Molecular bonding profiles [☆]

Milan Randić

Department of Mathematics and Computer Science, Drake University, Des Moines, Iowa 50311, USA

Received 9 February 1995; revised 21 December 1995

We present a refinement of recently proposed characterization of molecules based on a sequence of powers of interatomic separations referred to as molecular profiles. The molecular profiles and closely related shape profiles were based on the averaging contributions arising from different powers of interatomic distances for atoms in a molecule or atoms at the molecular periphery, respectively. Consequently, molecular models in which atoms have the same set of coordinates but different bonding patterns will result in identical molecular profiles. In this article we outline a refinement of molecular profiles in which the bonding pattern in a molecule is fully acknowledged. This is accomplished by adding "ghost" sites along chemical bonds. The distance-based invariants of the augmented matrix reflect the bonding pattern of a structure giving different molecular profiles for molecules having the same atomic coordinates but different bondings. The procedure is general and applies to two-dimensional and three-dimensional molecular skeletons. Equally, the approach can be applied to van der Waals-type molecular surfaces and molecular contours of equal electron densities in order to obtain characterization of more realistic molecular models.

1. Introduction

When we consider molecular models, immediately two basic problems arise: (i) How to represent a structure, and (ii) how to characterize a structure. A *representation* calls for a development of a suitable molecular code and atomic labels, or input of the structure (for computer use). By knowing the code one can not only retrieve the information on a structure (from a structure-library) but also fully reconstruct the structure. A trivial representation is the one given by the list of atomic coordinates of a structure, but even that needs to be standardized, and made unique, which is far from trivial. A good representation, besides being unique, satisfies additional requirements, as discussed by Read [1] and summarized in Table 1. In contrast, a *characterization* of a structure is based on structural invariants. It is generally believed that characterizations are inherently incomplete. Hence, two structures may have the same characterization, just as two molecules may have same

[☆] Dedicated to a leading graph theorist, Professor R.C. Read, for sharing his mathematical talents on chemical structures.

Table 1
List of desirable requirements for chemical codes and for molecular descriptors, respectively.

Codes

1 Codes should be linear
2 Codes should be unique
3 Reconstruction should be possible
4 Codes should be simple (if possible hand-made)
5 Decoding should be possible (possibly by hand)
6 Trivial names should be avoided
7 Properties should not be used in coding
8 Codes should be brief
9 Codes should be pronounceable
10 Codes should be easily understood
11 Only familiar symbols should be used
12 Coding and decoding should be efficient
13 [a] Similar structures should have similar codes

Descriptors

1 Should have structural interpretation
2 Should have good correlation with at least one property
3 Should preferably discriminate among isomers
4 Should be possible to apply to local structure
5 Should be possible to generalize to "higher" descriptors
6 [b] Descriptors should preferably be independent
7 Should be simple
8 Should not be based on properties
9 Should not be trivially related to other descriptors
10 Should be possible to construct efficiently
11 Should use familiar structural concepts
12 Should have the correct size dependence
13 Should change gradually with gradual change in structures

[a] Suggested by M. Randić [4].
[b] With the development of orthogonalization procedure this step can always be accomplished regardless how strongly two descriptors correlate, unless the correlation coefficient is exactly 1. For orthogonalization procedure see [39–42].

properties. A good characterization offers some insight into the modeling of the structure-property relationships. There are no restrictions on the design of structural invariants, the limiting factor is one's own imagination. Not surprisingly, this resulted in a proliferation of molecular descriptors, particularly topological indices [2], the most common structural invariants. However, several recent studies have shown that at most dozen molecular descriptors have found use in multivariate regression analysis [3–6]. In order to curb the proliferation of molecular descriptors, a set of requirements, similar to those proposed by Read for molecular codes, have been suggested [7]. As can be seen from Table 1 the requirements for the most part parallel the requirements proposed by Read for molecular codes.

The challenge of chemical graph theory is, on the one hand, to design descriptors that have a simple structural interpretation and parallel well some physicochemical property of molecules [8–12] and, on the other hand, to extend graph theoretical schemes to molecules viewed as three dimensional structures [13–19].

2. Molecular profiles and shape profiles

Very recently a novel approach to representation and characterization of 3-D molecules was outlined [20–26]. Briefly, a molecule is characterized by a sequence (D^1, D^2, D^3, D^4, ..., D^k, ...), where D^k is a suitably normalized average of the interatomic distances raised to the power k. The factor $1/k!$ is used as the normalization of the kth power of the averaged distances to ensure the convergence for D^k sequences. This is somewhat analogous to the presence of $1/k!$ in Taylor expansion of a function in calculus.

In our models of hydrocarbons the hydrogen atoms will be suppressed (as is customary in simplified models of benzenoids). In Fig. 1 (left) we illustrate carbon atom skeletons of pyrene, perylene and anthrathene. In Fig. 1 (right) we show the carbon atoms at the peripheries of the same smaller benzenoid hydrocarbons in order to emphasize the difference between the bonding pattern associated with a molecule as a whole and the bonding pattern of its periphery. To obtain the molecular profiles (or volume profiles) in the construction of the D^k matrices we use contributions arising from all the pairs of carbon atoms in the molecule. In contrast, in order to obtain the shape profiles (or periphery profiles) we use only the interatomic

