(3) E. Meyer, statement provided for the National Science Foundation Report, "Current Research and Development in Scientific Documentation," No. 9, 1961.
(4) M. Ye. Pantyukhina, "Machine Input and Output of Information on the Structure of Chemical Compounds," in Foreign Developments in Machine Translation and Information Processing, Office of Technical Services, No. 31, U. S. Dept. of Commerce, Washington, D. C., June 22, 1961, pp. 77–91 (JPRS 8479).
(5) *Chem. Eng. News*, **30**, 2622 (1952). Another keyboard, described by F. Murphy in T. E. R. Singer's "Information and Communication in Industry," Reinhold Publishing Co., New York, N. Y., 1958, pp. 293–296, yields esthetically less pleasing chemical structures.
(6) Personal communication.
(7) The invention is therefore credited to A. F.
(8) C. N. Mooers, "Ciphering Structural Formulas—The Zatopleg System," Zator Technical Bulletin No. 59, The Zator Co., Boston, Mass., July, 1950.
(9) M. Gordon, C. E. Kendall, and W. H. T. Davison, "Chemical Ciphering: A Universal Code as an Aid to Chemical Systematics," Royal Inst. of Chemistry of Great Britain and Ireland, 1948.
(10) W. H. T. Davison and M. Gordon, *Am. Doc.*, 8, 202 (1957).
(11) Dr. Gordon, in a private communication, dated April 2, 1962, made the following comment: "I might briefly mention one point about the Gordon, Kendall, Davison System concerning its topological basis, especially as authors of other systems have made considerable play concerning the topology of their systems. Our method of tracing a path through a structure is essentially that given by Wiener, who first considered the old problem of labyrinths as a mathematical problem in Mathematische Annalen, **6**, 29–30 (1873), as quoted in the famous book on combinatorial topology by D. Köning, "Theorie der Endlichen und Unendlichen Graphen," Akademische Verlagsgesellschaft M. B. H., Leipzig, 1936, p. 36. The reason I did not quote this in the original papers is that I innocently re-discovered Wiener's solution. There are very minor logical differences between our treatment and Wiener's solution, but essentially they are identical. I doubt if any real improvement on this problem has been made since 1837."
(12) L. C. Ray and R. A. Kirsch, *Science*, **126**, 814 (1957).
(13) This circuitry will be described in another publication.
(14) E. Meyer and K. Wenke, *Nachr. Dokument.*, **13**, 13 (1962).
(15) A. Opler, "A Topological Application of Computing Machines," AIEE Special Publication T-85, Proceedings of the Western Joint Computer Conference, February, 1956; *cf.* also: A. Opler and T. R. Norton, "A Manual for Programming Computers for Use with a Mechanized System for Searching Organic Compounds," The Dow Chemical Co., Midland, Mich., April 25, 1956.
(16) G. E. Vleduts and E. D. Stotskiy, "On Certain Systems for Recording Structural Formulas in Organic Chemistry," ref. 4, No. 30, June 2, 1961, pp. 36–48 (JPRS 8372).
(17) V. B. Borshchev, G. E. Vleduts, and V. K. Finn, "An Algorithm for Translating Structural Formulas in Organic Chemistry into Canonical Notation," ref. 4, No. 58, December 12, 1961, pp. 63–125 (JPRS 11483).

A Linear Notation for Organic Compounds

By JOHN A. SILK
Imperial Chemical Industries Limited, Jealott's Hill Research Station, Bracknell, Berkshire, England
Received May 9, 1963

While the notation outlined in this paper in some respects resembles the INPAC Notation,[1] its distinctive features justify its consideration as an independent and alternative system. Apart from semantic differences, it may be noted here that the notation uses only one alphabet and one set of numerals, so that the standard teleprinter keyboard and computer print-out system provide an adequate range of symbols. Since errors in ciphers are less obvious than those in names, the absence of subscript or superscript numerals or lower-case letters is also an important simplification in copying and checking ciphers. It will be appreciated, however, that additional symbols, such as subscript numbers for use as multipliers, could be introduced, if desired. It may also be pointed out that possibly the most attractive and legible ciphers are those with lower-case letters and small (but not subscript) numerals. Compare B4M(A)2.2F.AOQ—2A3N and b4m(a)2.2f.aoq—2a3n.

THE BASIC SYMBOLS

The first four letters of the alphabet are used to cipher the fundamental chains and ring systems of organic molecules, and for this reason they are referred to as *basic symbols*.

