Molecular Descriptors for Natural Diamondoid Hydrocarbons and Quantitative Structure-Property Relationships for Their Chromatographic Data

Alexandru T. Balaban*[,a], Douglas J. Klein[a], Jeremy E. Dahl[b] and Robert M. K. Carlson[b]

[a]*Texas A&M University at Galveston, 5007 Avenue U, Galveston, TX 775651, USA*

[b]*MolecularDiamond Technologies, Chevron Technology Ventures, 100 Chevron Way, Richmond, CA, USA*

Abstract: Diamond hydrocarbons (or diamondoids) are hydrocarbons that have a carbon skeleton superimposable on the diamond lattice and contain one or more adamantane units. Recently it was found that many higher diamondoids (containing four to eleven adamantane units) are present in petroleum and can be isolated by a series of methods that include HPLC and GC techniques. We develop QSPR equations using molecular descriptors derived from the topology and geometry of diamondoids, by means of dualist graphs consisting of vertices placed at the centers of adamantane cells forming the diamondoid, and of edges connecting vertices centered in adamantane cells sharing faces. From distance (or distance-distance) matrices encoding the topology and geometry of diamondoids one can obtain distance-sums, distance-distance sums, or eigenvectors as molecular descriptors characterizing the diamondoids. These descriptors afford satisfactory correlations with GC and HPLC retention data, and may also facilitate the identification of diamondoid isomers.

INTRODUCTION

The surprising discovery and isolation of higher diamondoids (or polymantanes, hydrocarbons whose carbon scaffold is part of the diamond lattice) was recently reported [1,2]. Consequently one desires theoretical methods for correlating structures with properties for these unique and useful compounds, only available from petroleum [3]. Many applications of "nanotube diamondoids" in nano-devices and medicine have been proposed. Chiral diamondoids may be used in stereoselective syntheses. Individual diamondoids with many potential applications (such as antiviral compounds, field emitters, components of molecular motors, nanorods as spacers in pharmaceuticals), being the most stable among their isomers, can be isolated *via* thermal treatments. Diamondoids are more thermally stable than nondiamondoid hydrocarbons (e.g., acyclic hydrocarbons) in petroleum systems, and therefore become concentrated in the natural gas condensates as they form over geologic time [3].Thermal treatments are similarly used to produce fractions extremely rich in diamondoids [1,2]. Moreover, diamondoids are more stable than graphite in the sub-5 nm size range (contrary to the stability relationship of macroscopic diamond and graphite).

MolecularDiamond Technologies of Chevron Technology Ventures measured various physical properties of the higher diamondoids during the development of methods for their isolation. These data are now used in this quantitative structure-property relationship (QSPR) study. Higher diamondoid hydrocarbons are 1 – 2 nm-sized structures that have carbon scaffolds corresponding to parts of the diamond lattice. Eleven higher diamondoids (not counting enantiomers) are shown in (Fig. **1**), four on a grey octahedral diamond lattice and also as separate carbon framework structures, with higher members of their respective series radiating out from the lattice.

*Address correspondence to this author at the Texas A&M University at Galveston, 5007 Avenue U, Galveston, TX 775651, USA;
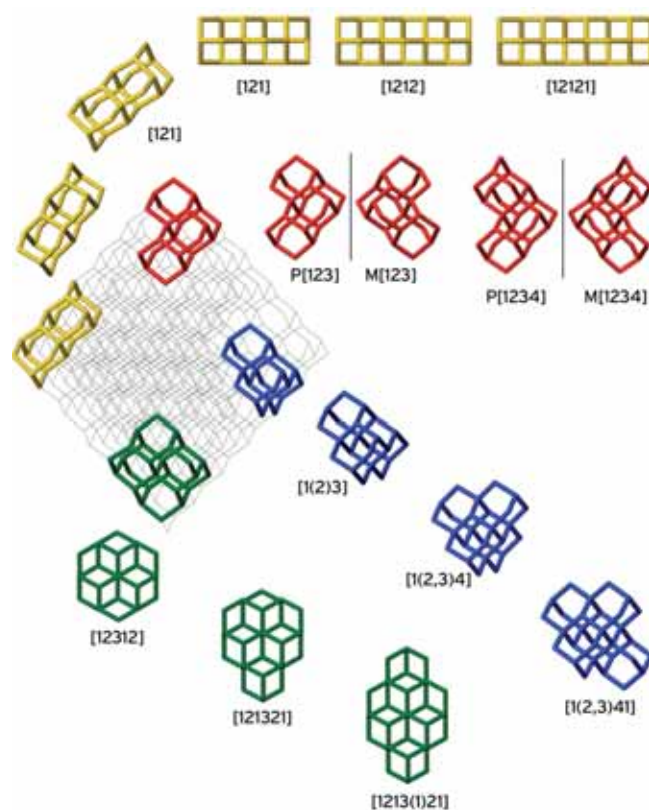E-mail: balabana@tamug.edu

Fig. (1). Examples of separate structure series of higher diamondoids indicated by four different colored carbon frameworks superimposed on a grey diamond lattice and shown separately with their notation, and with successive "adamantanologs". In the case of the red-colored chiral diamondoids (having primary helicity), enantiomers are displayed with **P/M** notations.

Thence it seems appropriate that a systematic study of diamondoid structures and diamondoid property characterization be developed. This is initiated here – in particular, using the systematic nomenclature based on dualist graphs, and the complicit structural characterization is further for-

malized with idealized coordinates, and MM2 optimized coordinates, all of which are used to compute distance-matrix-related invariants useful in correlations with chromatographic retention times.

STRUCTURAL CHARACTERIZATION OF DIA-MONDOIDS AND NOMENCLATURE

Several decades ago, for encoding the structure of condensed benzenoids, dualist graphs have been used: they consist of vertices placed at the centers of each benzenoid ring and of edges connecting vertices corresponding to rings sharing a CC bond. Unlike the usual graphs in which bond angles do not matter, in dualists they do. Cata-condensed benzenoids (that contain no carbon atom common to three rings and no internal hydrogen atoms) have acyclic dualists, peri-condensed benzenoids (that contain carbon atoms common to three rings) have dualists with 3-membered rings, and corona-condensed benzenoids (that have "holes" with internal hydrogen atoms) have larger rings in their dualists. A coding system was developed for cata-condensed benzenoids based on the geometry of their dualists.

By analogy, diamondoid hydrocarbons have dualist graphs consisting of vertices placed at the centers of adamantane cells forming the diamondoid and of edges connecting vertices of adamantane cells sharing faces. The dualists of diamondoids have structures corresponding to staggered alkane or cycloalkane rotamers. A coding system was proposed for specifying uniquely the structure of diamondoids by associating the tetrahedral directions with digits 1 through 4, indicating these directions from one end of the alkane chain to the other end, to and choosing among all possible notations that one which yields the smallest number on reading sequentially the digits. A few other simple rules were formulated for cyclic or polycyclic dualists, and these rules were also relaxed when wishing to indicate not only the constitution but also the stereoisomerism. Adamantane, diamantane and triamantane have but one constitutional isomer, but higher diamondoids have two or more isomers, as shown in Table **1**.
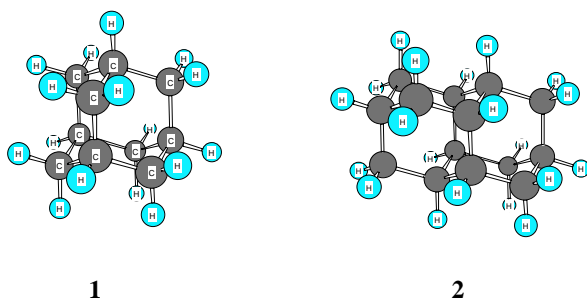

1   **2**

Fig. (2). Adamantane and diamantane with their carbon (grey) and hydrogen atoms (blue).

Adamantane ($C_{10}H_{16}$) (**1**) was isolated in the 1930s from petroleum where it is present in low amounts. Labor-intensive chemical syntheses followed in the 1940s. The breakthrough occurred in 1957 when the serendipitous catalyzed isomerization of many other $C_{10}H_{16}$ tricyclic hydrocarbons to yield adamantane was discovered by Schleyer [4,5]. The complicated mechanisms involving 1,2-rearrangements of carbocations were elucidated by computations [6,7] and reaction graphs [8,9]. Soon afterwards, the next homologues

up to triamantane were prepared *via* analogous isomerizations. Diamondoids are the most stable among all their isomers. However, no chemical syntheses are available for higher diamondoids than triamantane, other than one of the four tetramantanes, i.e., *C2h*-tetramantane, but many of them have now been isolated from petroleum and are available for scientific research from MolecularDiamond Technologies. Adamantane (**1**) and diamantane (**2**) are presented in (Fig. **2**).

