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

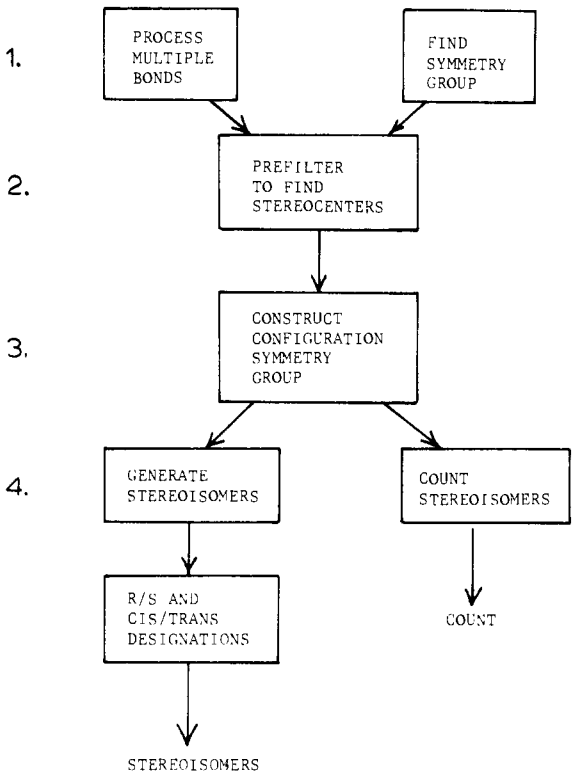


Figure 1. Flow diagram of the algorithm used to generate and count stereoisomers from a structure of defined constitution.

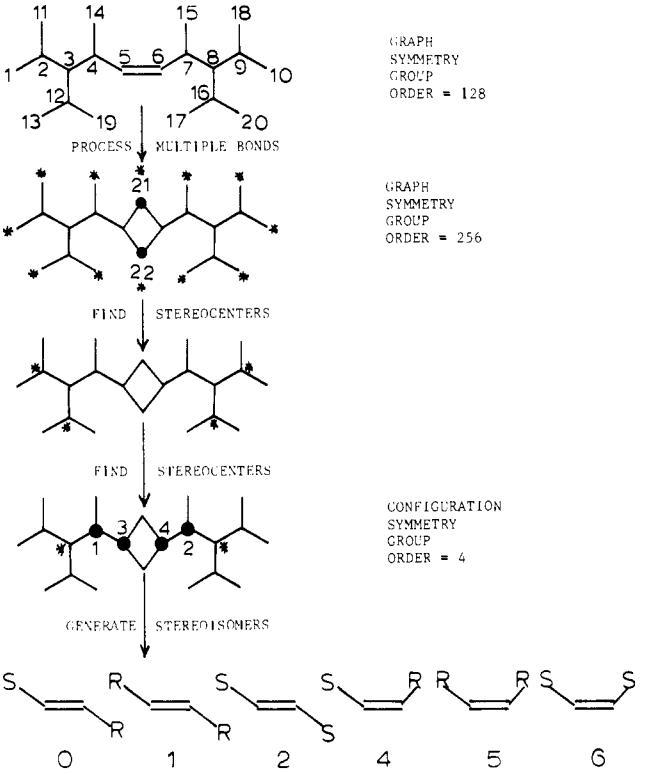


Figure 2. Sequence of the processing of **17** to generate the six possible stereoisomers. Structures A–D, top to bottom.

2 illustrate these processes.

(2) The stereocenters are found by the module "prefilter". The input is the symmetry group and the input structure with the information about multiple bonds. The output is the set of stereocenters and the atoms to which they are connected. The symmetry group remains unchanged here. Structure D in Figure 2 shows the stereocenters.

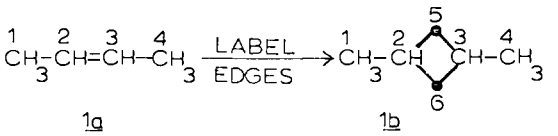
(3) The configuration symmetry group (CSG), which is the symmetry group represented on the configurations of the stereocenters,[3] is constructed (Figure 2, structure D).

(4) Using both the set of stereocenters and the CSG, the possible stereoisomers are generated by the module "generator". These are processed further to give cis/trans and *R/S* designations. The output is a list of stereoisomers (Figure 2). Using just the CSG, a count of the possible stereoisomers can be obtained by using the appropriate combinatorial equation.[3] The output is just the number of stereoisomers. The example shown in Figure 2 is discussed in greater detail later.

II. Method

Input Structure. The structure for which possible stereoisomers are generated must have a definite constitution, that is, a definite number of atoms and bonds. The algorithm and program consider this structure as a graph[2a] in which the atoms are nodes and the bonds are edges. Each atom is uniquely numbered, and for these purposes any numbering will suffice. For other purposes there is usually one which is preferred for some reason (e.g., one with numbering corresponding to standard nomenclature). The structure is represented as a connection table that has one numbered row that corresponds to each atom, and these rows consist of all the numbered atoms to which that atom is connected. Hydrogen atoms are not explicitly considered and are given the number 0. This is a space-saving feature of the CONGEN program. Each atom also carries a designation if it is part of an aromatic system and a designation for its atom type (C, N, O, etc).

