Some Notes on Randić−Razinger's Approach to Characterization of Molecular Shapes

Nikolai S. Zefirov* and Serge S. Tratch

Chemical Department, Moscow State University, Moscow 119899, Russia

Received March 27, 1997[⊗]

Several problems related to fundamentals of the recent Randić−Razinger's description of molecular shapes are briefly outlined. An alternative "geometry based" definition of molecular periphery and problems of degeneracy for molecular shape profiles are discussed; two examples of degenerate shape profiles are explicitly presented. It is shown that Randić−Razinger's periphery codes of polybenzenoids may be considered as characteristics of those ring 2D configurations which can be properly embedded on a regular hexagonal lattice.

In a series of recent publications,[1−4] M. Randić and M. Razinger have outlined the new approach which makes it possible to uniquely characterize the shapes of planar polybenzenoids and to discuss some of their specific properties such as similarity,[2,3] symmetry,[3] and chirality.[4] Additionally, well-defined invariants,[1,2] ordering procedures,[2,4] and chirality measures[4] were suggested for polybenzenoidal shapes. Although being illustrated only for benzenoidal systems, Randić−Razinger's approach was declared to be also applicable to any closed curvilinear contour which can be suitably approximated by periphery of some polyhex figure. Moreover, several ideas about further extension of the approach to 3D structures were also discussed.[2,3] It should be noted that the approach under discussion represents molecular shapes as discrete objects; an alternative, "continuous" approach elaborated by P. G. Mezey[5] is based on analysis of molecular electron density surfaces.

Being pioneering, the papers[1−4] are open for criticism, and the authors have pointed (see ref 3) at several important questions which cannot be completely answered in the present stage of study. In this paper, we are not planning to hurt the "newborn baby" but want to consider only fundamentals of the Randić−Razinger's approach. That is why the ordering procedures for benzenoidal shapes as well as new measures of their similarity and chirality will not be discussed at all. Instead, we shall concentrate our attention to such fundamental questions as the following: "*How to overcome some limitations which follow from the most natural definition of molecular periphery ?*", "*In what cases the suggested invariants (i.e., topographic index*[1] *and shape profile*[2]*) cannot discriminate planar or spatial shapes ?*", and "*What mathematical objects really correspond to periphery codes introduced in ref 3?*". In order to discuss the above-mentioned problems, the main notions of the Randić−Razinger's approach will be briefly illustrated below.

In the Randić−Razinger's approach,[1−4] molecular periphery is not precisely defined but is understood in a natural way, i.e., as consisting of all "boundary" bonds which form a closed contour of any polybenzenoidal system; the multiplicities of these bonds (as well as all bonds to hydrogens and substituents[6]) are not taken into account. The periphery bonds of three simplest benzenoidal frameworks are indicated by fatty lines in Figure 1a; note that all "interior" carbon atoms (such as the central atom in phenalene framework) are neglected in all considerations related to molecular periphery and molecular shape.

The *binary periphery codes* are introduced in ref 3 as main structural characteristics of planar molecular shapes; these codes are similar to the codes suggested earlier by A. T. Balaban.[7,8] The periphery codes are constructed out of digits 0 and 1; the number of digits is equal to the length of the cyclic periphery contour. The zero (or unity) digit in the code denotes the left (or the right) direction which must be chosen at any atom when walking around the contour in the clockwise manner. It is easy to see that the codes thus constructed really depend on the choice of the starting bond. In Figure 1b, all ten possible codes for numbered naphthalene framework of Figure 1a are explicitly demonstrated. The smallest of periphery codes (marked by an asterisk in Figure 1b) is chosen to uniquely characterize the shape of the polybenzenoidal system under discussion.

The periphery codes were also demonstrated[3] to properly reflect (two-dimensional) symmetry elements of any planar polybenzenoid. Thus, the presence of the center of symmetry (or *n*-fold rotation point,[10] in other terminology) in a benzenoidal system results in the fact that corresponding periphery code consists of *n* subunits (e.g., $1011110111 \equiv (10111)^2$ for the first code of Figure 1b). Additionally, only $1/n$ of all possible codes are pairwise nonidentical in this case; for example, for naphthalene framework, $n = 2$ and hence all periphery codes of the second column in Figure 1b coincide with corresponding codes of the first column.

The other symmetry elements—reflection lines (cf. note 10)—can also be recognized from periphery codes. In the papers,[3,4] the authors state that "two-dimensionally" achiral planar polybenzenoids (which necessarily contain reflection symmetry elements such as bond 3−8 and dashed line in naphthalene framework of Figure 1a) must be characterized by palindrome codes. (The word "palindrome" means that the code is the same whether it is read from the left to the right or from the right to the left.) The codes of the fifth row in Figure 1b, corresponding to a "dashed" line bisecting bonds 5−6 and 1−10, may serve as an example. We can refine here that the absence of palindrome codes *is not sufficient* however to state that corresponding polybenzenoid is in fact chiral. For example, all 12 periphery codes for achiral phenalene framework of Figure 1a are not palindromes. In order to understand a real situation, one must

[⊗] Abstract published in *Advance ACS Abstracts,* August 15, 1997.

S0095-2338(97)00018-8 CCC: $14.00

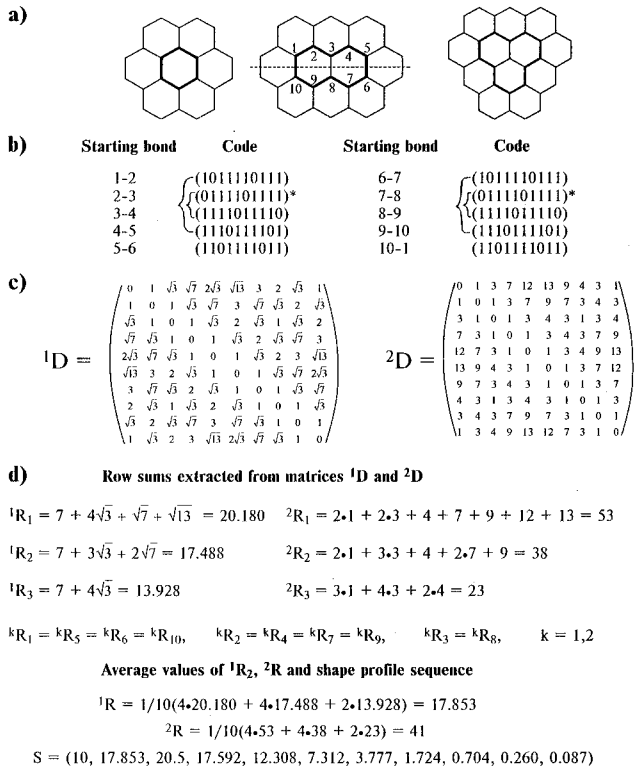

Figure 1. Illustrations to main notions of Randić−Razinger's approach: periphery bonds for benzene, naphthalene, and phenalene frameworks (a); periphery codes (b); distance matrices ^1D and ^2D (c); and row sums, averaged sums, and the resulting shape profile S (d) for naphthalene framework.

take into account that to some reflection lines "pseudopalindromic" pairs of codes can correspond; each code of such a pair is interconverted into its counterpart by reading from the right to the left. The "pseudopalindromic" pairs (related to the reflection line 3−8 of naphthalene framework) can be found in the first and fourth (or in the second and third) rows of Figure 1b.

In order to numerically characterize molecular shapes, the specific kind of distance matrices is considered in refs 1 and 2. The nondiagonal entries d_{ij}^k of matrices ^kD are kth powers, $k = 1, 2, ...$, of geometric distances for all pairs of atoms i and j being involved in the molecular shape. The matrices ^1D and ^2D corresponding to numbered naphthalene framework are explicitly shown in Figure 1c; in this example, the normal $C-C$ distances (or, in other words, the distances between adjacent nodes of the hexagonal lattice) are equated to unity. Note that for pericyclic polybenzenoids, all matrices ^kD, $k = 1, 2, ...$, *do not contain rows and columns related to "interior" atoms* (such as central atom of phenalene framework, cf. Figure 1a). This fact shows that in the general case, Randić−Razinger's distance matrices are submatrices of the "complete" (geometry) distance matrices.

The "powered" matrices, ^1D, ^2D, ^3D, ..., can obviously be used as a source of numerous geometry dependent shape invariants. One of the simplest "Wiener-like" [11] invariants, called topographic index D^2, was investigated[1] with more details. This index can be calculated by summing up all entries of the matrix ^2D; it was declared[1] to be the first topographic index with integer values for all polybenzenoids which can be properly embedded on the hexagonal lattice.

The more original and powerful shape invariant was formulated[2] as a sequence rather than unique number. In order to derive this sequence, one must initially count the row sums $^kR_i = \sum_{j=1}^{j=m} d_{ij}^k$ for all $i = 1, 2, ..., m$ and all $k = 1, 2, 3, ...$ (here are below m designates the number of bonds in the molecular periphery; note that in practice, the upper bound of k is also limited[2] to m). From the values ^kR_i, the average values $^kR = (1/m)\sum_{i=1}^{i=m} {}^kR_i$ are then calculated and finally normalized by means of division by $k!$. The resulting values $^kS = {}^kR/k!$ (together with the leading term $^0S \equiv m$ which can be formally associated with unity distance matrices $^0D = (d_{ij}^0)$) form the sequences

$$S = (^0S, {}^1S, {}^2S, ..., {}^mS)$$

which define *molecular shape profiles*. Just these sequences were used[2] for recognition of molecular similarity and partial order among polybenzenoidal shapes. The above calculation procedure is illustrated for naphthalene shape in Figure 1d. Symmetry considerations are used in deriving row sums of matrices ^1D, ^2D and corresponding averaged sums ^1R, ^2R; the values of $^kR/k!$, $k = 3, 4, ..., 10$, are taken from ref 2.

Now we summarize some standpoints of the Randić−Razinger's approach. Firstly, it is a "discrete" approach[12] in the sense that it is based on well-established positions of atoms in 2D (or 3D) space rather than on electron density surfaces. Secondly, it is a topology (or more strictly speaking, graph theory) based approach because any periphery contour is formed from a finite number, say m, of chemical bonds and corresponding m "exterior" atoms. Finally, this approach is shown[3] to be extended to characterization of any (planar or spatial) closed curvilinear contour; approximation of this contour by a boundary of hexagonal (or other type) lattice is needed in this case. As a result, Randić−Razinger's approach can be directly applied only *if the closed periphery contour is unambiguously defined*. In the 2D case, this requirement is evidently satisfied for polygon (not necessarily polyhex) systems which seem therefore to be the best models for the approach under consideration.

