

main coupled to one carbon atom and two remain coupled to the other. Moreover, all the spectra in Figure 5 indicate that the  $\alpha$ - and  $\beta$ -carbon atoms are nmr non-equivalent in toluene and benzene as well as in pyridine.

These results do not rule out the possibility that process 1 occurs, or even that it occurs rapidly, but, if it does, rotation of the olefin cannot be rapid. If process 1 occurred rapidly the  $^{13}\text{C}$  spectrum would not necessarily be very different from that expected if it did not occur at all. If the instantaneous concentration of the olefin intermediate were small, the magnitudes of the  $^1\text{H}$ - $^{13}\text{C}$  couplings would not be measurably affected. In any case the H- $\alpha$ -C couplings would be preserved. If spin correlation were preserved in the transfer of a  $\beta$ -H from the carbon atom to the nickel atom, which is entirely possible, the methyl  $^{13}\text{C}$  quartet would also be preserved essentially unchanged. Thus, we cannot rule out the occurrence of process 1 but we can say with certainty that *if it occurs* it is not followed by rapid rotation of the olefin.

The nmr equivalence of the methyl and methylene hydrogen atoms at 100 MHz is apparently due to accidental equivalence of their chemical shift values under the conditions of measurement. Their separa-

tion in the solvent pyridine may be due to coordination of the pyridine to the nickel atom, thereby altering its electronic structure and, in turn, the relative chemical shifts of the methyl and methylene protons. It is also possible that the polar nature and diamagnetic anisotropy of the pyridine molecule cause it to associate with molecules of 1 in such a way as to shift the two sets of protons differently, quite apart from any specific donor interaction toward the nickel atom.

**Acknowledgments.** We are grateful to the National Science Foundation for financial support under Grant No. 33142X and for matching funds used to purchase the JOEL spectrometer. We thank The Robert A. Welch Foundation for funds used to purchase the Varian spectrometer and the diffractometer.

**Supplementary Material Available.** A listing of structure factor amplitudes (Table I) and root-mean-square amplitudes of vibration (Table IV) will appear following these pages in the microfilm edition of this volume of the journal. Photocopies of the supplementary material from this paper only or microfiche (105  $\times$  148 mm, 24 $\times$  reduction, negatives) containing all of the supplementary material for the papers in this issue may be obtained from the Journals Department, American Chemical Society, 1155 16th St., N.W., Washington, D. C. 20036. Remit check or money order for \$4.00 for photocopy or \$2.00 for microfiche, referring to code number JACS-74-4820.

## Simulation and Evaluation of Chemical Synthesis. Computer Representation and Manipulation of Stereochemistry<sup>1a</sup>

W. Todd Wipke\* and Thomas M. Dyott<sup>1b</sup>

*Contribution from the Department of Chemistry, Princeton University, Princeton, New Jersey 08540. Received August 9, 1973*

**Abstract:** An ordered-list representation of double bond and asymmetric carbon atom stereochemistry is described which facilitates machine manipulation and analysis. Procedures have been developed and implemented to transform a three-dimensional structure or a two-dimensional structural diagram having wedged and hashed bonds into the ordered-list connection table, and to transform an ordered-list connection table into a stereochemically unambiguous structural diagram. Stereochemical requirements and selection rules of reactions, or their inverse, transforms, are represented in *ALCHEM*, a machine readable English-like language. Algorithms are outlined for manipulating the ordered-list connection table of a target structure according to the selection rules of a given transform, producing all stereochemically correct synthetic precursors of that target. From these results many types of problems involving stereochemistry are now amenable to computer analysis. The described algorithms have been utilized in a computer program for the design of organic syntheses.

The first work in applying computers to the problem of planning an organic synthesis<sup>2</sup> considered molecules in a topological sense, devoid of any shape or spatial arrangement. Although the structural diagram entered by the chemist, and produced as output by the program, appeared to carry three-dimensional information, this information was ignored by the program. In many synthetic targets, e.g., prostaglandins, stereochemistry is the primary feature to be controlled. Further, the

spatial arrangement of groups in a target can provide powerful heuristics to help select the applicable synthetic procedures for constructing that spatial arrangement. Many chemical reactions occur according to known stereochemical selection rules, e.g., Woodward-Hoffmann rules; a complete representation of chemistry should embody these principles. This paper describes techniques for machine representation and analysis of stereochemistry. While we are concerned here with incorporation of stereochemistry in a computer program for the design of organic syntheses (specifically the Simulation and Evaluation of Chemical Synthesis (SECS) program), these results are equally applicable to any computer approach which deals with chemical structures and changes in chemical structure.

(1) (a) Presented in part at the Symposium on Stereochemistry, 164th National Meeting of the American Chemical Society, New York, N. Y., Sept 1972; (b) NSF Trainee, 1969-1971.

(2) E. J. Corey and W. T. Wipke, *Science*, **166**, 178 (1969); E. J. Corey, W. T. Wipke, R. D. Cramer III, and W. J. Howe, *J. Amer. Chem. Soc.*, **94**, 421, 431 (1972); E. J. Corey, R. D. Cramer III, and W. J. Howe, *ibid.*, **94**, 440 (1972).

## Results and Discussion

**Outline of SECS Operation.** The SECS program operates on a PDP-10 computer with 48 K of user memory and a disk as auxiliary storage. The target molecule to be synthesized is entered by the chemist in one of the following ways. He may draw a two-dimensional diagram, or may trace a molecular model using a three-dimensional acoustic tablet.<sup>3</sup> Alternatively the connectivity and other necessary information may be entered *via* a teletype or read from a data file. When using a teletype, either 2-D or 3-D coordinates may be entered, but neither are required. SECS derives the necessary stereochemical information from any of the various input representations and then may use a model building module<sup>4</sup> to obtain a reasonable 3-D model from the initial input structure. With or without a 3-D model, SECS then analyzes the structure, selects appropriate chemical transforms, and applies the transforms to generate precursor structures in a "logic-oriented," retro-synthetic fashion. The precursors are evaluated by the machine and shown to the chemist for his evaluation. He or the machine selects a precursor as a sub-target and the process is repeated, building the familiar synthesis tree.<sup>2</sup> Further details on the general operation of SECS will be given in a separate paper.<sup>3a</sup>

In this paper we will restrict discussion to the treatment of stereochemistry in SECS which for convenience has been subdivided into five categories which will be discussed in order: (1) development of stereochemical input conventions, (2) development of a machine representation of stereochemistry which allows facile manipulation, correlation, and querying, (3) representation of the stereochemical consequences and requirements of a synthetic transformation, (4) generation of all possible stereoisomeric precursors which may be converted into the target by a given synthetic reaction, and (5) generation of stereochemically correct structural diagrams for output.

**Stereochemical Input Conventions.** Our goal was to select a set of conventions already in use so any organic chemist could easily enter a complex structure into the computer without stereochemical ambiguity.

