READABLE CHEMICAL NOTATION

J. Chem. Inf. Comput. Sci., Vol. 18, No. 4, 1978 **199**

little different from the literature of many different subfields of chemistry and other disciplines. The adherence to Bradford's law, the large number of authors contributing a single paper, and the large fraction of journal articles have been observed in many previous studies of the scientific literature. The unusual features of the noble gas compound literature are its sudden start, rapid growth, and great interest to a large number of investigators. Most of these chemists remained active in this field for only a short time. Only a relatively few (5%) remained active for 10 years or more. The same observation can be made for the overall rate of growth of this literature — a sudden spurt of interest following the initial discovery, followed by a decline and then a moderate growth. This pattern is shown in Figure 12, contrasted with the more usually observed pattern.[5,10] The past 15 years have indeed been exciting ones for chemists working on noble gas compounds.

REFERENCES AND NOTES

(1) N. Bartlett, "Xenon Hexafluoroplatinate", *Proc. Chem. Soc.*, 218 (1962).
(2) D. T. Hawkins and W. E. Falconer, "Noble Gas Chemistry and the Fluoride Literature--What Influences Research Directions?", *J. Chem. Inf. Comput. Sci.*, **17**, 219-20 (1977).
(3) D. T. Hawkins, W. E. Falconer, and N. Bartlett, "Noble Gas Compounds, A Bibliography, 1962-1977", Plenum Press, New York, N.Y., 1978.
(4) The BELDEX system has been copyrighted by Bell Laboratories. For further information, see W. K. Lowry, "Use of Computers in Information Systems", *Science*, **175**, 841-6 (1972).
(5) D. J. de S. Price, "Science Since Babylon", Yale University Press, New Haven, Conn., 1961, p 116.
(6) S. C. Bradford, "Documentation", Crosby Lockwood, Ltd., London, 1948.
(7) G. Pinski, "Influence and Interrelationship of Chemical Journals", *J. Chem. Inf. Comput. Sci.*, **17**, 67-74 (1977).
(8) M. P. Carpenter and F. Narin, "Clustering of Scientific Journals", *J. Am. Soc. Inf. Sci.*, **24**, 425-36 (1973).
(9) W. Goffman, "Mathematical Model for Analyzing the Growth of a Scientific Discipline", *J. Assoc. Comput. Mach.*, **18**, 173-85 (1971).
(10) B. C. Bennion and L. A. Neuton, "Epidemiology of Research on Anomalous Water", *J. Am. Soc. Inf. Sci.*, **27**, 53-6 (1976).

A Readable Chemical Notation

W. J. MARTIN

South Dakota School of Mines and Technology, Rapid City, South Dakota

Received November 23, 1977

A notation system is described which can be easily read as systematic names resembling IUPAC nomenclature or can be converted directly to connectivity matrices.

In the following a chemical notation system with a corresponding nomenclature and connectivity matrix is described which permits direct conversion between any of the four entities—name, notation, connectivity matrix, and structural formula. The notation is readable in the sense that no matter how complex the structure, the reader may read it either verbatim as a group of symbols or as the name represented by the symbols as easily as one would with a simple inorganic salt. The nomenclature system used is similar in many ways to the familiar IUPAC nomenclature system, but it has been drastically modified to obtain a one-to-one correspondence between each symbol or group of symbols in the notation and each word of the name in order to make possible the direct reading of the notation.

BASIC SYMBOLS OF THE NOTATION

Rule 1. Numerals. As in IUPAC nomenclature, arabic numerals are used as locants. In the names these are separated by commas, but in the notations the commas are omitted to save space. If it is necessary to indicate a number greater than nine in a notation, an asterisk is used to connect the digits. This is in accordance with a general convention of the notation system: adjacent symbols are independent of each other unless they are coupled by a connectivity symbol (in this case an asterisk).

If the numbers three, five and ten appear as a set of locants, they would be written as 351* 0 in the notation, but as 3,5,10 in the name. Note that zero can never be confused with the letter O in a notation, because a zero will always be coupled with another digit by an asterisk.

Rule 2. Atomic Symbols. Atomic symbols used in the notation are the usual symbols except that double letter symbols are represented by two capital letters connected by

Table I

alkane or alkyl group	notation symbol	mnenomic aid
methane, methyl	M	
ethane, ethyl	E	
propane, propyl	T	triple
butane, butyl	B	
pentane, pentyl	V	valeric, Roman five
hexane, hexyl	X	heXane
heptane, heptyl	K	seven-day weeK
octane, octyl	A	ate, *acht*
nonane, nonyl	G	greatest diGit

a dash. However, the symbols for eight elements have been changed as follows:

boron	B-N	tungsten	W-F
bromine	R	uranium	U-R
chlorine	L	vanadium	V-N
potassium	K-A	yttrium	Y-T

The symbols for bromine and chlorine have been shortened in order to save space. The other six symbols have been changed in order to reserve the letters B, K, U, V, and Y for other uses. Adjacent atomic symbols are not bonded to each other unless they are in a group of symbols enclosed in nondirectional single quotes.

