(7) T. H. Fife and L. H. Brod, *J. Am. Chem. Soc.*, **92**, 1681 (1970).
(8) E. Anderson and T. H. Fife, *J. Am. Chem. Soc.*, **91**, 7163 (1969).
(9) E. Anderson and T. H. Fife, *J. Am. Chem. Soc.*, **93**, 1701 (1971).
(10) R. F. Atkinson and T. C. Bruice, *J. Am. Chem. Soc.*, **96**, 819 (1974).
(11) E. Anderson and B. Capon, *J. Chem. Soc. B*, 1033 (1969).
(12) T. H. Fife, *Acc. Chem. Res.*, **5**, 264 (1972).
(13) T. H. Fife, *Adv. Phys. Org. Chem.*, **11**, 1 (1975).
(14) T. H. Fife and E. Anderson, *J. Am. Chem. Soc.*, **93**, 6610 (1971).
(15) E. Anderson and T. H. Fife, *J. Am. Chem. Soc.*, **95**, 6437 (1973).
(16) T. H. Fife and T. J. Przystas, *J. Am. Chem. Soc.*, **99**, 6693 (1977).
(17) B. Capon, M. I. Page, and G. H. Sankey, *J. Chem. Soc., Perkin Trans. 2*, 529 (1972).
(18) B. Capon, M. C. Smith, E. Anderson, R. H. Dahm, and G. H. Sankey, *J. Chem. Soc. B*, 1038 (1969).
(19) However, see B. Dunn and T. C. Bruice, *J. Am. Chem. Soc.*, **92**, 2410, 6589 (1970).
(20) T. C. Bruice and D. Piszkiewicz, *J. Am. Chem. Soc.*, **89**, 3568 (1967).
(21) 3-*O*-β-D-glucopyranosyl-L-gulonic acid is hydrolyzed more than 100 times faster than 4-*O*-β-D-glucopyranosyl-D-glucitol at pH 4. Intramolecular general acid catalysis was proposed and likewise for hydrolysis of alginate, which gives a sigmoidal pH–rate profile: O. Smidsrød, A. Haug, and B. Larson, *Acta Chem. Scand.*, **20**, 1026 (1966). However, the changing inductive effect due to carboxyl group ionization was not assessed nor were other possible mechanisms considered, such as electrostatic stabilization.
(22) E. E. Smissman and R. A. Mode, *J. Am. Chem. Soc.*, **79**, 3447 (1957).
(23) N. Mori, S. Omura, O. Yamamoto, T. Suzuki, and Y. Tsuzuki, *Bull. Chem. Soc. Jpn.*, **36**, 1401 (1963).
(24) Primed numerals refer to the methyl ester of the corresponding carboxylic acid.
(25) T. H. Fife, *J. Am. Chem. Soc.*, **87**, 271 (1965).
(26) R. L. Nath and H. N. Rydon, *Biochem. J.*, **57**, 1 (1954).
(27) A similar effect of solvent has been found in hydrolysis of the zwitterionic (neutral) species of phosphoramidates: J. D. Chanley and E. Feageson, *J. Am. Chem. Soc.*, **80**, 2686 (1958); **85**, 1181 (1963).
(28) T. H. Fife and L. K. Jao, *J. Org. Chem.*, **30**, 1492 (1965).
(29) The second-order rate constant k_H for hydronium ion catalyzed hydrolysis of benzaldehyde dicyclohexyl acetal at 30 °C in 50% dioxane/H_2O (μ = 0.1) is 66 M^{-1} s^{-1}. In contrast the analogous dimethyl acetal has k_H = 4.8 M^{-1} s^{-1}. T. J. Przystas and T. H. Fife, unpublished data.
(30) T. H. Fife and E. Anderson, *J. Org. Chem.*, **36**, 2357 (1971).
(31) B. Capon and K. Nimmo, *J. Chem. Soc., Perkin Trans. 2*, 1113 (1975).
(32) T. H. Fife and E. Anderson, *J. Am. Chem. Soc.*, **92**, 5464 (1970).
(33) Aliphatic ethers are generally more basic than phenolic ethers by 10^3: E. M. Arnett and C. Y. Wu, *J. Am. Chem. Soc.*, **82**, 4999, 5660 (1960).
(34) W. P. Jencks, *Chem. Rev.*, **72**, 705 (1972).

The Configuration Symmetry Group and Its Application to Stereoisomer Generation, Specification, and Enumeration[1a]

James G. Nourse

Contribution from the Departments of Computer Science[1b] and Chemistry, Stanford University, Stanford, California 94305. *Received June 9, 1978*

Abstract: The configuration symmetry group, a novel specification of the symmetry of an organic chemical structure of defined constitution, is formulated. The symmetry operations in this group are represented in part by their effects on the configurations of the stereocenters in the structure. A description follows of how this group is used: (a) to construct for the first time an algorithm that generates all the distinct stereoisomers of a chemical structure of defined constitution, (b) to specify the configuration of the stereocenters of a stereoisomer independent of geometry, and (c) to provide a general counting equation for the number of stereoisomers of a chemical structure that solves a problem dating back to van't Hoff. Several examples are provided.

I. Introduction

A number of ways exist for specifying the symmetry of a chemical structure, each suitable for different purposes. For most applications the familiar geometric point group is chosen.[2a] In some spectroscopic applications it is necessary to take internal motion into account and specify a nonrigid symmetry group.[2b] For applications in dynamic stereochemistry it is necessary to consider the group of all permutations of identical atoms and often several subgroups.[3] Symmetry groups that include the point group and operations that invert chiral centers are useful both in constructing chirality functions[4] and in specifying the pseudochirality of a structure.[5] An extension of the latter concept leads to a novel symmetry group for a chemical structure that includes, in part, features of all these cases. The purpose here is to describe the formulation of this group, which we term the "configuration symmetry group".