A denotes an Aliphatic (Alkane) carbon atom, together with the appropriate number of hydrogens.

B denotes the simple (unfused) Benzene ring. It is a specific symbol, and no other ring fused, hydrogenated, or heterocyclic, is ciphered with B.

C is the general symbol for a Cyclic structure. The number following it gives the number of rings in the structure. Ring systems are understood to contain the maximum number of noncumulative double bonds and rings to be six-membered, in the absence of other information.

D is used to cipher individual heterocyclic rings in a ring system. The number following it shows the size of ring, and it is followed by the symbols for each of the hetero atoms.

Thus, A*n* denotes a chain of *n* carbon atoms, C*n* a system of *n* fused rings, and D*n* an *n*-membered heterocyclic ring. A, C, D, and also H, are the main letter symbols for which numbers immediately following serve as multipliers. Multiplication of most other single letter symbols is shown by repeating the symbol.

The symbol A has been chosen for an aliphatic carbon atom because carbon is the prime element in organic compounds, and in alphabetical arrangements of ciphers it is

advantageous for it to come first. A benzene ring, denoted by B, is the most common cyclic structure, and the choice of C for other ring systems was determined by the advantage for indexing and classification of a letter near the beginning of the alphabet for these basic structures. The use of D for hetero rings completes the basic sequence.

A double bond is denoted by E, and a triple bond by EE. The symbol H is used in those situations where hydrogen atoms require to be specified, namely, on ring systems and with elements other than oxygen, nitrogen, and sulfur.

CARBON CHAINS

The ciphering of branched carbon chains exemplifies some principle features of cipher construction (Fig. 1).

(1) The locants for substituents precede the symbols to which they refer, and multipliers follow the symbols.

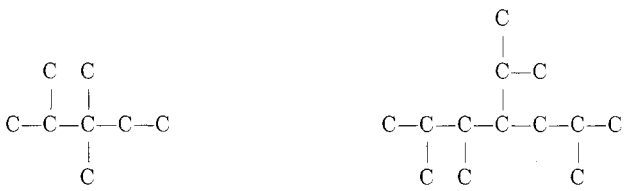
(2) In a series of locants no punctuation is used between numbers up to nine, but *every* two-digit number, including a single one, is preceded by a comma. This is simpler than the alternative of underlining two-digit numbers.

(3) The locant 1 is omitted when citing a single substituent at this position.

(4) A cipher commences with a basic symbol (A, B, C, of D), and substituents follow in order of diminishing seniority.

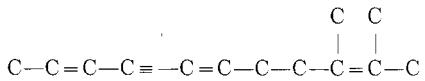
(5) Enumeration is determined by considering the attached substituents, one at a time, in order of seniority, and assigning each the lowest available locant until the enumeration has been uniquely determined. For identical items on a chain the set of locants giving the lowest sequence of numbers is preferred.

(6) Stops are used to separate parts of the cipher and break it up into more readily recognizable portions. No stop is required, however, for the first substituent cited after A, A*n*, or B. (This omission of the first stop allows simple branched side chains, such as *t*-butyl, to be ciphered without using parentheses.)

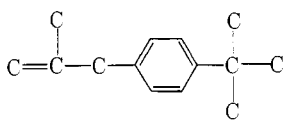


A5233A

A74A2A·236A



A, 12:4EE·26, 10E·, 10, 11A

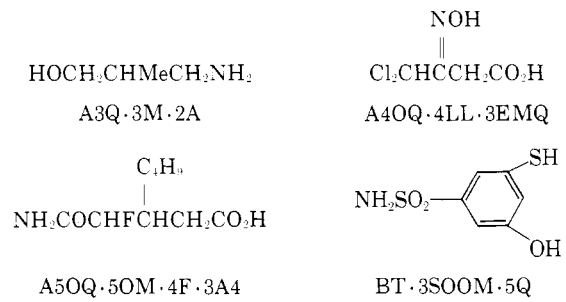


B(A32E·2A)·4A211A

Fig. 1.—Hydrocarbon skeletons exemplifying cipher construction.

(7) After the two-digit number for a chain of ten or more carbons, a colon is inserted (simply for clarity) between this number and a following locant.