The stereochemical information of primary interest to the synthetic chemist, and therefore to the SECS program, is the configuration of various "stereocenters" in the molecule. We define a "stereocenter" as a center which if inverted produces a different stereoisomer,<sup>5</sup> *i.e.*, all centers, whose *configuration* must be correctly controlled to synthesize the desired stereoisomer. The two types of stereocenters commonly found in organic chemistry are asymmetric carbon atoms,<sup>6</sup> and carbon-carbon double bonds which are capable of *cis-trans* isomerism.<sup>7</sup> This definition of a stereocenter is consistent with the

(3) (a) W. T. Wipke in "Computer Representation and Manipulation of Chemical Information," W. T. Wipke, S. R. Heller, R. J. Feldmann, and E. Hyde, Ed., Wiley, New York, N. Y., 1974, pp 147-174; (b) W. T. Wipke and A. Whetstone, *Comput. Graphics*, **5** (4), 10 (1971).

(4) W. T. Wipke, P. Gund, and J. M. Verbalis, to be submitted for publication.

(5) For further discussion of stereocenters and the naming of stereoisomers, see W. T. Wipke and T. M. Dyott, *J. Amer. Chem. Soc.*, **96**, 4834 (1974).

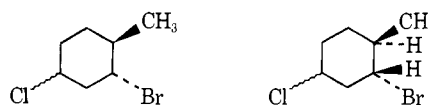
(6) K. Mislow, "Introduction to Stereochemistry," W. A. Benjamin, New York, N. Y., 1965, p 25. Note that while *trans*-decalin is achiral, it still has two stereocenters.

(7) Allenes and higher cumulenes can also be stereocenters if properly substituted. They can be treated in a manner analogous to double bond stereocenters.

needs of a synthetic chemist. We exclude from this discussion restricted rotation, helicity, and the stereochemistry of heteroatoms.

The myriad of notations<sup>8</sup> existing for representing the three-dimensional configuration of a stereocenter in two dimensions obscures any universal standard. In the carbohydrate field we find the Fischer and glycoside projection, in multicyclic systems we find large dots, open circles, wedges, bold lines, dotted lines, hashed lines, and often the use of mixed conventions.

The one system today which has the most generality uses hashed and wedged (or merely bold) bonds to indicate whether the bonds are directed behind or in front of, respectively, the plane of the diagram, or a parallel plane. The structural diagram below indicates that the six-membered ring is in the plane of the paper, and the methyl group and bromine are in front of, and in back of the plane of the paper, respectively.<sup>7</sup> Familiarity with the tetrahedral arrangement of bonds to a saturated carbon allows one to know the position of the implicit



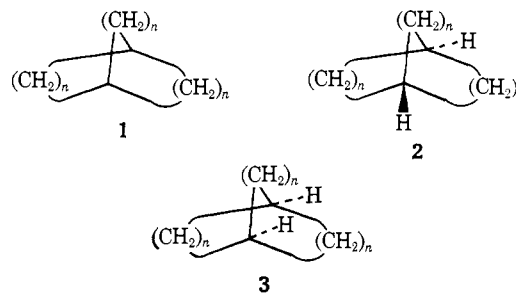
hydrogens as well. The wavy line indicates that the orientation of the chlorine is unknown.

Another often used convention is illustrated in the following structural diagram. If a saturated carbon is



drawn with only three substituents, and none of its bonds are wedged or hashed, then the implicit hydrogen is assumed to be down. This convention is violated almost as often as it is used.

Some structural diagrams require the reader to have inherent knowledge of the three-dimensional shape of the molecular skeleton. The bicyclic skeleton **1** as drawn is ambiguous, but is generally interpreted to mean structure **2** rather than structure **3**, although for



$n = 8$ , both structures have been made.<sup>9</sup> The chemist resolves the ambiguity of representation **1** on the basis that **3** would be too strained.

A consistent unambiguous system of stereochemical representation must be used to allow diagrams to be machine "understandable." The following conventions

(8) "IUPAC Tentative Rules for the Nomenclature of Organic Chemistry. Section E. Fundamental Stereochemistry," *J. Org. Chem.*, **35**, 2849 (1970).

(9) C. H. Park and H. E. Simmons, *J. Amer. Chem. Soc.*, **94**, 7184 (1972).

**Table I.** Atom Table<sup>a</sup> for Structure in Figure 1

Atom input no. <sup>b</sup>	ATYPE <sup>c,d</sup>	ASTEREO <sup>e</sup>	NBDS <sup>f</sup>	NATCH <sup>g</sup>	ATBD <sup>h</sup>	BD <sup>i</sup>
1	1	1	3	3	2 6 7	1 6 7
2	1	1	3	3	1 3 12	1 2 13
3	1	1	3	3	2 4 9	2 3 10
4	1	1	4	4	3 5 8 13	3 4 9 14
5	1	0	2	2	4 6	4 5
6	1	1	4	4	5 1 10 11	5 6 11 12
7	1	0	2	2	1 8	7 8
8	1	2	3	3	7 4 14	8 9 15
9	3	0	1	1	3	10
10	4	0	1	1	6	11
11	1	0	1	1	6	12
12	1	0	1	1	2	13
13	2	0	1	1	4	14
14	4	0	1	1	8	15

<sup>a</sup> Before ordering by perception. See text. <sup>b</sup> Atoms are numbered in the order they are entered. <sup>c</sup> Atom type C = 1, H = 2, N = 3, O = 4. <sup>d</sup> Charge and the *x*, *y*, and *z* coordinates for each atom are also represented but not shown here. <sup>e</sup> Atom stereochemistry; 0 = not stereocenter, 1 = stereocenter, 2 = unspecified stereocenter. <sup>f</sup> Number of valences used. <sup>g</sup> Number of atoms attached. <sup>h</sup> Atom input numbers of attached atoms. <sup>i</sup> Bond numbers of attached bonds in same order as atoms in ATBD. For bonds see Table II.

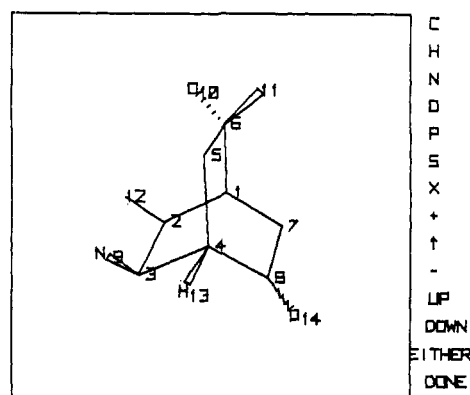
were adopted during the course of this work (also see Figure 1). (1) Bonds can be specified as proceeding in an upward or downward direction from an atom with respect to a plane passing through the atom and parallel to the plane of the diagram. The specification of such "stereobonds" is done with wedged bold and hashed bonds, respectively. The wedge is necessary to indicate which of the two atoms on the bond is the reference atom, from which the bond is up or down. The convention used is that the small end of the wedge is toward the reference atom. (2) An implicit hydrogen on a tertiary atom is assumed to be down if none of the explicit bonds to that atom is specified either up or down. (3) A bond can be specified by a wavy line to indicate the configuration at a center is either unknown, unimportant, or a mixture of both epimers.