In reality, the periphery contour is not self-evident even for molecular entities which can be embedded on the hexagonal lattice. In Figure 2a, we have depicted the biphenyl framework for which no closed periphery contour can be recognized. The other similar examples can be easily constructed if two or more polybenzenoidal systems are sequentially joined by "bridging" bonds similar to the bond 1−7 in Figure 2a). Additionally, closed periphery contours cannot be recognized if carbon or heteroatomic substituents (cf. 4-methylisopropenylbenzene skeleton in Figure 2b) and all hydrogen atoms (cf. naphthalene molecule represented in Figure 2c) are taken into account. It should be noted that consideration of substituents (and also hydrogens) is especially significant for molecular modeling and QSPR/QSAR studies; the shapes of real molecules (rather than of their skeletons or frameworks) are typically responsible for chemical and/or biochemical properties of corresponding organic substances.

The above examples convincingly show that the present state approach *shifts the main difficulties associated with the notion of shape to the notion of periphery contour*. In order to make this approach applicable to a broader class of organic structures, some additional formal rules should be elaborated. These rules must make it possible to easily construct the more expanded closed contours which consist not only of

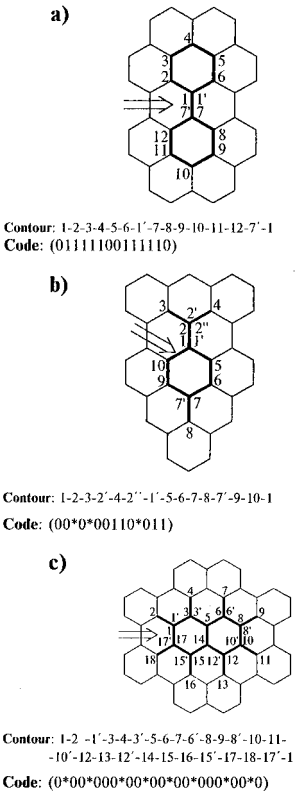

Figure 2. Periphery and pseudoperiphery bonds, expanded periphery contours, and corresponding codes for biphenyl framework (a), 4-methylisopropenylbenzene skeleton (b), and naphthalene molecule (c).

obvious periphery bonds but also of "pseudoperiphery" bonds, i.e., of bridging bonds and bonds forming acyclic appendages of any kind. The bonds to hydrogens and to other monoatomic substituents must be also included into these periphery contours if needed.

In the general case, pseudoperiphery bonds are those bonds of the planar structure which do not belong to any cycle and, additionally, are directed outside the Randić−Razinger's boundary contour. (The last restriction means that all acyclic bonds located within any planar cycle are not considered as pseudoperiphery ones.) In order to include these bonds into the closed contour, it is permitted to walk along these bonds twice, i.e., in two opposite directions. The expanded contours for biphenyl framework, 4-methylisopropenylbenzene skeleton, and naphthalene molecule are represented in Figures 2a−c; note that some atoms (atoms 1 and 7 in Figure 2a, atoms 1, 2, and 7 in Figure 2b, and atoms 1, 3, 6, 8, 10, 12, 15, and 17 in Figure 2c) are traced two or several times when walking around the contour in the clockwise manner.

The periphery codes related to expanded contours can evidently be constructed using Randić−Razinger's rules; an additional symbol is, however, needed in order to designate the backward motion after any endpoint (i.e., univalent atom) has been reached. The asterisk is used for that purpose in the ternary codes of Figure 2b,c; the starting bonds for all three codes of Figure 2 are indicated by arrows in pictorial representations of biphenyl framework, 4-methylisopropenylbenzene skeleton, and naphthalene molecule.

It is easy to see that in the case of bridged polybenzenoids (or polybenzenoids containing more or less complex substituents which can be properly embedded on a hexagonal lattice), the expanded periphery codes make it possible to easily reconstruct the original planar structures. These codes can be used in similarity and chirality measuring procedures, ordering procedures, etc. just in the same manner as was proposed in refs 3 and 4. The geometry distance matrices and corresponding invariants (e.g., topographic index ^2D and shape profile S) can also be calculated similarly; the single difference is that the number of rows and columns in a distance matrix depends not on the actual number of atoms in a contour but on the number of digits in the expanded periphery code. For example, the geometry distance matrices related to expanded contours of Figures 2a−c consist of 14, 14, and 26 (but not 12, 10, and 18) rows and columns, respectively.

The above modification of the Randić−Razinger's approach may be considered as an artificial tool which makes it possible to include acyclic bonds into the closed periphery contours. On the other hand, it is clear that formulation of the periphery contour is a crucial point of any possible "discrete" approach to characterization of the (planar) molecular shapes. For this reason, we want to suggest here a very natural alternative definition of the closed periphery contour which is completely independent of the bonding factor. This means that only atomic positions but not the connectivity of any planar molecule (or its skeleton, or its framework) must be known in order to recognize the "geometry based" molecular shape.

Let n points corresponding to n-atomic system be more or less uniformly situated in a plane; we additionally require the n-point system to be nonlinear. It is easy to see that the unique "geometry based" periphery contour always exists for such a system. This closed periphery contour is formed out of the minimal number, say m, $3 \leq m \leq n$, imaginary lines which *connect pairs of m points and do not dissect the set of other noncollinear points into those disposed to the left and to the right of the line under consideration.* In other words, for any peripheral line, all $n - 2$ points are disposed either on the line or in one of two half-planes defined by that line. The closed contour formed by m peripheral lines evidently corresponds to a convex (not necessarily regular) m-gonal geometrical figure[13] which may be thought to represent the "geometry based" shape of corresponding molecular species.

In Figure 3a, three polygonal figures, i.e., hexagon ($m = 6$), pentagon ($m = 5$), and octagon ($m = 8$), are explicitly shown by dashed lines; these convex figures evidently represent "geometry based" shapes of biphenyl framework, 4-methylisopropenylbenzene skeleton, and naphthalene molecule. It should be noted that in some cases, one or several atoms (e.g., four atoms of the biphenyl framework) can be situated *exactly on lines* which define the "geometry based" cyclic contour. In this paper, only m "boundary" points and m lines (i.e., those which correspond to vertices and edges of the polygonal figure) are considered as forming periphery contour;[14] for this reason, all atoms situated on edges on the convex polygon are neglected.

It is easy to see that in the case of "geometry based" periphery contours, the binary codes are no more useful tools for their characterization (because for all convex m-gonal planar figures any boundary code consists either of m zeros or of m units). The other Randić−Razinger's characteristics, i.e., those based on geometry distance matrix, can, however, be applied to "geometry based" planar shapes. In Figure 3b, the distance matrix ^2D related to octagonal contour of

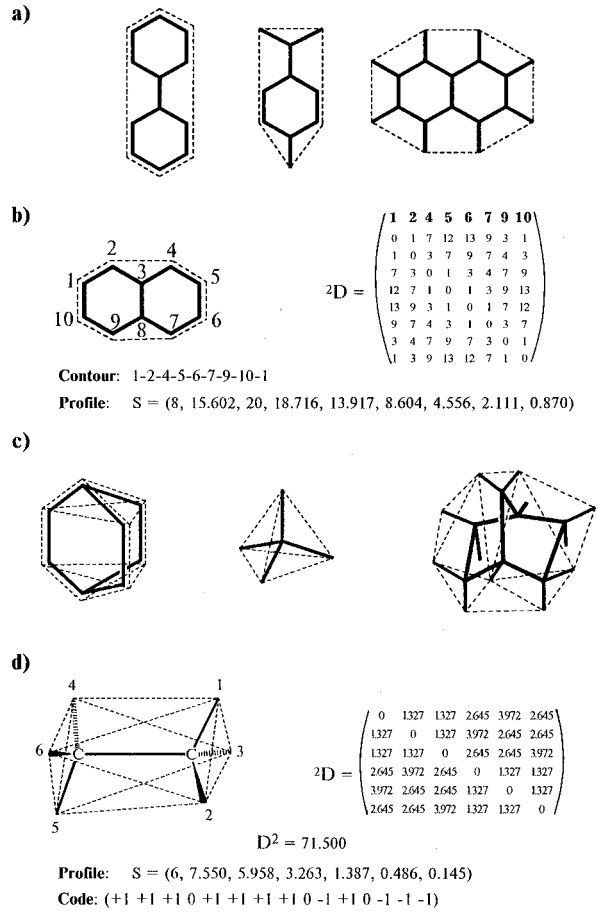

Figure 3. Geometry based characterization of molecular shapes: periphery contours for biphenyl framework, 4-methylisopropenylbenzene skeleton, and naphthalene molecule (a); distance matrix ^2D and shape profile sequence for naphthalene framework (b); 3D shapes for bicyclo[2.2.2]octane framework, neopentane skeleton, and norbornane molecule (c); and distance matrix $^2D = (d_{ij}^2)$, its invariants, and ternary code for 3D shape of staggered ethane rotamer (d).

the naphthalene skeleton is explicitly shown. This matrix differs from the corresponding matrix of Figure 1c by absence of rows and columns related to atoms 3 and 8 (which are considered as interior in the "geometry based" approach).

The differences between "topology based" and "geometry based" distance matrices lead, in turn, to different values of corresponding invariants. In the case of naphthalene framework, one can compare the values 410 and 320 for topographic index D^2 (calculated by summing up all entries of matrices ^2D in Figures 1c and 3b) and also the shape profile sequences S represented in Figures 1d and 3b. An additional interesting possibility consists in application of similarity measuring and ordering procedures[2] to "geometry based" shape profiles of polybenzenoids. In our opinion, there are no real reasons to anticipate that the results will be similar to those which were obtained by M. Randić[2] for the "topology based" shape profiles of polybenzenoidal frameworks.

The most significant feature of the "geometry based" approach to characterization of molecular shapes is that it can be directly applied to nonplanar molecules (or corresponding skeletons, or frameworks) embedded in 3D space. In this case, the "geometry based" shape can be associated *with the boundary of the convex polyhedral figure* formed by n atoms in 3D space. This figure is unambiguously defined if the set of m, $4 \le m \le n$, its vertices (or boundary atoms) is known. All other $n - m$ atoms which are situated either inside the polyhedron or exactly on its edges or faces are considered here as "interior" atoms (the alternative possibility is similar to that of note 14).

In Figure 3c, we explicitly demonstrate three polyhedral figures related to bicyclo[2.2.2]octane framework (8 carbon vertices, 15 edges, 9 faces), neopentane skeleton (4 carbon vertices, 6 edges, 4 faces), and norbornane molecule (12 hydrogen vertices, 27 edges, 17 faces of which 8 are made visible); the numbers of "interior" atoms are equal to 0, 1, and 7, respectively.

