

A Method for Generating Unique Computer Structural Representations of Stereoisomers*

ANTHONY E. PETRARCA, MICHAEL F. LYNCH†, and JAMES E. RUSH
Chemical Abstracts Service, The Ohio State University, Columbus, Ohio 43210

Received March 6, 1967

A method is described for generating unique computer structural representations of stereoisomers from simple, unambiguous tabular descriptions (topological coding) of their chemical structures. These tabular descriptions (or connection matrixes) can be produced clerically, or by other input techniques, from conventional stereochemical formulas. The proposed method successfully deals with both optical and nonoptical (*cis-trans*) stereoisomerism and it is particularly suited to computer techniques for generic (substructure) searching in a file of compounds for substructures with specified stereochemistry.

The development of techniques for translating chemical structural information into machine notation has been a subject of considerable interest during the past 15 years (1). Such techniques are necessary for the manipulation of large files of structural formulas by computer methods or for conducting generic (substructure) searches by computer for all compounds with related structural characteristics.

Recent developments in this field include reports of several methods (2) for generating unique non-stereochemical computer descriptions of chemical structures from simple tabular descriptions (3) of all atoms and bonds (chemical topology) shown in their conventional structural formulas. The tabular descriptions can be produced clerically or they can be generated by other methods such as by automatic generation from linear notations (4), automatic optical scanning of structural formulas (5), mechanical translation from nomenclature (6), and automatic encoding by chemical typewriters (7). An important advantage of storing structural information in the computer record by this tabular atom-by-atom approach is that it permits wide latitude in generic (or substructure) searching (8), even for structural features not foreseen on the basis of ordinary classification schemes, because no aspects of structure are subordinated to others.

However, while this degree of sophistication has been attained in handling nonstereochemical aspects of structure, the development of appropriate computer techniques for handling detailed aspects of stereoisomerism has received little attention. This paper describes methods for including stereochemical detail in tabular descriptions (topological coding) of chemical compounds and for generating unique computer structural representations of stereoisomers therefrom. The information required at input is derived from conventional stereochemical formulas and it can be supplied clerically or by other methods of input such as those mentioned previously (4-7). The proposed

method handles both optical and nonoptical (*cis-trans*) stereoisomerism and it is particularly designed to permit generic searching by computer for substructures with specified stereochemistry.

BASIC APPROACH

The basic approach is to describe the three-dimensional aspects of chemical structure in terms of appropriate stereodescriptors in accordance with the symmetry characteristics (9) of each particular system. A similar approach has recently been described by McDonnell and Pasternack (10), but our subsequent treatment differs from theirs. The treatment is best illustrated on a tetrahedral asymmetric atom for which the spatial arrangement of the peripheral atoms attached thereto can be described in terms of a set of descriptors—A, B, C, and D—as illustrated for the reference structure in Figure 1. If the peripheral atoms are ranked numerically according to some prescribed order of hierarchy [The ordering rules we prescribe are those used for unique numbering of the atoms in a two-dimensional computer description; see (2b) and (2c) for illustrations.], one possible description for each

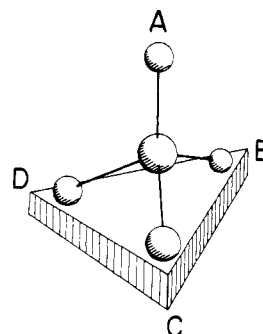


Figure 1. Reference structure for assigning stereodescriptors to peripheral atoms attached to a tetrahedral asymmetric atom.

* Presented in part before the Division of Chemical Literature, 150th National Meeting of the American Chemical Society, Atlantic City, N. J., September 1965.

† Present address: Postgraduate School of Librarianship, The University, Sheffield, England.

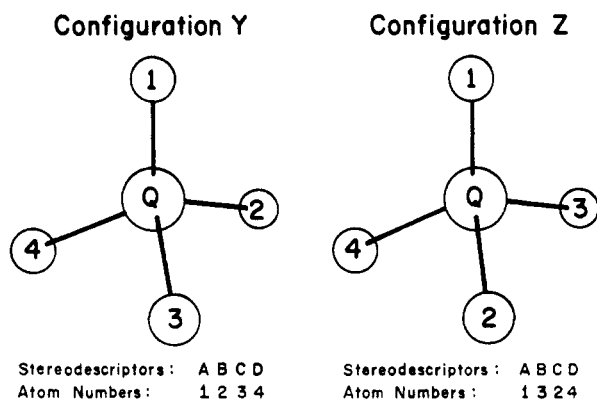


Figure 2. Descriptions of the two configurations for a hypothetical tetrahedral asymmetric atom Q.

configuration (designated here by the letters Y and Z) of a hypothetical asymmetric atom Q is illustrated in Figure 2. [The Y and Z descriptors are similar to the R and S descriptors of Cahn, Ingold, and Prelog (11). However, there is no one-to-one correspondence between the two sets of descriptors because the ordering rules are not the same.] The descriptions in Figure 2, however, are not unique, for there is a total of 12 possible permutations for the description of each configuration depending on how the asymmetric atom is oriented in space. Nevertheless, by applying the laws of permutations and inversions (12), a computer program can differentiate one configuration from the other in terms of the parity of the permutation. As illustrated in Figure 3, all of the permutations for configuration Y have even parity; conversely, all of those for configuration Z have odd parity.

	Spatial Orientation	Stereo-descriptors A B C D	No. of Inversions		Spatial Orientation	Stereo-descriptors A B C D	No. of Inversions
(a)		1 2 3 4	0	(g)		3 1 2 4	2
(b)		1 4 2 3	2	(h)		3 4 1 2	4
(c)		1 3 4 2	2	(i)		3 2 4 1	4
(d)		2 3 1 4	2	(j)		4 1 3 2	4
(e)		2 4 3 1	4	(k)		4 2 1 3	4
(f)		2 1 4 3	2	(l)		4 3 2 1	6

Figure 3. The twelve permutations (even parity) for description of configuration Y.

