Molecular Topology. 15. 3D Distance Matrices and Related Topological Indices

Mircea V. Diudea,*,† Dragos Horvath,† and Ante Graovac‡

Babes-Bolyai University, 3400 Cluj, Romania, and Ruder Boskovic' Institute Zagreb, 41001 Zagreb, Croatia

Received September 13, 1993⊗

3D-Metric distances supplied by the MMX calculations are used to build 3D and LM3D matrices[1,2] in a full analogy with the construction of the distance matrices D from the topological distances. The 3D matrices are used as a basis for the construction of two types of topological indices, *c* (centricity) and *x* (centrocomplexity), and the interconversion of the new indices is studied. Similar *c* and *x* indices are derived on the ground of LM3D matrices. It is shown that these indices are useful in the QSPR/QSAR studies.

INTRODUCTION

The construction of topological indices (TIs), as unique descriptors characterizing molecular graphs and suitable for QSAR/QSPR purposes, evolved during the years from the integer local vertex invariants (LOVIs) to real-number LOVIs.[1] Indeed, matrices making a basis for LOVIs attributing stage were defined over the real numbers.[3] Layer arrays[1,2,4,5] were also considered in order to devise real number LOVIs and TIs.

A layer matrix, LM, collects the properties (topological or chemical) of vertices u located on concentric shells (layers), $G(u)_j$, at distance j around the vertex i in the graph G, and can be defined[1,2,6] as

$$lm_{ij} = \sum_{u \in G(u)_j} m_u \qquad (1)$$

$$LM(G) = \{lm_{ij};\ i \in [1,n];\ j \in [0,d]\} \qquad (2)$$

where m and M are labels for a given property and the corresponding matrix, respectively; n is the number of vertices in the graph; and d stands for the diameter of the graph, (i.e., the largest topological distance in the graph).

A particular interest has been devoted to LOVIs capable of "seeing" the total graph environment of each vertex. This goal can be reached either by using quadratic[3] or layer matrices[1,2,8] for deriving such local invariants.

Thus, we proposed[1,2,6,8] two classes of LOVIs which give *c*- and *x*-type ordering when applied to a layer matrix:

$$c(LM)_i = [\sum_{j=1}^{ecc_i} (lm_{ij})^{j/dsp}]^{-1} \qquad (3)$$

$$x(LM)_i = [\sum_{j=0}^{ecc_i} lm_{ij} 10^{-zj} \pm l_i]^{\pm 1} t_i \qquad (4)$$

$$l_i = f_i(lm_{i0}/10 + lm_{i1}/100) \qquad (5)$$

$$f_i = \sum_u (c_{iu} - 1) \qquad (6)$$

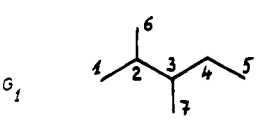
where ecc_i is the eccentricity of vertex i (the maximal topological distance from vertex i to any vertices in graph); dsp is a specified topological distance, usually larger than the diameter of the graph (here dsp = 10, unless otherwise specified); z is the number of digits of the (integer part of) max lm_{ij} value in the graph; l_i is a local parameter for multiple bonds; f_i is a multigraph factor, with c_{iu}—the conventional bond order (1, 2, 3, and 1.5 for single, double, triple, and aromatic bonds, respectively); t_i is a weighting factor, accounting for heteroatoms (e.g., a Sanderson type of electronegativity[9]).

Here, the *c*-symbol refers to the centricity of a vertex i (its location vs the center of graph), whereas the x−symbol refers to the centrocomplexity[1,2,8] (the location vs a vertex "of importance", i.e., a vertex with highest degree/valence, electronegativity, etc., we called the "center of complexity").

In this paper we present new *c*/*x*-type LOVIs and TIs which were derived both on quadratic and layer distance matrices (with 3D-metric distances supplied by an original MMX program of molecular geometry). A QSPR/QSAR study shows the proposed indices to have good correlating ability.

THREE-DIMENSIONAL DISTANCE MATRICES

We constructed the (three-dimensional) distance matrices, 3D, in a full analogy with the construction of distance matrices, D, from the topological distances: the entries in 3D are the actual 3D distances between the vertices of a graph which was geometrically optimized (i.e., by a MMX calculus). The 3D matrix for 2,3-dimethylpentane (23M2C5), G_1, is presented in Figure 1.



	1	2	3	4	5	6	7
1	0.0000	1.5414	2.5709	3.9411	4.5163	2.5178	3.0305
2	1.5414	0.0000	1.5543	2.5634	3.0891	1.5388	2.5821
3	2.5709	1.5543	0.0000	1.5468	2.5852	2.5930	1.5395
4	3.9411	2.5634	1.5468	0.0000	1.5364	3.0398	2.5461
5	4.5163	3.0891	2.5852	1.5364	0.0000	3.6199	3.9326
6	2.5178	1.5388	2.5930	3.0398	3.6199	0.0000	3.2366
7	3.0305	2.5821	1.5395	2.5461	3.9326	3.2366	0.0000

Figure 1. 3D matrix for 2,3-dimethylpentane, G_1 (optimized geometry).

By considering the mode of how the distance matrices collect the information on molecular structures, the matrices

† Babes-Bolyai University.
‡ Ruder Boskovic' Institute Zagreb.
⊗ Abstract published in *Advance ACS Abstracts,* December 1, 1994.

