

Using Protochirons for Three-Dimensional Coding of Certain Chemical Structures

Alexandru T. Balaban^{*,†} and Christoph Rücker[‡]

Department of Oceanography and Marine Sciences, Texas A&M University at Galveston, P.O. Box 1675, Galveston, Texas 77553-1675, and Department of Mathematics, Universität Bayreuth, D-95440 Bayreuth, Germany

Received October 2, 2000

For three-dimensional coding (including enantiomerism) of staggered paths and circuits on the diamond lattice, or paths/circuits with angles 90° or 180° on the cubic lattice, use is made of the previously defined paths-3 (paths of length three bonds defining two intersecting planes). The two cases mentioned above are examined and exemplified. In the diamond lattice there are three kinds of diamond-paths-3: one is achiral (Z) and two are chiral and enantiomeric (R and S). In the cubic lattice there are six kinds of orthopath-3, of which only two are chiral and enantiomeric (R and S) and four are achiral (I, L, U, and Z). The chiral paths-3 are the previously defined protochirons in the respective lattice. Coding ascribes to each bond the letter that would characterize it if it were the central bond of an isolated path-3. To obtain a unique code out of several equally correct ones it is proposed to use the convention of inverse alphabetic priority in the above system of letters.

INTRODUCTION

Stereochemical coding of a compound's configuration is based on the Cahn–Ingold–Prelog (CIP) Rules (incorporated into the IUPAC nomenclature) which ascribe an *R* or *S* descriptor to each asymmetrically substituted sp^3 -hybridized tetravalent atom such as carbon. Computer programs taking into account stereochemical features of molecules have been developed since the 1970s.^{1–11}

Far less effort has been invested by the Chemical Community in coding a compound's conformation. While pictorial words such as chair, boat, and twist are commonly used for some conformations of particularly prominent molecules, a general coding system for conformations seems not to be in wide use. We here undertake a step to fill this gap. While the basic idea of the proposed procedure is not new,¹² it is used here for cyclic compounds on the diamond lattice and for paths and cycles on a cubic grid for the first time.

In a previous paper *protochirons* were defined as the smallest chiral paths of length three bonds, chiral equivalents to hydrogen-depleted butane graphs, designated here as chiral *paths-3* (Figure 1).¹³ In that paper it was shown how one could construct chiral objects by *concatenation* (association such that two vertices merge into one) or by *assembly* (association such that two edges merge into one) of protochirons. The present paper takes one a further step and defines *overlap* as association of two paths-3 such that the last two edges of a path-3 merge with the first two edges of another path-3.

Whereas in graphs neither edge lengths nor any angles are defined, for the conformation of real molecules bond lengths, valence angles, and dihedral (interplanar, torsion) angles are crucial. If the former two are kept constant, then it is the latter only which define a molecule's conformation.

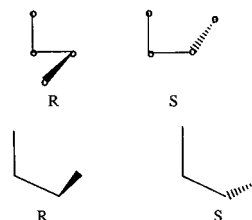


Figure 1. Protochirons on the cubic lattice (upper row) and diamond lattice (lower row).

Therefore, in the following we shall use the graph-theoretical term *path* with an added geometrical meaning: We here deal with the angle between two planes defined by the first two and the last two edges along a path-3.

For simplicity we shall examine in the present paper first paths-3 on the diamond lattice and then on the cubic lattice.

Paths and Circuits on the Diamond Lattice (Staggered Carbon Chains). On the diamond lattice, where all bond lengths are constant and all valence angles are tetrahedral (109.5°), there are three and only three geometrically different kinds of paths-3, which we shall call collectively *diamond-paths-3*. In chemistry, these correspond to the three staggered conformers of the carbon skeleton of *n*-butane. Recall that an achiral compound such as *n*-butane can adopt both achiral and chiral conformations. The zigzag planar and achiral path (interplanar angle 180° , corresponding to the *anti* conformer) is pictorially denoted by the letter Z, whereas the two chiral and enantiomeric *gauche* conformers (interplanar angle 60°) are denoted by R and S, since they may be considered parts of a right- or left-handed helix, respectively (Figure 2). In Eliel and Wilen's book,¹² these butane conformers are called a (*anti*), g^+ (*gauche*+), and g^- (*gauche*-), respectively. Note that the symbols Z, R, S introduced here for the conformation of paths-3 are to be written not in italics, to distinguish them from the italicized CIP symbols *R*, *S*.