Rule 3. Carbon Chains. In addition to arabic numbers, a system of alphabetic numbers is used in the notation to indicate the lengths of carbon chains. In the corresponding names the usual names of the carbon chains are used for chains shorter than eleven. The single letter symbols used to represent these names are usually the initial letter of the name, or, if this is not feasible, a letter that is a mnenomic aid. These group names and symbols along with mnenomic explanations are shown in Table I. These letters are treated as decimal digits, and, as with arabic numbers, multi-digit numbers are created

200 *J. Chem. Inf. Comput. Sci.,* Vol. 18, No. 4, 1978

MARTIN

by connecting them with asterisks. This requires a symbol for zero which is Z. For the names of larger groups normal Greek names are abandoned, and the names are produced by using the terms deka, dekyl for 10, dideka, didekyl for 20, etc.

Note that the true Greek spelling and pronounciation are used in place of deca, decyl in order to prevent any confusion with normal nomenclature. A 15-carbon chain is therefore represented as M*V with the name dekapentane or dekapentyl.

For each symbol or group of symbols in a notation there is a corresponding word appearing in the systematic name. For symbols introduced thus far, the corresponding parts of the systematic name are obvious. For each new symbol that is introduced the corresponding word will follow in parentheses.

NONCYCLIC SATURATED HYDROCARBONS

Rule 4. Straight-Chain Hydrocarbons. A straight-chain hydrocarbon is represented by the alphabetic number corresponding to its length. Thus *n*-heptane has the notation K and the name heptane. There is no need to designate it as normal heptane, because branched heptanes are named as hexane, pentane, or butane with side chains.

Rule 5. Side Chains. For branched hydrocarbons the longest straight chain is indicated, preceded by alphabetic symbols for the side chains. If there is more than one side chain of the same length, and locants are not needed, this is indicated in the name by the usual prefixes and in the notation by the proper arabic number and a dash preceding the notation for the side chain. Therefore, ethylpentane has the notation EV. Note that because there is no connecting asterisk it cannot be confused with didekapentane which would be represented by E*V. Diethylpentane would have the notation 2-EV.

Rule 6. Locants. When it is necessary to specify the atom to which a side chain is attached, an arabic locant precedes the symbol for the side chain as with IUPAC nomenclature. In the name the usual dashes are used, but not in the notation. If there is a choice in numbering the atoms of the main chain, that producing the lowest locant for the side chain is chosen. For example, isoheptane has the notation 2MX (2-methylhexane) not 5MX.

Rule 7. Priority of Side Chains. If there is more than one type of side chain, the side chains are listed in order of increasing length, with the lowest possible locant for the longest side chain. If there are any further options, the lower locant goes to the longer side chain. Therefore,

$$(Et)_2CHCHEt$$
$$|$$
$$CH_3$$

has the notation 4M3EX (4-methyl-3-ethylhexane, and the isomers produced by changing the position of the methyl group are 2M3EX, 3M3EX, and 5M3EX.

Rule 8. Multiple Locants. When a side chain appears more than once, the locants preceding the side chain symbol are listed in decreasing order without commas; both features contrary to IUPAC nomenclature. The commas are retained in the corresponding name. If in the preceding compound the two leading ethyl groups are replaced by methyl groups, the notation becomes 32MV (3,2-dimethylpentane). This use of decreasing locants produces a continuous increase of precedence to the right. Thus 32MV is the 3-methyl derivative of 2MV (2-methylpentane).

Rule 9. Multidigit Locants. In more complex structures requiring locants greater than nine, the connecting asterisk is used only for the lowest locant in each decade, with the exception that all locants containing a zero must have asterisks. All further use of locants will be subject to the same rules. Thus for a very large molecule all locants in the notation from 100 through 110 must use two asterisks. As examples of compounds with multidigit locants, 10,10-dimethyldide-

kapentane has the notation 1*01*0ME*V while the 11,10 compound has the notation 111*OME*V.

Rule 10. Locants in Side Chains. Unlike IUPAC nomenclature the indexing of the main chain is continued on through the side chains. The side chains are numbered in the reverse order in which they appear in the notation, each chain being numbered from root to end. For example, the highly branched hydrocarbon represented below by a skeleton of

$$1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9$$

locant numbers has the notation 111*073M4E5TG (11,10,-7,3-tetramethyl-4-ethyl-5-propylnonane).