This symmetry group is the key construction that leads to three interesting and important results: (a) an algorithm that exhaustively and irredundantly generates the possible distinct stereoisomers of a chemical structure of specified constitution, a result for which no satisfactory solution has been available; (b) a specification of the configuration of a stereoisomer of an organic molecule of specified constitution that is independent of any geometrical property and that is needed for computer-assisted structure elucidation; (c) a general equation for counting these stereoisomers, which represents the solution to this problem dating back to van't Hoff.[6]

This paper and the following one[7] describe the current effort to provide the CONGEN (for *con*strained structure *gen*eration) program with stereochemical capabilities.[8] This paper is primarily concerned with the chemical and mathematical theory necessary to this effort. The following paper is concerned primarily with novel algorithms and the computer implementation.[7] A third paper considers the theory in greater mathematical detail with some extensions to other topics.[9]

II. Chemical Graphs and Symmetry

Throughout this paper, chemical structures are considered as the graphs[10a] defined by their constitution.[10b-d] Thus, atoms correspond to nodes of the graph and bonds correspond to edges. Each node is numbered and each edge is labeled as in Figure 1. The numbering of the nodes is arbitrary but must be retained throughout the procedure.[9] Only atoms that are at most tetravalent are considered at present. Hydrogens are not explicitly considered and are given the number 0. This suppression of hydrogens is a convenient space-saving feature used throughout the CONGEN program,[8] but is not necessary to the algorithm. This suppression will simplify the presentation here without affecting the results in any way.

There are two standard groups[10a] that are used to describe

0002-7863/79/1501-1210$01.00/0

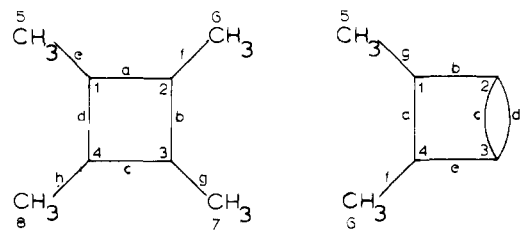


Figure 1. 1,2,3,4-Tetramethylcyclobutane and 3,4-dimethylcyclobutene with atoms and bonds labeled.

Table I. Permutations in the Node Symmetry Group and the Edge Symmetry Group for Tetramethylcyclobutane (Figure 1)

nodes[a]	edges[a]
(1)(2)(3)(4)(5)(6)(7)(8)	(a)(b)(c)(d)(e)(f)(g)(h)
(1)(24)(3)(5)(68)(7)	(ad)(bc)(e)(fh)(g)
(12)(34)(56)(78)	(a)(bd)(c)(ef)(gh)
(1234)(5678)	(abcd)(efgh)
(13)(2)(4)(57)(6)(8)	(ab)(cd)(eg)(f)(h)
(13)(24)(57)(68)	(ac)(bd)(eg)(fh)
(1432)(5876)	(adcb)(ehgf)
(14)(23)(58)(67)	(ac)(b)(d)(eh)(fg)

[a] Permutation conventions are given in Appendix A1.

Table II. The Node Symmetry Group and the Edge Symmetry Group for Dimethylcyclobutene (Figure 1)

nodes	edges
(1)(2)(3)(4)(5)(6)	(a)(b)(c)(d)(e)(f)(g)
(14)(23)(56)	(a)(b)(cd)(e)(f)(g)
	(a)(be)(c)(d)(fg)
	(a)(be)(cd)(fg)

the symmetry of a graph: the node symmetry group and the edge symmetry group. The node symmetry group is the group of all one to one mappings of the nodes of the graph onto themselves, which preserves connectivity of the graph. The edge symmetry group is the group of all one to one mappings of the edges, which preserves connectivity. Tables I and II give the node groups and the edge groups for the two graphs in Figure 1. Note that these two groups need not be the same size.

The graph symmetry group that is needed for the chemical problems discussed here contains elements that are in both the node and edge groups. The requisite group is a product[11] of two groups: (a) the node symmetry group mentioned above; (b) the group of all edge permutations that interchange the two edges corresponding to a double bond. The latter is a subgroup of the group of all edge permutations. Edge permutations that interchange edges in a triple bond are not included since there is no configuration stereochemistry associated with triple bonds. As an example, consider the hydrocarbon 4-methyl-2,5-heptadiene (Figure 2). The product[11] of the node symmetry group and the double-bond symmetry group is shown in Table III.

III. Configuration Symmetry Group

This group, which is the key construction in all the applications described below, is by definition the graph symmetry group represented by its action on the configurations of all stereocenters. For purposes here, a stereocenter is defined to be any non-triply bonded tetravalent or trivalent atom that has at most one hydrogen substituent. In practice, for the organic structures usually encountered, this corresponds to singly and doubly bonded carbon and noninverting nitrogen.[12a] Atoms with more than one hydrogen atom are excluded only because of the suppression of hydrogen indices discussed above. In the computer implementation[7] there is an algorithm that leads to

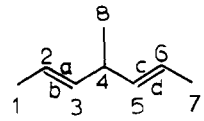


Figure 2. 4-Methyl-2,5-heptadiene with atoms and bonds labeled.

Table III. The Graph Symmetry Group of 4-Methyl-2,5-heptadiene (Figure 2)[a]

	(2)(3)(4)(5)(6)	(4)(26)(35)
(a)(b)(c)(d)	(2)(3)(4)(5)(6)(a)(b)(c)(d)	(4)(26)(35)(ac)(bd)
(ab)(c)(d)	(2)(3)(4)(5)(6)(ab)(c)(d)	(4)(26)(35)(adbc)
(a)(b)(cd)	(2)(3)(4)(5)(6)(a)(b)(cd)	(4)(26)(35)(acbd)
(ab)(cd)	(2)(3)(4)(5)(6)(ab)(cd)	(4)(26)(35)(ad)(bc)

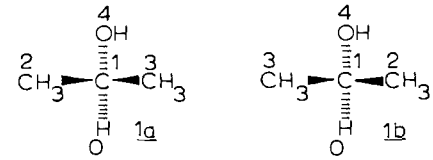
[a] This is displayed as a product[11] of the edge symmetry group (first column) and the node symmetry group (first row). The size of the product group is the product of the sizes of the component groups ($2 \times 4 = 8$). When computing the product it must be noted that permutations of the nodes imply permutations of the edges. Only permutations of the stereocenters and edges involved in double bonds are shown.