* Corresponding author e-mail: balabana@tamug.tamu.edu.

[†] Texas A&M University at Galveston.

[‡] Universität Bayreuth.

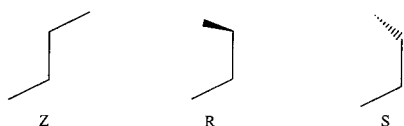


Figure 2. The three diamond-paths-3. In longer paths and circuits the respective symbols are written near the central edge of each path-3.

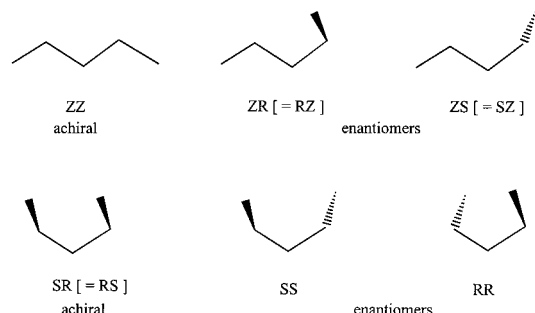


Figure 3. The six staggered conformers of *n*-pentane.

The procedure to assign stereodescriptors R or S to a protochiron closely resembles the construction of a Newman projection and the CIP procedure: First, view the path-3 along its central edge. Second, rotate the closer terminal edge along the 60° angle until it eclipses the more distant terminal edge. If this rotation is clockwise, the stereodescriptor is R, otherwise it is S. Note that the R/S assignment is independent of the viewing direction.

Any longer path or contour on the diamond lattice may be described as a series of *overlapping diamond-paths-3* (similarly to our earlier coding, by overlapping triplets, of *cata*-condensed polycyclic aromatic hydrocarbons or of diamond hydrocarbons using their corresponding dualist graphs).^{14–16} Overlapping diamond-paths-3 are defined as two diamond-paths-3 where the first two edges of the following one are merged with the last two edges of the preceding one. In enumerating the successive diamond-paths-3 on a longer path, each edge gets as stereodescriptor the letter by which it would be described if it were the central edge of an isolated diamond-path-3.

As the simplest example, there are six conformers of the achiral hydrocarbon *n*-pentane: ZZ, ZS, ZR, SS, SR, and RR (Figure 3). Conformers ZZ and SR (the latter is identical with RS) are achiral, and the remaining conformers are chiral (ZS and SZ are identical, and also ZR and RZ are identical). The four chiral ones exist as two pairs of enantiomers: RR with SS and ZS with ZR. In cases of more than one correct code for one and the same conformer, we propose to select the code with highest inverse alphabetic priority ($Z > S > R$), and this of course arbitrary convention was followed above already. The *n*-pentane conformers were coded earlier, e.g. in Eliel's book;¹² otherwise in previous enumerations enantiomeric conformers were not considered individually.^{14–18}

Coding the conformation of a branched chain does not pose any difficulty. We here propose to use the following general system of coding, which is a generalization of the above. Introducing a branching point somewhere in the middle of a chain has the effect that both main chain bonds incident with the branching point are characterized by *two* dihedral angles instead of one, so *two* symbols, glued together by a pair of parentheses, are required for each such bond.

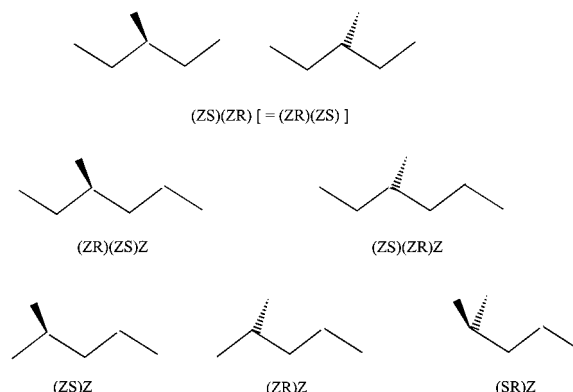


Figure 4. Top: Two drawings of one and the same main chain ZZ conformer of 3-methylpentane. Middle: The main chain ZZZ conformer of (*R*)-3-methylhexane and of (*S*)-3-methylhexane. Bottom: Two enantiomeric and the achiral conformers of 2-methylpentane. The codes proposed in this work are given for each conformer.