It is evident that the collection of codes (with each code relating to one of numerous closed contours formed by edges of a convex polyhedron) can be used in order to numerically characterize the shape of any convex figure; the encoding technique should be similar to that considered in ref 3 within the scope of Randić−Razinger's "topology based" approach. It is interesting, however, that any numbered m-point convex polyhedral figure can be more directly characterized by its unique ternary code which describes the relative disposition of m vertices of a polyhedron in 3D space. An example of such code is represented in the bottom of Figure 3d for staggered ethane rotamer; in this case, the shape of the convex polyhedral figure (i.e., elongated octahedron with 6 vertices, 12 edges, and 8 triangle faces) is formed out of "exterior" hydrogen atoms. Strictly speaking, ternary codes similar to that of Figure 3d are codes of *m-point 3D subconfigurations*; the explanation of how to construct such codes will be given below.

Contrary to codes of convex polyhedral figures, their numerical characteristics (e.g., topographic index D^2 and shape profile sequense S) can be produced in just the same manner as was suggested[1,2] for planar periphery contours. For that purpose, the "powered" distance matrices $^kD = (d_{ij}^k)$, $k = 0, 1, 2, ..., m$, must be initially constructed. As an example, the "squared" geometry distance matrix[15] (^2D), total of its entries (D^2), and shape profile sequence (S) are represented in Figure 3d. Note that in this example, $n = 8$, $m = 6$, and $n - m = 2$ "interior" carbon atoms are neglected in distance matrices and "geometry based" shape invariants.

An additional interesting problem associated with "geometry based" as well as "topology based" invariants consists in estimation of their discriminating ability. At first sight, topographic index D^2 and especially shape profile S seem to be extremely powerful shape invariants, and this conjecture is supported by the fact that no degeneracy was observed in the original papers[1,2] for the both "topology based" characteristics of molecular shapes. In this paper, we want to consider a very simple condition which necessarily results in coincidence of shape profiles related to planar or spatial structures of any kind. It should be noted that this condition also results in degeneracy of topographic index D^2 (because this index is derived from a single "squared" geometry distance matrix and hence is in any case less discriminating when compared with corresponding shape profile sequence S).

Suppose two distance matrices D and D' (corresponding to different planar or spatial structures) contain just the same numbers of equal entries, say, q_1 entries equal to a_1, q_2 entries equal to a_2, ..., q_r entries equal to a_r. In this case, the value of any row sum ^kR_i and corresponding value of the row sum

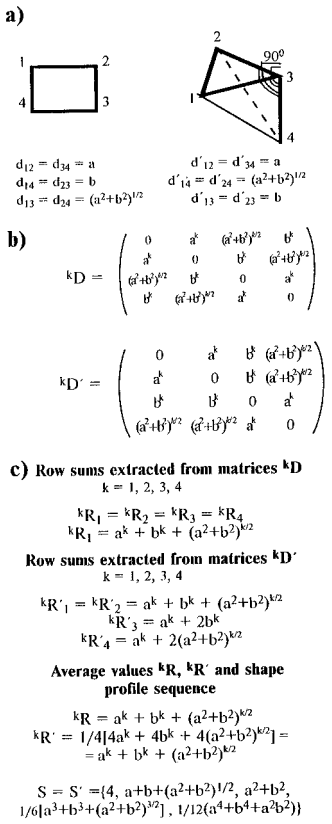

Figure 4. The simplest example of degenerate shape profiles: planar quadrangle and spatial C_s symmetric tetrahedron (a); "powered" geometry distance matrices ${}^kD$ and ${}^kD'$, $k = 1, 2, 3, 4$ (b); and row sums, averaged sums, and the resulting shape profile sequence $S = S'$ (c).

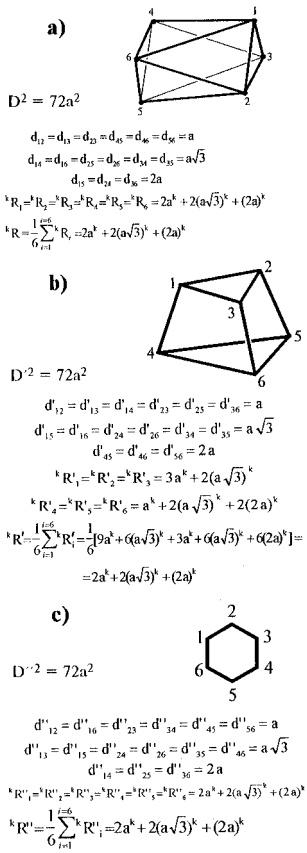

Figure 5. Coincidence of averaged sums (${}^kR$, ${}^kR'$, and ${}^kR''$) and topographic indices (D^2, D'^2, and D''^2) for elongated octahedral (a), truncated tetrahedral (b), and regular hexagonal (c) figures.

${}^kR_i'$ (both calculated by summing up all entries of the ith row, $i = 1, 2, ..., m$, of two "powered" matrices ${}^kD$ and ${}^kD'$, $k = 1, 2, ..., m$) must not be necessarily identical but the averaged sums ${}^kR$ and ${}^kR'$ will always coincide. The reason is that for any $k = 1, 2, ..., m$, these sums are defined by the total of all matrix entries which is equal to $\sum_{j=1}^{j=r} q_j a_j$.

Thus, the degeneracy condition can be formulated as follows: *the shape profiles (and also topographic indices) of two or several structures are identical if corresponding distance matrices contain the same numbers of equal entries*. It is important that this condition depends only on geometric distances between m boundary atoms but does not depend on the way these "exterior" atoms were preselected out of n atoms of the molecular structure. This fact shows the suggested degeneracy condition to be applicable both to "topology based" and to "geometry based" molecular shapes.

It may be that the simplest example of one planar and one spatial four-point figures (i.e., $m = 4$) with identical shape profiles is demonstrated in Figure 4. The "powered" distance matrices related to the quadrangle (with edge lengths equal to arbitrarily chosen a and b) and to the irregular tetrahedron (with edge lengths equal to a, b, and $(a^2 + b^2)^{1/2}$, cf. Figure 4a) are represented in Figure 4b in the generalized form: the nondiagonal entries of these matrices are kth powers of original geometric distances d_{ij} and d'_{ij}. It is easy to see that the both matrices ${}^kD$ and ${}^kD'$ contain the same number (i.e., 4) of entries equal to a^k, b^k, $(a^2 + b^2)^{k/2}$, and zero, respectively. The coincidence of shape profile sequences for both four-point figures is confirmed by direct calculations in Figure 4c; note that the values of row sums ${}^kR_i$ and ${}^kR_i'$ (related to individual rows, $i = 1, 2, 3, 4$) do not necessarily coincide, while the average values ${}^kR$ and ${}^kR'$ essentially do.

In general, we do not think that degenerate shape profiles are very typical for more complex planar or spatial figures and especially for those [other figures] which correspond to real molecules (because the coincidence of all interatomic distances seems to be quite improbable for nonisomorphic organic structures or substructures). Nevertheless, we could find an additional example which shows that the shape profile of elongated octahedron (geometrically very similar to that formed by hydrogen atoms of staggered ethane rotamer, cf. Figure 3d) is identical with the shape profiles related to truncated tetrahedral and regular hexagonal figures; the last two figures evidently resemble geometries of *cis*-1,2,3-trimethylcyclopropane and benzene skeletons. The elongated octahedron (formed from two equilateral triangles which vertices are joined by six edges of the length $a\sqrt{3}$), truncated tetrahedron (with edge lengths equal to a and $2a$), and regular hexagon (with edge lengths equal to arbitrary value of a) are visualized in Figure 5a−c together with some results of calculations. The geometry distance matrices and shape profile sequences are not explicitly shown in this case; the fact that all three shape profiles coincide can be immediately observed by comparison of geometric distances and by coincidence of "powered" averaged sumes (${}^kR$, ${}^kR'$, and ${}^kR''$, respectively). The values of topographic index (for their calculation, the average row sums for $k = 2$ must be multiplied by 6) are explicitly demonstrated in Figure 5a−c to be identical.

In our last comment, we shall clarify the mathematical nature of binary periphery codes having been used by M.

Randić and M. Razinger[3,4] for polybenzenoidal structures. These codes are, in fact, closely related to more general ternary codes which appear in our "configurational" approach to stereochemical problems[16−19] as a main tool making it possible to qualitatively characterize relative dispositions of points, or graph vertices, or atoms in 3D, 2D, or 1D space. Although the most complete description of "configurational" approach[16a] (as well as the starting publication[16b]) was written in Russian, the fundamentals of that approach were presented in two International Conferences[17a,b] and also considered in the English papers[18a,b] with necessary details. The main applications of our approach are associated with

−rigorous formulation of generation problems for gross, constitutional, and stereoformulas,[16a,b]

−elaboration of algebraic chirality criteria and classification scheme for chiral molecular systems,[18a,19b]

−derivation of geometry insensitive configurational and configuration-topological indices,[19b,c]

−characterization of atomic chains in 2D and 3D space,[19a,b]

−representation of planar 2D configurations by means of special directed graphs,[18b,19b] and

−solution of some related analytical enumeration problems.[18b,19a,b]

In this paper, only the fundamentals of "configurational" approach will be discussed in a simplified manner and to that extent which is needed to explain how configurational codes are constructed. Firstly, the notion of configuration is understood in a global sense and is not associated with disposition of some "ligands" around a "skeleton" of any kind. Instead, the "global" configuration is considered as *one of three fundamental structural characteristics* of organic molecules.

The first two of these characteristics, i.e., *composition and connectivity*, are traditionally associated with gross formulas and nondirected (multi) graphs which provide no information about localization of atoms in plane or in space. The third fundamental characteristic, i.e., *point configuration*, is thought to qualitatively describe the relative disposition of points (corresponding to atoms) in 2D or 3D space. An example is shown in Figure 6a for an arbitrarily numbered ketene molecule. The composition, connectivity and point configuration related to this molecule are separately visualized by an "expanded" gross formula (which differs from the "normal" formula C_2H_2O by explicit numbering of atoms), by corresponding numbered multigraph (which "geometry" should not be mixed with real ketene geometry), and by planar five-point figure reflecting 2D configuration of the molecule under discussion. Note that in this example (and also in Figures 6c, 7, and 8) "small point" lines are used with the sole purpose to facilitate the visualization of 2D and 3D configurations.