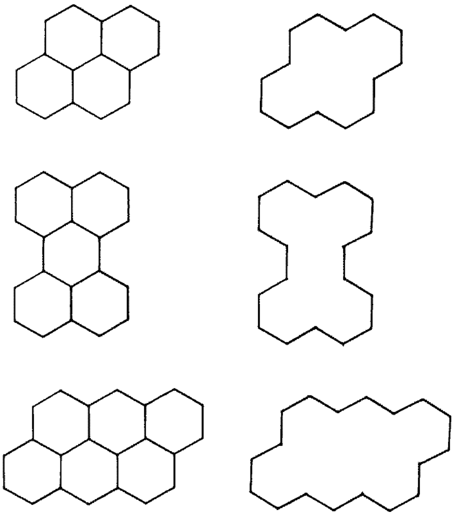

Fig. 1.

contributions arising from the carbon atoms at the molecular boundary. In the left part of Table 2 we show numerically the molecular profiles for the selection of smaller benzenoid hydrocarbons considered. In the right part of Table 2 we show the corresponding periphery profiles of the same molecule. The difference between the two types of profiles is not large, though it is significant. For smaller powers the molecular profiles are somewhat larger but as we consider large powers of interatomic distances the periphery profiles dominate. This is not surprising, since for higher powers of interatomic distance the atoms at the largest separations make dominant contributions, and these will be the atoms at the molecular periphery, not the interior atoms.

Strictly speaking, in both cases in the construction of the profiles only the locations of carbon atoms have been used, and not the connectivity, i.e., the bonding pattern itself. Molecular profiles and shape profiles illustrate a mapping of two-dimensional and three-dimensional objects to one-dimensional mathematical object (sequence). This mapping represents a novel breakthrough in characterization of three-dimensional molecules by structural invariants.

The shape profiles and the molecular profiles have been found useful in the discussion of molecular similarity and the discussion of selected molecular properties, e.g., chromatographic retention indices and the boiling points of benzenoid systems [25,26].

Table 2
Molecular profiles and periphery profiles for selected smaller benzenoids. (PYRE = pyrene, PERY = perylene and ANTH = anthanthrene.)

Molecule			Periphery		
PYRE	PERY	ANTH	PYRE	PERY	ANTH
36.48	52.48	61.18	33.93	49.24	54.29
52.00	86.00	107.50	51.50	84.00	103.5
55.41	105.70	143.83	57.53	106.51	147.40
48.08	105.62	158.69	51.58	109.00	170.26
35.48	89.39	150.24	38.90	93.97	166.71
22.87	65.75	125.06	25.44	70.10	142.27
13.13	42.79	93.05	14.74	46.12	107.83
6.80	24.97	62.64	7.68	27.13	73.59
3.21	13.20	38.50	3.64	14.43	45.71
1.39	6.38	21.77	1.59	7.01	26.05
0.56	2.84	11.40	0.64	3.13	13.71
0.21	1.17	5.55	0.24	1.29	6.71
0.07	0.45	2.53	0.08	0.50	3.06
0.02	0.16	1.08	0.03	0.18	1.31
0.01	0.05	0.43	0.01	0.06	0.53

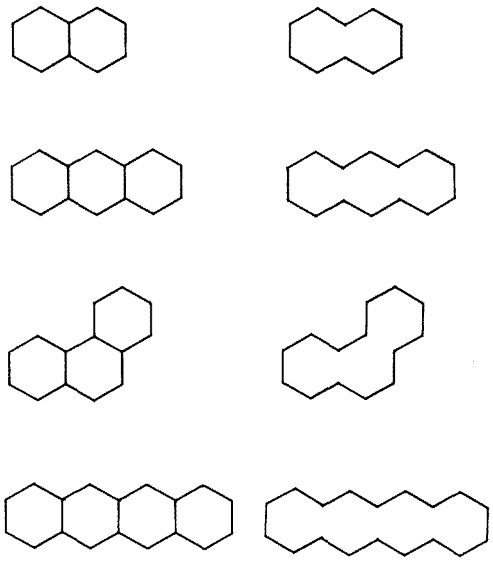

Fig. 2.

3. Limitations

In Fig. 2 we illustrate a few smaller benzenoids in which all atoms are on the molecular periphery. As a consequence, in these molecules the sequence of averaged powers of interatomic distances is the same for the shape profiles (or the surface profile) and the bulk profile (or the volume profile). Moreover, since only atomic coordinates enter the construction of the distance matrix from which the profiles are computed, we cannot even differentiate between polycyclic systems and the corresponding monocyclic structures devoid of the inner CC bonds. Such are naphthalene, anthracene and phenanthrene and shown in the left part of Fig. 2 and the monocyclic structures of the same periphery shown in the right part of Fig. 2. A similar situation occurs again for all *cis* conformation of hexatriene and the corresponding ring structure. In both cases all the carbon atoms considered have the same (idealized) coordinates.

One way to discriminate between *cis* hexatriene and the benzene ring is to combine the information from the geometric distances (that depends only on atomic coordinates) with the information from the topological distance matrix (which reflects molecular connectivity). By taking the ratios of the geometrical distance and the graph theoretical distance one can construct matrix, referred to as the D/D matrix [27]. The D/D matrix differentiates the bent path from the closed ring structure and monocyclic from polycyclic structures, even if all carbon atoms have the same coordinates.

