LITERATURE CITED

(1) Nubling, W., and Steidle, W., "The Documentationsring der Chemisch-pharmazeutischen Industrie; Aims and Methods," *Angew. Chem., Int. Ed. Engl.*, **9**, 596–8 (1970).

(2) Granito, C. E., Roberts, S., and Gibson, G. W., "The Conversion of Wiswesser Line Notation to Ring Codes. I. The Conversion of Ring Systems," *J. Chem. Doc.*, **12**, 190–196 (1972).

(3) Granito, C. E., "Chemtron and the Interconversion of Chemical Substructure Systems," *J. Chem. Doc.*, **13**, 72–74 (1973).

(4) Granito, C. E., Becker, G. T., Roberts, S., Wiswesser, W. J., and Windlinx, K. J. "Computer Generated Substructure Codes (Bit-Screens), *J. Chem. Doc.*, **11**, pp 106–110 (1971).

ALWIN—Algorithmic Wiswesser Notation System for Organic Compounds†

E. V. KRISHNAMURTHY,* P. V. SANKAR,** and S. KRISHNAN

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India

Received April 14, 1973

ALWIN, a new chemical notation system for organic compounds, based on the Wiswesser Line Notation, is described. Procedures and rules are given for constructing ALWIN for acyclic structures and cyclic structures, *viz.*, benzene and its derivatives, monocyclic, bicyclic, polycyclic, perifused, spiro, bridged ring, and ring of rings systems. A new method called "tessellation" is introduced for the topological description of fused and spiro ring systems. Also new concepts are introduced for describing bridged ring and ring of rings systems.

The enormous increase in the number of organic compounds, natural and synthetic, especially the latter, has necessitated the development of nonconventional chemical codes for the representation of organic compounds.[1-3] This paper introduces a new chemical coding or notation system called ALWIN (Algorithmic Wiswesser Notation) which retains the salient features of WLN.[3] The logical and information structure of ALWIN has been designed to algorithmize efficiently the encoding and decoding procedures.

ALWIN differs from WLN in the following aspects.

(1) The acyclic structures in ALWIN can be represented in two equivalent forms: the parentheses form and the parentheses-free form.

The former is useful for human comprehension; here the parentheses differentiate the side group symbols from the main chain symbols. (Note that the symbol ampersand "&" plays a similar role in WLN.)

The parentheses-free form of ALWIN resembles the Polish notation (Korfhage[4] and Knuth[5]) and is obtained by dropping the parentheses; this is economical from the point of view of storage and computation.

(2) When certain chemical groups occur repeatedly in the chemical structures, it is possible to contact the corresponding notations. For this purpose, a set of new rules for multiplier contraction, based on the symmetry of the structures, has been developed.

(3) The contraction of notation can be effected in some cases using the well-known concept of factorization in algebra. Here, whenever certain sets of symbols occur repeatedly in a string of symbols, they are factored out much the same way as in algebra.

(4) A novel method of delineating the cyclic structures as a tessellation (or tile-filling on a floor) is used.

It is to be noted that WLN makes use of a Hamilton path[3] passing through the chemical graph, whenever it exists, or a minimal spanning tree for the topological description of the chemical graph. For large chemical graphs such a description is not desirable, since the existence and the choice of a Hamilton path in these graphs appear to be very difficult problems in graph theory.[6]

(5) A new set of rules for effecting the uniqueness of the notation in fused cyclic structures has been proposed.

(6) The main ring and the bridging branches of a bridged ring system are encoded separately, and these are then cited with suitable punctuation; this eliminates the necessity for the introduction of concepts, such as the "pseudo-bridges" used in WLN.

(7) Also distinction is made between the two different types of ring of rings systems, *viz.*, a ring system connected back to itself and another in which many rings are fused together, enclosing an inner ring which shares two or more edges with at least one of the outer rings.

(8) Several novel algorithms for processing the chemical graphs have been developed,[7] and these serve as the basis for encoding into the decoding from ALWIN.

(a) An algorithm, similar to the parsing techniques of the Polish notation, which rapidly decodes the parentheses-free ALWIN.

(b) A procedure to detect a set of chemically significant fundamental rings which constitute a fused ring system; this can in general find planar projections of chemical graphs (graphs of maximal valency four).

(c) Another procedure which differentiates the bridged ring systems from the ring of rings systems.

(9) A new set of rules for representing the stereoisomers has also been developed.[8]

GLOSSARY

ALWIN comprises a set of symbols and a set of rules for the encoding and decoding of organic chemical structures. The symbols are listed first, and the rules are presented with examples later.

† Contribution No. 46, Molecular Biophysics Unit, Indian Institute of Science, Bangalore, India.

* Author to whom correspondence should be addressed.

** Presently at Tata Institute of Fundamental Research, Bombay 400005.

a. Character Set (Alphabet) of ALWIN

1. The capital letters of the English alphabet A through Z.
2. The decimal digits Ø, 1, 2, . . . , 9.
3. The special characters = # % ; : ( ) / * + , . $ & 〈 〉 @ -.
4. The blank space (denoted by ♭ for convenience).

b. Alphanumeric Hierarchy. (Highest order) = # % ; : ( ) / * + , . & 〈 〉 - $ ♭ Ø 1 2 3 4 5 6 7 8 9 @ A B C D E F G H I J K L M N O P Q R S T U V W X Y Z (lowest order). (Note: The numeral 0 (zero) is crossed to distinguish it from letter O).

c. Symbol Set. The WLN symbols are retained for the elements, branching atoms, functional groups, bonds, unsaturations, and rings. However, unbranched carbon atoms are denoted as below.