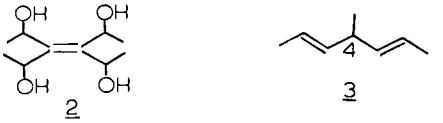
Process Multiple Bonds. The input structure is searched for atoms involved in multiple bonds that are potential stereocenters. At this stage only atoms involved in aromatic systems, triply bonded atoms, cumulenes with CH_2 ends, and rings of sp-hybridized carbon atoms are rejected as potential stereocenters. The latter hypothetical structure would be generated for an empirical formula with only carbon atoms. Next it is necessary to assign a configuration to these atoms. This is done by labeling the edges of the multiple bond with fictitious bi-



1a **1b**

valent nodes, **1a,b.** These nodes are given numbers larger than those already used to number the atoms.

Each multiply bonded atom now has four attached nodes with different numbers, so two configurations can be assigned to each atom. This manipulation serves two purposes. First, each multiply bonded stereocenter can be treated in the same manner as a tetracoordinated stereocenter so that no special handling of multiply bonded substructures is necessary. This is important for the generation of stereoisomers of structures such as **2** and **3** in which the configuration of a double bond depends on the configuration of attached tetracoordinate



2 **3**

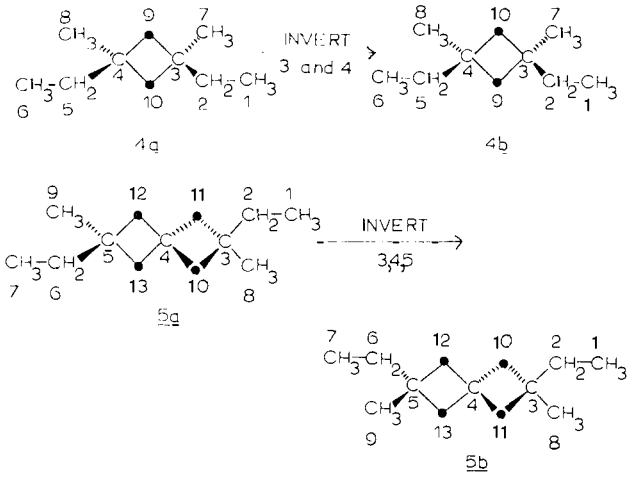
stereocenters, and vice versa. Second, the enantiomer of any cumulated multiply bonded structure is easily obtained by simply inverting the configurations of all the component stereocenters. This is illustrated by **4a,b** and **5a,b** for an olefin and an allene (atoms 9 and 10 and 10–13 are the fictitious bivalent nodes for **4** and **5**, respectively). The results hold for any cumulated system. Inversion of the configuration of the two

Table I. The Five Kinds (Distinct by Conjugation) of Permutations in the Symmetry Group of the Tetrahedron[a]

permutation cycle type[b]	example	even/odd	eliminate stereocenter?	example structure
1^4	(1) (2) (3) (4)	E	no	asymmetric carbon
$1^2 2$	(12) (3) (4)	O	yes	*gem*-dimethyl, **3**
2^2	(12) (34)	E	no	**8**
13	(123) (4)	E	no	**9**
4	(1234)	O	yes	**10**

[a] These may or may not eliminate a potential stereocenter if there are equivalent substituents related by this kind of permutation. [b] The permutation cycle type $1^a 2^b 3^c 4^d$ means that there are a cycles of length 1, b cycles of length 2, etc.

formal stereocenters of the olefin **4** is accomplished by exchanging nodes 9 and 10. This has no effect on the relative (i.e., cis or trans) configuration of the two stereocenters. Inversion of the three stereocenters of the allene, **5a** to **5b**, is accomplished by exchanging nodes 10 and 11 and nodes 6 and 9. This has the effect of changing **5a** into its enantiomer **5b**.



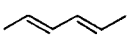
4a 4b

5a

5b

Find Symmetry Group. The symmetry group of the structure is required before further processing can be done. This is the group of all mappings of the nodes of the graph describing the chemical structure onto themselves that preserve connectivity. Because of the fictitious nodes labeling the edges of the multiple bonds, this group can be excessively large, even for structures with little overall symmetry. However, the entire group need never be explicitly constructed since there is a convenient factoring into two smaller groups. It is the product of these two smaller groups that gives the desired symmetry group.[3] One of these groups permutes only the fictitious nodes labeling the edges of double bonds. For each double bond so labeled, there is a group of order 2 that includes the permutation of the two nodes, corresponding to the two edges. If there are n double bonds this group will be of order 2^n, the product of n groups of order 2. This group is very easily constructed. The other group is the symmetry group of the graph corresponding to the input structure without the double bond labeling. The algorithm used to construct this group from the input structure is essentially the same as the one described previously.[4]

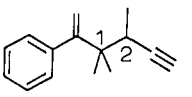
As an example consider the hydrocarbon **6.** Since there are two double bonds, the symmetry group that permutes the edges of the double bonds is of order $2^2 = 4$. The symmetry group of the graph corresponding to the structure is of order 2 and includes the C_2 rotation, relating to two equivalent halves of the structure. The overall symmetry group is the product of these and is of order 8. The symmetry of aromatic parts of any structure is determined by neglecting the unsaturation in aromatic rings—thus, the graph symmetry of benzene would be a group of order 12, rather than a group of order 6.



6

Find Stereocenters. An algorithm that finds the stereocenters in a structure of defined constitution is critical to reducing the size of the computational problem since the number of potential stereoisomers is 2^n, where n is the number of potential stereocenters. For purposes of this work, a stereocenter is defined by the algorithm to be described.