0095-2338/95/1635-0129$09.00/0

Table 1. Reversing *c/x* LOVIs: Labels and Exponent Values (cf. eq 8)

LOVI	*c*(D)	*x*(D)	*c*(3D)	*x*(3D)	*c*(3D, k)	*x*(3D, k)	*c*(3D, s)	*x*(3D, *s*)
a	−1	+1	−1	+1	−1	+1	−1	+1
b	+1	−1	+1	−1	+1	−1	+1	−1
c	0	0	0	0	−1/2	−1/2	−1/2	−1/2
p	1	1	1	1	*k*	*k*	*s*	*s*

D (or 2D) can be referred to as "through bond" type, whereas the matrices 3D are of "through space" type.[10]

According to eqs 1 and 2 and using the 3D distances, we constructed a new layer matrix, LM3D, with the following specifications

$$m_u = \sum_{\text{all } v \in G} 3d_{uv} = 3D_u \qquad (7)$$

where $3D_u$ is the 3D distance sum from vertex u to all the other vertices in the graph. Thus, the matrix LM3D is the 3D correspondent of the LMD matrix, earlier introduced[1,2] (old label R). Its dimensions are nd.

Basically, the LM3D matrix should differ for different conformations, i.e., it could be used for the characterization of molecular geometry. However, we do not have rigorous proof that LM3D is an injective function of molecular geometry. The LM3D matrix is examplified for 2,3-dimethylpentane, G_1; (missing entries are zeroes).

1	18.118	12.869	28.935	32.041	19.279
2	12.869	47.053	32.041	19.279	
3	12.390	44.910	53.943		
4	15.174	31.669	29.736	34.664	
5	19.279	15.174	12.390	29.736	34.664
6	16.546	12.869	30.507	32.041	19.279
7	16.867	12.390	28.043	53.943	

REVERSING *C/X* INDICES ON D AND 3D MATRICES

When a separation measure between the vertex i and any other vertices j in the graph is large enough (i.e., the cubic power of distances), the overall sum of the inverse values of such measure will give good information on the i location in graph. A precedent exists in the work of Hall and Kier.[3] In the view of exploring such topological information we defined a series of *c* and *x* LOVIs, $l(m,p)_i$, on the matrices D and 3D in the following way

$$l(m,p)_i = [\sum_{\text{all } j} [(m_{ij})^3 (p_i p_j)^c]^b]^a \qquad (8)$$

where l is labeling for c and x LOVIs, and m denotes the separation property and the correspondin matrix (i.e., D or 3D). The vertices could be either uncharacterized (in which case $p_u = 1$ for each vertex u) or characterized, i.e., by their degrees, *k*, or Sanderson adjusted electronegativities,[9] *s* ($p_u = k_u$ and $p_u = s_u$, respectively). The values for the exponents a, b, and c and the symbols of the corresponding LOVIs are listed in Table 1. Summation of $l(m,p)_i$ over all i vertices in G will give the corresponding *C* and *X* global indices. As the signs of a (and b) are opposite, this results in c and x LOVIs (and TIs). For this reason they are called the reversing indices. Examples will be given in a next section.

INDICES ON LAYER MATRIX LM3D

The centricity index, $c(LM3D)_i$, was defined in agreement with the eq 3. For the x LOVI, the eq 4 was modified as follows

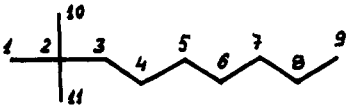
$$x(LM3D)_i = [\sum_{j=0}^{ecc_i} (lm_{ij}10^{-zj})/k_i - l_i]^{-1} t_i \qquad (9)$$

$$lm_{ij} = \sum_{u \in G(u)_j} 3D_u \qquad (10)$$

where $3D_u$ has the meaning given by eq 7. Summation over all i vertices in G will provide the corresponding global indices, donated *C*(LM3D) and *X*(LM3D), respectively.

INTRAMOLECULAR ORDERING

The character c or x of the local invariants resulted by applying reversing operators can be changed when the sign of a and b exponents in eq 8 interchanges. This is examplified on 2,2-dimethylnonane (22M2C9), G_2 (Figure 2).



c LOVIs (*10^2)

ver-tex	*c*(D)$_i$		*c*(3D)$_i$		*c*(3D, *k*)$_i$		*c*(3D, *s*)$_i$		*c*(LM3D)$_i$
4	0.31746	5	0.16345	4	0.25700	5	0.23785	5	6.11819
5	0.30488	4	0.16058	5	0.25667	4	0.23275	4	4.59918
3	0.21459	6	0.11117	6	0.17659	6	0.16125	6	4.09266
6	0.19569	3	0.10222	3	0.16914	3	0.14702	3	3.17255
2	0.12706	7	0.06230	2	0.14062	7	0.08997	7	2.67781
7	0.11338	2	0.05874	7	0.10067	2	0.07907	2	2.12072
1	0.07622	10	0.04981	8	0.06119	10	0.07710	8	1.70948
10	0.07622	11	0.04981	10	0.05992	11	0.07710	1	1.39544
11	0.07622	8	0.03729	11	0.05992	1	0.05716	10	1.39508
8	0.06798	1	0.03700	1	0.04495	8	0.05366	11	1.39508
9	0.04310	9	0.02302	9	0.02698	9	0.03579	9	1.08531

x LOVIs

ver-tex	*x*(D)$_i$		*x*(3D)$_i$		*x*(3D, *k*)$_i$		*x*(3D, *s*)$_i$		*x*(LM3D)$_i$
2	4.19321	2	1.16857	2	2.62915	2	1.60840	2	0.11833
3	2.56529	3	0.81636	3	1.74616	3	1.12819	4	0.08139
4	2.42177	4	0.77307	4	1.54225	4	1.07417	5	0.07818
5	2.38657	5	0.73848	5	1.46869	5	1.02403	3	0.07592
6	2.36370	6	0.72652	6	1.44006	6	1.00770	6	0.07143
7	2.32455	7	0.70717	7	1.37688	7	0.98547	7	0.06120
8	2.19904	8	0.64740	8	1.13150	8	0.92739	8	0.05136
1	1.44516	10	0.51401	10	0.83364	10	0.76152	10	0.03479
10	1.44516	11	0.51401	11	0.83364	11	0.76152	11	0.03479
11	1.44516	1	0.49294	1	0.80216	1	0.73028	1	0.03168
9	1.19907	9	0.37233	9	0.52595	9	0.55792	9	0.02492

Figure 2. c and x ordering of vertices in 2,2-dimethylnonane, G_2 (optimized geometry).

