {}^{0}\chi, {}^{1}\chi, {}^{t}\chi\}$. The modeling achieved by the linear combination $\{S(1), {}^{3}\chi_c, {}^{1}\kappa_\alpha\}$ is quite good with $Q = 3.47$, $F = 60$, $r = 0.943$, $s = 0.3$, $\mathbf{u} = (10, 11, 9.4, 3.9)$, and $\langle u\rangle = 8.6$. Let us start saying that pseudoconnectivity indices (see Table 12) are not able to achieve a better modeling with any linear combination of three indices or less, but things are completely different at the level of a single descriptor. The best single descriptor among the three given indices is *S*(1), with $Q = 0.861$, $F = 11$, $r = 0.560$, $s = 0.7$, $\mathbf{u} = (3.3, 53)$, and $\langle u\rangle = 28$. Now, the following I_S pseudoconnectivity term has the features of an overall dominant descriptor for this activity

$$Y_{\log K} = \left(\frac{{}^{S}\psi_E - 1.06\,{}^{S}\psi_I}{{}^{0}\psi_I}\right)^6 \qquad (27)$$

with $Q = 2.03$, $F = 61$, $r = 0.847$, $s = 0.4$, $\mathbf{u} = (7.7, 41)$, $\langle u\rangle = 24$. This is not the best overall term that can be found by a trial-and-error procedure; a more convoluted term, such as the one of eq 28, offers a somewhat better description with $Q = 2.07$, $F = 63$, $r = 0.852$, $s = 0.4$, $\mathbf{u} = (8, 41)$, $\langle u\rangle = 24$

$$Y'_{\log K} = \left(\frac{{}^{S}\psi_E - 1.1\,{}^{S}\psi_I - {}^{T}\psi_I}{{}^{0}\psi_I}\right)^{6.5} \qquad (28)$$

Clearly, the small improvement due to the second term does not decrease the importance, due to its simplicity (kind of Occam's razor) of the term of eq 27, and this goes also to show how the search procedure for the terms works. Interestingly enough is the fact that a linear combination of term 28 with the two aforementioned indices achieves a modeling similar to $\{S(1), {}^{3}\chi_c, {}^{1}\kappa_\alpha\}$

$$\{Y'_{\log K}, {}^{3}\chi_c, {}^{1}\kappa_\alpha\}: Q = 3.49, F = 60, r = 0.944, s = 0.3, \mathbf{u} = (11, 5.8, 4.5, 31), \langle u\rangle = 13$$

Only u_2-u_4 values differ consistently along the two combinations. The interrelation between $\{Y'_{\log K}, {}^{3}\chi_c, {}^{1}\kappa_\alpha\}$ indices is quite low with $r(Y'_{\log K}, {}^{3}\chi_c) = 0.28$, $r(Y'_{\log K}, {}^{1}\kappa_\alpha) = 0.20$, and $r({}^{3}\chi_c, {}^{1}\kappa_\alpha) = 0.80$. While the $\{Y'_{\log K}, {}^{3}\chi_c, {}^{1}\kappa_\alpha\}$ combination achieves a somewhat better modeling than the $\{S(1), {}^{3}\chi_c, {}^{1}\kappa\}$ combination, the combination $\{Y_{\log K}, {}^{3}\chi_c, {}^{1}\kappa_\alpha\}$, instead, has $Q = 3.42$ and $F = 58$, only a bit worse than the $\{S(1), {}^{3}\chi_c, {}^{1}\kappa\}$ combination. Molar masses, again, rate quite poorly with $Q = 0.042$, $F = 0.03$, and $r = 0.033$.

TABLE 14: Logarithm of the Minimum Anesthetic Concentration (MAC) of 11 Trifluoromethyl Ethanes (tFMeE), Calculated log MAC, P_{calc}, and Calculated log MAC, P_{jk}, with the Leave-One-Out or Jackknifing Method[a]

tFMeE	log MAC	P_{calc}	P_{jk}
CH_3CF_3	1.60	1.60	1.60
CH_2ClCF_3	0.90	0.89	0.89
CH_2BrCF_3	0.45	0.39	0.39
CH_2ICF_3	0.10	0.12	0.13
CHF_2CF_3	1.70	1.71	1.71
$CHFClCF_3$	1.18	1.13	1.12
$CHFBrCF_3$	0.70	0.70	0.70
$CHFICF_3$	0.30	0.44	0.45
$CHCl_2CF_3$	0.43	0.38	0.37
$CHClBRCF_3$	−0.10	−0.03	−0.02
$CHBr_2CF_3$	−0.40	−0.47	−0.51

[a] Compounds are taken from Table 13 and some from Table 12.