Balaban and Schleyer [10] published a system for the enumeration and nomenclature of diamondoids that is being currently used by MolecularDiamond Technologies Co.[1-3] The IUPAC name of adamantane (**1**) is tricy-clo[3.3.1.1^{3,7}]decane and the IUPAC name of diamantane (with two fused adamantane units) is pentacy-clo[7.3.1.1^{4,12}.0^{2,7}.0 ^{6,11}]tetradecane (**2**). The names and numbering schemes according to the IUPAC nomenclature for the higher diamondoids are exceptionally convoluted and prone to error. The coding and nomenclature system for diamondoids [10] uses the "dualist graph" and was first employed for benzenoids [11-14]. In the center of each unit (benzenoid ring or adamantane unit) one places a vertex of the dualist graph, and then one connects vertices sharing a bond (for benzenoids) or a face (for diamondoids). Dualist graphs
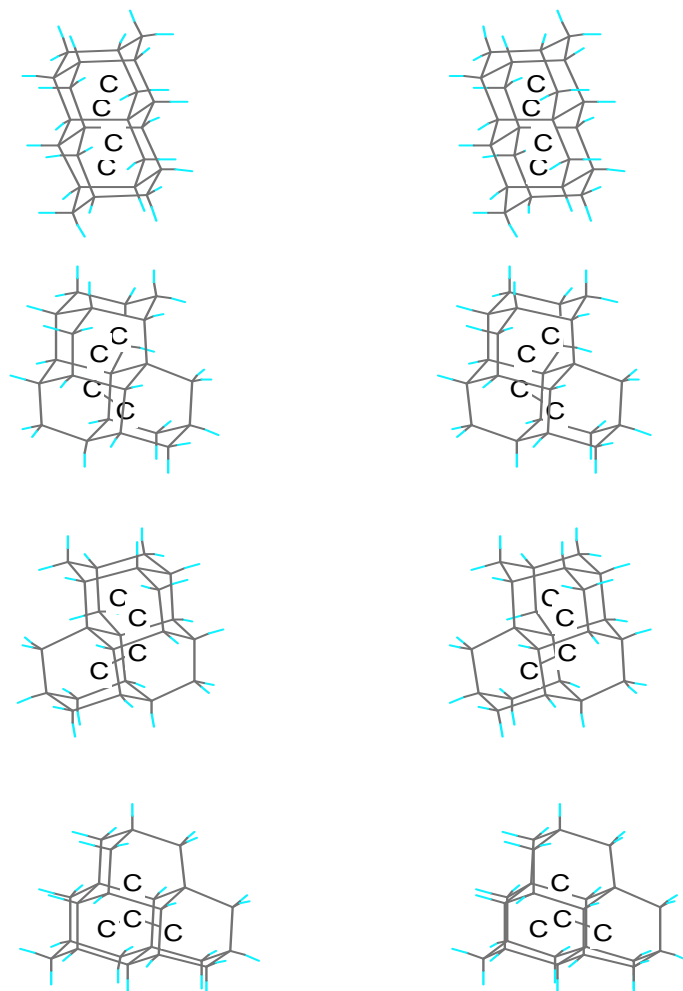

Fig. (3). Stereo-views of the tetramantane isomers. From top to bottom: the achiral [121]-, the two enantiomers [123]- and [124]-, and the achiral [1(2)3]-tetramantane.

differ from usual graphs in that the bond angles and bond lengths do matter. The stereo-view of one from the three possible tetramantane isomers (having four adamantane units) with dualist graph C-C-C-C in staggered conformation gives the new name for the dualist graph (one of the staggered rotamers of butane, i.e., *anti* [15]), namely "[121]", much simpler than that the name according to the IUPAC system. Its distance matrix will be shown later. In (Fig. **3**) we present as examples the two achiral constitutional isomers and the pair of enantiomeric tetramantanes with their dualists formulated as the carbon scaffold of the corresponding staggered isomers/staggered rotamers of butane.

The four tetrahedral directions are translated into digits 1–4 and the canonical name results as the smallest number formed from these digits along the longest linear carbon chain of the dualist graph; branches are included in brackets. We propose to specify substituent locations and geometries by position numbering, and substituent orientations by letters a–d (associated with numbers 1-4, see Figs. **4-6**). More detailed information on this topic will be presented in a separate paper.
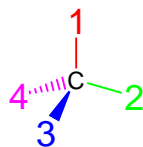

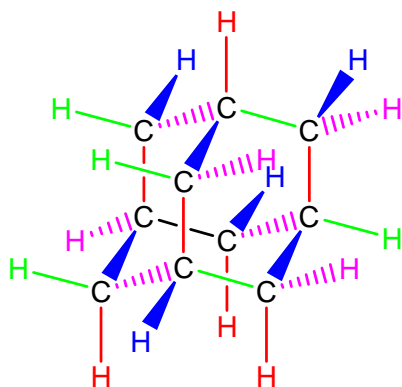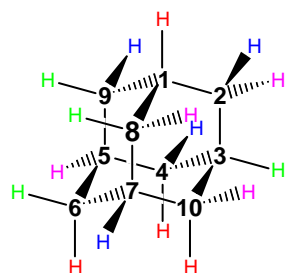
Fig. (4). The four tetrahedral directions.

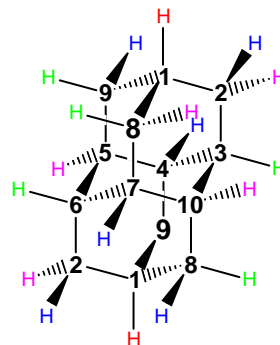**Fig. (5).** Adamantane with the 4 x 4 hydrogens oriented according to directions 1,2,3 and 4.

Table **1** presents all the possible constitutional isomers of diamond hydrocarbons having up to $n = 6$ adamantane units, with their codes and symmetry groups. Many of them are chiral, and give rise to enantiomers. Interestingly, separation of such enantiomers was possible for the naturally occurring diamondoids [1].

The dualist graphs of *catamantanes* may be represented by the carbon atoms of staggered rotamers of non-branched or branched alkanes (in the following figures, such carbon chains denote the dualist graph of diamondoids). In the *Tetrahedron* paper [10], perimantanes (whose dualist graphs contain rings, the smallest of which has 6 adamantane units) were also discussed. By analogy with benzenoids, diamondoids are classified into catamantanes (which are subdivided into *regular catamantanes* with acyclic dualist graphs, all systems $C_{4n+6}H_{4n+12}$ with the same number of carbon atoms and adamantane units being isomeric, and *irregular catamantanes* with quasi-cyclic dualist graphs, having lower numbers of C and H atoms) and *perimantanes* (with cyclic dualist graphs). Each of these classes may have a linear or a branched dualist graph. In the preceding table, all diamondoids with up to $n = 6$ adamantane units are described. Irregular catamantanes occur when their dualist graphs have sequences coded as 1231. Irregular branched catamantanes are possible starting with $n \geq 7$.

It is possible to code not only for the constitution (imposing the condition of the minimal number) but also for the stereochemistry of diamondoids (relaxing this condition with specified restrictions, as shown with one example in Fig. **3**).

The most fundamental physical properties are vapor pressures, melting points, and enthalpies of fusion. Although correlations with melting points are less reliable than those with boiling points, critical data, odd/even alternation of melting points, and enthalpies of fusion of alkanes have been modeled [16,17]. The gas chromatographic (GC) and high-performance liquid chromatography (HPLC) elution volumes for about 100 diamondoids were provided by ChevronTexaco, including data for 17 identified polymantanes: all tetramantanes and pentamantanes, and a few polymantanes with 6 – 11 adamantane units. An important task is to determine if it is possible to predict the structures of the larger components from their experimentally-determined properties.

The first properties to be examined will be GC retention times and HPLC elution volumes. For rod-like all-*anti* diamondoids it will be interesting to find QSPR equations; if

Fig. (6). A. Numbering of adamantane carbon atoms. **B**. Numbering of diamantane carbon atoms.

Table 1. Catamantanes with n=1-6 Adamantane Units

	Catamantanes									
	Regular					Irregular				
	linear			branched		linear			branched	
n	Code	Symm.	Formula	Code	Symm.	Code	Symm.	Formula	Code	Symm.
1		T_d	$C_{10}H_{16}$							
2	1	D_{3d}	$C_{14}H_{20}$							
3	12	C_{2v}	$C_{18}H_{24}$							
4	121	C_{2h}	$C_{22}H_{28}$	1(2)3	C_{3v}					
	123	C_2								
5	1212	C_{2v}		12(1)3	C_1					
	1213	C_1	$C_{26}H_{32}$	12(3)4	C_s	1231	C_s	$C_{25}H_{30}$		
	1234	C_2		1(2,3)4	T_d					
6	12121	C_{2h}		121(2)3	C_1					
	12123	C_1		12(1)32	C_1					
	12131	C_1		121(3)4	C_s				12(1)31	C_s
	12134	C_1	$C_{30}H_{36}$	12(1)34	C_1				123(1)2	C_1
	12321	C_i		12(1,3)4	C_1	12132	C_1	$C_{29}H_{34}$	123(1)4	C_1
	12324	C_2		12(3)12	C_s	12314	C_1		12(3)41	C_s
	12341	C_2		1(2)3(1)2	C_{2h}					
				1(2)314	C_1					
				12(3)14	C_1					
				1(2)3(1)4	C_{2h}					

mondoids it will be interesting to find QSPR equations; if there is any odd/even alternation, as for the melting points of alkanes [17], this could also be taken into account. It may be possible to explain why "cyclohexamantane" has a surprisingly high melting point that reverses the trend observed among lower diamondoids [2]. The coding of diamondoid dualist graphs (identical to the coding of staggered alkane rotamers that had been introduced earlier [15]) will be used for QSPR.

EUCLIDEAN DISTANCE MATRICES FOR OPTIMIZED GEOMETRIES

Molecular descriptors, adapted for describing stereochemical relationships [18], are the simplest tools that may be used in QSPR studies: topological-geometrical indices derived from distance matrices of the dualist graphs of diamondoids. Euclidean distance matrices have entries d_{ij} which are geometrical distances between vertices i and j (in Å). From such matrices one can obtain various molecular descriptors (*e.g.,* distance-sums of entries, distance-square sums, sums of ratios between geometrical and topological distances, eigenvalues, etc.). And using such matrices relevant to the dualist is simpler than using that for the whole dia-

mondoid. The vertex degrees of the dualist graphs indicate the number of adjacent carbon atoms in the alkane or cycloalkane rotamer that models the dualist graph. Thus, for the three possible tetramantanes the dualist graphs are modeled by the two rotamers of n-butane (*trans* and *gauche*) and by isobutane, as seen in (Fig. **3**) showing from top to bottom: [121]-, [123]-, [124]- and [1(2)3]-tetramantane. It should be noted that the *gauche* dualist graphs and the corresponding [123]- and[124]-tetramantanes are chiral. The two enantiomers of this tetramantane can be coded as [123] and [124], relaxing the condition for the smallest number. Alternatively, one may use **P** and **M** prefixes as in (Fig. **1**).