Unbranched carbon atom connected to two other non-hydrogen atoms	C
CH_3	A
CH	@
CH_2	1
CH_2-CH_2 . . . $(CH_2)_n$	*n*

d. Symbols to Denote Varying Valency. Usually the normal valency of any element is assumed and hence not cited in the notation; however, if an element exhibits a deviation from its normal valency, we indicate this as follows.

Cobalt	-CO-	(normal valency of 2)
	-CO3-	(valency 3)
Phosphorus	P	(normal valency of 3)
	-P5-	(valency 5)

ACYCLIC STRUCTURES

The linear notation in ALWIN is obtained from the structural formula of any organic compound by substituting suitable ALWIN symbols for the connected atoms and then ordering them according to certain well-defined rules (grammar). Hydrogen atoms, in general, are not cited in the notation.

a. Unbranched Structures. Unbranched structures have linear structural diagrams; consequently, the substitution of the notation symbols for the respective atoms gives rise to linear codes. These codes can be read from either end. Hence, in general, when two or more ALWIN codes are possible, uniqueness is achieved by the following rule.

*AC 1**

Choose that code which starts with a sequence of symbols having the lowest order.

Example 1

CH_3—CH=CH—Cl ALWIN: G@U@A

In this example there are two possible codes, *viz.*, A@U@G and G@U@A; the latter is chosen, since G has a lower order than A.

b. Branched Structures

AC 2

Choose that sequence of nodes in the chemical graph as the main chain satisfying the precedence given rule AC 2a. Form the notation of the main chain using rule AC 1. Now complete the notation by inserting the side group symbols within parentheses immediately after each branching symbol in the main chain using the subrule AC 2b.

AC 2a. Choice of the Main Chain

The main chain is defined as that chain which has (in the following order of precedence)

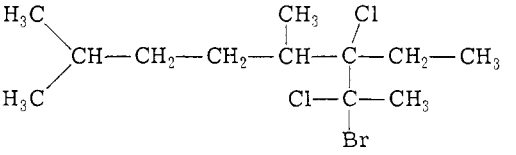
(1) the maximum number of branching atoms
(2) the maximum number of symbols in the notation
(3) the lowest symbol order

AC 2b. Precedence among the Branching Side Groups

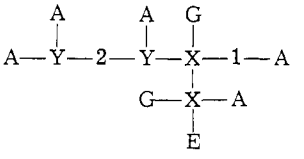
If the branching symbols have more than one side group, the side groups are cited in the following order of precedence

(1) with the minimum number of branch symbols
(2) with the minimum number of total symbols
(3) which start with the symbols in the lowest order.

Example 2



The structural formula is first translated into the following graphic form



The main chain is GXXY2YA. Thus, ALWIN is written as GX(E)(A)X(G)(1A)Y(A)2Y(A)A

c. Multiplier Contraction. When groups of symbols occur repeatedly, it is possible to contract the notation by citing each repeated group once and indicating the number of times that group occurs and also its location.

In some cases, the repeated occurrence of groups results in certain types of symmetry; to encode such structures we adopt the following conventions.

(1) Any numeral preceded by a blank space (♭) is called a "multiplier."

(2) The multiplied group and the multiplier always occur within a pair of brackets (〈 and 〉). The multiplied group is defined as the sequence of symbols between a multiplier and the symbol 〈 immediately preceding it.

(3) The multiplier numeral is always followed by the symbol 〉 and preceded by a blank and hence can never be confused with an alkyl numeral.

(4) The sign $ denotes an inverse insertion operation; this operator reverses the group of symbols that occurs between a pair of symbols 〈 and 〉 and inserts an inverted group of symbols in the place of $.

AC 3. Rule of Inversion

If $S = a_1 a_2 \ldots a_m$, then the inverse of S denoted by $\bar{S}$ is defined as follows.

(1) If a_1 = ((left parenthesis) and a_m =) (right parenthesis), then $\bar{S} = S$. The symbols enclosed within the parentheses are not inverted.

(2) If $a_1 \neq (, a_i = (\text{ and } a_{i+j} =)$ for $i + j \leq m$; *i.e.*, if

$$S = a_1 a_2 \ldots a_{i-1}(a_{i+1} \ldots a_{i+j-1}) a_{i+j+1} \ldots a_m$$

then

$$\bar{S} = a_m a_{m-1} \ldots a_{i+j+1} a_{i-1}(a_{i+1} \ldots a_{i+j-1}) a_{i-2} \ldots a_2 a_1$$

3. If there is no left or right parenthesis in the expression, then

$$\bar{S} = a_m a_{m-1} \ldots a_2 a_1$$

Example 3

$$S = (VQ); \bar{S} = (VQ)$$

Example 4

$$S = QVX(G)(G)G; \bar{S} = GX(G)(G)VQ$$

* Rules for acyclic and cyclic structures are prefixed by AC and C, respectively.

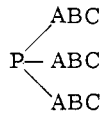
Example 5

$$S = 1UNVQ;\ \bar{S} = QVNU1$$

AC 4. n-Fold Symmetry

The notation $P(\Sigma)\ (\Sigma)\ \ldots\ n$ times is contracted as $P\langle(\Sigma)\not{b}n\rangle$ for $n \geq 3$, where Σ is a sequence of symbols; *i.e.*, $\Sigma = a_1a_2 \ldots a_m$ and P is a group of symbols or a single symbol.

