

On Characterization of Molecular Shapes

Milan Randić^{*,†} and Marko Razinger

National Institute of Chemistry, 61115 Ljubljana, P.O. Box 30, Republic of Slovenia

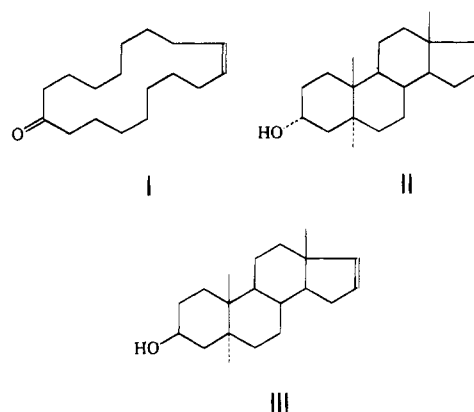
Received July 19, 1994[®]

We consider the problem of characterization of molecular shape. In particular we consider shapes of planar objects derived by fused regular hexagons, i.e., shapes that can be embedded on a regular hexagonal grid. We have introduced binary codes for the representation of molecular shapes and have limited our attention to planar benzenoid forms of a constant perimeter. The form of the code depends on the selection of the starting bond and the sense of circling around the molecular perimeter. We selected as the canonical code one that results in the smallest binary number for the periphery code. The binary canonical codes offer a systematic catalog of all possible shapes defined on the graphite lattice. The binary codes also fully reflect the symmetry of shapes, while chiral structures show different codes for their mirror forms. Finally, we discussed similarity of shapes (using Hamming distance between the codes as the measure of dissimilarity). The numerical magnitude of the similarity of codes depends on the length of the fragment of the code to be compared. With more entries (longer the local sequences of comparison) a better discrimination between similar objects is obtained. Calculating similarity by comparing codes using a longer "window" (longer local subsequence) was accomplished by a computer program written for that purpose (in FORTRAN).

INTRODUCTION

One way to characterize molecular structure is to use as a basis for a description of a structure a selection of mathematical invariants. From such one constructs sequences, the components of which are the selected invariants, descriptors, or topological indices.¹ Because of very strong inter-relatedness of molecular descriptors, although different authors often use different descriptors, alternative representations of a structure, if comprehensive, will ultimately reflect the same structural information. Similar considerations can be extended toward characterization of molecular shape. Here the situation is more difficult since while there are differences among chemists how to define the molecular structure, no attempts were made to define the molecular shape. Disagreements about how to define molecular structure are not serious. The differences are in the details and not in the substance. When one considers molecular shapes, there are hardly papers addressing the substance of the problem. Generally, many chemists have a "feeling" of what a molecular shape represents, but there is no useful, practical, or conceptual definition of what the molecular shape is, which would allow one to derive a quantitative characterization of this important concept.

Manifestations of molecular shape are well-known. For example, Ruzicka and Prelog² reported on a similar strong musk odor of a macrocyclic musk civetone (I) and steroids 3 α -hydroxy-5 α -androst-16-ene (II) and 3 β -hydroxy-5 α -androst-16-ene (III). The molecules apparently have a similar shape, but chemically macrocyclic civetone and the two steroids have quite distinct and unrelated chemical structures. The notion that similar molecules have similar properties goes back to the pioneering work of Emil Fischer³ and is the basis for many QSAR studies⁴ and molecular similarity analyses.⁵



The studies of drug–receptor interactions clearly point to a need to investigate more closely molecular shapes. The efficiency of the docking algorithms for molecular recognition critically depends on the characterization of the geometry of the guest and the host molecules,⁶ i.e., on the relative shapes of the interacting molecules. Connolly has given several arguments against calculations of the molecular energy for the molecular recognition.⁷ One of the factors against the use of the electron energy calculations for the studies of molecular conformations is the time necessary to perform the computations for large systems. Connolly has estimated that the brute force grid-search methods would require several years of computer time to predict the correct mode of the binding between macromolecules. However, it is not only the prohibitive time that is a crucial factor in such works. One wonders how much insight into the structure–property relationships offers such energy based calculations that produce "tons of numerical data" that one has difficulty to digest. Such data has to be reduced by either using standard data reduction techniques or developing novel concepts. The bottleneck to the progress apparently is of a conceptual nature, rather than caused by complexity of numerical work.

[†] On sabbatical leave from Dept. of Mathematics and Computer Science, Drake University, Des Moines, IA 50311.

[®] Abstract published in *Advance ACS Abstracts*, April 1, 1995.

There is an alternative approach to the numerical (number crunching) approach to structure–property relationships. The emphasis is shifted from the numerical calculations to various molecular components and relevant molecular fragments. One is interested in how various parts are connected and how various components are interrelated. Such approaches focus attention on *comparisons* between different structures, rather than being concerned with a single molecule and its properties. Often such analysis also leads to numerical and computer intensive computations. The focus is on a combined combinatorial and topological, i.e., graph theoretical, properties of molecules. Graph theory,⁸ which was known prior to 1920 as “combinatorial and topological mathematics”, is the prime mathematical tool for comparative study of large and complex systems. There is considerable literature to attest to the significant results of chemical graph theory,⁹ not only in problems that require chemical enumerations,¹⁰ constructions,¹¹ and structure manipulations¹² (by computers) but also when one considers structure–property and structure–activity relationships.¹³

Often it is assumed that graphs cannot model molecules as three-dimensional objects. This however is not the case, although most of chemical graph theory was applied to problems in which molecules were not treated as three-dimensional objects. Some schemes of graph theory have been successfully extended to molecules as three-dimensional objects.^{14,15} Our approach to molecular shape in this paper is an additional illustration of an extension of graph theoretical methodology to molecular geometry, rather than limiting attention to the molecular connectivity, as has been the tradition in the chemical graph theory.

MOLECULAR SHAPE

In a series of articles Mezey and Arteca tried studying topological properties of molecular surfaces in order to quantify molecular shape.¹⁶ In particular they considered curvatures of portions of the molecular surface that have distinct topological properties as a shape characteristic. Recently a published book of Mezey offers an introduction to the study of molecular shape, particularly from the point of view of topology.¹⁷ Haneef and collaborators⁶ considered the problem of molecular docking that critically depends on molecular shapes. Their approach is combinatorial rather than topological. They developed an automated method for determining the most likely mode of binding for two molecules solely from three-dimensional structures of the molecules. The method (based on graph theoretical techniques) of the docking was purely numerical; the results were found without the use of computer graphics or other visual aids.

We will focus attention on a characterization of the molecular shape by way of the binary molecular codes. Instead of considering the general problem of molecular shapes we will consider a simpler task: The characterization of the shapes of planar benzenoids. Although it may appear as a rather limited topic, as will be seen, our approach applies equally to characterization of the shape of an arbitrary closed planar curve and can be even extended to an arbitrary three-dimensional object in space.

It is important to emphasize the distinction between molecular invariants that are independent of the orientation

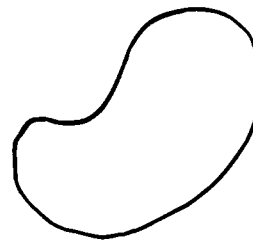


Figure 1. An arbitrary closed curve in the plane.

of an object and molecular codes that depended on orientation of the objects and choice of labels. Several geometry-dependent invariants, i.e., invariants that depend on molecular conformation, have been discussed in the literature.¹⁵ They were all derived as matrix invariants but from matrices the elements of which depend on the molecular geometry. Structural (geometry-dependent) invariants, as a rule, do not completely specify a structure. Two or more structures may have the same (finite) list of structural invariants. In contrast molecular codes, such as the binary periphery codes, completely specify an object (a molecule in three-dimensional space). Hence, from a code (knowing the rules for construction of the code) one can reconstruct the object. Invariants have other advantages. Once defined, they are independent of the labels used to specify the structure so different laboratories will derive identical numerical values for the invariants considered even if they use a different labeling system for their structures.

CHARACTERIZATION OF SHAPES

Our prime interest here is a characterization of molecular shape. We will seek codes that are sensitive to details of molecular shape and which allow one to reconstruct a molecular shape. In Figure 1 we show an arbitrary closed curve. A complete mathematical characterization of the curve of Figure 1 is possible using Fourier expansion, since moving along a closed curve represents a periodic motion. The coefficients of Fourier expansion suffice for reconstruction of the curve (shape). However, without a prior specification of the location of the origin of the Cartesian coordinate system relative to the closed curve and the relative orientation of its axes and the curve different laboratories will arrive at different Fourier expansions. Hence, Fourier expansion gives a code, while we would like to have invariants of such an expansion. For more details on Fourier transformation and regeneration of images see ref 18.

There are other ways to characterize shapes of closed planar curves. One can consider a set of equally spaced discrete points along the curve. Construction of a code for such a curve is then equivalent to a design of a code for a graph embedded in a plane. As a result one obtains for the code a sequence. The length of the sequence will depend on the size of the object and on the level of the resolution, i.e., on the relative size of the structure and the underlying mesh. Size dependence of such codes can be eliminated by a suitable normalization.