**Structural Input.** The fundamental operations of entering a structural diagram using interactive computer graphics have been described.<sup>2,3</sup> The chemist first uses the pen to position the cursor over the control word DRAW (Figure 1) and depresses the pen to enter the DRAW mode. Then a bond is drawn by moving the cursor to the start of the bond, depressing the pen, moving the cursor to the end of the bond, and releasing the pen. Atoms correspond to *pen-up* or *pen-down* events and not to the intersection of lines. Therefore there is no atom at the intersection of the lines between atoms 1-2 and 4-5 in Figure 1. Multiple bonds are obtained by drawing over a bond again. Atom types other than C are obtained by selecting the type with the cursor and designating the atom to be that type. Hashed and wedged bonds are obtained by selecting DOWN and UP, respectively, and then designating a bond and then the atom about which the stereochemistry pertains. There is no need to break one bond when it crosses behind another. The drawing conveys the absolute configuration of each stereocenter and the constitution of the structure including how the atoms are connected together, but not the conformation. It will be shown later that the structural diagram is totally unambiguous stereochemically.

**Machine Representation of Stereochemical Configuration.** The structural diagram as entered (Figure 1) is represented by two tables: an atom table (see Table I), and a bond table (see Table II). Stereochemical in-

## SIMULATION AND EVALUATION OF CHEMICAL SYNTHESIS

MOVE DELET DRAW ERASE



**Figure 1.** Graphical entry of structure with stereochemistry indicated. Numbers shown are the input sequence numbers of the atoms. See Tables I and II for corresponding machine representation.

**Table II.** Bond Table for Structure in Figure 1

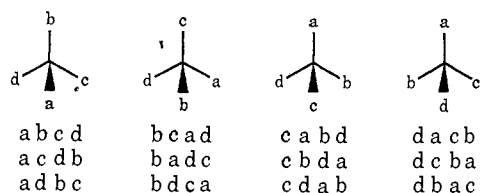
Bond input no. <sup>a</sup>	BTYPE <sup>b</sup>	BSTEREO <sup>c</sup>	AT1 <sup>d</sup>	AT2 <sup>d</sup>
1	1	0	1	2
2	1	0	2	3
3	1	0	3	4
4	1	0	4	5
5	1	0	5	6
6	1	0	6	1
7	1	0	1	7
8	1	0	7	8
9	1	0	8	4
10	1	1	3	9
11	1	6	6	10
12	1	1	6	11
13	1	0	2	12
14	1	1	4	13
15	1	4	8	14

<sup>a</sup> Bonds are numbered in the order they are entered. <sup>b</sup> Bond type, 1 = single, 2 = double. <sup>c</sup> Bond stereochemistry, 0 = none, 1 = AT2 up with respect to AT1, 6 = AT2 down wrt. AT1, 4 = AT2 either up or down wrt. AT1. <sup>d</sup> Atom number of one of the atoms involved in this bond. For atoms see Table I.

formation from the diagram is initially stored in the bond table as a code (BSTEREO) indicating the position

of the second atom of a bond (AT2) with respect to the first atom of that bond (AT1). Tables I and II correspond to the structure in Figure 1 immediately after input before perception has occurred. For internal processing another representation is utilized which is more systematic. The configuration of a tetrahedral carbon atom can be preserved by retaining a list of the attachments ordered in a specified manner.<sup>10</sup> Such an ordered list is inherent in the *RS* nomenclature system of Cahn, Ingold, and Prelog.<sup>11</sup> Since the connection table (Table I) contains an unordered list of attachments for each atom, the configuration of a stereocenter may be represented by simply ordering the corresponding list. The convention adopted in the SECS program orders the list of attachments so that viewing down the bond from the first attachment to the central atom, the other attachments are arranged in a clockwise manner.<sup>12</sup> This is equivalent to viewing from opposite the fourth attachment and citing the first three in a clockwise manner, which is the orientation used by Cahn, Ingold, and Prelog for an *R* configuration.<sup>11</sup>

A stereocenter with four explicit attachments to it can be represented by any of 12 possible list orderings as

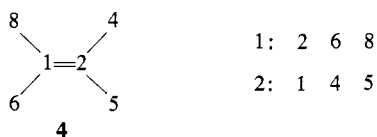


Equivalent orderings are related by an *even* number of pairwise interchanges of the attachments, since each pairwise interchange is a symbolic inversion of the stereocenter. Thus, the two of the orderings,  $\langle a, b, c, d \rangle$  and  $\langle d, a, c, b \rangle$  can be related as



If the stereocenter has an implicit hydrogen attachment,<sup>13</sup> the number of possible orderings is reduced to three since the implicit hydrogen is always considered to be fourth. (This prevents holes in the list of attachments.)

Double bond stereocenters are represented in a similar manner by ordering the attachments to the two atoms involved in the bond; attachments are ordered such that when viewed from the same side of the plane of the double bond they appear in a clockwise order, as shown below. Again equivalent orderings are gen-



(10) A. E. Petrarca, M. F. Lynch, and J. E. Rush, *J. Chem. Doc.*, 7, 154 (1967); J. E. Blackwood, C. L. Gladys, A. E. Petrarca, W. H. Powell, and J. E. Rush, *ibid.*, 8, 30 (1968); A. E. Petrarca and J. E. Rush, *ibid.*, 9, 32 (1969).

(11) R. S. Cahn, C. Ingold, and V. Prelog, *Angew. Chem., Int. Ed. Engl.*, 5, 385 (1966).

(12) W. T. Wipke, *Proc. Northeast Regional Electronics Meeting*, 12, 186 (1970).

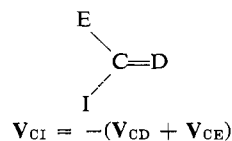
(13) While a chemist may explicitly draw hydrogen atoms during input, in general hydrogen atoms are not represented explicitly in order to reduce storage and processing requirements.



erated by an even total number of pairwise interchanges of the attachments in either or both lists. Note that the implicit H on the double bond must be last; hence, in **5** for atom 1: 2 6, not 6 2.

**Machine Perception of Configuration.** The task now remains to translate the two-dimensional structural diagram into the internal ordered list representation. A double bond, being two-dimensional and being represented in two-dimensional space, presents less problem than an asymmetric carbon atom. Perception of the spatial configuration of double bond stereocenters entails a simple algorithm based on the fact that *cis* attachments should be closer to each other than *trans* attachments. For example, in the structural diagram **4**, atoms 4 and 8 are *cis*, and 5 and 6 are *cis* since  $|\overline{48}| + |\overline{56}| < |\overline{46}| + |\overline{58}|$ . The connection table representation is then generated by placing the doubly bonded atoms first on the list, followed, in corresponding positions, by the attachments which are *trans*: 1: 2 6 8; 2: 1 4 5.