Among the c-type LOVIs, the $c(LM3D)^1$ gives the best centric ordering of vertices in 22M2C9, which alternates around the center of the longest chain in the graph (Bonchev's first centric criterion: minimum eccentricity[11]). The c(D)-induced ordering is improved by c(3D) which uses 3D-metric

Table 2. Values of Global TIs for 17 Geometrical Isomers of Heptane[a]

C Indices

graph	*C*(LK)	*C*(D)*10²	*C*(3D)*10²	*C*(3D,*K*)*10²	*C*(3D,*S*)*10²	*C*(LM3D)
C7	1.32495	4.56222	2.11553	3.24221	3.10206	0.58938
	1.32495	4.56222	2.42672	3.76273	3.54885	0.59800
	1.32495	4.56222	2.45578	3.76062	3.60175	0.59975
2MC6	1.47244	6.00622	3.54742	5.40582	5.22565	0.74061
	1.47244	6.00622	3.56967	5.37848	5.27559	0.74258
	1.47244	6.00622	3.04090	4.57529	4.49299	0.73001
3MC6	1.54056	7.24297	3.91708	6.00291	5.77210	0.78580
	1.54056	7.24297	3.66401	5.59445	5.40426	0.78246
24M2C5	1.73782	8.34368	4.58103	6.82365	6.81571	0.99623
22M2C5	1.78726	9.62898	4.89901	7.28299	7.32021	1.02187
3EC5	1.80268	9.44105	5.19622	8.05560	7.63888	1.03608
	1.80268	9.44105	5.25442	8.14142	7.72468	1.03730
23M2C5	1.83800	10.29196	5.86807	9.00538	8.68786	1.06819
	1.83800	10.29196	5.47451	8.29493	8.12380	1.06293
33M2C5	1.93840	12.60393	6.69405	10.30236	9.94592	1.14982
	1.93840	12.60393	6.36465	9.78058	9.46510	1.14236
223M3C4	2.14856	14.54199	7.12919	10.77349	10.68455	1.34806

X Indices

graph	*X*(LK)	*X*(D)	*X*(3D)	*X*(3D, *K*)	*X*(3D, *S*)	*X*(LM3D)
C7	14.39506	13.68131	4.10895	7.35749	5.85054	0.85993
	14.39506	13.68131	4.14010	7.39726	5.89811	0.88553
	14.39506	13.68131	4.14459	7.42188	5.90124	0.89694
2MC6	14.61504	13.92205	4.28829	7.78408	6.12085	0.97793
	14.61504	13.92205	4.30982	7.84379	6.14824	0.99767
	14.61504	13.92205	4.25306	7.73902	6.06682	0.94935
3MC6	14.63682	13.98012	4.30147	7.92740	6.12818	1.01281
	14.63682	13.98012	4.29778	7.92905	6.12245	1.00820
3EC5	14.65860	14.03819	4.38751	8.16986	6.24714	1.08685
	14.65860	14.03819	4.39871	8.18192	6.26502	1.08945
24M2C5	14.83680	14.17130	4.41880	8.16495	6.31601	1.05926
23M2C5	14.87640	14.25694	4.49663	8.43149	6.42484	1.12948
	14.87640	14.25694	4.47666	8.39795	6.39518	1.11269
22M2C5	15.05460	14.39005	4.49844	8.39053	6.45383	1.09245
33M2C5	15.09420	14.47569	4.58674	8.68318	6.57482	1.17568
	15.09420	14.47569	4.56549	8.64335	6.54519	1.16101
223M3C4	15.31200	14.69444	4.65702	8.84922	6.70325	1.20190

[a] C_n stands for the longest path in the graph; M and E are labels for the methyl and ethyl groups, respectively, with corresponding multiplicity and location, i.e., 22M2C5 = 2,2-dimethylpentane.

distances. Despite different orderings, the c-trend is obvious within this set of LOVIs.

The x character denotes a location versus a crucial vertex or subgraph (one of the highest LOVI). Most frequently, the LOVI is a function of vertex degree, so that the x character of LOVI values (supplied as a spectrum by an operator within the MOLORD algorithm, see ref 12) increases with the increasing order of L_n (iterative line dervatives). When the vertex ordering does not change along the spectrum of LOVI values, this is proof of a pure x character for the considered LOVI.

The best x-type LOVI tested in ref 1 was x(LK) (denoted therein as BX). In the case of 22MMC9 that LOVI gives the following vertex ordering: 2, 3, 4, 5, 6, 7, 8, 1, 10, 11, 9. Since 3D LOVIs cannot be computed for L_n (with the exception of the initial graph, n = 0, whose metric distances are computed by MMX calculations), we have taken the ordering induced by x(LK) LOVI as a standard x one. In Figure 2, one can see the good x ordering given by the operators based on D and 3D matrices.

A quite different ordering was induced by x(LM3D) index, due to the particular construction of LM3D matrices (the same ordering is given by Balaban's J index); however, the most important vertex 2, was correctly located.

INTERMOLECULAR ORDERING

The global TIs supplied by *C* and *X* indices in the set of 17 geometrical heptane isomers are listed in Table 2.