We should mention that molecular profiles can be constructed also for molecules by including hydrogen atoms. Nevertheless, we will continue to consider hydrogen-suppressed molecular skeletons. The characterization of molecules by using only atomic coordinates represents a limitation. Such characterization cannot, for example, reflect the presence of bent bonds in highly strained small ring compounds. Similarly, such characterization cannot describe adequately important details of molecular van der Waals surfaces, or that of equipotential surfaces around a molecule that can be of an arbitrary shape. In order to characterize molecular contours and molecular surfaces we have to increase the resolution of the approach. This can be done by considering distances between additional points on a molecular surface or molecular interior. Here we will outline the procedure to do just that. Hence, we need to go beyond atoms and extend our considerations to chemical bonds. In this paper we will outline a route to the characterization of molecules using molecular profiles that will discriminate between monocyclic and polycyclic molecular systems, such as decalin and naphthalene. It will be clear from the outline of the so-derived molecular bonding profiles that the approach is general and can be equally extended to molecular contours and even molecular surfaces, such as the already mentioned van der Waals molecular surface.

4. Bonding profiles

In order to differentiate between all *cis* hexatriene and a closed ring, the two structures in which carbon atoms have the same idealized coordinates, we have to go beyond simple molecular geometry restricted to atomic coordinates only. Instead of using the information on topological distance to differentiate between the two systems we will consider a direct approach based on molecular bonding pattern. We start again with a molecular diagram (embedded in 3-dimensional space) for which we know the positions of all atoms. We now select, besides the n sites of atoms that define the overall molecular geometry, additional sites along different CC bonds. In the case of bent bonds one could follow the line of the maximal electron density. A molecule, instead of being represented by $n(n-1)/2$ interatomic distances, where n is the number of atoms in a molecule, is represented by $N(N-1)/2$ distances, where one is at liberty to choose N, the size of the distance matrix.

If each bond is uniformly represented by m points, N is approximately given by the product nm. For acyclic structures, $N = (n-1)(m+1)+1$; for monocyclic structures (in which the number of atoms equals the number of bonds), $N = n(m+1)$; for bicyclic systems, $N = (n+1)(m+1)-1$; and so on. We may refer to m as an index of the resolution or the index of magnification of the approach. Thus $m = 0$ means no magnification, when $m = 1$ the bonds have been recognized, $m = 2$ indicates a better resolution of bonds or the higher magnification, etc. In Fig. 3 we illustrate the situation for hexatriene. In this way we can differentiate *cis* hexatriene and the benzene ring, or naphthalene and its monocyclic

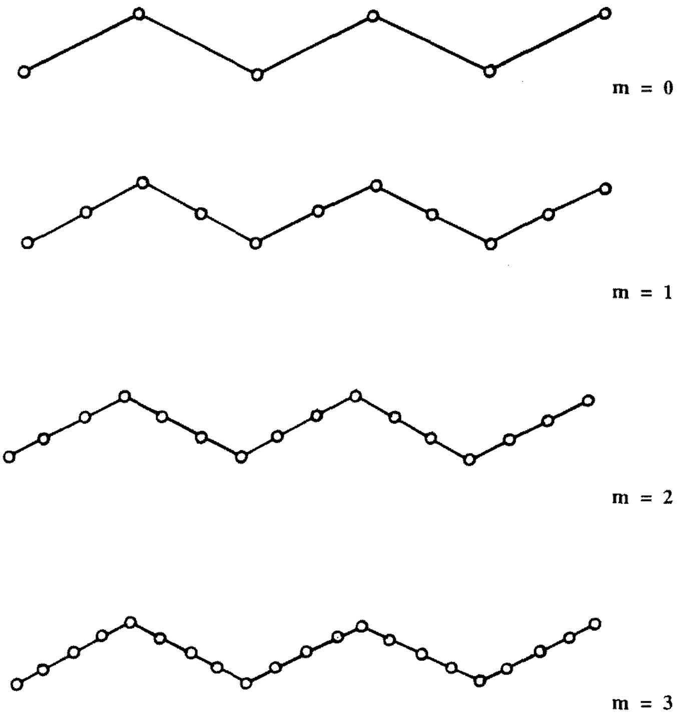

Fig. 3.

counterpart, or cyclohexane and decalin, since in each case the structures that have hitherto identical geometries but distinct bonding patterns will now manifest the presence of the bonds explicitly.

In Table 3 we illustrate the distance matrix for *all trans* hexatriene (top part) and the corresponding distance matrix for the same structure in which the "ghost atoms" were located half way across each CC bond. The original 6 × 6 matrix is now replaced by an 11 × 11 distance matrix. We see that the new 11 × 11 matrix is in fact the augmented 6 × 6 matrix in which all the entries already present in the 6 × 6 matrix are retained. The new matrix has additional entries, absent in the smaller matrix, which give distances between the bond mid-points and other bond mid-points or other atoms. This is apparent when we retain the labels used for 6 × 6 matrix and use labels 7–11 for the "ghost atoms".

The molecular profiles and the shape profiles are derived from such 11 × 11 distance matrices by considering different powers of the matrix elements and constructing the corresponding row sums. In Table 4 we list for the initial powers of k the row sums for the two matrices of Table 3. The molecular profile is obtained by averaging the row sums (RS) in a matrix D^k. As the 6 × 6 matrix was augmented to a 11 × 11 matrix, the row sums have increased – more than doubled – since there are many new matrix elements that have to be considered. In order to keep the cor-

Table 3
The 6 × 6 and 11 × 11 distance matrices for *all trans* hexatriene.