The algorithm starts out by assuming that every tri- and tetrasubstituted atom (i.e., nonhydrogen substituents, including those with fictitious nodes) is a stereocenter, apart from those already rejected during the processing of multiple bonds (vide supra), and proceeds in two stages to try to reject some of these as being incapable of exhibiting configurational stereochemistry. For example, only carbons number 1 and 2 remain to be considered as stereocenters in **7.**



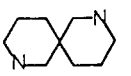
7

The first stage of the algorithm examines the symmetry at each potential stereocenter in order to find identical substituents. Intuitively, this corresponds to the familiar rule that tetravalent atoms with four different substituents can exhibit configurational stereochemistry while those with some identical substituents cannot. In reality, the problem is somewhat more complicated. There are actually five cases that must be considered. These five cases correspond to the five "kinds" of permutations or conjugacy classes[5] in the symmetry group of the tetrahedron, T_d. These are summarized in Table I. Once a symmetry represented as a permutation of identical substituents at a potential stereocenter is found, the question asked is whether this permutation is capable of eliminating this potential stereocenter from further consideration. The five possibilities are considered here in sequence (with reference to Table I).

1. If there are no equivalent substituents by this permutation then it is not capable of eliminating the potential stereocenter from further consideration. An example of such a case is a carbon with four different substituents. This case is trivial.

2. If there are two equivalent substituents by this permutation then it is possible that this permutation will eliminate the potential stereocenter. An example would be a *gem*-dimethyl substituted carbon. However, there are cases in which such carbons remain stereocenters (e.g., **3,** where the central carbon can be a stereocenter even though it has two constitutionally equivalent substituents). At this stage of the algorithm, these cases are merely marked for further consideration in the second stage.

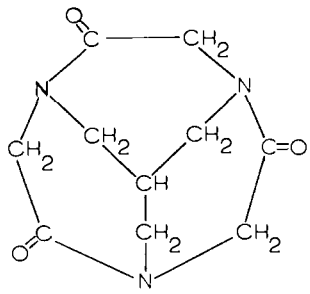
3. If there are two sets of two equivalent substituents by this permutation then it is not possible that this permutation alone will eliminate the stereocenter from further consideration. This property may be established with a single counterexample. For example, **8** has only a single C_2 axis of symmetry that exchanges two sets of equivalent substituents at the central carbon. Assuming that the nitrogens are free to invert, this structure exists in two enantiomeric forms. The central carbon



8

is the stereocenter that must be inverted to interconvert the enantiomers.

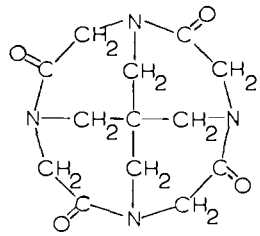
4. If there are three equivalent substituents by this permutation then it is not possible that this permutation alone will eliminate the stereocenter from further consideration. As an example, consider **9,** which has only C_3 symmetry. Assuming



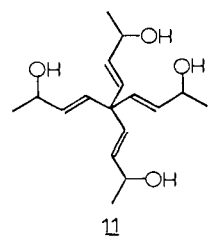
9

that the nitrogens are free to invert, this structure exists in two enantiomeric forms, with the central carbon being the stereocenter.

5. If there are four equivalent substituents by this permutation then it is possible that this permutation alone will eliminate the stereocenter from further consideration. As an example, consider **10,** which has only the S_4 symmetry axis.



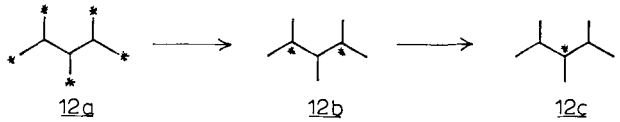
10



11

This structure exists in only one isomeric form. It is still possible for a carbon with four constitutionally identical substituents to be a stereocenter; an example is **11.** At this stage of the algorithm these cases are marked for further consideration in the second stage.

The second stage of this step of the algorithm for finding stereocenters considers cases 2 and 5 from the first stage. Potential stereocenters that have these two types of symmetry will actually be stereocenters only if the substituents themselves include stereocenters (e.g., **2, 3, 11**). Thus the substituents must be searched to see if they contain stereocenters; if not, then the central atom will not be a stereocenter. This search must be done iteratively; that is, if new nonstereocenters are found, then another search must be done. As an example, consider **12a–c.**



12a 12b 12c

In order to establish that the central carbon with two isopropyl substituents is not a stereocenter, it first must be established that the *gem*-dimethyl substituted carbons of the isopropyl groups are not stereocenters. This process is indicated by the sequence of structures. First, the methyl groups are indicated as nonstereocenters (by asterisks, **12a**). Second, the *gem*-dimethyl substituted carbons are indicated as nonstereocenters, **12b,** and finally the central carbon is indicated as a nonstereocenter, **12c.** The structure can therefore exist in only one isomeric form, since all its atoms are nonstereocenters.

A stereocenter can then be defined as an atom that "survives" this procedure, that is, it is never designated a nonstereocenter. This algorithmic definition is based solely on the constitution of the structure under consideration and is convenient for the problem of stereoisomer generation since all stereoisomers have the same designated stereocenters.[6] Nitrogen atoms that survive as stereocenters are noted, and in the computer implementation the user is asked whether a nitrogen atom that is a potential stereocenter is considered able to exist in stable configurations (i.e., not invert freely).