NONCYCLIC UNSATURATED HYDROCARBONS

Rule 11. Unsaturation Symbols. For unsaturated compounds the notation system uses the symbol Y (yne) for a triple bond and the symbol = (ene) for a double bond. These unsaturation symbols appear in the notation following the carbon skeleton. When both symbols are used they appear in order of unsaturation; Y precedes =. As examples, ethylene has the notation E= (ethene), and acetylene is EY (ethyne).

Rule 12. Unsaturation Locants. When locants are needed they are listed in ascending order after the unsaturation symbol, each locant indicating the *higher* indexed of the two atoms involved in the multiple bond. Thus 2-butene has the notation B=2 (butene-2), symmetrical butadiene is B=24 (butadiene-2,4), and CH_2=CHC≡CH is BY2=4 (butyn-2-ene-4). Note that this establishes a priority system similar to that for side chains, but in the opposite direction, giving a symmetrical system with priority decreasing with distance from the central symbol for the longest straight chain. As each structural feature is added to the front or back of a notation, it is denoting a derivative of the inner portion of the notation. The higher indexed atom is used as the locant for a multiple bond because in any noncyclic compound an atom is attached to only one lower indexed atom, but may be attached to several higher indexed atoms.

Rule 13. Locant Precedence. Side-chain indexing takes precedence over unsaturation indexing. Isopropylacetylene has the notation 2MBY4 (2-methylbutyne-4). In 1,1-diethylethylene the double bond is in the side chain, so its notation is 3MV=6 (3-methylpentene-6).

Rule 14. Hydrogen-Replacing Groups. Hydrogen-replacing groups are handled in the same way as side chains and are listed in front of the notation for the carbon skeleton. In indexing they yield precedence to unsaturation symbols. Unlike IUPAC nomenclature which accords special treatment to groups such alcohols, ketones, and aldehydes by listing them at the end of the name, all groups are treated alike.

As with the carbon chain, hydrogen atoms are not shown in the notation for these groups. Thus HO is represented by O (hydroxo), H_2N by N (amino), and HS by S (hydrosulfo). The notations for groups attached to the carbon chain by multiple bonds end in = or Y. O= has the notation O= (oxo), S= is S= (thio), HN= is N= (imido), and N≡ is NY (nitrilo). As some simple examples, 1-butanol is 10B (1-hydroxobutane), isopropylamine is 2NT (2-aminopropane), acetone is 2O=T (2-oxopropane), and 1-butanal is 1O=B (1-oxobutane).

Rule 15. Multifunctional Hydrogen-Replacing Groups. A carboxylic acid would call for O= and O with the same locant. To save space and allow ready recognition of this group, it is represented by the combination O,O= (hydroxoxo), the comma indicating that although the two groups are not bonded to each other they share the same locant. Analogous groups

with S or NH replacing the oxo group, or SH, NH_2, or a halogen replacing the hydroxo group are handled similarly. More complex groups based on inorganic acids will be treated separately. Thus isovaleric acid has the notation 4O,O=2MB (4-hydroxoxo-2-methylbutane), oxalic acid is 2-O,O=E (dihydroxoxoethane), and acetyl chloride is L,O=E (chloroxoethane).

Rule 16. Precedence of Hydrogen-Replacing Groups. Precedence of hydrogen-replacing groups is based primarily on the multiplicity of the bond to the carbon atom. Combined groups such as carboxylic acid take precedence before double bonds but after triple bonds. For groups with equal bond multiplicity precedence is by atomic number of each atom in the order in which it is attached to the carbon chain. As with side chains, higher precedence means later listing.

Rule 17. Triple Bonded Groups on Symmetrical Carbon Chains. Groups attached by triple bonds or combined double and single bonds can appear only on terminal carbon atoms. Therefore, when a single such group is attached to a symmetrical carbon chain, no locant is needed for it, and the atom to which it is attached becomes the number one atom. Thus propionic acid is O,O=T (hydroxoxopropane) and chloroacetic acid is 2LO,O=E (2-chlorohydroxoxoethane).

Rule 18. Substituted Methanes and Ethanes. For substituted methanes and singly substituted ethanes there is no need for locants. For polysubstituted ethanes the highest priority group identifies the number one atom; a single group needs no locant. Thus chloromethane is LM, ethanol is OE (hydroxoethane), chloroform is 3-LM (trichloromethane), and bromochloroform is 3-LRM (trichlorobromomethane). The two isomers of chlorobromoethane, 1LRE and 2LRE, need no locant for the bromine, while locants are required for the bromines in the notations for the three isomers of chlorodibromoethane, 1L11RE, 2L11RE, and 1L21RE.