The proposed indices give different orderings in the heptane isomers (Tables 2 and 3), depending on their c or x character. For comparison, we introduced the *C*(LK) and *X*(LK) indices[1] as representative ones of c and x character, respectively.

In the *C* set of indices, the *C*(D) induces only one inversion (3EC5 before 22M2C5) vs *C*(LK) and *C*(LM3D) (Table 3). The same ordering (as *C*(D)) was obtained by treating the path sequences[13] of heptane isomers according to the 1P–3P criteria of graph centricity[11] (Tables 3 and 4).

The best x ordering of heptane isomers is given by the *X*(LK), *X*(D), *X*(3D), and *X*(3D,S) indices, which is identical to that obtained by the lexicographic ordering of the path sequences (Tables 4 and 5). Notice that *X*(LK) index was computed taking the weighting factor $t_i = 1$. In *X*(3D,*K*) and *X*(LM3D) induced ordering, two inversions (24M2C5, 3EC5 and 22M2C5, 23M2C5) appeared vs *X*(LK) ordering. For comparison, J and DJ indices[1] induced only one inversion (24M2C5, 3EC5), Randic's ID number[14] one inversion (3EC5, 3MC6), but none was observed for DM^1 and DM^2 indices.[15]

Table 3. *C* and *X* Ordering in Heptane Isomers

					C Ordering				
C(LK):	C7,	2MC6,	3MC6,	24M2C5,	22M2C5,	3EC5,	23M2C5,	33M2C5,	223M3C4
C(D):	C7,	2MC6,	3MC6,	24M2C5,	3EC5,	22M2C5,	23M2C5,	33M2C5,	223M3C4
C(3D):	C7,	2MC6,	3MC6,	24M2C5,	22M2C5,	3EC5,	23M2C5,	33M2C5,	223M3C4
C(3D, *K*):	C7,	2MC6,	3MC6,	24M2C5,	22M2C5,	3EC5,	23M2C5,	33M2C5,	223M3C4
C(3D, *S*):	C7,	2MC6,	3MC6,	24M2C5,	22M2C5,	3EC5,	23M2C5,	33M2C5,	223M3C4
C(LM3D):	C7,	2MC6,	3MC6,	24M2C5,	22M2C5,	3EC5,	23M2C5,	33M2C5,	223M3C4
path seq:	C7,	2MC6,	3MC6,	24M2C5,	3EC5,	22M2C5,	23M2C5,	33M2C5,	223M3C4
					X Ordering				
X(LK):	C7,	2MC6,	3MC6,	3EC5,	24M2C5,	23M2C5,	22M2C5,	33M2C5,	223M3C4
X(D):	C7,	2MC6,	3MC6,	3EC5,	24M2C5,	23M2C5,	22M2C5,	33M2C5,	223M3C4
X(3D):	C7,	2MC6,	3MC6,	3EC5,	24M2C5,	23M2C5,	22M2C5,	33M2C5,	223M3C4
X(3D, *K*):	C7,	2MC6,	3MC6,	24M2C5,	3EC5,	22M2C5,	23M2C5,	33M2C5,	223M3C4
X(3D, *S*):	C7,	2MC6,	3MC6,	3EC5,	24M2C5,	23M2C5,	22M2C5,	33M2C5,	223M3C4
X(LM3D):	C7,	2MC6,	3MC6,	24M2C5,	3EC5,	22M2C5,	23M2C5,	33M2C5,	223M3C4
path seq:	C7,	2MC6,	3MC6,	3EC5,	24M2C5,	23M2C5,	22M2C5,	33M2C5,	223M3C4

Table 4. *X* and *C* Ordering in Heptane Isomers According to Path Sequences[13]

path sequence						*X* ord	*X*(LK)	*X*(D)	DM^1	*C* ord	*C*(D)*10²
6	5	4	3	2	1	C7	14.39506	13.68131	13.42462	C7	4.56222
6	6	4	3	2	0	2MC6	14.61504	13.92205	14.76562	2MC6	6.00622
6	6	5	3	1	0	3MC6	14.63682	13.98012	15.08212	3MC6	7.24297
6	6	6	3	0	0	3EC5	14.65860	14.03819	15.36658	24M2C5	8.34368
6	7	4	4	0	0	24M2C5	14.83680	14.17130	16.36313	3EC5	9.44105
6	7	6	2	0	0	23M2C5	14.87640	14.25694	16.94921	22M2C5	9.62898
6	8	4	3	0	0	22M2C5	15.05460	14.39005	17.94975	23M2C5	10.29196
6	8	6	1	0	0	33M2C5	15.09420	14.47569	18.48528	33M2C5	12.60393
6	9	6	0	0	0	223M3C4	15.31200	14.69444	20.54701	223M3C4	14.54199

Table 5. *X* and *C* Ordering in Octane Isomers According to Path Sequences[13]