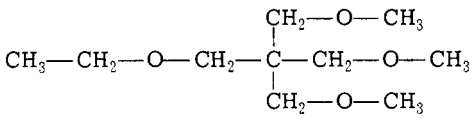
Example 6



Coded as P(ABC)(ABC)(ABC) ≡ P⟨(ABC)♭3⟩

(≡ denotes identically equal to)

Example 7

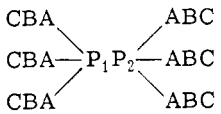


Coded as A101X⟨(10A)♭3⟩

AC 5. n-Fold Biaxial Symmetry

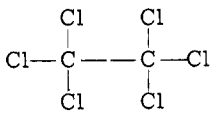
The notation $P(\Sigma)(\Sigma) \ldots n$ times $Q(\Sigma)(\Sigma) \ldots n$ times is contracted as $P\$Q\langle(\Sigma)\not{b}n\rangle$. Here $ indicates that the inverse of the multiplied groups should be inserted in its place while expanding the notation.

Example 8



Coded as $P_1(ABC)(ABC)(ABC)P_2(ABC)(ABC)(ABC) \equiv P_1\$P_2\langle(ABC)\not{b}3\rangle$

Example 9

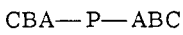


Coded as GX(G)(G)X(G)(G)G ≡ X$X⟨(G)♭3⟩

AC 6. Bilateral Symmetry

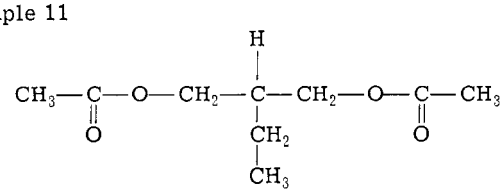
The notation $\bar{\Sigma}P\Sigma$ is contracted to $P\langle\Sigma\rangle$.

Example 10



Coded as CBAPABC ≡ $P⟨ABC⟩

Example 11

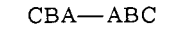


Coded as AV01Y(1A)10VA ◘ $Y(1A)⟨10VA⟩

AC 7. Bisymmetry

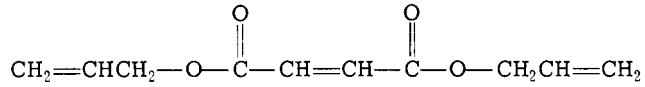
The notation $\bar{\Sigma}\Sigma$ is contracted to $\$\langle\Sigma\rangle$.

Example 12



Coded as CBAABC ≡ $⟨ABC⟩

Example 13



Coded as IU@10V@U@V01@U1 ◘ $U⟨@V01@U1⟩

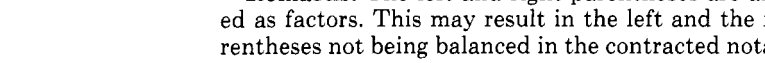
AC 8. Serial Replication

The notation $\Sigma\Sigma \ldots n$ times is contracted to $\Sigma\not{b}n$

Example 14

ABC—ABC—ABC—ABC is coded as ABCABCABCABC ◘ ⟨ABC♭4⟩

Example 15



Q202 ... 52 times Q ◘ Q2⟨02♭52⟩Q

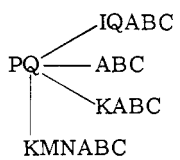
d. Factorization. The technique of factorization used in algebra is adopted in ALWIN for factoring repeated groups. The character "+" retains its conventional meaning inasmuch as it acts as a punctuation between a pair of groups to each of which the factored group is attached. The symbol * is placed before (after) the factored group in the case of right factorization (left factorization) so as to indicate that the sequence of symbols that follows (precedes) * up to the next following symbol ⟩ (preceding symbol ⟨) as the factored group.

AC 9. Right Factorization

The notation $PA_1QA_2QA_3Q \ldots A_nQ$ is contracted as $P\langle A_1+A_2+ \ldots +A_n*Q\rangle$, where P, A_i, and Q are sets of symbols.

Remarks. The left and right parentheses are also treated as factors. This may result in the left and the right parentheses not being balanced in the contracted notation.

Example 16

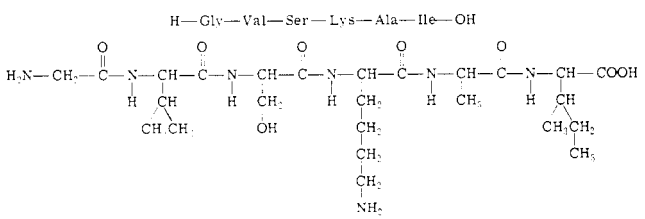


Coded as PQ⟨(IQ + (+ (K + (KMN*ABC))⟩

AC 10. Left Factorization

The notation $PQA_1QA_2 \ldots QA_n$ is contracted to $P\langle Q*A_1+A_2+ \ldots +A_n\rangle$.

Example 17



Coded as

Z1VMY(Y(A)A)VMY(1Q)VMY(4Z)VMY(A)VMY(VQ)Y(A)1A

and further contracted to

Z1⟨VMY(*Y(A)A) + 1Q) + 4Z) + A) + VQ⟩⟩Y(A)1A

AC 11. Compound Factorization

Rules AC 9 and AC 10 can be generalized into a "compound rule of factorization." The notation $PA_1QB_1A_2QB_2 \ldots A_nQB_n$ is contracted as $P\langle A_1+A_2+ \ldots A_n*Q*B_1+B_2+ \ldots B_n\rangle$.

AC 12

When there are several ways of contracting the notation, choose that code

(1) which gives the least number of symbols

(2) with the highest order

Criteria for Using the Multipliers. Since the multiplier contraction should decrease the total number of symbols, the following criteria can be observed for the use of multipliers.

Criterion 1. Rules AC 4 to AC 8 can be used if $mn \geq 8$, where n is the numerical value of the multiplier and m is the number of symbols in the multiplied group.