Rule 19. Highly Substituted Compounds. For highly substituted compounds such as perfluoro compounds, the usual "per" prefix is retained in the name. It is represented in the notation by a period preceding the symbol for the group. Any compound in which a hydrogen-replacing group has replaced at least two more than half the hydrogen atoms of the parent hydrocarbon is treated as a per compound, and hydrogen is treated as a substituent. In the notation the per group has highest priority followed by hydrogen.

Hexafluoroethane, therefore, has the notation .FE (perfluoroethane). The two heptafluoropropanes have the notations 1H.FT (1-hydroperfluoropropane) and 2H.FT. Hexafluoropropanes would be treated as per compounds, but not the pentafluoropropanes.

Rule 20. Heteroatoms. The concept of heteroatoms, commonly limited to cyclic compounds, is extended to include noncyclic compounds. A heteroatom is defined as any atom which may be regarded as having replaced a nonterminal carbon atom. Heteroatoms are listed after the unsaturation symbols. In the assignment of locants they take precedence over hydrogen-replacing groups, but yield precedence to unsaturation symbols. Priority among heteroatoms goes to the atom with the higher atomic number, and they are listed in order of precedence, i.e., in order of decreasing atomic number. As with unsaturation symbols locants follow the symbols of the heteroatoms in increasing numerical order.

Nonmetallic heteroatoms are named as simple negative ions, i.e., with an "ide" suffix. Metal heteroatoms will be treated separately. Thus ethers and esters will be considered to be compounds with hetero oxygen atoms, and nonprimary amines will have hetero nitrogen atoms. Examples of hetero compounds that need no locants are:

dimethyl ether TO (propane oxide)
dimethylamine TN (propane nitride)

trimethylamine MTN (methylpropane nitride)
CH_2=NCH_3 T=N (propene nitride)
Some more complex compounds which need locants are:
methyl isopropyl ether 2MBO3 (2-methylbutane oxide-3)
CH_2=CHOCH$_3$ B=2O3 (butene-2 oxide-3)
$CH_3OCH_2SCH_3$ VS2O4 (pentanesulfide-2 oxide-4)

Esters and acid anhydrides are not easily recognized. In order to identify them it is necessary to note the position of oxo groups relative to the oxide linkage. Adjacent oxo and oxide groups indicate an ester, and an oxide link between two oxo groups indicates an acid anhydride. Thus ethyl propionate is 4O=XO3 (4-oxohexane oxide-3) and acetic anhydride is 42O=VO3 (4,2-dioxopentane oxide-3).

MONOCYCLIC HYDROCARBONS

Rule 21. Cyclization. The presence of a ring is indicated by the symbol / (cyclo). Cyclization results in a loss of any reference point for locants. A single structural feature therefore requires no locant. Some simple cyclic compounds which require no locants are E/V (ethylcyclopentane), L/X (chlorocyclohexane), /V= (cyclopentene), and /VN (cyclopentane nitride).

Rule 22. Priority for Assigning Locants in a Ring. When there is more than one structural feature the atom associated with the highest priority feature is the number one atom of the ring. The highest priority feature is listed without a locant unless it appears more than once in the compound. The order of priority is (a) unsaturation and heteroatoms in the same order as with noncyclic compounds, (b) side chains attached to the ring and their branches, (c) unsaturation and heteroatoms in the side chains, and (d) hydrogen-replacing groups.

Rule 23. Conjugated Double Bonds in Rings. Because of the frequent occurrence of conjugated double bond systems in cyclic compounds, they are represented by the special symbol C (conjene) which is the highest priority unsaturation symbol. The term conjene does not necessarily imply aromatic character, and does not apply to side chains. In odd-numbered rings the conjugated system cannot extend over the entire ring, but the symbol C is still used with the nonparticipating atom being designated the number one atom. Therefore, cyclopentadiene is /VC (cyclopentaconjene), its methylene derivative is 1M/VC=6 (1-methylcyclopentaconjene-ene-6), and pyrazole is /VCN12 (cyclopentaconjene dinitride-1,2).

Rule 24. The Benzene Ring. Because of the extremely large number of compounds containing benzene rings, it is represented by the special symbol Q (benzene, phenyl). For example, toluene is MQ (methylbenzene), *p*-dichlorobenzene is 41LQ (4,1-dichlorobenzene), and pyridine is QN (benzene nitride).

Rule 25. Incompletely Conjugated Rings. If there are two or more nonconjugated atoms in a ring there are two possible notations: a polyalkene or a hydrogenated conjene, the shorter notation of the two being used. If both are of equal length the alkene notation is used. For example, the two cyclohexadienes are /6=25 (cyclohexadiene-2,5), and 2-HQ (dihydrobenzene) instead of /6=24 (cyclohexadiene-2,4).