The above outlines our route to a characterization of molecular shapes. Admittedly, here we consider only smaller planar benzenoids, but as it will become evident the methodology is general. Rather than considering a general

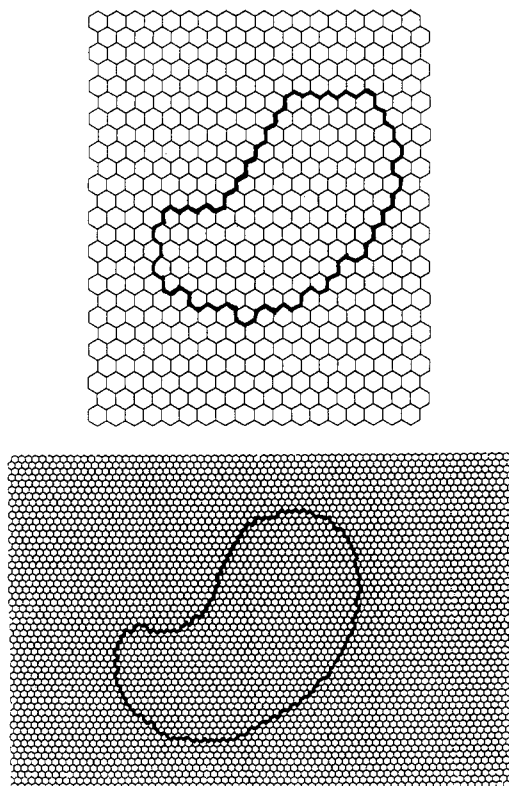


Figure 2. The curve from Figure 1 shown as a sequence of straight lines—a closed path on the hexagonal lattice. Two different resolutions are shown: the finer the tessellation, the better the fit of the shape.

shape, such as that of Figure 1, we will consider shapes that are drawn on a hexagonal lattice (graphite) (Figure 2). The quality of approximation of a smooth curve by an outline on a hexagonal lattice will depend on the relative size of the object and size of the elementary hexagon. In principle there is no limit to the power of resolution of a lattice, but for practical reasons one may restrict the length of the codes, that depends on the size of the lattice mesh.

Our first task is to derive a satisfactory code for the periphery of simple benzenoid shapes. Such codes should be linear, have structural origin, be simple, be of similar length for objects of similar size, be unique, and should allow reconstruction. By having a structural origin the codes will lead to some structural ordering of objects. Hence, one may derive a complete catalogue of shapes using such codes. Moreover, one expects that the codes of similar compounds will be similar. The codes that we will now introduce satisfy all the above conditions. Therefore we will use the binary periphery codes to discuss quantitatively the degree of similarity of shapes. In concluding remarks we will hint on an extension of the outline shape periphery codes to a characterization of the shapes of three-dimensional objects.

PERIPHERY CODES

In Figure 3 we have illustrated the simplest shapes possible on a hexagonal map, those given by the periphery of benzene, naphthalene, phenalene, anthracene, phenanthrene, and pyrene. The peripheries of these shapes are given by 6, 10, 12, and 14 edges (CC bonds). Apparently these are the only possible shapes with periphery length of 14 or fewer CC bonds. A simple way to represent the periphery of such objects

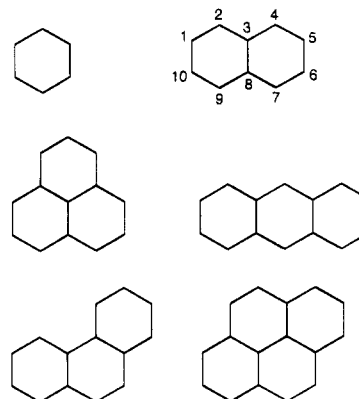
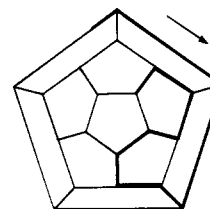


Figure 3. The six simplest shapes which can be drawn on a hexagonal lattice.

parallels the description of a route on a dodecahedron outlined in the book of Rouse Ball.¹⁹ As one moves over



the cubic graph of dodecahedron at each point there is a choice of left (L) and right (R) path. Hence, a sequence of L and R symbols will fully describe a history of movements over the edges of cubic graphs (such as Schlegel projection of dodecahedron). We use this approach to describe peripheries of benzenoid shapes that can be viewed as embedded on a graphite lattice. In order to label our moves around the molecular periphery, since a closed curve has “inside” and “outside” regions, instead of L and R we may use I (for inside) and O (for outside), or simply we may use the binary labels 1 and 0. Such binary code is reminiscent of “walk around” code of Read.²⁰

Knop, Trinajstić, and co-workers¹¹ introduced a code, DAST (an acronym for dualist angle-restricted spanning tree), which uniquely characterizes benzenoid systems. In DAST the entries in the code (digits from 0 to 7) uniquely label each hexagonal present. Such codes have been used to construct benzenoids having a given number of hexagons. We are interested in shapes having the periphery of a same length. Moreover, from our point of view molecular shape is given by the molecular boundary, i.e., by the molecular contour for planar molecules and the molecular surface for three-dimensional molecules. Molecules that have the same periphery may still differ in their interior structure. Thus the shapes given by polyhexes with holes are not discriminated from the shapes of the corresponding molecules without holes or differing internally. The binary codes we use here are not new in chemistry. Already in 1971 Balaban used binary codes to generate distinct annulenes.²¹ Periphery codes have already been introduced for shapes in the form of a complex code of square-cell configurations (animals) using complex plane representation.²²

The smallest benzenoids, benzene and naphthalene and hypothetical phenalene, have no isomers, and hence represent the only possible shape for objects of such size. In the case of benzene there is but a single binary code, since all six carbon atoms and all six CC bonds are strictly equivalent.

Thus for benzene we obtain the unique code:

	code	periphery length
benzene	1 1 1 1 1 1	6

In naphthalene there are three nonequivalent carbon atoms and two possible directions of circling around which results in a larger number of possible codes. In Figure 3 we have numbered carbon atoms on the periphery for naphthalene. The binary codes can be viewed as "walk around" type code with a distinction that we use 0 and 1 to represent moving left or right at each carbon atom as we move along the perimeter. If we follow the clockwise direction then moving right represents inside, 1 or 1, while moving left represents outside, 0 or zero (0). Depending on the selection of the starting atom we generate different codes for naphthalene periphery:

start	code	length
1	1 1 0 1 1 1 1 0 1 1	10
2	1 0 1 1 1 1 0 1 1 1	
3	0 1 1 1 1 0 1 1 1 1	
4	1 1 1 1 0 1 1 1 1 0	
5	1 1 1 0 1 1 1 1 0 1, etc.	

Because the (directed) edges 1–2 and 6–7 are symmetrical atom 6 will produce the same code as atom 1, and so on. Hence, the code reflects the presence of the center of symmetry in naphthalene. The code also reflects other symmetries of the naphthalene shape. From the code derived by starting at atom 1 we see that the molecule has a plane of reflection, because the first part of the code is the same as the last part of the code. The code can be written simply as $(1\ 1\ 0\ 1\ 1)^2$ with the repeating section also having symmetry.

We have to select one of the five codes of naphthalene and make a rule for the canonical code. In the case of a shape having n bonds (or n atoms) and no symmetry, there will be n different codes. The derived binary codes can be viewed as distinct binary numbers each having a different numerical value. In the case of naphthalene the code initiated by atom 3 corresponds to the smallest such number possible. On the other hand, the code derived by starting with atom 4 corresponds to the largest binary number. Either of the two is unique and can be used as the canonical label. We will use the smallest binary number of the code as its canonical form.

The next possible length for polihex periphery is 12, found in phenalene. Its canonical code is 0 1 1 1 0 1 1 1 0 1 1 1 or $(0\ 1\ 1\ 1)^3$.

Benzenoids having a larger periphery can assume more than one shape form. For the case $P = 14$ (i.e., periphery has 14 carbon-carbon bonds) there are three different shapes. They correspond to shapes defined by the periphery of anthracene, phenanthrene, and pyrene (Figure 3). Analogy with discriminating conjugated circuits of different shape is in place. Conjugated circuits,²³ which play an important role in resonance theory,²⁴ are circuits within individual Kekulé valence structures in which there is a regular alternation of CC double and CC single bonds. Klein and collaborators²⁵ raised the issue of the relative contributions of conjugated circuits of different shape by discriminating conjugated

Table 1. Canonical Binary Periphery Codes of All Benzenoids Having Periphery $P = 18$

benzphenanthrene	0 0 0 1 1 1 1 0 1 1 0 1 1 0 1 1 1 1
benzanthracene	0 0 1 0 1 1 1 1 0 1 0 1 1 0 1 1 1 1
benz[a]pyrene	0 0 1 1 0 1 1 1 1 0 1 1 0 1 0 1 1 1
chrysene	0 0 1 1 0 1 1 1 1 0 0 1 1 0 1 1 1 1
dibenzpyrene	0 0 1 1 1 0 1 1 1 0 1 1 0 1 1 0 1 1
benz[d]pyrene	0 0 1 1 1 0 1 1 1 0 1 1 1 0 0 1 1 1
perylene	0 0 1 1 1 0 1 1 1 0 0 1 1 1 1 0 1 1
triphenylene	0 0 1 1 1 1 0 0 1 1 1 1 1 0 0 1 1 1
tetracene	0 1 0 1 0 1 1 1 1 1 0 1 0 1 0 1 1 1
anthrathene	0 1 0 1 1 0 1 1 1 1 0 1 0 1 1 0 1 1
triangulene	0 1 0 1 1 1 0 1 0 1 1 1 1 0 1 0 1 1
coronene	0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1

Chiral Pairs	
benzanthracene	0 0 1 1 1 1 1 1 1 0 1 0 1 1 1 1 0 1
benz[a]pyrene	0 0 1 1 1 1 1 0 1 0 1 1 0 1 1 1 0 1
chrysene	0 0 1 1 1 1 1 0 1 1 0 0 1 1 1 1 0 1
anthrathene	0 1 0 1 1 0 1 1 1 1 0 1 0 1 1 0 1 1

circuits associated with the periphery of anthracene, phenanthrene, and pyrene, all of the same size.

The minimal binary codes for the three different shapes of periphery $P = 14$ are

phenanthrene	0 0 1 1 1 1 0 1 1 0 1 1 1 1
anthracene	0 1 0 1 1 1 1 0 1 0 1 1 1 1
pyrene	0 1 1 0 1 1 1 0 1 1 0 1 1 1

If we shift the origin of the code suitably the three codes can be written as

phenanthrene	(0 1 1 1 1 0 1)(1 0 1 1 1 1 0)
anthracene	(1 1 0 1 0 1 1)(1 1 0 1 0 1 1)
pyrene	(1 0 1 1 1 0 1)(1 0 1 1 1 0 1)

In this form the codes reflect the symmetry of the underlying shapes. The codes for anthracene and pyrene shapes, which possess a higher symmetry, have not only the two identical brackets, but within each of the brackets the code is symmetrical. In the case of phenanthrene, which is of lower symmetry, the part of the code within the bracket lacks such additional symmetry.