Criterion 2. Rules AC 9 to AC 10 can be used if $mn \geq 8$, where n is the number of groups to which the factored group is attached and m is the number of symbols in the factored group.

e. Organic Salts and Addition Compounds. Definition 1. All alkali and alkaline earth compounds where the metal is not bonded to a carbon atom and all inorganic salts of organic acids (carboxylic, sulfonic, sulfinic, phosphonic, phosphinic, boronic, etc.) are called ionic salts.

Definition 2. The ternary sulfur, the quaternary nitrogen, and analogous "onium" compounds are considered as organic salts.

AC 13

The ions (inorganic and organic) are encoded separately using the rules prescribed earlier; these are then cited together in the notation by inserting a suitable punctuation, namely, an ampersand & preceded by a blank space between any pair of such ions. As a convention, the organic ions are cited first in the notation followed by the inorganic ions, except H_2O (water) which is cited last. Hydrogen atoms are cited whenever clarity is needed.

AC 14

The proton transfers on organic base salts or salts of unsubstituted ammonia (ZH) are not cited, and such compounds are encoded as addition compounds.

Example 18

$H_2N—CH_2CH_2CH_2—NH_2 \cdot 2H_2O$ Z3Z ♭&QH♭&QH

f. Parentheses-Free ALWIN. According to rule AC 2, we represented the branched structures using parentheses. The parentheses procedure is very helpful for human comprehension and manual encoding. However, for large structures, such a procedure results in a large number of symbols. To avoid this, we now propose a parentheses-free ALWIN for acyclic structures.

The parentheses-free ALWIN is derived from ALWIN by deleting all the left and right parentheses and closing up the other symbols retaining their order; this resembles the Polish notation,[4] widely used for representing arithmetic expressions in computers; however, here we have only a single operator, *viz.*, the adjacency relation or connectivity between the atoms.

E.g., the Polish notation for example 2 is thus given by G X E A X G 1 A Y A 2 Y A A by deleting all the left and right parentheses.

However, the fact that we do not use any delimiter or punctuation symbol, like ampersand & in WLN, necessitates a distinction between CH_3, CH_2, and CH. It is for this reason that these are assigned the separate symbols A, 1, and @, respectively.

A simple algorithm is available for translating from the parentheses-free form of ALWIN to the connectivity table and *vice versa*, using the valencies of the symbols involved.[7]

In view of the simplicity, economy of storage, and availability of simple algorithms (encoding and decoding), we use only parentheses-free ALWIN for automatic retrieval purposes.

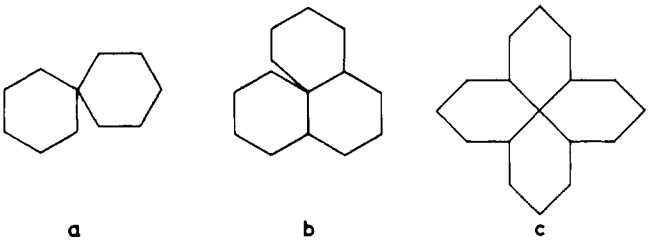


Figure 1.

GENERAL INFORMATION WORD FORMAT (OR DATA BASE) FOR THE CYCLIC STRUCTURES

Before considering benzene and other monocycles, we will first explain the ALWIN information format for general cyclic structures. Later, we will deal with benzene and monocycles as particular cases.

a. Categories of Fused Ring Systems. A system of rings in which two or more rings have a common node or edge is called a fused system; the common edge is called the fusion edge and the common atoms are called the fusion locants or fusion nodes. A multicyclic point is defined as a node which is common to at least three rings.

With these definitions one can, in general, classify the fused ring systems broadly into three categories.

(i) The bicyclic fused ring system consisting of only two rings.

(ii) The polycyclic fused ring system having more than two rings, but no multicyclic point.

(iii) The perifused ring system consisting of a minimum of three rings with at least one multicyclic point.

(iv) The spiro ring system in which a perifused ring system possesses a spiro atom—an atom (or node) is called a spiro atom if it is connected to four other ring atoms.

There are three kinds of spiro atoms.

1. A spiro atom shared by only two rings (nodesharing) (Figure 1a) (first kind).

2. A spiro atom shared by three rings (edge-sharing) (Figure 1b) (second kind).

3. A spiro atom shared by four rings (edge-sharing) (Figure 1c) (third kind).

(v) The bridged ring system in which two rings share more than a single edge, with the shared edges considered as the bridge. The locants of the ring to which the bridge is connected are called the bridgeheads. In general, all these systems have a general word format, with an increasing complexity in the code, from the monocycles onwards up to the bridged ring systems.

(vi) The type I ring of rings system in which a ring system is connected back to itself, through another ring system or a set of ring systems; the type II ring of rings system, in which many rings are fused together in such a way as to enclose completely another ring, which shares two or more edges with at least one of the above-mentioned rings.

The bridged ring systems and the ring of rings systems are classified here as poly-edge shared ring systems; in these systems there is at least a pair of rings which shares more than one edge.

b. Information Word Format for Cyclic Structures. The Information Word Format (IWF) for the cyclic structures is shown in Table I. The format is divided into 14 zones, each zone being separated from the next by suitable punctuation marks or implicit delimiters.

c. Zone Description

Ia. Zone of Ring of Rings. The notation for ring of rings, type I, starts with the symbols &1, while the notation for the ring of rings, type II, starts with the symbols &2.

Ib. Zone of Carbo/Hetero Symbolism. Depending on whether the cyclic system is carbocyclic or heterocyclic, a symbol L or T is cited at the beginning of the notation.