Subsequent examples (Fig. 2) involving functional substituents and chains on rings show that in a branched carbon chain the choice of parent chain is determined primarily by the attached functional groups and ring systems; the alkyl groups which are attached to it are considered last of all, because they are commonly the least significant features. Partly for this reason, but mainly in order to improve the classification of compounds in cipher indexes, carbon chains are divided into two classes, monosubstituted and other. Alkyl, alkenyl, and alkanoyl (acyl) chains, both normal and branched, provide the most important examples of monosubstituted chains. By definition, they are chains which are substituted on only *one* carbon atom by a function, a ring, or a senior carbon chain. Monosubstituted chains are, therefore, mainly those types of carbon which are normally regarded as being simple substituents, and, by separating them in this way from other carbon chains, we ensure that they are treated more simply in systematic ciphering.



A3Q·3M·2A

A4OQ·4LL·3EMQ

A5OQ·5OM·4F·3A4

BT·3SOOM·5Q

Fig. 2.—Ciphering of functional and simple substituents.

Enumeration of a monosubstituted chain begins at the atom bearing the substituent function or ring. When this is not an end of the chain, enumeration continues along the larger arm of it, and the smaller arm becomes a side chain attached at the 1-position. Enumeration of all other carbon chains, *i.e.*, those with two or more substituents (disregarding alkyl side chains), proceeds along the full length of the chain, and it starts from the end which gives the lowest locant(s) to the senior substituent(s).

FUNCTIONAL GROUPS

So great a number and variety of functional groups are formed by oxygen, nitrogen, and sulfur, sometimes also in combination with the halogens, that the introduction of three additional symbols, Q, M, and T, to represent oxygen, nitrogen, and sulfur in certain situations enables valuable improvements to be made in the systematic ciphering of functional groups. In addition, new single-letter symbols are used for chlorine and bromine, so that each of the halogens has its own *distinctive* symbol. The symbols are as follows.

Q always denotes –O–, *i.e.*, divalent oxygen linking two other atoms, one of which may be hydrogen. This hydrogen is implicit, so that AQ denotes methanol.

O denotes =O, or —O, *i.e.*, divalent oxygen joined to only one other atom, so that formaldehyde is ciphered AO, *or* it denotes oxygen in a higher valency state, *e.g.*, oxonium oxygen is ciphered O+.

T always denotes -S-, *i.e.*, divalent sulfur joined to two other atoms. This usage is exactly analogous to that of Q, so that methane thiol is ciphered AT.

S denotes =S, or —S, *i.e.*, divalent sulfur joined to only one other atom, *or* it denotes sulfur in a higher valency state.

M denotes amino-type nitrogen, and is used to cipher nitrogen in all groups related to NH_3, NH_2OH, NH_2SH, and NH_2NH_2 (but never —N=N—). This includes, for example, amines, imines, amides, amidines, guanidines, hydrazines, hydrazides, semicarbazones, and oximes.

N represents nitrogen in all other circumstances.

F denotes a fluorine atom.

J denotes an iodine atom.

K denotes a bromine atom.

L denotes a chlorine atom.

P denotes a phosphorus atom.

An atom which is joined to only one other atom is referred to as a *terminal* atom. Examples are =O, =S, ≡N, and the halogens in their monovalent states. The complement of terminal is *linking*, like oxygen in alcohols and ethers, sulfur in thiols and sulfides, and nitrogen in amines.

Multiplication of single atoms in functions is shown by repeating the symbol, instead of by using a multiplier. A peroxide is written QQ, a disulfide is TT, a hydrazine is MM, and azo is NN. Similarly, the nitro group is written NOO, not NO_2; because O is terminal, this sequence cannot mean N—O—O. Similarly, a sulfone is SOO, and a sulfonamide is SOOM. The same applies also for two or more halogen atoms on the *same* carbon atom. Methylene chloride is ALL, and benzotrichloride is BALLL. Parentheses are, however, used to show multiplication of di- and polyatomic units; *e.g.*, nitroform is A(NOO)3.

Because the symbol O denotes terminal oxygen and Q denotes linking oxygen, the important carboxyl group is systematically and concisely denoted by AOQ, which is, of course, comparable to the familiar COOH, where the distinction between terminal and linking oxygen is merely implied. Further, since this expression contains two different symbols, O and Q, the inversion of OQ to QO enables a clear distinction to be made between the isomeric esters R·CO·OR′ and RO·CO·R′.

All ·CO·X and ·CS·X groups can be ciphered in the same way as the carboxyl group.