Table IV. The Graph Symmetry Group (GSG) and the Configuration Symmetry Group for 2-Propanol (**1a,b**)

GSG	CSG
(1)(2)(3)(4)	(1)(2)(3)(4)
(1)(23)(4)	(1′)(23)(4)[a]

[a] The permutation (1′)(23)(4) indicates that the exchange of the two methyls numbered 2 and 3 inverts the configuration at stereocenter 1.

a more restrictive definition of a stereocenter. The configuration at a stereocenter is determined by the numbering associated with the attached atoms. Thus, even if some of the substituents are identical, two enantiomeric configurations can be formally defined for any stereocenter, such as the central carbon in 2-propanol (**1a,b**), since all the numbers of the at-



tached substituents will necessarily be different. This permits specification of a configuration for all tetravalent and trivalent atoms without further consideration to ensure complete generality.

The configuration symmetry group (CSG) for 2-propanol based on the numbered structures **1a,b** will now be explicitly constructed. The graph symmetry group for 2-propanol contains only one nontrivial permutation, which interchanges the two methyl groups and leaves everything else fixed (Table IV). Now, if one considers the two "enantiomeric configurations" of 2-propanol based on the atom numbering in **1a,b**, it is easy to see that this permutation interconverts these two configurations. Thus it can be said that the permutation (1) (23) (4) inverts the configuration at atom 1. This is designated (1′) (23) (4) and is shown in Table IV. An element in the configuration symmetry group is simply the permutation in the graph symmetry group augmented with superscripts (′) for the stereocenters that are inverted by the permutation. Thus, for 2-propanol, the configuration symmetry group (Table IV) is a group with two elements, one of which inverts the configuration at stereocenter 1.

From this simple example it may seem that the CSG is no different from the point group with reflective operations corresponding to those operations that invert stereocenters. However, a more complicated example shows that this is not the case. Consider tetramethylcyclobutane, whose atoms are

Table V. The Graph Symmetry Group (GSG) and Configuration Symmetry Group (CSG) for Tetramethylcyclobutane[a]

GSG	CSG	GSG	CSG
(1)(2)(3)(4)	(1)(2)(3)(4)	(13)(2)(4)	(13)(2′)(4′)
(1)(24)(3)	(1′)(24)(3′)	(13)(24)	(1′3′)(2′4′)
(12)(34)	(12)(34)	(1432)	(1′43′2)
(1234)	(12′34′)	(14)(23)	(1′4′)(2′3′)

[a] Only permutations of the stereocenters are shown.

numbered as in Figure 1. The graph symmetry group has eight permutations and is shown in Table V. There are four stereocenters and all of them are interconverted by the graph symmetry permutations. To construct the CSG for this example, it is necessary to determine the effect of permutations that exchange stereocenters on the configurations of all the stereocenters. Consider a permutation that takes stereocenter S_a to S_b. The four substituents on S_a are arranged in ascending order and associated with the labels a1-a4. These four atoms are mapped to the four atoms b1-b4 by the permutation. This mapping defines a permutation of the four numbers 1-4. If this permutation is even (see Appendix A1 for definitions of odd and even permutations), then there is no change of configuration at stereocenter S_b caused by the permutation. If the permutation is odd, there is a change of configuration at S_b and this is indicated with a superscript (′). As an example, consider the permutation (1234) on the tetramethylcyclobutane graph (Figure 1). This permutation takes stereocenter S_1 to S_2. The correspondence of substituents follows.

atom	number	label		label	number	atom
H	0	a1		b1	0	H
C	2	a2	→	b3	3	C
C	4	a3		b2	1	C
CH_3	5	a4		b4	6	CH_3

Since a1 goes to b1, a2 goes to b3, a3 goes to b2, and a4 goes to b4, the permutation of substituents is (1) (23) (4), which is an odd permutation; hence, the element of the CSG becomes (12′. . .). Performing this procedure for all the permutations in the graph symmetry group gives the CSG shown in Table V. Note that some of the permutations invert only two of the stereocenters; hence any intuitive connection between elements of the CSG and reflective operations in the point group is lost. A reflective operation would invert all four stereocenters. Furthermore, it is important to realize that this procedure does not depend on the geometry associated with the orientation of the atoms but only on the parity of a permutation that acts on the graph associated with the chemical structure.[12b]

As another example, consider the hydrocarbon 4-methyl-2,5-heptadiene (Figure 2). The graph symmetry group is given in Table VI. This group includes several permutations that exchange double edges. A permutation that exchanges a double edge has the effect of inverting the configurations at both stereocenters to which the edges are attached. The CSG for 4-methyl-2,5-heptadiene is given in Table VI. Only the permutation of stereocenters is given. Intuitively, the configuration symmetry group is the invariance group for a particular stereoisomer that takes account of the configurations of the stereocenters in the structure. See ref 5 for a more pictorial example.

IV. Applications

1. Stereoisomer Generator. The configuration symmetry group can be used to generate all the distinct stereoisomers of a given chemical structure of specified constitution. Each distinct stereoisomer corresponds to an equivalence class in the set of all 2^n possible stereoisomers where n is the number of stereocenters.

Table VI. The Graph Symmetry Group and the Configuration Symmetry Group for 4-Methyl-2,5-heptadiene (Figure 2)[a]

GSG	CSG
(4)(5)(3)(2)(6)(a)(b)(c)(d)	(4)(5)(3)(2)(6)
(4)(5)(3)(2)(6)(a)(b)(cd)	(4)(5)(3′)(2′)(6)
(4)(5)(3)(2)(6)(ab)(c)(d)	(4)(5′)(3)(2)(6′)
(4)(5)(3)(2)(6)(ab)(cd)	(4)(5′)(3′)(2′)(6′)
(4)(53)(26)(ac)(bd)	(4′)(53)(26)
(4)(53)(26)(adbc)	(4′)(5′3)(26′)
(4)(53)(26)(acbd)	(4′)(53′)(2′6)
(4)(53)(26)(ad)(bc)	(4′)(5′3′)(2′6′)

[a] Only the stereocenter permutations are shown.