For these three possible tetramantanes, the geometrical distance matrices for their dualists (in Å), computed with the CambridgeSoft Chem3DPro MM2 program (taking into account steric repulsions in [1231] and higher similar alkanes, with codes …*abca*…possibly including also round brackets), are indicated in Table **2**. There are also given derivative quantities, including the D-sums (as the sum of all the elements in the Euclidean distance matrix **D)** and the largest positive eigenvalue. It should be stressed that we are not dealing with topological distances (number of edges between vertices along the shortest path). One should note that it is

Table 2.   Euclidean Distance Matrices of the Three Isomers of Tetramantane

		Tetra-	121					
	1	**2**	**3**	**4**	**Sum**	**Degree**	**D-Sum**	**Eigenval.**
1	0	1.54	2.54	3.92	8.00	1	27.24	.6.980
2	1.54	0	1.54	2.54	5.62	2		.
3	2.54	1.54	0	1.54	5.62	2		.
4	3.92	2.54	1.54	0	8.00	1		.
		Tetra-	123					
	1	**2**	**3**	**4**				
1	0	1.54	2.54	3.14	7.22	1	25.68	6.498
2	1.54	0	1.54	2.54	5.62	2		
3	2.54	1.54	0	1.54	5.62	2		.
4	3.14	2.54	1.54	0	7.22	1		
		Tetra-	1(2)3					
	1	**2**	**3**	**4**				
1	0	1.54	2.54	2.54	6.62	1	24.48	5.897
2	1.54	0	1.54	1.54	4.62	3		
3	2.54	1.54	0	1.54	6.62	1		
4	2.54	2.54	1.54	0	6.62	1		

also possible to assign integer coordinates to carbon atoms in staggered rotamers of alkanes and cycloalkanes, as shown later. The derived geometric distance matrices are simpler, but the geometry is strictly that of the diamond lattice, ignoring steric repulsions that open up bond angles slightly in [1231] pentamantane and related higher diamondoids.

The vertex degree indicates how many vertices are adjacent to the specified vertex. Primary carbon atoms in the associated dualist graph have degree 1, secondary carbon atoms have degree 2, and tertiary carbon atoms have degree 3.

As shown in an earlier table, there are seven possible pentamantanes, one of which, [1231]-pentamantane, is irregular because it has a dualist graph which is quasi-cyclic leading to a different molecular formula ($C_{25}H_{30}$) than that of the other pentamantanes ($C_{26}H_{32}$) required by the general formula $C_{4n+6}H_{4n+12}$. Fig. (**7**) presents stereo-views of the four pentamantanes with non-branched dualist graphs. Their codes are (from top to bottom): [1212, 1213, 1231, 1234].

It may be observed that for the irregular quasi-cyclic isomer [1231], steric hindrance between hydrogen atoms causes a slight distortion of the molecule such that the terminal CC bonds of the dualist are no longer parallel.

Among the conventions adopted in devising the coding system [11-14], it was stipulated that for the code of branched systems, whenever several ends could be selected for beginning the numbering and resulting in the same set of digits, the preferred end should place the branch as far to the right as possible. Thus, instead of [1(2)31] one should prefer [12(1)3], and instead of [1(2)34] one should prefer [12(3)4].
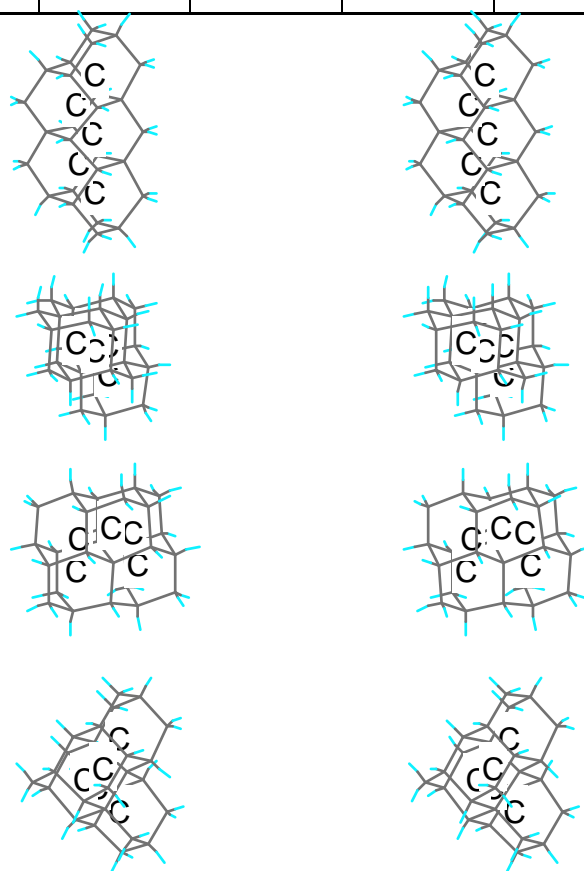

Fig. (7). Stereoviews of the "nonbranched" isomers of pentamantane.

In (Fig. **8**), the remaining pentamantanes with branched dualist graphs are shown as stereo-views. Their codes are (from top to bottom): [12(1)3], [12(3)4], and [1(2,3)4].
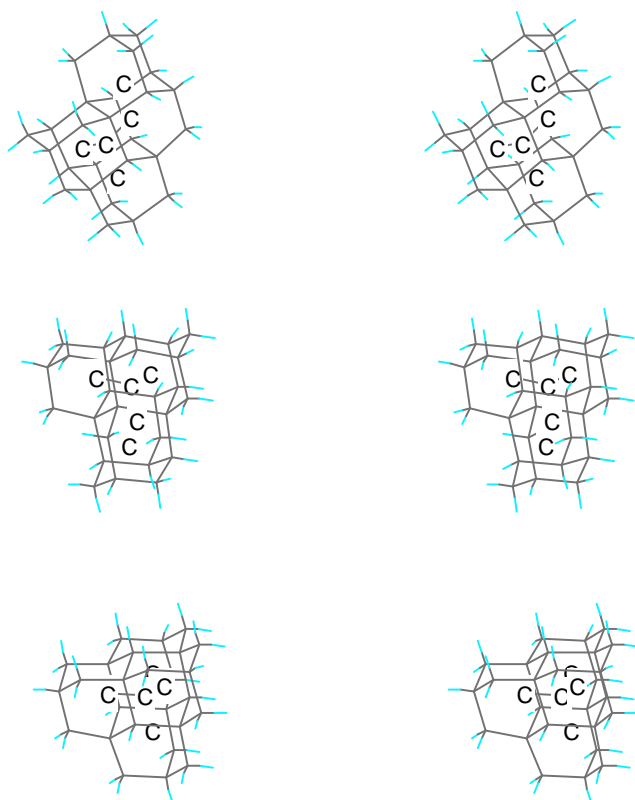

Fig. (8). Stereo-views of the "branched" pentamantanes.

For the six possible *regular* pentamantanes, the geometrical distance matrices are shown in Table **3**. Entries are geometrical distances (in Å) between carbon atoms in the staggered conformations of alkane rotamers, as they result by optimizing the geometry according to molecular mechanics force fields (MM2). From these distance matrices **D** one can obtain various molecular descriptors for use in QSPR equations.

INTEGER COORDINATES OF CARBON ATOMS IN THE DIAMOND LATTICE

One can consider that diamondoid hydrocarbons consist of adamantane units which have each a cubic unit cell (Fig. **9**), though one may also look at the diamond lattice in the canonical orientation (Fig. **10**) as consisting of diamantane units (**2**), all of which have the same orientation (Fig. **4**). Moreover, each unit cell as in (Fig. **9**) may be viewed as comprised from 8 smaller cubic (sub-unit) cells, which further can be partitioned into 8 smaller cubic cells, whereupon all the carbon atoms sit at the corners of these smallest unit cells, and so have integer coordinates with respect to the associated cubic lattice, though there are corners of this cubic lattice not occupied by atoms.

To be more precise about all this, let us start out by defining four sets of sites selected from the integer coordinates of a cubic lattice

$$S_\xi \equiv \{(a,b,c) : a,b,c \text{ all odd } \& \ a+b+c = 4n+\xi\} \quad ,\xi = 1,3$$

$$S_\xi \equiv \{(a,b,c) : a,b,c \text{ all even } \& \ a+b+c = 4n+\xi\} \quad ,\xi = 0,2$$

That is, the coordinates in any one set are to all be of the same parity and have a fixed modulo-4 sum. Clearly the sets are disjoint and partition the set S_{parity} of all cubic lattice sites with all coordinates of the same parity. We write this

$$S_{parity} = S_0 \oplus S_1 \oplus S_2 \oplus S_3$$

Table 3.   Euclidean Distance Matrices (Å) of the Regular Isomers of Pentamantane

		Penta-	1212							
	1	**2**	**3**	**4**	**5**	**Sum**	**Degree**	**D-Sum**	**Eigenval**
1	0	1.54	2.54	3.92	5.08	13.08	1	53.40	10.998
2	1.54	0	1.54	2.54	3.92	9.54	2		
3	2.54	1.54	0	1.54	2.54	8.16	2		
4	3.92	2.54	1.54	0	1.54	9.54	2		
5	5.08	3.92	2.54	1.54	0	13.08	1		
		Penta-	1213							
	1	**2**	**3**	**4**	**5**				
1	0	1.54	2.54	3.14	4.58	11.8	1	50.84	10.411
2	1.54	0	1.54	2.54	3.92	9.54	2		
3	2.54	1.54	0	1.54	2.54	8.16	2		
4	3.14	2.54	1.54	0	1.54	8.76	2		
5	4.58	3.92	2.54	1.54	0	12.58	1		

(Table 3). Contd…..