Table I

Ia	Ib	II	III	IV	V	VI	VII
&1 or &2	L or T	Total saturation symbol	Ring numerals with saturation symbols if necessary	Tessellation code	:	Knit code	;

VIII	IX	X	XI	XII	XIII	XIV
Multi-knit code	=	Heteroatoms functional groups and saturation or unsaturation specification	:	Bridge description	J	Substituents

II. Zone of Total Saturation Symbol. A saturation symbol T indicates that all the rings in the given ring system are saturated; the absence of T implies that they are unsaturated.

III. Zone of Ring Numerals. The ring numeral refers to the number of ring atoms in each ring of the given ring system. If the ring numeral is greater than 9, we put the ring numeral between two hyphens. The ring numerals are cited in a specified order (to be discussed in the section dealing with the formation of tessellation code).

C 1

If any ring has two or more saturated carbon atoms (carbon atom connected to four other atoms), then that ring is considered saturated; otherwise, it is unsaturated.

IV–IX. Zone of Tessellation–Knit Code. Principle of Code Construction. Since the construction of codes are based on graph theory, we will now define some of the important terms that will be used here.

Nodes or vertices	Points of connections (atoms)
Edges or arcs	Line segment connecting two nodes (bonds)
Graph	A collection of nodes connected through arcs
Chemical graph	A structural diagram of a molecule represented as an abstract graph
Degree of a node	The total number of arcs connected to a node
Tree	A connected graph, having at least two nodes, that has no circuits; there is a unique arc connecting any pair of nodes
Spanning tree	A tree which spans all the nodes of a given graph
Elementary path	A path which visits each of the nodes contained in it only once—a tree in which there are no branches
Minimal spanning tree	Any tree can be thought of as the union of a set of elementary paths; a minimal spanning tree is a spanning tree composed of the least number of elementary paths
Hamilton's path	An elementary path which passes through all the nodes of the given graph
Reduced graph	The reduced graph is an edge or node relational graph in which the rings in the original chemical graph (consisting of only rings) are represented as nodes and the sharing edges or nodes between them as edges in the reduced graph
Elementary tree	A maximal subtree derived from a spanning tree of the reduced graph, such that the nodes specified in this subtree do not constitute a cycle in the reduced graph

The codes for the fused and spiro ring systems can be constructed by the following two methods which describe the topology:

1. The Hamilton path method, or the minimal spanning tree method (when the Hamilton path does not exist).[3]
2. The tessellation method presently used in ALWIN.

The Hamilton path is a path through the graph which traverses all the nodes without revisiting any node over again. WLN is essentially based on the choice of paths (whenever available) in the chemical graphs; if such a path is not available, a set of long paths constituting a minimal spanning tree is chosen for the description. While the notation based upon this approach gives a complete topological description, it leads to difficulty inasmuch as there is no simple method to find the existence of Hamilton paths in large graphs. Even trial and error procedures for such large graphs would be computationally very laborious and time consuming; this complicates the computer encoding procedure.

In ALWIN, however, we treat the planar projection of the given chemical sturcture consisting of rings as a tessellation (or mosaic floor) or a tile-filling problem. The ordered sequence of such tile-filling operations completely describes the given chemical graph corresponding to the fused system.

The construction of this code is best explained by means of an example.

Consider the chemical graph shown in Figure 2a. Choose any particular ring and label its nodes (locants) clockwise (Figure 2a). The choice of the first ring and the sense of labeling are governed by certain rules to obtain a unique code; this aspect will be treated later. By convention, we name each edge by its cyclically preceding node label. Then, we choose a ring which shares a common edge with the first ring and label its nodes (continuing, from the previous ring), clockwise, starting from that node which has not been labeled and which would permit us to give a consecutive clockwise assignment of the locants (the sense of labeling is totally clockwise or anticlockwise so as to obtain a unique code). Reiterate this process until all the rings in the given structure are exhausted.

The topology is now described by first specifying the ring numerals in the order in which the locants have been assigned to them, and then the common edges (when a ring shares only one common edge with any other ring) and connectivities—called knit (when two or more edges are shared by other rings); this can be implemented easily by drawing the reduced graph.

The reduced graph is an edge or node relational graph in which the rings in the original ring system are thought of as the nodes and the sharing or common edges and nodes between the rings as connections between the corresponding set of nodes. In obtaining this reduced graph, priority is given to e edge relationship over the node relationship, when both these relationships coexist; this eliminates redundancy in the description.

The reduced graph for Figure 2a is shown in Figure 2b. Here the nodes 1, 2, and 3 correspond to the three rings in

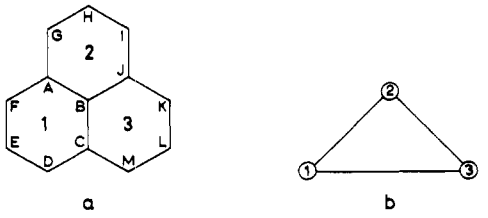


Figure 2.

