Stereoisomer Generation in Computer-Enhanced Structure Elucidation

Marko Razinger,† Krishnan Balasubramanian,‡ Marko Perdih,† and Morton E. Munk*,‡

Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85287-1604, and National Institute of Chemistry, P.O. Box 30, Ljubljana, Slovenia

Received May 11, 1993•

The existence of stereoisomerism can be related to the presence of structural features referred to as stereocenters. On the basis of this relationship, computer software capable of generating all stereoisomers of a given constitutional isomer has been developed. Input to the program is a canonical connection table. In an initial step the stereocenters present in the structure are identified and characterized either as true-stereocenters or para-stereocenters. In the presence of n different true-stereocenters, structure generation is trivial and leads to 2^n stereoisomers, each of which can be represented by a parity vector. In the presence of topological symmetry, the number of stereoisomers may be less than 2^n; i.e., some of the 2^n parity vectors correspond to equivalent configurations. An algorithm for determining equivalences among the set of 2^n parity vectors is described. Program output includes the parity vector of each valid stereoisomer and the relationships between them, i.e., enantiomeric or diastereomeric. A simple procedure for representing the parity vector of the stereoisomer as a three-dimensional representation is given.

INTRODUCTION

The characterization of the structure of an organic compound of synthetic or natural origin on the basis of its chemical and spectral behavior is of widespread importance in the chemical and biological sciences. In the case of compounds of considerable complexity, such as in commonly encountered in natural products, the process of characterization can be time-consuming, even in the hands of experienced chemists. In recent years, reports of long-term efforts in the development of software to augment the productivity of chemists engaged in this kind of work have appeared in the literature. Systems such as CHEMICS,[1] EPIOS,[2] ACCESS,[3] and SESAMI[4] integrate some spectrum interpretation capabilities with constrained structure generation. CONGEN[5] and GENOA,[6] although highly interactive and powerful, are structure generators only and therefore require the interpretation of spectral data and the input of structural information by the user.

The SESAMI project has as its goal the development of software capable of *directly* reducing the collective spectroscopic properties of a compound of unknown structure to no more than a manageable number of alternative structures, each of which is compatible with the data. This set of alternative structures is the entry point for the chemist who then must make the correct assignment from among them. Since this is a task at which the experienced chemist excels and usually completes relatively quickly, substantially increased productivity can be achieved. Furthermore, since SESAMI *exhaustively* generates structures compatible with the spectral data, the chemist has the assurance that no equally compatible structure has been overlooked.

SESAMI currently produces a set of isomers differing in *constitution*. Although software which discriminates between the enormous number of constitutional isomers corresponding to a given molecular formula can be of great value in structure elucidation, in practice an unknown is not fully characterized until its absolute configuration is ascertained. Thus, the enhancement of SESAMI's capabilities to include stereoisomer generation is a natural and necessary next step.

The output of the constrained structure generator of SESAMI is a set of connection tables (*2.0-D structures*) which describe the atom connectivity of the constitutional isomers produced.[7] Methods for both the *enumeration* of stereoisomers (i.e., calculation of the *number* of possible stereoisomers of a given constitutional isomer)[8-10] and the *generation* of stereoisomers (i.e., determination of the number of possible stereoisomers *and* the configuration (parity) at each stereocenter of each stereoisomer of a given constitutional isomer)[11-14] from atom connectivity information have been reported. Clearly, SESAMI's requirement is for the latter. The output of *stereoisomer generation* is a set of *2.5-D structures*, i.e., a set of connection tables augmented to include a parity symbol at each atom or set of atoms constituting a stereocenter. A *3.0-D structure* describes, in addition, the coordinates of each atom in space for a given stereoisomer and can be generated from a 2.5-D structure using model-building techniques.

Nourse was the first to consider the *generation* of stereoisomers from connectivity information only. The first step in his procedure[11,12] is the examination of the 2.0-D structure to identify all stereocenters. This is followed by stereoisomer generation, a process which is complicated only in the presence of molecular symmetry. Nourse's solution utilizes a new representation of symmetry called the configuration symmetry group. The more recent work of Zlatina and Elyashberg,[13] whose objective is the determination of an appropriate spatial model for each stereoisomer, combines the analysis of molecular topological symmetry, as developed by Nourse, with a consideration of geometric qualities. In Sasaki's approach,[14] stereocenter identification is followed by generation of *all* possible combinations of configurational parities in a process that uses defined "configuration molds". In a final step, a stereochemically unique name is assigned to each stereoisomer so generated, thereby permitting recognition and elimination of the redundancies arising from the presence of molecular symmetry. Wipke[15] and Balaban[16] have each developed computer-based methods for the general and unambiguous description of stereochemistry; the former was utilized by Sasaki.

No software for stereoisomer generation (2.5-D structures) is currently available. It is the purpose of this paper to detail

† National Institute of Chemistry.
‡ Arizona State University.
• Abstract published in *Advance ACS Abstracts*, November 1, 1993.

0095-2338/93/1633-0812$04.00/0

a *practical* approach to the creation of such software. Since the generation of 2.5-D structures does not require a knowledge of the spatial relationships of atoms or the constraints imposed by bonding considerations, chemically unstable isomers may be produced at this stage. Some "intelligence" has been added to eliminate some of these isomers (e.g., *E*-cyclohexene). More will be added at a later stage.

BASIC CONCEPTS

The generation of 2.5-D stereoisomers of constitutional isomers possessing no topological symmetry is straightforward. Complications arise in the presence of symmetry. In this section, pertinent definitions and concepts central to the development of a *general* computer-based approach to the generation of stereoisomers are presented.

Stereocenters. The origin of most stereoisomerism in chemical compounds can be described in terms of a small number of structural features, referred to herein as *stereocenters*.[17] The presence of such stereocenters is a *sufficient* condition for the existence of stereoisomers. In developing computer-based procedures for the generation of stereoisomers, the nature of the stereocenters to be considered must be precisely defined.

In organic compounds, carbon stereocenters are largely responsible for stereoisomerism. The discussion here, therefore, focuses on the properties of carbon, but its applicability to related elements (e.g., nitrogen) should be evident. In our approach, all non-hydrogen atoms and selected sets of atoms, e.g., atoms connected by double bonds, are considered as *potential stereocenters*. The program next classifies each as a *non-stereocenter*, *true-stereocenter* or *para-stereocenter*.

The process begins with the identification of all atoms with less than three neighbors (i.e., all mono and dicoordinate atoms) and atoms bound together by aromatic bonds. These are designated non-stereocenters and not considered further. Next, all tetracoordinate carbon atoms bearing four constitutionally different substituents, and all *sets* of two or more contiguously-joined, doubly-bonded carbon atoms (e.g., C=C, C=C=C), each terminus of which bears constitutionally different substituents, are classified as true-stereocenters. The remaining tetracoordinate carbon atoms and sets of contiguously-joined, doubly-bonded carbon atoms are examined for assignment of para-stereocenters. (We use the term "para-stereocenter" (para = resemble) to denote a structural feature which, although not a true-stereocenter, does influence the determination of the number and kinds of stereoisomers of a given compound in the procedure described herein.) Topological symmetry is a necessary condition for the existence of para-stereocenters, but does not ensure their presence in a structure. Para-stereocenters can be identified by means of the following rules:

1. Two or more potential stereocenters which are members of the same cycle are classified as para-stereocenters if the assemblage consists of sets of contiguously-joined, doubly-bonded carbon atoms and/or tetracoordinate carbon atoms, *and*

   a. if the terminal atom of each of these double bond systems that is the cycle member bears constitutionally identical substituents while the other terminal atom has either constitutionally different substituents or constitutionally identical substitutents with either one or more true-stereocenters or one or more potential stereocenters meeting these same conditions (i.e., rule 1), *and*

   b. if each of these cycle-membered, tetracoordinate carbon atoms bears two substituents which are either constitutionally different or constitutionally identical, but with each possessing either at least one true-stereocenter or one or more potential stereocenters meeting these same conditions (i.e., rule 1).

2. A tetracoordinate carbon atom is a para-stereocenter if

   a. it bears one or two pairs of constitutionally identical substituents, each of which contains at least one true-stereocenter or two parastereocenters as defined in rule 1 above, *or*

   b. it bears three or four constitutionally identical substituents, each of which contains at least two true-stereocenters or at least two *separate* assemblages of two or more para-stereocenters as defined in rule 1 above.

3. A set of two or more contiguously-joined, double-bonded carbon atoms is a para-stereocenter if one or both terminal atoms bear constitutionally identical substituents containing at least one true-stereocenter or two para-stereocenters as defined in rule 1 above.

After application of these rules, those atoms and sets of atoms not already classified as non-stereocenters, true-stereocenters, or para-stereocenters are added to the class of non-stereocenters.

There are noteworthy differences between para-stereocenters defined by rule 1 and those defined by rules 2 and 3. Rule 1 applies to *groups* of potential stereocenters whose stereogenicity is interdependent. Rules 2 and 3 deal with *individual* potential stereocenters. What has commonly been referred to in the past as a "pseudo-asymmetric" carbon atom is defined by rule 2. Different procedures are required for implementation of these rules. Rule 1 is applied recursively; rules 2 and 3 can be applied directly.

Although examples of non-stereocenters and true-stereocenters are easily envisaged, the structures in Figure 1 are of value in visualizing the nature of para-stereocenters.

(i) In structure a, two identified potential stereocenters are carbon–carbon double bonds (atoms 1–2 and 3–4). Atoms 1–2 comprise one para-stereocenter because one atom of the double bond is part of a cycle, while the other atom bears two different groups, and because there exists in the same cycle another double bond, atoms 3–4, that also meets the requirements for assignment as a para-stereocenter (rule 1a). The situation is similar in structure b except that the noncycle member of bond 1–2 bears two identical groups, each containing a true-stereocenter.

(ii) Carbon atoms 1 and 4 of structure c are defined as para-stereocenters in accord with rule 1b. Atom 1 bears two different substituents (H, OH), and there exists a second comparable atom (atom 4) in the same cycle. Structures d and e possess three and six para-stereocenters, respectively, based on the same rule.

(iii) Structure f requires both rules 1a and 1b and also illustrates the application of an important principle, that of *transitivity*. Atoms 1–2 and atom 4 are each para-stereocenters if atom 3 is one also. But, if atoms 1–2 and 4 are para-stereocenters, then it follows that atom 3 is indeed one (e.g., if atom 4 is a para-stereocenter, then

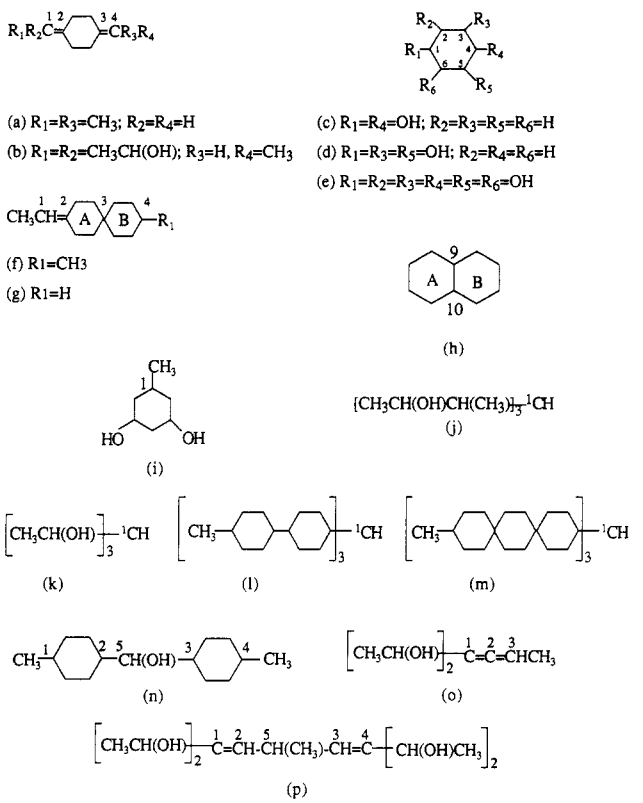


Figure 1. Examples of compounds containing para-stereocenters.

atom 3 is also one because atom 3 is a member of the same cycle as atom 2 and because, although atom 4 bears two identical groups (going around cycle B in either direction from atom 3 gives rise to exactly the same "path"), each possesses a para-stereocenter, atom 4. Application of this reasoning leads to the conclusion that structure (g) possesses no para-stereocenters. Since atom 4 now does not meet the requirements of rule 1, neither does atom 3, and therefore, neither do atoms 1–2. Thus, there are no stereocenters in structure g.