Rule 26. Precedence of the Hydro Symbol. The symbol H (hydro) takes precedence over side chains, and when four or more consecutive atoms are hydrogenated the series of locants is abbreviated to the first and last of the series separated by a - (to).

For example, the ten-atom conjugated ring /M*ZC (cyclodekaconjene) has a dihydro derivative which could have the notation 21H/M*ZC (2,1-dihydrocyclodekaconjene) or / M*Z=2468 (cyclodekatetrene-2,4,6,8), the first being chosen as being shorter. One of its tetrahydro derivatives could have the notation 4-1H/M*ZC (4-to-1-tetrahydrocyclodekaconjene)

or /M*Z=246 (cyclodekatriene-2,4,6). In this case the triene notation is shorter and is the proper notation.

NONCONDENSED POLYCYCLIC SYSTEMS

For simple noncondensed systems notations are straightforward. Biphenyl is QQ (phenylbenzene) and dicyclopentylmethane is 2-/VM. For more complex systems it is necessary to first identify the central feature of the compound which may be a ring, a pair of rings, a ring with side chain, or a straight chain. The notation of the compound is then the notation for this central feature modified by the surrounding shells, shell by shell, and completed by indicating unsaturation features, heteroatoms, side chains, and hydrogen-replacing groups.

Rule 27. The Central Feature of Noncondensed Polycyclic Systems. The systematic procedure for identifying the central feature of a complex system is as follows:

a. Eliminate all side chains and hydrogen-replacing groups to identify the skeleton of the system. It is to this skeleton that the term conjene will apply.

b. Eliminate all the peripheral rings.

c. If necessary continue removing peripheral chains and rings shell by shell until the central feature is obtained.

As an example consider the compound

$$R$$
$$/$$
$$RCH_2RCH$$
$$\backslash$$
$$CH_2R$$
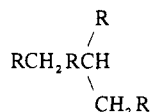
where the R's represent benzene rings. Removal of the peripheral rings gives $CH_3RCH_2CH_3$. Removal of the two side chains then leaves the benzene ring as the central feature. For the meta form the first fragment is 3MEQ (3-methylethylbenzene), and the complete compound is 987Q3MEQ (9,-8,7-triphenyl-3-methylethylbenzene).

SIMPLE CONDENSED POLYCYCLIC COMPOUNDS

Simple condensed polycyclic ring compounds are those which may be described as additional ring closures of a parent single ring.

Rule 28. The Parent Ring. Each simple polycyclic condensed ring system is described as a ring closure derivative of a parent ring that is the largest possible single ring. In numbering the atoms of the parent ring, the following conditions must be satisfied in the given order of priority.

a. The number one atom must be involved in a second ring closure.

b. The highest numbered atom, already involved in a ring closure to form the parent ring, must if possible be involved in a second ring closure.

c. If there is a choice, these two additional closures should be made with the lowest numbered atoms possible, with precedence for the number one atom.

d. Any further ring closures should involve the lowest indexed atoms possible.

e. Indexing priority for unsaturation, heteroatoms, and side chains is the same as for monocyclic compounds.

Rule 29. Single Locant Ring Closures. The second ring closure at the number one atom is indicated by placing a second / in front of the notation for the parent preceded by the locant of the other atom involved. This gives a pair of adjacent cyclo symbols read as "dicyclo".

If the system has an additional ring closure involving the last atom of the parent ring, this is indicated by the proper locant and / preceding the closure symbols for the first atom. Naphthalene is therefore 6//M*ZC (6-dicyclodekaconjene) and phenanthrene is 9/6//M*BC (9-cyclo-6-dicyclodekabutaconjene).

Rule 30. Double Locant Ring Closures. Further ring closures will require two locants in decreasing order to indicate the atoms involved. If there is more than one double locant ring closure, they are listed in decreasing order of the smaller locant of each pair. Anthracene, therefore, has the notation 1*38/6//M*BC (13,8-cyclo-6-dicyclodekabutaconjene), and naphthacene is 151*0/1*78/6/M*AC (15,10-cyclo-17,8-cyclo-6-dicyclodekaoctaconjene).

Rule 31. Double Bond Ring Closures. Any double bond appearing at a ring closure has to be identified by a pair of locants. These locant pairs are listed after the single locants. The higher locant is listed first separated from the second by a - (to). Two six-membered rings sharing a side which is a double bond would have the notation 6//M*Z=6-1 (6-dicyclodekene-6-to-1).

COMPLEX CONDENSED POLYCYCLIC COMPOUNDS

Complex polycyclic compounds are those which cannot be described as ring closures acting on a straight-chain compound. They are described as ring closures involving the end atoms of side chains on a simple condensed ring system.