Minimum Anesthetic Concentrations. Even if it has not been simulated, it has nevertheless been suggested that the minimum anesthetic concentration (MAC), in log units, of 11 trifluoromethylethanes could optimally be modeled with the *S*(F) values of the three fluoro atoms bonded to the same carbon atom.[34] The $\mathbf{E_S}$ and $\mathbf{I_S}$ matrices of these 11 compounds are collected in Table 11, and from them it is possible to read the *S*(F) values. In Table 13 are collected the pseudoconnectivity values of the six remaining CFCs included in the log MAC modeling but not taking part in the log *K* modeling. Finally, in Table 14 are collected the log MAC experimental values of the 11 compounds. Unexpectedly, the molar mass is a discrete descriptor for this property with $Q = 2.24$, $F = 23$, and $r = 0.847$, but the *S*(F) index achieves a very good improvement with $Q = 4.05$, $F = 74$, $r = 0.944$, $s = 0.2$, $\mathbf{u} = (8.6, 8.8)$, and $\langle u \rangle = 8.7$. While no single I_S molecular pseudoconnectivity index can compete with *S*(F) as a single descriptor, the following linear combination of two indices, instead, seems to be the optimal solution:

$$\{{}^0\psi_I, {}^S\psi_E\}: Q = 13.6, F = 419, r = 0.995, s = 0.1, \mathbf{u} = (29, 22, 15), \langle u \rangle = 22$$

The very good values of *Q*, *F*, and **u** seem to leave no doubt about the quality of this combination, whose two parameters are not so strongly correlated with $r({}^0\psi_I, {}^S\psi_E) = 0.73$. Notice that the best single pseudoindex, $\{{}^0\psi_I\}$, rates only $Q = 1.23$ and $F = 6.9$. But, to model 11 points with 2 parameters is a rather uneasy choice. A trial-and-error procedure discovers the following excellent $I_S - \psi$ term, with $Q = 15.3$, $F = 1065$, $r = 0.996$, $s = 0.06$, $\mathbf{u} = (33, 35)$, and $\langle u \rangle = 34$:

$$Y_{\log\ MAC} = \frac{{}^1\psi_I + {}^T\psi_I}{{}^S\psi_E} \qquad (29)$$

The correlation vector of the vector term $\mathbf{Y} = (Y_{\log\ MAC}, U_0)$, which allows the P_{calc} values given in Table 14 to be derived with the equation log MAC = $\mathbf{C}\cdot\mathbf{Y}$ (without bars, as two values are negative), is $\mathbf{C} = (-206.258, 7.66399)$. The *Q*-leave-one-out method and the jackknifing method both reveal the great stability of the modeling, as (i) P_{jk} values (see Table 14, last column) are nearly identical to the original calculated P_{calc} values, (ii) the **C** vector is highly constant with an error of ±2% for each parameter, and (iii) *Q* values are also highly constant, with the only exception for compound $CFICF_3$, which, when it is left out, gives rise to $Q = 22$, $F = 2155$, and $r = 0.998$. It is to be noticed that the term of eq 29 has been chosen following "Occam's razor" considerations. A more convoluted term, where

TABLE 15: Intrinsic State Value Matrices, I_S, and the Electrotopological E_S Matrices of $n = 22$ 2-Bromo-2-Phenethylamines (NCCBrPhY$_p$Z$_m$)[a]