(iv) Atoms 9 and 10 of decalin (h) are assigned as para-stereocenters since each meets the conditions in rule 1b. They are clearly not true-stereocenters, but if h is viewed as a substituted cyclohexane (cycle A), then atoms 9 and 10 are part of the same cycle and each bears two different groups (one of which is the path of cycle B in each case). Thus, the conditions of rule 1b are met.

(v) Atoms 1 of structures i and j are each identified as para-stereocenters in accord with rules 2a and 2b, respectively. In structure i, the potential stereocenter, atom 1, is a member of a ring containing two true-stereocenters. The two identical substituents are part of the same ring. Atom 1 of structure k, in which the three identical substituents of atom 1 possess only one true-stereocenter, is assigned as a non-stereocenter.

(vi) Atom 1 of structure l, which bears three constitutionally identical substituents, each of which contains two *separate* assemblages of two para-stereocenters (for a total of four para-stereocenters), is a para-stereocenter (rule 2b). Atom 1 of structure m would not be so assigned. Each of the three identical substituents of atom 1 contains four para-stereocenters, as in structure l, but in this case the four form a *single* assemblage.

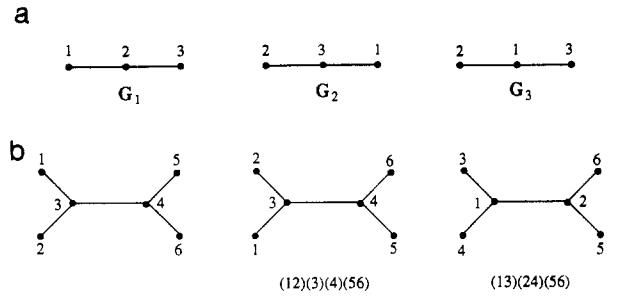


Figure 2. Graph permutation: (a) permutation notation; (b) automorphism group.

(vii) Structure n contains four atoms recognized as para-stereocenters (atoms 1–4) by the application of rule 1. Since atom 5 bears two identical substituents each of which contains two para-stereocenters, it too is assigned as a para-stereocenter based on rule 2a.

(viii) The atom set (1–2–3) of structure o is assigned as a para-stereocenter on the basis of rule 3.

(ix) Application of the rules leads to the detection of three para-stereocenters in structure p: atom 5 (rule 2) and atom sets 1–2 and 3–4 (rule 3).

Symmetry. The connectivity of a graph can be expressed as an adjacency matrix, **A**, whose elements, a_{ij}, are equal to 1 if vertices i and j are connected, and 0 if they are not. The automorphism (symmetry) group of a graph consists of permutations of the vertices such that the edge connections are preserved. Figure 2a shows three graphs which can be interconverted by permuting their vertices. The permutation $G_1 \rightarrow G_2$ can be symbolically represented in one of three ways:

(1 2 3)	1 2 3 2 3 1	2 3 1
cycle notation	standard mapping	shortened mapping

Cycle notation is extensively used in this paper, although the mapping notation also finds application.. In cycle notation, elements within parentheses are permuted in the order shown from left to right; i.e., vertex 1 becomes 2, 2 becomes 3, and 3 becomes 1. Cycle notation stresses that if an element returns to the beginning of the vector, the cycle is completed. A vertex that is not yet visited is the starting point to generate the next orbit in the cycle. For example, the permutation producing G_3 and G_1 (Figure 2a) is designated

(1 2)(3)

Standard mapping notation shows the original order of (identity) in the first row and the permuted elements in succeeding rows. Reading down in a column reveals the changes; i.e., 1 goes to 2, etc. The shortened mapping notation shows the new order imposed on a set by a permutation, but excludes the standard order of ascending numbering of the vertices. Consider a case where only the last two elements are interchanged within a set of twelve elements:

cycle notation:	(1) (2) (3) (4) (5) (6) (7) (8) (9) (10) (11 12)
standard mapping:	1 2 3 4 5 6 7 8 9 10 11 12 1 2 3 4 5 6 7 8 9 10 12 11
shrotened mapping:	1 2 3 4 5 6 7 8 9 10 12 11

This last notation is very practical for computer representation since any permutation of vertices can be represented by a vector.

The concept of automorphism can be illustrated with the graphs shown in Figure 2b. The permutation (1 2)(3)(4)(5 6) preserves all the original connections in that neither new connections are introduced nor existing ones destroyed, but the permutation (1 3)(2 4)(5 6) does not preserve the connectivity since it introduces a connection between vertices 1 and 2 which does not exist in the original graph. Likewise 3 and 4 are connected in the original graph and become disconnected in that permuted graph. Consequently, the permutation (1 2)(3)(4)(5 6) belongs to the automorphism group, while the permutation (1 3)(2 4)(5 6) does not.

The automorphism group of a graph is composed of *all* permutations of vertices which do not make or break any of the original connections between vertices. It can be defined in formal terms. To each permutation of vertices we can assign a matrix **P**, called the permutation matrix. The *i*th row of the permutation matrix contains all zeros except at the *j*th column if *i* goes to *j* in that permutation. In that case a "1" is entered. To illustrate, the permutation matrix for the permutation (1 2)(3 4)(5 6) is shown below

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

A permutation belongs to the automorphism group of a graph if and only if

$$\mathbf{PAP}^{-1} = \mathbf{A}$$

where $\mathbf{P}^{-1}$ is the inverse of the permutation matrix and **A** is the adjacency matrix of the graph in question.

The automorphism group of a graph is almost always a subgroup of S_n, where n is the number of vertices and S_n denotes the set of all $n!$ permutations of n objects. S_n *is* the automorphism group of a graph containing n vertices only if the graph is complete or trivial.[18] (A graph is complete if every vertex in the graph is connected to all other vertices of the graph. It is trivial if any vertex is not connected to any other vertex of the graph.) For other cases, the automorphism group is a subgroup of S_n.

At the present time, the algorithm for constructing the automorphism group of a compound (graph) is computationally intensive.[18-23] A totally "brute-force" approach involves checking all $n!$ permutations (where n is the number of vertices) for conservation of connectivity. However, the procedure used here is simplified by first carrying out the automorphism partitioning of the set of vertices of the graph.[24] Since only topologically equivalent vertices can be permuted within the automorphism group, only permutations within the automorphism partitioned sets need to be considered. For the graph in Figure 2b, automorphism partitioning leads to two sets of vertices, $Y_1 = \{1,2,5,6\}$ and $Y_2 = \{3,4\}$. Therefore, no permutation in the automorphism group can exchange vertices between sets Y_1 and Y_2. Consequently, prior automorphism partitioning simplifies a 6! (720 operations) problem to a 4! × 2! (48) problem. Now the brute-force approach would only require checking 48 permutations for conservation of connectivity. In practice, however, closure, inverse, and other properties of the group structure can be used to further reduce the number to be checked.

The compound shown in Figure 3 (compound **I**) is used to illustrate the step-by-step computer-based procedure for stereoisomer generation. The required automorphism group of that graph as generated by this procedure is shown as

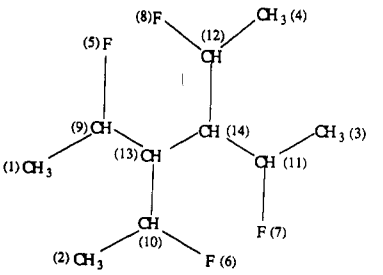


Figure 3. Structure (compound **I**) illustrating concepts and procedures of stereoisomer generation.

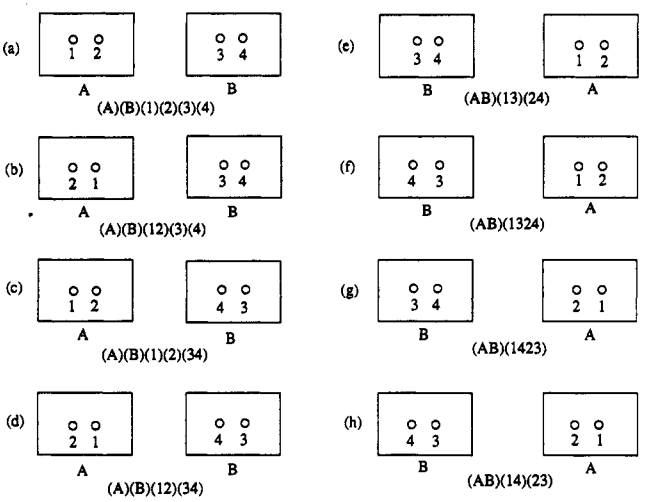


Figure 4. Particle-in-a-box illustration of wreath product $S_2[S_2]$.

follows: These eight operations and only these eight preserve the connectivity of the original graph shown in Figure 3. The first of the eight is the identify operation.

p_1 = (1)(2)(3)(4)(5)(6)(7)(8)(9)(10)(11)(12)(13)(14)

p_2 = (1 2)(3)(4)(5 6)(7)(8)(9 10)(11)(12)(13)(14)

p_3 = (1 2)(3 4)(5 6)(7 8)(9 10)(11 12)(13)(14)

p_4 = (1)(2)(3 4)(5)(6)(7 8)(9)(10)(11 12)(13)(14)

p_5 = (1 3)(2 4)(5 7)(6 8)(9 11)(10 12)(13 14)

p_6 = (1 4 2 3)(5 8 6 7)(9 12 10 11)(13 14)

p_7 = (1 4)(2 3)(5 8)(6 7)(9 12)(10 11)(13 14)

p_8 = (1 3 2 4)(5 7 6 8)(9 11 10 12)(13 14)

To ensure the validity of the computer-based procedure for constructing the automorphism group, a manual approach based on the wreath product formalism[9,25-28] was undertaken and the result was compared to the computer output. The discussion here also serves to provide additional insight into the nature of the automorphism group.

The wreath product of two groups G and H—written as $G[H]$—is illustrated for the simplest case of $S_2[S_2]$ (where S_2 is the group of all permutations of two objects) using the particles-in-a-box model shown in Figure 4. Each of the two boxes contains the same number of particles; two in this example. (In the event the boxes do not contain the same number of particles, a more complex procedure, the generalized wreath product, is required.[27] If G is the permutation group of the boxes A and B, and H is the permutation group of the

particles in each box, then the group of all permutations of particles in both boxes is given by the wreath product of G (the outer group) with H (the inner group). The number of permutations in the group $G[H]$ is

$$|G| \times |H|^n$$

where n is the number of boxes, $|G|$ is the number of elements in the group G and $|H|$ is the number of elements in the group H. For the case of two particles in each of two boxes, as in Figure 4, the group in the wreath product $S_2[S_2]$ contains $2 \times 2^2 = 8$ permutations. In Figure 4, the first four permutations (a–d) are generated by the action of the inner group H on the particles in the boxes, while the last four permutations (e–h) are generated by the action of G on the boxes followed by the action of H on the particles. It is important to note that group $G[H]$ is isomorphic to $(H_1 \times H_2 \times ... \times H_n) \cdot G$, where H_1, H_2, ..., H_n are copies of the same group H.

The two-particles-in-each-of-two-boxes model of Figure 4 is applicable to the construction of the automorphism group of the graph. The topologically equivalent vertices 13 and 14 can be the boxes and the vertices attached to 13 and 14, (1, 5, 2, 6, 9, 10) and (3, 7, 4, 8, 11, 12), respectively, the particles. It is evident that the permutations (1 2)(5 6)(9 10) and (3 4)(7 8)(11 12) each independently leave graph I invariant. Consequently, if the vertices 13 and 14 are mapped to boxes A and B, and the vertices attached to 13 and 14 are mapped to the particles in each box, the automorphism group of graph I is isomorphic with the group in Figure 4. In this case, the outer group is given by

$$G = \{(1)(2)(3)(4)(5)(6)(7)(8)(9)(10)(11)(12)(13)(14),$$
$$(1\ 4)(2\ 3)(5\ 8)(6\ 7)(9\ 12)(10\ 11)(13\ 14)\}$$

where the former corresponds to Figure 4a–d and the latter to Figure 4e–h, while the two copies of the inner group H are given by

$$H_1 = \{(1)(2)(5)(6)(9)(10), (1\ 2)(5\ 6)(9\ 10)\}$$

$$H_2 = \{(3)(4)(7)(8)(11)(12), (3\ 4)(7\ 8)(11\ 12)\}$$

where the two components of H_1 and H_2 correspond to the two particles in the boxes of Figure 4. The full expansion (multiplication of the wreath product group) gives exactly the same eight permutations as derived from the computer-implemented algorithm described earlier, thereby validating its output.