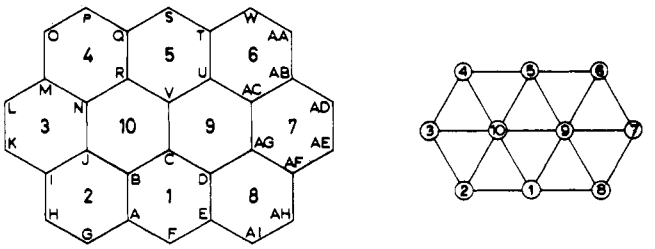


Figure 3.

the order in which they were assigned locants; the arcs 1–2, 2–3, and 3–1 in the reduced graph correspond to the common edges AB, JB, and CB, respectively. The code, in general, consists of the following parts.

i. The Tessellation Part. Here each common edge is represented by a single locant; this corresponds to the description of the edges of elementary trees–maximal subtrees derived from the spanning tree such that the nodes specified in the subtrees do not constitute cycles in the reduced graph. This gives the tessellation code.

ii. The Knit Part. After describing the tessellation corresponding to the elementary tree, the remaining rings in the structure are described by citing ring segments which complete the new rings. It will be observed that in some cases a single connected segment suffices to complete a ring, while in other cases two or more independent segments are required to complete a new ring. The completion of a ring by a single, connected segment is denoted by citing the beginning and ending locants of the particular segment; those new locants which lie in between this pair of locants are easily labeled by continuing from the shared edges of the previous rings, using the new ring numeral. The completion of a ring by two or more segments is a little more involved. In this case, there is no simple algorithm to obtain the number of locants that are included in each one of these segments; hence, we cite the number of included locants that lie between the beginning and the ending locants of each segment.

The process of completing a ring by a single segment is called a "knit" and that which involves more than a single segment is called a "multi-knit." The knit is separated from the tessellation by a colon (:), and the multi-knit is separated from the knit by a semicolon (;).

Both the knit and multi-knit correspond to those edges of a spanning tree (of the reduced graph) which are not realized in tessellation; using this, the given structure is encoded.

In Figure 2b one spanning tree is 1–2–3 and the corresponding elementary tree is 1–2, since 3 is connected to 1. Thus, the edge 1–2 of the reduced graph corresponds to the tessellation in which the rings 1 and 2 share the common edge AB (this is denoted in the notation by A) while the edge 2–3 corresponds to the knit; this corresponds to the segment JKLMC of ring 3 (Figure 2a) and hence is indicated by the pair JC. The code for this graph is 666A:JC.

In Figure 3b one spanning tree of the reduced graph of Figure 3a is 1–2–3–4–5–6–7–8–9–10. The elementary tree is 1–2–3–4–5–6–7; this corresponds to a tessellation code A, I, M, Q, T, AB. The remaining edges of the spanning tree 7–8–9–10 correspond to knit or multi-knit as explained below. The edges 7–8 and 8–9 correspond to the two ring segments AF–AH–AI–E and D–AG which close the ring 8 in Figure 3a. This gives the multi-knit part of the code AF2E, D, AG. The edge 9–10, however, stands for the single connected segment (C–V) which closes rings 9 and 10 simultaneously in the original graph. This gives the knit part (C–V). Thus, the code is

6666666666A,I,M,Q,T,AB;AF2E,D,AG:C,V

Since the algorithm we use is sequential in nature, it is possible that tessellation, knit, and multi-knit occur many times in any order. This will be seen from the repeated occurrence of ":" and ";" in the appropriate places which indicate that the succeeding set of symbols correspond to the knit or multi-knit, respectively.

Also, note that although corresponding to every tessellation there is an elementary tree, the converse need not be true, since it is possible that some of the other edges in the reduced graph might have been realized by knit or multi-knit during the intermediate steps in the sequential algorithm and would only permit us to knit or multi-knit rather than tessellate (see example 29). Thus, during the intermediate steps it is necessary to check which of the edges in the reduced graph have already been realized.

During the reverse process of decoding the entire code is fully read and tessellation, knit or multi-knit, is performed in the required order.

X. Zone of Ring Atom Description

C 2a

Assign the highest order locants to the cited ring segment symbols in the following order of precedence.

(1) Heteroatoms, functional groups (in their alphabetical order)

(2) Specification of locants of saturation (in an unsaturated ring) or unsaturation (in a saturated ring)

(3) Spiro atoms

(4) Bridge locants

(5) The side groups by AC2b

(*Note:* Locant "A" is not explicitly cited in the notation.)

C 2b

If after citing a locant its consecutive locants are to be cited, then we cite only the higher order locant followed by the consecutive ring atom symbols.

C 2c

This rule overrules C 2a if there is a string of consecutive heteroatoms or functional groups; in this case the higher order locants are given to that end of the sequence which has the higher order symbols.

C 3a,b

Saturation in an unsaturated ring is indicated by specifying the locants of the saturated atoms, immediately followed by the symbol H. Similarly, unsaturation in a saturated ring is indicated by the two locants across which there is a double (triple) bond, thus: Blank space (ϕ) cyclically preceding locant U (UU) succeeding locant. If the two locants are alphabetically consecutive, the succeeding locant is not cited.

XI and XII. Zone of the Bridge Description. Zone XI consists of the delimiter symbol colon which indicates the presence of bridges in the given ring system. Note that the symbol ":" has also been used in zone V as a delimiter between the tessellation and knit codes; this will not lead to ambiguity, since this colon appears before the "=" symbol, while the colon for a bridge appears after the "=" symbol; also the colon in zone V is followed by locants without blank space, while the colon in zone XI is followed by a blank space and then a locant which specifies the beginning of the bridge. Zone XII contains the description of the bridge.

XIII. Zone of J. The symbol "J" is used as a delimiter immediately following the ring description; this separates it from the substituents.

XIV. Zone of Substituents. The side groups are cited with their locants.

MONOCYCLES

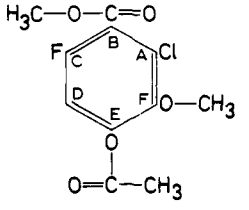
a. Benzene and Its Derivatives. The derivatives of benzene are so very numerous that the assignment of a single symbol "R" for benzene is statistically justifiable to obtain a compact notation. It may, however, be noted that the notation for benzene, corresponding to the general word format for cyclic structures is L6J.