Y_p / Z_m	I_S matrices	E_S matrices
H / H	4 1.5 1.33 1.67 2 2 2 2 2 0 0 2.75 ↦ 0 0 0 0 ↵	5.46 0.64 0.30 1.24 2.06 2.01 2.00 2.01 2.06 0 0 3.46 ↦ 0 0 0 0 ↵
F / H	4 1.5 1.33 1.67 2 2 1.67 2 2 0 0 2.75 ↦ 0 0 8 0 ↵	5.40 0.52 0.13 1.01 1.72 1.43 −0.22 1.43 1.72 0 0 3.36 ↦ 0 0 12.41 0 ↵
Cl / H	4 1.5 1.33 1.67 2 2 1.67 2 2 0 0 2.75 ↦ 0 0 4.11 0 ↵	5.46 0.60 0.24 1.16 1.97 1.86 0.76 1.86 1.97 0 0 3.41 ↦ 0 0 5.71 0 ↵
Br / H	4 1.5 1.33 1.67 2 2 1.67 2 2 0 0 2.75 ↦ 0 0 2.75 0 ↵	5.48 0.63 0.27 1.22 2.05 2.01 1.10 2.01 2.05 0 0 3.47 ↦ 0 0 3.37 0 ↵
I / H	4 1.5 1.33 1.67 2 2 1.67 2 2 0 0 2.75 ↦ 0 0 2.12 0 ↵	5.49 0.64 0.29 1.24 2.09 2.08 1.25 2.08 2.09 0 0 3.48 ↦ 0 0 2.29 0 ↵
Me / H	4 1.5 1.33 1.67 2 2 1.67 2 2 0 0 2.75 ↦ 0 0 2 0 ↵	5.50 0.64 0.29 1.25 2.10 2.10 1.28 2.10 2.10 0 0 3.48 ↦ 0 0 2.09 0 ↵
H / F	4 1.5 1.33 1.67 2 2 2 1.67 2 0 0 2.75 ↦ 0 0 0 8 ↵	5.38 0.48 0.05 0.88 1.84 1.68 1.42 −0.22 1.48 0 0 3.33 ↦ 0 0 0 12.60 ↵
H / Cl	4 1.5 1.33 1.67 2 2 2 1.67 2 0 0 2.75 ↦ 0 0 0 4.11 ↵	5.46 0.58 0.21 1.13 2.00 1.92 1.85 0.75 1.91 0 0 3.44 ↦ 0 0 0 5.78 ↵
H / Br	4 1.5 1.33 1.67 2 2 2 1.67 2 0 0 2.75 ↦ 0 0 0 2.75 ↵	5.49 0.62 0.26 1.21 2.05 2.00 2.00 1.09 2.06 0 0 3.47 ↦ 0 0 0 3.40 ↵
H / I	4 1.5 1.33 1.67 2 2 1.67 1.67 2 0 0 2.75 ↦ 0 0 0 2.12 ↵	5.50 0.64 0.29 1.25 2.08 2.04 2.07 1.25 2.13 0 0 3.49 ↦ 0 0 0 2.29 ↵
H / Me	4 1.5 1.33 1.67 2 2 2 1.67 2 0 0 2.75 ↦ 0 0 0 2 ↵	5.51 0.64 0.29 1.26 2.08 2.05 2.08 1.28 2.15 0 0 3.49 ↦ 0 0 0 2.08 ↵
F / Cl	4 1.5 1.33 1.67 2 2 1.67 167 2 0 0 2.75 ↦ 0 0 8 4.11 ↵	5.41 0.46 0.04 0.89 1.66 1.33 −0.41 0.13 1.57 0 0 3.34 ↦ 0 0 12.69 5.58 ↵
F / Br	4 1.5 1.33 1.67 2 2 1.67 1.67 2 0 0 2.75 ↦ 0 0 8 2.75 ↵	5.44 0.50 0.09 0.98 1.71 1.42 −0.25 0.47 1.72 0 0 3.38 ↦ 0 0 12.77 3.11 ↵
F / Me	4 1.5 1.33 1.67 2 2 1.67 167 2 0 0 2.75 ↦ 0 0 8 2 ↵	5.45 0.52 0.12 1.02 1.74 1.47 −0.17 0.66 1.81 0 0 3.40 ↦ 0 0 12.82 1.75 ↵
Cl / Cl	4 1.5 1.33 1.67 2 2 1.67 1.67 2 0 0 2.75 ↦ 0 0 4.11 4.11 ↵	5.47 0.54 0.15 1.05 1.90 1.72 0.57 0.56 1.82 0 0 3.42 ↦ 0 0 5.75 5.82 ↵
Cl / Br	4 1.5 1.33 1.67 2 2 1.67 1.67 2 0 0 2.75 ↦ 0 0 4.11 2.75 ↵	5.50 0.58 0.20 1.13 1.96 1.85 0.71 0.90 1.97 0 0 3.45 ↦ 0 0 5.84 3.35 ↵
Cl / Me	4 1.5 1.33 1.67 2 2 1.67 167 2 0 0 2.75 ↦ 0 0 4.11 2 ↵	5.51 0.60 0.23 1.18 1.99 1.90 0.80 1.09 2.05 0 0 3.48 ↦ 0 0 5.88 1.99 ↵
Br / Cl	4 1.5 1.33 1.67 2 2 1.67 1.67 2 0 0 2.75 ↦ 0 0 2.75 4.11 ↵	5.49 0.57 0.18 1.10 1.99 1.92 0.91 0.71 1.90 0 0 3.44 ↦ 0 0 3.32 5.91 ↵
Br / Br	4 1.5 1.33 1.67 2 2 1.67 1.67 2 0 0 2.75 ↦ 0 0 2.75 2.75 ↵	5.42 0.61 0.24 1.19 2.04 2.00 1.06 1.05 2.05 0 0 3.48 ↦ 0 0 3.41 3.44 ↵
Br / Me	4 1.5 1.33 1.67 2 2 1.67 167 2 0 0 2.75 ↦ 0 0 2.75 2 ↵	5.53 0.63 0.27 1.23 2.07 2.05 1.14 1.24 2.14 0 0 3.50 ↦ 0 0 3.45 2.07 ↵
Me / Me	4 1.5 1.33 1.67 2 2 1.67 167 2 0 0 2.75 ↦ 0 0 2 2 ↵	5.55 0.64 0.29 1.26 2.12 2.13 1.33 1.32 2.18 0 0 3.52 ↦ 0 0 2.12 2.12 ↵
Me / Br	4 1.5 1.33 1.67 2 2 1.67 1.67 2 0 0 2.75 ↦ 0 0 2 2.75 ↵	5.53 0.62 0.26 1.22 2.09 2.09 1.25 1.14 2.10 0 0 3.50 ↦ 0 0 2.07 3.48 ↵