		Penta-	1234						
	1	**2**	**3**	**4**	**5**				
1	0	1.54	2.54	3.14	3.92	11.14	1	47.96	9.732
2	1.54	0	1.54	2.54	3.14	8.76	2		
3	2.54	1.54	0	1.54	2.54	8.16	2		
4	3.14	2.54	1.54	0	1.54	8.76	2		
5	3.92	3.14	2.54	1.54	0	11.14	1		
		Penta-	12(1)3						
	1	**2**	**3**	**4**	**5**				
1	0	1.54	2.54	3.92	2.54	10.54	1	46.7	9.539
2	1.54	0	1.54	2.54	1.54	7.16	3		
3	2.54	1.54	0	1.54	2.54	8.16	2		
4	3.92	2.54	1.54	0	3.11	11.11	1		
5	2.54	1.54	2.54	3.11	0	9.73	1		
		Penta-	12(3)4						
	1	**2**	**3**	**4**	**5**				
1	0	1.54	2.54	3.14	2.54	9.76	1	45.2	9.172
2	1.54	0	1.54	2.54	1.54	7.16	3		
3	2.54	1.54	0	1.54	2.54	8.16	2		
4	3.14	2.54	1.54	0	3.14	10.36	1		
5	2.54	1.54	2.54	3.14	0	9.76	1		
		Penta-	1(2)(3)4						
	1	**2**	**3**	**4**	**5**				
1	0	1.54	2.54	2.54	2.54	9.16	1	42.8	8.709
2	1.54	0	1.54	1.54	1.54	6.16	4		
3	2.54	1.54	0	2.54	2.54	9.16	1		
4	2.54	1.54	2.54	0	2.54	9.16	1		
5	2.54	1.54	2.54	2.54	0	9.16	1		

(where $A = B \oplus C$ means that $A = B \cup C$ & $B \cap C = \varnothing$). Evidently each component of S_{parity} comprises 1/4 of all cubic lattice sites (with integer coordinates).

Next define two dual 4-sets of *interconnection* vectors

$$\Delta \equiv \{(1,1,1),(1,-1,-1),(-1,1,-1),(-1,-1,1)\}$$
$$\{(-1,-1,-1),(-1,1,1),(1,-1,1),(1,1,-1)\} \equiv -\Delta$$

Associated to each of these vectors define transformations τ_{+i} and τ_{-i} such that to the coordinates of one of the sets S_ξ (for $\xi = 0, 1, 2, 3$) one respectively adds the *i*th vectors of Δ and Δ_*. That is,

$$\tau_{+i}(a,b,c) = (a,b,c) + \vec{\delta}_i$$
$$\tau_{-i}(a,b,c) = (a,b,c) - \vec{\delta}_i$$

with

$$\vec{\delta}_0 \equiv (1,1,1), \ \vec{\delta}_1 \equiv (1,-1,-1), \ \vec{\delta}_2 \equiv (-1,1,-1), \ \vec{\delta}_3 \equiv (-1,-1,1).$$

constituting the members of Δ. Then

$$\tau_{\pm i} : S_\xi \to S_{\xi \mp 1}$$

where the subscripts ξ entail modulo-4 operations. That is, τ_{+i} cycles the site sets S_0, S_1, S_2, S_3 in a "counter-clockwise" direction amongst themselves, and τ_{-i} cycles them in a "clockwise" direction.

Now the *diamond lattice* $\mathsf{L}_{diamond}$ may be chosen to consist of site and edge sets
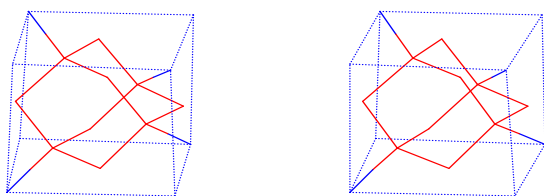
Fig. (9). Stereo-view of the cubic unit cell with an adamantane skeleton inside.
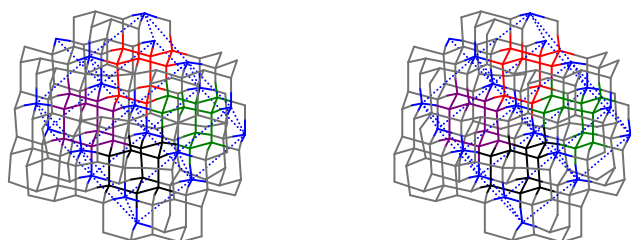

Fig. (10). Stereoview of the diamond lattice with cubic unit cells.

$$V(L_{\text{diamond}}) \equiv S_1 \oplus S_2$$
$$E(L_{\text{diamond}}) \equiv \{\{(a,b,c),(a,b,c)-\vec{\delta}\}:(a,b,c)\in S_1,\vec{\delta}\in\Delta\}$$
$$= \{\{(a,b,c),(a,b,c)+\vec{\delta}\}:(a,b,c)\in S_2,\vec{\delta}\in\Delta\}$$

while the *dual lattice* L_{dual} consists of site and edge sets

$$V(L_{\text{dual}}) \equiv S_0 \oplus S_3$$
$$E(L_{\text{dual}}) \equiv \{\{(a,b,c),(a,b,c)+\vec{\delta}\}:(a,b,c)\in S_3,\vec{\delta}\in\Delta\}$$
$$= \{\{(a,b,c),(a,b,c)-\vec{\delta}\}:(a,b,c)\in S_0,\vec{\delta}\in\Delta\}$$

Indeed any of the 4 "adjacent" pairs $S_\xi \oplus S_{\xi+1}$ can be viewed as the sites of a diamond lattice.

Following ideas outlined earlier, a general diamondoid may be conveniently specified in terms of the subset of dual lattice sites, each one of which is enclosed by an adamantane

unit. Thence given a site subset $s_{\text{dual}} \subseteq V(L_{\text{dual}})$ one may determine the associated diamondoid structure as the subgraph of L_{diamond} induced by the sites of $s_{\text{diamond}} \subseteq V(L_{\text{diamond}})$ suitably surrounding each of the sites of s_{dual}. With some further definitions this can be made more precise, and some matters of choice introduced as to where to translationally and rotationally locate a diamondoid on the diamond lattice.

For a subset $s \subseteq S_{\text{parity}}$ of N sites $\vec{r}_i \equiv (a_i, b_i, c_i)$, $i = 1, 2, ..., N$, identify an associated ordered set $\vec{s}$ where the site vectors $\vec{r}_i$ are ordered such that when each $\vec{r}_i$ is viewed as a number, then these numbers increase. That is, in $\vec{s}$ the sites $\vec{r}_i$ are lexicographically ordered, and may be viewed as a *code* for the set. Also for a set $s \in S_{\text{parity}}$ define

$$s_\xi \equiv s \cap S_\xi \quad, \xi = 1, 2, 3, 0$$

Further define

$$\Delta_* \equiv \{(0,0,-2),(0,0,2),(0,-2,0),(0,2,0),(0,0,-2),(0,0,2)\}$$

which consists of non-zero vectors in $\Delta + \Delta \equiv \{\vec{\delta} + \vec{\delta}': \vec{\delta}, \vec{\delta}' \in \Delta\} = -\Delta - \Delta$, and so when added to a site vector of $V(L_{\text{dual}})$ gives back another member of $V(L_{\text{diamond}})$, and conversely. If $(a,b,c) \equiv \vec{r}$ is a dual lattice site $\in S_0$, then the surrounding adamantane sites (in L_{diamond}) are $\vec{r} - \vec{\delta}$, $\vec{\delta} \in \Delta \oplus \Delta_*$; and if $\vec{r} \in S_3$, then the surrounding adamantane sites (in L_{diamond}) are $\vec{r} + \vec{\delta}$, $\vec{\delta} \in \Delta \oplus \Delta_*$.

One diamondoid conformation may be embedded in the diamond lattice in different ways, as mediated by different translations and rotations of any one embedding. Each different embedding corresponds to a different subset of dual-lattice sites, and it is the dual-lattice designation which involves fewer sites, especially for smaller less peri-condensed diamondoids, so that we here wish to use a dual-lattice based description. But still the problem of embeddings of the dual lattice sites remains.

And there is yet a further description of the diamond lattice. One may imagine the construction in terms of S_1 and S_2,
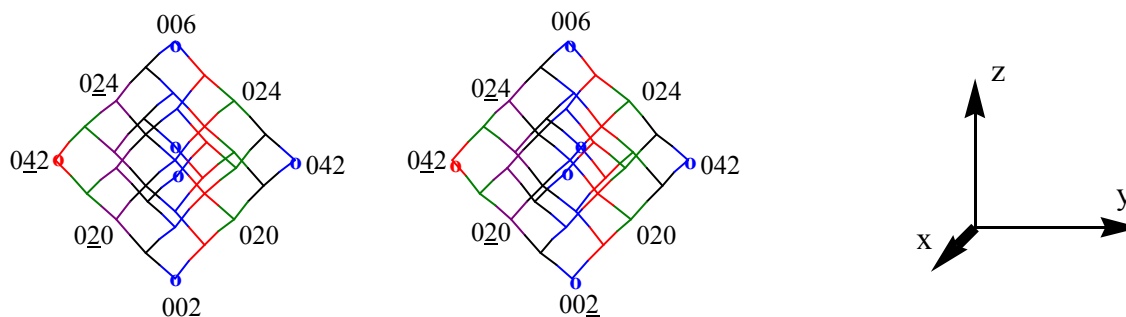

Fig. (11). Stereo-view of a few sites on the diamond lattice (different colors indicate different *y* values).

but without taking edges along one of the directions (say $\vec{\delta}_0$ = (1,1,1) of Δ, in which case the network falls into disjoint graphitic honeycomb nets, which however are crinkled with alternant sites a little on opposite sides of the mean plane of their honeycomb net. Adjacent nets have alternant sites set one over another, as viewed from the deleted direction, with the displacement of the sites from the mean planes of the nets towards one another. Thence one can imagine so arranging such honeycomb nets and interconnecting such nearer sites in adjacent pairs of nets, so as to obtain the diamond lattice.
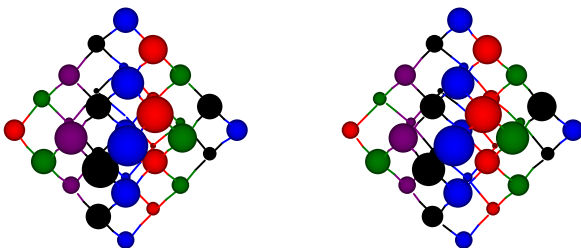

Fig. (12). Stereo-view of a few sites on the diamond lattice (different colors indicate different *y* values) seen from a point with high positive *x* value.

As an example we present in (Figs. **11** and **12**) a portion of the diamond lattice with the *x*, *y*, and *z* coordinates for each carbon atom. For brevity, negative numbers are underlined. Sites with the same *y* value have the same color.