SHAPES OF SMALLER BENZENOIDS

In Table 1 we give the smallest binary codes for all benzenoids having periphery $P = 18$ depicted in Figure 4. Four of the twelve benzenoid forms are chiral. These chiral shapes either have no symmetry (except the symmetry of the plane in which they are embedded) or have only the center of symmetry (but not planes of reflection perpendicular to the plane of the molecule). The former is the case with benzanthracene and benz[a]pyrene; the latter is the case with chrysene and anthanthrene. One enantiomer of these two-dimensional chiral shapes (molecules) is listed in the upper part of Figure 4, where the structures have been ordered according to the smallest binary code, while the other enantiomer is shown in the lower part of Figure 4.

Chirality in chemistry is associated with the three-dimensional molecular forms. It is a consequence of the asymmetrical substitution of tetrahedral carbons. However, already at the beginning of this century Kelvin has defined chirality for objects in n -dimensional spaces. Chiral objects are those that cannot be superimposed with their mirror image.²⁶ This includes chirality for objects in a two-dimensional space that recently Buda and Mislow discussed

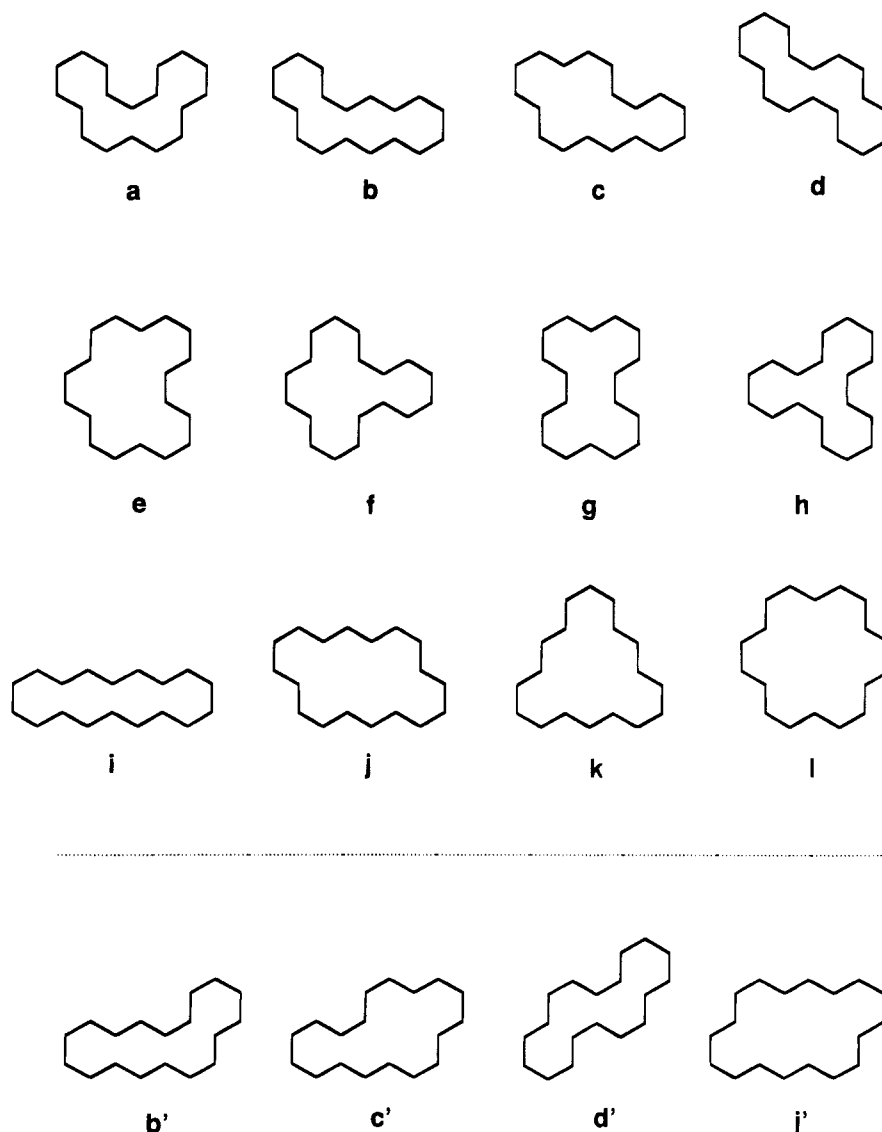
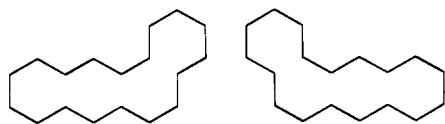


Figure 4. The shapes of all benzenoids having periphery $P = 18$; four of the eleven shapes are chiral, and their corresponding enantiomeric shapes are shown at the bottom.

for a case of asymmetrical triangles in a plane.²⁷ Hence we can distinguish



although it is arbitrary which of the two forms we call left or *l*-benzanthracene and which we call right or *d*-benzanthracene. The remaining eight shapes of Table 1 and Figure 4 have only a single form. That is, these benzenoid forms can be brought into complete overlap with their mirror images by sliding the molecules within the plane in which they are embedded. Hence, they are achiral. It should be noticed that in 2D space the presence of a center of symmetry does not preclude chirality, while in 3D it does.

Each form of the chiral shapes, as illustrated in the lower part of Table 1, has a different code. Since the corresponding codes do not show symmetry (beyond the center of symmetry and the plane of the molecule that are here irrelevant), one can deduce from the binary code whether a shape is chiral or not. If a code when read from the left to the right and

then from the right to the left gives a different answer when transformed to the canonical form the shape is chiral. If a code is palindrome, i.e., when read from the left to the right and then from the right to the left gives the same answer when the code is brought to its canonical form the structure is achiral. The codes of enantiomers are the mirror images of one another. If a code of an enantiomer is read from the right to the left (instead of the usual reading from the left to the right), it would give the code of the other enantiomer.

It is desirable sometimes that two enantiomers have nearly identical codes, differing only in plus/minus sign, since they certainly have the same shape, modulo \pm . A simple procedure to effect this is, of course, to walk around one enantiomer clockwise, that is usual, and around the other enantiomer in a counterclockwise sense (always writing 0 for outside and 1 for inside). One could then declare the smaller of the two codes as the canonical.

ORDERING OF CODES

The entries in Table 1 have been ordered so that the first shape, that of the periphery of benzphenanthrene, has the smallest binary code, while the last shape, that of the

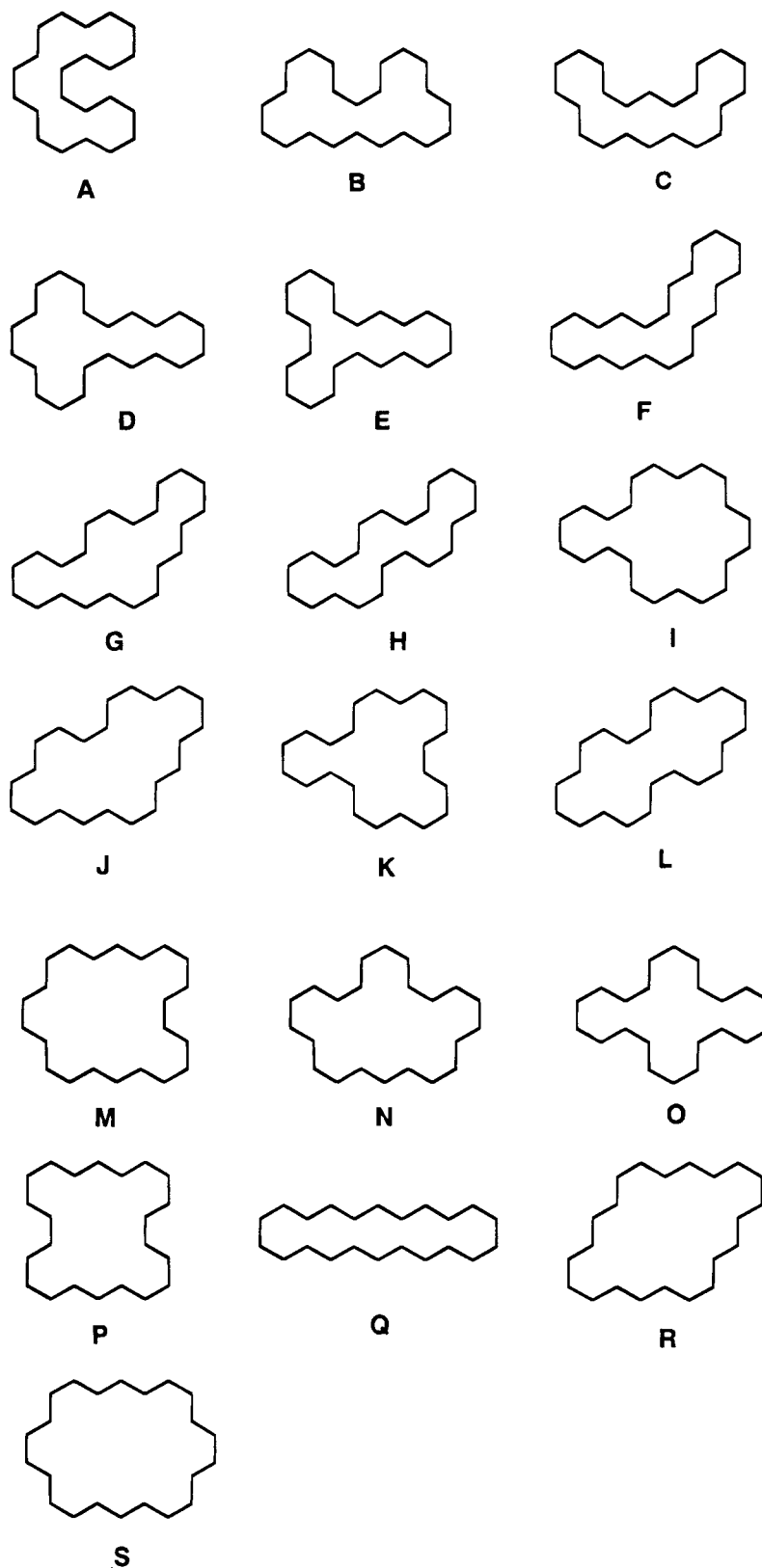


Figure 5. All 19 achiral shapes having periphery $P = 22$.

periphery of coronene, has the largest binary code among the structures that have the same perimeter.

