Kragujevac J. Sci. 22 (2000) 49–58.

SZEGED INDICES: VERTEX AND FRAGMENTAL DESCRIPTORS

Anton A. Kiss,[a] Irina E. Kacso,[a] Ovidiu M. Minailiuc,[a] Mircea V. Diudea,[a] Sonja Nikolić[b] and Ivan Gutman[c]

[a] *Facultatea de Chimie si Inginerie Chimica, Universitatea "Babes–Bolyai" Arany Janos 11, R–3400 Cluj, Romania,*
[b] *Rudjer Bošković Institute, P. O. Box 1016, HR-11001 Zagreb, Croatia, and*
[c] *Faculty of Science, P. O. Box 60, 34000 Kragujevac, Yugoslavia*

(Received December 15, 1999)

ABSTRACT. Novel fragmental Szeged indices, defined on unsymmetric property matrices, which collect various fragmental properties, are proposed. Classical vertex Szeged indices and fragmental descriptors are tested for correlating ability with physico–chemical properties of two sets of cycle–containing organic structures. Some QSPR models are proposed.

INTRODUCTION

Wiener index [1], W, one of the most studied topological indices, (see the recent reviews [2, 3, 4]) is defined, for acyclic structures, by

$$W = \sum_{e} N_{i,(i,j)}\, N_{j,(i,j)} \tag{1}$$

where $N_{i,(i,j)}$ and $N_{j,(i,j)}$ denote the numbers of vertices lying on the two sides of the edge/path e/p (having the endpoints i and j). The summation runs over all edges (i,j) in the respective graph. When (i,j) represents a path, then a *hyper–Wiener index*, WW, can be calculated as [5]

$$WW = \sum_{p} N_{i,(i,j)}\, N_{j,(i,j)} \tag{2}$$

The products $N_{i,(i,j)}\, N_{j,(i,j)}$ represent the edge/path contribution to the global index W/WW and are just the (i,j)-entries in the Wiener matrices [6, 7] $\mathcal{W}_{e/p}$, from which the index can be calculated as the half–sum of their entries:

$$W/WW = \frac{1}{2} \sum_i \sum_j [\mathcal{W}_{e/p}]_{ij} \tag{3}$$

Note that the vertices i and j must be adjacent in $\mathcal{W}_e$, otherwise its non-diagonal entries are zero. Relations (1)–(3) are valid only in acyclic graphs. In cycle–containing graphs, the Wiener indices are calculated by means of the distance–type matrices [8, 9].

In order to extend the validity of the above relations to cycle–containing graphs, Gutman has proposed [10] the *Szeged index*, SZ, as a Wiener index analog. The SZ-index is defined according to Eq. (1), but the quantities $N_{i,(i,j)}$ and $N_{j,(i,j)}$ are now

$$N_{i,(i,j)} = |\{v | v \in V(G)\,; D_{iv} < D_{jv}\}| \tag{4}$$

$$N_{j,(i,j)} = |\{v | v \in V(G)\,; D_{jv} < D_{iv}\}| \tag{5}$$

where $V(G)$ denotes the set of vertices in a graph and D_{iv}, D_{jv} are the topological distances (i. e., the number of edges on a shortest path joining the vertices i and j, respectively, with a vertex v). $N_{i,(i,j)}$ and $N_{j,(i,j)}$ represent the cardinalities of the sets of vertices closer to i and to j, respectively; vertices equidistant to i and j are not counted. Thus, the SZ index is calculated by summing all the edge contributions in the graph:

$$SZ = \sum_e N_{i,(i,j)}\, N_{j,(i,j)} \tag{6}$$

Since the SZ index is defined on edges, in the following it will be denoted by SZ_e.

By analogy to the Wiener matrices, Szeged matrices, $\mathcal{SZ}_{e/p}$, can be defined [11] by the aid of edge/path contributions:

$$[\mathcal{SZ}_{e/p}]_{ij} = N_{i,(i,j)}\, N_{j,(i,j)} \tag{7}$$

on which the Szeged indices can be calculated by

$$SZ_{e/p} = \frac{1}{2} \sum_i \sum_j [\mathcal{SZ}_{e/p}]_{ij} \tag{8}$$

When the Szeged matrix is defined on paths, the index calculated on it is the *hyper–Szeged index*, SZ_p [11].