For each example structure, the 2^n theoretically possible stereoisomers are represented by ordered (from left to right) n-tuples with "+" or "−" entries. [+ − − −] means stereocenter 1 has a "+" configuration, stereocenter 2 has a "−" configuration, etc. (See Appendix A3 for the correspondence of these "+" and "−" labels and the numbered configurations.) For a structure with n stereocenters the algorithm proceeds as follows: (1) Take an n-tuple and form its equivalence class by the operation of the permutations in the CSG. (2) Save one of the members of this equivalence class. (3) Choose another n-tuple that is not a member of a previously constructed equivalence class and iterate 1-3 until no n-tuples remain.

The operation of the permutations in the CSG on these n-tuples may change the n-tuple in two possible ways:

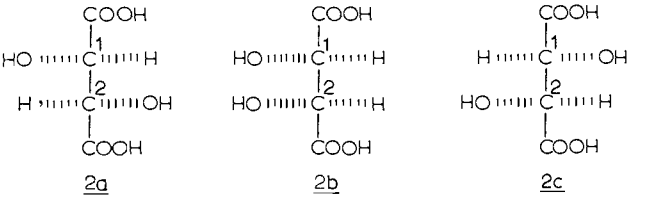
(1) CSG permutations that contain no stereocenter inversions simply permute the members of an n-tuple. For example:

$$[+\,+\,-\,-] \xrightarrow{(1234)} [-\,+\,+\,-]$$

(2) CSG permutations that include stereocenter inversions permute the members of the n-tuple and invert some of them. The permutation (12′34′) is read: 1 goes to 2 and inverts, 2 goes to 3, 3 goes to 4 and inverts, 4 goes to 1.

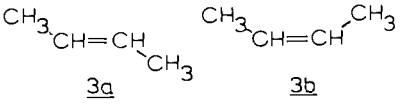
$$[+\,+\,-\,-] \xrightarrow{(12'34')} [-\,-\,+\,+]$$

As an example of the equivalence classes of the CSG consider tartaric acid (**2a-c**). This structure has two stereocenters



2a 2b 2c

and exists in three stereoisomeric forms, a meso form and a *dl* pair. The two elements in the CSG collect the four theoretically possible stereoisomers into three equivalence classes as shown by the multiplication table in Table VII. As a second example consider 2-butene (**3a,b**). This structure has two stereocenters



3a 3b

and exists in two isomeric forms (cis and trans). The four elements in the CSG collect the four possible stereoisomers into two equivalence classes, as shown in Table VIII. Finally, consider tetramethylcyclobutane (**4a-d**). This structure has four stereocenters and exists in four isomeric forms. The eight

Table VII. Generation of the Three Steroisomers for Tartaric Acid[a]

	(1)(2)	(12)	stereoisomer
[++]	[++]	[++]	*d*
[+−]	[+−]	[−+]	meso
[−+]	[−+]	[+−]	meso
[−−]	[−−]	[−−]	*l*

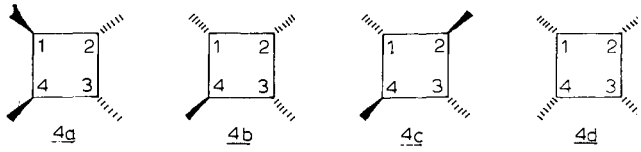
[a] The two permutations in the configuration symmetry group for tartaric acid act on the four possible stereoisomers to give three equivalence classes corresponding to the three distinct stereoisomers, **2a–c.**

Table VIII. Generation of the Two Stereoisomers of 2-Butene[a]

	(1)(2)	(1′)(2′)	(12)	(1′2′)	stereoisomer[a]
[++]	[++]	[−−]	[++]	[−−]	trans **3a**
[+−]	[+−]	[−+]	[−+]	[+−]	cis **3b**

[a] The four permutations in the configuration symmetry group for 2-butene act on the four possible stereoisomers to give two equivalence classes which correspond to the two distinct stereoisomers, **3a,b**. Each row is an equivalence class.

elements in the CSG collect the 16 possible stereoisomers into four equivalence classes, as shown in Table IX.



In summary, this algorithm yields a set of *n*-tuples, each a representative of a different equivalence class. Each of these *n*-tuples corresponds to one of the possible distinct stereoisomers. The computer implementation of this algorithm that yields the stereoisomers of an input structure of specified constitution is described in the following paper.[7]

2. Specification of Configuration. For the organic structures of defined constitution being considered here, an unambiguous specification of the configuration of stereoisomers is provided by establishing the equivalence class of the configuration symmetry group. Thus it can be stated that two chemical structures of identical constitution differ in configuration if and only if they are in different equivalence classes of the configuration symmetry group. This method of specifying configuration has features that are advantageous to both computer and traditional representations of chemical structures.

This method of specifying the configuration of a stereoisomer is independent of any geometrical considerations.[13a] A stereoisomer is represented by the graph describing its constitution augmented by parity labels ("+" or "−" designations) at each stereocenter. Even though these parity labels ultimately refer to geometrical orientation at stereocenters, no use is made of this geometric property when the parity labels are determined by the algorithm described above. The real geometry of the two configurations is considered only when the association of these parity labels with the two possible configurations is done, as shown in Appendix A3. This is an important property for a computer representation of a chemical structure since it is easier to store and manipulate graphical representations augmented with parity labels than to manipulate three-dimensional structures given by spatial coordinates. No preferred viewing directions or reference coordinates are needed. The computer implementation of this method of specifying configuration provides a "canonical" or unique name for each stereoisomer.[7]

This method of specifying the configuration of a stereoisomer simultaneously considers the local property of the substituents at each stereocenter and the global property of the overall symmetry of the structure. It is important to consider both these kinds of properties to obtain a convenient and unique representation of a stereoisomer. A method, such as the *R* and *S* naming system,[13b] that specifies configuration by considering each stereocenter locally becomes extremely complicated for highly symmetrical structures.[13b] Considering each stereocenter locally is necessary, however, since merely specifying the overall symmetry (i.e., a global property) of a chemical structure would not give a unique designation, particularly since most chemical structures have little or no symmetry. The method described above provides a local designation at each stereocenter ("+" or "−") and uses the overall symmetry (the configuration symmetry group) to determine these local stereocenter designations unambiguously. Once the configuration specification has been made, the symmetry group is no longer needed and is not part of the specification.