0	1	1.7321	2.6456	3.4641	4.3589
1	0	1	1.7321	2.6456	3.4641
1.7321	1	0	1	1.7321	2.6456
2.6456	1.7321	1	0	1	1.7321
3.4641	2.6456	1.7321	1	0	1
4.3589	3.4641	2.6456	1.7321	1	0

0	0.5000	1	1.3229	1.7321	2.1794	2.6458	3.0414	3.4641	3.9051	4.3589
0.5000	0	0.5000	1	1.3229	1.7321	2.1794	2.6458	3.0414	3.4641	3.9051
1	0.5000	0	0.5000	1	1.3229	1.7321	2.1794	2.6458	3.0414	3.4641
1.3229	1	0.5000	0	0.5000	1	1.3229	1.7321	2.1794	2.6458	3.0414
1.7321	1.3229	1	0.5000	0	0.5000	1	1.3229	1.7321	2.1794	2.6458
2.1794	1.7321	1.3229	1	0.5000	0	0.5000	1	1.3229	1.7321	2.1794
2.6458	2.1794	1.7321	1.3229	1	0.5000	0	0.5000	1	1.3229	1.7321
3.0414	2.6458	2.1794	1.7321	1.3229	1	0.5000	0	0.5000	1	1.3229
3.4641	3.0414	2.6458	2.1794	1.7321	1.3229	1	0.5000	0	0.5000	1
3.9051	3.4641	3.0414	2.6458	2.1794	1.7321	1.3229	1	0.5000	0	0.5000
4.3589	3.9051	3.4641	3.0414	2.6458	2.1794	1.7321	1.3229	1	0.5000	0

responding row characteristics as much constant as possible we need an adequate normalization. For that reason we will consider RS/n (Table 5). Alternatively one could use $RS/(n-1)$. In this way we obtain the part of the row sum that corre-

Table 4
The row sums for the two matrices of Table 3. Only the symmetry nonequivalent rows are shown.

	D^1	D^2	D^3	D^4	D^5
Row 1	13.20080	42.00000	149.10471	564.00000	2218.62343
Row 2	9.84190	24.00000	67.28563	204.00000	646.06091
Row 3	8.10985	15.00000	30.91256	69.00000	162.81873
Average	10.38419	27.00000	82.43430	279.00000	1009.16769
	D^1	D^2	D^3	D^4	D^5
Row 1	24.14963	73.25000	249.58306	907.81250	3440.2936
Row 2	20.10909	54.00000	165.55525	543.00000	1854.9825
Row 3	17.38561	40.25000	108.33583	315.31250	959.57673
Row 4	14.92876	29.25000	67.39743	170.06250	452.50471
Row 5	13.93505	23.75000	46.14501	97.81250	220.15767
Row 6	13.20080	21.00000	37.27618	70.50000	138.66396
Average	17.65610	42.00000	119.20994	376.22727	1272.15404

Table 5
The averaged row sums and the normalized averaged row sums for the two matrices of Table 3.

D^n	6×6	11×11	$6 \times 6/n!$	$11 \times 11/n!$
D^1	10.3842	17.6561	10.3842	17.6561
D^2	27.0000	42.0000	13.5000	21.0000
D^3	82.4343	119.2099	13.7391	19.8683
D^4	279.0000	376.2273	11.6250	15.6761
D^5	1009.1677	1272.1540	8.4097	10.6013
D^6	3819.0000	4514.1478	5.3042	6.2696
D^7	14928.0543	16599.3278	2.9619	3.2935
D^8	59775.000	62734.731	1.4825	1.5559
D^9	243780.94	242294.99	0.7458	0.7412
D^{10}	1008387.0	952339.07	0.3085	0.2913
D^{11}	4217371.4	3797399.7	0.1057	0.0951
D^{12}	17791239	15324041	0.0930	0.0801

sponds to an atomic contribution, or the part that corresponds to bond contribution, respectively.

When RS/n normalization is applied to *all trans* hexatriene, instead of molecular profile:

$$10.384, 13.500, 13.739, 11.625, 8.410, 5.304, 2.962, 1.483, 0.672, 0.278, \ldots (1)$$

we obtain the normalized profile:

$$1.731, 2.250, 2.290, 1.938, 1.402, 0.884, 0.494, 0.247, 0.112, 0.046, \ldots . \qquad (2)$$

Alternatively, we may consider bond-normalized profile:

$$2.077, 2.700, 2.748, 2.325, 1.682, 1.061, 0.592, 0.297, 0.134, 0.056, \ldots . \qquad (3)$$

The profiles (1)–(3) differ only in the scaling. When the constructions of the profiles are based on atomic coordinates only the *per atom* normalization does not introduce novelties. However, in the case of $N \times N$ augmented matrices the scaling is necessary so that the derived profiles, based on the matrices of different size, can be compared.

In the case of *all trans* hexatriene based on the 11×11 distance matrix we obtain the following new *per atom* profile, or A-profiles:

$$2.943, 3.500, 3.311, 2.613, 1.767, 1.045, 0.549, 0.259, 0.111, 0.044, \ldots . \qquad (4)$$

If we use the *per bond* normalization, i.e., instead of $1/n$ we use $1/(n-1)$ as the normalization factor, we similarly obtain B-profiles:

$$3.531, 4.200, 3.974, 3.135, 2.120, 1.254, 0.659, 0.311, 0.134, 0.053, \ldots . \qquad (5)$$

A comparison of (2) and (4), or alternatively (3) and (5), shows that the difference between the molecular profile derived from the 6 × 6 matrix and the molecular profile derived from the 11 × 11 matrix are now less dramatic than without the scaling. Nevertheless, the difference between the scaled molecular profiles when using 6 × 6 and 11 × 11 matrices are considerable and call for further analysis.