Within a pair of parentheses, the first symbol describes the main chain conformation, the second symbol the branch.¹⁹

Consider for illustration the achiral hydrocarbon 3-methylpentane and the chiral hydrocarbon 3-methylhexane, particular conformers of which are shown in Figure 4, top and middle rows. Each of their conformations can be oriented so that the branching atom (C3) and its two main chain neighbors (C2 and C4) are in the paper plane. Then there seems to be one degree of freedom for the orientation of the side chain (methyl group), up or down. For the depicted conformer of 3-methylpentane (main chain conformation ZZ) these two possibilities turn out to be identical (Figure 4, top); in fact the two drawings represent one and the same achiral conformer, which by the above rule has to be coded (ZS)-(ZR). This code clearly indicates this conformer's achirality.

Altogether there exist three achiral staggered rotamers and three pairs of enantiomeric staggered rotamers of 3-methylpentane: achiral: (ZS)(ZR), (SR)(RS), and (SZ)(RZ); chiral: (ZS)(RS) and (ZR)(SR), (ZS)(SZ) and (ZR)(RZ), and (SZ)(SR), and (RZ)(RS).

A chiral compound such as (*R*)- or (*S*)-3-methylhexane has all its conformations necessarily chiral, the orientation of the side chain in every conformation is determined by the *R*- or *S*-configuration of atom C3 (recall that *italic* symbols *R* and *S* correspond to the CIP rules for coding configurations). The conformers of 3-methylhexane with ZZZ main chain conformation depicted in Figure 4, middle row, are corresponding conformers of the *R* and the *S* enantiomeric hydrocarbons. Their codes according to the above rule, (ZR)-(ZS)Z and (ZS)(ZR)Z, respectively, clearly mirror their enantiomerism.²⁰ Altogether there are 27 pairs of enantiomeric staggered rotamers of 3-methylhexane (not shown).

If branching is introduced at the end of a chain, only one of the two main chain bonds incident with the branching point (the interior one) is to be described by *two* dihedral angles, while the exterior one still has no dihedral angle and thus still does not get a stereodescriptor. In fact in such a case it is not defined which methyl group is part of the main chain and which is the branch; therefore, the first and second position within a pair of parentheses need not be distinguished.

As a case of a terminally branched chain consider the achiral hydrocarbon 2-methylpentane. There are three dif-

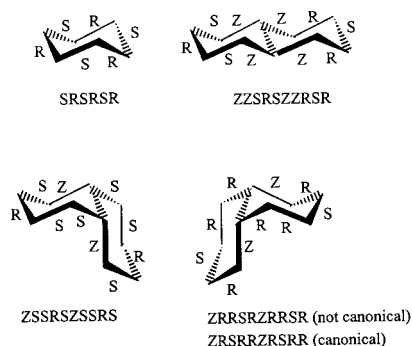


Figure 5. Upper row: codes for *chair*-cyclohexane and *trans*-decalin; lower row: codes for the *cis*-decalin enantiomers.

ferent rotamers with respect to the C2–C3 bond for each C3–C4 rotamer, and this is of course true for the backbone ZZ conformation illustrated in Figure 4 (bottom). In the chiral rotamer drawn on the left, bond C2–C3 is the central edge of a Z path-3 with respect to one methyl group and of an S path-3 with respect to the other. It is therefore given a double symbol in parentheses (ZS), so that the complete code of this conformer is (ZS)Z. Its enantiomer (middle) accordingly is (ZR)Z, the achiral conformer (right) is (SR)Z. In addition to the conformers shown in the last row of Figure 4, there are three more pairs of staggered chiral rotamers of 2-methylpentane: (ZS)R and (ZR)S; (SR)S and (SR)R; and (ZS)S and (ZR)R. These examples in our opinion are representative of many branched chains.