Table IX. Generation of the Four Stereoisomers of Tetramethylcyclobutane[a]

	4a	**4b**	**4c**	**4d**
(1)(2)(3)(4)	[++++]	[−+++]	[−−++]	[−++−]
(1′)(24)(3′)	[−+−+]	[++−+]	[++−−]	[+−−+]
(12)(34)	[++++]	[+−++]	[−−++]	[+−−+]
(12′34′)	[+−+−]	[+++−]	[++−−]	[−++−]
(13)(2′)(4′)	[+−+−]	[+−−−]	[++−−]	[+−−+]
(1′3′)(2′4′)	[−−−−]	[−−+−]	[−−++]	[−++−]
(1′43′2)	[−+−+]	[−+−−]	[++−−]	[−++−]
(1′4′)(2′3′)	[−−−−]	[−−−+]	[−−++]	[+−−+]

[a] The eight permutations in the configuration symmetry group for tetramethylcyclobutene act on the 16 possible stereoisomers to give four equivalence classes which correspond to the four distinct stereoisomers, **4a–d.** Each column is an equivalence class. The operation of the CSG on only the top entry of each column is shown.

3. Stereoisomer Counting. The configuration symmetry group can be used to give a single equation for counting the number of stereoisomers of an organic structure of defined constitution. This isomer counting problem dates back to the time of van't Hoff[6] and has heretofore lacked a general solution, although a variety of special cases have been treated successfully (vide infra). The objective was to obtain an equation that gives the number of stereoisomers for a structure requiring only a knowledge of the symmetry group of that structure.

The problem of counting the number of various kinds of isomers of chemical structures has fascinated chemists and mathematicians for a long time.[14] A successful approach for counting substitution isomers makes use of the Polya enumeration theorem.[15] The formulation of the problem used in this approach can be used with modification in the present effort. An example of a substitution isomer counting problem is to determine the number of possible chlorobenzenes (i.e., from mono- to hexa-). The problem is formulated as a mapping problem in which the six possible sites on the benzene skeleton are mapped to the two possible ligands (Cl and H). This method can be used for enumeration of stereoisomers if the stereoisomers can be considered somehow as substitution isomers.[16] Thus, for a given chemical structure for which the number of stereoisomers is to be counted, part of the structure must be considered as a skeleton with substitution sites and part must be considered as ligands. This is generally an arbitrary distinction. As an example, consider the substituted cyclobutadiene structure **5**. The ligands are the four constitutionally identical chiral secondary alcohols C_2H_5O-, and the skeleton is the cyclobutadiene fragment with four substitution sites. It is assumed that cyclobutadiene has D_4 symmetry. Each ligand can have two absolute configurations; thus the problem has been reduced to a substitutional isomerism problem with four

Table X. Computation[a] of the Number of Stereoisomers of the Cyclobutadiene Derivative, **5**

	E		$2C_4$		C_2		$2C_2'$		$2C_2''$
permutation	(1)(2)(3)(4)		(1234)		(13)(24)		(12)(34)		(1)(24)(3)
polynomial	$(f_1)^4$	+	$2(f_4)^1$	+	$(f_2)^2$	+	$2(f_2)^2$	+	$2(f_1)^2(f_2)$
counting term	2^4	+	2×2	+	2^2	+	2×2^2	+	$2 \times 2 \times 2 \times 2$

[a] The total number of stereoisomers is the total of the bottom row divided by 8, the order of the CSG.

Table XI. Computation[a] of the Number of Stereoisomers of Tetramethylcyclobutane (**4a–d**)

permutation	(1)(2)(3)(4)		(12′34′) (1′43′2)		(1′3′)(2′4′)		(1)(2)(34) (1′4′)(2′3′)		(1′)(24)(3′) (13)(2′)(4′)
polynomial	$(f_1)^4$	+	$2(f_4)^1$	+	$(f_2)^2$	+	$2(f_2)^2$	+	0
counting term	2^4	+	2×2	+	2^2	+	2×2^2	+	0

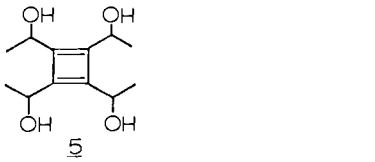
[a] The number of stereoisomers is the total of the bottom row divided by 8, the order of the CSG.

Table XII. Computation of the Number of Stereoisomers of the Tetrahedron Homologue (Figure 3)[a]

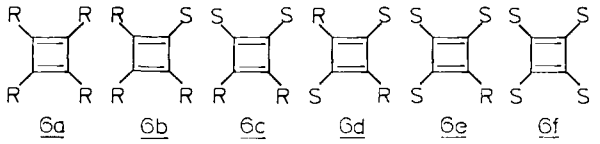
permutation	(1)(2)(3)(4)		(1)(23′4′)		(1′2′)(3′4′)		(1′)(2′)(34)		(1234)
class order	1		8		3		6		6
counting term	2^4	+	$8 \times 2 \times 2$	+	3×2^2	+	0	+	6×2

[a] Only one representative member of each conjugacy class of the configuration symmetry group is shown. The number of stereoisomers is equal to the total of the bottom row divided by 24.