More strictly speaking, the point 2D and 3D configurations are associated with *types of polygonal or polyhedral figures* formed by point systems; all "interior" points must be taken into account in this case. It is important that the total amount of (unnumbered or numbered) polygonal or polyhedral figures, and consequently the number of point configurations, is *always finite* in spite of any possible violations of distances, angles, or dihedral angles. For example, four points can evidently form the tetrahedral figure in 3D space, quadrangular, "centrated" triangular, and "edge-centrated" triangular figures in 2D space and "linear" figure in 2D space. This fact shows that there can exist exactly one (unnumbered)

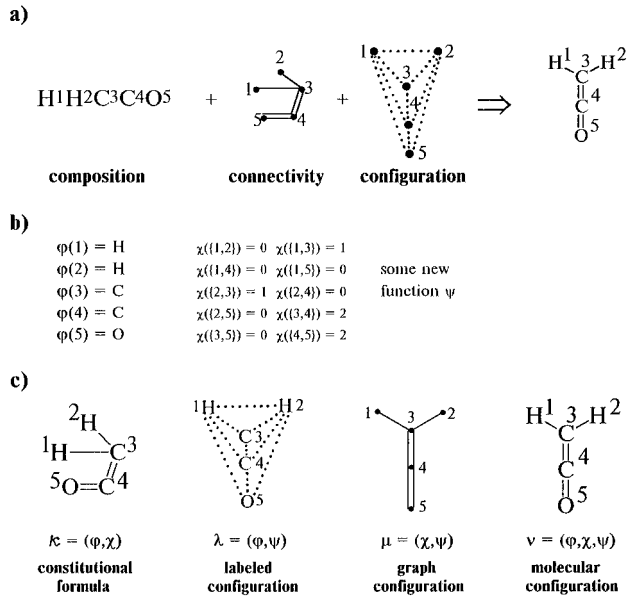

Figure 6. Fundamental structural characteristics of ketene molecule: "expanded" gross formula, nondirected multigraph, and relative disposition of atoms in plane (a); functions φ and χ (and ψ) representing its composition, connectivity (and 2D configuration) (b); and four possible superpositions of functions φ, χ, and ψ (c).

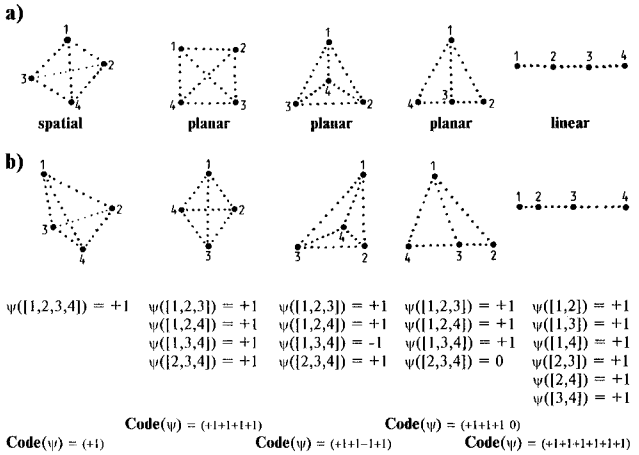

Figure 7. Regular (a) and irregular (b) four-point systems characterized by one 3D, three 2D, and one 1D configurations.

3D configuration, three 2D configurations, and one 1D[20] configuration for four-point systems. These configurations are visualized by numbered regular four-point figures in Figure 7a.

Up to this moment, we considered point configurations in a qualitative manner (i.e., as types of point figures); in this interpretation, 2D and 3D configurations are quite similar to (planar or spatial) *molecular topological forms* having been introduced[21] as geometry based qualitative characteristics of molecular structures. The most significant features of our approach are, however, based on the observation that the notion of configuration can be "algebraized", i.e., numerically represented *by a function (or mapping) from the well-defined domain set into the well-defined range set*.

The functions between finite sets will be initially illustrated here by functions φ and χ which describe the composition and connectivity of any numbered molecular structure. In Figure 6b, one can see that the domain set of the function φ (corresponding to "expanded" gross formula of Figure 6a) consists of atom numbers (1, 2, 3, 4, and 5), while the range set consists of three *labels*, i.e., atom symbols H, C, and O.

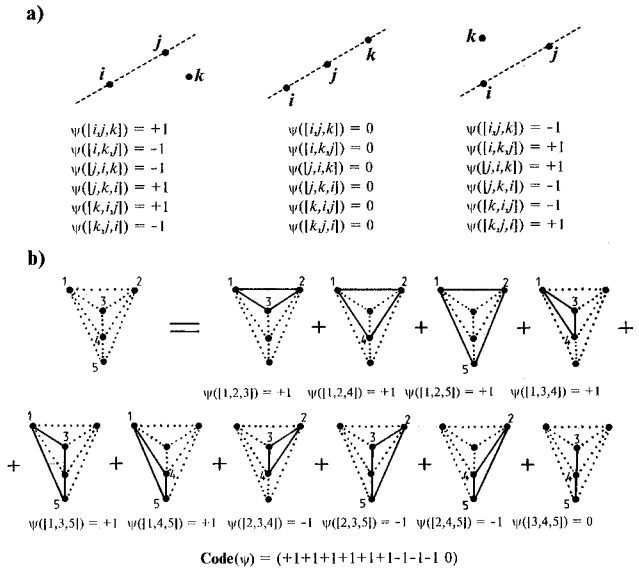

Figure 8. Point 2D configurations: values of the function ψ for ordered triples (a) and decomposition of planar five-point system into triples, values of ψ, and the resulting code for numbered five-point 2D configuration (b).

Similarly, the arguments of the function χ (relating to numbered multigraph of Figure 6a) are unordered pairs [22] {1,2}, {1,3}, ..., {4,5} or 2-subsets of the set of atom numbers, and the values of χ can be equal to integers 0, 1, 2, and 3, i.e., to possible bond multiplicities in organic structures. In both cases, the functions can be characterized by their *linear codes* (*HHCCO* and 0100100202 in the example of Figure 6b) which evidently depend on actual numbering of atoms.

In our "configurational" approach, the functions ψ describing relative dispositions of points in 2D or 3D space are to some extent similar to functions χ and can consequently be regarded as "graph-like" *combinatorial objects*.[23] The formal similarity between functions φ, χ, and ψ makes it also possible to use combinations (or, in other words, superpositions) of these functions for even more detailed characterization of molecular structures. In Figure 6c, four possible combinations $\kappa = (\varphi,\chi)$ ["composition" + "connectivity" = "constitution"], $\lambda = (\varphi,\psi)$ ["composition" + "point 2D configuration" = "labeled 2D configuration"], $\mu = (\chi,\psi)$ ["connectivity" + "point 2D configuration" = "graph 2D configuration"], and $\nu = (\varphi,\chi,\psi)$ ["composition" + "connectivity" + "point 2D configuration" = "molecular 2D configuration"] are pictorially represented for ketene molecule. Additionally, the function φ and two combinations $\kappa = (\varphi,\chi)$ and $\nu = (\varphi,\chi,\psi) = (\kappa,\psi)$ rigorously characterize the traditional hierarchy levels ("composition" − "constitution" − "stereochemistry") in description of organic molecules. This hierarchy was just discussed in literature[25,26] but only in a nonformalized verbal manner.

Let us turn to functions ψ which represent numbered 2D and 3D configurations. Similarly to functions χ (being defined for unordered pairs $\{i,j\}$ of graph vertices i and j, $i \neq j$), the functions ψ need unordered triples $\{i, j, k\}$ or quadruples $\{i, j, k, l\}$ to be known (in 2D case or 3D case, respectively). These triples and quadruples may be considered as "elementary units" of 2D and 3D configurations and hence are denoted here as (2D and 3D) configurons.[27a] It is important, however, that contrary to the values of χ, the values of functions ψ depend not only on the choice of points

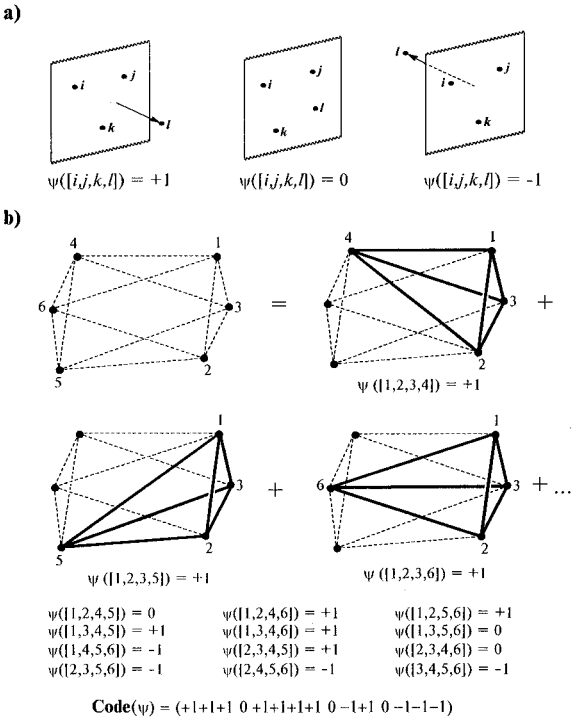

Figure 9. Point 3D configurations: values of the function ψ for ordered quadruples (a) and decomposition of spatial six-point system into quadruples, values of ψ, and the resulting code for numbered six-point 3D configuration (b).

(*i*, *j*, *k* or *i*, *j*, *k*, *l*) but also on their "orientation" or, more strictly, on the order in which the numbers of points are listed. This fact shows that ordered triples [*i*, *j*, *k*] or quadruples [*i*, *j*, *k*, *l*] are actual arguments of functions ψ.[27b]

In the simplest example of Figure 8a, the points *i*, *j*, and *k* are arranged clockwise (or in a line, or anticlockwise), and the value of ψ assigned to an ordered triple [*i*, *j*, *k*] is defined to be equal to +1 (0 or −1, respectively). From Figure 8a, one can, however, observe that clockwise or anticlockwise orientation of points is also retained for ordered triples [*j*, *k*, *i*] and [*k*, *i*, *j*] but is opposite for triples [*i*, *k*, *j*], [*j*, *i*, *k*], and [*k*, *j*, *i*].[29] Thus, the values of ψ for 3! = 6 ordered triples (explicitly shown in Figure 8a) are interdependent, and only one of them must be selected for representation purposes. Conventionally, the value of ψ related to a smallest ordered quadruple [*i*, *j*, *k*], $i < j < k$, is chosen to represent an arbitrary 2D configuron which consists of points numbered by *i*, *j*, and *k*.

A very similar situation is observed in 3D case. Four points numbered by *i*, *j*, *k*, and *l* form 3D configuron {*i*, *j*, *k*, *l*}. Of 4! = 24 possible ordered quadruples, the quadruple [*i*, *j*, *k*, *l*], $i < j < k < l$, is selected as a "standard" one. The value of ψ for this quadruple depends on localization of the point *l* with respect to the plane in which three other points *i*, *j*, and *k* are arranged clockwise (cf. Figure 9a; only the values of ψ related to a "standard" quadruple [*i*, *j*, *k*, *l*] are explicitly shown in this case).