[a] I_S and E_S values of Br, Y_p, and Z_m are in the second row.

${}^0\psi_I + 0.4{}^S\psi_E$ replaces ${}^S\psi_E$ in the denominator and where the whole expression is elevated to the 0.8 power, shows a somewhat improved statistical value, with $Q = 16.1$, $F = 1171$, $r = 0.996$, $s = 0.06$, and $\mathbf{u} = (34, 36)$.

Antagonism of Adrenalin by 2-Bromo-2-Phenethylamines (BrPhAm). The hydrogen-suppressed formula of BrPhAm derivatives is NCCBrPhY$_p$Z$_m$, where Y_p and Z_m are substituents in para (p) and meta (m) positions of the phenyl Ph ring.[35] Their $\mathbf{I_S}$ and $\mathbf{E_S}$ matrices and corresponding ψ value are collected in Tables 15 and 16, respectively, while their molar mass, *M*,

TABLE 16: I_S Molecular Pseudoconnectivity Values for $n = 22$ 2-Br-2-Phenethylamines[a] Whose Heteroatomic Formula is $NCCBrPhY_pZ_m$

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	0.112980	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	0.219480	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 the Method section).

TABLE 17: Antagonism of Adrenalin by 2-Br-2-Phenethylamines ($NCCBrPhY_pZ_m$) in PED_{50} Units, Their Molar Mass M Values, Calculated pED_{50}(calc) Values, and Percent Residual Relative to the Experimental pED_{50} Value, $\Delta\% = |(P_{ex} - P_{calc})| \times 100/P_{exp}$

Y_p/Z_m	M	pED_{50}	pED_{50}(calc)	$\Delta\%$
H/H	200.1	7.46	7.45	0.1_4
F/H	218.1	8.16	8.18	0.2
Cl/H	234.5	8.68	8.29	4.5
Br/H	279	8.89	8.58	3.5
I/H	326	9.25	9.17	0.8
Me/H	214	9.30	8.79	5.5
H/F	218.1	7.52	7.55	0.4
H/Cl	234.5	8.16	8.28	1.5
H/Br	279	8.30	8.58	3.3
H/I	326	8.40	8.75	4.2
H/Me	214	8.46	8.79	3.9
F/Cl	252.5	8.19	8.34	1.9
F/Br	297	8.57	8.62	0.6
F/Me	232	8.82	8.83	0.06
Cl/Cl	269	8.89	8.77	1.3
Cl/Br	313.4	8.92	8.97	0.6
Cl/Me	248.6	8.96	9.12	1.8
Br/Cl	313.4	9.00	9.17	1.9
Br/Br	357.9	9.35	9.14	2.3
Br/Me	293	9.22	9.27	0.6
Me/Me	228.1	9.30	9.39	1.0
Me/Br	293	9.52	9.27	2.6