path sequence							*X* ord	*X*(LK)	*X*(D)	DM^1	*C* ord	*C*(LM3D)
7	6	5	4	3	2	1	C8	16.83951	16.06772	15.61028	C8	0.44479
7	7	5	4	3	2	0	2MC7	17.05950	16.31189	17.02015	2MC7	0.56848
7	7	6	4	3	1	0	3MC7	17.08148	16.37670	17.56044	3MC7	0.60339
7	7	6	5	2	1	0	4MC7	17.08346	16.39195	17.56044	4MC7	0.63236
7	7	7	5	2	0	0	3EC6	17.10544	16.45677	17.91494	25M2C6	0.71165
7	8	5	4	4	0	0	25M2C6	17.27968	16.55937	18.60840	22M2C6	0.73761
7	8	6	5	2	0	0	24M2C6	17.30344	16.63269	19.20822	3EC6	0.75937
7	8	7	4	2	0	0	23M2C6	17.32324	16.67552	19.60890	24M2C6	0.76325
7	8	8	4	1	0	0	34M2C6	17.34502	16.73359	20.00839	23M2C6	0.76779
7	8	8	5	0	0	0	3E2MC5	17.34700	16.74884	20.10744	34M2C6	0.81452
7	9	5	4	3	0	0	22M2C6	17.49946	16.79337	20.51561	33M2C6	0.82412
7	9	7	4	1	0	0	33M2C6	17.54302	16.90952	21.42983	224M3C5	1.01891
7	9	8	4	0	0	0	234M3C5	17.56480	16.96759	22.07279	3E2MC5	1.04532
7	9	9	3	0	0	0	3E3MC5	17.58460	17.01042	22.13693	234M3C5	1.05845
7	10	5	6	0	0	0	224M3C5	17.72320	17.05787	22.80578	3E3MC5	1.11623
7	10	8	3	0	0	0	223M3C5	17.78260	17.18634	24.14856	223M3C5	1.08177
7	10	9	2	0	0	0	233M3C5	17.80240	17.22917	24.49869	233M3C5	1.13486
7	12	9	0	0	0	0	2233M4C4	18.23800	17.66667	29.75000	2233M4C4	1.39893

The set of octane isomers was similarly tested. The path sequence ordering can be seen in Table 5.

The best c index *C*(LM3D) (one inversion vs the path sequence c ordering: 223M3C5, 3E3MC5) is followed by the *C*(D) index (two inversions: 24M2C6, 3EC6 and 223M3C5, 3E3MC5), whereas other 3D indices give very different ordering. This is probably due to the increasing flexibility of the longest carbon chain in octanes and to the geometry optimizing procedure (see below).

The x indices, particularly the 3D ones, again show a large variety of ordering. In contrast, the pure topological x indices *X*(LK) and *X*(D) give an ordering which is identical to that given by DM^1 superindex[15] or to that given by the lexicographic ordering of path sequences in octanes (see Table 5). By comparison, the index *X*(3D,S) shows two inversions (25M2C6, 3EC6 and 22M2C6, 3E2MC5).

Topological indices are frequently intercorrelated. The intercorrelations generally change with the number and type of the molecules studied. The intercorrelation matrices contain along with the c- and x-type indices two van der Waals (VDW) parameters (computed as described in the section Molecular Geometries) for the set of 17 geometric isomers of heptane (presented in Table 6). Further discussion on the intercorrelations is presented in the section QSPR/QSAR studies.

PROPERTIES OF *C* AND *X* ORDERINGS

While the c ordering in a set of molecules is easily conceivable, the x ordering requires some comments. Within the MOLORD algorithm, which works on the ground of L_n, the intermolecular ordering can be made by taking into account subgraphs larger than one vertex. For example, the

Table 6. Intercorrelation matrices in 17 Geometric Isomers of Heptane (Index of Correlation Given as r^2)

(a) van der Waals Parameters and *C* Global Indices

	VDW area	SPI	*C*(3D)	*C*(3D, *K*)	*C*(3D, *S*)	*C*(LM3D)
VDW area	1.00000	0.91192	0.99178	0.99005	0.99135	0.97707
SPI		1.00000	0.92355	0.90885	0.92906	0.93059
C(3D)			1.00000	0.99918	0.99986	0.98196
C(3D, *K*)				1.00000	0.99849	0.97659
C(3D, *S*)					1.00000	0.98311
C(LM3D)						1.00000

(b) van der Waals Parameters and *X* Global Indices

	VDW area	SPI	*X*(3D)	*X*(3D, *K*)	*X*(3D, *S*)	*X*(LM3D)
VDW area	1.00000	0.91192	0.97011	0.97508	0.96087	0.98594
SPI		1.00000	0.98186	0.97010	0.98835	0.93806
X(3D)			1.00000	0.99629	0.99882	0.98477
X(3D, *K*)				1.00000	0.99252	0.99176
X(3D, *S*)					1.00000	0.97580
X(LM3D)						1.00000

X(D) index in heptanes preserves the ordering C7, 2MC6, 3MC6, 3EC5, 24M2C5, 23M2C5, 22M2C5, 33M2C5, 223M3C4 for subgraphs of 0, 1, and 2 edges, but one inversion appeared (24M2C5, 3EC5) for subgraphs of three edges. This is exactly the sequence of Bertz's index.[16] The other indices could give more varied ordering, according to their vertex separation ability. This is not a problem, since, according to Bertz,[16] the number of edges in L_n (the Bertz index, *B*) within a set of molecules can vary along a string of Ln.

From the above examples some conclusions emerge: (1) The c ordering given by *C*(LM3D) index is very close to that given by applying the Bonchev's 1P−3P criteria to the graph path sequences. The question of the "true" ordering remains open. (2) The x ordering implies that the vertex degree is the most important local invariant. (3) An ordering which follows the branching in graphs will always remain a matter of intuition as it was already pointed out by Bertz.