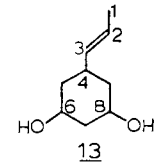
Table II. Renumbering of Five Stereocenters

stereocenter	1	2	3	4	5
input number	4	6	8	2	3

Construct Configuration Symmetry Group. The configuration symmetry group is constructed using the algorithm previously described.[3] Each element of this group consists of two parts, a permutation of the stereocenters and a list of the stereocenters whose configuration is inverted by the permutation. This group will generally be smaller than the graph symmetry group because permutations that invert only nonstereocenters are discarded, leaving the smaller group of all the permutations which invert stereocenters.[7]

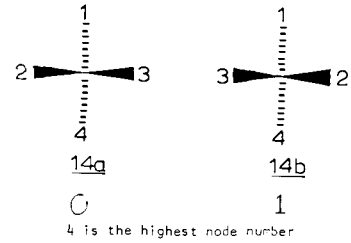
Count Stereoisomers. Using the configuration symmetry group, a count of the possible stereoisomers can be obtained using the recently derived combinatorial equation.[3] This can be done much faster than the independent generation of stereoisomers.

Generate Stereoisomers. The distinct stereoisomers can be generated from the list of stereocenters and the configuration symmetry group, using the algorithm described previously.[3] To do this efficiently, we define an ordering of the n stereocenters and the 2^n potentially distinct stereoisomers. The ordering of the stereocenters is accomplished by using the numbering of the input structure. All the stereocenters (as defined by the above algorithm) except those involved in multiple bonds are collected together and ordered based on their input ordering. This set is followed by the set of stereocenters involved in multiple bonds, also ordered by their input ordering. This ordering is shown for **13** in Table II.



13

With the stereocenters ordered, the 2^n potential stereoisomers can also be ordered. The two possible configurations of each stereocenter, based on the numbering of the nodes, correspond to the binary numbers 0 and 1 as shown, **14a,b.** Each



14a 14b

C 1

4 is the highest node number

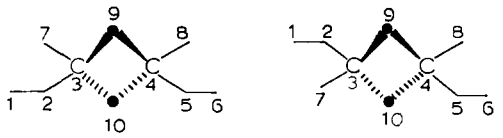
stereoisomer then corresponds to a string of n zeros and ones, which is just the binary representation of an integer. Thus the 2^n potential stereoisomers correspond to the integers from 0 to $2^n - 1$. The ordering of the integers is the ordering of the stereoisomers. This representation of stereoisomers as single integers permits a very compact machine representation for efficient storage. For any stereoisomer in this representation, the enantiomer (the structure with all stereocenters inverted) is just the bit-wise complement; for any integer i, $(2^n - 1) - i$.

Table III. Correspondence between the *R/S* Convention and the Convention Used Here with Node Numbers

	0	1
R	even	odd
S	odd	even

The first stereoisomer is generated by starting with the integer 0 (i.e., the stereoisomer with all stereocenters in configuration 0) and multiplying by each of the elements in the configuration symmetry group, as detailed in ref 3. For any group element this will either give back the integer 0 or a larger integer. If 0 is returned, the group element is a symmetry element of this stereoisomer and this information is saved.[6] If a higher integer is returned, that integer is "crossed off" the list of possible distinct stereoisomers since it is symmetrically equivalent to 0. Should this higher integer be the enantiomer of 0 (i.e., $2^n - 1$) then two facts are noted. The first is that this stereoisomer is achiral, i.e., equivalent to its enantiomer. The second is that this symmetry operation corresponds to a reflective operation on this stereoisomer. Generation of the next distinct stereoisomer proceeds by taking the lowest integer not already crossed off the list and repeating the same generation procedure as was done for 0. This process continues until all $2^m - 1$ integers are accounted for, where m is the number of tetravalent (not multiply bonded) stereocenters. For each distinct stereoisomer, the lowest integer is kept as a representative.

For multiply bonded stereocenters the process differs somewhat. Each double bond is considered as consisting of two stereocenters (allenes and higher cumulenes consist of three or more). However, it suffices to consider only one when generating stereoisomers since the two possible stereoisomers of a double bond differ in the configuration of just one stereocenter. This is shown by example, **15a,b**, in which the cis and

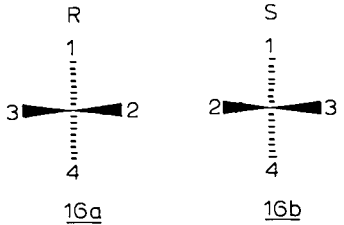


15a 15b

trans isomers differ by the configuration of stereocenter number 3. Generation proceeds by considering only one stereocenter for each cumulene, which is chosen to be the lowest numbered one. When one of the higher numbered stereocenters is reached, it is "skipped"; that is, if it is stereocenter number p, 2^{p-1} integers are skipped in the generation procedure. This leads to a considerable saving of time in the generation of double-bond stereoisomers. However, the real importance of this method is that this is a constrained generation of stereoisomers; only certain stereocenters are allowed to vary in configuration. An analogous constrained generation problem would be to generate all the stereoisomers of a structure with two tetravalent stereocenters in a fixed relative configuration (i.e., erythro or threo for sugars, cis or trans for two stereocenters in a cyclic system, etc.). The generation scheme described here admits such constraints easily and prospectively. That is, unwanted stereoisomers are rejected before they are generated or, in fact, even considered.