5. Higher resolution profiles

In order to better understand the difference between the molecular profiles based only on interatomic distances and molecular profiles using bonding information, shortly A profiles and B-profiles, respectively, we examined several models of all *trans* hexatriene using an increasing number of "ghost" sites along CC bonds. In Table 6 we show the leading terms of the resulting profiles when m changes from $m = 0$ to $m = 9$. While the initial differences between the corresponding members of the profile sequence (for small m) are considerable, as m increases they slow down suggesting a convergence. The last row in Table 6 is based on the distance information in a 51 × 51 distance matrix, the elements of which are all raised to powers d^k up to $k = 10$. From such matrices the row sums were first calculated for all rows, then normalized and subsequently averaged. Computer program written in BASIC and run on Apple IIe personal computer takes 30 minutes to extract the bond profile from the 51 × 51 matrix raised to all powers up to $k = 10$. As input we took the coordinates given by the hexagonal grid system. Atoms were placed so as to have integers as grid coordinates, while bond sites assume fractional grid coordinates. Once the atomic and the "ghost" grid coordinates are input, the program first derives the corresponding Cartesian coordinates for each input site. From the computed Cartesian coordinates all the pairs of interatomic distances are evaluated and the $N \times N$ distance matrix is constructed. This step is followed by calculation of all powers of distances till the priorily selected values of k. The row sums for all these matrices are found, averaged and normalized.

6. Illustrations

It is desirable to have the limiting values for molecular bond profiles, that is, independent of m, the number of sites along a bond. However, this is not a necessary requirement for applications of the refined molecular profiles as outlined in this paper. It suffices that one is consistent, that is, one uses the same m value for all bonds and all molecules considered. We will illustrate the use of bond profiles on *cis* hexatriene and the closed hexagonal ring of benzene mentioned earlier. We selected $m = 7$, i.e., we represented each CC bond by eight shorter segments with fractional coordinates:

$$0.000, 0.125, 0.250, 0.375, 0.500, 0.625, 0.750, 0.875, 1.000\,.$$

Table 6

Variation of the profile with the number of bond interpolation segments for *all trans* hexatriene. The first two columns of Table 6 are derived from the last two columns of Table 5 by dividing by 6 and 11, respectively.

m:	0	1	2	3	4	5	6	7	8	9
Size:	6 × 6	11 × 11	16 × 16	21 × 21	26 × 26	31 × 31	36 × 36	41 × 41	46 × 46	51 × 51
D^1	1.73070	1.60500	1.55998	1.53684	1.52278	1.51334	1.50656	1.50148	1.49748	1.49429
D^2	2.25000	1.90909	1.79861	1.74405	1.71154	1.68996	1.67460	1.66315	1.65419	1.64706
D^3	2.28984	1.80621	1.65502	1.58142	1.53791	1.50918	1.48880	1.47363	1.46180	1.45239
D^4	1.93750	1.42510	1.27063	1.19661	1.15324	1.12477	1.10465	1.08972	1.07811	1.06890
D^5	1.40162	0.96375	0.83657	0.77661	0.74181	0.71909	0.70311	0.69128	0.68211	0.67485
D^6	0.88403	0.56997	0.48196	0.44113	0.41765	0.40242	0.39174	0.38387	0.37778	0.37297
D^7	0.49465	0.29941	0.24679	0.22276	0.20906	0.20023	0.19407	0.18953	0.18604	0.18328
D^8	0.24709	0.14145	0.11372	0.10125	0.09420	0.08969	0.08655	0.08425	0.08248	0.08108
D^9	0.11197	0.06970	0.04763	0.04184	0.03860	0.03653	0.03510	0.03405	0.03325	0.03262
D^{10}	0.04631	0.02386	0.01829	0.01585	0.01450	0.01364	0.01305	0.01262	0.01229	0.01204
D^{11}	0.01761	0.00865	0.00648	0.00554	0.05030	0.00471	0.00448	0.00432	0.00420	
D^{12}	0.00619	0.00291	0.00213	0.00180	0.00162	0.00151	0.00143	0.00137	0.00133	

The six atomic coordinates of *cis* hexatriene superimposed on a hexagonal graphite lattice are, for example,

$$(0,0,0),\quad (1,0,0),\quad (1,1,0),\quad (1,1,1),\quad (0,1,1),\quad \text{and} \quad (0,0,1)\,.$$

Notice that we have oriented the unit directions such that the above hexagonal ring avoids negative coordinates. This is not essential. If the origin of the hexagonal grid used coincides with the terminal carbon atom of *cis* hexatriene, the list of 41 input coordinates begins with

$$(0,0,0),\quad (0.125,0,0),\quad (0.25,0,0),\quad (0.375,0,0),\quad \ldots,$$

and so on. After reaching the atom $(1,0,0)$ we continue to move through the segments of the next CC bond:

$$(1,0,0),\quad (1,0.125,0),\quad (1,0.25,0),\quad (1,0.375,0),\quad \ldots,$$

and so on, till we reach point $(1,1,0)$. The input (which itself can be programmed, [28]) continues till all CC bonds have been traversed. The order in which the coordinates are listed is immaterial. However, it is advisable to apply some systematic way of listing sites so that entries in the distance matrix can be easily located and inspected if necessary. With each bond having eight segments for hexatriene we obtain a 41×41 distance matrix. The corresponding matrix for the hexagonal ring structure is 48×48, since in the ring there is an additional CC bond with seven bond sites. The molecular profile for *cis* hexatriene and the benzene ring structure are the same (listed in the first column of Table 7). Bond profiles for the two structures are different as can be seen by comparing the entries in the last two columns of Table 7. The difference between the bond profiles of *cis* hexatriene and the benzene ring is not large. However, also the two idealized structures are not very different! The profile of benzene dominates that of hexatriene, i.e., the corresponding entries for benzene ring are greater than those for hexatriene.