QSPR/QSAR STUDIES

Two van der Waals parameters were tested in the set of 17 geometric isomers of heptane: area and surface potential index, SPI, (see section entitled Molecular Geometries). The areas correlate well with the c indices (*C*(LK), 0.97689; *C*(D), 0.96638; *C*(3D,*K*), 0.99178; *C*(3D,S), 0.99005; *C*(LM3D), 0.99135—see Tables 3 and 6), whereas the x indices correlate better with the SPI parameter (*X*(LK), 0.99407; *X*(D), 0.99498; *X*(3D), 0.98186; *X*(3D,S), 0.98835). This holds for the indices with well differentiated c and x characters (e.g., *C*(LK) vs *X*(LK), 0.91863) but becomes less pronounced as the correlation c vs x increases (*C*(D) vs *X*(D), 0.96041; *C*(3D,S) vs *X*(3D,S), 0.977218; *C*(3D) vs *X*(3D), 0.97621; *C*(LM3D) vs *X*(LM3D), 0.97669; *C*(3D,*K*) vs *X*(3D,*K*), 0.97698). This behavior can be partly explained by the fact that the c indices "feel" better the remote points (i.e., those located on the van der Waals envelope) in molecular graphs than the x indices do. Conversely, the x indices express better the identity of atoms, sometimes associated with their electronegativity. Notice that *X*(LK) and *X*(D) do not use electronegativities ($t_i = 1$). Some QSPR equations are given in Table 7.

Further we tested the correlation ability of the indices introduced here with selected physicochemical and biological properties within a set of 25 ethers[17] (Tables 8 and 9).

Table 7. QSPR Equations in the Set of 17 Geometric Isomers of Heptane

van der Waals Area	
$y = 235.43228 - 501.17591C(3D)$	$r^2 = 0.99178$
	$s = 1.00021$
$y = 235.31912 - 327.03903C(3D, K)$	$r^2 = 0.99005$
	$s = 1.10039$
$y = 235.11990 - 333.75613C(3D, S)$	$r^2 = 0.99135$
	$s = 1.02614$
SPI (Surface Potential Index)	
$y = 0.00030 + 0.02023(X(3D)$	$r^2 = 0.98186$
	$s = 0.00063$
$y = 0.03147 + 0.00709X(3D, K)$	$r^2 = 0.97010$
	$s = 0.00081$
$y = 0.00599 + 0.01325X(3D, S)$	$r^2 = 0.98835$
	$s = 0.00051$

The physicochemical parameters considered are van der Waals area, SPI, and relative conformational energy, and the biological property studied is the toxicity of ethers on mice.

In Table 10 we present some QSPR and QSAR equations for the set of ethers in Table 8.

From Table 10 one can see that the van der Waals areas show direct correlation with the number of carbon atoms, NC, in ethers (which is modulated by the topological indices), while the SPI and the relative conformational energy are inverse correlated (by means of normalized topological indices, TI/NC) with this parameter. The toxicity of ethers on mice shows not only parabolic dependence on NC or TIs but also an excellent linear correlation with SPI and a combination of the normalized TIs, SUM*X*/NC.

Notice that the above QSAR results are very close to that obtained by Mekenyan et al.[17] by using three variables: NC, NC^2 and the electropy index ($r = 0.9784$).

MOLECULAR GEOMETRIES

The molecular geometries used in the 3D indices calculation were obtained from molecular mechanics calculations, using an updated version written in Turbo Pascal 6.0 for the PC environment of our MM-workstation described in ref 18. The empirical force field used was Allinger's MM2.[19] In order to obtain the absolute minimum of potential energy, a simple procedure of mapping the conformational subspace was used.

According to the multiplicity of each torosional barrier, the workstation assumes a set of optimal values for the

Table 8. Toxicity on Mice (pC), Number of Carbon Atoms (NC), Conformational Energy, and van der Waals Parameters in Ethers (List Taken from Ref 17)

no.	graph	pC	NC	conformat. energy (kcal/mol)	VDW area (Å²)	SPI (euc/Å)
1.	dimethyl ether	1.43	2	4.7018	105.6058	0.2586
2.	ethyl methyl ether	1.74	3	5.8029	134.1241	0.2407
3.	propyl methyl ether	2.45	4	6.4223	160.2592	0.2264
4.	isopropyl methyl ether	2.26	4	8.5567	156.2889	0.2299
5.	cyclopropyl methyl ether	2.75	4	18.8305	143.5338	0.2343
6.	butyl methyl ether	2.70	5	6.8500	183.4762	0.2111
7.	isobutyl methyl ether	2.79	5	7.9113	180.5055	0.2174
8.	*sec*-butyl methyl ether	2.79	5	9.2527	181.3286	0.2190
9.	*tert*-butyl methyl ether	2.79	5	11.2903	175.0331	0.2192
10.	pentyl methyl ether	2.88	6	7.7073	211.2203	0.2018
11.	ethyl ethyl ether	2.22	4	6.8643	162.9201	0.2291
12.	propyl ethyl ether	2.60	5	7.4748	188.3811	0.2159
13.	isopropyl ethyl ether	2.60	5	9.6103	184.9085	0.2193
14.	cyclopropyl ethyl ether	3.00	5	19.8335	172.5336	0.2233
15.	butyl ethyl ether	2.82	6	8.1156	214.7855	0.2055
16.	isobutyl ethyl ether	2.82	6	7.8514	209.7583	0.2094
17.	*sec*-butyl ethyl ether	2.85	6	10.2410	209.3766	0.2076
18.	*tert*-butyl ethyl ether	2.92	6	12.3225	203.2090	0.2112
19.	pentyl ethyl ether	3.00	7	8.7536	240.2706	0.1962
20.	neopentyl ethyl ether	3.15	7	8.1182	227.6520	0.2032
21.	vinyl ethyl ether	2.34	5	12.2599	147.3691	0.2379
22.	propyl propyl ether	2.79	6	8.0819	214.2605	0.2071
23.	isopropyl propyl ether	2.82	6	10.1934	211.0182	0.2093
24.	isopropyl isopropyl ether	2.82	6	12.2733	205.7203	0.2109
25.	divinyl ether	2.33	6	20.9689	135.1905	0.2566