sites related by D_4 symmetry and two kinds of ligands. This problem can now be solved using the Polya enumeration theorem,[16] and the solution is indicated in Table X. The top row shows the symmetry operations in the D_4 group collected by conjugacy class. (See Appendix A2 for a definition of conjugacy class and for a summary of properties needed for isomer counting.) The second row gives a representative element expressed as a permutation on the four skeletal sites. The third row gives the polynomial (termed the cycle index) that results from each type of permutation. The subscript gives the length of the cycles that compose the permutation and the exponent gives the number of cycles of that length. Thus, $(f_1)^2(f_2)$ indicates that the permutation of the kind (1) (24) (3) has two cycles of length 1 and one of length 2. Since there are two kinds of ligands possible (i.e., the two absolute configurations), the



5



6a 6b 6c 6d 6e 6f

number of stereoisomers is obtained by substituting 2 into the polynomial for every f term and dividing by the number of elements in the group (eight, in this case). This gives six stereoisomers, which is the correct answer. The six stereoisomers are symbolized by structures **6a–f.**

There are two reasons why this method cannot be used for the general case. First, factoring the problem into sites and ligands is not always possible (vide infra); second, the Polya enumeration theorem in its original form only considers symmetries of the sites and not of the ligands. Thus, in the example above, no account was taken of the fact that the two kinds of ligands being substituted on the skeleton were enantiomeric. Had the two ligands been CH_3– and Cl–, there would still have been six isomers.

A modification of the Polya theorem, termed the power group enumeration theorem,[17] is somewhat more versatile in that some symmetries that exchange ligands are considered. The additional symmetry is the one that converts all the ligands of one type into another type. This permits the application of this theorem to a wide variety of isomer counting problems, properly formulated.[18]

A general solution to the stereoisomer counting problem is obtained by using the exponentiation group enumeration theorem[19] and the configuration symmetry group. The additional versatility of the exponentation theorem as compared with the power theorem and the original Polya theorem is best demonstrated by an example. Consider the problem of enumerating the stereoisomers of tetramethylcyclobutane (**4a–d**). This structure has four stereocenters that comprise the entire molecule; hence, a factoring into sites and ligands is not apparent. Instead, the problem is formulated as a mapping of stereocenters to configurations.

stereocenters (1–4) → configurations ("+" or "−")

Stereocenters and configurations were defined in section 3. The solution to the enumeration problem is indicated in Table XI. The eight elements in the configuration symmetry group are given on the top two rows of Table XI. With one exception, the polynomial terms are computed as in the Polya method and are indicated in the third row of Table XI. The exception is that any permutation in the CSG with an odd number of inversions in a cycle contributes zero to the total. Thus, permutations of the type (1′) (24) (3′) have an odd number of inversions in two cycles (1′) and (3′) and hence contribute zero, as indicated in Table XI. The number of stereoisomers is obtained by substituting 2 for f in the polynomial and dividing by the size of the CSG (eight, in this case). This gives the correct total of four stereoisomers, as shown in structures **4a–d.** Note the superficial resemblance between the cyclobutadiene problem and this cyclobutane problem. The number of stereocenters and the size of the symmetry groups are the same in both cases; however, the number of stereoisomers differs. The necessary feature of the exponentiation group enumeration theorem is that symmetries that invert the configuration at some but not all stereocenters can be correctly accommodated. It was noted earlier that the CSG will usually contain permutation inversions that invert the configurations at some, but not all, stereocenters. The power group theorem can only accommodate symmetries which invert the configurations at all stereocenters, and the original Polya theorem considers only site symmetries—as discussed above.

The stereoisomer enumeration equation can be given as

$$N = \frac{1}{g} \sum_{i=1}^{c} h_i \prod_{j=1}^{p} 2^{n_j} \prod_{k=0}^{n_j} (n_{ijk} + 1) \bmod 2 \qquad (1)$$

where g is the size of the CSG, c is the number of conjugacy classes in the CSG, h_i is the size of the ith conjugacy class, p is the number of stereocenters, n_j is the number of cycles of length j, and n_{ijk} is the number of inversions in the kth cycle of length j in the ith conjugacy class. n_{ij0} is defined to be equal to 0.[20]

As a second example, consider the problem of enumerating the stereoisomers of the tetrahedrane homologue (Figure 3). This structure has four stereocenters, numbered as shown. The calculation using eq 1 is indicated in Table XII. The first row gives a representative element from each conjugacy class of the CSG for this structure. The second row gives the size of each conjugacy class, which is the coefficient h_i in (1). The third row gives the computed term for each conjugacy class. Note that there is only one zero term. Dividing by $g = 24$ gives three stereoisomers. These are shown in Figure 3. Note that the structure with all hydrogens pointing inside the cage is identical with the structure with all hydrogens pointing out. These structures can turn "inside out" in analogy with the bicyclic structures reported by Simmons and Park.[21]

In conclusion, the construction of the configuration symmetry group for a chemical structure of defined constitution leads to the solution of three diverse problems concerning the stereochemistry[22] of the structure. These are the problems of generation of all distinct stereoisomers, specification of the configuration of each stereoisomer, and enumeration of the stereoisomers from a single equation.

Acknowledgment. Encouragement and helpful suggestions provided by the members of the DENDRAL project in the Departments of Chemistry, Computer Science, and Genetics at Stanford are gratefully acknowledged, as is financial support from the National Institutes of Health (Grant 2R24 RR 00612-08).

Appendix A1. Permutations

Permutations of the type (123) are read: 1 goes to 2, 2 goes to 3, 3 goes to 1. Multiplication of two permutations proceeds from left to right: (123) (1234) = (1324). A cycle in a permutation is the set of numbers between a left and right parentheses. The permutation (12) (34) has two cycles of length 2. A cycle is even if its length is odd. A cycle is odd if its length is even. To determine whether a permutation is odd or even simply take the sum of the odd or even characters of the cycles by the rules:

$$\text{odd} + \text{odd} = \text{even}$$

$$\text{even} + \text{even} = \text{even}$$

$$\text{even} + \text{odd} = \text{odd}$$

Thus (1) (2) (34) is even + even + odd = odd; (12) (34) is even; (1) (234) is even; (1234) is odd; etc. Permutations with inversions of the type (12'34') are read: 1 goes to 2 and inverts configuration; 2 goes to 3; 3 goes to 4 and inverts configuration; 4 goes to 1.