Table 7
The profiles for the six member ring and its spanning tree based on $m = 8$ bond interpolation segments.

	$m = 0$	Hexatriene	Ring
D^1	1.244017	1.141722	1.164096
D^2	1.000000	0.818077	0.835938
D^3	0.566453	0.424257	0.432889
D^4	0.250000	0.172176	0.175232
D^5	0.090523	0.057375	0.058221
D^6	0.027778	0.016219	0.016401
D^7	0.007392	0.003981	0.004014
D^8	0.001736	0.000864	0.000868
D^9	0.000365	0.000168	0.000168
D^{10}	0.000064	0.000030	0.000030

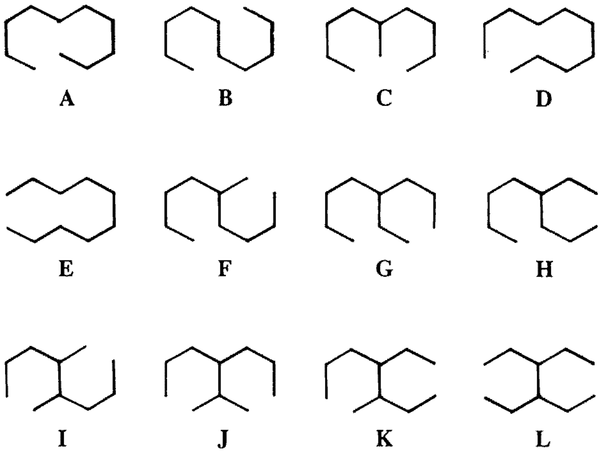

Fig. 4.

As another illustration we will derive the bond profiles for twelve distinct spanning trees of naphthalene, which are illustrated in Fig. 4. All these structures, including naphthalene itself, would have the same A-profile, the profile based solely on the coordinates of carbon atoms. For computation of their B-profile we have selected $m = 4$. In Table 8 we show the corresponding profiles. The differences between different spanning trees are again not very large but nevertheless significant. The leading member in the profile sequence suffices to discriminate any pair of spanning trees, and hence may suffice to label individual trees uniquely. Consequently, the leading term induces ordering of the spanning trees (illustrated in Fig. 4).

7. Discussion

A question can be raised: Are the B-profiles unique? Are there nonisomorphic 3-dimensional structures that have an identical B-profile? Such structures may have atoms (vertices) at the same positions, i.e., have the same geometry but have different connectivity, i.e., different bonding pattern. The same questions can be raised concerning A-profiles: Are there nonisomorphic structures that have an identical A-profile? Such structures will differ at least in the coordinates of a single one vertex (atom). Alternatively one can ask: Can a molecule be reconstructed from the information given by its B-profile (or A-profile, respectively)?

The uniqueness of the profiles and the reconstruction problem would be settled if one finds two nonisomorphic structures that have identical profiles. Recently we examined two pairs of closely related structures that could produce the same molecular A-profile [28,29]. These are the semiregular polyhedra shown in Fig. 5. The first is truncated octahedron and its twist form obtained by a rotation of half of the

Table 8
The profiles for the twelve symmetry nonequivalent (embedded) spanning trees of naphthalene.

	A	B	C	D	E	F	G	H	I	J	K	L
D^1	1.6747	1.6996	1.7595	1.5592	1.5825	1.6395	1.6029	1.6595	1.7161	1.6091	1.6680	1.7291
D^2	1.7664	1.8373	1.9791	1.5283	1.5878	1.7069	1.6417	1.7664	1.9026	1.6587	1.7835	1.9196
D^3	1.3983	1.5012	1.6755	1.1302	1.2098	1.3445	1.2848	1.4344	1.6037	1.3083	1.4519	1.6158
D^4	0.9011	1.0025	1.1502	0.6860	0.7594	0.8670	0.8323	0.9587	1.1033	0.8542	0.9708	1.1094
D^5	0.4934	0.5698	0.6662	0.3569	0.4093	0.4765	0.4642	0.5474	0.6425	0.4796	0.5539	0.6449
D^6	0.2354	0.2822	0.3338	0.1632	0.1938	0.2286	0.2278	0.2728	0.3239	0.2365	0.2756	0.3247
D^7	0.0996	0.1238	0.1472	0.0666	0.0818	0.0972	0.0996	0.1204	0.1437	0.1038	0.1214	0.1440
D^8	0.0378	0.0486	0.0580	0.0245	0.0311	0.0371	0.0392	0..476	0.0569	0.0409	0.0479	0.0569
D^9	0.0130	0.0173	0.0206	0.0082	0.0108	0.0128	0.0140	0.0170	0.0203	0.0146	0.0171	0.0203
D^{10}	0.0041	0.0056	0.0067	0.0025	0.0034	0.0040	0.0046	0.0055	0.0066	0.0048	0.0056	0.0066
D^{11}	0.0012	0.0017	0.0020	0.0007	0.0010	0.0012	0.0014	0.0017	0.0020	0.0014	0.0017	0.0020
D^{12}	0.0003	0.0005	0.0005		0.0002	0.0003	0.0004	0.0005	0.0005	0.0004	0.0005	0.0005