Table 9. 3*D* Topological Indices for 25 Ethers in Table 8

G	*X*(3D, *K*)	*X*(3D, *S*)	*X*(3D)	*X*(3D, *K*)/NC	*X*(3D, *S*)/NC	*X*(3D)/NC
1	2.1185	2.4446	1.5426	1.0592	1.2223	0.7713
2	3.6004	3.4498	2.2715	1.2001	1.1499	0.7572
3	5.0505	4.4635	3.0020	1.2626	1.1159	0.7505
4	5.4570	4.6870	3.1468	1.3643	1.1717	0.7867
5	7.6192	5.1757	3.6472	1.9048	1.2939	0.9118
6	6.5355	5.5198	3.7604	1.3071	1.1040	0.7521
7	6.9053	5.7138	3.8880	1.3811	1.1428	0.7776
8	7.1365	5.7918	3.9448	1.4273	1.1584	0.7890
9	7.6069	6.1102	4.1323	1.5214	1.2220	0.8265
10	7.9554	6.4850	4.4612	1.3259	1.0808	0.7435
11	5.2003	4.5129	3.0431	1.3001	1.1282	0.7608
12	6.6436	5.5048	3.7608	1.3287	1.1010	0.7522
13	7.1011	5.7435	3.9188	1.4202	1.1487	0.7838
14	9.3066	6.2316	4.4216	1.8613	1.2463	0.8843
15	8.1062	6.5201	4.4938	1.3510	1.0867	0.7490
16	8.5393	6.7789	4.6642	1.4232	1.1298	0.7774
17	8.8285	6.8746	4.7365	1.4714	1.1458	0.7894
18	9.3296	7.1968	4.9289	1.5549	1.1995	0.8215
19	9.5731	7.5371	5.2286	1.3676	1.0767	0.7469
20	10.7574	8.2583	5.6928	1.5368	1.1798	0.8133
21	5.8788	5.1956	3.4942	1.4697	1.2989	0.8736
22	8.1284	6.5257	4.4989	1.3547	1.0876	0.7498
23	8.6044	6.7760	4.6653	1.4341	1.1293	0.7775
24	99.1264	7.0566	4.8516	1.5211	1.1761	0.8086
25	6.6080	5.8495	3.9349	1.6520	1.4624	0.9837

Table 10. QSPR and QSAR Equations in 25 Ethers

van der Waals Area	
$y = 46.5146 + 26.9513NC$	$r^2 = 0.97781$ $s = 6.96178$
$y = 44.1547 + 38.0474NC - 7.3692X(3D, K)$	$r^2 = 0.99267$ $s = 4.01784$
$y = 54.1935 + 45.7464NC - 25.5681X(3D)$	$r^2 = 0.99839$ $s = 1.88436$
$y = 56.2506 + 46.5791NC + 1.8354X(3D, K) - 30.4604X(3D)$	$r^2 = 0.99851$ $s = 1.81203$
SPI (Surface Potential Index)	
$y = 0.1867 + 0.0721X(3D)/NC - 0.0101NC$	$r^2 = 0.98524$ $s = 0.00276$
$y = 0.2482 - 0.0004$VDW area $+ 0.0393X(3D, S)/NC$	$r^2 = 0.97706$ $s = 0.00343$
Conformational Relative Energy[a]	
$y = -34.5923 + 15.9551X(3D, K)/NC + 18.5892X(3D, S)/NC$	$r^2 = 0.95724$ $s = 1.21944$
$y = -36.1417 + 11.4469X(3D, K)/NC + 37.3173X(3D)/NC$	$r^2 = 0.96181$ $s = 1.15370$
$y = -36.1693 + 13.5831$SUMX/NC	$r^2 = 0.96027$ $s = 1.17634$
Toxicity (pC)	
$y = 0.1796 + 0.1289X(3D, K) + 0.5552NC - 0.0481(NC)^2$	$r^2 = 0.96724$ $s = 0.10035$
$y = 1.9848 - 4.9526$SPI $+ 0.3539X(3D, K) - 0.0149[X(3D, K)]^2$	$r^2 = 0.97165$ $s = 0.09345$
$y = 5.4395 - 24.1071$SPI $+ 0.7328$SUMX/NC	$r^2 = 0.97699$ $s = 0.08432$

[a] With SUM$X/NC = X(3D, K)/NC + X(3D, S)/NC + X(3D)/NC$. [b] With SUM$X/NC$ as specified in the last section.

torsional angles of the molecule. The user is free to modify both the multiplicity and the minimum positions of each torsional axis or to ignore several degress of freedom if the user supposes that these do not have any great influence on the molecular potential or if the rotation around an axis leads to degenerate conformations. The next step consists in generating all the possible conformations (of the initial geometries). All possible combinations of the minimum values for the torsional angles are obtained by rotations around the corresponding axes, and a simple criterion is used to test the validity of each conformation. Since this conformational setup does not use any molecular energy evaluations and provides only rough initial geometries, only the structures showing a very large degree of overlapping of the van der Waals spheres can be ignored.

The degree of overlapping between the nonbonded atoms i and j is simply defined as

$$OVR_{i,j} = (R_{wi} + R_{wj})/d_{ij} \quad (11)$$

The user is asked about the maximal overlapping degree he/she wishes to admit in the conformations to be obtained, and all structures showing stronger interactions will be ignored. A good threshold value was found to be about 0.5, yielding a reasonable limitation on the valid conformations and also ensuring that at least one of the initial geometries leading to the absolute minimum would be obtained. Since optimization can dramatically rearrange the molecular structure, there is no correlation between the degree of overlapping encountered in these initial conformations and the energy of the stable conformations obtained by their optimization. An overlapping index was used to characterize each valid conformation

$$IOVR = \max(OVR_{i,j}) + 1/10(\sum OVR_{i,j}/n) \quad (12)$$

This basically not only accounts for the maximal overlapping degree encountered in the molecule but also includes information about the mean overlapping degree (n is the number of nonbonded pairs), so that it is able to distinguish degenerate a chiral conformations (having the same IOVR value). Different conformations might have the same maximal overlapping degree, so that only the first term cannot be used to decide whether these are actually identical or not.