Now we shall examine the coding of contours (circuits) on the diamond lattice. The chair-shaped cyclohexane is achiral; it contains three planes of symmetry intersecting along the C_3 axis that is orthogonal to the median ring plane (Figure 5). The edges in turn are given R and S descriptors. Of course, the correct code for a cyclic structure is cyclic itself. To linearize it one can cut it at any position, and the resulting string of letters can be read in both directions. All the resulting codes are equally correct. Here it is again proposed to use inverse alphabetic priority to single out a preferred (canonical) code, which for *chair*-cyclohexane becomes SRSRSR. It contains equal numbers of R and S enantiomeric diamond-paths-3 and is in obvious accord with the molecular symmetry.

The *trans* diastereomer of decalin is also achiral; it has one plane of symmetry passing through the central bond and being perpendicular to the median plane of the molecule. The code of the contour (i.e. ignoring the central bond) after cutting and applying the canonicalization rule is ZZSRSZZRSR. Again one observes an equal number of R and S enantiomeric protochirons, and the enantiomeric relation of the two molecular halves is clearly mirrored in the code.

On the other hand, *cis*-decalin exists as a racemic mixture of two equilibrating chiral conformers, each of which has only a binary symmetry axis C_2 bisecting the central bond. The codes of the two enantiomers of course reflect the molecular symmetry; they are ZSSRSZSSRS and ZRRSRZRRSR, respectively, by the rule of inverse alphabetic priority. Notice that this rule obscures the enantiomerism, which would be apparent if the equally correct but not canonical code ZRRSRZRRSR had been selected for the second enantiomer. In each of these codes for *cis*-decalins, one enantiomeric diamond-path-3 (R or S) outnumbers the other (S or R, respectively) by four. Therefore, by the same

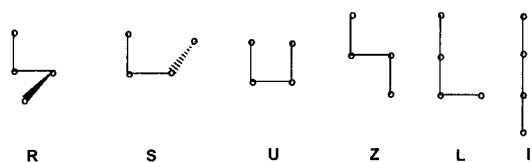


Figure 6. The six orthopath-3s and their notation (L has several subtypes). In longer paths and circuits the respective symbols are written near the central edge of each path-3.

convention that was applied to the R and S diamond-paths-3, the former conformer of *cis*-decalin should be called the S enantiomer, and the latter should be the R enantiomer.

Decoding is easy and unambiguous provided information is given whether a particular string of symbols codes a cycle or a path. Thus, SRSRSR as given above for *chair*-cyclohexane could as well code a particular conformation of *n*-nonane (alternatively, a special symbol for ring closure could be included in the code). If a given code stands for a path, then the first symbol is to be translated into the dihedral angle of the second edge, i.e., it determines the direction of the third edge after the first and second edges are drawn arbitrarily (but of course with a tetrahedral valence angle).

Paths and Circuits on the Cubic Lattice. On the cubic lattice we have *orthogonal objects*, i.e., all valence and interplanar angles are 0° , 90° , or 180° . We call a path-3 in the cubic lattice an *orthopath-3*. In analogy to the above there are two enantiomeric protochirons R and S as well as four achiral arrangements of three edges in a path under the above restrictions. These four types are denoted by capital letters I, L, U, and Z using the evident mnemonic shape analogy. In Figure 6 all six orthopath-3s are shown. The interplanar angles are 90° in R and S, 0° in U, 180° in Z, and undefined in L and I.

The procedure to assign R/S descriptors to the protochirons is exactly as described above, with the adjustment that the expression “ 60° ” is replaced by “ 90° ”. Again, a stereodescriptor R or S is obtained for a particular orthopath-3 independent of the direction of viewing and independent of a direction that may be inherent in the path (e.g. in the protein example below).