INTRODUCTION TO STEREOISOMER GENERATION

The number of stereoisomers in which a given constitutional isomer may exist can be related to the nature and number of stereocenters it possesses. The maximum possible number of stereoisomers (N) for a compound with n stereocenters is 2^n. This is achieved when all of the stereocenters are true-stereocenters and all are substituted differently. In the presence of topological symmetry, the number of stereoisomers may be less than 2^n. The extent of reduction is structure-dependent, and no general mathematical relationships are known which predict the number *and* kinds of stereoisomers in all cases.

Consider a molecule with n ($n > 1$) *different* true-stereocenters. Each stereocenter exists in one of two configurations which is accordingly assigned a parity value of +1 or −1. Thus, the configuration of a given stereoisomer can be expressed as a vector of length n consisting of +1 and/or −1 as elements. Mathematically, there are 2^n different vectors, the maximum number of stereoisomers possible (N). This information can be conveniently expressed as an $N \times n$ matrix which is called the *stereoparity matrix* (**M**) and takes the form

$$\mathbf{M} = \begin{array}{r|cccccccc} & 1 & 2 & 3 & 4 & \cdots & & & n \\ 1 & 1 & 1 & 1 & 1 & \cdots & 1 & 1 & 1 \\ 2 & 1 & 1 & 1 & 1 & \cdots & 1 & 1 & -1 \\ & 1 & 1 & 1 & 1 & \cdots & 1 & -1 & 1 \\ \vdots & 1 & 1 & 1 & 1 & \cdots & 1 & -1 & -1 \\ & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ N-1 & -1 & -1 & -1 & -1 & \cdots & -1 & -1 & 1 \\ N & -1 & -1 & -1 & -1 & \cdots & -1 & -1 & -1 \end{array}$$

The presence of topological symmetry in a compound with n stereocenters can lead to topological equivalence among true-stereocenters and/or the introduction of para-stereocenters and consequently, a number of stereoisomers less than N. In this case, some of the row vectors of the $N \times n$ stereoparity matrix are mathematically equivalent and, therefore, configurationally equivalent.

The automorphism group of such a compound is at the heart of establishing the equivalence of the configuration vectors. Although permutations of the automorphism group conserve connectivity of the graph of a compound, they do not necessarily conserve parity (configuration) at all stereocenters. Parity at true-stereocenters *is* conserved in all cases because they bear no topologically equivalent substituents. Therefore, the substituents cannot be interchanged by a permutation of the automorphism group. However, since para-stereocenters do bear equivalent substituents, which may be interchanged by a permutation, parity may not be conserved at these centers. The concepts of *stereoindex vectors* and *signed permutation matrices* provide a pragmatic means to follow if and where such changes occur and parallel Nourse's use of the configurational symmetry group.[11]

Consider a structure characterized by topological symmetry with n stereocenters. For each permutation p of the automorphism group of that structure, a *stereoindex vector* $\mathbf{S}^p$ is assigned. The stereoindex vector is a column vector containing n elements, each of which corresponds to one of the stereocenters. An element is given a value of +1 if the stereocenter to which it corresponds is a true-stereocenter. If the stereocenter is a para-stereocenter, its corresponding element in the vector is given a value of −1 if its *equivalent neighbors* are interchanged by the permutation p and a value of +1 if they are not interchanged.

Compound **I** serves to illustrate this concept. The automorphism group consists of the eight permutations described earlier in cycle notation. Only six of the vertices—9, 10, 11, 12, 13, 14—are stereocenters; therefore, the stereoindex vector for each of the eight permutations will contain six elements. Consider the *contracted* permutations p_2, p_3, p_6, and p_7 (i.e., minus non-stereocenter vertices) which are shown below in shortened mapping notation

$$\begin{aligned} p_1 &= 9\ 10\ 11\ 12\ 13\ 14 \\ p_2 &= 10\ 9\ 11\ 12\ 13\ 14 \\ p_3 &= 10\ 9\ 12\ 11\ 13\ 14 \\ p_6 &= 12\ 11\ 9\ 10\ 14\ 13 \\ p_7 &= 12\ 11\ 10\ 9\ 14\ 13 \end{aligned}$$

For each of these four permutations, the stereoindex vector is assigned a value of +1 for elements corresponding to the true stereocenters 9, 10, 11, and 12. The mapping notation clearly reveals that permutations p_2 and p_3 lead to an interchange of the equivalent neighbors 9 and 10 attached to para-stereocenter 13 (i.e., in p_2 and p_3, 9 and 10 are in

descending order reading left to right). Thus, element 13 in both S^{p_2} and S^{p_3}

$$S^{p_2} = \begin{array}{r|c} 9 & 1 \\ 10 & 1 \\ 11 & 1 \\ 12 & 1 \\ 13 & -1 \\ 14 & 1 \end{array} \quad S^{p_3} = \begin{vmatrix} 1 \\ 1 \\ 1 \\ 1 \\ -1 \\ -1 \end{vmatrix} \quad S^{p_6} = \begin{vmatrix} 1 \\ 1 \\ 1 \\ 1 \\ -1 \\ 1 \end{vmatrix} \quad S^{p_7} = \begin{vmatrix} 1 \\ 1 \\ 1 \\ 1 \\ -1 \\ -1 \end{vmatrix}$$

is assigned a value of −1. However, of these two permutations, only p_3 interchanges the equivalent neighbors 11 and 12 of para-stereocenter 14. Element 14, therefore, is assigned +1 and −1 for S^{p_2} and S^{p_3}, respectively. If the para-stereocenters themselves are interchanged in the permutations, as 13 and 14 are in p_6 and p_7, then the elements assigned to these vertices are also interchanged. Thus, for p_6, 9 and 10 are in natural order, but 11 and 12 are not. Elements 13 and 14 are accordingly assigned values of +1 and −1, respectively, but since these elements are interchanged by p_6, their positions in the stereoindex vector are also interchanged, as shown. The same interchange of the elements for vertices 13 and 14 is made in the stereoindex vector for p_7, but since both elements have the same sign, no actual change results. Note that the eight stereoindex vectors are not distinct since different permutations can yield the same stereoindex vector. For example, such is the case with permutations p_2 and p_6. The maximum number of possible stereoindex vectors for the permutations in an automorphism group is 2^m where m is the number of para-stereocenters. For I the maximum number of different stereoindex vectors is therefore 4.

The stereoindex vector is needed in the conversion of the permutation matrix to its corresponding *signed* permutation matrix. The process involves multiplying the mth row of the permutation matrix $\mathbf{P}_i$ (representing permutation p_i in the automorphism group) by +1 or −1, depending on the sign of the mth element of its stereoindex vector S^{p_i}. Consider the permutation p_2 of compound I. Eliminating non-stereocenters 1, 2, ..., 8, the permutation matrix is constructed as

$$\mathbf{P}_2 = \begin{array}{r|cccccc} & 9 & 10 & 11 & 12 & 13 & 14 \\ 9 & 0 & 1 & 0 & 0 & 0 & 0 \\ 10 & 1 & 0 & 0 & 0 & 0 & 0 \\ 11 & 0 & 0 & 1 & 0 & 0 & 0 \\ 12 & 0 & 0 & 0 & 1 & 0 & 0 \\ 13 & 0 & 0 & 0 & 0 & 1 & 0 \\ 14 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}$$

The corresponding stereoindex vector S^{p_2} is

$$S^{p_2} = \begin{vmatrix} 1 \\ 1 \\ 1 \\ 1 \\ -1 \\ 1 \end{vmatrix}$$

Hence the signed permutation matrix $\mathbf{P}_2^s$ is

$$\mathbf{P}_2^s = \begin{array}{r|cccccc} & 9 & 10 & 11 & 12 & 13 & 14 \\ 9 & 0 & 1 & 0 & 0 & 0 & 0 \\ 10 & 1 & 0 & 0 & 0 & 0 & 0 \\ 11 & 0 & 0 & 1 & 0 & 0 & 0 \\ 12 & 0 & 0 & 0 & 1 & 0 & 0 \\ 13 & 0 & 0 & 0 & 0 & -1 & 0 \\ 14 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}$$

The signed permutation matrix is the key to reveal equivalence between rows of the stereoparity matrix. Specifically, two rows of the stereoparity matrix, $\mathbf{r}_i$ and $\mathbf{r}_j$, are equivalent, i.e., represent the same configuration, if there exists a signed permutation matrix $\mathbf{P}^s$ such that

$$\mathbf{P}^s \mathbf{r}_i^T = \mathbf{r}_j^T$$

where $\mathbf{r}_i^T$ and $\mathbf{r}_j^T$ are transposes of row vectors in the stereoparity matrix to column vectors. To illustrate, consider the signed permutation matrix $\mathbf{P}_2^s$. The rows

1 1 1 1 1 1

and

1 1 1 1 -1 1

of the stereoparity matrix M that would be generated for a six-stereocenter compound (rows 1 and 3, respectively) are equivalent since the signed permutation matrix $\mathbf{P}_2^s$ converts the transpose of the former row vector to the transpose of the latter, i.e.,

$$\mathbf{P}_2^s \times \mathbf{r}_1^T = \mathbf{r}_3^T$$

$$\begin{vmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{vmatrix} \times \begin{vmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{vmatrix} = \begin{vmatrix} 1 \\ 1 \\ 1 \\ 1 \\ -1 \\ 1 \end{vmatrix}$$

Note that the converse leads to the same conclusion, i.e., $\mathbf{P}_2^s \times \mathbf{r}_3^T = \mathbf{r}_1^T$. In principle, each signed permutation matrix is applied to every row of the stereoparity matrix M to partition them into equivalence classes. (In practice, a smaller number of operations is required. See below.) Each of the equivalence classes represents one of the stereoisomers of the molecule.

STEREOISOMER GENERATION

In structures with only true-stereocenters, all of which are substituted differently, the generation of stereoisomers is a simple process: construct the $N \times n$ binary stereoparity matrix, each row of which is a representation of the configuration of one of the 2^n stereoisomers. In the presence of topological symmetry *among the stereocenters*, an additional step —partitioning the rows of the stereoparity matrix into equivalence classes—is required. This section details the process of stereoisomer generation for such a compound.

The *canonical* connection table (CCT), such as that currently produced as output by SESAMI, is the starting point for the process and the source of the required 2.0-D structural information. The program will accept arbitrarily numbered connection tables, but must be instructed to canonicalize them before proceeding. Figure 5a shows the CCT for compound I.