Canonical Name. The lowest integer kept as a representative for each distinct stereoisomer provides a canonical or unique name for that stereoisomer. This name can be appended to a canonical numbering of the input structure to provide a canonical name for the structure that includes the configuration of all stereocenters. In the CONGEN program the stereoisomers are generated after the constitutional isomers so it is important that the canonical name for the configurations of the stereocenters can be appended to the canonical name for the constitution. In this respect, this canonicalization procedure differs from a previously reported one.[6]

***R/S* and Cis/Trans Designations.** Once a stereoisomer has been generated, it is possible to compute these designations from the integer representations and the original node numbers. The cis/trans designations for pairs of substituents on opposite sides of a double bond are easily computed by noting the relative configurations of the two stereocenters of the double bond and the ordering by node number of the attached substituents. The two substituents with larger numbers are trans if the two stereocenters have the same configuration (e.g., **1b, 15a,b**). The other possible designations follow similar rules. *R/S* designations are computed only for those stereocenters whose configuration depends on constitutional differences according to the Cahn–Ingold–Prelog convention.[8] These are the stereocenters that do not have symmetry elements passing through them. The algorithm used is the one described in ref 8; that is, comparisons of atomic numbers based on extended connectivity are made for each pair of substituents at each stereocenter. Once all the comparisons have been done, the substituents can be ordered by the *R/S* priorities at each stereocenter. This ordering is then compared with the ordering based on node numbers. That is, starting with the four substituents at a stereocenter ordered by increasing node numbers, convert this ordering to the ordering by increasing Cahn–Ingold–Prelog priorities. This conversion is done by a permutation of the four substituents. If this permutation is even, then *R* corresponds to 0 and *S* corresponds to 1. If this permutation is odd, then *R* corresponds to 1 and *S* corresponds to 0. (See Appendix for a discussion of even and odd permutations.) This information is summarized in Table III and structures **16a,b**.



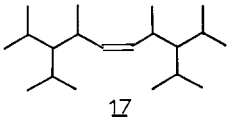
16a 16b

1 is the highest C-I-P priority

This conversion, to give the correspondence between *R/S* designations and 0/1 designations for all the stereocenters, is done only once for each structure since the priorities based on stereochemical differences (and chiral axes and planes) are not computed. All the stereoisomers of a single constitutional structure make use of the same correspondence.

III. Example

The algorithm will be illustrated with the hydrocarbon **17** as an example. The processing of the structure to find its



17

stereocenters is illustrated in Figure 2. The input structure (A) with the atoms numbered is shown at the top. The graph symmetry group of structure A is computed and is found to be a group of order 128. The edges of the double bond are labeled with fictitious bivalent nodes, and these are given numbers. All methylenes and methyls are designated as nonstereocenters and are indicated by asterisks in structure B of Figure 2. The symmetry that corresponds to switching the two fictitious nodes labeling the edges of the double bond is computed. This is a group of order 2 and is combined with the previously determined graph symmetry group to yield a combined group of order 256. Next, the carbons with dimethyl substituents are

Table IV. Configuration Symmetry Group for **17**

group element[a]	(1) (2) (3) (4)		(1′2′) (34)		(1) (2) (3′) (4′)		(1′2′) (3′4′)
no. of orbits	4		2		4		2
counting term	16	+	4	+	0	+	4

[a] The permutations are for the numbered stereocenters in Figure 2. Superscripts on the group numbers indicate stereocenters whose configuration is inverted by the permutation.

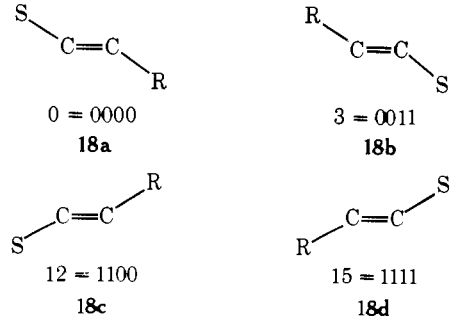
Table V. Generation of the Possible Distinct Stereoisomers of **17**[a]

group elements	stereoisomers					
	0	1	2	4	5	6
(1) (2) (3) (4)	0	1	2	4	5	6
(1′2′) (34)	3	1	2	7	9	10
(1) (2) (3′) (4′)	12	13	14	8	9	10
(1′2′) (3′4′)	15	13	14	11	5	6

[a] Each column corresponds to one stereoisomer. Each element of the configuration symmetry group acts on the integer (stereoisomer) on the top row to yield an equivalent integer (stereoisomer) below it.

found and marked as nonstereocenters, since these carbons have symmetrically equivalent substituents that themselves contain no stereocenters (structure C). This procedure is repeated, and it is found that the two diisopropyl substituted carbons (3 and 8) are nonstereocenters since the isopropyl groups now are known to not include stereocenters. These are indicated by asterisks in structure D. This leaves atoms 4, 5, 6, and 7 as stereocenters, as indicated by large dots in structure D of Figure 2. These are renumbered as shown, the tetrasubstituted carbons being given the lower numbers. Now that the stereocenters are found, the configuration symmetry group is computed. This group contains only four permutation-inversions and is shown in Table IV. The permutations are of the stereocenters only, and the superscripts indicating stereocenter inversion follow the notation defined in ref 3. The number of stereoisomers is computed using the configuration symmetry group and the combinatorial equation given in ref 3. The second row of Table IV gives the number of orbits (see Appendix) in each permutation and the final row gives the contribution of each term. The number of stereoisomers is this total divided by the order of the group, $\frac{1}{4}(16 + 4 + 4) = 6$.