Acid chloride	Amide	Hydrazide	Thionacid
CO·Cl	CO·NH_2	CO·NH·NH_2	CS·OH
AOL	AOM	AOMM	ASQ

For other groups containing nitrogen the situation is slightly more complex, because there are three trivalent states to distinguish —N<, =N—, and ≡N. The first is always ciphered with M (amino-type nitrogen) and the third, which is terminal, with N. In the intermediate state, =N—, it is ruled that when the double bond is to a carbon atom EM is used, and in other circumstances N is used (apart from two logical exceptions given below).

Imine	Oxime	Azine
(C)=NH	(C)=NOH	(C)=N—N=(C)
EM	EMQ	EMME

Groups of the type $-C(=X)-N-$ are ciphered with the aid of a colon (or semicolon, if preferred) as a terminating symbol for EM.

Iminoacid	Amidoxime
C(:NH)OH	C(:NOH)NH_2
AEM:Q	AEMQ:M

Similar ciphers can be constructed for $X-C(=Y)-Z$ groups, such as carbonates, ureas, thioureas, and guanidines. These frequently form the link between two other carbon systems, and, for reasons given later, the =Y part, as well as the —X part, always precedes A in the ciphers.

Carbamic acid	Urea	Thiosemicarbazide
NH_2COOH	NH_2CONH_2	$NH_2CSNHNH_2$
QOAM	MOAM	MSAMM

Dithiocarbamic acid	Guanidine	Biuret
SH·$CSNH_2$	NH_2·C(=NH)·NH_2	$NH_2CONHCONH_2$
TSAM	M:MEAM	MOAMOAM

Other acidic groups and their derivatives, such as sulfonic, sulfinic, phosphonic, and other acids, as well as their derivatives, like amides, are analogous to carboxylic acids in which the central atom is S, P, etc., instead of C. The same considerations apply, therefore, to their ciphering, and EM, not N, is used for imino nitrogen in them. Inorganic acids and their derivatives are treated similarly.

Sulfinic acid	Sulfonyl chloride	Sulfamic acid
SO·OH	SO_2Cl	OH·SO_2·NH_2
SOQ	SOOL	QSOOM

Phosphoric acid	Phosphoramidothiolic acid
OH·$PO(OH)_2$	SH·PO(OH)NH_2
QPOQ:Q	TPOQ:M

Finally, the ciphering of "cyan" and "azo" types may be illustrated.

Benzonitrile	Phenyl thiocyanate	Phenyl isocyanate
PhCN	PhSCN	PhNCO
BAN	BTAN	BMEAO

Benzene-diazonium	Diazo-methane	Phenyl azide
PhN_2^-	CH_2N_2	PhN_3
BNN+	AENN	BNNN

(The diazo group is the only C=N group which is not ciphered with EM, and this exception is logical because the group is not formally derived from amino-type nitrogen.)

Seniority of Functions.—For systematic purposes it is advantageous to classify groups into four ranks. These are based on the number of valencies of the carbon atom which form bonds with noncarbon atoms (ignoring hydrogen), and the greater this number the greater is the rank and seniority.

First	Second	Third	Fourth
—C—X	> C=X	—C—Y ‖ X	X—C—Z ‖ Y
	> C < X Y	—CX_3	CX_4
		—C≡X	etc.

A third-rank function is senior to all functions of first or second rank, regardless of the elements present. Among functions of equal rank, the one coming latest when they are listed in alphabetical order is senior (Fig. 3).

$EtOCH_2CH_2NMe_2$	$Me_2C{:}NNHCOCH_2CO_2Et$
A2QA2·2M(A)2	A3OQA2·3OMMEA2A
	$(C_4H_9)_3N^+MeI^-$
BQOA2·4M(OA2)A3	(A4)3M+A.,J–

Fig. 3.—Simple derivatives of functions.

Among functions of fourth-rank structure, not all are usefully regarded as having this high seniority; fourth rank is, therefore, limited to those with two substitutable atoms which enable the function to form a link between two other carbon structures. Examples are carbonate, carbamate, carbazide, guanidine, and biuret, as well as orthocarbonate and carbodiimide. Monovalent terminal groups, like -SCN, -NC, -OCOCl, and $-SCCl_3$, have only first rank.

Groups in which the central atom is not carbon, but another element, are all regarded as first-rank functions, *e.g.*, sulfonate, phosphate, sulfenamide, and arsinite.

SIMPLE DERIVATIVES

Simple derivatives of functions (Fig. 3) are ciphered by following the symbols for the function with those for the substituent; a stop at the end of this sequence then provides the reference back to the parent chain or ring for subsequent operations without the need for parentheses to enclose the cipher for the substituent.