With the same color convention, (Fig. **13**) displays adamantane (**1**) and diamantane (**2**). In (Fig. **14**) we present the perifused decamantane with an adamantane-like dualist graph, but hydrogen atoms are omitted for a better view of the carbon atoms and the dualist graph.
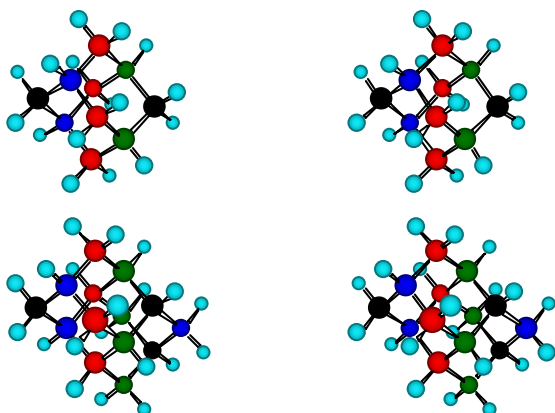

Fig. (13). Stereo-views of adamantane (upper row) and diamantane (lower row) with the same color convention as in Figs. **9** and **10**.

In (Fig. **15**) we display a portion of the diamond lattice (using the same color convention as in Figs. **11-14**) with the values of *xyz* coordinates for each carbon atom.
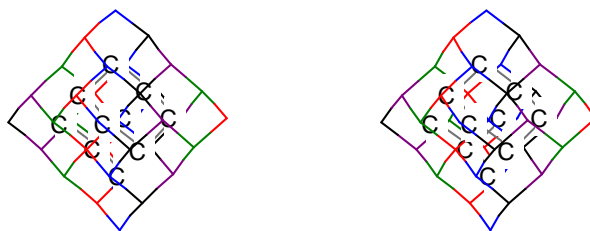
Fig. (14). Stereoview of the decamantane whose dualist graph (symbolized by C atoms) is isomorphic to adamantane (hydrogen atoms have been omitted for clarity).

With the integer-coordinate representation one can easily compute Euclidean geometric distances either between carbon atoms or "dualist" network sites in the idealized geometry for diamondoids. This distance between sites *i* and *j* generally is

$$d_{ij} = [(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2]^{1/2}.$$

Thence with each coordinate difference =1, one obtains a C–C bond length of $\sqrt{3}$ (in our current units). And the same formula applies for arbitrary centers *i* and *j*. The sum

$$1/n \sum_{i<j}^{D} d_{ij}^2 \equiv I_D$$

gives the diamondoid's mean moment of inertia (neglecting the contributions from the less massive hydrogen atoms, and taking the carbons of unit mass). And the sum over the dualist network pairs of sites gives a mean inertia estimator,

which (especially for smaller diamondoids) is even simpler. Instead of the distance matrices for MM2 optimized geometry presented earlier, one may use distance matrices based on idealized integer coordinates for the dualist network choice that are presented in Table **4**. With the use of the distance matrices various invariants may be computed. It may be mentioned that the integer-coordinate distance matrix of a diamondoid (or even its dual-distance matrix) faithfully (*i.e.*, uniquely) specifies the diamondoid up to enantiomerism. Further comparisons [19-21] may be made here, as in [20], utilizing both geometric distances d_{ij} and shortest-path (or through-bond) graphical distances ρ_{ij}. The maximum eigenvalue of the distance-distance matrix (**DD**-matrix) with elements d_{ij}/ρ_{ij} provides a useful graphical-geometric invariant, the **DD**-eigenvalue.

EUCLIDEAN DISTANCE MATRICES FOR IDEALIZED GEOMETRIES WITH INTEGER COORDINATES

Instead of the distance matrices for MM2 optimized geometry presented earlier, one may use distance matrices based on idealized integer coordinates for the dualist lattice choice; they are presented in Table **4**.

GC RETENTION TIMES, HPLC ELUTION VOLUMES, AND INTERRELATIONSHIPS BETWEEN THESE DATA

Experimentally-determined gas-chromatographic (GC) retention times and elution volumes for high-performance

Add to existing coordinates, according to eight directions:

Fig. (15). A portion of the diamond lattice with integer coordinates (with the same color convention as in Figs. **11-14**). The origin (000) is in the center of the lower leftmost adamantane unit. Negative signs are indicated by underlined digits. Antipodal vertices for the three pairs of vertices in this unit differ in the sign of *xyz* coordinates (± 2).
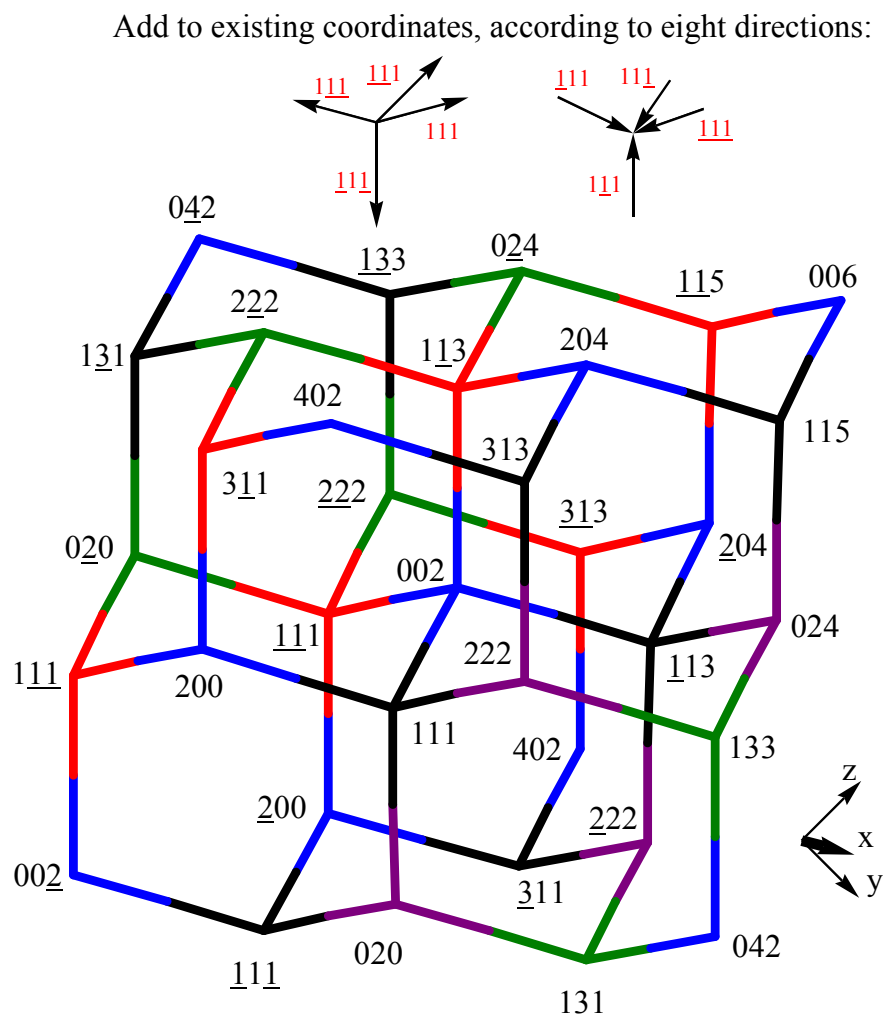
liquid chromatography (HPLC) of diamondoids are presented in Table **5**.

GC retention times were determined using a combined gas-chromatograph-mass spectrometer (GC-MS) consisting of a Agilent model 5973 Network Mass Selective Detector equipped with an Agilent model 6890 gas chromatograph set-up to run capillary columns in splitless sample injection mode with an Agilent model 7683 autosampler. The GC column used was an Agilent HP-MS5 (30m by 0.25 mm I.D., 0.25μ phase thickness) run with a helium carrier gas flow rate of 1.2 mL/min. (constant flow mode) and inlet pressure of 16 psi. During GC-MS analyses, the GC oven had a 1.0 min initial hold time at 150°C followed by oven programming at 10°C/min to 320°C with a final hold time of 15 min. The GC-MS transfer line temperature was maintained at 320 °C. A solvent delay of 2.5 min was applied before GC-MS data accumulation. In selected ion monitoring (SIM) mode dwell time = 100 msec. Scan Mode ran from m/z 50 to m/z 550 at 2.94 scans/sec. HPLC was performed on a Waters Prep 4000 Series instrument equiped with Whatman octadecyl silane, ODS-2 (C18), 10 micron particle size, two 500 mm by 22 mm I.D. columns (two in series), using acetone mobile phase at 5 ml/min. The detector was a Waters Model 2410 Differential refractometer. Fractions where taken every 2.00 minutes using a Waters Fraction Collector II. Each fraction was analyzed by GC/MS to identify higher diamondoids present in the fraction using GCMS elution order data described above. Further purification of individual higher diamondoids was achieved using Hypercarb HPLC columns (Thermo Electron Corp.) [1]. Analysis of the chromatographic properties of higher diamondoids on Hypercarb HPLC columns will be the subject of a subsequent report. Molecular weights were determined by electron-impact mass-spectrometry (at 70eV), and thus regular diamondoids were discriminated from irregular ones, as well as perimantanes from catamantanes. Dualist graphs of perimantanes included in Table **5** are displayed in (Fig. **16**).

If one plots the GC versus the HPLC data, one obtains a very interesting diagram (Fig. **17**). There are many points corresponding to most of the diamondoids from Table **5**, but only a few of the structures could be identified by means of their molecular weights and chromatographic data: diamondoids with the same number of carbon atoms with a rod-like shape (cf. [22]) lead to higher retention/elution values than than those with a bundle-like shape; regular diamondoids have higher molecular weights than irregular diamondoids with the same number of carbon atoms.