One can now apply the overlap operation for coding any path or circuit on the cubic grid. This coding has obvious advantages over the trivial specification of the points on the path or circuit via their Cartesian coordinates, because the code is independent of the orientation on the grid. One can recall the similarity with a previously used overlapping triplet code for specifying staggered conformations of alkanes^{14,16} or structures of cata-condensed polycyclic aromatic hydrocarbons via their dualist graphs.^{21–25}

It is interesting to look at more complex paths. A $3 \times 3 \times 3$ cubic lattice is frequently used as a frame for compact 27-mers in model studies of protein folding.²⁶ Randic presented five such models and proposed quantitative measures for their respective extent of folding.²⁷ These five patterns are shown in Figure 7. In each case the origin is denoted by O and the final point by F. We here present codes for such arrangements.

As always, the first edge in a longer path is not the central edge of a path-3 and therefore does not get a stereodescriptor. The second edge gets a descriptor according to the angle between the plane formed by edges 1 and 2 and the plane formed by edges 2 and 3, and so on. When an extended

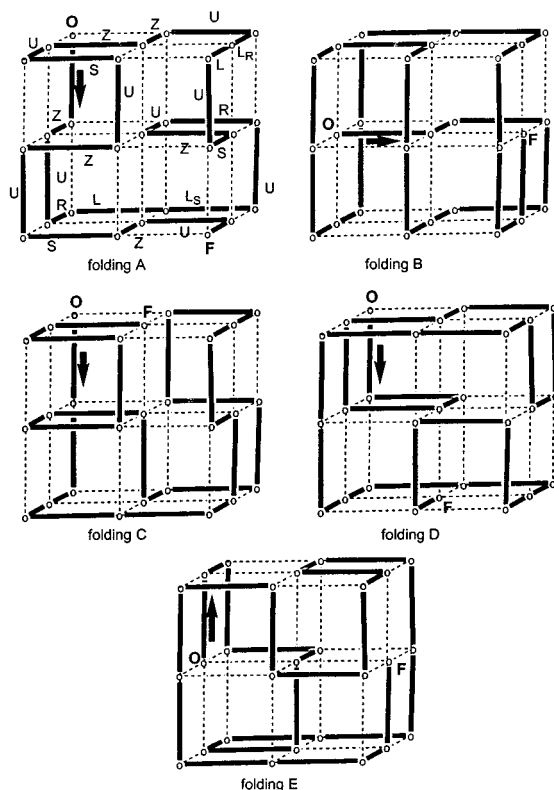


Figure 7. Five model protein foldings of Randic.

region is encountered (an L region in the examples), its first symbol is simply L (for the central edge of the L), meaning for the decoding procedure “for the next edge go ahead”. This is followed by one of the symbols L_R , L_S , L_U , L_Z (for the terminal edge in the long arm of the L, which at the same time is the central edge of another L arrangement), meaning in decoding “for the next edge turn by 90° so that a centrally elongated R, S, U, or Z is formed”, respectively.²⁸ The last edge in the path again does not get a stereodescriptor, since its direction is coded by the descriptor ascribed to the second-last edge.

In a $3 \times 3 \times 3$ cubic grid (Figure 7) no four vertices are arranged in a straight line, an I descriptor therefore cannot appear, while in the general case one or several I may be enclosed in a pair of L descriptors, or may precede a single initial L, or may follow a final single L. The codes for the foldings designated **A** through **E** in Figure 7 and the corresponding numbers of L descriptors are as follows: **A**: ZURLL_SURUZSULL_RUZZUSUZUSZU 4 L descriptors; **B**: ZRUSLL_RLL_URUSLL_ULL_ULL_RLL_SLL_RLL_UU 16 L descriptors; **C**: LLL_SLL_URUZULL_RSUZUSRRRURUSU 7 L descriptors; **D**: RUSURZRSUZSLL_RLL_SLL_RLL_ULL_SRUS 10 L descriptors; and **E**: USRRSRULL_SLL_URUZRSUZU-LL_ULL_SR 8 L descriptors.

The “foldedness/extendedness” of such arrangements can be roughly quantified simply by counting the L descriptors or in the general case the L and I descriptors, where one I is equivalent to a pair of L descriptors.^{29,30} The ordering of the above foldings according to the numbers of L descriptors fully agrees with that obtained according to Randic’s Φ values,²⁷ while Estrada’s folding measure results in a slightly different order.³¹ Use of these codes as “names” for arrangements and as tools for evaluation of similarity (partial identity) can be envisaged.