Generation of distinct stereoisomers is accomplished by computing the symmetry equivalence classes of the 16 potential stereoisomers as discussed in ref 3. These 16 potential stereoisomers correspond to the integers from 0 to 15. The six resulting equivalence classes are shown in the columns of Table V. Starting with stereoisomer 0, it is found—by multiplication of the four elements of the configuration symmetry group—that stereoisomers 3, 12, and 15 are equivalent to 0. These four, **18a–d,** correspond intuitively to rotated structures. Stereoisomers 3, 12, and 15 need not be considered further. Stereoisomer 1 is considered next since this is the lowest integer (stereoisomer) not already generated, and the process continues until all 16 stereoisomers are accounted for. Each equivalence class (distinct stereoisomer) is represented by the lowest integer, and these are indicated with the stereoisomer shown in Figure 2. The correspondence between the configurations of the stereocenters and the *R/S* designations is shown by **19a,b.** For both tetracoordinate stereocenters, the 0 configuration is shown. Note that this corresponds to *S* for one and *R* for the other.



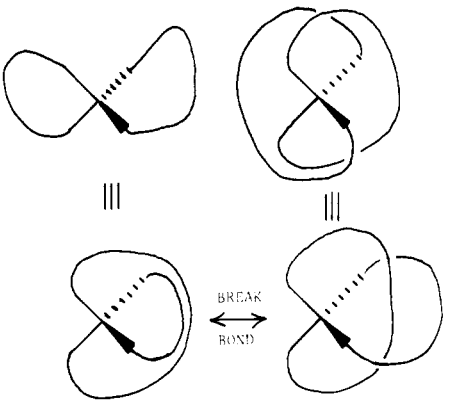
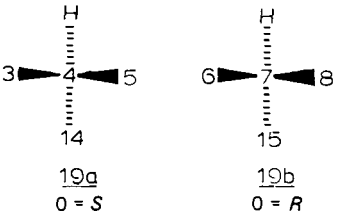


Figure 3. Illustration of the interconversion of two possible spiranes by breaking one bond.



Limitations. Intrinsic limitations of the stereoisomer generation algorithm and its computer implementation exist because the method contains no information about overall molecular geometry or energetics.

The final representation of a stereoisomer consists of the atom to atom connectivity with parity labels for the stereocenters.[3] This includes nothing about the positions of the atoms in space, bond angles, or dihedral angles, etc. Hence, there is no specification of the conformation of the structure. Structures that differ by rotations about single bonds receive the same configurational specification. This includes interesting cases of polycyclic structures that can be turned "inside out".[9] Structures that can be interconverted by passing bonds through bonds are also not differentiated, and these include catenanes[10] and interesting structures such as the inverted spiran shown in Figure 3, which can be interconverted with the usual spiran as shown.

No account is taken of energetics, hence the relative stability of the possible stereoisomers. This is a particularly severe problem for structures with a great deal of unsaturation, since many of the possible stereoisomers will require trans double bonds in small rings or trans ring junctions, etc., which are not likely to be observed as stable ground-state structures. In many cases the resultant strain may force a severe distortion in bond angles so that a resemblance to tetrahedral coordination is lost. In all of these cases the number of stereoisomers generated is simply an upper limit to the number that is likely to ever be observed. One objective of a constrained stereoisomer gener-

Table VI. Number of Isomers

emp form	constitutional	stereoisomers
C_4H_4	11	22
$C_5H_{11}(OH)$	8	11
$C_5H_9(OH)$	35	68
C_6H_8	159	514
C_6H_{12}	25	38
C_6H_{14}	5	5
$C_6H_{13}(OH)$	17	28
$C_6H_{11}(OH)$	100	238
$(CH)_6$	6	33[a]
C_7H_{14}	56	101
C_7H_{16}	9	11
$C_7H_{15}(OH)$	39	74
C_8H_{16}	139	299
C_8H_{18}	18	24
$C_8H_{17}(OH)$	89	199
t-$C_8H_{17}(OH)$	17	27
$(CH)_8$	20	476
C_9H_{20}	35	55
$C_{10}H_{22}$	75	136
$C_{10}H_{16}$	21	162[b]

[a] Aromatic benzene. [b] Constrained; see ref 14 and 15.

ator would be to eliminate unrealistic cases early in the generation process. Many of these unrealistic cases result from structural features for which only certain of the possible stereoisomers are likely to be observed. Examples are rings with double bonds that have fewer than seven atoms and polycyclic structures with small rings. These structural features could be found using a feature analogous to the "BADLIST" feature of CONGEN,[2c] and the stereoisomer generation could be constrained to produce only those stereoisomers with the desired relative configurations of the stereocenters.

Two other limitations are that only atoms which are at most tetravalent are considered at present and that the computer implementation only allows structures with 72 or fewer nonhydrogen atoms. The current computer implementation recognizes D, C, N, O, F, Si, S, Cl, and Br. Silicon is treated like carbon, and sulfur is treated as a stereocenter depending on valence. Nitrogen stereochemistry is discussed above. Any other atom that is at most tetravalent can be easily added.

Verification. The danger always exists that the implementation of an algorithm as a computer program will contain undetected errors since it is not yet possible to prove large computer programs. The program described here has been checked in several ways.