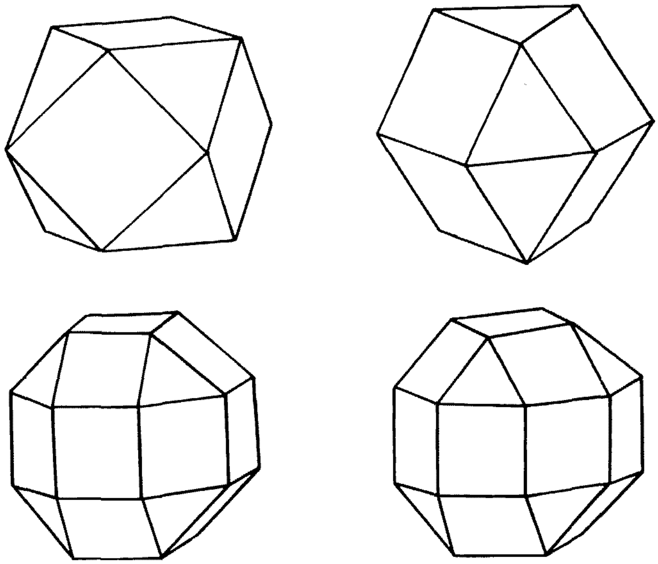

Fig. 5.

polyhedron on an axis through the center by $\pi/3$. The other pair of the structures having the same A-profile is small rhombicuboctahedron and its twist form.

The profiles of cuboctahedron and its twist form, however, are different even though the difference is small. ${}^1D$ values are 15.7566 and 15.7559 for cuboctahedron and twist-"cuboctahedron", respectively. Molecular bond profiles (B-profiles) have more information and are likely to be unique.

Existence of structures that have an identical characterization would show that A-profiles do not allow reconstruction. Such a result may be expected. Structural invariants are associated with loss of information and it is generally believed that a finite list of invariants is not likely to discriminate among all structures. However, the negative result is not necessarily detrimental. Structures that have similar magnitudes for their invariants are likely to have similar physicochemical properties. Hence, such invariants could be suitable for structure-property studies and could adequately describe structures even if occasionally duplications occur since such structures may have the same magnitude for several properties. In the cases of truncated octahedron and rhombicuboctahedron the polyhedra and their corresponding twist forms have the same surface, the same volume, the same respective moments of inertia about the axis of the twist, and possibly same additional properties.

Let us comment on the form of the normalization of d_{ij}^k entries of the distance matrices. Clearly $1/k!$ is not the only possibility and one could equally use $1/d_{ij}$ or some other such quantity. Alternative normalization is possible but will not be considered here. If we want to emphasize the importance of atomic interactions at the short distance then the exponential normalization of the type $1/k!$ that we have already selected is better than $1/d_{ij}$. The advantage of the factorial normalization is that it assures absolute convergence of the power expansion for any conceivable interatomic separations.

Are the A-profiles just a modification of Crippen's distance geometry [30]? The distance geometry is concerned with the constraints imposed by interatomic separations on the geometry of the molecular structure. Here we are interested in molecular *invariants* derived from the distance matrix, we are not interested in molecular geometry as such. Even when we consider the elements of the distance matrix ^1D, we do not use them directly (as components of vectors), but use them to extract structural invariants.

The distance geometry does not consider explicitly the molecular bonding. In distance geometry the input information on a structure is given by the atomic coordinates and no information on the bonding is directly contained in the distance matrix. The bonding profiles have been introduced to record the connectivity as reflected by the bonding pattern. Atomic separations and atomic coordinates need not necessarily indicate the connectivity, although in most cases one can infer the bonding pattern from known interatomic separations. However, when ambiguities arise, such as in some small highly strained systems, the bonding pattern has to be deduced from quantum chemical calculations that give electron densities, not the atomic coordinates.

Do matrices ^2D, ^3D, ^4D, ^5D, ... introduce novel information that the geometry matrix ^1D does not contain?

If D^k is unwarranted, one could then argue similarly that the higher powers of the adjacency matrix A^n are unwarranted, and in parallel do not introduce novel information. However, it is not easy to count the walks of length n in a graph not using the matrix A^n. It is difficult to identify and recognize graphs that have integers eigenvalues, or even a single integer eigenvalue, from examination of the characteristic polynomial without solving it. An inspection of the characteristic polynomials does not yield such information easily. The characteristic polynomial apparently has its advantages, the coefficients are integers, and as was hinted by Coulson [31] and later fully outlined by Sachs [32] and others, they enumerate qualified subgraphs in a structure [33]. Nevertheless, the graph eigenvalues have found useful in various applications of graphs in chemistry. For example, Lovasz and Pelikan [34] suggested that the first eigenvalue for trees (acyclic graphs) is an index of molecular branching. Similarly, the first eigenvalue of D/D matrices has been suggested as an index of the molecular folding [27]. Without examination of eigenvalues, i.e., by confining attention only to the characteristic polynomials, such indices would have never been considered.

8. Concluding remarks

Molecular bond profiles are to be viewed as a novel tool to represent molecules.

Their prime advantage is that they equally apply to simple molecular models, such as the molecular skeletons represented by graphs, and elaborate 3-dimensional molecular models. Molecular profiles were already used in discussing molecular similarity of planar benzenoids [23] and the similarity among 3-dimensional nine-membered puckered rings [26]. Molecular profiles were also used in quantitative structure-property studies (e.g., the boiling points in planar benzenoids [25] and the chromatographic retention data [24]). The molecular profiles, in particular the bonding profiles, and their extensions for characterization of contours of arbitrary form, appear particularly suitable for a quantitative approach to molecular shape – an elusive concept that has recently received considerable attention [20–25,34–36]. Our prime motive for the development of the molecular profiles and the molecular bond profiles is in anticipation of their use in the studies of drug–receptor interactions. The efficiency of the docking procedure for molecular recognition critically depends on the representation of the geometry of the guest and the host molecules. As is known, nonvisual approaches are time consuming as they imply point by point matching, which have often to be backtracked. There are several arguments against approaches that use calculated molecular energies for establishing molecular recognition [37]. Apparently the brute force grid-search methods are impractical, because of excessive time consumption that they would require. Visual approaches offered by some computer graphic packages, on the other hand, imply a trial and error approach, which is, to say the least, inefficient, not necessarily reliable, and qualitative rather than quantitative. A robust and efficient automated docking for molecular recognition based on use of vectors to represent structures is possible as has been recently reported [38]. We hope that our molecular profiles and bond profiles will develop into an alternative efficient automated docking approach for molecular recognition based on geometry dependent molecular invariants that may turn out even more efficient, and certainly more user friendly than hitherto used schemes. The heightened efficiency follows precisely because we use invariants, quantities independent of labels (and orientations of molecules) and thus avoid point by point comparisons and time consuming backtracking search algorithms.