If there are implied hydrogens on the double bond, their positions are calculated using the formula



or

$$V_I = -(V_D + V_E) + 3V_C$$

where  $V$ 's are position vectors,  $I$  is the implicit hydrogen,  $C$  is the doubly bonded atom bearing the implicit hydrogen,  $E$  is the explicit attachment bonded to  $C$ , and  $D$  is the other atom on the double bond. Any holes in the attachment list due to implicit hydrogens are moved to the end by performing an even number of interchanges. To correctly handle even very crude diagrams, the lengths of the bonds adjacent to the double bond should be normalized prior to the calculation of the positions of any implicit hydrogens, but we do not normalize these bonds and have experienced no difficulties with interpretation in actual use by chemists.

Perception of the spatial configuration of asymmetric carbon stereocenters is performed in one of two ways depending on whether the input is two or three dimensional. If the initial synthetic target has 3-D coordinates, then the perception of configuration is a rather straightforward mathematical process, using only the coordinates of the stereocenter and its attachments. If the stereocenter bears an implicit hydrogen, its position is calculated from the positions of the stereocenter and the three explicit attachments using the formula

$$V_H = -(V_1 + V_2 + V_3) + 4V_C$$

The attachment list in the connection table is initially in the arbitrary order of input. If this order correctly represents the configuration at the stereocenter, the vector cross-product  $\overline{12} \times \overline{13}$  will lie on the same side of the 123 plane as attachment 4, as illustrated. If the cross-product is not on the same side of the plane as atom 4, the symbolic configuration in the connection

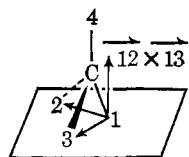
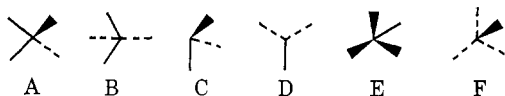


table is inverted, by interchanging two of the attachments on the list, to obtain the correct representation. Requiring only one vector cross-product and one vector dot-product, this algorithm efficiently assigns the configuration of the tetrahedron "closest" to the actual arrangement in all cases, even when the tetrahedron is greatly distorted.

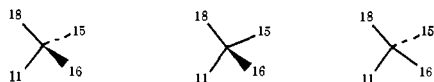
From a two-dimensional structural diagram the configuration is perceived using the *X* and *Y* coordinates and any significant stereobonds. This algorithm temporarily assigns positive *Z* coordinates to wedged attachments and negative *Z* coordinates to hashed attachments, other attachments remaining *Z* = 0. The 3-D perception algorithm just described is then employed to order the attachment list. Finally, the *Z* coordinates are restored to zero. This method is the most general one, being able to correctly interpret any unambiguous wedged/hashed representation, e.g.



Ambiguous representations include the following.



Prior to discovery of the above general algorithm, the following less general algorithm was utilized. This method uses the stereobonds symbolically in conjunction with the two-dimensional coordinates. If a bond is specified *up* with respect to the stereocenter of interest, its attachment is placed first on the list; if a bond is specified *down*, its attachment is placed fourth on the list. The free positions are filled by citing the remaining attachments in clockwise order. As usual, holes in the list due to implicit hydrogens are permuted to the end by two pairwise interchanges. The three partial structures



all lead to the same ordering of the attachment list, (16, 11, 18, 15), or an equivalent ordering. Ambiguous diagrams G, H, and I are also ambiguous to this algorithm but additionally any diagrams with more than one wedged (diagram E) or more than one hashed bond (B, D, F) to a stereocenter are ambiguous to this method.

In either algorithm, ambiguity also arises if none of the bonds to a stereocenter have been designated as *up* or *down*. If the center has four explicit attachments (case I) the ambiguity cannot be resolved, but if the center bears an implicit hydrogen, the ambiguity can be resolved by the implicit down hydrogen rule stated previously. The implicit hydrogen, being *down*, goes in the fourth position. The other three positions are filled as usual by taking the explicit attachments in clockwise

order. For the following partial diagram, one of the three possible orderings is (11, 18, 15). The implicit



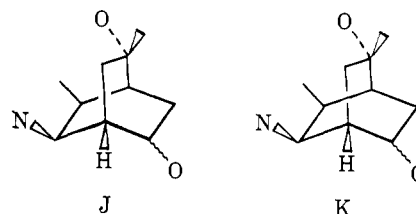
down hydrogen rule is used only when none of the bonds to the stereocenter have been designated as *up* or *down*.

If, when using either of the two methods described above for determining configuration from two-dimensional input, a stereobond about a center is found *up* or *down* with respect to the attachment instead of the stereocenter, then the sense of the bond is inverted for use with respect to the center in question. For example, the following two partial diagrams are considered to be equivalent.

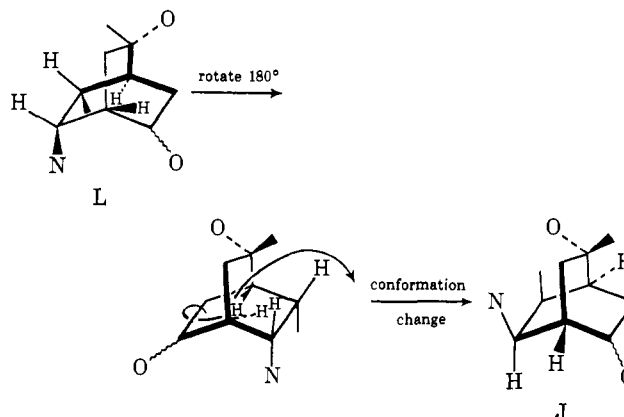


### Optical Illusions and Conformational Ambiguity.

Earlier in this paper it was stressed that it is not necessary to indicate which bond is in front when two bonds cross as in Figure 1. It might appear Figure 1 is ambiguous since it can be interpreted as the normal bicyclo[2.2.2] system J with both bridgehead hydro-



gens directed outward and is viewed from the top, or as the highly strained structure K in which both bridgehead hydrogens are directed inward and is viewed from the bottom. Surprisingly, J and K represent the same stereoisomer but two different conformations as is now shown. Because the vertical bridge and horizontal ring in K are viewed from sides opposite to that in J, the

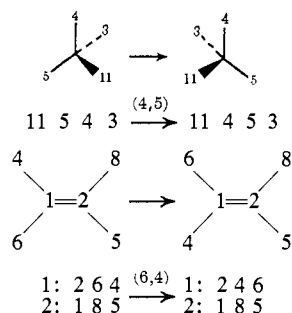


meaning of K is shown in L. Note the absolute configurations in J, K, and L are the same. L may then be turned inside out producing conformer J. Thus Figure 1 is ambiguous regarding conformation but is unambiguous regarding configuration and constitution (connectivity, and atom and bond types). Neglecting conformation, J and K represent the same structure because

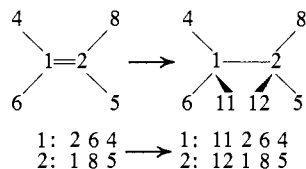
they have the same constitution and configuration at each stereocenter.