The perception of stereocenters is the initial step. In distinguishing stereocenters from non-stereocenters, atom and bond type information serves as an initial filter. Tetracoordinate carbon, silicon, germanium, and (positively charged) quaternary nitrogen, phosphorus, and arsenic are considered potential stereocenters, but tricoordinate nitrogen is not, due to its facile inversion. (Bridgehead tricoordinate nitrogen could act as a stereocenter but would not be identified as such by the program at this time.) Sets of two or more contiguous, doubly-bonded carbon atoms are also flagged as potential stereocenters, but sets of triply-bonded carbon atoms are not. The carbon–carbon double bonds of non-aromatic ring systems of less than eight members and atoms of aromatic or

a

Sequence Number	Atom Type	Neighboring Atoms				Corresponding Bond Type			
1	C	9	0	0	0	1	0	0	0
2	C	10	0	0	0	1	0	0	0
3	C	11	0	0	0	1	0	0	0
4	C	12	0	0	0	1	0	0	0
5	F	9	0	0	0	1	0	0	0
6	F	10	0	0	0	1	0	0	0
7	F	11	0	0	0	1	0	0	0
8	F	12	0	0	0	1	0	0	0
9	C	13	5	1	0	1	1	1	0
10	C	13	6	2	0	1	1	1	0
11	C	14	7	3	0	1	1	1	0
12	C	8	14	4	0	1	1	1	0
13	C	14	10	9	0	1	1	1	0
14	C	12	13	11	0	1	1	1	0

b

Assignment	Sequence Number
Non-stereocenter	1,2,3,4,5,6,7,8
Para-stereocenter	13,14
True-stereocenter	9,10,11,12

c

Stereocenter	Permutation							
	1	2	3	4	5	6	7	8
9	1	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1	1
11	1	1	1	1	1	1	1	1
12	1	1	1	1	1	1	1	1
13	1	-1	-1	1	1	-1	-1	1
14	1	1	-1	-1	1	1	-1	-1

d

NO. 1:

1	0	0	0	0	0
0	1	0	0	0	0
0	0	1	0	0	0
0	0	0	1	0	0
0	0	0	0	1	0
0	0	0	0	0	1

NO. 2:

0	1	0	0	0	0
1	0	0	0	0	0
0	0	1	0	0	0
0	0	0	1	0	0
0	0	0	0	-1	0
0	0	0	0	0	1

NO. 3:

0	1	0	0	0	0
1	0	0	0	0	0
0	0	0	1	0	0
0	0	1	0	0	0
0	0	0	0	-1	0
0	0	0	0	0	-1

NO. 4:

1	0	0	0	0	0
0	1	0	0	0	0
0	0	0	1	0	0
0	0	1	0	0	0
0	0	0	0	1	0
0	0	0	0	0	-1

NO. 5:

0	0	1	0	0	0
0	0	0	1	0	0
1	0	0	0	0	0
0	1	0	0	0	0
0	0	0	0	0	1
0	0	0	0	1	0

NO. 6:

0	0	0	1	0	0
0	0	1	0	0	0
1	0	0	0	0	0
0	1	0	0	0	0
0	0	0	0	0	-1
0	0	0	0	1	0

NO. 7:

0	0	0	1	0	0
0	0	1	0	0	0
0	1	0	0	0	0
1	0	0	0	0	0
0	0	0	0	0	-1
0	0	0	0	-1	0

NO. 8:

0	0	1	0	0	0
0	0	0	1	0	0
0	1	0	0	0	0
1	0	0	0	0	0
0	0	0	0	0	1
0	0	0	0	-1	0

e

Vector Row	9	10	11	12	13	14
1	1	1	1	1	1	1
2	1	1	1	1	1	-1
3	1	1	1	1	-1	1
4	1	1	1	1	-1	-1
5	1	1	1	-1	1	1
6	1	1	1	-1	1	-1
7	1	1	1	-1	-1	1
8	1	1	1	-1	-1	-1
9	1	1	-1	1	1	1
10	1	1	-1	1	1	-1
11	1	1	-1	1	-1	1
12	1	1	-1	1	-1	-1
13	1	1	-1	-1	1	1
14	1	1	-1	-1	1	-1
15	1	1	-1	-1	-1	1
16	1	1	-1	-1	-1	-1
17	1	-1	1	1	1	1
18	1	-1	1	1	1	-1
19	1	-1	1	1	-1	1
20	1	-1	1	1	-1	-1
21	1	-1	1	-1	1	1
22	1	-1	1	-1	1	-1
23	1	-1	1	-1	-1	1
24	1	-1	1	-1	-1	-1
25	1	-1	-1	1	1	1
26	1	-1	-1	1	1	-1
27	1	-1	-1	1	-1	1
28	1	-1	-1	1	-1	-1
29	1	-1	-1	-1	1	1
30	1	-1	-1	-1	1	-1
31	1	-1	-1	-1	-1	1
32	1	-1	-1	-1	-1	-1

Vector Row	9	10	11	12	13	14
33	-1	1	1	1	1	1
34	-1	1	1	1	1	-1
35	-1	1	1	1	-1	1
36	-1	1	1	1	-1	-1
37	-1	1	1	-1	1	1
38	-1	1	1	-1	1	-1
39	-1	1	1	-1	-1	1
40	-1	1	1	-1	-1	-1
41	-1	1	-1	1	1	1
42	-1	1	-1	1	1	-1
43	-1	1	-1	1	-1	1
44	-1	1	-1	1	-1	-1
45	-1	1	-1	-1	1	1
46	-1	1	-1	-1	1	-1
47	-1	1	-1	-1	-1	1
48	-1	1	-1	-1	-1	-1
49	-1	-1	1	1	1	1
50	-1	-1	1	1	1	-1
51	-1	-1	1	1	-1	1
52	-1	-1	1	1	-1	-1
53	-1	-1	1	-1	1	1
54	-1	-1	1	-1	1	-1
55	-1	-1	1	-1	-1	1
56	-1	-1	1	-1	-1	-1
57	-1	-1	-1	1	1	1
58	-1	-1	-1	1	1	-1
59	-1	-1	-1	1	-1	1
60	-1	-1	-1	1	-1	-1
61	-1	-1	-1	-1	1	1
62	-1	-1	-1	-1	1	-1
63	-1	-1	-1	-1	-1	1
64	-1	-1	-1	-1	-1	-1

f

STEREOPARITY MATRIX ROW 1
VALID STEREOISOMER

$r_1 = 1\ 1\ 1\ 1\ 1\ 1$

$r_1^T \cdot P_2^s = 1\ 1\ 1\ 1\ -1\ 1 = r_3$
$r_1^T \cdot P_3^s = 1\ 1\ 1\ 1\ -1\ -1 = r_4$
$r_1^T \cdot P_4^s = 1\ 1\ 1\ 1\ 1\ -1 = r_2$
$r_1^T \cdot P_5^s = 1\ 1\ 1\ 1\ 1\ 1 = r_1$
$r_1^T \cdot P_6^s = 1\ 1\ 1\ 1\ -1\ 1 = r_3$
$r_1^T \cdot P_7^s = 1\ 1\ 1\ 1\ -1\ -1 = r_4$
$r_1^T \cdot P_8^s = 1\ 1\ 1\ 1\ 1\ -1 = r_2$

VALID STEREOISOMER(S)

$E_1 = 1\ 1\ 1\ 1\ 1\ 1 = r_1$
$E_{-1} = -1\ -1\ -1\ -1\ -1\ -1 = r_{64}$

REDUNDANT STEREOISOMER(S):
r_2/r_{63}; r_3/r_{62}; r_4/r_{61}

STEREOPARITY MATRIX ROW 5
VALID STEREOISOMER

$r_5 = 1\ 1\ 1\ -1\ 1\ 1$

$r_5^T \cdot P_2^s = 1\ 1\ 1\ -1\ -1\ 1 = r_7$
$r_5^T \cdot P_3^s = 1\ 1\ -1\ 1\ -1\ -1 = r_{12}$
$r_5^T \cdot P_4^s = 1\ 1\ -1\ 1\ 1\ -1 = r_{10}$
$r_5^T \cdot P_5^s = 1\ -1\ 1\ 1\ 1\ 1 = r_{17}$
$r_5^T \cdot P_6^s = -1\ 1\ 1\ 1\ -1\ 1 = r_{35}$
$r_5^T \cdot P_7^s = -1\ 1\ 1\ 1\ -1\ -1 = r_{36}$
$r_5^T \cdot P_8^s = 1\ -1\ 1\ 1\ 1\ -1 = r_{18}$

VALID STEREOISOMER(S)

$E_2 = 1\ 1\ 1\ -1\ 1\ 1 = r_5$
$E_{-2} = -1\ -1\ -1\ 1\ -1\ -1 = r_{60}$

REDUNDANT STEREOISOMER(S):
r_7/r_{58}; r_{10}/r_{55}; r_{12}/r_{53}; r_{17}/r_{48}; r_{18}/r_{47}; r_{29}/r_{36}; r_{30}/r_{35}

STEREOPARITY MATRIX ROW 6
VALID STEREOISOMER

$r_6 = 1\ 1\ 1\ -1\ 1\ -1$

$r_6^T \cdot P_2^s = 1\ 1\ 1\ -1\ -1\ -1 = r_8$
$r_6^T \cdot P_3^s = 1\ 1\ -1\ 1\ -1\ 1 = r_{11}$
$r_6^T \cdot P_4^s = 1\ 1\ -1\ 1\ 1\ 1 = r_9$
$r_6^T \cdot P_5^s = 1\ -1\ 1\ 1\ -1\ 1 = r_{19}$
$r_6^T \cdot P_6^s = -1\ 1\ 1\ 1\ 1\ 1 = r_{33}$
$r_6^T \cdot P_7^s = -1\ 1\ 1\ 1\ 1\ -1 = r_{34}$
$r_6^T \cdot P_8^s = 1\ -1\ 1\ 1\ -1\ -1 = r_{20}$

VALID STEREOISOMER(S)

$E_3 = 1\ 1\ 1\ -1\ 1\ -1 = r_6$
$E_{-3} = -1\ -1\ -1\ 1\ -1\ 1 = r_{59}$

REDUNDANT STEREOISOMER(S):
r_8/r_{57}; r_9/r_{56}; r_{11}/r_{54}; r_{19}/r_{46}; r_{20}/r_{45}; r_{31}/r_{34}; r_{32}/r_{33}

STEREOPARITY MATRIX ROW 13
VALID STEREOISOMER

$r_{13} = 1\ 1\ -1\ -1\ 1\ 1$

$r_{13}^T \cdot P_2^s = 1\ 1\ -1\ -1\ -1\ 1 = r_{15}$
$r_{13}^T \cdot P_3^s = 1\ 1\ -1\ -1\ -1\ -1 = r_{16}$
$r_{13}^T \cdot P_4^s = 1\ 1\ -1\ -1\ 1\ -1 = r_{14}$
$r_{13}^T \cdot P_5^s = -1\ -1\ 1\ 1\ 1\ 1 = r_{49}$
$r_{13}^T \cdot P_6^s = -1\ -1\ 1\ 1\ -1\ 1 = r_{51}$
$r_{13}^T \cdot P_7^s = -1\ -1\ 1\ 1\ -1\ -1 = r_{52}$
$r_{13}^T \cdot P_8^s = -1\ -1\ 1\ 1\ 1\ -1 = r_{50}$

VALID STEREOISOMER(S)

$D_1 = 1\ 1\ -1\ -1\ 1\ 1 = r_{13}$

REDUNDANT STEREOISOMER(S):
r_{14}/r_{51}; r_{15}/r_{50}; r_{16}/r_{49}; r_{52}

STEREOPARITY MATRIX ROW 21
VALID STEREOISOMER

$r_{21} = 1\ -1\ 1\ -1\ 1\ 1$

$r_{21}^T \cdot P_2^s = -1\ 1\ 1\ -1\ -1\ 1 = r_{39}$
$r_{21}^T \cdot P_3^s = -1\ 1\ -1\ 1\ -1\ -1 = r_{44}$
$r_{21}^T \cdot P_4^s = 1\ -1\ -1\ 1\ 1\ -1 = r_{26}$
$r_{21}^T \cdot P_5^s = 1\ -1\ 1\ -1\ 1\ 1 = r_{21}$
$r_{21}^T \cdot P_6^s = -1\ 1\ 1\ -1\ -1\ 1 = r_{39}$
$r_{21}^T \cdot P_7^s = -1\ 1\ -1\ 1\ -1\ -1 = r_{44}$
$r_{21}^T \cdot P_8^s = 1\ -1\ -1\ 1\ 1\ -1 = r_{26}$

VALID STEREOISOMER(S)

$D_2 = 1\ -1\ 1\ -1\ 1\ 1 = r_{21}$

REDUNDANT STEREOISOMER(S):
r_{26}/r_{39}; r_{44}

STEREOPARITY MATRIX ROW 22
VALID STEREOISOMER

$r_{22} = 1\ -1\ 1\ -1\ 1\ -1$

$r_{22}^T \cdot P_2^s = -1\ 1\ 1\ -1\ -1\ -1 = r_{40}$
$r_{22}^T \cdot P_3^s = -1\ 1\ -1\ 1\ -1\ 1 = r_{43}$
$r_{22}^T \cdot P_4^s = 1\ -1\ -1\ 1\ 1\ 1 = r_{25}$
$r_{22}^T \cdot P_5^s = 1\ -1\ 1\ -1\ -1\ 1 = r_{23}$
$r_{22}^T \cdot P_6^s = -1\ 1\ 1\ -1\ 1\ 1 = r_{37}$
$r_{22}^T \cdot P_7^s = -1\ 1\ -1\ 1\ 1\ -1 = r_{42}$
$r_{22}^T \cdot P_8^s = 1\ -1\ -1\ 1\ -1\ -1 = r_{28}$