The calculation of the van der Waals surface was made using polar coordinates. Each elemental area on an atomic van der Waals sphere was tested whether it belongs to the external surface or is placed in an overlapping zone. If the first situation occurs, the area is simply added to the total van der Waals area. By trial and error, we have found that a division of π by 30 for both angular coordinates suffices to provide the stability of the method. Van der Waals radii were taken from ref 20.

The fractional atomic charges for the studied molecules were obtained by DelRe[21] calculations.

In addition to the van der Waals area, an electrostatic index on the van der Waals area was calculated as follows: the sum of the squared atomic contributions to the electrostatic potential was calculated on each elemental area dA found

$$\phi(dA) = \sum_i [q_i/d(i,dA)]^2 \quad (13)$$

of the van der Waals surface. Notice that this treatment adds up all the atomic contributions, irrespective of the sign of the atomic fractional charges. The square root of the sum of these squared values on the entire surface was divided by the van der Waals surface and used as an index of the overall polarity of the van der Waals area, named surface potential index, SPI

$$SPI = [1/A\sum\phi(dA)dA]^{1/2} \quad (14)$$

CONCLUSIONS

By replacing the topological distances in the matrices D and LMD with the 3D-metric distances, matrices 3D and LM3D are introduced. 3D-Metric distances are obtained through the MMX molecular geometry calculations. On the basis of these matrices, a series of new local and global topological indices, for characterizing the centricity, c, and centrocomplexity, x, in graphs, are generated. It is shown that the indices introduced are useful in the QSAR/QSPR studies. The testing was performed on heptane and octane isomers and in a set of ethers.

REFERENCES AND NOTES

(1) Balaban, A. T.; Diudea, M. V. Real number vertex invariants: regressive distance sums and related topological indices. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 421−428.

(2) Diudea, M. V. Layer matrices in molecular graphs, *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1064−1071.

(3) Hall, L. H.; Kier, L. B. Determination of topological equivalence in molecular graphs from the topological state, *Quant. Struct.-Act. Relat.* **1990**, *9*, 115−131.

(4) Skorobogatov, V. A.; Dobrynin, A. A. Metric analysis of graphs, *MATCH* **1988**, *23*, 105−151.

(5) Diudea, M. V.; Parv, B.; A new centric connectivity index (CCI). *MATCH* **1988**, *23*, 65−87.

(6) Diudea, M. V.; Minailiuc, O. M.; Balaban, A. T. Regressive vertex degrees (new graph invariants) and derived topological indices. *J. Comput. Chem.* **1991**, *12*, 527−535.

(7) Diudea, M. V.; Kacso, I. E.; Minailiuc, O. M. *Y* indices in homogeneous dendrimers. *MATCH* **1992**, *28*, 61−99.

(8) Diudea, M. V.; Horvath, D.; Kacso, I. E.; Minailiuc, O. M.; Parv, B. Centricities in molecular graphs. The MOLCEN algorithm. *J. Math. Chem.* **1992**, *11*, 259−270.

(9) Diudea, M. V.; Silaghi-Dumitrescu, I. Valence group electronegativity as a vertex discriminator. *Rev. Roumaine Chim.* **1989**, *34*, 1175−1182.

(10) Randić, M. Generalized molecular descriptors. *J. Math. Chem.* **1991**, *7*, 155−168.

(11) (a) Bonchev, D.; Balaban, A. T.; Randić, M. The graph center concept for polycyclic graphs. *Int. J. Quantum. Chem.* **1981**, *19*, 61−82. (b) Bonchev, D.; Mekenyan, O.; Balaban, A. T. Iterative procedure for the generalized graph center in polycyclic graphs. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 91−97.

(12) Diudea, M. V.; Horvath, D.; Topan, M. MOLORD algorithm and real number subgraph invariants. *Croat. Chem. Acta*, in press.

(13) Randić, M.; Wilkins, C. L. Graph-theoretical ordering of structures as a basis for systematic searches for regularities in molecular data. *J. Phys. Chem.* **1979**, *83*, 1525−1540.

(14) Randić, M. On molecular identification numbers. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 164−175.

(15) Balaban, A. T.; Ciubotariu, D.; Ivanciuc, O. Design of topological indices. Part 2. Distance measure connectivity indices. *MATCH* **1990**, *25*, 41−70.

(16) Bertz, S. H. Branching in graphs and molecules. *Discr. Appl. Math.* **1988**, *19*, 65−83.

(17) Mekenyan, O.; Bonchev, D.; Sabljić, A.; Trinajstić; N. Applications of topological indices to QSAR. The use of Balaban index and the electropy index for correlations with toxicity of ethers on mice. *Acta. Pharm. Jugosl.* **1987**, *37*, 75−86.

(18) Horvath, D.; Silaghi-Dumitrescu, I. An interactive workstation for molecular mechanics modelling de chemical structures. *Rev. Roum. Chim.* **1992**, *37*, 1165−1174.

(19) Burkert, U.; Allinger, N. L. *Molecular Mechanics*; American Chemical Society: Washington, DC, 1982.

(20) Labanowski, J.; Motoc, I.; Dammkoehler, R. A. The physics meaning of topological indices. *Computers Chem.* **1991**, *15*, 47−53.

(21) DelRe, G. A simple MO-LCAO method for the calculation of charge distributions in saturated organic molecules. *J. Chem. Soc.* **1958**, 4031−4040.

CI930134E