Now we can formulate the simple rules which make it possible to construct 2*D* and 3*D configurational codes* for any (planar or spatial) numbered *n*-point system. For this purpose, one must

(1) recognize all $n(n-1)(n-2)/6$ 2D configurons or all $n(n-1)(n-2),(n-3)/24$ 3D configurons; this subproblem may be considered as "decomposition" of *n*-point system into configurons;

(2) for each configuron, i.e., for each unordered triple or quadruple, arrange point numbers in an ascending order thus forming an ordered triple or quadruple of point numbers;

(3) for each ordered triple $[i, j, k]$, $i < j < k$, or ordered quadruple $[i, j, k, l]$, $i < j < k < l$, calculate[30] the value of function ψ; and

(4) list the obtained values of ψ in the order which corresponds to lexicographic (or other[31]) order of involved triples or quadruples; the resulting code completely describes the numbered n-point 2D or 3D configuration.

The application of the above procedure to a planar five-point system (related to ketene molecule, cf. Figure 6a) is illustrated in Figure 8b; all $5 \cdot 4 \cdot 3/6 = 10$ possible 2D configurons are visualized by solid lines in this example. In a very similar manner, the spatial six-point system can be decomposed into $6 \cdot 5 \cdot 4 \cdot 3/24 = 15$ possible 3D configurons; only three of them are explicitly shown in Figure 9b. It is easy to see that in this case, the resulting 3D configurational code, Code(ψ), is identical to the code of Figure 3d which was shown above to characterize the "geometry based" shape of the staggered ethane rotamer. This fact leads to a conclusion that in order to describe the "geometry based" 3D shape of any numbered n-atomic system, one must initially recognize its m-point, $m \leq n$, convex polyhedral figure (cf. above) and then calculate 3D configurational code of that figure.

It should be noted here that in the examples of Figures 8b and 9b, both configurational codes contain zeros (related to one linear triple and three planar quadruples) and hence are ternary codes; such codes evidently characterize numbered *ternary 2D and 3D configurations*. The examples of *binary 3D, 2D, and 1D configurations* can be found in Figure 7b. Four point figures of that figure (i.e., spatial tetrahedral figure, quadrangular and "centrated" triangular figures, and linear four-point figure) are characterized by binary codes with the components equal either to -1 or to $+1$. The examples of Figure 7 additionally demonstrate that point configurations and corresponding codes are, in fact, combinatorial (but not geometrical!) characteristics which depend on *the type of point figure* rather than on real distances, angles, etc. Thus, one can easily see that the values of function ψ (and hence the codes) are just the same for regular tetrahedral figure of Figure 7a and for irregular tetrahedral figure of Figure 7b, for square figure of Figure 7a and for rhombic figure of Figure 7b, etc. This fact shows that "combinatorial symmetries" (or more strictly speaking, automorphism groups; cf. refs 16a, 18a, 19b) of functions ψ can be associated with the highest possible symmetries of corresponding point figures.

An important property of functions ψ and configurational codes consists of the fact that both always depend on *numbers of points*. The numbering system which results in the lexicographically smallest code of 2D or 3D configuration is considered as *canonical*; this word is also applicable to individual functions ψ and corresponding codes. (Note that point numbers in Figures 7b, 8b, 9b, and 10a−c are chosen arbitrarily and hence do not produce canonical codes.) The canonicalization problem for functions ψ is somewhat similar to that for functions χ (and hence for undirected graphs) and also needs the induced actions of several permutation groups, i.e., power groups,[16a,18a,19b] to be thoroughly discussed. This problem as well as other interesting problems associated with 2D and 3D configurations (e.g., their enumeration,[18b] chirality,[18a,19b] and realizability[32]) will not be specially considered in this paper. The additional notions associated with some "significant parts" of point configurations are, however, closely related to Randić−Razinger's periphery codes. These notions will be discussed in detail.

The notions of subconfiguration and partial (sub)configuration can be regarded as analogs of corresponding graph-theoretical notions, i.e., subgraph and partial (sub)graph.[33] Thus, for any n-point configuration, its *m-point subconfiguration* is defined by all possible $m(m-1)(m-2)/6$ (in 2D case) or $m(m-1)(m-2)(m-3)/24$ (in 3D case) values of ψ related to preselected m points. It is evident that the lower bound of m is equal to 3 (in any 2D subconfiguration) or to 4 (in any 3D subconfiguration). An example can be found in Figure 3d: six hydrogens form a subconfiguration of eight-point 3D configuration which additionally includes two carbon atoms of the staggered ethane rotamer. The values of ψ and *Code*(ψ) corresponding to this six-point subconfiguration were considered above (cf. Figure 9b).

In contrust to subconfigurations, the *partial m-point configurations and subconfigurations* are defined if the values of ψ are known only for some part of 2D or 3D configurons forming n-point or m-point ($m < n$) system. Thus, the collection of three values $\psi([1,2,3]) = +1$, $\psi([1,2,4]) = +1$, and $\psi([1,2,5]) = +1$ (related to the first, second, and third figures of Figure 8b) can serve as an example of partial five-point 2D configuration; any two of these values may be said to represent the partial four-point 2D subconfiguration.

It is evident that among numerous partial (sub)configurations, only those are practically significant which are constructed according to some well-defined rules. The very obvious way to construct the "significant" partial m-point, $m \leq n$, (sub)configurations is based on disclosed or closed sequences formed out of m different point numbers. Thus, the disclosed sequence ($i_2, i_2, ..., i_m$) may be associated with a "chain" $i_1-i_2-...-i_m$; the corresponding numbered m-point *chain configuration*[19a−c] is defined if the values of ψ are known for $m-2$ ordered triples $[i_1, i_2, i_3]$, $[i_2, i_3, i_4]$, ..., $[i_{m-2}, i_{m-1}, i_m]$ (in 2D case) or for $m-3$ quadruples $[i_1, i_2, i_3, i_4]$, $[i_2, i_3, i_4, i_5]$, ..., $[i_{m-3}, i_{m-2}, i_{m-1}, i_m]$ (in 3D case). In Figure 10a, the visualized 10-point chain 2D configuration ψ_C (related to a sequence (3, 4, 5, 6, 7, 8, 9, 10, 1, 2)) corresponds to one of 12 Hamiltonian chains of the naphthalene framework graph; the values of ψ_C for eight ordered triples ("lying along the chain") and corresponding code, *Code*(ψ_C), are demonstrated in Figure 10a.

It should be noted here that the codes of (binary or ternary, 2D or 3D) chain configurations were shown[19b] to describe various isomers and stereoisomers belonging to relatively broad classes of organic structures. For example, the ternary codes of chain 3D configurations were identical to the codes having been introduced by M. Randić[34a] for characterization of staggered alkane rotamers; the formally similar binary codes of chain 2D configurations were demonstrated to uniquely characterize geometrical (i.e., cis/trans and s-cis/s-trans) isomers of unbranched conjugated polyenes. In some instances, the chemical structures were initially substituted by auxiliary "characteristic graphs" which contained just the same information as corresponding partial 2D or 3D configurations. These graphs were initially suggested by A. T. Balaban[34b] in order to describe constitutional isomers of cata- and pericondensed polybenzenoids; the similar graphs were

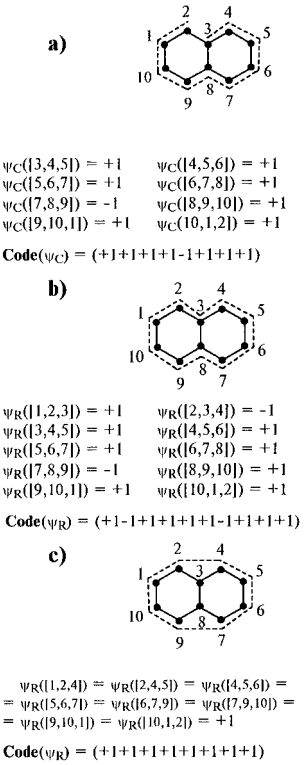

Figure 10. Examples of chain (a) and ring (b and c) configurations corresponding to traditional numbering of naphthalene framework. The values of ψ_R and *Code*(ψ_R) relate to "topology based" (b) and "geometry based" (c) periphery contours.

also applied to stereoisomeric unbranched spiro-condensed cyclopropanes (triangulanes).[19a,b] In the above-mentioned situations, the codes assigned to unbranched characteristic graphs were identical with codes of ternary 2D and binary 3D chain configurations, respectively.

In this paper, we introduce the new kind of partial (sub)-configuration which will be referred to as *ring configuration*. The numbered *m*-point ring configuration ψ_R corresponds to the closed sequence (i_1, i_2, ..., i_m) or, in other words, to the ring $i_1 - i_2 - ... - i_{m-1} - i_m$ formed by *m* points. This configuration is said to be defined if the values of the function ψ are known for *m* triples [i_1, i_2, i_3], [i_2, i_3, i_4], ..., [i_{m-1}, i_m, i_1], [i_m, i_1, i_2] (in 2D case) or for *m* quadruples [i_1, i_2, i_3, i_4], [i_2, i_3, i_4, i_5], ..., [i_{m-1}, i_m, i_1, i_2], [i_m, i_1, i_2, i_3] (in 3D case). The examples of 10-point and eight-point binary ring 2D configurations (related to "topology based" and "geometry based" periphery contours for naphthalene framework, cf. with Figures 1a,b and 3b) are visualized by dashed lines in Figure 10b,c together with the values of the function ψ_R and corresponding codes, *Code*(ψ_R). Note that the ring configuration of Figure 10b corresponds to the 10-membered sequence (1, 2, 3, 4, 5, 6, 7, 8, 9, 10) and is considered as partial configuration, while the ring configuration of Figure 10c corresponds to the eight-membered sequence (1, 2, 4, 5, 6, 7, 9, 10) and hence must be regarded as partial subconfiguration.

At this step, we can (at last!) state that Randić−Razinger's codes and codes of binary ring 2D configurations are "two kinds of water from one well". To be convinced, the reader can compare the binary code of Figure 10b with the first periphery code of Figure 1b. Both codes correspond to the same numbering of points (or atoms of naphthalene framework, cf. Figures 10b and 1a) and can be interconverted by substitution of 0 by −1, or vice versa.[35] This fact shows that there exists a one-to-one correspondence between Randić−Razinger's periphery contours of polybenzenoids and those binary ring 2D configurations which points can be properly embedded on a regular hexagonal lattice.

In their fourth paper,[4] M. Randić and M. Razinger have mentioned that "Construction of all benzenoid forms having the same perimeter *P* has apparently not yet been considered in the literature.". Bearing in mind this statement and also difficulties in nonduplicate generation of large polybenzenoids,[4] we can declare that *any effective procedure for computer-assisted enumeration of nonisomorphic binary ring 2D configurations can be easily modified in order to sequentially produce canonical periphery codes of medium-size and even large polybenzenoidal systems*. For that purpose, a single (but effective!) subalgorithm making it possible to recognize "improper" *partially constructed codes* (i.e., codes to which no properly embeddable "complete" ring 2D configurations correspond) must be additionally elaborated.