VALID STEREOISOMER(S)

$D_3 = 1\ -1\ 1\ -1\ 1\ -1 = r_{22}$

REDUNDANT STEREOISOMER(S):
r_{23}/r_{42}; r_{25}/r_{40}; r_{28}/r_{37}; r_{43}

STEREOPARITY MATRIX ROW 24
VALID STEREOISOMER

$r_{24} = 1\ -1\ 1\ -1\ -1\ -1$

$r_{24}^T \cdot P_2^s = -1\ 1\ 1\ -1\ 1\ -1 = r_{38}$
$r_{24}^T \cdot P_3^s = -1\ 1\ -1\ 1\ 1\ 1 = r_{41}$
$r_{24}^T \cdot P_4^s = 1\ -1\ -1\ 1\ -1\ 1 = r_{27}$
$r_{24}^T \cdot P_5^s = 1\ -1\ 1\ -1\ -1\ -1 = r_{24}$
$r_{24}^T \cdot P_6^s = -1\ 1\ 1\ -1\ 1\ -1 = r_{38}$
$r_{24}^T \cdot P_7^s = -1\ 1\ -1\ 1\ 1\ 1 = r_{41}$
$r_{24}^T \cdot P_8^s = 1\ -1\ -1\ 1\ -1\ 1 = r_{27}$

VALID STEREOISOMER(S)

$D_4 = 1\ -1\ 1\ -1\ -1\ -1 = r_{24}$

REDUNDANT STEREOISOMER(S):
r_{27}/r_{38}; r_{41}

SUMMARY

TOTAL STEREOISOMERS	10
CHIRAL	6
ACHIRAL	4

Figure 5. Stereoisomer generation in presence of para-stereocenters: (a) canonical connection table of compound **I**, Figure 3; (b) stereocenter assignments; (c) stereoindex vectors; (d) signed permutation matrices ($\mathbf{P}_i^s$); (e) stereoparity matrix (**M**); (f) summary of process of matrix–vector multiplication.

anti-aromatic rings are designated as non-stereocenters. Non-aromatic, polycyclic systems are examined to distinguish atoms common to two or more cycles, as found, for example, in spiro, fused, and bridged systems.

Next, each stereocenter is classified as either *true* or *para* using algorithms derived from their definitions. As expected from these definitions, cycle perception is an important component of this step.[29] The assignments for the atoms of compound I are summarized in Figure 5b. This information is carried internally by augmenting the CCT. Recognition of the topological equivalence of atoms and groups of atoms, which is central to the process, is facilitated by the automorphism partitioning of the atoms. In particular, an atom-by-atom comparison of substituents needs to be undertaken only when the atoms immediately adjacent to the sites bearing the substituents belong to the same topological equivalence class. The atom-by-atom search is necessary since the substituents, all of whose atoms and bonds must be identical, may include additional stereocenters that can affect "equivalence".

Generating the stereoisomers of a constitutional isomer cannot depend in any way on the spatial arrangements of the atoms since that information is not known. Thus, initially, symmetry considerations are limited to molecular topology. The most informative symmetry description at this level is given by the vertex automorphism group of the "chemical graph". The algorithmic procedure currently implemented includes the following steps:

1. All possible permutations within automorphism orbits (equivalence classes) are generated.
2. All possible interorbit combinations of orbit vertex permutations are obtained.
3. Vertex permutations are transformed into binary permutation matrices ($\mathbf{P}_n$).
4. Permutation $\mathbf{P}_i$ is tested to determine if $\mathbf{P}_i\mathbf{A}_i\mathbf{P}_i^{-1} = \mathbf{A}_i$. Permutation $\mathbf{P}_i$ belongs to the automorphism group only if that condition is met.

This procedure gives rise to the eight permutations of the automorphism group shown earlier in cycle notation. Stereoindex vectors (Figure 5c) are next assigned by a routine that uses the rules described earlier. Since atoms 1 through 8 are non-stereocenters, they require no stereochemical assignment and play no role in subsequent steps of the procedure. Deleting them simplifies and speeds up the algorithm procedure. Thus, the permutation matrices are contracted to 6 × 6 and the stereoindex vectors contain only 6 elements. Next, each row of each permutation matrix is multiplied by the corresponding element of its stereoindex vector to produce the set of 6 × 6 signed permutation matrices (Figure 5d).

To establish the equivalency that exists among the parity row vectors of the computer-generated, 64 × 6 stereoparity matrix (Figure 5e), the column vector transpose ($\mathbf{r}_i^T$) of each row vector ($\mathbf{r}_i$) is multiplied by each of the seven significant signed permutation matrices ($\mathbf{P}_2^s - \mathbf{P}_8^s$; multiplication by the signed permutation matrix $\mathbf{P}_1^s$ produces no change in any vector and, hence, is not executed). The product of multiplication, $\mathbf{r}_j^T$, is equivalent to $\mathbf{r}_i^T$, and therefore the corresponding row vector $\mathbf{r}_j$ is *redundant*. In practice, it is not necessary to examine the action of each signed permutation matrix on each of the 64 stereoparity matrix vectors since, once an equivalence is established, the redundant stereoparity matrix vector is eliminated from further consideration.

To begin, the first row vector (1 1 1 1 1 1) is designated a "valid stereoisomer". Multiplication of its transpose ($\mathbf{r}_1^T$) by signed permutation matrices $\mathbf{P}_2^s - \mathbf{P}_8^s$ yields the following result:

$$\mathbf{P}_2^s\cdot\mathbf{r}_1^T = 1\ 1\ 1\ 1\ -1\ \ 1 = \mathbf{r}_3$$
$$\mathbf{P}_3^s\cdot\mathbf{r}_1^T = 1\ 1\ 1\ 1\ -1\ -1 = \mathbf{r}_4$$
$$\mathbf{P}_4^s\cdot\mathbf{r}_1^T = 1\ 1\ 1\ 1\ \ 1\ -1 = \mathbf{r}_2$$
$$\mathbf{P}_5^s\cdot\mathbf{r}_1^T = 1\ 1\ 1\ 1\ \ 1\ \ 1 = \mathbf{r}_1$$
$$\mathbf{P}_6^s\cdot\mathbf{r}_1^T = 1\ 1\ 1\ 1\ -1\ \ 1 = \mathbf{r}_3$$
$$\mathbf{P}_7^s\cdot\mathbf{r}_1^T = 1\ 1\ 1\ 1\ -1\ -1 = \mathbf{r}_4$$
$$\mathbf{P}_8^s\cdot\mathbf{r}_1^T = 1\ 1\ 1\ 1\ \ 1\ -1 = \mathbf{r}_2$$

Thus, stereoparity matrix vectors $\mathbf{r}_2$, $\mathbf{r}_3$, and $\mathbf{r}_4$ represent redundancies (i.e., stereoisomers 1–4 are members of the same configuration equivalence class). These three vectors are no longer considered in the matrix multiplication step of the algorithmic procedure. A stepwise record of the matrix multiplication as output by the program is shown in Figure 5f.

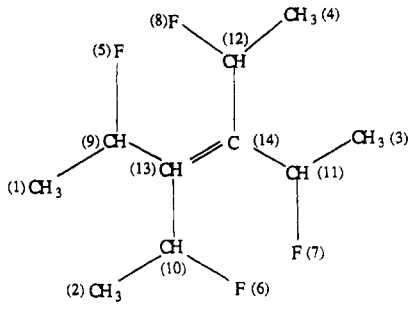
Upon completion of the multiplication of a given row vector of the stereoparity matrix by all seven signed permutation matrices, the algorithm identifies that stereoisomer as either an *enantiomer* (E) or a *diastereomer* (D). In the case of $\mathbf{r}_1$ (1 1 1 1 1 1), its parity "mirror image" is $\mathbf{r}_{64}$ (−1 −1 −1 −1 −1 −1). Therefore, the configurations represented by rows 1 and 64 are said to be enantiomerically related; i.e., a pair of enantiomers has been generated. Note that the configuration represented by vector 64 is indeed *different* than that of vector 1 since vector 64 is not one of the *redundancies* produced in the matrix–vector multiplication step. Since configuration vectors 61, 62, and 63 are parity mirror images of vectors 4, 3, and 2, respectively, they must be configurationally equivalent to 64 and as such are eliminated by the program from further consideration as well.

The algorithm then proceeds to the next valid row in the stereoparity matrix, $\mathbf{r}_5$, and commences multiplication by each of the seven signed permutation matrices. A second enantiomeric pair is generated—configuration vector 5 and its mirror image, vector 60—and redundancies 7, 10, 12, 17, 18, 29, 30, 58, 55, 53, 48, 47, 36, and 35 (the latter seven being enantiomerically related, respectively, to the former seven) are identified. The process resumes with row 6 of the stereoparity matrix which gives rise to the third and final enantiomeric pair to be produced (vectors 6 and 59).

When the next valid vector, row 13 (1 1 −1 −1 1 1), is processed, the step in which it is multiplied by signed permutation matrix $\mathbf{P}_5^s$ gives rise to vector 52 (−1 −1 1 1 −1 −1) which is consequently deleted from further consideration. However, since this equivalent vector is also the parity mirror image of vector 13, the enantiomer of 13 will not be found in the search that follows completion of the vector–matrix multiplication step. The program considers a stereoisomer whose enantiomer is not found to be a diastereomer. In the case of compound I considered here, the diastereomer is actually a *meso* form. The remaining valid rows that are processed—21, 22, and 24—also give rise to diastereomers. Thus, a total of 10 stereoisomers—three enantiomeric pairs and four *meso* forms—are correctly predicted for the compound.

The procedure for structures with stereocenters comprised of contiguously-joined, doubly-bonded carbon atoms is similar but not identical to that for compounds without such features. It is illustrated in Figure 6 for compound II (Figure 6a). The C=C unit is considered to be a *single* element for stereocenter assignment. Thus, this compound has five rather than six stereocenters. Atoms 9–12 are true-stereocenters, while the C=C unit (atoms 13 and 14) is a single para-stereocenter.

a



b

Sequence	Atom Type	Neighboring Atoms				Corresponding Bond Type			
1	C	9	0	0	0	1	0	0	0
2	C	10	0	0	0	1	0	0	0
3	C	11	0	0	0	1	0	0	0
4	C	12	0	0	0	1	0	0	0
5	F	9	0	0	0	1	0	0	0
6	F	10	0	0	0	1	0	0	0
7	F	11	0	0	0	1	0	0	0
8	F	12	0	0	0	1	0	0	0
9	C	5	13	1	0	1	1	1	0
10	C	2	6	13	0	1	1	1	0
11	C	3	7	14	0	1	1	1	0
12	C	4	8	14	0	1	1	1	0
13/14	C	10	9	14	0	1	1	2	0
	C	12	11	13	0	1	1	2	0

c

Stereocenter	Permutation							
	1	2	3	4	5	6	7	8
9	1	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1	1
11	1	1	1	1	1	1	1	1
12	1	1	1	1	1	1	1	1
13/14	1	-1	1	-1	1	-1	1	-1

d

NO. 1:

1	0	0	0	0
0	1	0	0	0
0	0	1	0	0
0	0	0	1	0
0	0	0	0	1

NO. 2:

0	1	0	0	0
1	0	0	0	0
0	0	1	0	0
0	0	0	1	0
0	0	0	0	-1

NO. 3:

0	1	0	0	0
1	0	0	0	0
0	0	0	1	0
0	0	1	0	0
0	0	0	0	1

NO. 4:

1	0	0	0	0
0	1	0	0	0
0	0	0	1	0
0	0	1	0	0
0	0	0	0	-1

NO. 5:

0	0	1	0	0
0	0	0	1	0
1	0	0	0	0
0	1	0	0	0
0	0	0	0	1

NO. 6:

0	0	0	1	0
0	0	1	0	0
1	0	0	0	0
0	1	0	0	0
0	0	0	0	-1

NO. 7:

0	0	0	1	0
0	0	1	0	0
0	1	0	0	0
1	0	0	1	0
0	0	0	0	1

NO. 8:

0	0	1	0	0
0	0	0	1	0
0	1	1	0	0
1	0	0	0	0
0	0	0	0	-1

e

Vector Row	9	10	11	12	13/14
1	1	1	1	1	1
2	1	1	1	1	-1
3	1	1	1	-1	1
4	1	1	1	-1	-1
5	1	1	-1	1	1
6	1	1	-1	1	-1
7	1	1	-1	-1	1
8	1	1	-1	-1	-1
9	1	-1	1	1	1
10	1	-1	1	1	-1
11	1	-1	1	-1	1
12	1	-1	1	-1	-1
13	1	-1	-1	1	1
14	1	-1	-1	1	-1
15	1	-1	-1	-1	1
16	1	-1	-1	-1	-1

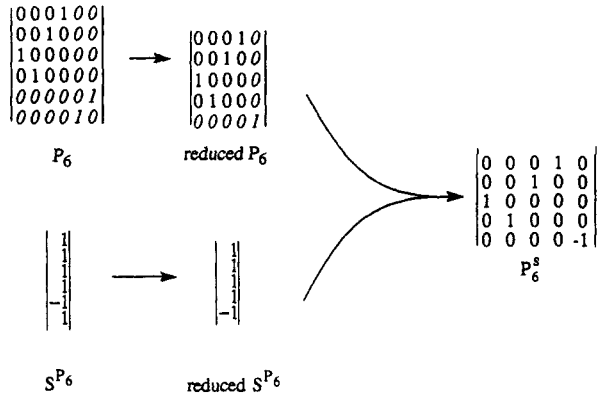
Vector Row	9	10	11	12	13/14
17	-1	1	1	1	1
18	-1	1	1	1	-1
19	-1	1	1	-1	1
20	-1	1	1	-1	-1
21	-1	1	-1	1	1
22	-1	1	-1	1	-1
23	-1	1	-1	-1	1
24	-1	1	-1	-1	-1
25	-1	-1	1	1	1
26	-1	-1	1	1	-1
27	-1	-1	1	-1	1
28	-1	-1	1	-1	-1
29	-1	-1	-1	1	1
30	-1	-1	-1	1	-1
31	-1	-1	-1	-1	1
32	-1	-1	-1	-1	-1

Figure 6. Stereoisomer generation in presence of multiatom stereocenters: (a) structure of compound **II**; (b) canonical connection table; (c) stereoindex vectors; (d) signed permutation matrices ($\mathbf{P}_i^s$); (e) stereoparity matrix (**M**).

The CCT (Figure 6b) differs from that of compound **I** only in the multiplicity of the bond joining atoms 13 and 14. The permutations of the automorphism group are *identical* to those of the previous problem (Figure 5c). The six-element stereoindex vectors (Figure 5d) derived from these permutation vectors are algorithmic precursors of the five-element stereoindex vectors. Multiplication of elements 5 and 6 (elements for atoms 13 and 14) of each of these eight stereoindex vectors gives the *reduced* stereoindex vectors used in this problem (Figure 6c). Note that now there are only two distinct stereoindex vectors.

The *signed* permutation matrices necessary in the pruning of the stereoparity matrix first require reduction of the *normal* 6 × 6 *permutation* matrices to 5 × 5. The last two rows/columns, corresponding to the terminal atoms of the para-stereocenter in each case (atoms 13 and 14), are replaced by a single row/column (row/column "5/6") in the permutation matrix. In this "new" row/column, the fifth row/column of the reduced matrix, *only* the fifth element is assigned a "1". This is true for all five permutation matrices since no permutation of the automorphism group will exchange vertices 13 and 14 with those of the other stereocenters (9–12).

Multiplying the mth row of each reduced permutation matrix by ±1 depending on the sign of the mth element of its reduced stereoindex vector produces the set of signed 5 × 5 permutation matrices (Figure 6d). The process of generating the signed permutation matrices is illustrated below starting with the (6 × 6) permutation matrix $\mathbf{P}_6$ corresponding to permutation p_6.



Note that reduction of the permutation matrix $\mathbf{P}_6$ affects *only* elements corresponding to the para-stereocenter and that, in the process, rows/columns 5 and 6 are replaced by a single row/column of the form (0, 0, 0, 0, 1).

Consistent with the above, a 32 × 5 stereoparity matrix is generated for compound **II** (Figure 6e). The process of vector–matrix multiplication to identify configurational equivalences follows that described for the previous example, with one important difference. In the earlier example, one parity vector (configuration) of the stereoparity matrix was designated the enantiomer of another if it was its exact parity inversion (e.g. 1 1 1 1 1 1 and −1 −1 −1 −1 −1 −1). In the present case, where one "element" serves to represent the set of the two atoms of a carbon–carbon double bond, the following rule applies: one parity vector is the enantiomer of another if the parities of all

"single atom" stereocenters are inverted, but all "two-atom" stereocenters have the same parity.

To illustrate, multiplication of the transpose of row vector r_1 (1 1 1 1 1) by each of the seven signed permutation matrices identifies only parity vector r_2 as equivalent in configuration to r_1. The algorithm next selects parity vector r_{31} (−1 −1 −1 −1 1) as enantiomeric to vector r_1. Parity vector r_2 and its enantiomer r_{32} are eliminated from further consideration. Parity vectors r_3 and r_{29} are next identified as a second enantiomeric pair, and the remaining "surviving" vectors 7, 11, and 12 are identified as diastereomers since their "enantiomers" (vectors 20, 21, and 25, respectively) are generated by multiplication by the signed permutation matrices. Thus, the program correctly predicts seven stereoisomers—two pairs of enantiomers and three *meso* forms—for this compound.

Compounds with stereocenters consisting of more than two, contiguously-joined, doubly-bonded carbon atoms (i.e., cumulenes) are treated similarly. The stereoindex vector value of the stereocenter is the product of the values of is terminal atoms; intervening atoms (non-stereocenters) are dropped from consideration. The permutation matrices are likewise reduced such that only the terminal atoms are represented in the matrix, and by a *single* row/column. That row/column will be the same in all permutation matrices unless there is a topologically equivalent multiple bond unit in the structure with which "exchange" occurs in the automorphism group.

The rule for identifying an enantiomeric relationship between two stereoparity vectors varies with the number of double bonds in the cumulene unit. For an *odd* number of double bonds (e.g., compound **32**, Figure 7), the rule is the same as that for a single carbon–carbon double bond (Figure 6). For an *even* number of double bonds (e.g., compound **31**, Figure 7), the rule is the same as that for a single atom stereocenter (Figure 5). Thus, although the two stereoparity vectors surviving the matrix multiplication step are exactly the same for compound **1**, **30**, **31**, **32**, and **33** (+1 for one, −1 for the other), the program correctly distinguishes enantiomeric and diastereomeric relationships because of this rule.

RESULTS AND DISCUSSION

The program for exhaustive, irredundant stereoisomer generation was tested extensively with a broad spectrum of structure classes, each of whose stereoisomerism can be related to the presence of *stereocenters* that can be identified from the 2.0-D connection table. Compound types without such stereocenters can still exhibit stereoisomerism, e.g., biphenyls (astropisomers), helical structures, catenanes, and knots,[30] but these would not be so identified by the program at this time.

The 69 compounds against which the program was tested are shown in Figure 7. The number and kinds of stereoisomers generated by the program for each one are summarized in Table I. The chiral stereoisomers exist as pairs of enantiomers; e.g., the eight stereoisomers of compound **4** are related as four pairs of enantiomers. If there is more than one achiral stereoisomer of any given compound, they are related as diastereomers of one another. Achiral stereoisomers are *meso* forms if at least two of the stereoisomers of the compound are chiral; otherwise, they are designated as diastereomers. Thus, compound **3** exists as six pairs of enantiomers and four *meso* forms; compounds **22** and **35** exist as 4 and 3 diastereomers, respectively, each of which is commonly referred to as a geometric or *cis–trans* isomer.

Currently, the program does not identify strained, unstable stereoisomers. Thus, the program identifies eight stereoisomers (four pairs of enantiomers) for compound **55** when in reality only four will exist under normal conditions. The four stereoisomers with a *trans*-linked, one-carbon bridge across the 1 and 4 positions of the six-membered ring are too strained to exist.

The program correctly identifies the stereocenters in compounds *without* symmetry (**1**, **10**, **12**, **55**, **56**, **61**, **62**, **63**) and recognizes $2^n/2$ pairs of enantiomers in each case. The number and kinds of stereoisomers of symmetrical acyclic compounds with both even (e.g., **2**) and odd (e.g., **3**) numbers of stereocenters are well-known[30] and were correctly predicted. As the degree of topological symmetry increases in a series of structurally-related compounds (e.g., **10** → **9** → **8** → **7** → **6**), the number of possible stereoisomers decreases.

Odd- and even-membered saturated monocyclic systems with and without para-stereocenters (**14–29**) give answers that are "correct" at normal temperatures. Thus, compound **16** with two true-stereocenters is predicted to exist as a pair of enantiomers and one *meso* form, consistent with observation. This result corresponds to that obtained by "manually" evaluating all possible configurations of this compound in which the six-membered ring is considered to be planar even though it is known to exist in the chair conformation.[30] (The isomer designated a *meso* form is actually a pair of rapidly equilibrating (racemizing) enantiomers and, therefore, displays no chirality. In a very low temperature world, four stable stereoisomers of compound **16**, each displaying chirality, could be isolated.) The isomeric compound **18** with two para-stereocenters is correctly predicted to exist as two diastereomers.

The treatment of the terminal atoms of a set of atoms joined together through one or more contiguous double bonds as a single stereocenter leads to correct answers for conjugated and unconjugated alkenes (with and without the presence of tetracoordinate carbon true- and para-stereocenters) and cumulenes (compounds **30–40**). The program properly distinguishes those cumulene linkages that can induce chirality (even number of double bonds) and those which in of themselves will not (odd number of double bonds). Compounds **41–47** illustrate those with double bonds both *endo* (e.g., **44**) and *exo* (e.g., **45**) to a non-aromatic ring. Some possess true tetracoordinate carbon stereocenters either within the ring (e.g., **41**, **44**) or external to it (e.g., **42**). One ring carbon atom of **45** is a para-stereocenter.

The only chemical constraint on the stereoisomer generation process currently enabled allows *cis–trans* isomerism of a carbon–carbon double bond only in cycles of eight members and more. Thus, two diastereomers are predicted for **46**, but only one configuration is allowed for the double bond of **44**.

The results for fused ring systems (**48–52**) are consistent with theory. The program predicts two diastereomers for decalin (**48**) as expected at normal temperatures. As indicated earlier, currently the program does not recognize the strain introduced by a *trans*-linked short bridge (**53–56**) and predicts too high a number of stereoisomers for such compounds. Substituted spiranes are correctly perceived (**58–60**).

Most naturally occurring compounds (e.g., **61–63**) have no or very low symmetry, and the process is largely one of detecting stereocenters, of which there can be many. Compound **61** gives rise to 128 pairs of enantiomers, some of which will be unstable due to strain in their fused rings. These of course cannot be eliminated solely on the application of symmetry considerations.