We ignore this conformational ambiguity in 2-D structural diagrams. However, when the conformation of a structure is needed, as it is in proximity or steric congestion measurements, our model builder generally produces the more acceptable conformation J, or a model may be traced in 3-D.<sup>3</sup>

**Symbolic Manipulation of the Machine Stereochemical Representation.** The usefulness of the ordered-list representation of configuration lies in the facility with which it can be manipulated for chemical reactions and stereochemical analysis. The simplest transformation is inversion of a stereocenter (cis-trans isomerization in the case of a stereocenter of the double bond type). On a molecular model, inversion is performed by physically interchanging two substituents; on the ordered-list representation inversion is performed symbolically by interchanging any two list entries.



Interconversion of the two types of stereocenters is important in, for example, the cis addition to a double bond. Symbolically, cis addition involves adding each

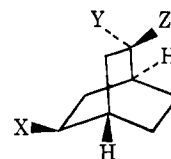


of the two new substituents to the beginning of the attachment list for the center to which the substituent is to be bonded. If the addition is to the opposite face of the double bond, the new attachments would be placed at the ends of the lists. Trans addition becomes a symbolic cis addition followed by symbolic inversion of one stereocenter.

Analysis of the stereochemistry of one group relative to another, establishing relative stereochemical relations (cis or trans), is an important perception process necessary to the planning of stereoselective syntheses. An implicit step in this analysis is the establishment of plane of reference.<sup>8</sup> For a double bond, the reference plane contains the doubly bonded atoms and is perpendicular to the plane of the double bond. In comparing two substituents with respect to this reference plane, the doubly bonded atoms are first permuted by an even number of pair interchanges to the same column; then if, in the respective ordered lists, the two substituents appear in the same relative position, the substituents are trans, otherwise cis.

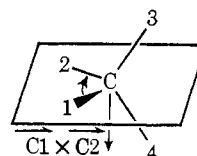
Other questions concerning the cis or trans relationship between two substituents generally assume as the reference plane the "plane" of a ring. Consider this problem working from the structural diagram. If the

representation of the ring were planar and convex, one could simply apply the rule that wedged substituents are on one side of the ring plane and hashed substituents are on the other; however, if the ring is concave or involves bonds crossing, as in bicyclo[2.2.2]octane,

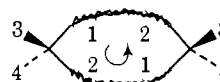


then such a simple rule would incorrectly imply that X and Y are trans, and X and Z are cis with respect to the ring they share. To solve this problem properly, the configuration of the stereocenters involved must be related not to the plane of the paper, but to an idealized "plane" of the ring.

Let us examine the stereocenter C with ordered list

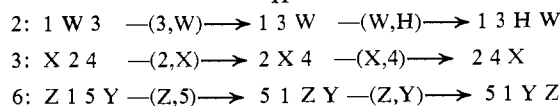
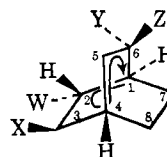


(1,2,3,4). Stereocenter C and the first two attachments form a plane. Attachment 4 is on the same side of that plane as  $C1 \times C2$ ; 3 is on the opposite side. If, for each center involved in the analysis, the  $C12$  plane is made to be the "plane" of the ring, and the lists of attachments are ordered so  $C1 \times C2$  lies on the same side of the ring plane for each center, then the attachments in position 3 are on one side of the ring plane, and those in position 4 are on the opposite side. All of the  $C1 \times C2$  vectors will be on the same side of the ring plane if at each center the 1 and 2 attachments are ordered as



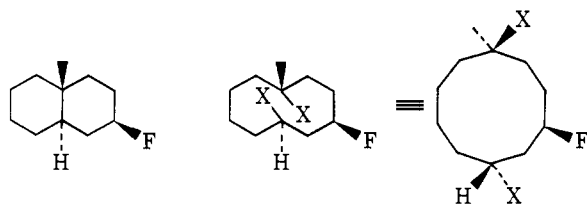
they are found going around the ring in a common *basis* direction. (The number of pairwise interchanges at each center must always be even to preserve the configuration.)

Applying this algorithm to the previous problem, let an arbitrary direction around the ring common to X and Y be the basis: 1,2,3,4,5,6,1,2... The lists are permuted to place the ring atoms in the first two positions, ordered in the *basis order*. Note the basis order for atom 6 is 5,1 and not 1,5. If after the ring atoms are ordered correctly an odd number of interchanges have been performed, as is the case in the example for atoms 2 and 6, then one more permutation is required to



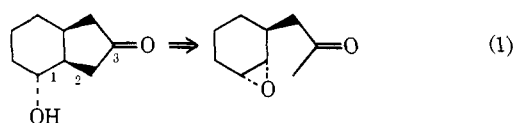
preserve the configuration. From the final ordered lists we conclude X and Y are cis with respect to this ring as are W and Z.

Care must be exercised to select the proper ring when there is more than one. The relationship obtained is that which would result from opening and flattening the ring into a planar convex form. In the case shown here, H and F are trans with respect to the six-membered ring, but they are cis with respect to the ten-membered ring. The proper basis which is used in SECS is the smallest ring containing both stereocenters in question.



**Stereochemical Description of Reactions.** The analysis methods and manipulations just described provide a set of fundamental operations for stereochemistry. A chemical reaction may consist of a sequence of these operations, may require a specific stereochemical relationship between reactive groups or bonds, and may produce new bonds or groups in the product in a specific stereorelationship. If we define a *transform* to be the inverse of a chemical reaction, then product stereorelationships become transform stereorequirements; stereorequirements of the reaction become the precursor stereorelationships created by the transform; and the sequence of operations for the reaction become inverted for the transform. Our goal was to allow the chemist to represent existing chemical principles and observations in the form of transforms in whatever degree of detail he chose. The following discussion pertains specifically to the stereochemical aspects of transforms.