The question can be raised: Does the lexicographic ordering of the shapes, based on the relative magnitudes of the canonical codes of Table 1, reflect some inherent structure characteristic? In Figure 4 we have ordered molecular diagrams of the smaller benzenoids according to their shape code. Inspection of Figure 4 does not immediately reveal

an obvious regularity. Difficulty to discern some regularities in the ordering of structures are in part due to their being too few structures and too many structural diversities to allow one to recognize a dominant regularity if present. There are too many competing structural factors: bending of the periphery (the number of "kinks"), size (as measured by the number of fused hexagons), branching, and different modes of fusion of hexagons. To make things worse four out of

Table 2. Canonical Binary Periphery Codes of All Achiral Benzenoids Having Periphery $P = 22$

molecule	periphery code																				
A	0	0	0	0	1	1	1	1	0	1	1	0	1	1	0	1	1	1	1	1	1
B	0	0	0	1	1	1	0	1	1	1	0	1	0	1	0	1	1	1	0	1	1
C	0	0	1	0	0	1	1	1	1	0	1	1	0	1	0	1	1	0	1	1	1
D	0	0	1	0	1	1	1	1	0	1	0	0	1	1	1	0	1	1	0	1	1
E	0	0	1	0	1	1	1	1	0	1	0	0	1	1	1	1	0	0	1	1	1
F	0	0	1	0	1	1	1	1	0	1	0	1	1	0	1	0	1	1	1	0	1
G	0	0	1	1	0	0	1	1	1	1	0	1	0	1	1	0	1	0	1	1	1
H	0	0	1	1	0	0	1	1	1	1	0	1	1	0	0	1	1	0	1	1	1
I	0	0	1	1	0	1	1	0	1	1	0	1	1	0	1	1	0	0	1	1	1
J	0	0	1	1	0	1	1	1	0	1	0	1	1	0	1	1	0	1	1	0	1
K	0	0	1	1	0	1	1	1	0	0	1	1	1	0	1	1	0	0	1	1	1
L	0	0	1	1	0	1	1	1	0	1	1	0	0	1	1	0	1	1	1	0	1
M	0	0	1	1	1	0	1	0	1	1	0	1	1	0	1	1	0	1	0	1	1
N	0	0	1	1	1	0	0	1	1	1	0	1	1	0	1	0	1	1	0	1	1
O	0	0	1	1	1	0	0	1	1	1	1	0	0	1	1	1	0	0	1	1	1
P	0	0	1	1	1	0	1	0	1	1	1	0	0	1	1	1	0	1	0	1	1
Q	0	1	0	1	0	1	0	1	1	1	1	0	1	0	1	0	1	0	1	1	1
R	0	1	0	1	1	0	1	0	1	1	1	0	1	0	1	1	0	1	0	1	1
S	0	1	0	1	1	0	1	1	0	1	1	0	1	0	1	1	0	1	1	0	1

the twelve structures have enantiomers and thus have two codes to choose from. Perhaps some regularity may be found when considering bigger benzenoids that will lead to a larger number of shapes. However, when one combines inspection of Figure 4 with that of Table 1 an apparent regularity stands out. Since leading zeros in the code correspond to a local concavity, and consecutive ones correspond to a local convexity, the shapes at the beginning of the list (having many initial zeros) indicate molecules with the periphery that is locally most concave. The structures at the end of the list similarly indicate molecular forms having the most convex shapes among those enlisted.

LARGER BENZENOIDAL SHAPES

Already for smaller benzenoids it is not easy to construct by an exhaustive trial and error approach *all* the forms having the same perimeter. With larger benzenoids the task to construct *all* possible molecular shapes of a given perimeter is becoming tedious, unless some algorithmic approach is developed for that purpose. Algorithms that describe construction of all benzenoids having the same number of ring were reported.¹¹ Schemes have been outlined for construction of all nonexistent, so called elusive, benzenoids which have no Kekulé valence structure.²⁸ Construction of all benzenoids with a given number of Kekulé valence structures has also been outlined.²⁹ However, construction of all benzenoids having the same perimeter apparently has not yet been considered. Brak, Enting, and Guttmann³⁰ considered construction of all forms having the same perimeter but not differentiating relative orientation of such structures. We are interested in planar shapes derived by fusion of polihex (benzene) rings such that we differentiate a shape and its mirror image if they could not be superimposed by sliding the shapes within the plane of embedding. Hence our shapes can occur in two forms, "left" and "right".

In Figure 5 we depicted all achiral shapes having perimeter $P = 22$. In addition to the 19 shapes shown in Figure 5 there are 28 2-D chiral shapes that we discuss at more length elsewhere.³¹ In Table 2 we give the canonical (minimal) perimeter codes for all achiral benzenoids of Figure 5. The shapes are ordered so that the first shape has the smallest binary code, while the last shape has the largest canonical (minimal) codes.

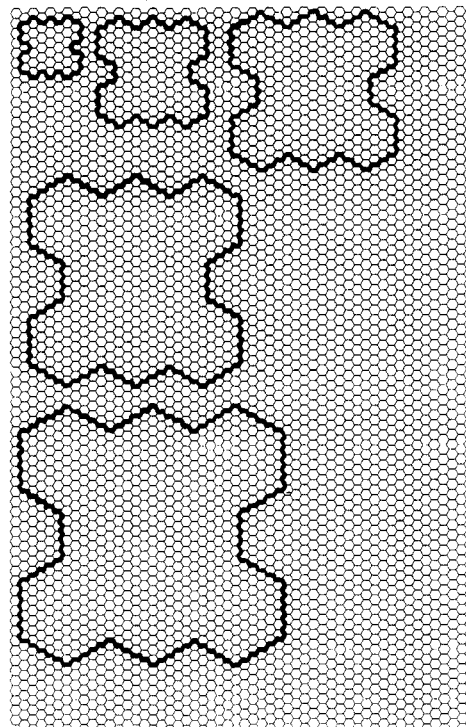
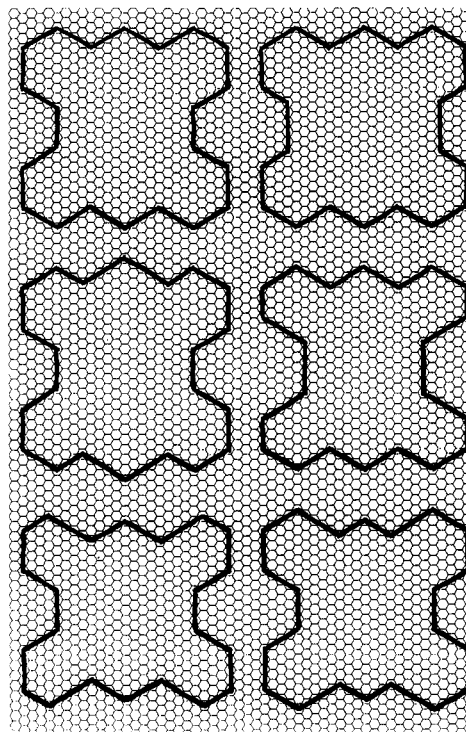
**Figure 6.** Shape "P" from Figure 5 shown at increasing resolution, i.e., at increasing shape/lattice size ratio.**Figure 7.** Six variations of shape "P" from Figure 5; higher resolution permits coding of minor shape variations, shown as unequal edge lengths of the six shapes.

Figure 5 as well as Table 2 can be viewed as part of a hexagonal shape catalog. They list for the given perimeter all possible achiral shapes obtained by fusing regular hexagons. Clearly, when one increases the perimeter the number of the hexagonal shapes will increase, while the "resolution" power of the codes will be improved. In Figure 6 we illustrate this on the shape "P" of Figure 5 that is shown at an increasing resolution. The resolution power relates to the increased size of the shape relative to the size of the

Table 3. Similarity/Dissimilarity Matrix of the 19 Achiral Shapes of Figure 5^a

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
A	0	8	6	6	4	8	6	8	4	8	8	8	4	8	8	8	8	6	4
B		0	8	8	8	8	6	8	8	6	8	6	6	6	8	8	8	8	8
C			0	8	6	8	6	6	8	6	6	8	6	4	8	8	8	6	4
D				0	4	4	6	8	8	6	6	6	8	6	8	8	8	10	8
E					0	6	8	8	6	8	6	8	8	6	6	8	8	8	6
F						0	4	8	8	4	8	8	8	6	8	8	8	8	10
G							0	4	6	6	8	6	6	6	6	6	8	4	6
H								0	4	4	6	8	6	4	8	8	8	8	10
I									0	6	4	10	4	6	8	8	8	8	8
J										0	6	4	6	6	8	6	6	4	8
K											0	8	8	6	8	8	8	8	8
L												0	8	10	8	8	8	8	8
M													0	4	8	4	6	4	4
N														0	8	8	8	8	8
O															0	4	8	8	8
P																0	4	4	8
Q																	0	4	8
R																		0	4
S																			0

^a The codes of two shapes (two cyclic binary sequences) slide one against the other. The Hamming distance counts the number of different components, i.e., of noncoincidental sites in the code. The larger the distance, the less similar are the codes. The matrix elements represent the minimal Hamming distance of each corresponding shape-pair. The minimal (4) and maximal (10) distances are emphasized.

underlying grid used as template. Because of the limitations governed by the hexagonal geometry of the plane tessellation we can increase the size of the shape at discrete steps. Nevertheless, we can view the different representations of Figure 6 as essentially the same shape viewed at different resolution. With the increasing resolution we do not see structural novelty of the shape-objects as the size has been increased. The "novelty" is that with increased resolution we can differentiate shapes that differ little from the one considered. Hence, numerically the similarity measure between two shapes, such as P and say S, need not change, at least not dramatically, with the increased resolution. What will change is that there will be many "in-between" shapes. At the lower resolution, there was only one obvious case between P and S (shape M, Figure 5), but now one can have dozens, hundreds, and even thousands of shapes "in-between," depending on the resolution power, i.e., the scale of the hexagons. In addition, with an increase in the resolution we can consider minor modifications of a given shape as illustrated in Figure 7 where the length of the edges appearing in the shape "P" are no longer equal.