$\mathcal{SZ}_e$ can be obtained by the Hadamard product [12] (i. e., $[\mathcal{M}_a \bullet \mathcal{M}_b]_{ij} = [\mathcal{M}]_{ij}[\mathcal{M}]_{ij}$) between $\mathcal{SZ}_p$ and $\mathcal{A}$ (the adjacency matrix, whose entries are 1 if two vertices are adjacent and zero otherwise):

$$\mathcal{SZ}_e = \mathcal{SZ}_p \bullet \mathcal{A} \tag{9}$$

By analogy to the Cluj matrix [13], $\mathcal{CJ}_u$, a Szeged unsymmetric matrix, $\mathcal{SZ}_u$, was defined by Diudea [11] (cf. Fig. 1):

$$[\mathcal{SZ}_u]_{ij} = N_{i,(i,j)} \tag{10}$$

$\mathcal{SZ}_u$ is a square array of dimension $n \times n$, in general unsymmetric. It allows the construction of the symmetric matrices $\mathcal{SZ}_e$ and $\mathcal{SZ}_p$ by relation

$$[\mathcal{SZ}_{e/p}]_{ij} = [\mathcal{SZ}_u]_{ij}\,[\mathcal{SZ}_u]_{ji} \tag{11}$$

and the derivation of two Wiener–type indices, as

$$SZ_{e/p} = \sum_{e/p} [\mathcal{SZ}_u]_{ij}\,[\mathcal{SZ}_u]_{ji} \tag{12}$$

For tree graphs, $SZ_e = CJ_e = W$. Note that, in the above discussion, the Cluj matrix, $\mathcal{CJ}_u$, was defined [14] on the shortest path (i,j).

Analytical relations for calculating the Szeged indices in paths, P_n, and simple cycles, C_n, are derived [15]:

$$SZ_e(P_n) = \frac{1}{6} n(n^2-1) \tag{13}$$

$$SZ_p(P_n) = \frac{1}{48}(5n^4 - 10n^3 + 16n^2 - 8n - 6zn + 3z) \tag{14}$$

$$SZ_e(C_n) = \frac{1}{4} n(n-z)^2 \tag{15}$$

$$SZ_p(C_n) = \frac{1}{8} n(n-1)^{2z+1}(n^2 - 2n + 4)^{1-z} \tag{16}$$

where $z = 0$ if n is even and $z = 1$ if n is odd.

In the present paper some extensions of the Szeged index are presented, which account for fragments and their chemical nature.

SZEGED PROPERTY MATRICES

By analogy to $\mathcal{SZ}_u$, see Eqs. (4) and (10), Szeged property matrices are defined as [16]

$$[\mathcal{SZ}_u\mathcal{P}]_{ij} = P_{i,(i,j)} \tag{17}$$

$$P_{i,(i,j)} = m \sum_v P_v \tag{18}$$

$$P_{i,(i,j)} = \left(\prod_v X_v \right)^{1/N_{i,(i,j)}} \tag{19}$$

Entries in a Szeged property matrix are evaluated on the set of vertices v which obey the Szeged index condition (see Eq. (4)). In fact, such a set of vertices can be viewed as a fragment (i. e., a subgraph) since a molecular graph is always connected. The summation on the right-hand side of Eq. (18) goes over all vertices of the graph G which have the property $D_{iv} < D_{jv}$. A similar consideration holds for the product in Eq. (19).

Some special cases of the above definition deserve particular attention:

(a) $P_v = 1$; $m = 1$ (classical matrix, $\mathcal{SZ}_u$)

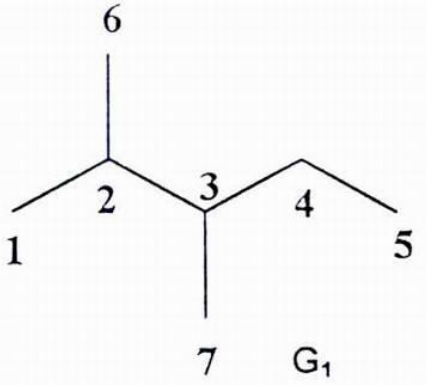
(b) $P_v = \sum_u A_u$; $m = 1/12$ (mass matrix, $\mathcal{SZ}_u\mathcal{A}$)

(c) X_v = group electronegativities [17] (electronegativity matrix, $\mathcal{SZ}_u\mathcal{X}$)

The case (a) is obvious: $P_{i,(i,j)}$ represents the cardinality of the set of vertices v (see above). In the case (b) A_u is the atomic mass and the summation runs over all atoms u which are represented by the same vertex v. The factor $m = 1/12$ indicates that $P_{i,(i,j)}$ is a fragmental mass, relative to the carbon atomic mass. In the case (c) $P_{i,(i,j)}$ is just the geometric mean of vertex values X_v, of group electronegativities.