A transform describes the factors which aid or hinder a reaction, the condition, and the structural changes needed to generate the precursor(s) from the target structure. This description is written in a machine-interpretable English-like language called ALCHEM.<sup>14</sup> The structure testing statement of ALCHEM allows the chemist to describe the transform stereorequirements. In the sequence



the inference that the epoxide may be a synthetic precursor to the alcohol is only valid if the hydroxyl group and the bond  $\beta$  to the ketone are trans. In ALCHEM this requirement can be stated as

```
IF BOND 1 IS A RING BOND THEN
BEGIN IF BOND 2 AND GROUP 1 ARE CIS THEN KILL
DONE
```

where the bonds are numbered from 1 to 3 from the atom bearing the alcohol to the carbonyl carbon, GROUP 1 is the alcohol, and KILL means eliminate this transform from consideration. Thus, through ALCHEM one can easily express the needed stereorequirements of a transform and prevent the generation of precursors *via* transforms which are stereochemically inappropriate.

(14) W. T. Wipke, T. M. Dyott, and C. Still, to be submitted for publication.

Additionally, the ALCHEM description of a reaction contains structure manipulation instructions, which describe the fundamental operations necessary to convert the target to the precursor. The basic manipulation instructions are

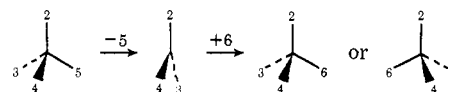
```
DELETE atom or charge
ADD atom or charge to another atom
BREAK BOND
MAKE BOND from one atom to another
INVERT ATOM n
LOSE STEREOCHEMISTRY AT ATOM n
```

DELETE removes a charge from an atom or removes an atom, breaking all bonds to that atom. ADD, the opposite of DELETE, puts a charge on an atom, or creates a new atom and bonds it to a specified atom. BREAK and MAKE BOND are self-explanatory. These four instructions have an assumed *suprafacial* course, *i.e.*, displacements proceed with *retention*; additions and eliminations proceed *cis*. The assumed course may be overridden by specifying an appropriate INVERT instruction, which inverts the configuration of the specified stereocenter, regardless of whether the center is an asymmetric carbon or a double bond stereocenter. The structure manipulation instructions for the *antarafacial* epoxide alkylation transform (eq 1) then should be

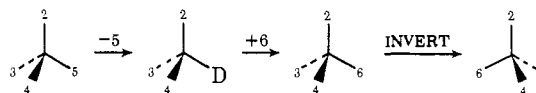
```
BREAK BOND 2
MAKE BOND FROM ATOM 1 IN GROUP 1 TO ATOM 2
INVERT ATOM 2
```

The remaining manipulation instruction, LOSE STEREOCHEMISTRY AT ATOM *n*, indicates that the specified center in the precursor can be *either* configuration, *i.e.*, either configuration would lead to the desired synthetic target.

Since a stereospecific reaction is a continuum, rather than a series of discrete steps, care must be taken to avoid the loss of stereochemical information by the use of a discrete step description. An ALCHEM description is a discrete step description, *e.g.*, an S<sub>N</sub>2 displacement is described by the BREAK BOND, MAKE BOND, and INVERT ATOM instructions, all discrete steps. Were the central atom allowed to collapse to essentially an sp<sup>2</sup> configuration in the connection table, the stereochemistry at the central atom would be lost, as illustrated. This loss of

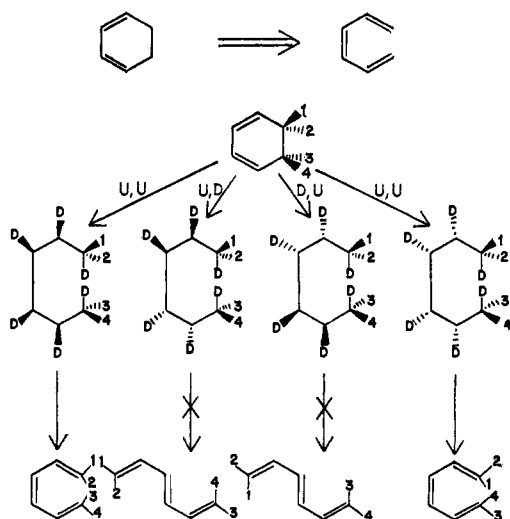


stereochemical information is avoided by replacing in the connection table any bond that is broken by a dummy to hold the configuration. The displacement then gives the desired results, as shown below.



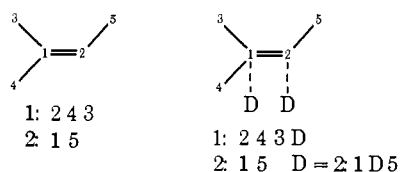
Stereochemical continuity throughout the transform is achieved in the sense of an assumed *suprafacial* course, by applying the following rules during execution of the structure manipulation instructions.

(1) Replace each broken bond with a dummy to hold the configuration. Multiple bonds which are completely broken are replaced with multiple dummies, *e.g.*, if a double bond is cleaved, two dummies are added to each of the atoms that formerly comprised the double bond.



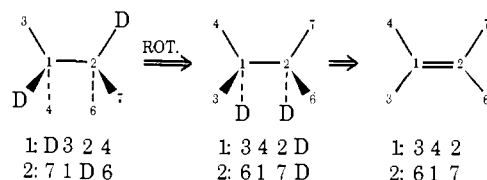
**Figure 2.** Thermally allowed hexatriene-cyclohexadiene rearrangement. The ALCHEM PERICYCLIC command adds dummies to the same face of the "ring," preventing absurd precursors.

(2) If breaking only the  $\pi$  bond of a double bond, add a dummy to each atom from the same face of the double bond, *i.e.*, place the dummies in the same columns. Application of this rule results in cis addition of dummies.



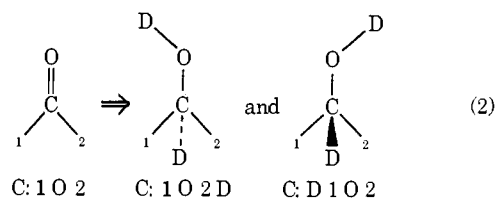
(3) When making new bonds, replace dummies, if there are any on the atoms in question, otherwise replace implicit hydrogens.

(4) If replacing a single bond with a double bond, remove the dummies and/or implicit hydrogens from the same side. Application of this rule results in cis elimination of dummies and/or implicit hydrogens.

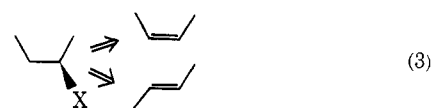


**Stereochemical Multiplicities.** Many transforms imply more than one possible precursor, as exemplified by eq 2-5. Rather than rely on the writer of the transform to foresee all possible multiplicities, a procedure which would most certainly be prone to error, we chose to have the program automatically detect multiplicities and generate all precursors implied by that transform. The algorithm that accomplishes this recognizes choice points encountered as the manipulation instructions are executed. There are two basic types of choice points. The first type occurs when a  $\pi$  bond is broken and an  $sp^2$  center becomes an  $sp^3$  center, a situation covered by rule 2. The dummies are to be added from the same side of the double bond, but which side shall we choose? An example of this type of choice is shown in eq 2. This choice corresponds to placing the

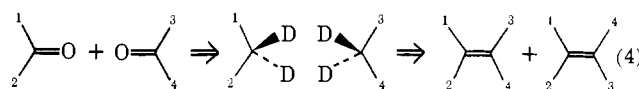
dummies in either the first or fourth position of the attachment list.