Table 4.   Idealized Euclidean Distance Matrices of Diamondoids

Tetra-	121					
1	**2**	**3**	**4**	**Sum**	**D-Sum**	**Degree**
0	1.73	2.83	4.36	8.92	30.42	1
1.73	0	1.73	2.83	6.29		2
2.83	1.73	0	1.73	6.29		2
4.36	2.83	1.73	0	8.92		1

Tetra-	123					
1	**2**	**3**	**4**	**Sum**	**D-Sum**	**Degree**
0	1.73	2.83	3.32	7.88	28.34	1
1.73	0	1.73	2.83	6.29		2
2.83	1.73	0	1.73	6.29		2
3.32	1.73	2.83	0	7.88		1

Tetra-	1(2)3					
1	**2**	**3**	**4**	**Sum**	**D-Sum**	**Degree**
0	1.73	2.83	2.83	7.39	27.36	1
1.73	0	1.73	1.73	5.19		3
2.83	1.73	0	2.83	7.39		1
2.83	2.83	1.73	0	7.39		1

Penta-	1212						
1	**2**	**3**	**4**	**5**	**Sum**	**D-Sum**	**Degree**
0	1.73	2.83	4.36	5.66	14.58	59.58	1
1.73	0	1.73	2.83	4.36	10.65		2
2.83	1.73	0	1.73	2.83	9.12		2
4.36	2.83	1.73	0	1.73	10.65		2
5.66	4.36	2.83	1.73	0	14.58		1

Penta-	1213						
1	**2**	**3**	**4**	**5**	**Sum**	**D-Sum**	**Degree**
0	1.73	2.83	4.36	4.90	13.82	58.06	1
1.73	0	1.73	2.83	4.36	10.65		2
2.83	1.73	0	1.73	2.83	9.12		2
4.36	2.83	1.73	0	1.73	10.65		2
4.99	4.36	2.83	1.73	0	13.82		1

(Table 4). Contd…..

Penta-	1231						
1	**2**	**3**	**4**	**5**	**Sum**	**D-Sum**	**Degree**
0	1.73	2.83	3.32	2.83	10.71	49.76	1
1.73	0	1.73	2.83	3.32	9.61		2
2.83	1.73	0	1.73	2.83	9.12		2
3.32	2.83	1.73	0	1.73	9.61		2
2.83	3.32	2.83	1.73	0	10.71		1

Penta-	1234						
1	**2**	**3**	**4**	**5**	**Sum**	**D-Sum**	**Degree**
0	1.73	2.83	3.32	4	11.88	52.1	1
1.73	0	1.73	2.83	3.32	9.61		2
2.83	1.73	0	1.73	2.83	9.12		2
3.32	2.83	1.73	0	1.73	9.61		2
4	3.32	2.83	1.73	0	11.88		1

Penta-	12(1)3						
1	**2**	**3**	**4**	**5**	**Sum**	**D-Sum**	**Degree**
0	1.73	2.83	4.36	2.83	11.75	53.92	1
1.73	0	1.73	2.83	1.73	8.02		3
2.83	1.73	0	1.73	2.83	9.12		2
4.36	2.83	1.73	0	4.36	13.28		1
2.83	1.73	2.83	4.36	0	11.75		1

Penta-	12(3)4						
1	**2**	**3**	**4**	**5**	**Sum**	**D-Sum**	**Degree**
0	1.73	2.83	3.32	2.83	10.71	49.76	1
1.73	0	1.73	2.83	1.73	8.02		3
2.83	1.73	0	1.73	2.83	9.12		2
3.32	2.83	1.73	0	3.32	11.2		1
2.83	1.73	2.83	3.32	0	10.71		1

Penta-	1(2,3)4						
1	**2**	**3**	**4**	**5**	**Sum**	**D-Sum**	**Degree**
0	1.73	2.83	2.83	2.83	10.22	48.2	1
1.73	0	1.73	2.83	1.73	7.32		4
2.83	1.73	0	1.73	2.83	10.22		1
2.83	2.83	2.83	0	2.83	10.22		1
2.83	1.73	2.83	2.83	2.83	10.22		1

(Table 4). Contd.....

Hexa	12121								
1	**2**	**3**	**4**	**5**	**6**	**Sum**	**D-Sum**	**Degree**	
0	1.73	2.83	4.36	5.66	7.14	21.72	103.02	2	
1.73	0	1.73	2.83	4.36	5.66	16.31		2	
2.83	1.73	0	1.73	2.83	4.36	13.48		2	
4.36	2.83	1.73	0	1.73	2.83	13.48		2	
5.66	4.36	2.83	1.73	0	1.73	16.31		2	
7.14	5.66	4.36	2.83	1.73	0	21.72		2	

Hepta	121212									
1	**2**	**3**	**4**	**5**	**6**	**7**	**Sum**	**D-Sum**	**Degree**	
0	1.73	2.83	4.36	5.66	7.14	8.49	30.21	163.44	1	
1.73	0	1.73	2.83	4.36	5.66	7.14	23.45		2	
2.83	1.73	0	1.73	2.83	4.36	5.66	19.14		2	
4.36	2.83	1.73	0	1.73	2.83	4.36	17.84		2	
5.66	4.36	2.83	1.73	0	1.73	2.83	19.14		2	
7.14	5.66	4.36	2.83	1.73	0	1.73	23.45		2	
8.49	7.14	5.66	4.36	2.83	1.73	0	30.21		1	

Cyclohex	12312								
1	**2**	**3**	**4**	**5**	**6**	**Sum**	**D-Sum**	**Degree**	
0	1.73	2.83	3.32	2.83	1.73		74.64	2	
1.73	0	1.73	2.83	3.32	2.83			2	
2.83	1.73	0	1.73	2.83	3.32			2	
3.32	2.83	1.73	0	1.73	2.83			2	
2.83	3.32	2.83	1.73	0	1.73			2	
1.73	2.83	3.32	2.83	1.73	0			2	

One may observe that in (Fig. **17**) all three pentamantanes appear at the lower left (GC around 1.1), followed by an isolated spot corresponding to C-6, the peri-hexamantane with a chair-hexagon as dualist (GC around 1.3). The next cluster has 6 points corresponding to the pentamantanes (GC around 1.5); the topmost point (with HPLC around 1.4) is due to the rod-like pentamantane. The unique irregular pentamantane [1231] does not appear in Table **5** and this plot. Catamantanes from the hexamantane group and perimantanes from the heptamantane group form the next cluster between GC 1.75 and 2.0; the topmost point with HPLC 2.04 is the rod-like hexamantane; also two perifused heptamantanes with dualists isomorphic to methylcyclohexane were identified in this cluster: the equatorial [121321]heptamantane with GC around 1.85, and the axial [123124]heptamantane with GC around 1.92. The next cluster with GC between

ter with GC between 2.1 and 2.35 contains diamondoids from the octamantane and nonamantane groups, mostly perimantanes, including the nonamantane having a dualist graph isomorphic with bicyclo[3.3.1]nonane. The last two points with GC around 2.5 and 2.6, correspond, respectively, to the decamantane shown in (Fig. **14**) and to its undecamantane homolog having one extra vertex attached to a "tertiary-carbon" of the adamantane-like dualist.

Older literature data for diamondoid chromatographic retention data and for their normal boiling points, NBPs (estimated from the retention times, RT) are available in the published literature [23], as seen in Table **6** (M and E are abbreviations for methyl and ethyl, respectively; "ad" stands for adamantane, "di" for diamantane, and "tr" for triamantane). The molecular weight (MW) is given in Daltons.

Table 5.   Diamondoids with their Dualist-Graph Codes, Molecular Formulas, Molecular Weights, Relative GC Retention Times and Relative HPLC Elution Volumes

Higher Diamondoid Code	Compound Reference Number	Molecular Formula	M⁺ (m/z) Base Peak	GC/MS Relative Retention Times*	ODS HPLC Relative Elution Volumes*
[1(2)3]	4-1	$C_{22}H_{28}$	292	1.000	1.000
[121]	4-2	$C_{22}H_{28}$	292	1.069	1.055
[123]	4-3	$C_{22}H_{28}$	292	1.126	1.036
[1(2,3)4]	5-1	$C_{26}H_{32}$	344	1.284	1.126
[12(1)3]	5-2	$C_{26}H_{32}$	344	1.473	1.269
[1212]	5-3	$C_{26}H_{32}$	344	1.479	1.431
[1213]	5-4	$C_{26}H_{32}$	344	1.529	1.323
[12(3)4]	5-5	$C_{26}H_{32}$	344	1.543	1.269
[1234]	5-6	$C_{26}H_{32}$	344	1.570	1.287
[12312]	C-6	$C_{26}H_{30}$	342	1.523	1.341
	6-1	$C_{30}H_{36}$	396	1.785	1.520
	6-2	$C_{30}H_{36}$	396	1.804	1.449
	6-3	$C_{30}H_{36}$	396	1.848	1.704
	6-4	$C_{30}H_{36}$	396	1.850	1.520
	6-5	$C_{30}H_{36}$	396	1.857	1.556
[12121]	6-6	$C_{30}H_{36}$	396	1.868	2.039
	6-7	$C_{30}H_{36}$	396	1.879	1.591
	6-8	$C_{30}H_{36}$	396	1.891	1.627
	6-9	$C_{30}H_{36}$	396	1.904	1.627
	6-10	$C_{30}H_{36}$	396	1.907	1.802
	6-11	$C_{30}H_{36}$	396	1.912	1.573
	6-12	$C_{30}H_{36}$	396	1.918	1.724
	6-13	$C_{30}H_{36}$	396	1.926	1.573
	6-14	$C_{30}H_{36}$	396	1.952	1.627
	6-15	$C_{30}H_{36}$	396	1.962	1.743
	6-16	$C_{30}H_{36}$	396	1.981	1.724
	6-17	$C_{30}H_{36}$	396	1.985	1.822
[121321]	7-1	$C_{30}H_{34}$	394	1.847	1.743
[123124]	7-2	$C_{30}H_{34}$	394	1.917	1.664
	7-3	$C_{34}H_{40}$	448	2.140	2.003
	7-4A	$C_{34}H_{40}$	448	2.185	2.110
[121212]	7-4B	$C_{34}H_{40}$	448	2.185	2.521
	7-5	$C_{34}H_{40}$	448	2.186	2.271
	7-6	$C_{34}H_{40}$	448	2.196	2.110
	7-8	$C_{34}H_{40}$	448	2.221	1.973

(Table 5). Contd…..