We have drawn Figures 6 and 7 in two versions. The first version is the standard form in which peripheries are outlined by emphasizing edges at the boundary of a shape. In the other version the small "wiggles" caused by zig-zag movements are "overlooked" by drawing boundaries through the centers of the hexagons on the boundary. In this way we can see how the approach of hexagons will approximate general shapes, such as that of Figure 1.

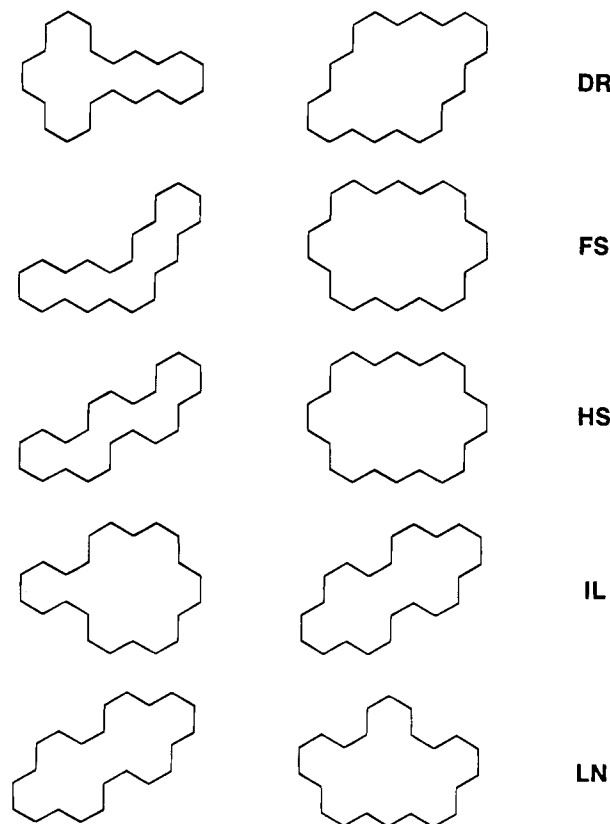
SIMILARITY OF SHAPES

Different molecular characterizations will reflect different aspects of the similarity among objects. We are interested to find how the similarity is reflected when objects are characterized by their binary periphery codes. We will consider the 19 achiral shapes of Figure 5 having periphery $P = 22$. We will use Hamming distance as the **index of**

Table 4. Similarity/Dissimilarity Matrix of the 19 Achiral Shapes of Figure 5^a

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
A	0	12	10	10	8	12	12	12	8	12	13	13	8	12	13	13	13	11	7
B		0	11	12	13	13	12	10	12	12	11	12	10	9	11	12	13	13	13
C			0	13	10	12	10	10	13	11	10	14	10	8	13	12	12	11	8
D				0	5	8	12	12	10	10	10	13	11	14	13	12	14	12	12
E					0	10	12	12	10	13	9	13	12	12	11	13	13	14	12
F						0	5	10	12	6	12	11	11	11	12	11	10	9	13
G							0	5	10	9	12	11	10	10	10	10	11	6	10
H								0	8	6	10	9	10	8	12	13	12	11	15
I									0	9	5	14	6	10	10	11	12	12	10
J										0	9	5	9	9	13	9	10	5	11
K											0	12	10	9	11	11	13	14	14
L												0	14	14	12	12	12	10	12
M													0	6	11	5	10	6	5
N														0	10	11	12	12	10
O															0	6	12	12	14
P																0	8	6	10
Q																	0	8	12
R																		0	6
S																			0

^a The Hamming distance of the codes is calculated by comparing pairs of consecutive digits in the two codes and counting noncoincidences. The minimal (5) and maximal (15) distances are emphasized.

**Figure 8.** The five "most dissimilar" (Hamming distance 10) pairs of shapes out of 171 possible pairs of shapes A to S.

similarity/dissimilarity. Hamming distance³² is the standard parameter that counts the coincidences between binary sequences (of equal length). The larger the distance between two codes the less similar are the codes and the less similar are the corresponding shapes. The Hamming distances shown in Tables 3 and 4 are obtained not by comparing the pairs of canonical codes but by shifting the codes one against the other so as to produce the maximal overlap.

In Table 3 we show the similarity/dissimilarity matrix for the 19 achiral shapes of Figure 5 labeled as A–S. First we observe relatively small variations in the magnitude of the

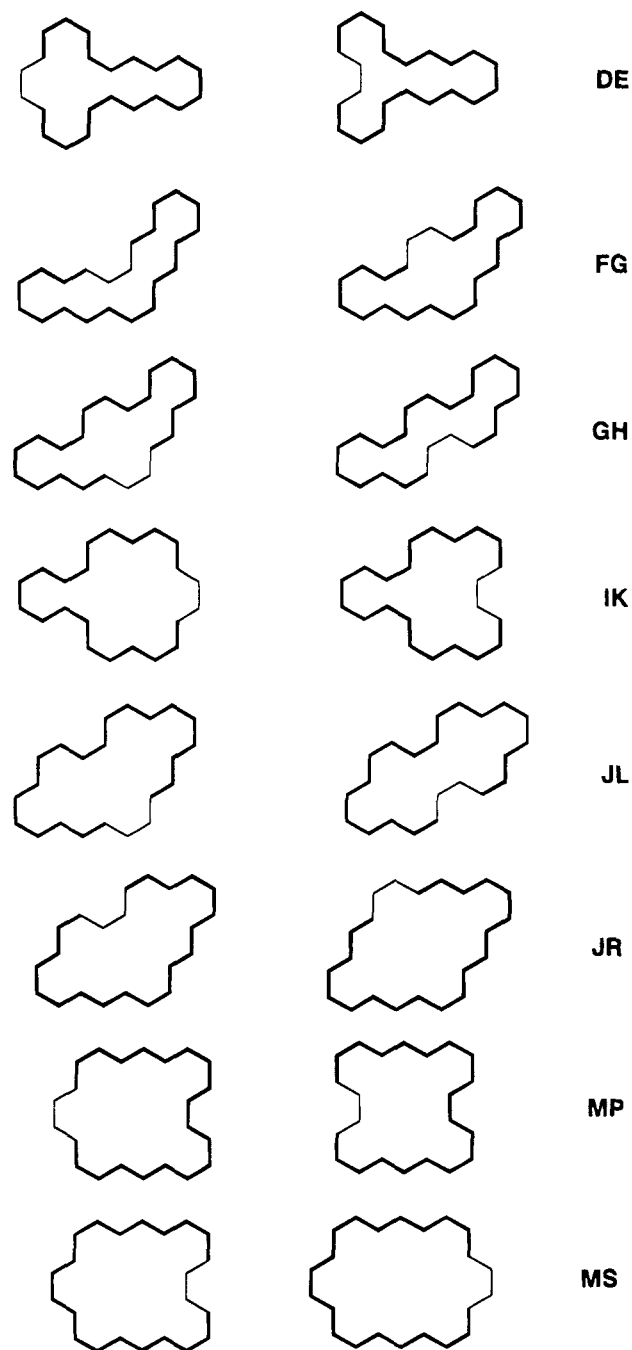


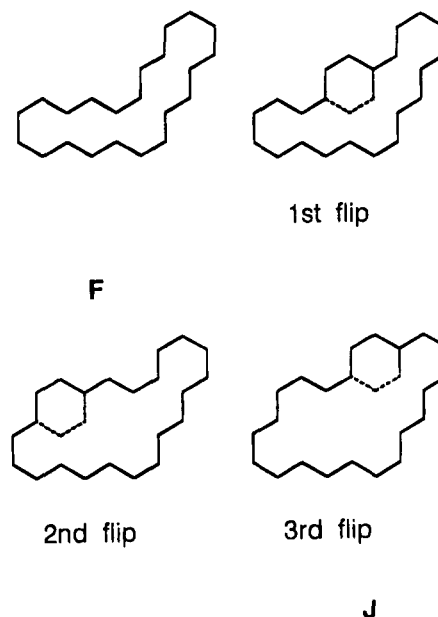
Figure 9. The first eight pairs from the group of 28 "most similar" (Hamming distance 4) pairs of shapes; common features can be identified by inspection—"single flip" similarity.

similarity indices in Table 3. The most similar shapes are associated with the similarity distance of 4, while the most dissimilar shapes are at the distance 10. Moreover, while there are only five pairs of shapes (from 171 possibilities) that are "the most dissimilar" (shown in Figure 8), there are 28 pairs of "the most similar" shapes, depicted in Figures 9–11. Inspection of Figures 8–11 is instructive. We see that, indeed, "the least similar" shapes do agree with our common perceptions of dissimilarity, by being rather distinct. These pairs of "the least similar" shapes show limited similarities in their periphery. The common parts of their codes vary from six to nine consecutive digits (or from seven to ten overlapping edges out of 22 when the corresponding figures are optimally superimposed). Indeed, the "least similar" shapes have very little in common.