In the preceding discussion, chain and ring configurations were associated with the preselected disclosed and closed sequences formed by *m* points (corresponding to atoms) in 2D or 3D space. Graph structures of molecules and hence functions χ were not taken into account in the above definitions. There exists, however, an alternative possibility, i.e., to define chain and ring configurations as numerical characteristics of corresponding graphs which *m* vertices are disposed in 2D or 3D space. Thus, chain configurations can be regarded as *superpositions* $\mu_C = (\chi,\psi_C)$ *relating to "path" graphs* P_m, while ring configurations can be regarded as *superpositions* $\mu_R = (\chi,\psi_R)$ *relating to "cycle" graphs* C_m. In this interpretation, the both superposition are, in fact, partial *graph configurations*; we recall here that "connectivity" + "point configuration" = "graph configuration" (cf. above).

Both definitions are very similar in essence but, strictly speaking, superpositions μ_C and μ_R should be preferred if one deals with "isolated" *m*-atomic chains or rings which need not be considered as parts of more complex molecular graphs. In this case, the vertices of graphs P_m and C_m can always be sequentially numbered by integers 1, 2, ..., *m* (rather than by arbitrary integers i_1, i_2, ..., i_m, cf. above), and this fact significantly facilitates computer-assisted manipulations with chain and ring configurations. It in the rest of this paper, mainly the ring configurations $\mu_R \equiv (C_m, \psi_R)$ corresponding to "isolated" cyclic graphs C_m will be considered, and for this reason, the dashed lines will be substituted by (ordinary or multiple) graph edges in all pictorial representations.

In Figure 11a, one binary chain 2D configuration (μ_C) and two binary ring 2D configurations (μ_R) of Figure 10a−c are visualized by means of embeddings of graphs P_{10}, C_{10}, and C_8 on the hexagonal lattice. The *canonically numbered*[36] graph vertices and corresponding lexicographically minimal codes *Code*(ψ_C) and *Code*(ψ_R) are shown in this example. In order to provide a real example of ring 3D configurations μ_R, three spatial embeddings of the graph C_6 (corresponding to principal conformations of cyclohexane framework) are pictorially represented in Figure 11b together with their canonical codes. This example convincingly shows that one ternary and two binary codes of functions ψ_R (relating to superpositions $\mu_R = (C_6, \psi_R)$) are unique and actually

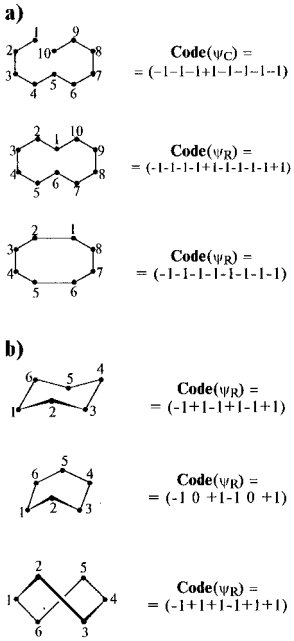

Figure 11. Characterization of planar (a) and spatial (b) embeddings of graphs P_m and C_m. The chair, boat, and one enantiomeric twist form of cyclohexane framework are discriminated by canonical codes of corresponding ring 3D configurations.

differentiate between the chair, boat, and chiral twist forms of six-membered cycle.

Several additional properties of ring 2D configurations will be finally discussed. Firstly, these partial configurations are in no case associated with embeddings of graphs C_m on hexagonal (or any other) lattice. This fact leads to a conclusion that the functions ψ_R and their codes may be equally applied in order to describe the shapes of planar annulenes and dehydroannulenes of any size and also the shapes of cata- and peri-condensed polycyclic systems constructed from cycles of any size. As an example, the canonically numbered embeddings of graphs C_{14}, C_{12}, and C_{10}, related to 1,8-bisdehydro[14]annulene, perinaphthylene, and azulene frameworks, are represented (by fatty lines) in Figure 12a. The first of the corresponding canonical codes, *Code*(ψ_R), exemplifies the applicability of ternary 2D ring configurations to characterization of planar molecular shapes.

On the other hand, one must clearly understand that ring 2D configurations (and functions ψ_R) are *much less informative* when compared with corresponding "complete" configurations (and functions ψ). The example of Figure 12b demonstrates that "trigonal", "tetragonal", and "pentagonal" planar embeddings of the graph C_5 cannot be discriminated by five values of the functions ψ_R and hence are characterized by the same ring 2D configuration. These embeddings are however differentiated by "complete" 2D configurations. For that purpose, all values of the functions ψ (i.e., five values of ψ_R plus five additional values for ordered triples [1, 2, 4], [1, 3, 4], [1, 3, 5], [2, 3, 5], and [2, 4, 5]) must be taken into account (cf. Figure 12b). Even a more striking example, that of Figure 12c, shows that configurational codes of the "star"-like and two other nontraditional embeddings of the graph C_5 are identical to canonical code of the regular pentagonal shape. Thus, ring 2D configurations are sensitive only to relative disposition of graph vertices along cyclic contours $\overline{1-2-\ldots-m}$ but (in contrast to 2D configurations ψ) do not completely characterize planar polygonal figures.

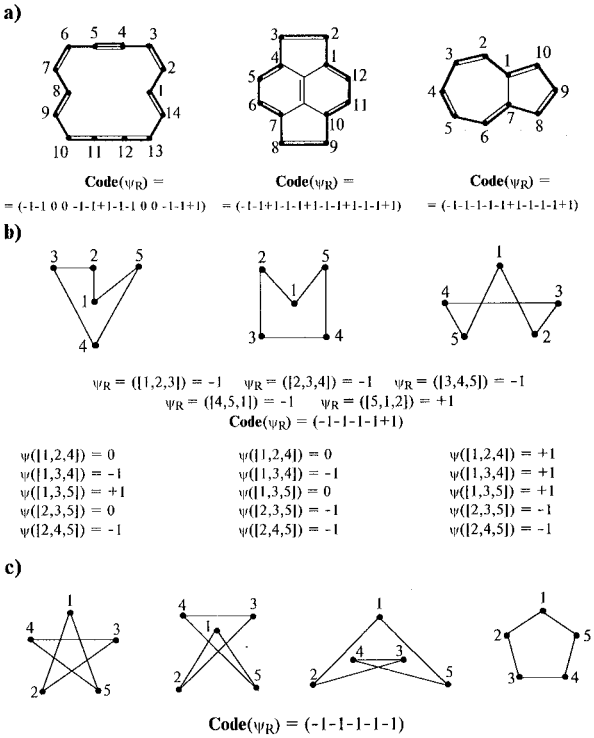

Figure 12. Planar realizations of some ring 2D configurations: chemically realizable embeddings of graphs C_{14}, C_{12}, and C_{10} (a) and nonidentical embeddings characterized by the same configurational code, *Code*(ψ_R) (b and c).

The examination of Figures 12b,c clearly shows that among various planar embeddings of graphs C_m there exist many embeddings in which two or more lines (corresponding to graph edges) intersect. Any ring 2D configuration is *graphically nonrealizable*[37] if intersecting lines can be found in all its planar embeddings; such configurations are obviously of no interest to an organic chemist. For this reason, the following question arises: *"In what instances are binary and ternary ring 2D configurations graphically nonrealizable?"*. The similar question may be posed if to remind one (cf. notes 32 and 37) that *geometrically nonrealizable* "complete" point configurations also exist; these configurations are characterized by functions ψ to which no possible dispositions of points in 2D or 3D space correspond.[16a,18b]

In a binary case, the answer to both questions is very simple: *all binary ring 2D configurations related to graphs C_m, $m \neq 4$, are geometrically and graphically realizable*. A single geometrically realizable but graphically nonrealizable ring 2D configuration is visualized in the left part of Figure 13a together with the values of the function ψ_R. (Note that canonical numberings of graph vertices are used in all examples of Figure 13.) The geometrically nonrealizable binary ring 2D configuration also corresponds to the graph C_4 and is also unique. In order to understand the nonrealizability of this configuration, one must compare the values of functions ψ_R and ψ represented in the right part of Figure 13a. The fact that these values coincide (due to alternativity property of the function ψ cf. note 29) shows that nonrealizability is really associated with four-point 2D configuration ψ. The schematic drawing in the four part of Figure 13a completely clarifies the situation: if points 1, 2, and 3 are arranged in an anticlockwise manner (according to the value of ψ([1, 2, 3]), then point 4 must be located to the right of the line 1−2 (because ψ([1, 2, 4]) = +1), to the left of the

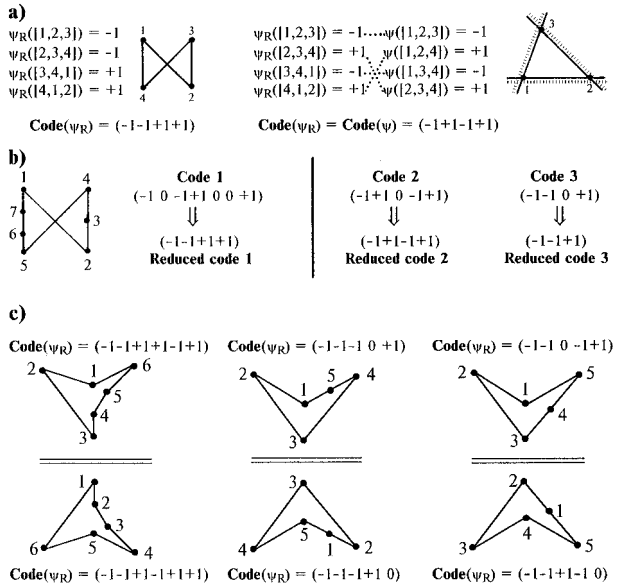

Figure 13. Realizability and chirality of 2D ring configurations: geometrically and graphically nonrealizable binary (a) and ternary (b) functions ψ_R and the smallest examples of chiral ring 2D configurations (c).

line 1−3 (because $\psi([1, 3, 4]) = -1$), and to the right of the line 2−3 (because $\psi([2, 3, 4]) = +1$). But these three half-planes do not intersect in any case, and this proves the geometrical nonrealizability of the both functions, ψ and ψ_R.