	7-9A	$C_{34}H_{40}$	448	2.238	1.988
	7-9B	$C_{34}H_{40}$	448	2.238	2.020
	7-9C	$C_{34}H_{40}$	448	2.238	2.307
	7-10	$C_{34}H_{40}$	448	2.240	2.450
	7-13A	$C_{34}H_{40}$	448	2.258	1.959
	7-13B	$C_{34}H_{40}$	448	2.258	2.235
	7-13C	$C_{34}H_{40}$	448	2.258	2.557
	7-14	$C_{34}H_{40}$	448	2.262	2.164
	8-1	$C_{34}H_{38}$	446	2.136	2.360
	8-2	$C_{34}H_{38}$	446	2.144	2.396
	8-3	$C_{34}H_{38}$	446	2.150	2.056
	8-5	$C_{34}H_{38}$	446	2.187	2.038
	8-6	$C_{34}H_{38}$	446	2.200	2.325
	8-7	$C_{34}H_{38}$	446	2.204	2.182
	8-8	$C_{34}H_{38}$	446	2.250	2.414
	8-9	$C_{34}H_{38}$	446	2.279	2.235
	8-10	$C_{34}H_{38}$	446	2.303	2.360
	8-11	$C_{34}H_{38}$	446	2.316	2.432
[12312412]	9-1	$C_{34}H40$	444	2.292	2.396
[1231241(2)3]	10-1	$C_{35}H_{36}$	456	2.451	2.503
[123(1,2)42143]	11-1	$C_{39}H_{40}$	508	2.599	2.987

However, these older data are for other chromatographic systems, and moreover they refer only to smaller diamondoids with one to three adamantane units, having 0, 1, 2, 3, or 4 methyl groups, or one ethyl group, as well as a combination of ethyl and methyl groups.

QUANTITATIVE STRUCTURE-PROPERTY RELATIONSHIP (QSPR) CORRELATIONS WITH GC AND HPLC DATA

With the help of Euclidean distance matrices **D**, one can obtain various molecular descriptors for QSPR correlations. The simplest ones are modeled after topological indices (such as the oldest one, namely the Wiener index, *W*) [24-

26]. Here we will use D-Sums, which are twice as large as *W*. Also the eigenvalue of **D**, and $(D-Sum)^2/N^2$ which is proportional to the moment of inertia of the molecule. It should be stressed that these descriptors take into account geometrical factors such as the distinction between staggered alkane rotamers, i.e., dualist graphs of diamondoids. A different procedure involves distance-distance matrices (**DD**-matrices) [19-21], whose entries are quotients between geometrical (Euclidean) distances and topological distances. The geometrical distances can be computed either from MM2 calculations with a C – C bond distance of 1.54 Å (154 pm), or from integer coordinates with a C – C bond distance of 1 conventional unit. In Table **7** one may see 17 diamondoids with
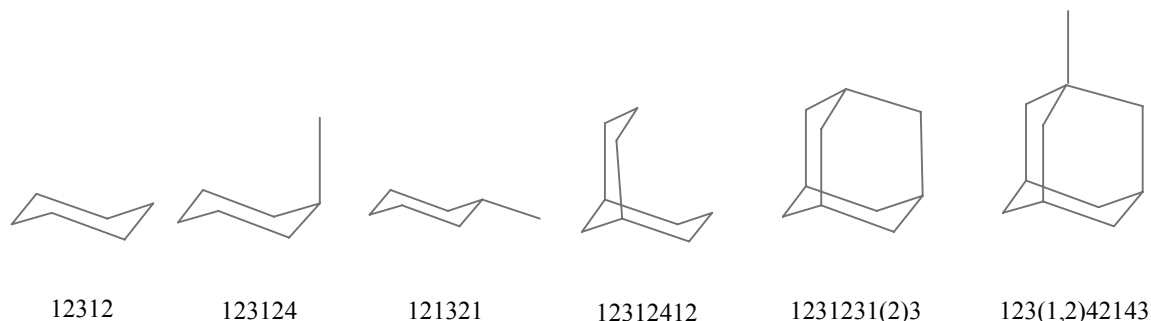

| 12312 | 123124 | 121321 | 12312412 | 1231231(2)3 | 123(1,2)42143 |

Fig. (16). Dualist graphs of bundle-shaped higher perimantanes with their 4-digit codes.
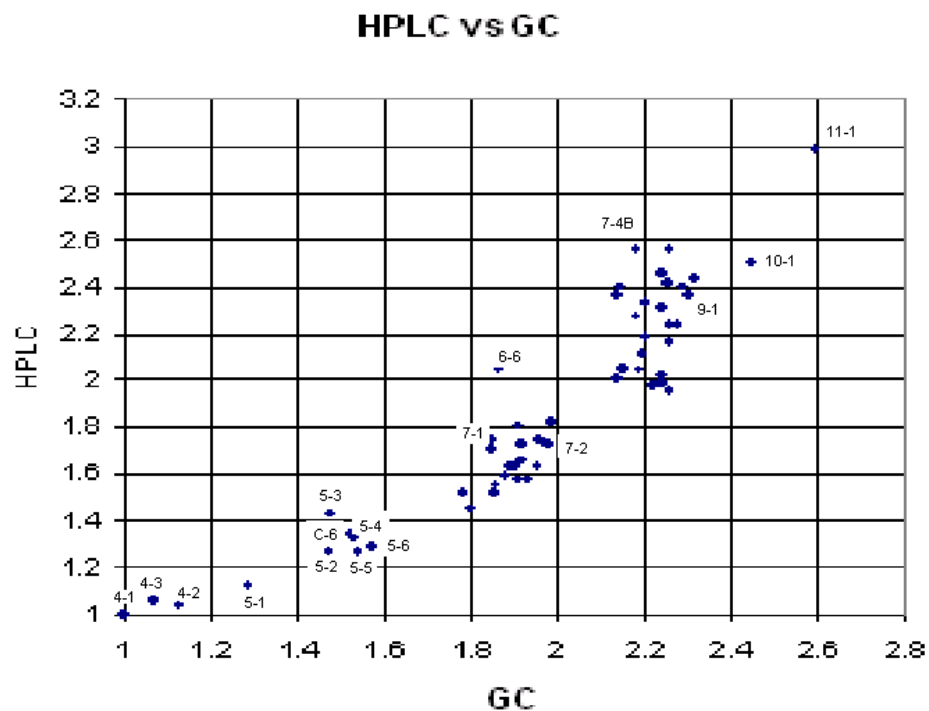
Fig. (17). Plot of the HPLC eluent volumes versus the GC retention times of diamondoids.

Table 6.   Retention Times (RTs) and NBPs Estimated [23] on the Basis of RTs

No.	Compound	MW (Da)	RT (min)	NBP,°C
1	Adamantane (ad)	136	19.81	190
2	1-M-ad	150	20.57	194
3	1,3-MM-ad	164	21.20	197
4	1,3,5-MMM-ad	178	21.66	199
5	1,3,5,7-MMMM-ad	192	22.03	201
6	2-M-ad	150	23.23	207
7	*cis*-1,4-MM-ad	164	23.66	210
8	*trans*-1,4-MM-ad	164	23.90	211
9	1,3,6-MMM-ad	178	24.18	212
10	1,2-MM-ad	164	24.92	216
11	*cis*-1,3,4-MMM-ad	178	25.23	217
12	*trans*-1,3,4-MMM-ad	178	25.42	218
13	1,2,5,7-MMMM-ad	192	25.58	219
14	1-E-ad	164	25.91	221
15	1-E-3-M-ad	178	26.41	223
16	1-E-3,5-MM-ad	192	26.77	225
17	2-E-ad	164	27.12	227

18	Diamantane (di)	188	36.02	272
19	4-M-di	202	36.47	274
20	4,9-MM-di	216	36.86	276
21	1-M-di	202	37.73	280
22	1,4-MM-di and 2,4-MM-di	216	37.82	281
23	4,8-MM-di	216	37.99	282
24	MMM-di	230	38.06	282
25	3-M-di	202	38.56	285
26	3,4-MM-di	216	38.91	286
27	Triamantane (tr)	240	48.00	333
28	9-M-tr	254	48.13	334
29	MM-tr	268	48.22	335

Table 7.   Diamondoids and their Data Based on MM2-Optimized Geometry

Diamondoid	N	Formula	(m/z)	D-Sum	$(DSum)^2/N^2$	D-Eigenv.	GC/MS	HPLC	Type
[1(2)3]	4	$C_{22}H_{28}$	292	24.48	37.45	5.9	1	1	4-cata
[121]	4	$C_{22}H_{28}$	292	27.24	46.38	6.98	1.069	1.055	4-rod
[123]	4	$C_{22}H_{28}$	292	25.68	41.22	6.5	1.126	1.036	4-cata
[1(2,3)4]	5	$C_{26}H_{32}$	344	42.8	73.27	8.71	1.284	1.126	5-bundle
[12(1)3]	5	$C_{26}H_{32}$	344	46.7	87.24	9.54	1.473	1.269	5-cata
[1212]	5	$C_{26}H_{32}$	344	53.4	114.06	10.99	1.479	1.431	5-rod
[1213]	5	$C_{26}H_{32}$	344	50.84	103.39	10.41	1.529	1.323	5-cata
[12(3)4]	5	$C_{26}H_{32}$	344	45.2	81.72	9.172	1.543	1.269	5-cata
[1234]	5	$C_{26}H_{32}$	344	47.96	92.01	9.732	1.57	1.287	5-cata
[12121]	6	$C_{30}H_{36}$	396	92.4	237.16	11.1	1.868	2.039	6-rod
[12312]	6	$C_{26}H_{30}$	342	67.8	127.69	15.94	1.523	1.341	6-bundle
[121321]	7	$C_{30}H_{34}$	394	104.66	223.55	15.15	1.847	1.743	7-bundle
[123124]	7	$C_{30}H_{34}$	394	100.56	206.37	14.45	1.917	1.664	7-bundle
[121212]	7	$C_{34}H_{40}$	448	146.66	438.96	21.72	2.185	2.521	7-rod
[12312412]	9	$C_{34}H_{36}$	444	180.1	400.44	20.07	2.292	2.396	9-bundle
[1231241(2)3]	10	$C_{35}H_{36}$	456	219.42	481.45	21.15	2.451	2.503	10-bundle
[123(1,2)42143]	11	$C_{39}H_{40}$	508	278.35	640.32	26.57	2.599	2.987	11-bundle