Appendix A2. Conjugacy Classes

Conjugacy classes in a group are defined in ref 2a and 12a. Two elements a and b are conjugate in G if for some g in G, $gag^{-1} = b$. Two properties of conjugacy classes are important for this paper:

(1) All elements in a conjugacy class have the same permutation cycle structure (see Appendix A1 for definitions).

(2) All elements in a conjugacy class have the same number and length of cycles with an odd or even number of inversions. The actual number of inversions may vary as long as the number remains odd or even. References 4 and 18 include a discussion of this property, of odd or even number of inversions in permutation cycles. See Table XI for examples.

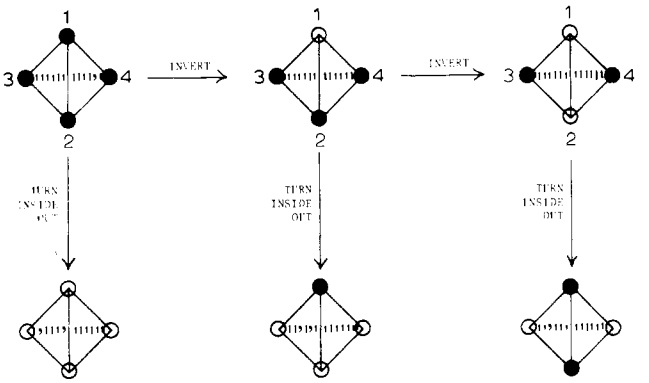


Figure 3. The three possible stereoisomers of a tetrahedrane homologue and their interconversion. Dots on the vertices of the tetrahedrons indicate that the remaining substituent points outside the cage of the tetrahedron. Open circles indicate that this substituent points inside the cage. The three distinct stereoisomers are in each of the three columns of the figure. Starting from the left, the stereoisomer with all four substituents pointing outside the cage is equivalent to the stereoisomer with all four pointing inside, by the operation that turns the structure "inside out." If stereocenter 1 is inverted, a distinct stereoisomer is obtained that has one substituent pointing inside the cage (middle column). If stereocenter 2 is also inverted, the final distinct stereoisomer is obtained (right column).

Appendix A3. Configurations

The correspondence between the parity labels ("+" or "−") and the configuration of a stereocenter based on the numbering of the atoms is shown below.

References and Notes

(1) (a) Applications of Artificial Intelligence for Chemical Inference. 28. For preceding paper see T. H. Varkony, R. E. Carhart, D. H. Smith, and C. Djerassi, *J. Chem. Inf. Comput. Sci.*, **18**, 168 (1978). (b) Address correspondence to the Department of Computer Science.
(2) (a) F. A. Cotton, "Chemical Applications of Group Theory", 2nd ed., Wiley-Interscience, New York, 1971; (b) H. C. Longuet-Higgins, *Mol. Phys.*, **6**, 445 (1963).
(3) (a) M. Gielen and N. Vanlautem, *Bull. Soc. Chim. Belg.*, **79**, 679 (1970); (b) P. Meakin, E. L. Muetterties, F. N. Tebbe, and J. P. Jesson, *J. Am. Chem. Soc.*, **93**, 4701 (1971); (c) W. G. Klemperer, *J. Chem. Phys.*, **56**, 5478 (1972); (d) W. Hasselbarth and E. Ruch, *Theor. Chim. Acta*, **29**, 259 (1973); (e) D. J. Klein and A. H. Cowley, *J. Am. Chem. Soc.*, **97**, 1633 (1975); (f) J. G. Nourse and K. Mislow, *ibid.*, **97**, 4571 (1975).
(4) A. Mead, *Top. Curr. Chem.*, **49**, 1 (1974).
(5) J. G. Nourse, *J. Am. Chem. Soc.*, **97**, 4594 (1975).
(6) (a) J. H. van't Hoff, *Arch. Neerl. Sci. Exactes Nat.*, **9**, 445 (1874); (b) R. W. Robinson, F. Harary, and A. T. Balaban, *Tetrahedron*, **32**, 355 (1976).
(7) J. G. Nourse, R. E. Carhart, D. H. Smith, and C. Djerassi, *J. Am. Chem. Soc.*, following paper in this issue.
(8) R. E. Carhart, D. H. Smith, H. Brown, and C. Djerassi, *J. Am. Chem. Soc.*, **97**, 5755 (1975), and references cited therein.
(9) J. G. Nourse, to appear.
(10) (a) F. Harary, "Graph Theory", Addison-Wesley, Reading, Mass., 1969, Chapter 14; (b) A. T. Balaban, Ed., "Chemical Applications of Graph Theory", Academic Press, New York, 1976; (c) L. M. Masinter, N. S. Sridharan, J. Lederberg, and D. H. Smith, *J. Am. Chem. Soc.*, **96**, 7702 (1974); (d) K. Mislow, *Bull. Soc. Chim. Belg.*, **86**, 595 (1977). See this paper for a discussion of the limitations of the concept of constitution.
(11) This product of groups is called a semidirect product. (a) M. Hall, Jr., "The Theory of Groups", Macmillan, New York, 1959; (b) M. G. Hutchings, J. G. Nourse, and K. Mislow, *Tetrahedron*, **30**, 1535 (1974).
(12) (a) sp-hybridized carbons (i.e., in cumulenes) are only formally considered as stereocenters for purposes of generation of stereoisomers (section 4A) but need not be considered for purposes of counting (section 4C). (b) There is a sense of arbitrariness here since the permutation inversions of the CSG