In Fig. 1 the above matrices are illustrated for the graph G_1 (the molecular graph of 2,3-dimethylpentane).

Figure 1. Szeged Property Matrices; Vertex and Fragmental Indices in the Graph G_1



$\mathbb{SZ}_u$; $P_v = 1$; m = 1

0	1	1	3	3	1	3
6	0	3	3	5	6	3
4	4	0	5	5	4	6
4	2	2	0	6	4	2
2	2	1	1	0	2	2
1	1	1	3	3	0	3
4	1	1	1	5	4	0

$SZ_e = 46$; $SZ_p = 151$

$\mathbb{SZ}_u\mathbb{A}$; $P_v = A_v$; m = 1/12

0.000	1.250	1.250	3.583	3.583	1.250	3.583
7.083	0.000	3.583	3.583	5.917	7.083	3.583
4.750	4.750	0.000	5.917	5.917	4.750	7.083
4.750	2.417	2.417	0.000	7.083	4.750	2.417
2.417	2.417	1.250	1.250	0.000	2.417	2.417
1.250	1.250	1.250	3.583	3.583	0.000	3.583
4.750	1.250	1.250	1.250	5.917	4.750	0.000

$SZ_eA = 66.736$; $SZ_pA = 217.729$

$\mathbb{SZ}_u\mathbb{X}$; $P_v = EVG_v$; m = 1

0.000	0.958	0.958	0.962	0.962	0.958	0.962
0.963	0.000	0.962	0.962	0.963	0.963	0.962
0.962	0.962	0.000	0.963	0.963	0.962	0.963
0.962	0.960	0.960	0.000	0.963	0.962	0.960
0.960	0.960	0.958	0.958	0.000	0.960	0.960
0.958	0.958	0.958	0.962	0.962	0.000	0.962
0.962	0.958	0.958	0.958	0.963	0.962	0.000

$SZ_eX = 5.538$; $SZ_pX = 19.383$

Indices are calculated on the basis of these matrices by the general relation

$$I_{e/p} = \sum_{e/p} [\mathcal{SZ}_u \mathcal{P}]_{ij} [\mathcal{SZ}_u \mathcal{P}]_{ji} \quad ; \quad I = SZ\,;\, SZA\,;\, SZX \tag{20}$$

Indices calculated in the cases (b) and (c) are useful for discriminating chemical graphs which contain heteroatoms and multiple bonds.

APPLICATIONS

Two correlation tests have been performed; (a) for the vertex descriptors and (b) for both vertex and fragmental descriptors.

(a) A dozen of cycloalkanes (Table 1) were selected for correlating their boiling points, BP, and chromatographic retention indices, RI, with the classical hyper–Szeged index, SZ_p. For comparison, Wiener and hyper–Wiener indices are included. Statistics of single–variable regression is shown in Table 2. In such a set, the simplest descriptor is, of course, the number of carbon atoms, NC. Any acceptable correlation must

surpass the correlation shown by NC. Thus, in Table 2 one can see that the correlation coefficient, r, shown by the topological indices is inferior to that supplied by NC (0.965). However, by applying the logarithmic function, all descriptors showed increased values of r with a maximum for $\ln SZ_p$ (0.980) and a significant drop in the standard error of estimates, s. The crude descriptor, that is NC, is far less sensitive to logarithmation, proving a low structural information content.