The treatment for geometrical isomerism about a double bond is somewhat similar, except that the doubly bound atom-pair and all of the peripheral atoms attached thereto constitute the structural group which is to be described stereochemically. Because of the coplanarity of all of the atoms in this group, only two stereodescriptors, A and B, are needed to describe the spatial arrangement of the peripheral atoms relative to the axis coincident with the double bond (Figure 4). Also, there are only four possible

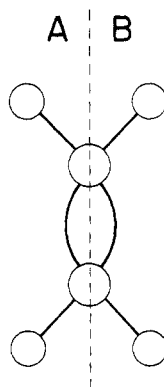


Figure 4. Reference structure for assigning stereodescriptors to peripheral atoms attached to a doubly bound atom-pair.

permutations for the description of each configuration—*e.g.*, *cis* or *trans*—in terms of the peripheral-atom subsets attached to the stereogenic atom-pair. [A stereogenic atom is defined as an atom bearing several groups of such nature that an interchange of two groups will produce a stereoisomer (13).] The number of permutations is reduced to two by ordering rules (2b, 2c) which would always cause the peripheral-atom subset attached to the higher ranked stereogenic atom to appear first. Once again, a computer program can differentiate between the two configurations (designated by the letters W and X in Figures 5 and 6, respectively) on the basis of the parity of the permutations for each combination of peripheral-atom subsets attached to the stereogenic atom-pair. The parity in this case is based on the sum of the number of inversions for each peripheral-atom subset in the combination, the permutations for configuration W being even (Figure 5), while those for configuration X are odd (Figure 6). (These descriptors are similar to the traditional descriptors *cis*- and *trans*-, or *syn*- and *anti*-. However, there is no one-to-one correspondence between W and X and the traditional descriptors because of the different basis for assigning them.)

Application of the W, X, Y, and Z configurational descriptors to obtain unique computer structural representations of specific stereoisomers will be illustrated later.

INPUT PROCEDURES

The operations just described are to be performed within the computer after the necessary information has been

	Spatial Orientation	Stereo-descriptors A B	Inversion Sum
(a)		1 2 3 4	0
(b)		3 4 1 2	0
(c)		2 1 4 3	2
(d)		4 3 2 1	2

Figure 5. Permutations (even parity) for description of configuration W on the hypothetical doubly bound atom-pair α and β .

supplied at input. There are three basic requirements for adequate clerical input of this information from conventional stereochemical formulas. These are: (a) recognition of a stereochemical formula, (b) recognition of the atoms which must be described stereochemically, and (c) procedures for describing the stereochemistry.

Three simple input rules will now be described to illustrate how these requirements can be met for most types of stereoisomerism.

1. The Haworth Formula Rule. The first rule is for input of stereoisomerism from Haworth formulas (14). These

	Spatial Orientation	Stereo- descriptors AB	Inver- sion Sum
(a)		1 2 4 3	1
(b)		4 3 1 2	1
(c)		2 1 3 4	1
(d)		3 4 2 1	1

Figure 6. Permutations (odd parity) for description of configuration **X** on the hypothetical doubly bound atom-pair α and β .

stereochemical formulas are easily recognized by the manner in which they are drawn with their ring attachments projecting up or down, and usually parallel to each other, in the plane of the paper. Any ring atom which has two different nonring attachments drawn in this fashion is usually a stereogenic atom (check routines can be written into the computer program to preclude erroneous treatment of nonstereogenic atoms which were interpreted to be stereogenic at input), and the stereodescriptors to be used for each of its peripheral atoms are as follows (Figure 7): (1) the **A** atom is *up*; (2) the **B** atom is *down*; (3) the **C** atom is *clockwise* in the ring; and (4) the **D** atom

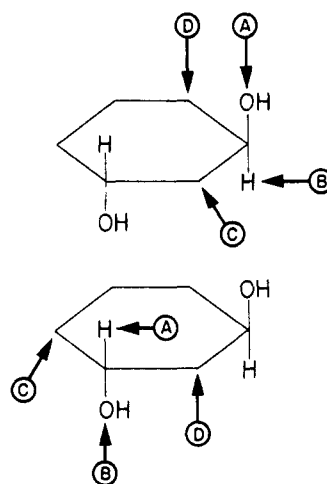


Figure 7. Assignment of stereodescriptors by the Haworth formula rule.

is *anticlockwise* in the ring. These assignments are made by recording peripheral-atom numbers (or H, for a hydrogen atom) in the appropriate stereodescriptor columns of a connection table (3) such as that illustrated for (+)- α -D-glucopyranose in Figure 8. (Conversion of these representations into their appropriate three-dimensional tetrahedral representations will show that the stereodescriptor assignments are in accord with those of the reference structure in Figure 1.) Connection tables of this type (without the stereodescriptor columns) have been described by Gluck (2b), Leiter (15), and others. Hydrogen must be identified by its elemental symbol in this treatment because most notational systems are hydrogen suppressed—*i.e.*, only nonhydrogen atoms are numbered for computer treatment. (Hydrogen atoms have arbitrarily been assigned a lower rank than all of the nonhydrogen peripheral atoms for the purpose of determining the configurations of stereogenic atoms in such situations.)

II. The Clockwise Sequence Rule. For input purposes, the clockwise sequence rule treats all of the atoms in a tetrahedral system as if they were coplanar. The peripheral atoms in the **A**, **B**, **C**, and **D** positions are then treated as if they were arranged in clockwise sequence about the central stereogenic atom (Figure 9); hence, by defining the position of one peripheral atom, the locations of all of them are fixed. The application of this rule to two widely used types of stereochemical formulas is described below.