depend on the numbering of the atoms and a different numbering of the atoms will give a different CSG. However, the results of operations and computations using the CSG do not change as they are invariant to renumbering. In group theoretic terms, the overall result is invariant to conjugation and a renumbering merely gives a conjugate CSG.
(13) (a) The term configuration is generally used in reference to the geometric property of the positions of the atoms in space. See K. Mislow, "Introduction to Stereochemistry", W. A. Benjamin, New York, 1965, p 82. (b) R. S. Cahn, C. Ingold, and V. Prelog, *Angew. Chem., Int. Ed. Engl.*, **5**, 385 (1966). In particular see the footnote on p 397. We are not proposing a new system of nomenclature but merely pointing out a difficulty involved in designing such a system.
(14) For a recent review see D. H. Rouvray, *Chem. Soc. Rev.*, **3**, 355 (1974), and the references in ref 18 of the present paper.
(15) (a) G. Polya, *Acta Math.*, **68**, 145 (1937); (b) C. Berg, "Principles of Combinatorics", Academic Press, New York, 1971, Chapter 5.
(16) (a) B. A. Kennedy, D. A. McQuarrie, and C. H. Brubaker, Jr., *Inorg. Chem.*, **3**, 265 (1964); (b) J. G. Nourse, Ph.D. Thesis, California Institute of Technology 1974; (c) J. E. Leonard, Ph.D. Thesis, California Institute of Technology, 1971.
(17) N. G. deBruijn, *Indagationes Math.*, **21**, 59 (1959), and ref 11 and 15b.
(18) R. A. Davidson, Ph.D. Thesis, The Pennsylvania State University, 1976. We thank Dr. Davidson for making this available.
(19) (a) A. Kerber, "Representations of Permutation Groups," Vol. II, Springer-Verlag, New York, 1975; (b) A. Kerber, *Discrete Math.*, **13**, 13 (1975); (c) F. Harary, *Am. Math. Monthly*, **66**, 572 (1959).
(20) Proof of this formula relies on these properties: (1) The CSG is a subgroup of the exponentiation group.[19] (2) A permutation–inversion with an odd number of inversions per cycle cannot fix any state of the configurations of the stereocenters. (3) The result is invariant to conjugation (renumbering). See ref 9 for further discussion and extensions.
(21) C. H. Park and H. E. Simmons, *J. Am. Chem. Soc.*, **94**, 7184 (1972).
(22) The term *stereology* has been suggested as a substitute for *stereochemistry* in this context.[18]

Exhaustive Generation of Stereoisomers for Structure Elucidation[1]

James G. Nourse,* Raymond E. Carhart, Dennis H. Smith, and Carl Djerassi

Contribution from the Departments of Chemistry, Computer Science, and Genetics, Stanford University, Stanford, California 94305. *Received June 9, 1978*

Abstract: An algorithm and its implementation as a computer program are described that, for the first time, permit the enumeration and construction of all possible distinct stereoisomers consistent with a given empirical formula. The algorithm finds the stereocenters in a chemical structure, takes full account of any symmetry, and produces the stereoisomers with cis/trans and *R/S* designations along with a canonical (unique) name. Examples of its use and a discussion of potential applications are given.

To determine the structure of an unknown compound from an empirical formula is one of the oldest problems in chemistry. A second very old problem is to determine the number of possible structures for a given empirical formula. A third problem is to generate and display these possible structures. The latter two problems are the focus of this work. Particular emphasis is placed on stereoisomers since the most significant limitation of our current effort in computer-assisted structure elucidation[2a-c] has been the inability to recognize the stereochemical features of chemical structures. Indeed, the wide application of computer methods to structure elucidation depends on the successful solution of the problem of isomer enumeration and generation.

These problems, of isomer enumeration (computation of the total number) and generation (construction of all possibilities), have proved to be very difficult,[2d] and it was not until 1974 that the problem of generating the possible *constitutional* isomers from a given empirical formula was finally solved.[2a,b] The only deficiency to the solution at that time was that stereochemistry was not considered so no *stereoisomers* were generated. The purpose of this paper is to describe an algorithm and its concomitant implementation as a computer program which can generate or enumerate the possible stereoisomers of a structure of given constitution. The algorithm makes use of the novel group theoretical and combinatorial results described in the preceding paper.[3] The computer program has been combined with the program CONGEN (for *con*strained *gen*eration),[2c] which generates all constitutional isomers, to yield a program which is now capable of generating all the possible stereoisomers from a given empirical formula.

It is important to be able to exhaustively generate all the possible isomers for a given structural problem to assure that none have been overlooked. However, the complete collection of possible isomers can be extremely large so it is important that the method of generation of these possibilities can be constrained to only a subset of possibilities, if partial structures are known. The algorithm presented here for generation of stereoisomers is capable of admitting certain constraints that reduce the number of stereoisomers generated.

I. Overview and Flow Diagram

When a chemist is faced with the problem of determining the number of stereoisomers of a structure of given constitution, he will probably break the problem into two parts. First, he will try to find the features of the structure which give rise to configurational stereochemistry, such as asymmetrically substituted carbon atoms and double bonds. Symmetrically substituted atoms such as methylenes or *gem*-dimethyls will be rejected as potential stereocenters. Second, having found n stereocenters, he will assume that there are 2^n possible stereoisomers, unless the structure has some overall symmetry—in which case this total may be reduced. In cases with overall symmetry, the distinct stereoisomers will probably be found by trial and error—by varying the configuration of stereocenters in turn and seeing if new stereoisomers are generated.

The algorithm to solve the problem of stereoisomer generation is summarized in the flow diagram shown in Figure 1. Just as the chemist, the algorithm faces two key problems: to determine the potential stereocenters and to correctly gauge the effect of any structural symmetry. A brief overview of this algorithm follows (numbers correspond to those on Figure 1).

(1) The input structure is processed to find multiply bonded atoms that are potential stereocenters (e.g., olefins, allenes) by the module "process multiple bonds". The symmetry group of the input structure is also determined at this stage by the module "find symmetry group". Structures A and B in Figure

0002-7863/79/1501-1216$01.00/0