values and their adrenalin antagonism measure in pED_{50} units, which measures the biological response as the negative logarithm of the estimated dose for 50% response, are given in Table 17. Many of these molecules not only show the same subset of $\{\chi\}$ values but also the same subset of $\{\chi^v\}$ values, i.e., interchanging the place of substituents Y_p and Z_m not only χ-type indices do not change, but also the χ^v-type indices are not altered. Further, χ-type indices also stay constant whenever Y_p or Z_m is replaced by a substituent different from H. This means that the general $\{\chi\}$ representation of these molecules is highly redundant, with the consequence of a poor modeling. Kier and Hall achieved the following optimal modeling of pED_{50} with a set of three E-state indices: $Q = 4.75$, $F = 49$, $r = 0.95$, $s = 0.2_0$, $\langle u\rangle = 5.8$, $\mathbf{u} = (1.0, 11, 11, 0.3)$. The single E-state index, instead, achieved a rather bad modeling with $Q = 0.75$, $F = 4$, $r = 0.4$, $s = 0.5$, and $\langle u\rangle = 3.8$.[35] The modeling achieved by $I_S-\psi$ indices is much more satisfactory at the level of the single index with the exclusion of the utility $\langle u\rangle$ value

$$\{^0\psi_I\}: Q = 2.17, F = 32, r = 0.785, s = 0.36, \langle u\rangle = 2.9$$

But it is with the following pseudoconnectivity term that the $I_S-\psi$ modeling becomes optimal and in some insight even better than the three E-state indices modeling

$$Y_{pED} = [^0\psi_I + (1.1^T\psi_I)^{0.9} - {}^0\psi_E - 3.3]^{-0.01} \qquad (30)$$

This term has $Q = 4.00$, $F = 108$, $r = 919$, $s = 0.23$, $\langle u\rangle = 11$, and $\mathbf{u} = (10.5, 11)$. The correlation vector by the aid of which the calculated values of Table 17 have been obtained is $\mathbf{C} = (-114.643, 124.083)$. In the last column of this table are reported the percentage residuals relative to the experimental value, from which the level of the achieved modeling is evident. Notice (i) the low average interrelation value for the overall $I_S-\psi$ indices with $\langle r\rangle = 0.647$, (ii) the maximum and minimum r values, $r_{max}(^0\psi_I, {}^1\psi_I) = 0.999_4$ and $r_{min}(^T\psi_I, {}^T\psi_E) = 0.162$, and (iii) the bad modeling power of M with $Q = 0.994$, $F = 6.7$, and $r = 0.50$.

Before concluding let us come back to the sign problem underlined in the Method section and solved with the rescaling procedure. Tentatively, to introduce a negative sign at the ψ level, that is, at the S_i, $(S_i)^{-0.5}$, $(S_iS_j)^{-0.5}$, and $^T\psi_E$ levels every time S_i is negative (i.e., $-(S_i)^{-0.5}$, and so on; see eqs 3−6) has a very negative consequence for the modeling. Let us here also meet the criticism that properties modeled with descriptors are adimensional properties, and in fact they are, as what is modeled is the adimensional ratio P/P^0, where P^0 is the unitary property with unit value.[36]

Conclusion

The electrotopological state (E_S) model for atoms in molecules originated in 1990−1991,[24,25] and it aimed to derive structure information at the atomic level to use in QSPR/QSAR studies. Actually, the I-state molecular pseudoconnectivity indices here developed, and based on the electrotopological model, compact

again the information at the molecular level and try to use this information for QSAR/QSPR studies. Through the E_S model the pseudodescriptors are not only indirectly related to graph and pseudograph properties, but also enclose electronic information, a fact that could render them interesting rivals of other graph theoretical descriptors.