APPLICATION TO BOLD LINE-BROKEN LINE REPRESENTATIONS. Stereochemical formulas of this type, and the stereogenic atoms depicted therein, are easily recognized by the presence of bold lines (or wedges) and broken lines. The usual conventions apply for three-dimensional representations of this type—namely, that bold lines represent bonds projecting out of the plane of the paper, and broken lines represent bonds projecting behind the plane of the paper. The clockwise sequence rule is applied to each this fashion, the assignment of stereodescriptors (Figure 10) being based on the following conventions: (1) the bold being based on the following conventions: (1) the bold line attachment is always **A**; (2) the broken line attachment is always **B**, except when a conflict arises between use of con-

Name: (+)- α -D-GlucopyranoseM.F. $C_6H_{12}O_6$

Stereo: Absolute

Sign of Rotation: (+)

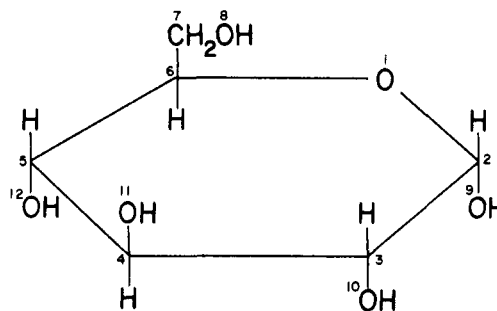


Figure 8. Connection table illustrating application of the Haworth formula rule.

Atom No.	Element	Group	Bond 1	Attachment 1	Bond 2	Attachment 2	Bond 3	Attachment 3	Bond 4	Attachment 4	Charge	Abnormal Valence	Abnormal Mass	H Count	Stereodescriptors				Atom No.
															A	B	C	D	
1	O		1	2	1	6								1	H	9	3	1	1
2	C		1	3	1	9	1	1						1	H	10	4	2	2
3	C		1	4	1	10	1	2						1	H	11	5	3	3
4	C		1	5	1	11	1	3						1	H	12	6	4	4
5	C		1	6	1	12	1	4						1	H	7	1	5	5
6	C		1	7	1	1	1	5						1					6
7	C		1	8	1	6								2					7
8	O		1	7										1					8
9	O		1	2										1					9
10	O		1	3										1					10
11	O		1	4										1					11
12	O		1	5										1					12

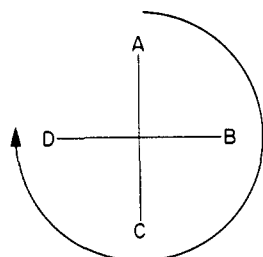


Figure 9. Reference chart for assigning stereodescriptors by the clockwise sequence rule.

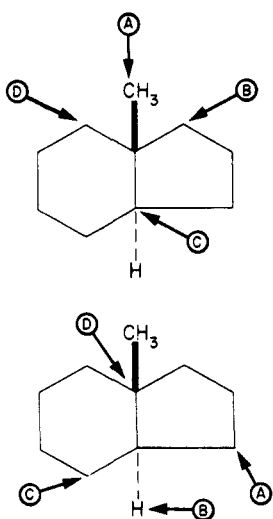


Figure 10. Assignment of stereodescriptors by the clockwise sequence rule on bold line-broken line representations.

tions 1 and 2; under such circumstances convention 1 has precedence. This decision is purely arbitrary. Its necessity is dictated by the fact that the relative order in which the bold line and broken line attachments are drawn varies with individual chemists. Application of the clockwise sequence rule for input of stereoisomerism from a typical bold line-broken line representation is illustrated in the connection table for (+)-cholesterol (Figure 11). A stipulation that the rule must only be applied on atoms which have at least three bonds shown explicitly to their peripheral atoms enables clerks to recognize that atoms such as 1, 8, 20, and 28 in Figure 11 are not stereogenic. Hydrogen atoms whose bonds to stereogenic atoms are not shown explicitly need not be recorded in the stereodescriptor columns, since their spatial arrangements are implicit in the descriptions of the remaining peripheral atoms. This treatment is analogous to that for an electron pair—*e.g.*, on stereogenic Group V and Group VI atoms in asymmetric phosphines, arsines, and sulfoxides, etc.

APPLICATION TO FISCHER PROJECTION FORMULAS. Fischer projection formulas have no special characteristics which enable a clerk (or an optical scanner) to recognize which atoms have to be described stereochemically at input. Consequently, as illustrated in Figure 12, a chemist must (1) identify each such atom (methods other than the check mark could also be used), and (2) show the direction of the A position before the assignment of stereodescriptors can be made by the clockwise sequence rule. For a formula which follows the true Fischer convention [see (3) 3rd ed., p. 251], the A position is to the right as illustrated in Figures 12 and 13. For a true Fischer formula which has been rotated 90° by the rigid-body convention [see (3) 3rd ed., pp. 251-52], the A position

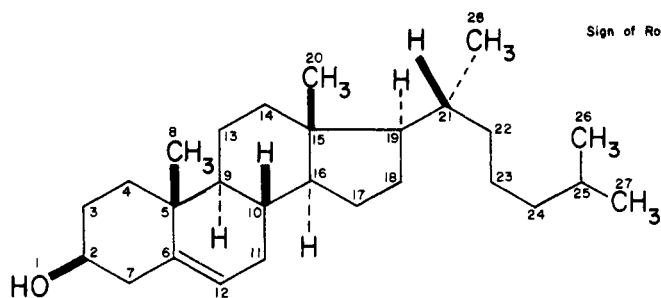
GENERATING UNIQUE COMPUTER STRUCTURAL REPRESENTATIONS OF STEREOISOMERS

Name: (+)-Cholesterol

M.F. $C_{27}H_{46}O$

Stereo: Absolute

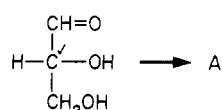
Sign of Rotation: (+)