In ring structures the nodes (atoms) constituting the ring are called the locants. These locants are denoted by the capital letters of the English alphabet A through W (A through F in the case of benzene), each preceded by a blank space.* We have chosen capital letters, since the numerals have already been used to denote a chain of alkyl carbon atoms, while the numerals preceded by a single blank space represent multipliers.

C 4

Treat R as a branch symbol and use the precedence rule AC2b to assign the higher order locants to those ring atoms with substituents.

Example 19



ALWIN: R␣AG␣BVOA␣CF␣EOVA␣FOA

In example 19, the side groups F and Cl have the least number of symbols, *viz.*, only one symbol; among them Cl is given the highest order locant since G is later than F. In order to give F a higher locant, the anticlockwise labeling has been chosen.

If there is a chain of benzene rings, one arrives at a unique notation using the rules C 6a and C 6b, proposed later for monocycles.

b. Factorization Rule for Cyclic Structures. When many of the substituent side groups are identical, it is possible to contract the notation by modifying the rule of right factorization described earlier (AC 9).

C 5a

The notation

$$P\not b\lambda_1 A_1 Q\not b\lambda_2 A_2 Q \ldots \not b\lambda_n A_n Q$$

is contracted as

$$P\langle \not b\lambda_1 A_1 + \not b\lambda_2 A_2 + \ldots + \not b\lambda_n A_n{}^* Q\rangle$$

where P, in general, is any ring system (P = R for benzene), A_i any sequence of symbols, Q the repeated substituent group, and $\not b\lambda_i$ the available locants of the ring system.

Example 20

Notation R␣A1V01A␣C1V01A␣E1V01A contracted to

R⟨␣A + ␣C + ␣E*1V01A⟩

* When the ring system has more than 23 atoms, the locants are numbered in base 23 with A through W as digits, *e.g.*, 24 is denoted by AA, 25 is denoted by AB, 553 is denoted by AAA, and so on. Whenever such locants occur together, a comma is placed between them. Also while citing the locants for ring atoms or substituents, those locants which have two or more letters are followed by a comma before citing the ring atoms or substituents.

C 5b

If three or more consecutive locants in the alphabetical order are present within symbols ⟨ ⟩ in rule C 5a, the notation is contracted further, by citing only the first and the last locants and a period "." in between these locants.

Example 21

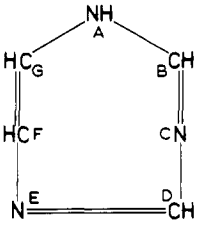
Notation R⟨␣A + ␣B + ␣C + ␣D*010A⟩␣E101VQ

contracted to R⟨␣A.␣D*010A⟩␣E101VQ

c. Monocycles Other than Benzene

In the IWF (Information Word Format), the zones Ia, II, IV, V, VI, VII, VIII, IX, XI, and XII are absent. We shall consider a few examples.

Example 22



ALWIN: T7M␣CN␣ENJ

Here we start the notation with the symbol T as the given structure is heterocyclic; since it is unsaturated, we do not cite the symbol T after the ring numeral. The heteroatom M (M being higher than N) is assigned the higher order locant "A"; the other ring atoms are labeled so as to give the higher order locants to the nitrogen atoms (rule C 2a).

C 6a. Cyclic Precedence Rule

In a chain of monocyclic rings, one of the rings is chosen as the main ring (the ring that is cited first in the notation) *by the following order of precedence.*

(1) The heterocyclic ring
(2) The ring of maximum size
(3) The unsaturated ring
(4) The ring with the maximum number of substituents
(5) The ring with the substituents having the lowest order symbols.

The other rings are treated as branches; where there are ambiguities in the choice of the substituent branching rings, the ring that satisfies rule C 6a is cited next; this process is reiterated until there are no more alternatives. In case there are still alternate ways of writing the notation, then the one which has the lower order symbols is chosen.

C 6b

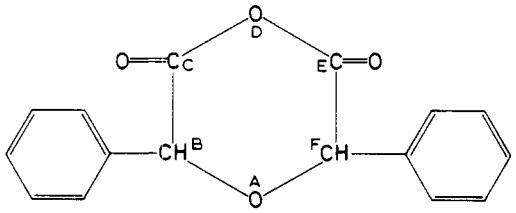
If a locant of some ring leads to a locant of another ring except benzene, then a hyphen is placed in the notation immediately after the connecting locant of the former ring; this is followed by the connected locant of the latter ring. In case the connected locant is A, it is not cited.

C 6c

It is necessary to enclose branching rings within parentheses (where required) to distinguish the locants belonging to a main ring from those of the substituent rings. See example 23.

Here the central ring is chosen as the first ring since it is heterocyclic (rule C 6a). This ring consists of an oxygen atom followed by a sequence of atoms VOV. Among these, the former is given the locant A and the other atoms are given the locants C, D, and E, respectively. In the code, the locant F (the last but one symbol in the notation) belongs to the main T6 ring and does not denote the locant of the

Example 23



ALWIN: T60ϕCVOVJϕB(R)ϕFR

benzene connected to locant B of the main ring; to make this unambiguous we enclose this R within parentheses.

BICYCLIC AND POLYCYCLIC FUSED RING SYSTEMS

a. Bicyclic Ring Systems. In the case of bicyclic fused ring systems, there is only one common edge; this is taken to be the first edge AB; hence the locant A can be considered implicit in the notation (without explicit citation).

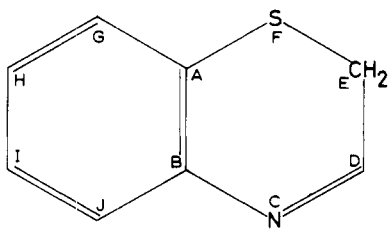
In the IWF for bicyclic systems, the zones Ia, IV, V, VI, VII, VIII, IX, XI, and XII are irrelevant.