Rule 32. Choice of Parent Compound. It is frequently possible to describe a complex system using several different parent compounds. To eliminate this difficulty the partent compound is selected in accordance with the following criteria in the order given: (a) maximum number of rings, (b) largest ring, (c) fewest ring closure locants, (d) lowest locants in order from right, (e) fewest side chains involved in additional ring closures, and (f) longest side chain as short and simple as possible, and the same criteria for successively shorter chains.

Rule 33. Ring Closures Involving Side Chains. The ring closures involving the ends of side chains of the parent compound are indicated by a / preceded by the locant of the second atom involved in the closure. These ring closures are listed in front of the notation of the parent compound in the same order as the side chains. If the end of a side chain is involved in a double ring closure, two locants are listed in descending order followed by two cyclization symbols.

Two cyclohexanes joined by a spiro linkage would be described as a ring closure acting on V/X (pentylcyclohexane) and would have the notation 1/V/X (1-cyclopentylcyclohexane). A system of three cyclohexanes joined by spiro linkages at the para positions of the central ring could be represented as ring-closure derivatives of either 41V/X (4,1-dipentylcyclohexane) or 9EA/X (9-ethyloctylcyclohexane). These are identical for the first five criteria given. The sixth criterion that the longest side chain should be as short and simple as possible eliminates the second compound. The complete notation is therefore 4/1/41V/X (4-cyclo-1-cyclo-4,1-dipentylcyclohexane). Spiro compounds are easily identified by the matching of a cyclization locant with its corresponding side chain locant.

SPATIAL ISOMERISM

Rule 34. Spatial Isomerism Symbols. Spatial isomerism is denoted by the use of cis, trans, recto, and sinister symbols. The cis prefix is represented by an opening parenthesis, (, and trans by a closing parenthesis,). Recto or *R* will be represented by + and sinister or *S* by -. In all cases these symbols will precede a locant for some structural feature. If the absolute configuration is unknown or of no interest, optical activity is indicated in the name by the usual prefixes, and in the notation by a capitalized prefix followed by a blank instead of a dash.

Rule 35. Orientation at a Double Bond. Geometrical isomerism due to a double bond that does not involve a terminal carbon atom is described by giving the orientation of

the two portions of the chain relative to each other. The compound
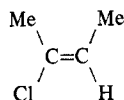
is 2LB=(3 (2-chlorobutene-*cis*-3). The chlorine atom has nothing to do with the geometrical isomerism. The compound is the 2-chloro derivative of B=(3 (butene-*cis*-3).

Rule 36. Optional Indexing of Double Bonds. If two different notations are possible for a compound depending on the end of the molecule from which indexing starts, the notation that assigns a trans configuration to the lowest locant is used. For example, 1,4-dimethylbutadiene with one cis and one trans bond has the notation X=)3(5 (hexadiene-*trans*-3-*cis*-5) instead of X=(3)5.

Rule 37. Double Bond at the Root of a Side Chain. When a double bond is at the root of a side chain, the orientation of the side chain relative to the lower indexed portion of the main chain is given. The compound
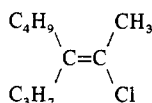
therefore has the notation 9L4EA=)9 (9-chloro-4-ethyl-octene-*trans*-9).

Rule 38. Double Bond Involving Terminal Carbon Atom. For a double bond involving the terminal atom of a chain to produce geometrical isomerism there must be a hydrogen-replacing group on the terminal atom. The orientation of this group, or the higher priority group if there are two, relative to the carbon chain is specified. For substituted ethenes with no chain for reference, the orientation of the higher priority group on the number two atom relative to the higher priority group on the number one atom is given. For example,
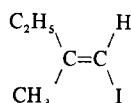
has the notation)1I2MB=2 (*trans*-1-iodo-2-methylbutene-2).

Rule 39. Geometrical Isomerism in Cyclic Compounds. If a cyclic compound is sufficiently symmetrical it is possible to have geometrical isomerism due to groups being attached to the same (cis) or opposite (trans) sides of the ring. Again the highest priority group attached to the ring serves as the reference. For example, a methyl and ethyl group attached to opposite carbon atoms of cyclohexane give the isomer pair (4ME/X and)4ME/X (*cis*- or *trans*-methylethylcyclohexane).