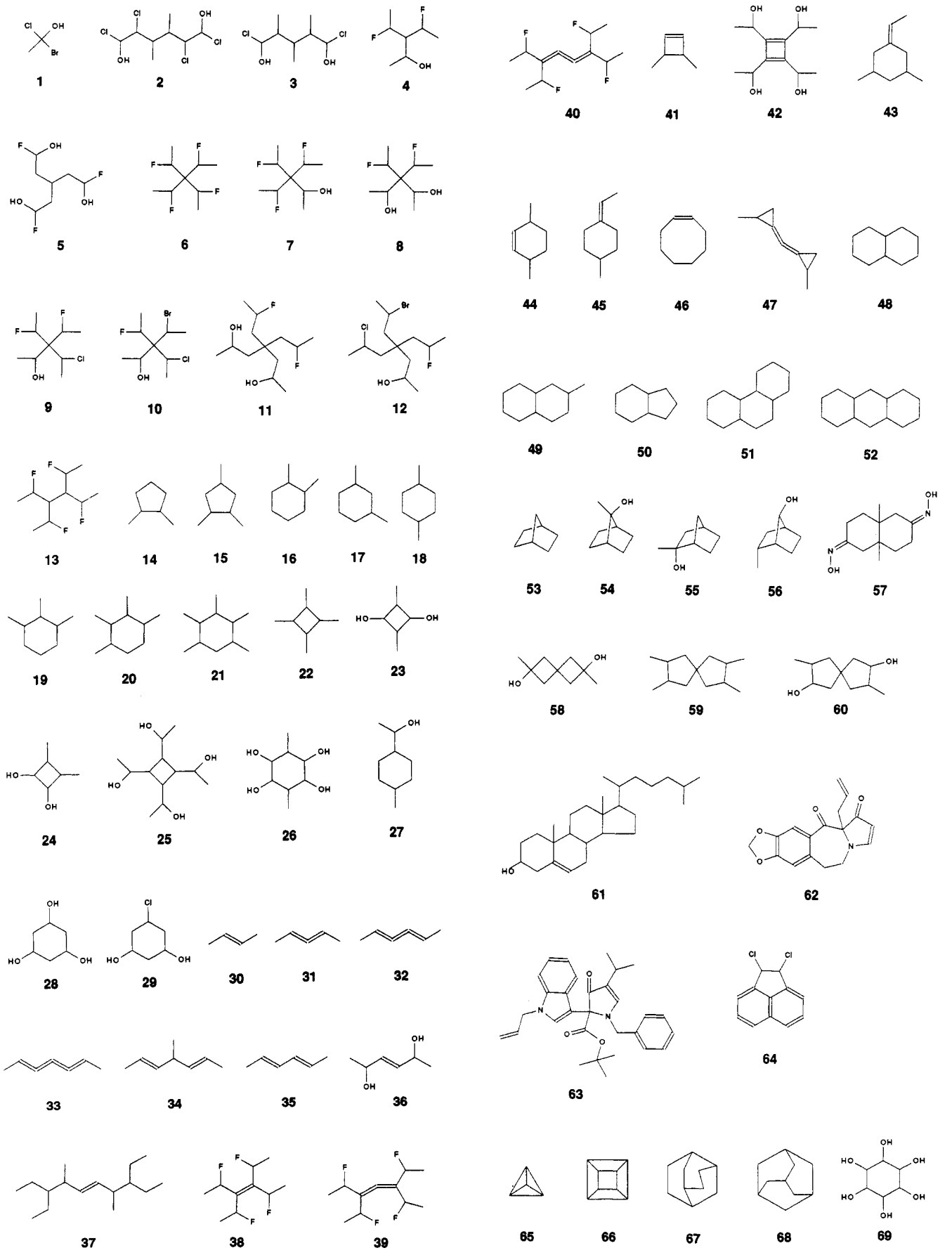


Figure 7. Test compounds.

Finally, a group of structures with very high symmetry were considered. Structures such as tetrahedrane (**65**), cubane (**66**), twistane (**67**), adamantane (**68**) and inositol (**69**) are commonly cited examples. The program provides correct results in each case.

STEREOISOMER ENUMERATION

Combinatorial methods based on Pólya's theorem[31] can be used to *enumerate* stereoisomers in some cases.[9,31-33] For this purpose the cycle index of the automorphism group

Table I. Number and Kinds of Stereoisomers of Compounds in Figure 7

structure no.	no. of stereoisomers			structure no.	no. of stereoisomers		
	total	chiral[a]	achiral[b]		total	chiral[a]	achiral[b]
1	2	2	0	41	3	2	1
2	20	16	4	42	7	4	3
3	16	12	4	43	4	4	0
4	8	8	0	44	3	2	1
5	4	4	0	45	2	2	0
6	5	4	1	46	2	0	2
7	8	8	0	47	6	6	0
8	10	10	0				
9	16	16	0	48	2	0	2
10	32	32	0	49	8	8	0
11	10	10	0	50	3	2	1
12	32	32	0	51	10	8	2
13	10	6	4	52	5	2	3
				53	2	0	2
14	3	2	1	54	2	0	2
15	4	2	2	55	8	8	0
16	3	2	1	56	16	16	0
17	3	2	1	57	10	8	2
18	2	0	2				
19	4	2	2	58	2	2	0
20	10	8	2	59	7	6	1
21	16	12	4	60	20	20	0
22	4	0	4				
23	5	0	5	61	256	256	0
24	10	8	2	62	2	2	0
25	39	32	7	63	2	2	0
26	20	12	8	64	3	2	1
27	4	4	0				
28	2	0	2	65	3	0	3
29	4	2	2	66	14	0	14
				67	7	4	3
30	2	0	2	68	3	0	3
31	2	2	0	69	9	2	7
32	2	0	2				
33	2	2	0				
34	4	2	2				
35	3	0	3				
36	6	4	2				
37	6	4	2				
38	7	4	3				
39	7	6	1				
40	7	4	3				

[a] Exist as pairs of enantiomers. [b] All achiral stereoisomers are diastereomers of one another.

associated with the structure is needed. The cycle index $\mathbf{P}_G$ of an automorphism group G is defined as

$$\mathbf{P}_G = \frac{1}{|G|}\sum_{g\in G} x_1^{b_1}x_2^{b_2}\ldots x_n^{b_n}$$

if a typical $g\in G$ generates b_1 cycles of length 1, b_2 cycles of length 2, ..., b_n cycles of length n upon its action on the true-stereocenters of the topological structure. For example, if there are four true-stereocenters and the action of a $g\in G$ permutes these centers as (12)(34), then the cycle representation is x_2^2. All such cycle representations are collected and divided by the number of elements in the group to obtain the cycle index.

Pólya's theorem provides a method to enumerate the stereoisomers once the cycle index is constructed. Since each true-stereocenter gives rise to two possible configurations, the total number of stereoisomers (I) is given by Pólya's theorem as

$$I = \mathbf{P}_G(x_k \rightarrow 2)$$

where the arrow stands for replacing every x_k by 2 since there are two possible configurations for each true-stereocenter in the structure.

The theorem can now be applied to enumerate the stereoisomers for some of the classes of structures considered in this investigation. For an unbranched chain containing n true-stereocenters, the cycle indices are given by the following expressions depending on whether n is odd ($\mathbf{P}^o_G$) or even ($\mathbf{P}^e_G$).

$$\mathbf{P}^o_G = {}^1/_2[x_1{}^n + x_1x_2{}^{(n-1)/2}]$$

$$\mathbf{P}^e_G = {}^1/_2[x_1{}^n + x_2{}^{n/2}]$$

Hence the stereoisomer counts for compounds with an odd and even number of unbranched chains containing n true-stereocenters (I^o and I^e, respectively) are given by

$$I^o = {}^1/_2[2^n + 2.2^{(n-1)/2}]$$
$$= 2^{n-1} + 2^{(n-1)/2}$$

$$I^e = {}^1/_2[2^n + 2^{n/2}] = 2^{n-1} + 2^{(n-2)/2}$$

Consider compound **6** (Figure 7). The cycle index of the tetrahedral point group (T_d) or the automorphism group of compound **6** is given by

$$\mathbf{P}_G = {}^1/_{24}[x_1{}^4 + 8x_1x_3 + 3x_2{}^2 + 6x_1{}^2x_2 + 6x_4]$$

Hence the stereoisomer count I is

$$I = {}^1/_{24}[2^4 + 8.2.2 + 3.2^2 + 6.2^2.2 + 6.2]$$
$$= 120/24 = 5$$

This result agrees with that reported by the program (Table I). The same was found to be true for compounds **7** and **8** as well.

This technique is rigorous for cases which contain only true-stereocenters. In the presence of para-stereocenters, the direct application of Pólya's theorem, with n equal to the number of *all* stereocenters (true plus para), usually predicts too many stereoisomers. However, this can be rectified in some cases by making a correction to isomers which are "interconverted" by certain symmetry operations, e.g., those operations that invert a ring substituent from above the "plane" of the ring to below. This is illustrated with inositol (**69**). The ordinary cycle index of structure **69** is

$$\mathbf{P}_G = {}^1/_{12}[x_1{}^6 + 2x_6 + 2x_3{}^2 + 4x_2{}^3 + 3x_1{}^2x_2{}^2]$$

The stereoisomer count, assuming all stereocenters to be true, is given by

$$I = {}^1/_{12}[2^6 + 2.2 + 2.2^2 + 4.2^3 + 3.2^4] = 13$$

However, for inositol, considering a planar ring, there are six C_2 operations that exchange the positions of the substituents on the six para-stereocenters, among which three involve an odd number of exchanges. Modifying the cycle index accordingly leads to a correct stereoisomer count of nine.

$${}^1/_{12}[2^6 + 2.2 + 2.2^2 + 2^3 - 3.2^3 + 3.2^4] = 9$$

2.5-D CONFIGURATIONAL REPRESENTATION

Program output is the set of stereoisomers in which a given constitutional isomer may exist. Each stereoisomer has an assigned stereoparity vector and therefore can be conveniently expressed as a connection table which is augmented at each stereocenter site by its corresponding stereoparity vector element. The stereoparity vector itself can be considered as a 2.5-D representation of that particular configuration. However, although this is a simple and straightforward

expression of configuration, it is not commonly used. The Cahn–Ingold–Prelog (CIP) descriptor,[34] which can be readily related to a 3.0-D representation by chemists themselves using a set of published rules, is the accepted standard. The augmented CCT is indeed convenient for the storage and manipulation of stereochemically defined structures, e.g., for direct computer conversion to 3.0-D representations using model-building software, but a generally applicable 2.5-D configuration must also be readily reducible to a 3.0-D representation by the chemist, just as in the CIP system. Since the latter system is well-known, one approach would be to convert the stereoparity vector to a CIP descriptor. In fact, in compounds with only true-stereocenters, e.g., compounds **2**, **36**, and **61** of Figure 7, a simple relationship is workable: parity values of +1 and −1 can be assigned CIP descriptors *R* and *S* (or *Z* and *E*), respectively. However, no simple relationship exists in the case of compounds containing para-stereocenters. The lack of a straightforward connection between the stereoparity vector and the CIP descriptor is not surprising as the former is a symmetry-based concept and the latter derives from a set of sequence rules dependent in part on the property of atomic number. (In fact, the CIP system itself is not inviolate. Prelog proposed modifications in 1982,[34] and later, some additional deficiencies were noted[35,36] which suggest the need to revisit the sequence rules once again.)

A convenient procedure has been developed for the expression of the stereoparity vector as a 3.0-D representation which is similar to, but actually simpler than, that used by the chemist in the CIP system. In this approach, the rather elaborate set of sequence rules used in assigning CIP group priority is replaced by a single, simple priority-ordering rule based on the canonical numbering of the atoms, information that is found in the CCT, a component of the output. That rule, which in contrast to the CIP system only requires an examination of atoms joined *directly* to stereocenters, is as follows: The ranking of atoms contiguous to a stereocenter *decreases* as their canonical sequence numbers *increase*; i.e., that atom with the smallest canonical sequence number is assigned the highest priority (corresponding to the group assigned an "a" in the CIP system). Hydrogen atoms, which are implicitly, not explicitly, defined in the CCT, are always assigned the *lowest* priority.

The next step in the process again parallels that of the CIP system. For a tetracoordinate carbon stereocenter, either true or para, the eye, the stereocenter, and the atom of *lowest* priority, *in that order*, are aligned along the viewing axis. A parity of +1 requires the priorities of the remaining three atoms attached to the stereocenter (which now occupy a plane perpendicular to the viewing axis) to decrease in the *clockwise* sense; a parity of −1 requires a decrease in priority in the *counterclockwise* sense. In the case of a carbon–carbon double bond as the stereocenter, a parity of +1 requires a *cis* relationship between atoms of highest priority and a parity of −1, a *trans* relationship between those atoms.

Application of the proposed system is illustrated in Figure 8. The canonical numbering of atoms for these compounds is taken directly from their CCTs. Program output for the simple compound in Figure 8a—the four, two-element stereoparity vectors shown—correctly reveals two pairs of enantiomers. The stereoparity-based 2.5-D configurations—which could be written, as, for example, [2(1), 3(−1)]-1,2-dichloro-1-propanethiol (i.e., enantiomer E_2)—are readily converted to the 3.0-D representations shown using the procedure described above. The absence of correspondence between the stereoparity-based 2.5-D configuration and that assigned by the CIP system is illustrated by enantiomer E_2.