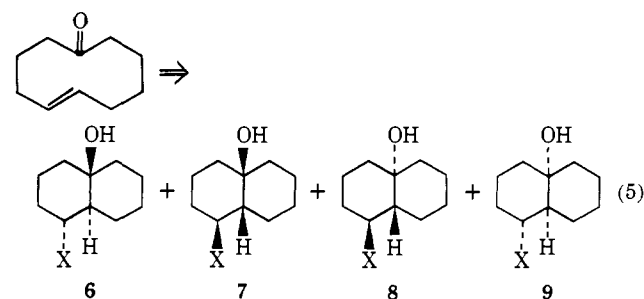
The second type of choice may occur when a bond is made (see rules 3 and 4). Two such situations are: making a bond to an atom which bears two dummies, or two implicit hydrogens (see eq 3), and forming a



double bond between two previously nonbonded carbons (eq 4).



All implied precursors are then generated by taking all combinations of possible choice points. Most transforms present only one such choice point or none at all, but the base-initiated ring opening (eq 5) illustrates the



complexity that can arise. Note in 6 and 7 that the relative orientation of the hydroxyl group is different, but the relationship of the leaving group to the bridging bond being broken is the same, being trans-periplanar in both cases. Experimentally it has been verified that the latter stereochemical relationship is required and that the orientation of the hydroxyl group is unimportant in this reaction.<sup>15</sup> Thus, both structures 6 and 7 would stereospecifically produce the trans olefin. In this case 8 and 9 are enantiomers of 6 and 7, respectively, so 8 and 9 would be deleted by SECS.

For most reactions, the "take all possible choices" technique generates exactly the set of precursors desired, but for certain pericyclic reactions, this technique leads to some undesired precursors. The thermal cyclohexadiene-hexatriene transform depicted in Figure 2 is such a reaction. The dummies may be added from either above or below each of the two double bonds, generating  $2^2$  or 4 precursors. These correspond to the two  $[\sigma_{2s} + \pi_{4s}]$  and two  $[\sigma_{2s} + \pi_{2a} + \pi_{2a}]$  "symmetry allowed" processes. Although all four precursors are theoretically allowed, only two are reasonable: the two

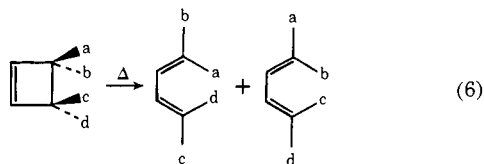
(15) E. J. Corey, R. B. Mitra, and H. Uda, *J. Amer. Chem. Soc.*, **86**, 485 (1964); J. A. Marshall, W. F. Huffman, and J. A. Ruth, *ibid.*, **94**, 4691 (1972).



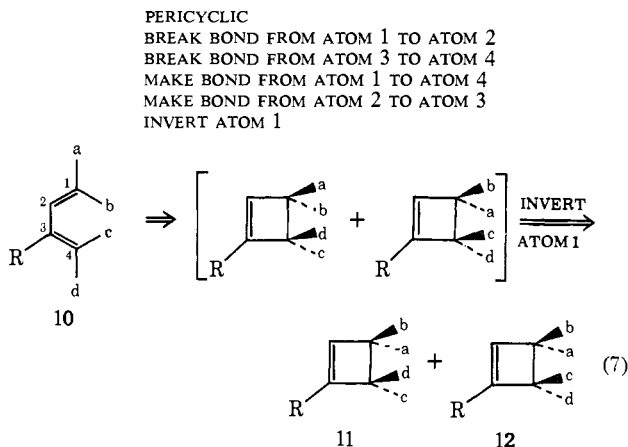
derived by the suprafacial addition of dummies to the entire  $\pi$  system. Such a transform is specified in ALCHEM to be pericyclic by the command PERICYCLIC preceding the manipulation instructions. This command causes the program to assign dummies to the entire  $\pi$  system in a cis manner. Since the reaction pathway describes a ring, there exists a simple basis or reference "plane" used in controlling the addition of dummies in a manner similar to that previously described for determining the relative stereochemistry of substituents on a ring.



Antarafacial pericyclic pathways may be specified with the aid of the appropriate INVERT commands. Consider the conrotatory ring opening reaction of a cyclobutene to a butadiene (eq 6). In the analytical



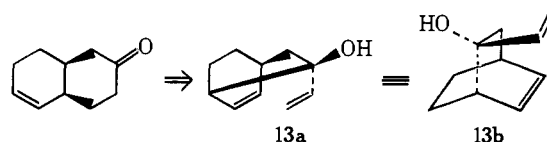
direction, the transform is a conrotatory pericyclic ring closure (eq 7). The ALCHEM manipulation instructions for this transform would be (see 10 for atom reference numbers)



During execution of this transform, dummies will be added first to one side of the reference plane of the diene, then to the other side. Bonds will be broken or made in a suprafacial manner producing two intermediates which are converted by the INVERT instruction to 11 and 12 as shown in eq 7. Note that inversion of either atom 1 or atom 4 leads to the same precursors, 11 and 12. If R is achiral, one precursor would be discarded, because 11 and 12 would be enantiomers.<sup>5</sup> But if R is chiral, both would be retained, because the two precursors would be diastereomers. Thus the complications of transform multiplicities, even in pericyclic reactions, are automatically considered by the program, leaving the writer of the transform responsible for only the fundamental chemical and stereochemical information.

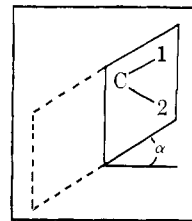
**Computer Generation of Stereochemically Valid Structure Diagram.** We have described how the structural diagram is transferred to the machine

representation, the connection table (CT), and how the CT can be manipulated by stereospecific chemical transforms. After such a transform has been applied to the CT of the target structure, the resulting CT correctly represents the precursor, constitutionally and configurationally. The CT, however, is not a useful end product for the chemist. Therefore, it is desirable to convert the precursor CT to the form most easily interpreted by the chemist, *i.e.*, an unambiguous structural diagram. The 2- or 3-D coordinates of the atoms in the target can be used as coordinates for the atoms in a precursor, but it must be remembered that the synthetic transform may have greatly changed the connectivity, and added, deleted, or inverted atoms. Coordinates are computed for atoms added and all atoms are connected according to the CT of the precursor. Appropriate



bonds in this possibly distorted diagram must then be selected for wedging and hashing to properly represent the configuration of stereocenters.

The key to this problem lies in relating the stereocenter to the projection surface, and the viewer. A stereocenter (C: 1 2 3 4) and its first two attachments may be used to define a plane (C12) which intersects the plane of the paper (or screen) with dihedral angle  $\alpha$ .