Of more interest is to examine "the most similar" shapes. We have grouped the "most similar" shapes into three groups, shown in Figures 9–11. In the first group we included the shapes for which some common features can be identified by inspection. In the second and the third group we show structures that are similar for less apparent reasons. For the similar shapes of Figure 9 in all cases we could align their common periphery so that it coincides with all but three edges (or four digits in the code). The parts in which two shapes differ can be brought into full coincidence by an operation "flip", which consists in "flipping over" the three bonds from left to right or vice versa, using the adjacent common bonds as the rotation axis. Hence, the similarity based on the binary periphery code for the shapes of Figure 9 can be deduced easily by inspection. The index of "common periphery" for such shapes is 19/22, i.e., 19 out of 22 edges can be brought in coincidence without "flipping".

Shapes of Figure 10 have also been found numerically (using the Hamming distance measure) as "the most similar", yet it is less apparent why this is the case. Is this an artifact of the particular approach? Before trying to answer this question let us examine some of the cases illustrated in Figure 10 more closely. Similarity between pairs (F, J), (G, R), and (H, J) can be related to a "composite flip" in which first one replaces a larger concave region with a convex one and then follows with a smaller flip that indents the convex region by forming a smaller concave region. Thus from F by a "large" flip we first obtain R and then from R by a "small" flip we obtain J.

Alternatively we may view the high similarity between F and J as a result of three simple flips as illustrated below:



The same mechanism of "composite flips" applies to several other pairs from Figure 10, like (I, M) and (M, N). In considering such an operation sometimes the order in which various "local" flips are applied may be important.

Similarity among shapes in Figure 11 is even less apparent, while a visual inspection of various pairs of shapes may be misleading. Thus the pair (A, I) may appear interesting as it reminds one of "the lock and the key" model of Emil Fischer.³ Here the "inside" part of one shape corresponds to the "outside" part of the other shape. However, when

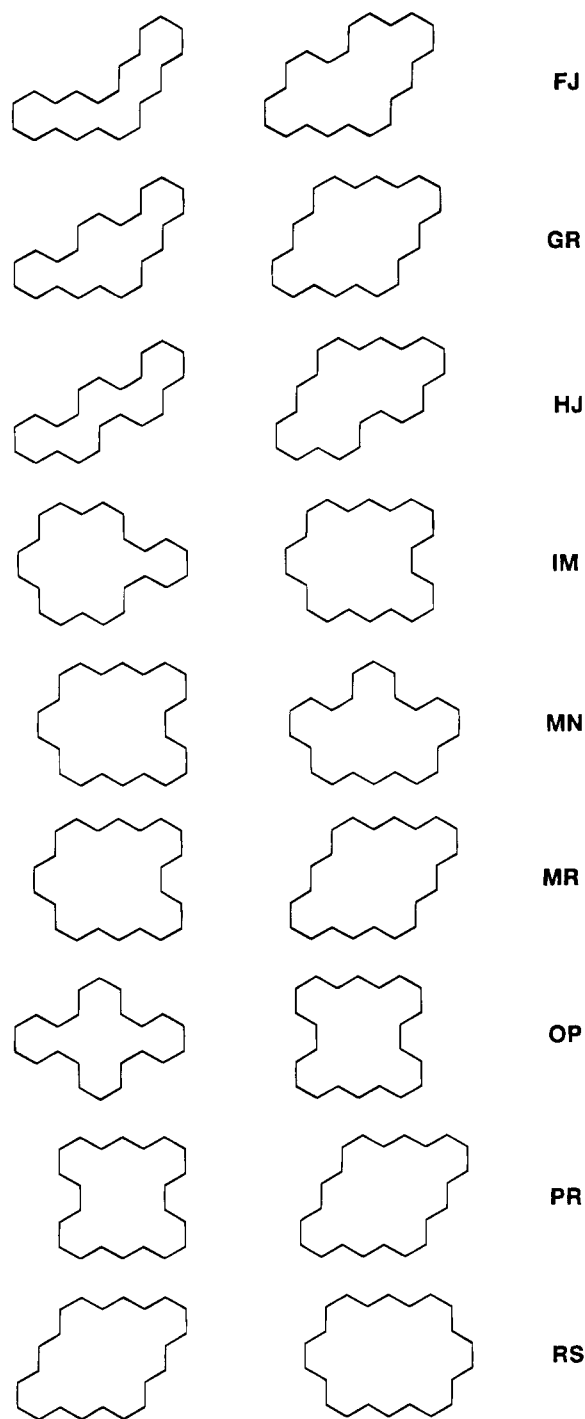
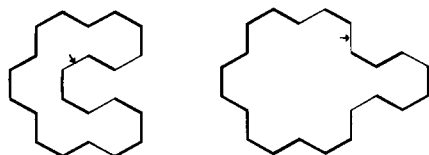


Figure 10. The nine pairs from the group of 28 "most similar" pairs of shapes; similarities are less apparent—"composite flip" similarity.

we compare the corresponding codes we see that the similarity of the two shapes is mostly due to the parts of the code that describe a portion of "coronene" periphery present in both shapes as illustrated below:



The edges that coincide in their code are emphasized. As we see there is little overlap of the codes for the part of the

two shapes that correspond to "the lock and the key" model. We indicated by an arrow the CC bonds in the two structures, the alignment of which gives the maximum overlap of the two codes. The mismatch of the lock and the key parts of the code of two shapes is a consequence of the nature of the codes: The lock (concave) region of molecule A is coded by four zeros in a row, while the key (convex) region in I is described by four ones in a row. Thus, of course, there is a correspondence, but Hamming distance is not a useful measure to perceive it. If one is interested in such complimentary regions local matching of zeros and ones should be sought, which can be achieved by first converting a code to its complement and then searching for the overlap of the codes. The complement of a code is defined by replacing zeros by ones and vice versa.

There are several pairs of structures in Figure 10, such as (O, P), (P, R), and finally (R, S), viewing of which does not suggest an apparent reason for observed high similarity. This is even more pronounced for some structures shown in Figure 11. Pairs of structures like (A, E), (D, F), (H, I), (H, N), and particularly (P, Q) look more like the least similar than the most similar! For the most similar pairs of shapes of Figure 9 one can visually find the overlapping segments of the code, because the overlapping segments are in one piece. This is no longer the case with structures shown in Figures 10 and 11. It is therefore important to find not only pairs of structures for which suitably shifted codes differ as little as possible but also for which an uninterrupted fragment of the maximal length of the code occurs in two codes that are compared. The length of the uninterrupted fragment of the code is therefore an important parameter for establishing truly the most similar shapes.

The considered measure of similarity does not reflect whether high coincidence is due to a contiguous portion of a code or results from accumulation of several short-length local coincidences among disjoint fragments of the periphery. A way to discriminate among the most similar shapes of Figure 9 and those of Figures 10 and 11, as will be seen in the next section, is to compare longer "windows" of the codes.

SIMILARITY BASED ON COMPARISON OF LONGER PERIPHERY FRAGMENTS

In order to arrive at a more selective measure of similarity that will reflect the shape measure based on the overlapping of consecutive edges we have extended the comparison of codes by considering **pairs of consecutive digits** in two codes and counting their coincidences. In order to find an optimal overlap one "slides" the codes one against the other by a single site at a time, to the right or to the left. While counting the number of instances at which digits differ in two codes was not particularly tedious (though it is prone to errors), the count of the binary coincidences in groups of two and more digits is not only error prone but also not practical by hand. The computer program used for finding similarities based on the comparison of single entries in two codes has been extended to handle any desired number of consecutive places in the code. From the comparison of codes we again construct the similarity/dissimilarity table (Table 4). The new table shows a greater discrimination among the shapes. The range of similarity/dissimilarity values is now between 5 and 15. Of the 28 pairs of "the

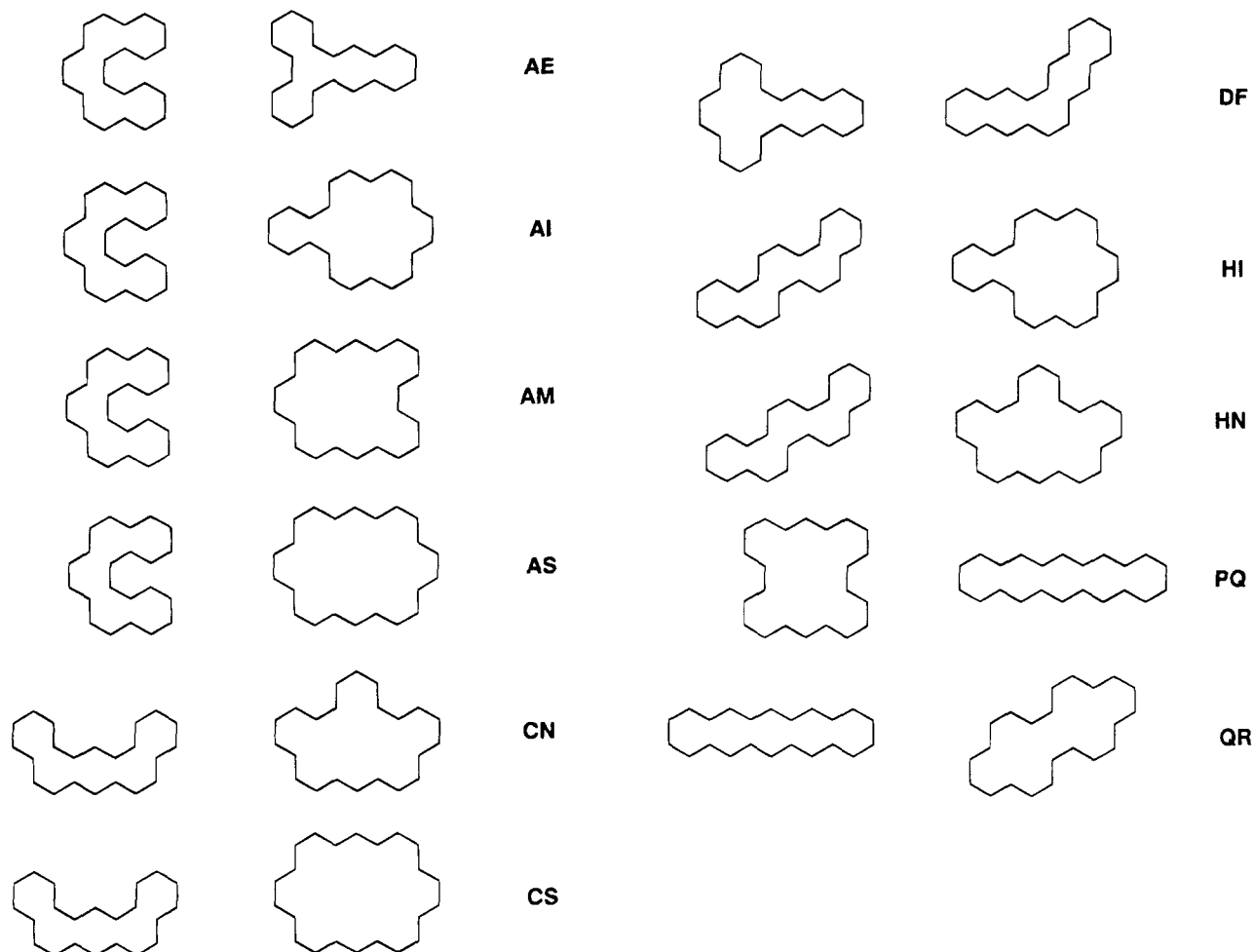


Figure 11. The 11 pairs from the group of 28 “most similar” pairs of shapes; similarities are even less apparent since the overlapping code is present in several fragments.