First, there exist a number of known results for the number of stereoisomers for given empirical formulas and structures. In particular, enumeration formulas and a tabulation of results have been given for empirical formulas corresponding to acyclic systems.[2d] Substantial collections of these have been given.[2d,11] In all cases, the program has agreed with these enumerations. Some of the results are given in Table VI. In many cases it is possible to verify the results manually, particularly for the enumeration and generation of the stereoisomers of a single structure. This has been done for most of the single structures given in Table VII. This can be difficult, especially for polycyclic systems, since unusual conformational processes are possible that interconvert different stereoisomers.[9]

Second, the program independently generates and counts the stereoisomers so that in effect the program checks itself by ensuring that these two numbers are the same. This equality has been observed in all cases.

Finally, the algorithm itself is invariant to any change of the numbering of the atoms,[12] and this property can be checked by varying the input ordering of the atoms of the structure. This property has been verified for many of the test cases shown in Table VII.

Table VII. Number of Stereoisomers Possible for Various Structures

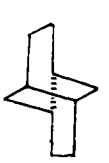
structure	total	chiral	achiral
cubane **22**	14	0	14
twistane **20**	7	4	3
inositol **21**	9	2	7
cyclo-$(Ala)_6$	14	12	2[a]
cyclo-$(Ala)_8$	36	32	4[a]
cyclo-$(Ala)_{10}$	108	104	4[a]
cyclo-$(Ala)_{12}$	352	344	8[a]

[a] See ref 16.

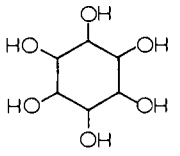
IV. Results and Discussion

Tabulation of the numbers of stereoisomers, along with other information, is given for various empirical formulas in Table VI and for individual structures in Table VII. The totals for empirical formulas that can only yield acyclic structures have been checked against the totals from a combinatorial formula.[2d] The number of possible isomers increases dramatically as the amount of unsaturation increases (e.g., C_6H_8). Many of the possible constitutional isomers for C_6H_8 are given in ref 2a. $(CH)_n$ systems have been subjected to extensive study and review,[13] largely with only considerations of constitutional isomerism. For these heavily unsaturated structures the number of possible stereoisomers is very large, although many of these are likely to be too strained to have more than a transient existence. The total, 162, in Table VI for $C_{10}H_{16}$ is for those isomers, such as adamantane, that have no multiple bonds, no methyl groups, and no three- and four-membered rings. This system has been the subject of considerable mechanistic study[14,15] in which stereochemistry was not considered to simplify the problem. Clearly, the number of possible stereoisomers is very large, and, while many of these are likely to be excessively strained, it is probable that several stereoisomers of one constitutional structure will be more stable than all the stereoisomers of another. That is, a ranking of the possible structures by increasing energy will likely encounter several stereoisomers of one structure before encountering any stereoisomer of another.

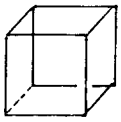
The stereoisomer totals of 14, for cubane, **22**, and other polycyclic systems given in Table VII, are most interesting in the context of higher homologues that have chains long enough



20



21



22

to allow the interesting conformational processes that these structures would exhibit.[9] The computer program is more conveniently applied to smaller molecules that yield the same total stereoisomers. Symmetric cyclic peptides (Table VII) have been studied extensively for their interesting stereoisomeric possibilities by Prelog,[16] who has given compilations for some cases. These totals agree in all cases that overlap (up to *cyclo*-$(Ala)_{10}$).

Experimental Section

The stereoisomer generator program is written in SAIL, an ALGOL-like language, and is part of the CONGEN program (for *con*strained *gen*eration) that generates the possible constitutional isomers. This program, the computer facility, and possibilities for user access have been described in ref 2a-c.

Acknowledgment. The encouragement of the members of the DENDRAL group, particularly Professors Bruce Buchanan, Edward Feigenbaum, and Joshua Lederberg, is appreciated.

The algorithm and program for symmetry group finding were provided by Dr. Harold Brown. Programming assistance was provided by Mr. Frederick Fisher. Editorial assistance was provided by Helen K. Tognetti. Manuscript preparation assistance was provided by Janet Kay Friendshuh. Financial support was provided by the National Institutes of Health Grant 2R24 R00612-08.

Appendix

Constitution and Constitutional Isomer. The constitution of a chemical structure is defined as the set of atoms and the set of bonds connecting them. For any pair of atoms there is an integral number of bonds from 0 to 3 connecting them. Constitutional isomers have the same set of atoms but differ in the set of bonds.

Order of a Group. The order is the number of elements in the group.

Orbits of a Permutation. The orbits are the sets of atoms made equivalent by the permutation. The permutation (123) (4) has two orbits, one of length 3 and one of length 1.

Even or Odd Permutations. For the symmetry group of the tetrahedron, even permutations are of the type (1) (2) (3) (4), (12) (34), (123) (4), and odd permutations are of the type (12) (3) (4) and (1234). (See ref 3 for further details.)