References

[1] R.C. Read, J. Chem. Inf. Comput. Sci. 23 (1983) 135.

[2] N. Trinajstić, *Chemical Graph Theory* (CRC Press, Boca Raton, FI, 1992).

[3] M. Randić, Croat. Chem. Acta 66 (1993) 289.

[4] A.R. Katritzky and E.V. Goordeeva, J. Chem. Inf. Comput. Sci. 33 (1993) 835.

[5] S.C. Basak, V.R. Magnuson, G.J. Niemi, R.R. Regal and G.D. Veith, Math. Model. 8 (1986) 300.

[6] D.E. Needham, I-C. Wei and P.G. Seybold, J. Amer. Chem. Soc. 110 (1988) 4186.

[7] M. Randić, J. Math. Chem. 7 (1991) 155.

[8] M. Randić, J. Math. Chem. 9 (1992) 97.

[9] M. Randić and N. Trinajstić, J. Mol. Struct. (Theochem) 300 (1993) 551.
[10] M. Randić and N. Trinajstić, J. Mol. Struct. (Theochem) 284 (1993) 209.
[11] M. Randić, P.J. Hansen and Jurs, J. Chem. Inf. Comput. Sci. 28 (1988) 60.
[12] A.T. Balaban, I. Motoc, D. Bonchev and Ov. Mekenyan, Top. Curr. Chem. 114 (1983) 21.
[13] M. Randić, Studies Phys. Theor. Chem. 54 (1988) 101.
[14] M. Randić, Int. J. Quant. Chem.: Quant. Biol. Symp. 15 (1988) 201.
[15] M. Randić, B. Jerman-Blazic and N. Trinajstic, Comp. & Chem. 14 (1990) 237.
[16] Z. Mihalić and N. Trinajstić, J. Mol. Struct. (Theochem) 232 (1991) 65.
[17] B. Bogdanov, S. Nikolić and N. Trinajstić, J. Math. Chem. 3 (1983) 299.
[18] O. Mekenyan, D. Peitchev, D. Bonchev, N. Trinajstić and I. Bangov, Drug. Res. 30 (1986) 176.
[19] O. Mekenyan, D. Peitchev, D. Bonchev, N. Trinajstić and J. Dimitrova, Drug. Res. 36 (1986) 629.
[20] M. Randić and M. Razinger, J. Chem. Inf. Comput. Sci. 35 (1995) 140.
[21] M. Randić and M. Razinger, J. Chem. Inf. Comput. Sci. 35 (1995) 594.
[22] M. Randić and M. Razinger, J. Chem. Inf. Comput. Sci. 36 (1996) 429.
[23] M. Randić, J. Chem. Inf. Comput. Sci. 35 (1995) 373.
[24] M. Randić, New J. Chem. 19 (1995) 781.
[25] M. Randić, New J. Chem., in press.
[26] M. Randić, Int. J. Quant. Chem.: Quant. Biol. Symp. 22 (1996) 61.
[27] M. Randić, A.F. Kleiner and L.M. DeAlba, J. Chem. Inf. Comput. Sci. 34 (1994) 277.
[28] G.M. Crippen, J. Comput. Phys. 24 (1977) 96.
[29] M. Randić and G. Krilov, Int. J. Quant. Chem.: Quant. Biol. Symp., in press.
[30] M. Randić and G. Krilov, Croat. Chem. Acta, to be submitted.
[31] C.A. Coulson, Proc. Camb. Phil. Soc. 46 (1950) 202.
[31] H. Sachs, Publ. Math. (Debrecen) 11 (1964) 119.
[32] N. Trinajstić, *Chemical Graph Theory* (CRC Press, Boca Raton, FI, 1992) chap. 5, pp. 61–83.
[33] L. Lovasz and A. Pelikan, J. Period. Math. Hung. 3 (1973) 175.
[34] P.G. Mezey, *Shape in Chemistry: Introduction to Molecular Shape and Topology* (VCH, 1993).
[35] P.G. Mezey, J. Chem. Inf. Comput. Sci. 32 (1992) 650.
[36] P.G. Mezey, J. Comput. Chem. 8 (1987) 462.
[37] M.L. Connolly, Biopolymers 25 (1992) 1229.
[38] N. Kasinos, G.A. Lilley, N. Subbarao and I. Haneef, Protein Eng. 5 (1992) 69.
[39] M. Randić, J. Chem. Inf. Comput. Sci. 31 (1991) 311.
[40] M. Randić, New J. Chem. 15 (1991) 517.
[41] M. Randić, Croat. Chem. Acta 64 (1991) 43.
[42] M. Randić, Int. J. Quant. Chem.: Quant. Biol. Symp. 21 (1994) 215.