Atom No.	Element	Group	Bond 1	Attachment 1	Bond 2	Attachment 2	Bond 3	Attachment 3	Bond 4	Attachment 4	Charge	Abnormal Valence	Abnormal Mass	H Count	Stereodescriptors				Atom No.
															A	B	C	D	
1	O		1	2										1					1
2	C		1	3	1	7	1	1						1	1	3	7		2
3	C		1	4	1	2								2					3
4	C		1	5	1	3								2					4
5	C		1	6	1	8	1	9	1	4					8	9	6	4	5
6	C		1	7	2	12	1	5											6
7	C		1	2	1	6								2					7
8	C		1	5										3					8
9	C		1	10	1	13	1	5						1	10	H	5	13	9
10	C		1	11	1	16	1	9						1	H	16	11	9	10
11	C		1	12	1	10								2					11
12	C		2	6	1	11								1					12
13	C		1	14	1	9								2					13
14	C		1	15	1	13								2					14
15	C		1	16	1	19	1	20	1	14					20	19	16	14	15
16	C		1	17	1	10	1	15						1	17	H	10	15	16
17	C		1	18	1	16								2					17
18	C		1	19	1	17								2					18
19	C		1	21	1	15	1	18						1	15	H	21	18	19
20	C		1	15										3					20
21	C		1	22	1	28	1	19						1	H	28	22	19	21
22	C		1	23	1	21								2					22
23	C		1	24	1	22								2					23
24	C		1	25	1	23								2					24
25	C		1	26	1	27	1	24						1					25
26	C		1	25										3					26
27	C		1	25										3					27
28	C		1	21										3					28

Figure 11. Connection table illustrating application of the clockwise sequence rule on a bold line-broken line representation.

Fischer formula supplied by chemist:



Assignment of stereodescriptors:

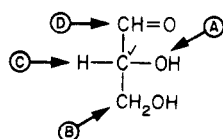


Figure 12. Assignment of stereodescriptors by the clockwise sequence rule on Fischer projection formulas.

is down. Application of the rule on a typical Fischer projection formula is illustrated in the connection table for (+)-tartaric acid (Figure 13).

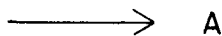
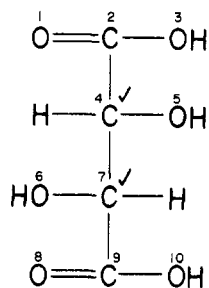
III. The Double Bond Rule. When stereoisomerism about a double bond has to be described, the doubly bound atom-pair must also be identified by a chemist (Figure 14), since structural formulas for compounds containing such double bonds are often written as though the stereochemistry is known, even though it is not. Stereodescriptors for the peripheral atoms attached to a doubly bound stereogenic atom-pair are assigned as follows (Figures 4 and 14): (1) the axis coincident with the double bond is used as an arbitrary dividing line; (2) those peripheral atoms on one side of the dividing line are assigned A descriptors, while those on the other side are assigned B descriptors. Use of the double-bond rule is illustrated in the connection table for Vitamin A (Figure 15).

Name: (+)-Tartaric Acid

 M.F. $C_4H_6O_6$

Stereo: Absolute

Sign of Rotation: (+)



Atom No.	Element	Group	Bond 1	Attachment 1	Bond 2	Attachment 2	Bond 3	Attachment 3	Bond 4	Attachment 4	Charge	Abnormal Valence	Abnormal Mass	H Count	Stereodescriptors				Atom No.
															A	B	C	D	
1	O		2	2															1
2	O		1	3	1	4	2	1											2
3	O		1	2										1					3
4	O		1	5	1	7	1	2						1	5	7	H	2	4
5	O		1	4										1					5
6	O		1	7										1					6
7	O		1	9	1	6	1	4						1	H	9	6	4	7
8	O		2	9															8
9	O		1	10	1	8	1	9											9
10	O		1	9										1					10

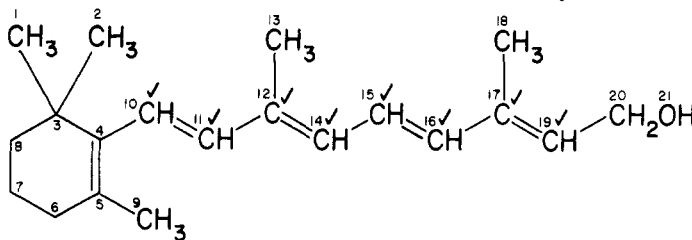
Figure 13. Connection table illustrating application of the clockwise sequence rule on a Fischer projection formula.

Name: Vitamin A

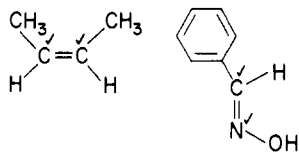
 M.F. $C_{20}H_{30}O$

Stereo: Absolute

Sign of Rotation: (0)



Formulas supplied by chemists



Assignment of stereodescriptors

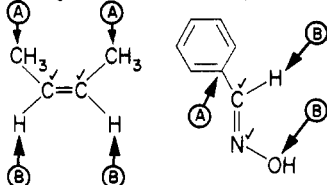


Figure 14. Assignment of stereodescriptors by the double bond rule.