most similar” shapes when the comparison was based on matching of single entries of two codes (Table 3) only eight still remain the most similar shapes. Similarly, instead of five of the “most dissimilar” cases now we find only a single such case (pair H, S). The most similar shapes are the shapes already shown in Figure 9, that is, the shapes that differ one from another by one “flip” (or by one local exchange of 1 0 1 0 into 0 1 0 1). The next most similar cases are the shapes shown in Figure 10, which now differ by six entries in their codes. The shapes of Figure 11, that differ by seven and eight entries in their codes, show even more reduced similarity.

This is gratifying, since as we have seen with this extension of the length of the fragment (“window”) for the comparison of codes we can discriminate cases that appeared more similar from comparisons based on single entries in the code (Figures 10 and 11). One can extend the outlined similarity/dissimilarity tests by considering coincidences among longer fragments of the code. Elsewhere we discuss the results of such comparisons in more detail.³³ It suffices here to state that the shapes that differ by a single “flip” (Figure 9) stay equally similar, when longer “windows” of comparisons are used. Use of longer entry fragments or “windows” helps one to differentiate among the less similar shapes (including also those Figures 10 and 11). By comparing longer “windows” in the codes one can better discriminate less and less similar structures that appeared initially as similar.

We can understand why when the “window” of comparison is increased by one, the similarity/dissimilarity measure determined by the overlapping of the codes increases by one. This happens for the shapes of Figure 9 that differ in Table 3 by four bonds and in Table 4 by five bonds. If the maximal overlap between two codes is due to two fragments, the dissimilarity measure will increase by at least two, and possibly more, which will depend on the separation between the disjoint overlapping regions in the codes.

CONCLUDING REMARKS

We have suggested use of simple binary codes for a characterization of the molecular boundaries or molecular shapes. In particular, we considered smaller benzenoid systems that can serve as a model for more general planar structure that can be represented as a contour on a hexagonal tessellation of a plane. By an exhaustive construction of all legitimate binary codes possible one can derive a listing of all possible shapes of a given perimeter P . We have seen also that the derived binary codes reflect the symmetry (and chirality) of the embedded shapes and offer a source of structural shape invariants.

Similarity based on the introduced shape codes has been examined, and apparently similar shapes show smaller variations in their binary codes. This was particularly evident when the “window” of comparison was extended to longer lengths. Dissimilarity among enantiomers, which have

different codes corresponding to circumscribing the same shape in opposite direction, can serve as a measure of the "degree of chirality".³⁴

The approach raises a number of questions that remain open. These include considerations of extending the present approach to objects in three-dimensional space. Nevertheless, we have opened a novel route to the quantitative characterization of molecular shapes. Previous works, particularly the pioneering studies of Mezey and Arteca¹⁶ and the work on 3-D topological approach to QS(hape)AR of Mezey,³⁵ were mainly concerned with topological aspects of molecular shape. In contrast, our work is mainly concerned with combinatorial aspects of shapes, extending graph theoretical concepts to objects composed of fused hexagons embedded in regular hexagonal grid. Extensions to other types of lattices are possible. For example, if one uses Cartesian two-dimensional square lattice, then at each grid point there are three possibilities to continue to move: to go to the left, to go to the right and to go straight. The same is true for "walking" on the diamond grid in three-dimensional space, where "straight" is to be interpreted as *trans* (relative to previous two steps), while the left and the right are two alternative *gauche* directions. Contours on such grids can be encoded by using triple labels, such as +1, -1, and 0. Hence the corresponding codes will appear as ternary codes. The formal equivalence of Cartesian planar grid and three-dimensional diamond grid allows one to "translate" spatial curves (embedded on the diamond grid) into planar "equivalents". This may be of interest for characterization of knots or conformations of long chains (such as *n*-alkanes or various polymers).

It is premature to address all these, and other, topics at this moment although we anticipate expansion of our work in that direction. In this we report on a rather modest goal: To characterize planar shapes based on hexagonal cells. Skeptics, as well as those who wait for "ready-to-use" packages for analysis of shapes, may be disappointed, since a number of questions even pertaining to planar structures are unanswered. How are we to treat heteroatoms? Are there "basic" shapes that would allow construction of other shapes as "composite"? Why not use distance-based matrices instead of the binary matrices for codes? Why not use Laplacian matrices and their eigenvalues? Is Hamming distance the most suited similarity measure for discrimination of shapes? Are the binary shape codes sufficiently general? Can they be extended to three-dimensional structures? And so on.

We cannot answer these questions since they have not yet been considered. We will, however, consider only one of the questions here in a brief outline. This is the last question above, which is highly important. In a way probably this is the most important question among those mentioned to be resolved. It is well-known in graph theory that generalizations from trees to cyclic graphs and from planar to nonplanar graphs are sometimes possible but more often not. It may appear that a generalization from characterization of planar contours to characterization of surfaces in a three-dimensional space is not possible. There have been, however, several methods already published which treat the problem of 3-D codes. For more details see original references cited in the book of Mezey.¹⁷

In order to extend our approach to three-dimensional objects, we could first derive for a 3-D surface a set of 2-D

contours. Such are, for example, the electrostatic maps that have found use in QSAR in the well-known CoMFA method.³⁶ Once we have the contours, each contour can be encoded by the binary code as illustrated in this paper. Lists of such contours, with specifications as to their relative positions, will lead to a three-dimensional code for molecular shapes. In this way while the contours of 2-D objects will be coded as binary sequences, a surface of a 3-D object will be represented by a sequence of codes that can be organized in a binary matrix. Comparisons of 3-D shapes will be now more involved, more cumbersome to compute. The results may depend to a degree on the relative orientation of the planes that bisect the shape and generate the corresponding contours. With an increase in the resolution of the model (i.e., by a decrease of the relative size of the hexagonal cells and the object) a comparison of two objects will be less and less dependent of the relative orientations of the bisecting planes that generate shape contours, but will be more and more tedious, possibly becoming computationally impractical.

We anticipate that practical detailed comparisons of 3-D objects are not likely to follow from such sets of surface codes. They are more likely to be based on a selection of suitable invariants that can be extracted from such codes. Then the calculations will be relatively simple even if a set of invariants needs not fully characterize a structure. Preliminary results based on use of distance-based structural invariants are very encouraging.³⁷ They show that indeed one can directly extend distance-based characterizations of molecular contours to distance-based characterizations of 3-D objects.³⁸

These and other such questions may provoke new directions yet to be considered, but at this initial stage one should not forget the anecdote of Michael Faraday, who at the end of his lecture about the discovery of the electromagnetic induction was asked by someone in the audience: "Sir, for what good is your new discovery?" Michael Faraday answered: "For what good is a new born baby?"

ACKNOWLEDGMENT

The financial support of the Ministry of Science and Technology of the Republic of Slovenia is gratefully acknowledged. The authors appreciate the comments of Dr. Ch. Rücker and P. G. Mezey that led to a significant improvement of the presentation.

REFERENCES AND NOTES

- (1) Randić, M. Chemical Structure—What is "she"? *J. Chem. Educ.* **1992**, 69, 713.
- (2) Prelog, V.; Ruzicka, L. *Helv. Chim. Acta* **1994**, 27, 66.
- (3) Fischer, E. The Influence of Configuration on the Activity of Enzyme. *Ber. Deut. Chem. Ges.* **1894**, 27, 2985–2993.
- (4) Hansch, C. A quantitative approach to biochemical structure—activity relationships. *Acc. Chem. Res.* **1969**, 2, 232–239. See, also: *QSAR in Drug Design and Toxicology*; Hadži, D., Jerman-Blažič, B., Eds.; Elsevier: Amsterdam, 1987.
- (5) Randić, M. Design of molecules with desired properties. A molecular similarity approach to property optimization. In *Concepts and Applications of Molecular Similarity*; Johnson, M. A., Maggiora, G. M., Eds.; Wiley: New York, 1990. Randić, M. Similarity methods of interest in Chemistry. In *Mathematical methods in Contemporary Chemistry*; Kutchanov, S. I., Ed.; Gordon & Breach Publ.: Langhorne, PA, in press.
- (6) Kasinos, N.; Lilley, G. A.; Subbarao, N.; Haneef, I. A robust and efficient automated docking for molecular recognition. *Protein Eng.* **1992**, 5, 69–75.
- (7) Connolly, M. L. *Biopolymers* **1986**, 25, 1229–1247.