Table 1. Boiling Points, BP[18] Chromatographic Retention Index, RI and Topological Indices

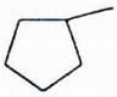
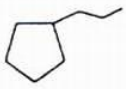
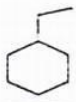
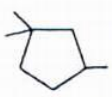
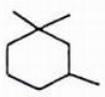
Graph	BP/N	RI	SZ_p	W	WW	Graph	BP/N	RI	SZ_p	W	WW
1 	35.95	510.0	40	17	26	7 	100.97	725.8	182	42	71
2 	49.35	565.0	40	15	20	8 	103.57	733.8	159	43	75
3 	70.76	621.1	92	29	49	9 	126.48	812.1	247	61	110
4 	71.86	627.9	79	26	39	10 	131.08	830.3	230	58	99
5 	80.76	662.7	105	27	42	11 	131.88	834.3	296	61	110
6 	105.08	723.6	308	64	122	12 	136.69	840.4	447	82	152

Table 2. Statistics of Single Variable Regression (y=a +bx) of Parameters from Table 1

Y	X	r	s	F
BP	SZ_p	0.9305	12.9426	64.5455
	W	0.9553	10.4518	104.3108
	WW	0.9533	10.6742	99.5966
	NC	0.96518	9.2434	136.1513
	ln SZ_p	0.9840	6.2945	305.172
	ln W	0.9757	7.7414	198.369
	ln WW	0.9721	8.2956	171.454
	ln NC	0.9725	8.2247	174.599
RI	SZ_p	0.9240	44.6232	58.384
	W	0.9455	37.9846	84.375
	WW	0.9470	37.4987	86.837
	NC	0.9512	36.0009	95.062
	ln SZ_p	0.9733	26.7664	180.061
	ln W	0.96241	31.6953	125.545
	ln WW	0.9597	32.8101	116.490
	ln NC	0.9580	33.4446	111.737

(b) A set of 15 structures of explosives (Fig. 2 and Table 3) was tested for the correlation of two properties: with the topological descriptors SZ_e, SZ_p, SZ_eA, SZ_pA, SZ_eX and SZ_pX (Table 4). Statistics of single–variable regressions and cross validation test ("leave one out") are shown in Table 5.

Figure 2. Molecular Formulas of Explosive Compounds

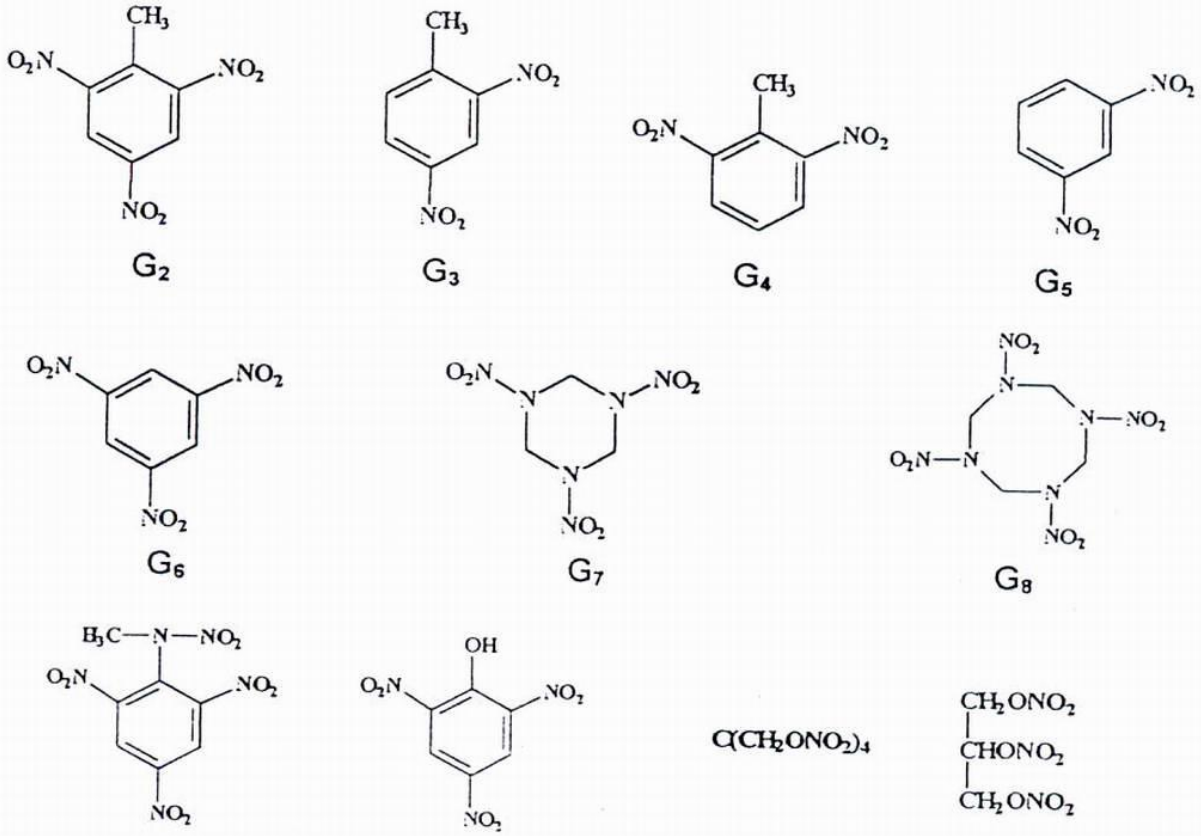
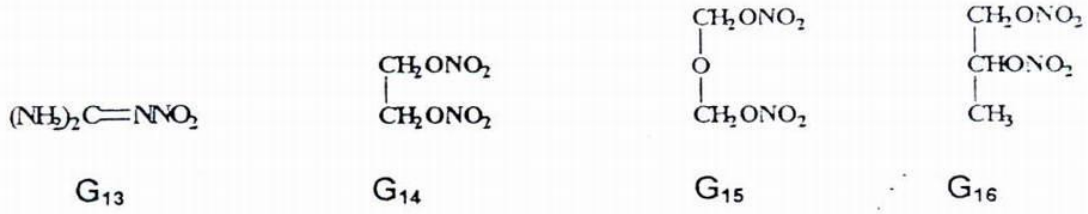