Atom No.	Element	Group	Bond 1	Attachment 1	Bond 2	Attachment 2	Bond 3	Attachment 3	Bond 4	Attachment 4	Charge	Abnormal Valence	Abnormal Mass	H Count	Stereodescriptors				Atom No.
															A	B	C	D	
1	C		1	3										3					1
2	C		1	3										3					2
3	C		1	4	1	8	1	1	1	2									3
4	C		2	5	1	10	1	3											4
5	C		1	6	1	9	2	4											5
6	C		1	7	1	5								2					6
7	C		1	8	1	6								2					7
8	C		1	3	1	7								2					8
9	C		1	5										3					9
10	C		2	11	1	4								1	4				10
11	C		1	12	2	10								1		12			11
12	C		1	13	2	14	1	11						1	11	13			12
13	C		1	12										3					13
14	C		1	15	2	12								1		15			14
15	C		2	16	1	14								1	14				15
16	C		1	17	2	15								1		17			16
17	C		1	18	2	19	1	16							16	18			17
18	C		1	17										3					18
19	C		1	20	2	17								1		20			19
20	C		1	21	1	19								2					20
21	O		1	20										1					21

Figure 15. Connection table illustrating application of the double bond rule.

GENERATING UNIQUE COMPUTER STRUCTURAL REPRESENTATIONS OF STEREOISOMERS

UNIQUE COMPUTER STRUCTURAL REPRESENTATIONS OF STEREOISOMERS

The uniqueness of the computer representations which would be generated by the proposed treatment can best be seen by comparing the results for several groups of

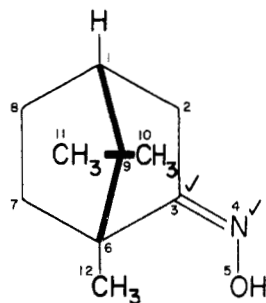
closely related stereoisomers. First, compare the treatment for a pair of diastereomers, as typified by the *syn*- and *anti*-camphor oximes whose input connection tables are illustrated in Figure 16. (These particular stereoisomers being used for illustrative purposes have not been reported in the literature.) The only difference in the structures

Name: (1R,4R)-syn-Camphor Oxime

M.F. C₁₀H₁₇NO

Stereo: Absolute

Sign of Rotation: (?)



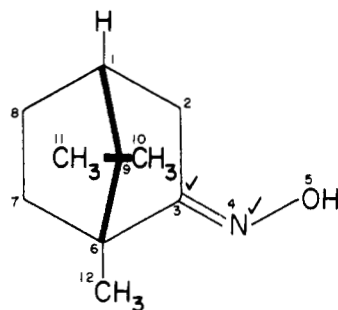
Atom No.	Element	Group	Bond 1	Attachment 1	Bond 2	Attachment 2	Bond 3	Attachment 3	Bond 4	Attachment 4	Charge	Abnormal Valence	Abnormal Mass	H Count	Stereodescriptors				Atom No.
															A	B	C	D	
1	C		1	2	1	8	1	9						2	9	8	H	2	1
2	C		1	3	1	1								2					2
3	C		2	4	1	6	1	2						2	2	6			3
4	N		1	5	2	3								1		5			4
5	O		1	4										1					5
6	C		1	7	1	9	1	12	1	3				2	9	3	12	7	6
7	C		1	8	1	6								2					7
8	C		1	1	1	7								2					8
9	C		1	10	1	11	1	1	1	6				3					9
10	C		1	9										3					10
11	C		1	9										3					11
12	C		1	6										3					12

Name: (1R,4R)-anti-Camphor Oxime

M.F. C₁₀H₁₇NO

Stereo: Absolute

Sign of Rotation: (?)



Atom No.	Element	Group	Bond 1	Attachment 1	Bond 2	Attachment 2	Bond 3	Attachment 3	Bond 4	Attachment 4	Charge	Abnormal Valence	Abnormal Mass	H Count	Stereodescriptors				Atom No.
															A	B	C	D	
1	C		1	2	1	8	1	9						2	9	8	H	2	1
2	C		1	3	1	1								2					2
3	C		2	4	1	6	1	2						2	2	6			3
4	N		1	5	2	3								1		5			4
5	O		1	4										1					5
6	C		1	7	1	9	1	12	1	3				2	9	3	12	7	6
7	C		1	8	1	6								2					7
8	C		1	1	1	7								2					8
9	C		1	10	1	11	1	1	1	6				3					9
10	C		1	9										3					10
11	C		1	9										3					11
12	C		1	6										3					12

Figure 16. Input connection tables for diastereomeric camphor oximes.