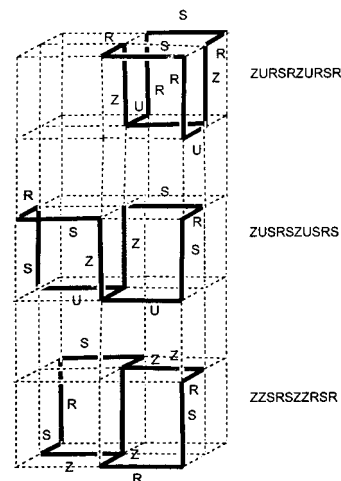


Figure 8. Models approximating *cis*-decalin enantiomers (top) and *trans*-decalin (bottom) on a cubic lattice.

Now we present a few circuits on the cubic lattice. Contours and codes corresponding to *trans*-decalin and the two enantiomeric *cis*-decalins (all severely distorted) are shown in Figure 8. Note the difference between the codes for the *cis*-decalins given here and those given in Figure 5. Such a difference was to be expected, since the conformations on the diamond and the cubic lattices are clearly different, the decalins being forced into different Procrustes’ beds. The identity of the codes for *trans*-decalin in the two lattices therefore should be considered accidental.

Again, it is evident that for *trans*-decalin the coding rules lead to equal numbers of R and S orthopaths-3. For the first of the *cis*-decalin enantiomers, on the other hand, R-type orthopaths-3 outnumber the S-type orthopaths-3, and for the second one the reverse situation holds. As required for enantiomers, again the two codes can be interchanged by reverting each R/S assignment.

Limitations of the present approach are obvious. Most real molecules cannot be superimposed on the diamond or cubic lattice, i.e., dihedral angles in molecular conformations have to be described by real numbers. Even *n*-butane and cyclohexane exist as conformers geometrically slightly different from the ideal conformations considered here.¹²

The present approach is very simple due to the restrictions imposed by the lattice, rendering the dihedral angles discrete with few values allowed. When the restrictions are relaxed, codes similar to those introduced above for the diamond-lattice case should be applicable to the conformations of at least all compounds made of sp^3 centers only, provided that a quantitative descriptor (say, the dihedral angle given in $^\circ$) is added to a qualitative descriptor where appropriate. In the tetrahedral case not restricted to the diamond lattice a U path-3 also is viable (eclipsed butane conformation), so that the achiral *boat*-cyclohexane can be described as USRUSR and the two enantiomeric *twist*-cyclohexane conformers as $SR_{40}R_{40}SR_{40}R_{40}$ and $S_{40}S_{40}RS_{40}S_{40}R$, respectively (assuming dihedral angles of 40°). Topologically interesting molecules such as knots, catenanes, and rotaxanes could then be envisaged, but more elaborate techniques would be needed.^{32–36} The search for improved coding systems for three-dimensional chemical structures continues.^{3,4,31,37} The present type of approach will be developed in a future paper for specifying protein folding motifs in terms of backbone dihedral angles

ϕ and ψ and side chain dihedral angles χ , all of which influence a peptide's biological activity.³⁸

In conclusion, we have proposed a method for coding the conformations of paths and circuits on the diamond and cubic lattices. Both the coding and the decoding procedures are unambiguous as well as simple. Using examples it was demonstrated that the codes are enantiomorphic for enantiomeric conformations and completely different for diastereomeric conformations.