- (8) Harary, F. *Graph Theory*; Addison-Wesley: Reading, MA, 1969.
- (9) Trinajstić, N. *Chemical Graph Theory*, 2nd ed.; CRC Press, Boca Raton, FL, 1992.
- (10) Knop, J. V.; Müller, W. R.; Jeričević, Z.; Trinajstić, N. Computer enumeration and generation of trees and rooted trees. *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 91–99. Dias, J. R.; Enumeration, theoretical properties, and other aspects of helical aromatic hydrocarbons. *J. Mol. Struct. (Theochem)* **1993**, *28*, 11–22. Herndon, W. C. *J. Am. Chem. Soc.* **1990**, *112*, 4546. Cyvin, B. N.; Guo, X.; Cyvin, S. J.; Zhang, F. Enumeration of helicenes. *Chem. Phys. Lett.* **1992**, *188*, 537–542.
- (11) Knop, J. V.; Müller, W. R.; Jeričević, Z.; Trinajstić, N. *Computer Generation of Certain Classes of Molecules*; SKHT/Kemija u industriji; Zagreb, Croatia, 1985. Müller, W. R.; Szymanski, K.; Knop, J. V.; Nikolić, S.; Trinajstić, N. On the enumeration and generation of polyhex hydrocarbons. *J. Comput. Chem.* **1990**, *11*, 223–235. Jerman-Blazić Džonova, B.; Trinajstić, N. Computer-aided enumeration and generation of the Kekule structures in conjugated hydrocarbons. *Comput. Chem.* **1982**, *6*, 121–132.
- (12) Zupan, J. *Algorithms for Chemists*; Wiley: New York, 1989.
- (13) Sabljic, A.; Trinajstić, N. Quantitative structure–activity relationships: The role of topological indices. *Acta Pharm. Jugosl.* **1981**, *31*, 189–214. Trinajstić, N.; Randić, M.; Klein, D. J. On the quantitative structure–activity relationship in drug research. *Acta Pharm. Jugosl.* **1986**, *36*, 267–279. Romanowska, K. The application of the graph-theoretical method in the QSAR scheme: Possibilities and limits. *Int. J. Quant. Chem.* **1992**, *43*, 175–195.
- (14) Randić, M. Graphical enumeration of conformations of chains. *Int. J. Quant. Chem.: Quant. Chem. Symp.* **1980**, *7*, 187–197.
- (15) Randić, M. Molecular topographic descriptors. *Studies Phys. Theor. Chem.* **1988**, *54*, 101–108. Randić, M. On characterization of three-dimensional structures. *Int. J. Quant. Chem.: Quant. Biol. Symp.* **1988**, *15*, 201–208. Randić, M.; Jerman-Blazić, B.; Trinajstić, N. Development of three-dimensional molecular descriptors. *Comput. Chem.* **1990**, *14*, 237–246. Mihalić, Z.; Trinajstić, N. The algebraic modelling of chemical structures: On the development of three-dimensional molecular descriptor. *J. Mol. Struct. (Theochem)* **1991**, *232*, 65–78. Bogdanov, B.; Nikolić, S.; Trinajstić, N. On the three-dimensional Wiener number. A comment. *J. Math. Chem.* **1990**, *5*, 305–306. Balasubramanian, K. Geometry dependent characteristic polynomial of molecular structures. *Chem. Phys. Lett.* **1990**, *169*, 224–230. Ivanov, J.; Karabunarliev, St.; Mekenyan, O. 3DGEN: A system for exhaustive 3D molecular design proceeding from molecular topology. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 234–243. Randić, M.; Kleiner, A. F.; DeAlba, L. M. Distance/distance matrices. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 277–286. Pogliani, L. On a graph theoretical characterization of cis/trans isomers. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 801–804.
- (16) Mezey, P. G. Molecular Surfaces, Chapter 7. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publ., Inc.: New York, 1990. Arteca, G.; Mezey, P. G. Shape characterization of some molecular model surfaces. *J. Comput. Chem.* **1988**, *9*, 554–563. Arteca, G.; Mezey, P. G. Molecular conformation and molecular shape: A discrete characterization of continua of van der Waals Surfaces. *Int. J. Quant. Chem.* **1988**, *34*, 517–526. Walker, P. D.; Mezey, P. G. Representation of square-cell configurations in the complex plane: Tools for the characterization molecular monolayers and cross section of molecular surfaces. *Int. J. Quantum Chem.* **1992**, *43*, 375–392. Mezey, P. G. Iterated similarity sequences and shape ID numbers for molecules. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 244–247.
- (17) Mezey, P. G. Shape in Chemistry: Introduction to Molecular Shape and Topology; VCH: 1993.
- (18) Razinger, M.; Novič, M. Reduction of the information space for data collections. Chapter 5. In *PC for Chemists Data Handling in Science and Technology 5*; Zupan, J., Ed.; Elsevier: Amsterdam, 1990; pp 89–103. Leicester, S.; Finney, J.; Bywater, R. A quantitative representation of molecular surface shape. I: Theory and development of the method. *J. Math. Chem.* **1994**, *16*, 315–341. Leicester, S.; Finney, J.; Bywater, R. A quantitative representation of molecular surface shape. II: Protein classification using Fourier shape descriptors and classical scaling. *J. Math. Chem.* **1994**, *16*, 343–365.
- (19) Rouse Ball, W. W. *Mathematical Recreations and Essays*; Macmillan: New York, 1967; 5th printing, p 262.
- (20) Read, R. C. The coding of various kinds of unlabelled trees. In: *Graph theory and computing*; Read, R. C., Ed.; Academic Press: New York, 1972; p 153.
- (21) Balaban, A. T. Chemical graphs. Part 12. Configuration of annulenes. *Tetrahedron* **1971**, *27*, 6115–6131.
- (22) Walker, P. D.; Mezey, P. G. Representation of square-cell configurations in the complex plane: Tools for characterization of molecular monolayers and cross sections of molecular surfaces. *Int. J. Quant. Chem.* **1992**, *43*, 375–392.
- (23) Randić, M. Conjugated circuits and resonance energies of benzenoid hydrocarbons. *Chem. Phys. Lett.* **1976**, *38*, 68–70. Randić, M. Aromaticity and conjugation. *J. Am. Chem. Soc.* **1977**, *99*, 444–450. Randić, M. A graph theoretical approach to conjugation and resonance energies of hydrocarbons. *Tetrahedron* **1977**, *33*, 1905–1920. Randić, M. Resonance energies in very large benzenoid hydrocarbons. *Int. J. Quant. Chem.* **1980**, *17*, 549–586.
- (24) Herndon, W. C. Resonance energies of aromatic hydrocarbons. A quantitative test of resonance theory. *J. Am. Chem. Soc.* **1973**, *95*, 2404–2406. Herndon, W. C.; Ellzey, M. L., Jr. Resonance theory V. Resonance energies of benzenoid and non benzenoid π systems. *J. Am. Chem. Soc.* **1973**, *95*, 2404–2406. Simpson, W. T. Internal dispersion forces. *J. Am. Chem. Soc.* **1951**, *73*, 5363–5367. Simpson, W. T. Resonance force theory of carotenoid pigments. *J. Am. Chem. Soc.* **1955**, *77*, 616–6168.
- (25) Klein, D. J.; Trinajstić, N. Foundations of conjugated circuits models. *Pure Appl. Chem.* **1989**, *61*, 2107. Nikolic, S.; Trinajstić, N.; Klein, D. J. The conjugated circuit model. *Comput. Chem.* **1990**, *14*, 313–322.
- (26) Lord Kelvin, *Baltimore lectures on molecular dynamics and the wave theory of light*; C. J. Clay and Sons: London, 1904, pp 439, 619.
- (27) Buda, A. B.; Mislow, K. On geometric measures of chirality. *J. Mol. Struct. (Theochem)* **1991**, *232*, 1–12.
- (28) Hosoya, H. How to design non-Kekulé polyhex graph? *Croat Chem. Acta* **1986**, *59*, 583–590. Guo, X. F.; Zhang, F. J. A construction method for concealed non-Kekulé benzenoid systems with $h = 12$, 13 rings. **1989**, MATCH, *24*, 85. Gutman, I. Some topological properties of benzenoid systems. *Croat. Chem. Acta* **1974**, *46*, 209–215. Balaban, A. T. Challenging problems involving benzenoid polycyclics and related systems. *Pure Appl. Chem.* **1982**, *54*, 1075. Cyvin, S. J. Search for concealed non-Kekulé benzenoids and coronoids. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 236.
- (29) Randić, M.; Klein, D. J. Kekulé valence structures revisited. Innate degrees of freedom of π -electron couplings. In *Mathematics and Computational Concepts in Chemistry*; Trinajstić, N., Ed.; Ellis Horwood Publ.: Chichester, England, 1986.
- (30) Brak, R.; Guttmann, A. J.; Enting, I. G. Exact solution of the row-convex polygon perimeter generating function. *J. Phys. A: Math. Gen.* **1990**, *23*, 2319–2326.
- (31) Randić, M.; Razinger, M. Molecular shape and chirality of planar molecules. *New J. Chem.*, to be submitted for publication.
- (32) Hamming, R. W. Error detecting and error correcting codes. *Bell Syst. Tech. J.* **1950**, *29*, 147–160.
- (33) Randić, M.; Razinger, M. Reported at MATH/CHEM/COMP 1994; Dubrovnik, Croatia, June 26–July 2, 1994.
- (34) Randić, M. Molecular Profiles—Novel geometry dependent molecular descriptors. *New J. Chem.*, in press. Randić, M. Molecular Shape profiles. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 373–382. Randić, M. Distance dependent labels. *J. Chem. Inf. Comput. Sci.*, submitted for publication. Randić, M. Molecular bond profiles. *J. Math. Chem.*, submitted for publication. Randić, M. Characterization of conformations of eight membered rings. *Chem. Phys. Lett.*, to be submitted for publication. Randić, M. Characterization of conformations of nine membered rings. *Int. J. Quant. Chem.: Quant. Biol. Symp.*, in press.
- (35) Mezey, P. G. Shape-similarity measures for molecular bodies: A three-dimensional topological approach to quantitative shape-activity relations. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 650–656.
- (36) Cramer III, R. D.; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on bonding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.

CI940088Y