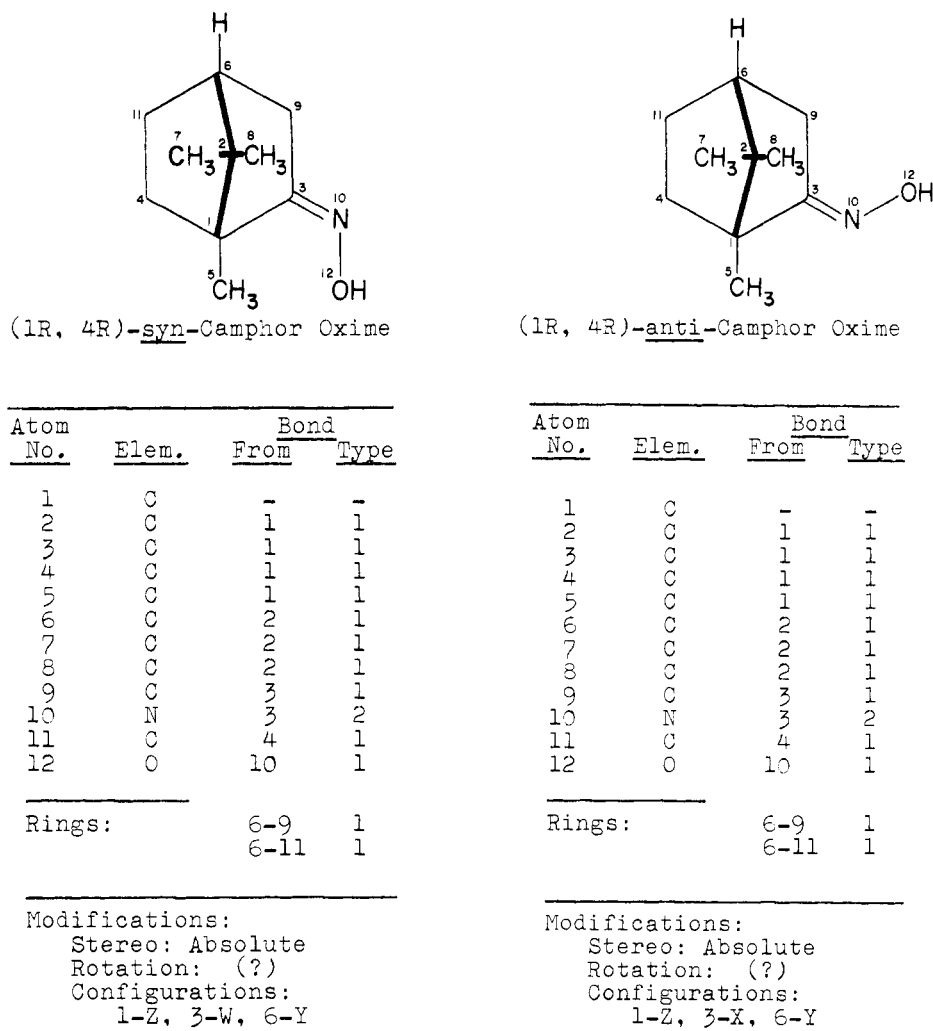


Figure 17. Unique computer structural representations for the diastereomeric camphor oximes illustrated in Figure 16.

Table I. Derivation of Configurational Descriptors From Information Supplied at Input for the Diastereomeric Camphor Oximes Illustrated in Figures 16 and 17.

Compound	Stereogenic ^a Atom No.	Peripheral Atom Numbers ^a				No. of Inversions ^b	Configuration
		A	B	C	D		
(1R, 4R) <i>syn</i> -Camphor oxime	1(6)	2(9)	3(3)	5(12)	4(7)	1	Z
	6(1)	2(9)	11(8)	H	9(2)	2	Y
	3(3)	9(2)	1(6)			2	W
	10(4)		12(5)				
(1R, 4R) <i>anti</i> -Camphor oxime	1(6)	2(9)	3(3)	5(12)	4(7)	1	Z
	6(1)	2(9)	11(8)	H	9(2)	2	Y
	3(3)	9(2)	1(6)			1	X
	10(4)	12(5)					

^aThe unparenthesized number is the unique number derived by the technique of Morgan (2c) (Figure 17), while the parenthesized number is that used at input (Figure 16). ^bBased on the unique numbering of Morgan (2c).

of these two diastereomers is in the configuration at the carbon-nitrogen double bond. Yet, this single difference in stereochemistry would be evident in the proposed computer-generated unique structural representations for these compounds (Figure 17), the double bond in one diastereomer being assigned the W configuration, and the double bond in the other, the X configuration. The form

of the computer-generated unique structural representation used in Figure 17 is that which would be generated by the technique of Morgan (2c), except that stereochemistry has been added to his two-dimensional descriptions by the techniques proposed herein. The lower numbered atom of each doubly bound atom-pair is used as the locant for the assignment of the descriptors. Deriva-

GENERATING UNIQUE COMPUTER STRUCTURAL REPRESENTATIONS OF STEREOISOMERS

tion of the configurational descriptors from the information supplied at input for these diastereomeric oximes is illustrated in Table I.

The treatment of enantiomers and racemates would be somewhat similar, as illustrated for the enantiomeric and racemic *endo*-norbornylamines in Figures 18 and 19. The input connection table for the racemate would be similar to that for either one of the enantiomers (Figure 18) except for the auxiliary information coded at input (upper right hand corner of the connection table)—that is, "relative stereo" and the sign (\pm) would replace "absolute stereo" and the particular sign of rotation at the sodium D line, unless otherwise noted. In the computer-generated unique structural representations for the two enantiomers and the racemate (Figure 19), uniqueness is attained by a combination of the computer-

generated configurational descriptors for each stereogenic atom and the auxiliary information coded at input. The configurational descriptors in the racemate are chosen on the basis of lexicographic ordering of the two possibilities coinciding with those for each of the enantiomers. The auxiliary information coded at input then functions as an operator which permits differentiation between the two levels of stereochemical information—*i.e.*, relative or absolute stereochemistry—conveyed by the same set of configurational descriptors.

GENERAL APPLICABILITY OF THE METHOD

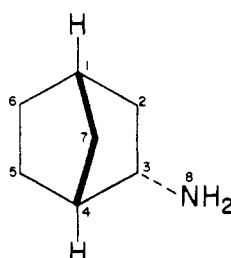
As suggested earlier, this treatment of stereoisomerism would permit generic searching by computer for substructures with specified stereochemistry. For example,

Name: (+)-*endo*-Norbornylamine

M.F. $C_7H_{13}N$

Stereo: Absolute

Sign of Rotation: (+)