REFERENCES AND NOTES

- (1) Wipke, W. T.; Dyott, T. M. Stereochemically unique naming algorithm. *J. Am. Chem. Soc.* **1974**, *96*, 4834–4842.
- (2) Balaban, A. T.; Mekenyan, O.; Bonchev, D. Unique description of chemical structures based on hierarchically ordered extended connectivities (HOC procedures). III. Topological, chemical and stereochemical coding of molecular structure. *J. Comput. Chem.* **1985**, *6*, 562–569.
- (3) Balaban, A. T. From chemical graphs to 3D molecular modeling. In *From Chemical Topology to Three-Dimensional Geometry*. Balaban, A. T., Ed.; Plenum Press: New York, 1997; pp 1–24.
- (4) Satoh, H.; Koshino, H.; Funatsu, K.; Nakata, T. Novel canonical coding method for representation of three-dimensional structures. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 622–630.
- (5) Herndon, W. C.; Leonard, J. E. Canonical numbering, stereochemical descriptors, and unique linear notations for polyhedral clusters. *Inorg. Chem.* **1983**, *22*, 554–557.
- (6) Akutsu, T. A new method of computer representation of stereochemistry. Transforming a stereochemical structure into a graph. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 414–417.
- (7) Hahn, M. Three-dimensional shape-based searching of conformationally flexible compounds. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 80–86. Tasi, G.; Mizukami, F. Quantum algebraic-combinatoric study of the conformational properties of *n*-alkanes. *J. Math. Chem.* **1999**, *25*, 55–64.
- (8) Goldman, B. B.; Wipke, W. T. Quadratic shape descriptors. 1. Rapid superposition of dissimilar molecules using geometrically invariant surface descriptors. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 644–658.
- (9) Sadowski, J. Three-dimensional structure generation: automation. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III., Schreiner, P. R., Eds.; Wiley: Chichester, 1998; pp 2976–2988.
- (10) Güner, O. F.; Henry, D. R. Three-dimensional structure searching. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III., Schreiner, P. R., Eds.; Wiley: Chichester, 1998; pp 2988–3003.
- (11) Balaban, A. T. Topological and stereochemical molecular descriptors for databases useful in QSAR, similarity/dissimilarity and drug design. *SAR QSAR Environ. Res.* **1998**, *8*, 1–21.
- (12) Eliel, E. L.; Wilen, S. H. *Stereochemistry of Organic Compounds*; Wiley: New York, 1994; p 603. See also the references given therein.
- (13) Balaban, A. T. From chemical topology to 3D geometry. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 645–650.
- (14) Balaban, A. T. Chemical graphs. XXVII. Enumeration and codification of staggered conformations of alkanes. *Rev. Roum. Chim.* **1976**, *21*, 1049–1071.
- (15) Funck, E. Symmetry and numbers of rotation isomers for unbranched polymethylene chains. *Z. Elektrochem.* **1958**, *62*, 901–905 (in German).
- (16) Balaban, A. T. Enumeration of catafusenes, diamondoid hydrocarbons, and staggered alkane C-rotamers. *Math. Chem. (MATCH)* **1976**, *2*, 51–61.
- (17) Randic, M. Graphical enumeration of conformations of chains. *Int. J. Quantum Chem., Quantum Biol. Symp.* **1980**, *7*, 187–197.
- (18) Balaban, A. T.; Schleyer, P. v. R. Systematic classification and nomenclature of diamond hydrocarbons. 1. Graph-theoretical enumeration of polymantanes. *Tetrahedron* **1978**, *34*, 3599–3609.
- (19) The main chain is the longest linear chain defined so as to have the largest number of branches; atoms and bonds in the main chain are numbered in that direction that results in the lowest locants for branching points.
- (20) Another coding system is possible for branched compounds carrying a CIP descriptor, such as (R)- or (S)-3-methylhexane; it profits from the fact that in every conformation the direction of the side chain is determined by the configuration of the branching atom. It therefore suffices to code the conformation of the main chain and to add the CIP descriptor. In this system the two conformers of the enantiomeric 3-methylhexanes depicted in Figure 4, middle row, are ZZZ-(3R)-3-methylhexane and ZZZ-(3S)-3-methylhexane, respectively. We do not favor this nomenclature system, since it is obviously useless in cases lacking a CIP descriptor, such as 3-methylpentane and 2-methylpentane, and generally for branched chains in the cubic lattice, where analogues to the CIP descriptors are not defined.