It has been said that "anything can be defined provided it is consistent: usefulness is the key";[37] well, I_S pseudoconnectivity descriptors are both consistent and useful. In fact, the molecular pseudoconnectivity descriptors defined and used in this work possess quite interesting modeling aptitudes. They are discrete but poorer descriptors than χ indices of amino acid properties, even if the melting point and crystal density of amino acids are better but not optimally described by molecular pseudoconnectivity terms. It is with the motor octane number and melting points of alkanes and the lattice enthalpy of inorganic salts that molecular pseudoconnectivity terms achieve an optimal description, nearly doubling the descriptive power of the corresponding molecular connectivity terms, a fact which is probably due to their capacity to model electronic characteristics at a finer level than normally done by graph and/or pseudograph invariants. This last characteristic can, surely, explain the enhanced modeling of the singlet energies of DNA/RNA bases achieved by a single I_S pseudoconnectivity term. Good mixed $\{\chi, \psi\}$ descriptions have been detected with the solubility of amino acids, with the specific rotations, and especially with the crystal density of amino acids, which overrates any other description. A mixed description for the melting points of alkanes can also be mentioned, even if it rates poorer than a corresponding two-χ index LCCI. The fact that mixed $\{\chi, \psi\}$ descriptions are practically never much better than the homogeneous descriptions seems to overrule the use of such mixed sets for future modeling. The interesting modeling power of pseudoconnectivity descriptors is underlined by three QSAR studies: the rate of extraction and the minimum anesthetic concentration of CFCs, and the antagonism of adrenalin by 2-Br-2-phenethylamines. Here I_S pseudoconnectivity descriptors offer an interesting alternative to the solution of the modeling problem. Further, pseudoconnectivity indices seem to be optimal descriptors of the molar masses of amino acids ($^0\psi_E$), and, by far, of the molar mass of alkanes ($^1\psi_E$).

The trend already detected with molecular connectivity indices, that is, the much better quality of higher-level trial-and-error descriptors or terms, is here thoroughly confirmed for pseudoconnectivity terms also. The construction of this term allows an enormous freedom to be achieved with a small set of indices, with the consequence that single molecular pseudoconnectivity terms rate normally better than any linear combinations of molecular pseudoconnectivity indices (LCpCI).

What remains to be seen is the modeling power of I_S molecular pseudoconnectivity descriptors relative to a set of compounds which strongly differ from each other by van der Waals interactions and/or hydrogen bonds. Would semiempirical molecular pseudoconnectivity terms fill the gap as their counterparts the semiempirical molecular connectivity terms, already did?[6]

Facing the global modeling capability of (δ, δ^v)-related descriptors, a sentiment of amazement cannot be hidden; it is surely something more than a strange coincidence, as, for example, it is held by astronomers the strange Bode law[38] which explains all the distances of the planets and planetoids from the sun with the exclusion of Neptune and Pluto.

Acknowledgment. I thank Professors L. B. Kier of the Virginia Commonwealth University, Richmond, VA, and L. H. Hall of the Eastern Nazarene College, Quincy, MA, for their constant support and help.