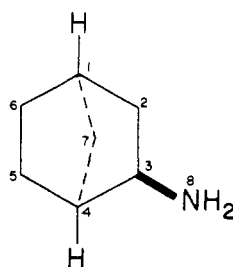
Atom No.	Element	Group	Bond 1	Attachment 1	Bond 2	Attachment 2	Bond 3	Attachment 3	Bond 4	Attachment 4	Charge	Abnormal Valence	Abnormal Mass	H Count	Stereodescriptors				Atom No.
															A	B	C	D	
1	C		1	2	1	6	1	7						1	7	6	H	2	1
2	C		1	3	1	1								2					2
3	C		1	4	1	8	1	2						1		8	4	2	3
4	C		1	5	1	7	1	3						1	7	3	H	5	4
5	C		1	6	1	4								2					5
6	C		1	1	1	5								2					6
7	C		1	1	1	4								2					7
8	N		1	3										2					8

Name: (-)-*endo*-Norbornylamine

M.F. $C_7H_{13}N$

Stereo: Absolute

Sign of Rotation: (-)



Atom No.	Element	Group	Bond 1	Attachment 1	Bond 2	Attachment 2	Bond 3	Attachment 3	Bond 4	Attachment 4	Charge	Abnormal Valence	Abnormal Mass	H Count	Stereodescriptors				Atom No.
															A	B	C	D	
1	C		1	2	1	6	1	7						1	2	7	6	H	1
2	C		1	3	1	1								2					2
3	C		1	4	1	8	1	2						1	8	4	2		3
4	C		1	5	1	7	1	3						1	5	7	3	H	4
5	C		1	6	1	4								2					5
6	C		1	1	1	5								2					6
7	C		1	1	1	4								2					7
8	N		1	3										2					8

Figure 18. Input connection tables for the enantiomeric *endo*-norbornylamines.

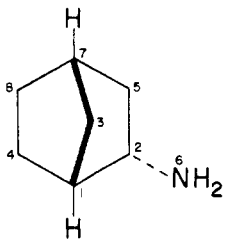
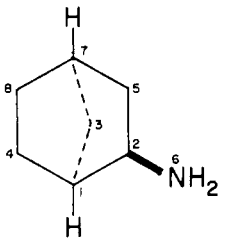
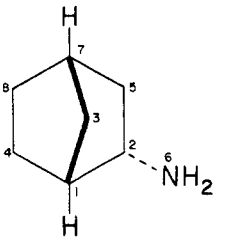
											
(+)- <u>endo</u> -Norbornylamine				(-)- <u>endo</u> -Norbornylamine				(±)- <u>endo</u> -Norbornylamine			
Atom No.	Elem.	Bond		Atom No.	Elem.	Bond		Atom No.	Elem.	Bond	
		From	Type			From	Type			From	Type
1	C	-	-	1	C	-	-	1	C	-	-
2	C	1	1	2	C	1	1	2	C	1	1
3	C	1	1	3	C	1	1	3	C	1	1
4	C	1	1	4	C	1	1	4	C	1	1
5	C	2	1	5	C	2	1	5	C	2	1
6	N	2	1	6	N	2	1	6	N	2	1
7	C	3	1	7	C	3	1	7	C	3	1
8	C	4	1	8	C	4	1	8	C	4	1
Rings:		5-7	1	Rings:		5-7	1	Rings:		5-7	1
		7-8	1			7-8	1			7-8	1
Modifications:				Modifications:				Modifications:			
Stereo: Absolute				Stereo: Absolute				Stereo: Relative			
Rotation: (+)				Rotation: (-)				Rotation: (±)			
Configurations:				Configurations:				Configurations:			
1-Y, 2-Z, 7-Y				1-Z, 2-Y, 7-Z				1-Y, 2-Z, 7-Y			

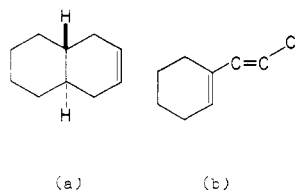
Figure 19. Unique computer structural representations for the enantiomeric and racemic *endo*-norbornylamines.

Figure 20. Examples of stereospecific substructures which could be searched generically by the proposed treatment.

if one were interested in retrieving compounds containing a *trans*- Δ^2 -octalin system (Figure 20a), (+)-cholesterol (Figure 11) would be obtained as an answer. Similarly, a search for compounds containing a 1-(*trans*-1-alkenyl)-cyclohexene system (Figure 20b) would retrieve Vitamin A (Figure 15) as an answer.

Although the proposed treatment has been illustrated only for the two most prevalent types of stereochemical systems encountered in organic compounds, the treatment for tetrahedral atoms is also applicable to systems such as the allenes and restricted biphenyls which may be treated as elongated tetrahedral systems (16). The techniques have also been partially extended to the stereochemistry of other types of systems—*e.g.*, square coplanar, octahedral, and trigonal bipyramidal systems, etc.—which are characteristic of many of the coordination compounds. Description of the treatment of some of these other systems will be reported separately.

ACKNOWLEDGMENT

Financial support by the National Science Foundation (Grant GN-382) is gratefully acknowledged.