- (21) Balaban, A. T.; Harary, F. Chemical graphs. V. Enumeration and proposed nomenclature of benzenoid cata-condensed polycyclic aromatic hydrocarbons. *Tetrahedron* **1968**, *24*, 2505–2516.
- (22) Smith, F. T. Capacitive energy and the ionization of aromatic hydrocarbons. *J. Chem. Phys.* **1961**, *34*, 793–801.
- (23) Balaban, A. T. Chemical graphs. VII. Proposed nomenclature of branched cata-condensed benzenoid polycyclic hydrocarbons. *Tetrahedron* **1969**, *25*, 2949–2956.
- (24) Balaban, A. T. Challenging problems involving benzenoid polycyclics and related systems. *Pure Appl. Chem.* **1982**, *54*, 1075–1096.
- (25) Balaban, A. T. Enumeration of isomers. In *Chemical Graph Theory. Introduction and Fundamentals*; Bonchev, D., Rouvray, D. H., Eds.; Abacus Press – Gordon and Breach: New York, 1991; pp 177–234.
- (26) (a) Shakhnovich, E.; Gutin, A. Enumeration of all compact conformations of copolymers with random sequence of links. *J. Chem. Phys.* **1990**, *93*, 5967–5971. (b) Socci, N. D.; Onuchic, J. N. Kinetic and thermodynamic analysis of proteinlike heteropolymers: Monte Carlo histogram technique. *J. Chem. Phys.* **1995**, *103*, 4732–4744. (c) Li, H.; Helling, R.; Tang, C.; Wingreen, N. Emergence of preferred structures in a simple model of protein folding. *Science* **1996**, *273*, 666–669.
- (27) (a) Randic, M.; Krilov, G. Characterization of 3-D sequences of proteins. *Chem. Phys. Lett.* **1997**, *272*, 115–119. (b) Randic, M. Topological indices. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III., Schreiner, P. R., Eds.; Wiley: Chichester, 1998; pp 3018–3032.
- (28) Thus, L descriptors always occur in pairs “LL_N”, N ∈ {R, S, U, Z}. As an exception, at the beginning of a code after zero, one, or several I a single L can appear, meaning “turn by 90°”, e.g. in folding C in Figure 7; likewise, for symmetry reasons, at the end of a code, followed by zero, one, or several I a single L may appear, meaning “go ahead”. If a given code is to be read in the opposite direction, the subscript in each LL_N pair has to be moved, e.g. XXXLL_NYYY becomes YYYLL_NXXX.
- (29) Thus, the procedure is equivalent to simply counting the number of vertices at which the path does not turn.
- (30) By the way, the lengths of the printed codes (measured e.g. in cm) for paths of equal lengths (measured in vertices or edges) due to the subscripts are positively correlated with the numbers of LL_N pairs included and thus with the extendedness of the path: the most extended (most folded) path has the longest (shortest) code. This is of course true only in the absence of I descriptors.
- (31) Estrada, E. Characterization of 3D molecular structure. *Chem. Phys. Lett.* **2000**, *319*, 713–718.
- (32) Liang, C.; Mislow, K. Classification of topologically chiral molecules. *J. Math. Chem.* **1994**, *15*, 245–260.
- (33) Walba, D. M. A topological hierarchy of molecular chirality and other tidbits in topological stereochemistry. In *New Developments in Molecular Chirality*; Mezey, P. G., Ed.; Kluwer Academic: Dordrecht, 1991; pp 119–129.
- (34) Simon, J. Topological chirality of certain molecules. *Topology* **1986**, *25*, 229–235.
- (35) Flapan, E. Intrinsic chirality. *J. Mol. Struct. (THEOCHEM)* **1995**, *336*, 157–164; *When Topology Meets Chemistry. A Topological Look at Molecular Chirality*. Cambridge University Press, 2000, p 23.
- (36) Balaban, A. T. Chemical graphs. Part 49. Open problems in the area of condensed polycyclic benzenoids: topological stereoisomers of coronoids and congeners. *Rev. Roum. Chim.* **1988**, *33*, 699–707.
- (37) Nicolaou, K. C.; Boddy, C. N. C.; Siegel, J. S. Does CIP nomenclature adequately handle molecules with multiple stereoelements? A case study of vancomycin and cognates. *Angew. Chem., Int. Ed. Engl.* **2001**, *40*, 701–704.
- (38) Hruby, V. J. Design in topographical space of peptide and peptidomimetic ligands that affect behavior. A chemist's glimpse at the mind-body problem. *Acc. Chem. Res.* **2001**, *34*, 389–397, and references therein.