Rule 40. Asymmetric Atoms. The absolute configuration of an asymmetric atom is denoted by + or − signs. In the corresponding names these are written as *R* and *S* and read as recto and sinister. The configuration of a tetrahedrally bonded asymmetric atom in a chain is described by using the bond to the next lower indexed atom as a rotation axis pointing away from the observer, the remaining three bonds extending obliquely toward the observer. The bond to the next higher indexed atom is used as a rotational reference indicating the six o'clock position. The remaining bonds are at the ten o'clock and two o'clock positions, the first representing a negative or left-hand rotation from twelve, and the second a positive or right-hand rotation. The orientation of only the higher priority group is given in the notation as a + or − sign in front of its locant.

Dihydroxysuccinaldehyde is a simple example of optical isomerism having one meso form and a pair of enantiomorphs. The two enantiomorphs are −3+2O41O=B ((*S*)-3-(*R*)-2-

dihydroxo-4,1-dioxobutane), but the meso form has two possible notations, +3+2O41O=B or −3−2O41O=B depending on the choice of the number one atom.

Rule 41. Optional Indexing of Optical Isomers. The problem of a compound having more than one possible notation is eliminated by adopting the following conventions.

a. For meso forms with optional indexing the configuration of the lowest indexed group is always +.

b. For enantiomorphs with optional indexing the proper notation is the one that continues the configuration of the lowest indexed asymmetric atom as long as possible, and if this is not decisive, the same criterion is applied to each following configuration.

The proper choice of notation for the previous example is, of course, the one beginning with the +. A more complex case would be one of the pairs of enantiomorphs of penta-hydroxypimaldehyde. They may be represented as +6−5−4+3−2O71O=K and −6+5+4−3+2O71O=K, or by reversing the indexing as +6−5+4+3−2O71O=K and −6+5−4−3+2O71O=K. In both pairs the configuration of the third atom differs from the second. In the first pair the configuration of the fourth atom again differs from the third, while in the second pair the third and fourth atoms have the same configuration, making this pair the accept notations.

Rule 42. Asymmetric Terminal Atom. For an asymmetric atom at the end of a chain the highest priority group attached to it is used to replace whatever main chain segment is missing. For optically active substituted methanes the highest priority group serves as the lower indexed chain; the next lower priority group serves as the higher indexed chain, and the configuration of the third highest priority group is given. In such cases since no locants are used, the + or − directly precedes the symbol of the group.

Rule 43. Asymmetric Atoms in Rings. For cyclic compounds the ring closing bond is considered to be an extension of the chain in either direction. For example, for a compound with a cyclohexane ring having a methyl group and an ethyl group on adjacent atoms, the ethyl group is on the number one atom. Therefore, the bond to the sixth atom serves as the missing lower indexed chain, giving two sets of enantiomorphs; + 2M+E/X, −2M−E/X and −2M+E/X, +2M−E/X.

OTHER STRUCTURAL FEATURES

Rule 44. Metal Heteroatoms. If a metal atom in an organic compound is involved in more than one metal–carbon bond it is treated as a heteroatom, but, unlike nonmetals, the ide suffix is not used in the name. The presence of square or octahedral bonds for some metals introduces further complications in the stereochemistry of these compounds, which can, however, be managed by use of the tetrahedral symbols. Treating the lead in tetraethyllead as a heteroatom gives the notation 2-EVP-B3 (diethylpentanelead-3).

Rule 45. Metals in Hydrogen-Replacing Groups. The best known and most important examples of metals replacing hydrogen atoms are the various Grignard reagents. In the notations for these compounds the metal and the halogen attached to it are grouped together by nondirectional single quotes to indicate that they are an integral group. As an example, the iodo Grignard reagent for the isopropyl group is 2'IM-G'T (2-iodomagnesium–propane).

Rule 46. Notations for Inorganic Compounds. There is little incentive for devising a structural notation for inorganic compounds, but in a mixed list it would be desirable to have a compatible notation for both types of compounds. The notation used is simply the normal empirical formula with the following changes.

a. All symbols and numbers are in accord with the usage for organic compounds.

b. Parentheses are replaced by nondirectional single quotes.

c. Subscripts are changed to on-line numbers.

In the context of this notation true carbides, carbonates, and carbon oxides are considered to be inorganic compounds.

Rule 47. Inorganic Hydrogen-Replacing Groups. When complex inorganic groups appear as hydrogen-replacing groups in organic compounds, it is necessary to show the internal structure of the groups. The last symbol of the group is the bond attaching it to the carbon skeleton. These are the usual symbols Y, =, and no entry for a single bond. The next to the last symbol indicates the atom involved in this bond, and the remaining symbols are then listed in order of attachment from right to left. The entire group is set off by nondirectional single quotes, and any side chains within the group are set off in a similar fashion. A common feature in such groups is an oxygen atom attached to the group by a double bond or by a coordinate covalent bond. Such an atom is indicated by the symbol U (for unsaturated) appearing in front of the atom to which it is attached. Two such groups attached to the same atom are indicated by the symbol W (for double U). If instead of oxygen, the atom is sulfur or some other element, the U is preceded by the symbol of the element and the group is set off by quotes to indicate a side chain.