LITERATURE CITED

- (1) a. "Survey of Chemical Notation Systems," Publication 1150, National Academy of Sciences-National Research Council, Washington, D. C., 1964.
b. Huber, M. L., *J. CHEM. DOC.* 5, 4 (1965).
c. Frear, D. E. H., "Survey of European Nonconventional Chemical Notation Systems," Publication 1278, National Academy of Sciences-National Research Council, Washington, D. C., 1965.
- (2) a. Spialter, L., *J. CHEM. DOC.* 4, 261 (1964).
b. Gluck, D. J., *ibid.* 5, 43 (1965).
c. Morgan, H. L., *ibid.* 5, 107 (1965).
- (3) These tabular descriptions are usually referred to as "connection matrixes" (2a) or "connection tables" (2c). For one of the earliest illustrations of such tabular descriptions, see Wheland, G. W., "Advanced Organic Chemistry," 2nd ed., John Wiley and Sons, Inc., New York, 1949, p. 87; 3rd ed., John Wiley and Sons, Inc., New York, 1960, p. 41.
- (4) a. Dyson, G. M., Cossum, W. E., Lynch, M. F., and Morgan, H. L., *Inform. Storage Retrieval* 1, 69 (1963).
b. Bouman, H., *J. CHEM. DOC.* 5, 14 (1965).
- (5) a. Cossum, W. E., Hardenbrook, M. E., and Wolfe, R. N., *Proc. Amer. Doc. Inst.* 1, 270 (1964).
b. Meyer, E., *Angew. Chem. Intern. Ed. Engl.* 4, 347 (1965).
- (6) a. Seifer, A. L., *Inform. Storage Retrieval* 1, 29 (1963).
b. Dyson, G. M., *ibid.* 2, 159 (1964).
c. Vander Stouw, G. G., Naznitsky, I., and Rush, J. E., *J. CHEM. DOC.* 7, 165 (1967).

- (7) a. Feldman, A., Holland, D. B., and Jacobus, D. P., *J. CHEM. DOC.* **3**, 187 (1963).
b. Mullen, J. M., *ibid.*, **7**, 88 (1967).
- (8) a. Cossum, W. E., Krakiwsky, M. L., and Lynch, M. F., *J. CHEM. DOC.* **5**, 33 (1965).
b. Sussenguth, E. H., Jr., *ibid.*, **5**, 36 (1965).
- (9) a. Jaffe, H. H., and Orchin, M., "Symmetry in Chemistry," John Wiley and Sons, Inc., New York, 1965, Chapters 1-3.
b. Mislow, K., "Introduction to Stereochemistry," W. A. Benjamin, Inc., New York, 1965, Chapter 1.
- (10) a. McDonnell, P. M., and Pasternack, R. F., *J. CHEM. DOC.* **5**, 56 (1965).
b. Pasternack, R. F., and McDonnell, P. M., *Inorg. Chem.* **4**, 600 (1965).
- (11) Cahn, R. S., Ingold, C. K., and Prelog, V., *Angew. Chem. Intern. Ed. Engl.* **5**, 385 (1966).
- (12) Hohn, F. E., "Elementary Matrix Algebra," 2nd ed., The MacMillan Co., New York, 1965, pp. 42-45.
- (13) McCasland, G. E., "A New General Method for the Naming of Stereoisomers," Chemical Abstracts Service, Columbus, Ohio, 1953, p. 2.
- (14) Noller, C. R., "Chemistry of Organic Compounds," 2nd ed., W. B. Saunders Co., Philadelphia, 1957, pp. 383-84.
- (15) Leiter, D. P., Jr., Morgan, H. L., and Stobaugh, R. E., *J. CHEM. DOC.* **5**, 238 (1965).
- (16) a. See (11) p. 388.
b. Cahn, R. S., *J. Chem. Educ.* **41**, 116 (1964).

Procedures for Converting Systematic Names of Organic Compounds into Atom-Bond Connection Tables*

G. G. VANDER STOUW, I. NAZNITSKY, and JAMES E. RUSH
Chemical Abstracts Service, The Ohio State University, Columbus, Ohio 43210

Received June 14, 1967

Simultaneously with its development of a computer-based Chemical Compound Registry System, Chemical Abstracts Service is devising procedures for automatically converting systematic names of organic compounds into atom-bond connection tables which can be manipulated by computer. A study of systematic *Chemical Abstracts* (CA) index names has resulted in a dictionary of word roots used in the names and in step-by-step procedures for converting names to connection tables. Statistical studies of nomenclature in CA indexes show that these procedures are applicable not only to current nomenclature, but also generally to names in past indexes. Procedures have been written which are applicable to the majority of names of carbon compounds, and the preparation of computer programs is now under way.

NOMENCLATURE TRANSLATION AND ITS ROLE AT CAS

In 60 years, some 3.5 to 4 million chemical compounds associated with 15 to 20 million references have been reported in the world's scientific and technical literature and have been indexed in *Chemical Abstracts* by their systematic names. Each year, nearly 10% of this number—about 350,000 compounds—are processed for CA indexes; this figure includes both compounds that appear in the literature for the first time and old compounds that are reappearing in the literature. Identification of these compounds depends on a knowledge of their molecular structure, and the ability to describe molecular structure continues to be the basic factor assuring the value of the literature to today's practicing scientist.

In order to process this large and growing collection of data, Chemical Abstracts Service (CAS) established in 1965 a Chemical Compound Registry System. This computer-based system, which depends upon molecular structure as the means of identifying a compound, includes a computer record of the structure, names, molecular for-

mula, and bibliographic citations for each compound registered. In less than two years of operation nearly 600,000 different chemical compounds have been identified and recorded in this system. Almost all of these compounds have been input *via* structural diagrams which were hand-drawn by professionals and translated into machine language through clerical effort (1).

One requirement that must be met before the full benefit of this Registry System can be derived is that of including in the record information on all chemical compounds that have been reported in the literature. Only with a complete record, for example, will it be possible to determine with certainty that a given compound is or is not new to chemistry. Thus, the tremendous body of compound-oriented information that appears in pre-1965 indexes to CA constitutes an important collection for future registration. Nevertheless, this registration task will not be feasible—in terms of time, manpower, or dollars—if it is to require the preparation of a structural diagram for each compound and the subsequent conversion of that structural diagram to machine language for registration. Fortunately, past indexes to CA have indexed compounds through a systematic nomenclature that is based directly on the diagrammatic representation of structure which

*Presented in part before the Division of Chemical Literature, Symposium on Progress Report on Modern Methods, 153rd National Meeting of the American Chemical Society, Miami Beach, Fla., April 1967.