This method is limited specifically to certain types of monosubstituted chains, namely, (i) alkyl groups, (ii) alkylidene groups, (iii) alkenyl and alkynyl groups with only one multiple bond, and (iv) alkanoyl and alkanthioyl groups. This definition covers all the common terminal aliphatic substituents, as well as many others. It allows simple functional derivatives to be ciphered in a similar manner to the parent, unsubstituted compound, and thus to be brought together with the parent in a cipher index.

Parentheses are used conventionally to show duplication, etc., of a substituent on a nitrogen or other polyvalent atom. An ordinary, not a subscript, numeral following the closing parenthesis signifies multiplication, and when different substituents occur on a polyvalent atom the junior one(s) goes in parentheses. When parentheses are used to shorten the cipher for a structure of the type $(R)_nX$, the opening parenthesis is omitted, *i.e.*, $R)_nX$ is used.

Quaternary nitrogen is denoted by M+ (acyclic) or N+ (cyclic), the sign being written on the line. The cipher for a salt is continuous between the acidic and basic parts when salt formation has eliminated water, *e.g.*, sodium acetate, but when only proton transfer has taken place, *e.g.*, amine salts, each compound is ciphered separately with two stops .. linking the ciphers; molecular compounds are treated similarly. Metal chelates are ciphered as salts, since this is simpler than treating them as cyclic systems and is adequate for identification purposes.

RING SYSTEMS

The three ring symbols B, C, and D have been defined. A specific symbol for the single benzene ring, which occurs more frequently than all other ring systems combined, is used not only for brevity, but also in order to improve the differentiating power of the ciphers. All other six-membered ring systems are ciphered with the aid of C, the general ring symbol.

Because six-membered rings are the most common in organic molecules, all rings are understood to be of this size in the absence of a specific indication of ring size, and rings are also assumed to be "aromatic," *i.e.*, to contain the maximum number of noncumulative double bonds, unless additional hydrogen atoms are ciphered.

Complete hydrogenation is shown by H following Cn(...), the expression giving the ring structure. Intermediate states are ciphered from this by removal of pairs of hydrogens with E or from the aromatic state by addition of hydrogens, according to which method is more concise when the following devices are taken into account.

(i) When hydrogen atoms need to be added at positions bearing other substituents, their citation may be combined with that of the substituent. This is specially useful with =O, =S, etc., on a hetero ring, where the addition is merely a formal one.

(ii) Addition of hydrogen to *n* atoms numbered consecutively is shown by using Hn with the locant of the first-numbered position.

(iii) En may be used similarly to substract hydrogen and form a complete ring or ring adduct with *n* conjugated double bonds.

Five of the most common fused-ring systems are ciphered and enumerated as shown in Fig. 4, which shows that naphthalene, phenanthrene, indene, and fluorene (and, of course, many heterocyclic analogs of them) are unchanged in enumeration of their unshared atoms, and anthracene differs only in the 10-position being renumbered 12. These are the systematic enumerations of this notation, and they are based on previously described

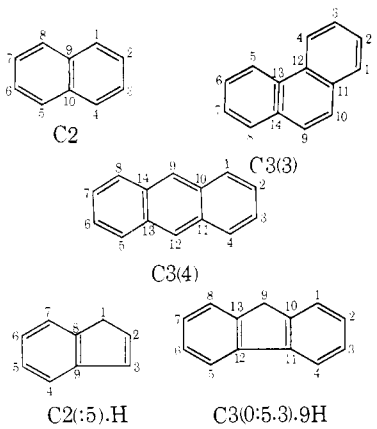


Fig. 4.—Common fused-ring systems.

developments[2] of the Taylor–Patterson–Dyson methods. The principle features of this modified system are as follows.

(1) Enumeration starts in the most central ring of a system, regardless of its size.

(2) It proceeds throughout in one direction (clockwise or anticlockwise).

(3) It proceeds in the manner which gives the lowest *total* of fusion-point locants, not the lowest sequence.

(4) *Peri*-type, or reticular, fusion is recognized as a distinct and characteristic class. With the aid of the Y-atom concept these systems can be described with great power and simplicity, and molecular formulas can be computed directly from the ciphers.

(5) The enumeration derived to describe the structure of the ring system is reversed end-to-end for describing the positions of hetero atoms and substituents.

A consequence of the last rule has already been mentioned, namely, that some of the common, and statistically very important, systems with two or three rings retain their customary enumerations. A general consequence is that in all systems the lower-numbered positions are assigned predominantly to the unshared atoms in the end rings of a system. The steroid skeleton has, however, been made a complete exception, since any change from established practice would be an unjustifiable complication; the special cipher C4(ST) is used.