Some common inorganic hydrogen-replacing groups are "UN" (nitroso), "WN" (nitro), "UNO" (nitrito), "HOWSO" (sulfuric acid), and "HOWS" (sulfonic acid).

Rule 48. Inorganic Hetero Groups. Some nonmetal heteroatoms, particularly sulfur and phosphorus, can form additional coordinate covalent bonds. In such cases the entire group enclosed in quotes is used as a hetero symbol. The sulfoxide group is therefore "SU", and the sulfone group is "SW". Dimethyl sulfoxide has the notation T'SU' (propane sulfoxide).

Rule 49. Organic Salts. If one of the ions of an organic salt is inorganic, it is listed as in an inorganic salt. In general positive organic ions are formed by gaining hydrogen ions and negative ions are formed by losing them. A positive ion is indicated by placing an H after the symbol of the group gaining the hydrogen, and a negative ion is indicated by placing an H before the symbol of the group losing the hydrogen.

The name of the positive ion is obtained by adding the suffix onium to the name of the group involved. "Amine" becomes "ammonium" and "nitride" becomes "nitronium". The name of the negative ion is obtained by deleting "hydroxo" from the name of the group involved. "Hydroxoxo" becomes "oxoxo".

The notations of the two ions are separated by a nondirectional quote. If subscripts are needed, they appear as arabic numerals set off from the notations of the ions by as many single quotes as are needed.

Thus lead acetate is P-B'HO,O=E'2 (lead oxoxoethane), sodium ethoxide is N-A'HOE (sodium oxoethane), and tetramethylammonium chloride is 2-MTNH'L (dimethylpropanenitronium chloride).

CONNECTIVITY MATRICES

A notation or its name furnishes explicity all the information needed to construct a connectivity matrix for the compound. Starting with the central feature of the notation and moving to the left, each segment of the notation is entered along the matrix diagonal from left to right. For each segment the index number of its first atom and its locant are the coordinates of the matrix element which is to be used to indicate its attachment to the portion already entered.

At the end of the notation the unsaturation symbols and locants show how and where the matrix is to be modified to show unsaturation, and the heteroatom data indicate which carbon atoms along the diagonal are to be replaced with heteroatoms.

Using the convention that the digit 1 indicates a single bond; 2, a double bond; and 0, no bond, the matrix for 1N2MV=5O3 (1-amino-2-methylpentene-5 oxide-3) is as shown in (I).

	1	2	3	4	5	6	7	
1	C	1	0	0	0	0	1	
2	1	C	1	0	0	1	0	
3	0	1	O	1	0	0	0	
4	0	0	1	C	2	0	0	(I)
5	0	0	0	2	C	0	0	
6	0	1	0	0	0	C	0	
7	1	0	0	0	0	0	N	

A Fortran IV Program for Finding the Smallest Set of Smallest Rings of a Graph

BURGHARDT SCHMIDT and JÖRG FLEISCHHAUER*

Department of Theoretical Chemistry, RWTH Aachen, Templergraben, D-5100 Aachen, West Germany

A Fortran IV program to find the set of smallest rings is described. The algorithm consists of two parts. First the nodes, which are on rings, are found; then in the second part the smallest set of smallest rings (SSSR) is found considering these nodes only.

INTRODUCTION

To use the computer for solving chemical problems, particularly for the design of organic synthesis, analysis, and the storage of chemical information, it is necessary to translate the chemist's formulas of molecules into a form which can be manipulated by the computer. To handle molecules in a computer, graphs and their adjacency matrices[1] are very well suited. (Other well-known representations of chemical formulas in a computer program are connection-tables or the WLN notation.[2]) The nodes of the graphs represent the atoms and the edges the bonds. By definition the entries $a_{ij} = a_{ji}$ of the adjacency matrix **A** are only unequal to zero if the nodes *I* and *J* are connected by an edge.[1] Among other things a

computer program must now recognize special structural features of the graphs. Very important is the finding of cyclic structures, because they differ strongly from acyclic ones with regard to their chemical and physical behavior. Especially for chemical synthesis work the so-called smallest set of smallest rings (SSSR) of a chemical structure must be recognized. This was pointed out by Corey and Wipke[3] in their fundamental work about computer-assisted design of organic synthesis. The SSSR is the set of those rings which are not the envelopes of smaller rings. This is equivalent to Corey's and Petersson's[4] definition of the synthetically important ring.

In this paper a program is described for finding the indices of the nodes which form the set of the smallest rings of the