The situation with ternary configurations is somewhat more complicated. Any ternary ring 2D configuration is realizable if its code can be reduced (by deleting all zeros) either to the l-component, $l \geq 4$, code of realizable binary ring configuration, or to one of sequences $(-1,-1,-1)$, $(+1,+1,+1)$ which uniquely correspond to configurational codes, $Code(\psi_R) = \pm 1$, for the graph C_3. (In reality, instead of three values $\psi_R([i, j, k]) = \psi_R([j, k, i]) = \psi_R([k, i, j]) = \pm 1$, only one is needed to codify any nonlinear embedding of this graph.) An example of graphically nonrealizable ternary ring 2D configuration is visualized in the left part of Figure 13b; it is evident that the reduced *Code 1* coincides with the left code of Figure 13a. This example shows that linear parts of the graph embedding are actually shortened (by deleting all "interior" graph vertices) in the course of the reduction procedure. The geometrically nonrealizable ternary ring 2D configurations are represented in the right part of Figure 13b by their canonical codes. The first of these codes, *Code 2*, is reduced to the code of geometrically nonrealizable binary configuration (cf. the right part of Figure 13a) while the second, *Code 3*, is reduced to the three-component sequence $(-1,-1,+1)$ which cannot be associated with any planar embedding of the graph C_3.

Similarly to "complete" configurations ψ, the ring 2D configurations can also be classified into achiral and chiral. This, "two-dimensional" chirality can be recognized by examination of configurational codes in just the same manner as was described above for Randić−Razinger's codes. It should be noted, however, that chirality property is an intrinsic property of the functions ψ_R which also may be regarded as combinatorial objects (cf. note 23). In order to completely understand the chirality (and also symmetry) properties of ring 2D configurations, *the normal and expanded automorphism groups*[18a,19b] of the functions ψ_R

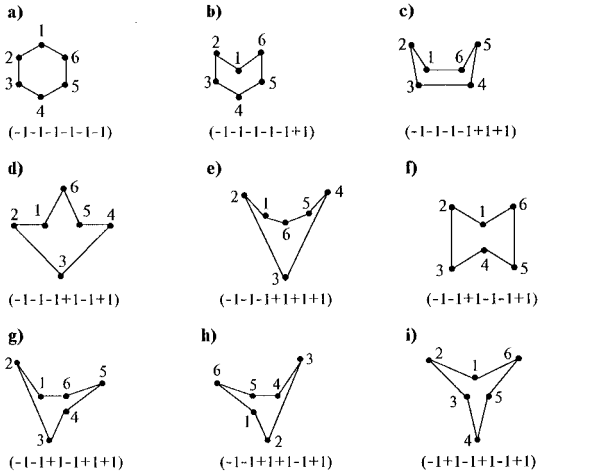

Figure 14. The complete list of binary six-point ring 2D configurations: canonical numberings and configurational codes (a−i).

must be considered in detail. These groups are, in fact, subgroups of the dihedral groups D_m which unequivocally describe all symmetries associated with "isolated" cyclic graphs C_m. For the sake of brevity, we demonstrate here only the smallest chiral ring 2D configurations. One enantiomeric pair of six-point binary and two enantiomeric pairs of five-point ternary ring 2D configurations are visualized in Figure 13c together with their canonical codes.

The last (and may be the most important!) feature of ring 2D configurations consists of the fact that canonical m-component codes of functions ψ_R can be effectively constructed by computer programs even for large values of m. All possible canonical codes of binary configurations represented in Figure 14 (together with planar embeddings of corresponding graph C_6) were produced by the simplest of such programs. From this figure, it is evident that there exists seven achiral (a−f,i) plus one pair of enantiomeric chiral (g,h) binary ring 2D configurations; one can also see that six vertices of the graph C_6 form hexagonal (a), pentagonal (b), 3 tetragonal (c,d,f), and 4 trigonal (e,g−i) figures in two-dimensional space. According to results obtained by a computer, the total number of corresponding graphically realizable ternary ring 2D configurations is equal to 33; of these configurations, 11 relate to achiral and 22 to chiral embeddings of the graph C_6.

The computer-assisted nonduplicate generation of canonical codes for (binary or ternary, achiral or chiral, graphically realizable or related to hexagonal lattice, etc.) functions ψ_R is based on *rigorous mathematical models of ring 2D configurations*; in some instances these models make it also possible to analytically enumerate ring configurations without their actual generation. According to general mathematical models, the unnumbered ring 2D configurations (or, in other words, equivalence classes consisting of isomorphic functions ψ_R) are *orbits of the doubly induced actions of dihedral groups D_m on complete sets of individual functions*. The comprehensive exposition of these models requires the (quite nontraditional for chemists) mathematical formalisms to be considered in detail and hence cannot be described here in a simple, nonformal style.

This main notions related to mathematical models of ring 2D configurations (and some results obtained by enumeration and generation techniques) were briefly summarized by one

of us at MATH/CHEM/COMP'96 Conference;[38a] the corresponding detailed paper[38b] is now in preparation and will be published elsewhere. The second part of the present paper (in which we tried to build a bridge between Randić−Razinger's codes and our configurational codes) may be regarded as an introductory survey to the rigorous theory of ring configurations.

REFERENCES AND NOTES

(1) Randić, M.; Razinger, M. Molecular Topographic Indices. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 140−147.

(2) Randić, M. Molecular Shape Profiles. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 373−382.

(3) Randić, M.; Razinger, M. On Characterization of Molecular Shapes. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 594−606.

(4) Randić, M.; Razinger, M. Molecular Shapes and Chirality. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 429−441.

(5) Mezey, P. G. *Shape in Chemistry: An Introduction to Molecular Shape and Topology*; VCH Publ.: New York, 1993. The broader list of references on results of P. G. Mezey's group may be found in refs 2−4.

(6) Strictly speaking, neglecting hydrogens and substituents means that molecular periphery (and hence molecular shape, its codes, invariants, etc.) does not characterize real planar molecules but relates to planar embeddings of corresponding hydrogen-depleted, or skeleton, or framework graphs. In the case of *polybenzenoidal hydrocarbons*, these three sorts of graphs coincide but in the general case they can be differentiated, see sections 7 and 9 in ref 18a.

(7) Balaban, A. T. Chemical Graphs. Part 12. Configuration of Annulenes. *Tetrahedron* **1971**, *27*, 6115−6131.

(8) Numerous other codes were used in order to identify or generate polybenzoidal systems: three-digit codes associated with unbranched characteristic graphs,[9a] binary codes of boundary (= periphery) degree sequences,[9b] six-digit boundary codes,[9c] and pairs of binary codes representing half-boundaries of (unbranched) polybenzenoids[9d] may serve as examples.

(9) (a) Balaban, A. T.; Harary, F. Chemical Graphs. Part 5. Enumeration and Proposed Nomenclature of Benzenoid Cata-Condensed Polycyclic Aromatic Hydrocarbons. *Tetrahedron* **1968**, *24*, 2505−2516. (b) Mgel'skaya, E. V. Graphs of Polycyclic Compounds. In *Problems of Algorithmic Analysis for Structural Information (Computational Systems, 119)*; Novosibirsk, 1987; pp 71−90 [in Russian]. (c) Knop, J. V.; Szymanski, K.; Jeričević, Ž.; Trinajstić, N. Computer Enumeration and Generation of Benzenoid Hydrocarbons and Identification of Bay Regions. *J. Comput. Chem.* **1983**, *4*, 23−32. (d) Dobrynin, A. A. An Effective Generation Algorithm for Graphs of Unbranched Hexagonal Systems. In *Mathematical Problems of Chemical Informatics (Computational Systems, 130)*; Novosibirsk, 1989; pp 3−38 [in Russian].

(10) For any planar molecule (localized in a plane σ_h), to two three-dimensional symmetry elements a single symmetry element in 2D space evidently corresponds. Thus, an identity symmetry element E and reflection plane σ_h correspond to two-dimensional identity element. Similarly, the main n-fold rotation axis C_n, $n \geq 2$ (perpendicular to σ_h), and corresponding rotation-reflection axis S_n, $n > 2$ (or inversion center i if $n = 2$), are transformed into a single n-fold rotation point or center of symmetry;[3] the rotation around this point by $2\pi/n$ leads to coincidence of the molecule with itself. Finally, to axes C_2' (located in σ_h) and to other reflection planes (i.e., σ_v and σ_d), one or several reflection lines correspond; just these "mirror lines" must be absent in two-dimensionally chiral planar molecules.

(11) (a) In 1971, H. Hosoya[11b] have stated that the known Wiener's number[11c] W could be calculated by summing up all entries of the (topological) distance matrix: $W = {}^1/_2\sum_{i=1}^{i=n}\sum_{j=1}^{j=n} d_{ij}$; in this case, d_{ij} denotes graph-theoretical distance between atoms i and j in a molecular graph of order n. (b) Hosoya, H. Topological Index. A Newly Proposed Quantity Characterizing the Topological Nature of Structure Isomers of Saturated Hydrocarbons. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332−2339. (c) Wiener, H. Structure Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, *69*, 17−20.

(12) The interrelations between "discrete" and "continuous" approaches to description of molecular shapes are discussed in ref 2. The "continuous" approaches[5] are based on a somewhat fuzzy concept of molecular surface which, in turn, significantly depends on the assumed threshold for electron density. For this reason, "continuous" approaches are less properly adapted for organic chemists which traditionally operate with molecular graphs and their embeddings in 2D or 3D space.

(13) In more rigorous mathematical terminology, this polygonal figure is a *convex hull* of the planar n-point system. Similarly, in the case of nonplanar n-point system in 3D space, its convex hull is represented by m-point ($4 \leq m \leq n$) convex polyhedral figure.

(14) Strictly speaking, there is an alternative possibility: *all atoms disposed on edges of the convex polygon* can be included into the periphery contour (and hence taken into account in corresponding distance matrices and their invariants). The same problem appears however in any "topology based" approach to characterization of molecular shapes if periphery contour contains linear fragments such as −C≡C−, >C=C=C<, or −N=C=N− fragments. In such cases, the ternary (but not binary!) codes should be used in Randić−Razinger's description[3] of planar as well as spatial molecular shapes.

(15) The entries d_{ij} are real distances between hydrogen atoms *divided by 1.54* (i.e., by the length of normal $C-C$ distance). All $H\cdots H$ distances in the staggered ethane rotamer are, in turn, calculated from coordinate values represented in T. Clark's *Handbook of Computational Chemistry*; John Wiley & Sons,Inc.; New York, 1985.

(16) (a) Tratch, S. S. Mathematical Models in Stereochemistry. I. Combinatorial Characteristics of Composition, Connectivity, and Configuration of Organic Molecules. *Zh. Organ. Khim.* **1995**, *31*, 1320−1351 [in Russian]. (b) Tratch, S. S.; Zefirov, N. S. Combinatorial Models and Algorithms in Chemistry. Ladder of Combinatorial Objects and Its Application to Formalization of Structural Problems of Organic Chemistry. In *Principles of Symmetry and Systemology in Chemistry*; Stepanov, N. F., Ed.; Moscow State University Press: Moscow, 1987; pp 54−86 [in Russian].