References and Notes

(1) Applications of Artificial Intelligence for Chemical Inference. 29. For part 28 see J. G. Nourse, *J. Am. Chem. Soc.*, preceding paper in this issue.
(2) (a) L. M. Mastiner, N. S. Sridharan, J. Lederberg, and D. H. Smith, *J. Am. Chem. Soc.*, **96**, 7702 (1974); (b) L. M. Masinter, N. S. Sridharan, R. E. Carhart, and D. H. Smith, *ibid.*, **96**, 7714 (1974); (c) R. E. Carhart, D. H. Smith, H. Brown, and C. Djerassi, *ibid.*, **97**, 5755 (1975); (d) R. C. Read in "Chemical Applications of Graph Theory", A. T. Balaban, Ed., Academic Press, New York, 1976, Chapter 4.
(3) J. G. Nourse, *J. Am. Chem. Soc.*, preceding paper in this issue.
(4) H. Brown, *SIAM J. Appl. Math.*, **32**, 534 (1977).
(5) F. A. Cotton, "Chemical Applications of Group Theory", 2nd ed., Wiley-Interscience, New York, 1971.
(6) W. T. Wipke and T. M. Dyott, *J. Am. Chem. Soc.*, **96**, 4834 (1974).
(7) Technically, this smaller group is a homomorphic image of the larger one.
(8) R. S. Cahn, C. Ingold, and V. Prelog, *Angew. Chem., Int. Ed. Engl.*, **5**, 385 (1966).
(9) C. H. Park and H. E. Simmons, *J. Am. Chem. Soc.*, **94**, 7184 (1972).
(10) G. Schill, "Catenanes, Rotaxanes, and Knots", Academic Press, New York, 1971.
(11) R. A. Davidson, Ph.D. Thesis, The Pennsylvania State University, 1976.
(12) J. G. Nourse, to appear.
(13) L. T. Scott and M. Jones Jr., *Chem. Rev.*, **72**, 181 (1972).
(14) H. W. Whitlock and M. W. Siefken, *J. Am. Chem. Soc.*, **90**, 4929 (1968).
(15) T. H. Varkony, R. E. Carhart, and D. H. Smith in "Computer-Assisted Organic Synthesis", W. T. Wipke and W. J. Howe, Eds., ACS Symposium Series No. 61, 1977, p 188.
(16) V. Prelog and H. Gerlach, *Helv. Chim. Acta*, **47**, 2288 (1964).

The Equilibrium Constant and Rate Constant for Allyl Radical Recombination in the Gas Phase

M. Rossi,† K. D. King,‡ and D. M. Golden*

Contribution from the Thermochemistry and Chemical Kinetics Group, SRI International, Menlo Park, California 94025. *Received May 4, 1978*

Abstract: The equilibrium and recombination 2 allyl ⇄ 1,5-hexadiene at $\langle T \rangle$ = 950 K and the recombination reaction at T = 625 K have been studied in a VLPP (very low-pressure pyrolysis) apparatus. The van't Hoff plot yields $\ln K_{r,d}/M^{-1} = (-34.56 + 10.60)/R + (56\,100/RT)$, which gives $\Delta H°_f$(allyl) = 39.1 ± 1.5 kcal/mol, a bond dissociation energy BDE(C_3H_5–H) = 86.3 ± 1.5 kcal/mol, and an allyl resonance energy ARE = 11.7 ± 2.0 kcal/mol. The recombination rate constant k_r at $\langle T \rangle$ = 900 K is found to be (1.90 ± 0.80) × 10^9 M^{-1} s^{-1}, and at T = 625 K, k_r is (6.50 ± 1.0) × 10^9 M^{-1} s^{-1}. RRKM calculations indicate a degree of fall-off k_r/k_r^∞ = 0.37 at 625 K and 0.045 at 900 K.

I. Introduction

Allyl radical is the prototype of a resonance-stabilized radical whose thermochemistry and reactivity in the gas phase have been the subject of numerous investigations.[1-3] There has been considerable controversy[1a] about the correct value of the heat of formation and the allyl resonance energy (ARE).[4] Quoted values for ARE range from 10 to 25 kcal/mol, but recent experimental values for ARE now seem to fall around 11 ± 2 kcal/mol. The correct value for ARE is certainly fundamental for a thorough understanding of chemical bonding and ground-state properties of conjugated radicals. Theoretical calculations on open-shell species, such as allyl radical, are hampered somewhat by intrinsic difficulties,[5] but, despite this problem, the generalized valence-bond (GVB) concept appears to be quite successful.[6]

Furthermore, the need for the accurate determination of radical–radical recombination rate constants has been clearly pointed out.[7] Absolute rate constants for many disproportionation reactions and radical–molecule reactions are critically dependent on the rates for combination of the radicals, since many rate constants have been measured relative to the radical combination rate constants. Thermochemical parameters of free radicals can also be obtained from the Arrhenius parameters of free-radical combination in cases where the reverse reaction (dissociation) has been studied. Radical recombination rate studies in the case of resonance-stabilized radicals are sparse and the rates for allyl and 2-methallyl recombination have only been measured at ambient temperature by flash photolysis.[8,9] It seemed appropriate to extend the rate measurements in order to get a reliable set of Arrhenius parameters for allyl recombination at higher temperatures where this reaction is the prototype for important chain terminations.

In this paper, we report (a) the equilibrium and kinetics of allyl radical recombination at 844 K ≤ T ≤ 1061 K, and (b) the recombination kinetics of allyl radical at T = 625 K. The method used is very low-pressure pyrolysis (VLPP) employing a newly designed molecular beam sampling apparatus, thereby eliminating complicated secondary reactions of radicals on the walls of the mass spectrometry chamber.[10]

† Postdoctoral Research Associate.
‡ On leave from Department of Chemical Engineering, University of Adelaide, Adelaide, South Australia 5001.

0002-7863/79/1501-1223$01.00/0