Figure 2. (continued).



G_{13} G_{14} G_{15} G_{16}

Table 3. Name of Compounds in the Graphs G_2-G_{16}.

Graph	Name	DC water $*10^{-6}$ (cm^2/s)	DC air (cm^2/s)
2	Trinitrotoluene	6.71	0.0639
3	2,4-Dinitrotoluene	7.31	0.0670
4	2,6- Dinitrotoluene	7.31	0.0670
5	1,3- Dinitrobenzene	7.94	0.0729
6	1,3,5- Trinitrobenzene	7.20	0.0679
7	Hexahydro-1,3,5-trinitro-1,3,5-triazine	7.15	0.0739
8	Octahydro-1,3,5,7-tetranitro-1,3,5,7-tetrazocine	6.02	0.0629
9	N-2,4,6-Tetranitro-N-methylaniline	5.99	0.0590
10	Picric acid	7.03	0.0660
11	Pentaerythrytol tetranitrate	5.61	0.0570
12	Nitroglycerin	6.95	0.0700
13	Nitroguanidine	10.40	0.1019
14	Ethylene glycol dinitrate	8.72	0.0839
15	Diethylene glycol dinitrate	7.05	0.0689
16	Propylene glycol dinitrate	7.93	0.0769

Table 4. Topological Indices for Explosives (Figure 2 and Table 3).

Graph	SZ_e	SZ_p	Sz_eA	SZ_pA	SZ_eX	SZ_pX
2	594	4348	601.049	4169.958	74.949	510.786
3	360	2050	378.451	2049.049	57.524	333.915
4	348	1993	366.007	1978.160	49.430	290.101
5	296	1542	303.292	1500.896	49.076	268.231
6	516	3450	509.917	3239.271	66.837	431.457
7	516	3450	555.083	3528.729	68.038	439.675
8	1156	11794	1228.833	12083.000	100.785	824.478
9	1014	10342	1022.236	9969.632	99.907	820.577
10	594	4348	613.896	4248.014	78.289	528.620
11	968	11514	1083.000	12479.986	94.923	832.585
12	424	3677	475.729	3961.931	68.805	455.598
13	48	159	61.389	194.514	36.330	133.953
14	151	827	172.771	904.382	43.192	206.946
15	344	2518	408.569	2913.472	45.670	270.017
16	184	1153	214.278	1274.937	50.820	262.697

Table 5. Statistics of Single Variable Regression: Y = a + bX for Explosives and Cross Validation Test (Leave One Out)