Table 8.   Diamondoids with their Data and Molecular Descriptors Based on Integer Coordinates

Diamondoid	N	D-Sum	$(DSum)^2/N^2$	D-Eigenv.	DD-Sum	DD-Eigenv.	GC/MS	HPLC	Type
[1(2)3]	4	27.36	46.79	6.87	27.36	4.714	1	1	4-cata
[121]	4	30.42	57.84	7.79	30.42	5.103	1.069	1.055	4-rod
[123]	4	28.34	50.2	7.15	28.34	4.575	1.126	1.036	4-cata
[1(2,3)4]	5	48.2	92.93	10.58	47.8	6.17	1.284	1.126	5-bundle
[12(1)3]	5	53.92	116.29	11.08	53.92	6.191	1.473	1.269	5-cata
[1212]	5	59.58	141.99	12.27	59.58	6.187	1.479	1.431	5-rod
[1213]	5	58.24	135.68	11.88	58.24	6.127	1.529	1.323	5-cata
[12(3)4]	5	49.76	99.04	10.04	49.76	5.939	1.543	1.269	5-cata
[1234]	5	52.1	108.58	10.53	52.1	5.773	1.57	1.287	5-cata
[12121]	6	103.02	294.81	17.74	45.78	7.632	1.868	2.039	6-rod
[12312]	6	74.64	154.75	12.44	44.34	7.39	1.523	1.341	6-bundle
[121321]	7	116.66	277.75	16.89	116.64	8.814	1.847	1.743	7-bundle
[123124]	7	110.7	250.09	15.88	110.7	8.57	1.917	1.664	7-bundle
[121212]	7	163.44	545.16	24.17	63.44	9.068	2.185	2.521	7-rod
[12312412]	9	196.24	475.43	21.84	196.24	21.834	2.292	2.396	9-bundle
[1231241(2)3]	10	247.08	610.49	24.74	247.08	24.738	2.451	2.503	10-bundle
[123(1,2)42143]	11	323.02	862.33	29.65	323.02	29.655	2.599	2.987	11-bundle
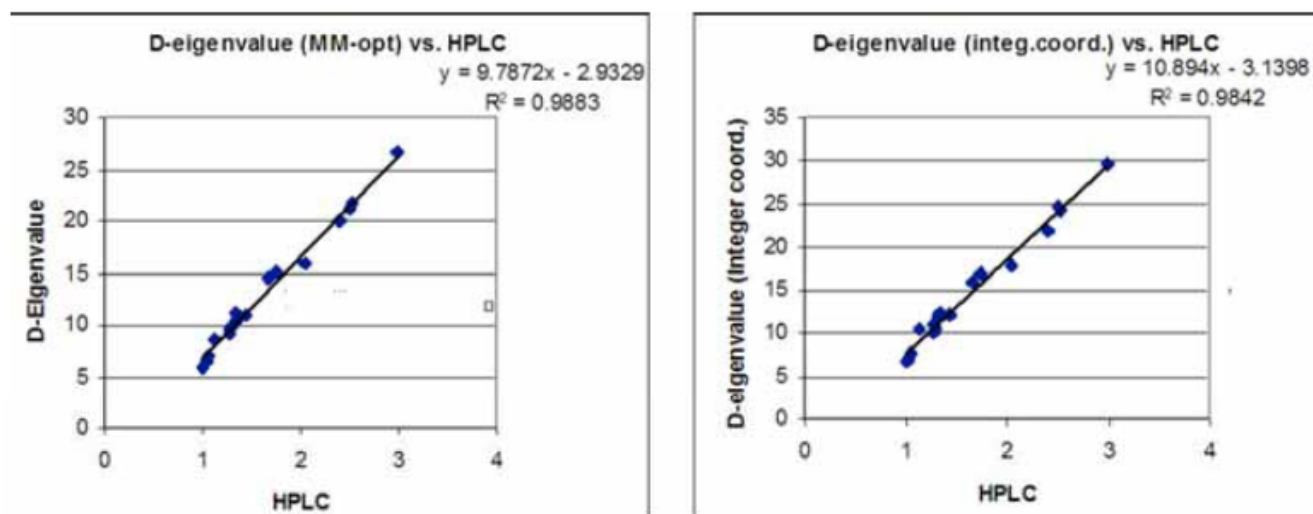
Fig. (18). Plot of eigenvalues of distance matrices for MM2-optimized geometries or idealized integer-coordinate geometries versus HPLC elution volumes for the 17 diamondoids listed in Tables **7** and **8**.

certified structures and their GC or HPLC data along with molecular descriptors based on the MM2-optimized geometry, and in Table **8** the same diamondoids with descriptors based on idealized integer coordinates.

It should be stressed that the dualist graph contains all the information about the diamondoid structure, and that the descriptors take all this information into account, ignoring however chirality.

Among the many possible QSAR correlations, we present in (Fig. **18**) only two global correlations for all 17 dia-

mondoids from the above two tables, namely linear correlations between the D-eigenvalue and the GC or HPLC data: $R^2 = 0.9883$ for GC retention time and 0.9842 for HPLC elution volume. Because of the small number of points, validation procedures have not been introduced. Also, separate correlations for the elongated (rod-like catamantanes) or spherical (bundle-like) diamondoids are not displayed. The DD-matrix derived descriptors do not improve the correlations over those obtained from the Euclidean D-matrices. The GC data vary less linearly than the HPLC data with most molecular descriptors.

CONCLUSIONS

For diamondoid hydrocarbons occurring naturally in petroleum, GC-MS and HPLC data are reported for the first time. With the help of dualists and their topological/geometrical mathematical transforms we developed molecular descriptors which can be used for QSPR. Linear correlations between chromatographic data and such descriptors have been found for 17 diamondoids with known structures.

ACKNOWLEDGEMENT

REFERENCES

[1]   Dahl, J. E.; Liu, S. G.; Carlson, R. M. K. *Science*, **2003**, *299*, 96.
[2]   Dahl, J. E.; Moldowan, J. M.; Peakman, T. M.; Clardy, J. C.; Lobkovsky, E.; Olmstead, M. M.; My, P. W.; Davis, T. J.; Steeds, J. W.; Peters, K. E.; Pepper, A.; Ekuan, A.; Carlson, R. M. K. *Angew. Chem. Int. Ed.,* **2003**, *42*, 2040.
[3]   Dahl, J. E.; Moldowan, J. M.; Peters, K. E.; Claypool, G. E.; Rooney, M. A.; Michael, G. E.; Mello, M. R.; Kohnen, M. L. *Nature*, **1999**, *399*, 54.
[4]   Schleyer, P. v. R. *J. Am. Chem. Soc.,* **1957**, *79*, 3292.
[5]   Schleyer, P. v. R. In: *Cage Hydrocarbons* (G. A. Olah, Ed.), Wiley-Interscience, New York, **1990**, pp. 1-38.
[6]   Whitlock, Jr., H. W.; Siefken, M. W. *J. Am. Chem. Soc.,* **1968**, *90*, 4929.
[7]   Gund, T. M.; Schleyer, P. v. R.; Gund, P. H.; Wipke, W. T. *J. Am. Chem. Soc.,* **1975**, *97*, 743.
[8]   Balaban, A. T.; Farcasiu, D.; Banica, R. *Rev. Roum. Chim.*, **1966**, *11*, 1205.
[9]   Balaban, A. T. In: *Graph Theoretical Approaches to Chemical Reactivity*, (Bonchev, D.; Mekenyan, O. Eds.), Kluwer Academic Publishers, Dordrecht, Netherlands, **1994**, pp.137-180.
[10]   Balaban, A. T.; Schleyer, P. von R. *Tetrahedron,* **1978**, *34*, 3599.
[11]   Balaban, A. T.; Harary, *Tetrahedron,* **1968**, *24*, 2505.
[12]   Balaban, A. T. *Tetrahedron,* **1969**, *25*, 2949.
[13]   Balaban, A. T. *Pure Appl. Chem.*, **1982**, *54*, 1075.
[14]   Balaban, A. T. *MATCH*, *Commun. Math. Comput. Chem.,* **1976**, *2*, 51.
[15]   Balaban, A. T. *Rev. Roum. Chim.,* **1976**, *21*, 1049.
[16]   March, N. H.; Balaban, A. T.; Leys, F. E.; Klein, D. J.; Seitz, W. A. *Phys. Chem. Liquids,* **2003**, *41*, 303.
[17]   Ivanciuc, O.; Ivanciuc, T.; Klein, D. J.; Seitz, W. A.; Balaban, A. T. *J. Chem. Inf. Comput. Sci.,* **2001**, *41*, 536.
[18]   Balaban, A. T. In: *From Chemical Topology to Three-Dimensional Geometry*, (Balaban, A. T. Ed.), Plenum Publishing Corporation, New York, **1997**, pp. 1-24.
[19]   Klein, D. J. *Intl. J. Quantum Chem. S*, **1986**, *20,* 153.
[20]   Randić, M.; Kleiner, A. F.; De Alba, L. M. *J. Chem. Inf. Comput. Sci.,* **1994**, *34*, 277.
[21]   Klein, D. J. *J. Math. Chem.,* **1995,** *18,* 321.
[22]   Schwab, P. F. H.; Levin, M. D.; Michl, J. *Chem. Rev.,* **1999**, *99*, 1863.
[23]   Wingert, W. S. *Fuel,* **1992**, *71*, 37.
[24]   Devillers, J.; Balaban, A. T. (Eds.), *Topological Indices and Related Descriptors in QSAR and QSPR*, Gordon and Breach, The Netherlands, **1999**.
[25]   Karelson, M. *Molecular Descriptors in QSAR/QSPR*, Wiley-Interscience, New York, **2000**.
[26]   Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*, Wiley-VCH, New York, **2000**.