References and Notes

(1) Pogliani, L. *J. Chem. Inf. Comput. Sci*. **1996**, *36*, 1082.
(2) Pogliani, L. *J. Phys. Chem*. **1996**, *100*, 18065.
(3) Pogliani, L. *Med. Chem. Res*. **1997**, *7*, 380.
(4) Pogliani, L. *J. Chem. Inf. Comput. Sci*. **1999**, *39*, 104.
(5) Pogliani, L. *J. Mol. Struct.: THEOCHEM* **1999**, *466*, 1.
(6) Pogliani, L. *J. Phys. Chem*. **1999**, *103*, 1598.
(7) Balaban, A. T., Ed. *Chemical applications of Graph theory*; Academic Press: London, 1986.
(8) Trinajstič, N. *Chemical graph theory*, 2nd ed., CRC Press: Boca Raton, FL, 1992.
(9) Reinhard, M.; Drefahl, A. *Handbook for Estimating Physicochemical Properties of Organic Compounds*; Wiley: New York, 1999.
(10) Basak, S. C.; Magnuson, V. R.; Niemi, G. J.; Regal, R. R. *Discr. Appl. Math*. **1988**, *19*, 17.
(11) Basak, S. C.; Niemi, G. J.; Veith, G. D. *J. Math. Chem*. **1991**, *7*, 243.
(12) Katrizky, A. R.; Mu, L. *J. Chem. Inf. Comput. Sci*. **1998**, *38*, 28, 293 and 300.
(13) Randič, M.; Basak, S. Variable Molecular Descriptors. In *Some Aspects of Mathematical Chemistry*; Sinha, D. K., Basak, S. C., Mohanty, R. K., Basumallik, I. N., Eds. In press.
(14) Gutman, I.; Popovic, Lj; Estrada, E.; Bertz, S. H. *ACH−Models Chem*. **1998**, *38*, 819.
(15) Amič, D.; Beslo, D.; Lucič, B.; Nikolič, S.; Trinajstic, N. *J. Chem. Inf. Comput. Sci*. **1998**, *38*, 819.
(16) Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*, Wiley: New York, in press.
(17) Basak, S.; Balaban, T.; Grunwald, G. D. *J. Chem. Inf. Comput. Sci*. To be published.
(18) Basak, S.; Nikolič, S.; Trinajstič, N.; Amič, D.; Beslo, D. *J. Chem. Inf. Comput. Sci*. To be published.
(19) Rosen, K. H. *Discrete mathematics and its applications*; McGraw-Hill: New York, 1995.
(20) Kier, L. B.; Hall L. H. *Molecular connectivity in structure−activity analysis*; Wiley: New York, 1986.
(21) Randič, M. *J. Am. Chem. Soc*. **1975**, *97*, 6609.
(22) Pogliani, L. *Croat. Chem. Acta* **1997**, *70*, 803.
(23) "The significance of a law of nature ... is the finding of an invariant, that is to say, a quantity which remains unchanged even when all the other determining elements vary within the possible limit imposed by the law. Thus, we perceive that the historical development of scientific concepts is ever associated with the discovering and working out of such invariants." Ostwald, W. In *The Question of the Atom: from the Karlsruhe Congress to the First Solvay Conference*, 1860−1911; Nye, M. J., Ed.; Tomash: Los Angeles, 1986.
(24) Kier, L. B.; Hall, L. H. *Pharm. Res*. **1990**, *7*, 801.
(25) Hall, L. H.; Mohney, B.; Kier, L. B. *J. Chem. Inf. Comput. Sci*. **1991**, *31*, 76.
(26) Kier, L. B.; Hall, L. H. *Molecular Structure Description. The Electrotopological State*; Academic Press: New York, 1999 (see also references therein).
(27) Kier, L. B.; Hall, L. H. *Molecular Structure Description. The Electrotopological State*; Academic Press: New York, 1999; pp 65−67 and 230.
(28) Estrada, E. *J. Chem. Inf. Comput. Sci*. **1990**, *35*, 31.
(29) Nikolic, S.; Trinajstic, N.; Ivanis, S. *Croat. Chim. Acta* **1999**, *72*, 875.
(30) Randic, M. *Int. J. Quantum Chem.: Quantum Biol. Symp*. **1994**, *21*, 215.
(31) Pogliani, L. *J. Phys. Chem*. **1993**, *97*, 6731.
(32) Gil, V. M. S.; Geraldes, C. F. G. C. *Ressonância Magnética Nuclear*; Fundação Calouste Gunbelkian: Lisboa, 1987.
(33) Kier, L. B.; Hall, L. H. *Molecular Structure Description. The Electrotopological State*; Academic Press: New York, 1999; pp 170−172.
(34) Kier, L. B.; Hall, L. H. *Molecular Structure Description. The Electrotopological State*; Academic Press: New York, 1999; pp 163−164.
(35) Kier, L. B.; Hall, L. H. *Molecular Structure Description. The Electrotopological State*; Academic Press: New York, 1999; pp 90−93.
(36) Berberan-Santos, M. N.; Pogliani, L. *J. Math. Chem*. **1999**, *26*, 255
(37) Garrett, A. J. M. In *Maximum Entropy in Action*, Buck, B., Macaulay, V. A., Eds. Clarendon Press: Oxford, 1992; p 142.
(38) Bode's law proposed by the 18th century astronomer Titus-Bode held that the distances of all known planets from the sun (inclusive planetoids between Mars and Jupiter) in terms of Earth's distance followed from a sequence of doubled numbers: 0, 3, 6, 12, 24, ... Adding 4 to each number and dividing by 10 gives the desired distance. This law fails totally for Neptune and Pluto and shows some deviations for Saturn and Uranus. Many astronomers held Bode's law for a strange coincidence. Gamow, G. *One, Two, Three.....Infinite*; Dover: New York, 1988; pp 307−308.