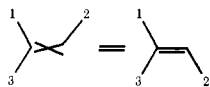


Assuming the viewer is a significant distance from the plane of the diagram and roughly perpendicular to it, unless  $\alpha$  is exactly  $90^\circ$ , we can assign one side of the plane as being toward the viewer. In a "right-handed" (*i.e.*,  $\hat{i} \times \hat{j} = \hat{k}$ , not  $-\hat{k}$ ) coordinate system, the  $X, Y$  plane is the plane for the diagram, and the viewer is on the  $+Z$  side of that plane. Thus, if the  $Z$  component of the vector  $C1 \times C2$  is on the viewer's side of the C12 plane, the C-4 bond should be wedged and the C-3 bond hashed; if  $C1 \times C2$  is on the side of the plane away from the viewer, the C-3 bond should be wedged and the C-4 bond hashed. This algorithm works equally well for 2-D structures as for 3-D structures.

When a stereocenter has two bonds within a ring and the remainder to appendages, the chemist prefers to hash or wedge the appendage bonds. This is easily accomplished by first permuting (by an even number of pair interchanges) the attached ring atoms into the first two positions before applying the above algorithm.

Since a bond can only be hashed or wedged once, it is necessary to check that the bonds to attachments 3 and 4 are not already hashed or wedged from some other atom. If they are, the algorithm can either use the existing designations in the inverse sense and designate other bonds to be consistent, or it can select the existing hashed or wedged bonds as the reference plane (C12) and proceed as normal.

The configuration about double bonds must also be conveyed to the viewer. If the  $X,Y$  projection of a double bond, as viewed by the user, is in agreement with the symbolic representation contained in the connection table, then the double bond is displayed normally; otherwise the disagreement is indicated by crossing the double bond, indicating the true configuration is opposite to that shown.



Using these techniques, SECS assures the chemist that the stereochemistry he sees is precisely that which is in the connection table of the precursor. From the distorted diagram **13a**, it is difficult to determine if the implied oxy-Cope is reasonable, but after atoms are moved as in **13b**, it is more apparent that the transform is stereochemically plausible.<sup>16</sup> As the chemist moves atoms manually to obtain an alternate view of the structure, the program dynamically modified the hashing and wedging to maintain the correct stereochemical representation.

### Conclusions

We have developed an unambiguous method for describing stereochemical configurations to a computer and for representing them within a computer in a way that facilitates stereochemical analysis. It was shown that stereochemical selection rules for common chemical reactions can be represented and applied by machine to

(16) W. L. Scott and D. A. Evans, *J. Amer. Chem. Soc.*, **94**, 4779 (1972).

a chemical structure. Finally, a simple algorithm was described that transforms the machine representation of a structure into an unambiguous structural diagram including proper stereochemical designations. This representation of stereochemistry provides the basis for naming stereoisomers uniquely and for recognizing enantiomers.<sup>5</sup> The same configurational information facilitates the generation of a stereochemically correct three-dimensional model,<sup>4</sup> which can be utilized in evaluating steric congestion,<sup>17</sup> and reaction mechanisms.<sup>18</sup> The described algorithms not only increase the selectivity of transforms, but also increase evaluation capabilities, allowing recognition of strained precursors, e.g., those containing a trans double bond in a small ring, or containing a *transoid* bridged ring system. Thus, it is now possible for a computer to assist in synthetic design, not only in the crude connectivity of molecules, but also in the fine details of stereochemistry. Subsequent papers in this sequence will illustrate actual syntheses produced by the SECS program using the principles described here.

**Acknowledgment.** The authors thank Dr. W. Clark Still and Peter Friedland for their contributions in the development of *ALCHEM* and Dr. Peter Gund for his work on the model builder which led to stimulating discussions concerning this work. The work was supported by the National Institutes of Health Research Resource Grant RR00578 and in part by the Merck Company Foundation and Eli Lilly and Company.

(17) W. T. Wipke and P. Gund, *J. Amer. Chem. Soc.*, **96**, 299 (1974).

(18) T. M. Gund, P. v. R. Schleyer, P. H. Gund, and W. T. Wipke, to be submitted for publication in *J. Amer. Chem. Soc.*

## Stereochemically Unique Naming Algorithm

W. Todd Wipke\* and Thomas M. Dyott

*Contribution of the Department of Chemistry, Princeton University, Princeton, New Jersey 08540. Received August 9, 1973*

**Abstract:** An algorithm has been developed and implemented to generate for each chemical structure a unique and invariant linear name which includes double bond and asymmetric carbon isomerism. A logical proof is given for the one-to-one correspondence between name and structure. By inspection of the linear names of two structures, one can determine if the two structures are identical, nonisomeric, constitutionally isomeric, diastereomeric, or enantiomeric. The algorithm determines the true stereocenters and calculates a reduced set of chiral centers,  $S_{RC}$ . It is proven that if there are any centers in  $S_{RC}$  that the compound must be chiral; an achiral compound must have  $S_{RC} = \text{null}$ . Extensions of the algorithm are outlined to allow uniquely naming conformational isomers.

Nonunique representations of chemical structures are useful for many things: general chemical nomenclature and discourse, chemical synthesis by computer, substructure searches, etc. Registry and storage-retrieval systems, however, require exact structural matches. Searching for such matches is greatly simplified if a canonical name can be assigned, since only one search of the structure file is then required. A canonical name means for each structure there is one name and for each name there is only one structure. This is of great importance in systems, such as the Chemical Abstracts Service registry system, where information pertaining to a compound is stored with an identify-

ing name. In our own work with the Simulation and Evaluation of Chemical Synthesis (SECS) program,<sup>1,2</sup>

(1) W. T. Wipke, P. Gund, J. G. Verbalis, and T. M. Dyott, Abstracts, 162nd National Meeting of the American Chemical Society, Washington, D. C., Sept 1971, No. ORGN-17; W. T. Wipke, T. M. Dyott, P. Gund, and C. Still, Abstracts, 164th National Meeting of the American Chemical Society, New York, N. Y., Aug 1972, No. CHED-39; W. T. Wipke in "Computer Representation and Manipulation of Chemical Information," W. T. Wipke, S. R. Heller, R. J. Feldmann, and E. Hyde, Ed., Wiley (1974). For related work see E. J. Corey and W. T. Wipke, *Science*, **166**, 178 (1969); E. J. Corey, W. T. Wipke, R. D. Cramer III, and W. J. Howe, *J. Amer. Chem. Soc.*, **94**, 421, 431 (1972).

(2) Other stereochemical aspects of the SECS program, including the perception of stereochemistry from two-dimensional structural diagrams, are described in the preceding paper, W. T. Wipke and T. M. Dyott, *J. Amer. Chem. Soc.*, **96**, 4825 (1974).