In the ciphers for ring systems the fusion-point locants are cited in sequence within parentheses following C*n*, the expression showing the total number of rings. Since the first such locant must always be 1, this can be omitted, and subsequent ones are cited with a stop between each. Where necessary, the symbols Y (reticular fusion), X (spiro), and U*n* (bridge with *n* atoms) are inserted after the appropriate locants, and so are the symbols :n, to denote an *n*-membered ring, when *n* is not 6.

Heterocyclic Systems.—Each hetero ring in a system is ciphered by an operation with D, and these citations precede those with C, which describe the basic ring structure. The intention here is to facilitate the location of all systems containing a specific hetero ring, such as pyrimidine, and to improve the classification of related types of heterocyclic compound in a cipher index. The size of the hetero ring, when it is not six-membered, is shown by a number following D. Next comes the citation of hetero atoms in standard order (*cf. The Ring Index*), the symbol being repeated for each like hetero atom. Their locants then follow in the same order, and for systems with two or more hetero rings, this sequence of operations is carried out for each in turn. The resulting ciphers, therefore, show each type of heterocyclic ring which is present in a system (Fig. 5).

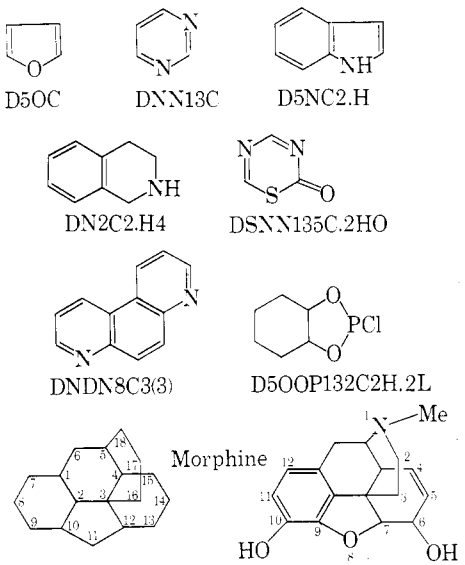


Fig. 5.—Heterocyclic systems.

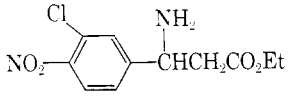
ANALYSIS OF STRUCTURES

Components and Simple Substituents.—Each time that a basic symbol A, B, or C is used we are potentially, if not actually, specifying a unit of structure with its own system of enumeration, and when the position(s) of attachments to this needs to be specified, the enumeration must be clearly distinguishable from that of adjoining chains and rings. Structures are, therefore, analyzed into *components* and *simple substituents*. The latter are the monosubstituted chains discussed above; *i.e.*, they are mainly the simple substituents, like alkyl and acyl derivatives of functions, and they are distinguished from other carbon chains because it is generally advantageous not to start a cipher with a part representing a simple terminal substituent when a more complex structure is present. All other carbon chains rank as components, and so do all ring systems.

The cipher for a simple molecule, *i.e.*, one with only one component, always commences with the description of this chain or ring. For more complex molecules, ciphering starts at the senior end of the longest and most senior sequence of components (unless this is a function of fourth rank), and it continues progressively through to the other end. This sequence is, in the first place, simply that of the basic symbols A, B, C, and D, so that it is frequently determined in an extremely simple manner, even for complex molecules.

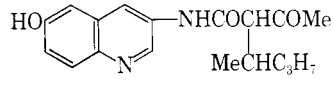
The general situation, in which a component R is joined at position *m* to another component R′ at position *n* is represented by R*m*-*n*R′. The hyphen marks the transition

to the next component, so that all locants following it must refer to this component. When there is a noncarbon atom(s) forming a *linking group* X between the two components, the symbol(s) for it can be inserted between *m* and *n*, giving R*m*X-*n*R′ or R*m*-X*n*R′. The first is more common; the second is used when a function of higher rank is encountered in reversed form (Fig. 6).