No	Y	X	a	b	r	s	F	r_{cv}	s_{cv}
1	CD_{air}	ln SZ_e			0.943	0.0037	105.09		
2		1/ln SZ_eA	0.00073	0.4157	0.960	0.0031	155.21	0.951	0.0035
3		1/ln SZ_pA			0.954	0.0034	133.61		
4		1/W			0.942	0.0038	102.87		
5	CD_{water}	ln SZ_e	15.60561	- 1.3934	0.973	0.2801	231.72	0.960	0.338
6		ln SZ_eA	16.31334	- 1.4915	0.979	0.2443	310.57	0.968	0.302
7		ln SZ_pA	15.82862	-1.0766	0.984	0.2123	413.59	0.975	0.266
8		1/SZ_pX			0.961	0.3321	160.95		

Table 5 indicates a good estimative and predictive ability of the regression equations. It appears that the fragmental descriptors are more suitable for QSPR studies than the original Wiener and Szeged indices.

CONCLUSIONS

The novel Szeged fragmental indices are defined for both acyclic and cycle–containing molecular structures. They take into account multiple bonds as well as heteroatoms, by means of fragmental mass or by fragmental electronegativities. The presented correlations demonstrate their ability in modeling various molecular properties. They are promising tools in correlating the biological activity with the chemical structures, as shown in a forthcoming paper [19].

References

[1] H. Wiener, *Structural determination of paraffin boiling points*, J. Am. Chem. Soc. **69** (1947), 17–20.

[2] I. Gutman, Y. N. Yeh, S. L. Lee, Y. L. Luo, *Some recent results in the theory of the Wiener number*, Indian J. Chem. **32A** (1993), 651–661.

[3] S. Nikolić, N. Trinajstić, Z. Mihalić, *The Wiener index: Development and applications*, Croat. Chem. Acta **68** (1995), 105–129.

[4] M. V. Diudea, I.Gutman, *Wiener-type topological indices*, Croat. Chem. Acta **71** (1998), 21–51.

[5] M. Randić, *Novel molecular descriptor for structure–property studies*, Chem. Phys. Lett. **211** (1993), 478–483.

[6] M. Randić, X. Guo, T. Oxley, H. Krishnapriyan, *Wiener matrix: Source of novel graph invariants*, J. Chem. Inf. Comput. Sci. **33** (1993), 709–716.

[7] M. Randić, X. Guo, T. Oxley, H. Krishnapriyan, L. Naylor, *Wiener matrix invariants*, J. Chem. Inf. Comput. Sci. **34** (1994), 361–367.

[8] H. Hosoya, *Topological index. A newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons*, Bull. Chem. Soc. Japan **44** (1971), 2332–2339.

[9] M. V. Diudea, *Walk numbers ^eW_M : Wiener numbers of higher rank*, J. Chem. Inf. Comput. Sci. **36** (1996), 535–540.

[10] I. Gutman *A formula for the Wiener number of trees and its extension to graphs containing cycles*, Graph Theory Notes New York **27** (1994), 9–15.

[11] M. V. Diudea, O. M. Minailiuc, G. Katona, I. Gutman, *Szeged matrices and related numbers*, Commun. Math. Chem. (MATCH) **35** (1997), 129–143.

[12] R. A. Horn, C. R. Johnson, *Matrix Analysis*, Cambridge Univ. Press, Cambridge, 1985.

[13] M. V. Diudea, *Cluj matrix invariants*, J. Chem. Inf. Comput. Sci. **37** (1997), 300–305.

[14] M. V. Diudea, B. Parv, I. Gutman, *Detour–Cluj matrix and derived invariants*, J. Chem. Inf. Comput. Sci. **37** (1997), 1101–1108.

[15] M. V. Diudea, *Indices of reciprocal properties or Harary indices*, J. Chem. Inf. Comput. Sci. **37** (1997), 292–299.

[16] O. M. Minailiuc, G. Katona, M. V. Diudea, M. Strunje, A. Graovac, I. Gutman, *Szeged fragmental indices*, Croat. Chem. Acta **71** (1998), 473–488.

[17] M. V. Diudea, I. E. Kacso, M. I. Topan, *Molecular topology. 18. A QSPR/QSAR study by using new valence group carbon–related electronegativities*, Rev. Roum. Chim. **41** (1996), 141–157.

[18] G. Rücker, C. Rücker, *On topological indices, boiling points, and cycloalkanes*, J. Chem. Inf. Comput. Sci. **39** (1999), 788–802.

[19] I. Schirger, M. V. Diudea, Studia Univ. Babes–Bolyai, Ser. Chem., to appear.