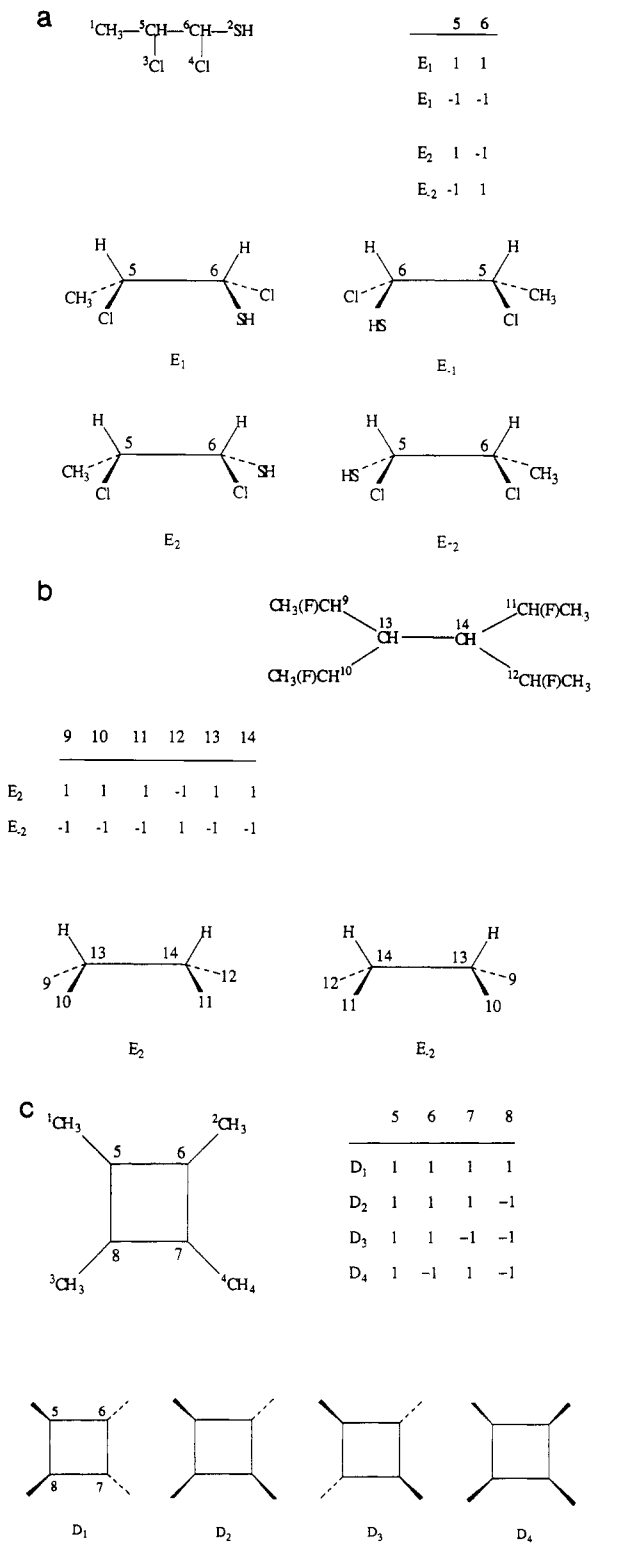


Figure 8. Conversion of stereoparity vectors to conventional 3-D structural Representations: (a) compound with true-stereocenters only; (b) compound with true- and para-stereocenters; (c) compound with para-stereocenters only.

In the former system, the two stereocenters have opposite "configurational assignments" (+1, −1); in the latter system they are the same (*R*,*R*). In either system, the corresponding stereocenters of two stereoisomers related as enantiomers will have opposite configuration designations; e.g., E_2 and E_{-2} are (+1, −1)/(*R*,*R*) and (−1, +1)/(*S*,*S*), respectively.

The procedure described above is completely general. It is applicable to compounds containing both true- and para-stereocenters (Figure 8b) and those with only para-stereocenters (Figures 8c). Figure 8b shows only the spatial

arrangement of atoms about the para-stereocenters in enantiomers E_2 and E_{-2} of compound **I** used earlier to describe the process of stereoisomer generation. Figure 8c displays the four possible stereoisomers (diastereomers) of compound **22** (Figure 7) in dashed–wedged line drawings.

CONCLUSIONS

A generally applicable, computer-executable procedure has been described for the exhaustive and irredundant generation of the stereoisomers of a compound of given topological structure. Although no proof of the universality of the algorithm or its computer implementation can be offered at this time, the absence of a single incorrect prediction among a wide range of structure types (Figure 7) suggests the effectiveness of the program. At the present time, the program contains little intelligence to exclude stereoisomers which are too structurally strained to exist under normal conditions of pressure and temperature. Such refinements will be added.

The program is currently running on a DEC VaxStation 3500. Although most of the structures shown in Figure 7 lead to answers in seconds of CPU time, some of those of very high symmetry (e.g., **25** and **69**) can currently take considerably longer because of the inefficiency of the algorithm now being used in the construction of the automorphism group. A more efficient procedure is being developed to replace it.

Although the procedure described for "visualization" by the chemist of stereoparity vector-based 2.5-D configuration is simple, this designation is not to be viewed as a substitute for the CIP system. In contrast to the CIP system, its utilization requires a canonically numbered structure which is not conveniently derived without computer assistance. However, it does offer the advantage of simplicity of the priority-ordering sequence rule.

ACKNOWLEDGMENT

The financial support of this research by the National Institutes of Health (Grant GM 37963), Sterling Drug, Inc., and Allied Signal, Inc., is gratefully acknowledged. The assistance of Dr. Mark Madison of the Department of Chemistry and Biochemistry at Arizona State University was indispensable in testing the performance of the program. The authors are appreciative of his many contributions.

REFERENCES AND NOTES

(1) Funatsu, K.; Miyabayashi, N.; Sasaki, S. Further Development of Structure Generation in the Automated Structure Elucidation System CHEMICS. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 18–28.
(2) Dubois, J.-E.; Carabedian, M.; Dagane, I. Computer-aided Elucidation of Structures by Carbon-13 Nuclear Magnetic Resonance. The DARC-EPIOS Method: Characterization of Ordered Structures by Correlating the Chemical Shifts of Their Branched Carbon Atoms. *Anal. Chim. Acta* **1984**, *158*, 213–233.
(3) Bremser, W.; Fachinger, W. Multidimensional Spectroscopy. *Magn. Reson. Chem.* **1985**, *23*, 1056–1071.
(4) Christie, B. D.; Munk, M. E. The Role of Two-Dimensional Nuclear Magnetic Resonance Spectroscopy in Computer-Enhanced Structure Elucidation. *J. Am. Chem. Soc.* **1991**, *113*, 3750–3757.
(5) Carhart, R. E.; Smith, D. H.; Brown, H.; Djerassi, C. An Approach to Computer-Assisted Elucidation of Molecular Structure. *J. Am. Chem. Soc.* **1975**, *97*, 5755–5762.
(6) Carhart, R. E.; Smith, D. H.; Gray, N. A. B.; Nourse, J. G.; Djerassi, C. GENOA: A Computer Program for Structure Elucidation Utilizing Overlapping and Alternative Substructures. *J. Org. Chem.* **1981**, *46*, 1708–1718.
(7) Christie, B. D.; Munk, M. E. Structure Generation of Reduction: A New Strategy for Computer-Assisted Structure Elucidation. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 87–93.
(8) Nourse, J. G. The Configuration Symmetry Group and Its Application to Stereoisomer Generation, Specification, and Enumeration. *J. Am. Chem. Soc.* **1979**, *101*, 1210–1216.
(9) Balasubramanian, K. A Generalized Wreath Product Method for the Enumeration of Stereo and Position Isomers of Polysubstituted Organic Compounds. *Theor. Chem. Acta* **1979**, *51*, 37–54.
(10) Davidson, R. A. Unified Combinatorial Molecular Analysis. Ph.D. Thesis, The Pennsylvania State University, University Park, PA, 1977.
(11) Nourse, J. G.; Carhart, R. E.; Smith, D. H.; Djerassi, C. Exhaustive Generation of Stereoisomers for Structure Elucidation. *J. Am. Chem. Soc.* **1979**, *101*, 1216–1223.
(12) Nourse, J. G.; Smith, D. H.; Carhart, R. E.; Djerassi, C. Computer-Assisted Elucidation of Molecular Structure with Stereochemistry. *J. Am. Chem. Soc.* **1980**, *102*, 6289–6295.
(13) Zlatina, L. A.; Elyashberg, M. E. Generation of Stereoisomers and Their Spatial Models Corresponding to the Given Molecular Structure. *MATCH* **1992**, *No. 27*, 191–207.
(14) Abe, H.; Hayasaka, H.; Miyashita, Y.; Sasaki, S. Generation of Stereoisomeric Structures Using Topological Information Alone. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 216–219.
(15) Wipke, W. T.; Dyott, T. M. Simulation and Evaluation of Chemical Synthesis. Computer Representation and Manipulation of Stereochemistry. *J. Am. Chem. Soc.* **1974**, *96*, 4825–4834.
(16) 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 Structures. *J. Comput. Chem.* **1985**, *6*, 562–569.
(17) Mislow, K.; Siegel, J. Stereoisomerism and Local Chirality. *J. Am. Chem. Soc.* **1984**, *106*, 3319–3328.
(18) Balasubramanian, K. Symmetry Groups of Chemical Graphs. *Int. J. Quantum Chem.* **1982**, *21*, 411–418.
(19) Randić, M. On the Recognition of Identical Graphs Representing Molecular Topology. *J. Chem. Phys.* **1974**, *60*, 3920–3928.
(20) Randić, M. On Discerning Symmetry Properties of Graphs. *Chem. Phys. Lett.* **1976**, *42*, 283–287.
(21) Randić, M. Symmetry Properties of Graphs of Interest in Chemistry II: Desargues-Levigraph. *Int. J. Quantum Chem.* **1979**, *15*, 633–682.
(22) Harary, F. On the Group of the Composition of Two Graphs. *Duke Math. J.* **1959**, *26*, 29–34.
(23) Liu, X. Y.; Balasubramanian, K. Computer Generation of Character Tables of Generalized Wreath Products. *J. Comput. Chem.* **1990**, *11*, 589–602.
(24) Razinger, M.; Balasubramanian, K.; Munk, M. E. Graph Automorphism Perception Algorithms in Computer-Enhanced Structure Elucidation. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 197–201.
(25) Randić, M.; Oakland, D. O.; Klein, D. J. Symmetry Properties of Chemical Graphs. IX. The Valence Tautomerism in the P_7^{3-} Skeleton. *J. Comput. Chem.* **1986**, *7*, 35.
(26) Balasubramanian, K. The Symmetry Groups of Non-rigid Molecules as Generalized Wreath Products and Their Representations. *J. Chem. Phys.* **1980**, *72*, 665–677.
(27) Balasubramanian, K. Symmetry Operators of Generalized Wreath Products and Their Applications to Chemical Physics. *Int. J. Quantum Chem.* **1982**, *22*, 1013.
(28) Balasubramanian, K. Computer Generation of Symmetry Elements of Non-rigid Molecules. *J. Comput. Chem.* **1983**, *4*, 302–307.
(29) Baumer, L.; Sala, G.; Sello, G. Ring Perception in Organic Structures: A New Algorithm for Finding SSSR. *Comput. Chem.* **1991**, *15*, 293–299.
(30) Testa, B. *Principles of Organic Chemistry*; Marcel Dekker: New York, 1979.
(31) Pólya, G.; Read, R. C. *Combinatorial Enumeration of Groups, Graphs and Chemical Compounds*; Springer: New York, 1987.
(32) Leonard, J. E.; Hammond, G. S.; Simmons, H. E. The Apparent Symmetry of Cyclohexane. *J. Am. Chem. Soc.* **1975**, *97*, 5052–5054.
(33) Balasubramanian, K. Applications of Combinatorics and Graph Theory to Spectroscopy and Quantum Chemistry. *Chem. Rev.* **1985**, *85*, 599–618.
(34) Prelog, V.; Helmchen, G. Basic Principles of the CIP-System and Proposals for a Revision. *Angew. Chem., Int. Ed. Engl.* **1982**, *21*, 567–583.
(35) Custer, R. H. Mathematical Statements About the Revised CIP-System. *MATCH* **1986**, *No. 21*, 3–31.
(36) Mata, P.; Lobo, A. M.; Marshall, C.; Johnson, A. P. The CIP Sequence Rules: Analysis and Proposal for a Revision. *Tetrahedron: Asymmetry* **1993**, *4*, 657–668.