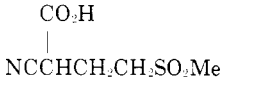
(B-A)

B4NOO·3L·-3A3OQA2·3M



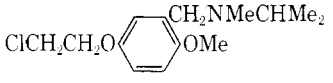
(D-A)

DNC2·7Q·3-MOA4·3O·2A4A



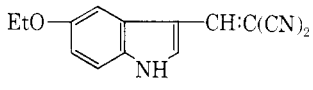
(A3-A2)

A3OQ·3N·2-2A2SOOA



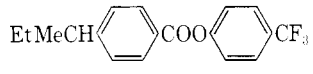
(A2-B-A)

A22L·Q-5B2QA·AM(A)A2A



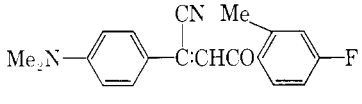
(D-A-A)

D5NC2·H·5QA2·3AE-2A313N

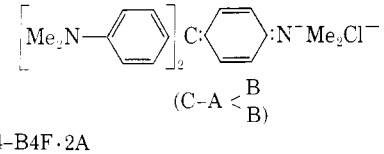


(B-A-B-A)

B4A3A·AOQ-4BAFFF



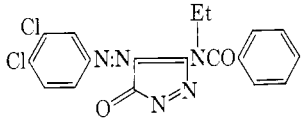
(B-A-B)



(C-A < B B)

B4M(A)2·-2A4N·4O·2E·4-B4F·2A

C·4HEM+(A)2·HE-A11-B4M(A)2··L-

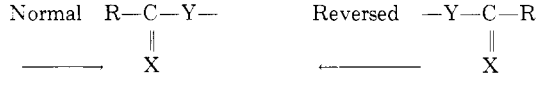


(B-D-A-B)

B24L·NN-4D5NN12C·3HO·4-(A2)MOA-B

Fig. 6.—Multicomponent systems.

Reversal of Functions of Higher Ranks.—An ester may be encountered either as R·CO·O·R′ or as R·O·CO·R′, and this situation is a general one for functions of this rank.



The normal form will generally be ciphered by XY, where X is either a terminal atom or =NH, and Y is a linking atom or group, and the reversed form will be YX.

In fourth-rank functions, which are frequently encountered as links between two other carbon systems, the linking atom Y is always encountered before the carbon or other atom on which the function is centered. In effect, therefore, these functions are always encountered in the reversed form, so that it is logical to use MOAM for urea, rather than MAOM, and in general to use YXAZ, not YAXZ. This reversal is also advantageous for indexing, since it leads to complete separation of different classes of derivatives, such as the alkyl, alkylidene, acyl, and ureido derivatives of an amine, whose ciphers, respectively, are RMA*n*, RMEA*n*, RMOA*n*, and RMOAM; if the O symbol followed A or A*n*, the principle dividing factor would be the value of *n*, and for each value of this the alkyl and acyl derivatives would alternate, with all the alkylidene derivatives coming later.

One-Carbon Components.—In many instances one-carbon components, other than fourth-rank functions, can be treated simply as linking groups. When no reference-back is needed, the hyphen can be omitted and ciphering made continuous. This applies for such groups as $-CH_2-$, -CHCl-, $-CCl_2-$, -C-, or -C-, and combinations of these

‖ ‖
O S

with -O-, -S-, -NH-, or =N-, as in COO, CONH, CH_2O (Fig. 6).

SENIORITY PRINCIPLES

For systematic ciphering, structures need to be analyzed into parts and the seniorities of these assessed. Although it is possible to decide seniorities entirely by alphabetical and numerical criteria, the resulting enumerations and ciphers for related compounds, as well as the index order of related compounds, are less satisfactory from a chemical viewpoint than when they are based primarily on broad classification principles, with the alpha numerical rule used only in the later stages for deciding between fairly similar alternatives.

The analysis of carbon skeletons into components and simple substituents has been described, and so has the classification of functional groups.

Seniority of Carbon Structures.—Five divisions are recognized, and in order of increasing seniority these are

(1) The simple substituents	(A)
(2) Other carbon chains	(A)
(3) The single benzene ring	(B)
(4) Other carbocyclic systems	(C)
(5) Heterocyclic systems	(D and C)

For both types of carbon chain the multiplier of A is the first criterion of seniority, and further assessment is based on the attachments considered in the order as for enumeration.

For rings a stepwise procedure is employed. The multipliers of D and C are considered first (counting D6 for a hetero six-ring, even though the "6" is omitted from the cipher):

C, C2, C3, . . .
D3C, D3C2, . . . D4C, D4C2, . . .
D5C, D5C2, . . . D6C, D6C2, . . .
D6D5C2, D6D6C2, D6D6C3, . . .