(17) (a) Tratch, S. S. Combinatorial Approach to Stereochemical Problems: A Brief Survey. Abstracts of the Sixth International Conference on Mathematical Chemistry; Pitlochry, Scotland, July 10−14, 1995. (b) Tratch, S. S. Combinatorial Objects in Chemistry: Formal Models, Enumeration Techniques, and Generation Algorithms. Abstracts of the Eleventh International Course & Conference on the Interfaces Among Mathematics, Chemistry, and Computer Sciences; Dubrovnik, Croatia, June 24−29, 1996.

(18) (a) Tratch, S. S.; Zefirov, N. S. Algebraic Chirality Criteria and Their Application to Chirality Classification in Rigid Molecular Systems. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 448−464. (b) Klin, M. H.; Tratch, S. S.; Zefirov, N. S. 2D-Configurations and Clique-Cyclic Orientations of the Graphs L(K_p). *Rep. Mol. Theory* **1990**, *1*, 149−163.

(19) (a) Zefirov, N. S.; Kozhushkov, S. I.; Kuznetsova, T. S.; Kokoreva, O. V.; Lukin, K. A.; Ugrak, B. I.; Tratch, S. S. Triangulanes: Stereoisomerism and General Method of Synthesis. *J. Am. Chem. Soc.* **1990**, *112*, 7702−7707. (b) Tratch, S. S. *Logical−Combinatorial Approaches to Design Problems for Organic Structures, Reactions, and Configurations* (Doctoral Dissertation Thesis); Moscow, 1993; Vol. 2, pp 56−144 [in Russian]. (c) Tratch, S. S.; Devdariani, R. O.; Zefirov, N. S. Combinatorial Models and Algorithms in Chemistry. Configuration−Topological Analogs of Wiener Index. *Zh. Organ. Khim.* **1990**, *26*, 921−932 [in Russian].

(20) Linear or 1D point configurations are not considered in this paper because relative dispositions of points in 1D space can be trivially characterized by ordered sequences of point numbers.

(21) Drozd, V. N.; Zefirov, N. S.; Sokolov, V. I.; Stankevitch, I. V. Topological Definition of the Notion "Stereochemical Configuration". *Zh. Organ. Khimii* **1979**, *15*, 1785−1793 [in Russian].

(22) All nondirected graphs and multigraphs can also be described by *symmetric functions* χ from the sets of ordered pairs [1, 2], [2, 1], [1, 3], [3, 1], ... into the set of allowed edge multiplicities; the symmetry property means that $\chi([i, j]) = \chi([j, i])$ for all pairs of vertices i and j, $i \neq j$.

(23) The individual functions between arbitrary finite sets (and also the equivalence classes of these functions) can be regarded as combinatorial objects.[24a] A. Kerber[24b] was the first in mathematical chemistry who has treated graphs in this way.

(24) (a) Faradgev, I. A. Constructive Enumeration of Combinatorial Objects. In *Algorithmic Investigations in Combinatorics*; Faradgev, I. A., Ed.; Nauka Press: Moscow, 1978; pp 3−11 [in Russian]. (b) Kerber, A. On Graphs and Their Enumeration. *MATCH* **1975**, *1*, 5−10.

(25) For example, W. Hässelbarth[26a] states "On the lowest level of description only the composition of molecules with respect to certain types of fragments is recorded. The next level specifies molecular constitution, that is, the bonding connectivity among the fragments." and "On the next level, opening the door to what is called stereochemistry, molecules are recognized to live in 3-dimensional space. On this level the description records qualitative geometrical features...". The very similar (and also qualitative) ideas were discussed in refs 26b,c.

(26) (a) Hässelbarth, W. A. Generalisation of the Pólya/de Bruijn Enumeration Theory and Its Application to "Chemical Combinatorics". *MATCH* **1986**, *20*, 241−250. (b) Turro, N. J. Geometric and Topological Thinking in Organic Chemistry. *Agnew. Chem., Int. Ed. Engl.* **1986**, *25*, 882−901. (c) Maggiora, G. M.; Johnson, M. A. Introduction to Similarity in Chemistry. In *Concepts and Applications of Molecular Similarity*; Johnson, M. A.; Maggiora, G. M., Eds.; John Wiley & Sons, Inc.: New York, 1985; pp 1−13.

(27) (a) Other terms such as 3-subsets and 4-subsets can be applied to unordered triples and quadruples. The nonlinear triples and nonplanar quadruples can also be treated as vertices of two- and three-dimensional simplices, respectively. (b) A. S. Dreiding and K. Wirth[28a] were probably the first chemists who had used ordered triples and quadruples in discussion of stereochemical problems; these authors introduced the term chiron for "the vertices of a mobile oriented simplex, which retains its orientation". In a very similar situation, the name "tetrad" was also used[28b] for ordered quadruples related to four-atomic chains.

(28) (a) Dreiding, A. S.; Wirth, K. The Multiplex. A Classification of Finite Ordered Point Sets in Oriented d-Dimensional Spaces. *MATCH* **1980**, *8*, 341−352. (b) Johnson, M.; Tsai, C.; Nicholson, V. Chemical Stereographs: An Extension of Chemical Graphs for Representing Stereochemical and Conformational Structures. *J. Math. Chem.* **1991**, *7*, 3−38.

(29) In a more rigorous formulation, all even permutations of *i, j, k* retain the *sign* of ψ, while all odd permutations of point numbers convert -1 into $+1$ and vice versa. Just the same situation is observed for ordered pairs [*i, j*] (related to 1D configurons {*i, j*}) and for ordered quadruples [*i, j, k, l*] (related to 3D configurons {*i, j, k, l*}). This fact shows that in all cases ψ is *an alternating function*, and this is the main difference between functions χ and ψ (χ may be regarded as a symmetric function, cf. note 22).

(30) The values of functions ψ can be calculated analytically, i.e., without visual examination of three-point or four-point figures. These values (i.e., -1, 0, and $+1$) were shown[16a,18b] to be defined by reversed sign of the determinant which corresponds to 3×3 (or 4×4) matrix with columns related to points *i, j, k* (and *l*). The first rows of the both matrices consist of units while the second, third (and fourth) rows are formed out of *x*-, *y*- (and *z*-) coordinates of the point system under consideration.

(31) In this and other papers, we use the lexicographic order of triples and quadruples. According to this order (in the case $n = 5$), the values of ψ in Figure 8b correspond to the following list of 10 triples: [1, 2, 3], [1, 2, 4], [1, 2, 5], [1, 3, 4], [1, 3, 5], [1, 4, 5], [2, 3, 4], [2, 3, 5], [2, 4, 5], [3, 4, 5]. The "reversed" lexicographic order results however in more effective solution of some computational problems and hence is well-adapted for computer representation of functions ψ. According to this order, the same 10 triples form the following sequence: [1, 2, 3], [1, 2, 4], [1, 3, 4], [2, 3, 4], [1, 2, 5], [1, 3, 5], [2, 3, 5], [1, 4, 5], [2, 4, 5], [3, 4, 5].

(32) Although to any disposition of numbered points in 2D or 3D space the unique function ψ evidently corresponds, the reverse statement is not true due to existence of *geometrically nonrealizable* point configurations. For this reason, the general realizability criterion for functions ψ is a crucial point in computer-assisted solution of several generation problems related to different kinds of configurations. At present, we can only assert that to any realizable *n*-point 2D or 3D configuration, the convex hulls for all *m*-point, $m = 4, 5, ..., n$, subconfigurations must necessarily exist.

(33) Slightly different definitions of subgraphs and partial (sub)graphs can be found in mathematical literature. In this paper, we follow the terminology having been used in N. Christophide's handbook (*Graph Theory: An Algorithmic Approach*; Academic Press: New York, 1985).

(34) Randić, M. Graphical Enumeration of Conformations of Chains. *Int. J. Quant. Chem., Quant. Biol. Symp.* **1980**, *7*, 187−197. (b) Balaban, A. T.; Harary, F. Chemical Graphs. V. Enumeration and Proposed Nomenclature of Benzenoid Cata-Condensed Polycyclic Aromatic Hydrocarbons. *Tetrahedron* **1968**, *24*, 2505−2516.

(35) The codes formed out of zeros and units (or out of integers 0, 1, 2 in ternary case) can be treated as binary (or ternary) numbers to which the decimal equivalents uniquely correspond. On the other hand, the codes consisting of -1 and $+1$ (and 0 in ternary case) more clearly reflect relative orientations of points in corresponding 2D or 3D configurons. The choice between coding rules is not, however, important because any of possible coding systems can easily be converted into another one.

(36) The attentive reader will note that there is no correspondence between the digits of the Randić−Razinger canonical code for the naphthalene framework (the code marked by an asterisk in Figure 1b) and of the second code of Figure 11a. In the Randić−Razinger description, the canonical code is the smallest code which can be obtained when walking around the periphery contour *in a clockwise manner*. In contrast, the canonicity of the ring 2D configuration is not associated with some preselected (clockwise or anticlockwise) direction; the codes corresponding to all 2*m* possible numberings must be compared in order to recognize the smallest, i.e., canonical code. Both the canonicalization systems can be shown to differentiate between (achiral or chiral, binary or ternary) ring configurations as well as between corresponding planar or spatial embeddings of graphs C_m. The "circulation independent" canonicalization rules seem, however, to be more general in the sense that they do not depend on the choice between two possible "walking directions" and hence can be applied to all 2D as well as 3D ring configurations. Note that in 3D space, clockwise and anticlockwise directions cannot, in principle, be differentiated.

(37) We differentiate between geometrical, graphical, and chemical realizability of ("complete" or partial) 2D and 3D configurations. The point (or labeled, or graph, or molecular) configuration is geometrically nonrealizable if there exist no dispositions of points (or graph vertices, or atoms) to which the prescribed values of ψ correspond. The geometrically realizable graph (or molecular) configuration is graphically nonrealizable if in any of its embeddings at least two graph edges necessarily intersect. The less precisely defined notion of chemical realizability relates to graphically realizable molecular configurations; this kind of realizability evidently depends on allowed values of interatomic distances and bond and dihedral angles which can, in turn, be affected by numerous other factors.

(38) Tratch, S. S. On Molecular Shapes and Ring 2D Configurations. Abstracts of the Eleventh International Course & Conference on the Interfaces Among Mathematics, Chemistry, and Computer Sciences; Dubrovnik, Croatia, June 24−29, 1996. (b) Tratch, S. S.; Zefirov, N. S. Mathematical Models of Ring Configurations and Their Application to Enumeration and Generation Problems. Manuscript in preparation.

CI9700182