Next, the hetero atoms and their locants are inserted and the resulting expressions compared alphabetically and numerically. Thirdly, among ciphers identical at this

stage, the fusion locants describing the ring structure are compared, and, finally, the attached rings, functions, etc., in the order in which they are considered for enumeration.

For rings and chains with identical substituents at different positions, the one with the substituent at the higher-numbered position is senior.

ENUMERATION

Attachments to chains and rings are considered in the following order to determine their enumeration.

Carbon chains	Ring systems
(1) Functions	(1) Other ring systems linked (a) directly, and (b) *via* carbon atoms
(2) Branches with functions	(2) Chains linked directly and bearing functions
(3) Ring systems linked directly (not through functions)	(3) Functions on ring atoms
(4) Unsaturation and unsaturated branches	(4) Unsaturation or additional hydrogen
(5) Alkyl groups	(5) Unsaturated chains and alkyl groups

The appropriate seniority rules are applied when two or more items in the same category have to be considered. Functions are considered first in their unsubstituted forms. Thus, an ether is split to give two hydroxy compounds, and an ester to give an acid and a hydroxy compound; in the latter case it is, of course, only the acid component which possesses a third-rank function when seniority is under consideration. When like functions have different substituents the seniorities of the latter are compared.

The *order of citation* of attachments on a ring or chain is not quite the same as that which applies for enumeration, mainly because substituents associated with one component must, where possible, be ciphered before going on to the next component.

Chains	Rings
Third-rank functions	Hydrogenation or unsaturation
Second-rank functions	Second-rank functions
First-rank functions	First-rank functions
Unsaturation	Alkyl groups and other simple substituents
Alkyl groups and other simple substituents	Next component
Next component	

OTHER FEATURES

Other Elements.—Apart from phosphorus, for which P is available, the customary atomic symbols are prefaced by Z, and a capital letter is used for the second, as well as the first, letter of two-letter symbols. There are, however, four special symbols, namely, ZN zinc, ZE selenium, ZI silicon, and ZO sodium.

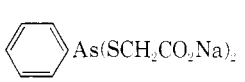
Stereoisomers.—An asterisk * denotes a D-configuration, and an oblique /, an L-configuration, and it is inserted between the locant and the substituent symbols. In parentheses at the end of a cipher these symbols denote simply "dextrorotatory" and "levorotatory." In complex structures, like steroids, the asterisk may be used to denote a substituent above the plane of the ring system, and the stroke, a substituent below.

For geometrical isomers V denotes a *cis* configuration and W, a *trans* configuration. They are also used in appropriate cases to denote *endo* and *exo*, respectively.

Isotopes.—Deuterium (ZD) and tritium (ZT) are ciphered as normal substituents. For carbon atoms in a chain or ring the symbol I is used alone as if a substituent were being ciphered, and for other elements, I is added immediately after the usual symbols. It is understood in the absence of further indication that isotopes of the following mass numbers are referred to: C-14, O-18, N-15, S-35, Cl-36, P-32.

Polymers.—Square brackets or braces are used to denote repetition of units in linear sequence; in typescript { and } are suitable. The repeating unit is then ciphered in the normal manner, as if it were a component or a sequence of components. Ethylenic polymers are ciphered in terms of the polymer unit, not of the monomer. Components which form appendages on the polymer chain are enclosed in parentheses, in the same way as complex branches in monomeric compounds. For copolymers the ciphers for the individual units can both be enclosed within the same set of brackets and linked by a plus(+) sign between them; the proportions in mole per cent may be added in parentheses after each unit. In addition to the normal optical and geometrical isomerism, RR may be used to denote a sequence in which all units have the same orientation, such as isotactic, and R/R for regular alternation, such as syndiotactic.

Some typical examples of the features mentioned in this section are shown in Fig. 7.



B-ZAS(T-2A2OQZO)2

A2TPS(QA)2·2TA2

A3OM·3TA·2*M

B4AI··A2OI·2ZD

Polyvinyl alcohol

[12A2·Q·]

Polyethylene glycol

[2A2Q-]

Polystyrene

[12A2(B)·]

Nylon 66

[A616OM-16A6]

Fig. 7.—Some other features.

A detailed account of this system is available on request. An experiment in structure searching of a set of about 1,500 of these ciphers by computer is in progress.

REFERENCES

(1) "Rules for IUPAC Notation for Organic Compounds," Longmans, Green and Co., Ltd., London, 1961.
(2) J. A. Silk, *J. Chem. Doc.*, 1 (3), 58 (1961).