C 7

In order to arrive at a unique code (for a given planar projection), form the notation for the mirror image of the chemical graph in addition to the original. This is equivalent to labeling all the locants uniformly in the anticlockwise direction and forming the code.

Among these codes, choose the codes which give the higher order locants for the chosen precedence in C 2a. If the codes are still unresolved, then choose that code which has the lower order symbols.

Example 24



ALWIN: T66ϕCNϕEHSJ

In Example 24 the heterocyclic ring has to be cited first, and hence the atoms in this ring are first assigned the locants. Since N has to be given a higher order locant than S (rule C 2a), the anticlockwise labeling was adopted. Note that the saturated carbon atom at locant E and S at F are cited as ϕEHS (rule C 2b).

b. Polycyclic Fused Ring Systems. Polycyclic fused ring systems consist of three or more rings with no multicyclic point. As a consequence, the reduced graphs of such systems themselves are the elementary trees; hence the notation consists of only the tessellation part. The information format for the polyfused rings comprises all zones except Ia, V, VI, VII, VIII, XI, and XII of IWF.

C 8a

The reduced graph of the polyfused ring system is obtained first. The first ring to be cited in the notation is chosen from the terminal nodes of the reduced graph which satisfies the precedence rule C 6a. With this as the starting point, tessellation codes are formed for both the original chemical graph and its mirror image; among them, choose that code which satisfies C9. If the reduced graph is not an elementary path but contains branches (tree), then go to rule C 8b.

C 8b

Here, since the reduced graph is a tree and the choice of the starting point is unique, there are several ways of assigning the locants to the subsequent rings. This is obviated by forming a unique code resulting in an ordering of the nodes of the reduced graph using precedence rules analogous to AC 2a and AC 2b.

Choice of the Main Chain. Choose that chain in the reduced graph

(1) which contains the maximum number of branching atoms in the reduced graph
(2) which contains the maximum number of nodes
(3) which contains those nodes corresponding to the rings satisfying the higher precedence in C 6a. Cyclic Precedence Rule.

Now the rest of the nodes can be considered as substituents and hence included in the code using precedence in C 8c. Once the code is formed, the rings are assigned locants in the order in which they occur in the code.

C 8c

Choose that substituent in the reduced graph

(1) which has the minimum number of branching nodes
(2) which has the minimum number of nodes
(3) which has the nodes corresponding to rings satisfying the higher precedence in C 6a.

C 8d

If, after the application of C 6a in C 8a, there are equal ways of writing the notation from two or more starting points, then form all the possible tessellation codes for the original chemical graph as well as its mirror image and among them choose the one which satisfies rule C9.

C 9

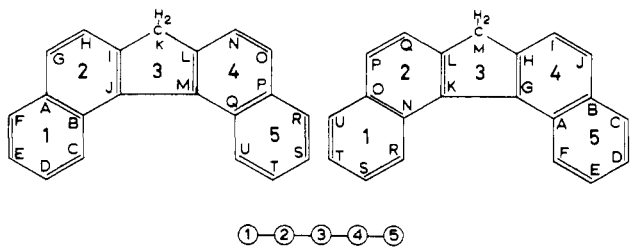
1. Among the several tessellation codes, choose that code which contains the maximum number of locants; among these tessellation codes of equal number of locants, choose that code which gives the least sum for the cited locants (by assigning integral weights in the natural order for A, B, C, etc., in lexicographic order).

2. If there are still equal alternatives, then choose that tessellation code which has the higher order symbols.

If the tessellation codes are identical, then compare the entire tessellation-knit code (only for perifused systems) which has the higher order symbols.

3. If still the codes are identical then choose that code which gives the higher order locants for the precedence cited in C 2a.

Example 25



ALWIN: L66566AGKN=ϕMHJ

In example 25, the terminal rings 1 and 5 are identical and either can be chosen as the first ring for the tessellation code. By C9 we choose that code which has the least sum of the cited locants in the tessellation code.

The tessellation code, with ring 1 as the first ring is AILP with a sum of 1 + 9 + 12 + 16 = 38; the tessellation code with ring 5 as the first ring is AGKN with a sum of 1 + 7 + 11 + 14 = 33.

PERIFUSED AND SPIROFUSED RING SYSTEMS

a. Perifused Ring System. Perifused ring systems are assemblages of rings (with a minimum of three) having at least one multicyclic point. As already mentioned, the code for the perifused ring system consists of two parts, *viz.*, the tessellation code and the knit code, the two codes being separated by the delimiter colon. The information format for these systems consists of all zones except the zones Ia, XI, and XII in IWF.

C 10

If the reduced graph of the perifused ring system consists of many branches (apart from the central core which will have a cycle of length ≥3), then the terminal node of the longest branch is chosen as the first ring. If the branches are of equal length, then the terminal node of that branch which satisfies the precedence in C 6a is chosen as the first ring. In the case of spiro ring systems, choose the spiro ring (terminal node) as the starting point. In case the precedence constraints are not capable of selecting the first ring (which means the terminal nodes of the branches are indistinguishable), then choose each one of the terminal nodes as the first ring to be cited in the notation; the spanning trees and elementary trees are formed for the starting points as given in rules C 11, C 12, and C 13.

C 11

Given a starting point in the reduced graph, form a spanning tree according to the following precedence rules:

(1) Choose that connected node in the reduced graph with the least degree, degree being defined as the number of edges connected to a node.

(2) Choose that node in the reduced graph corresponding to a ring in the original chemical graph satisfying C 6a.

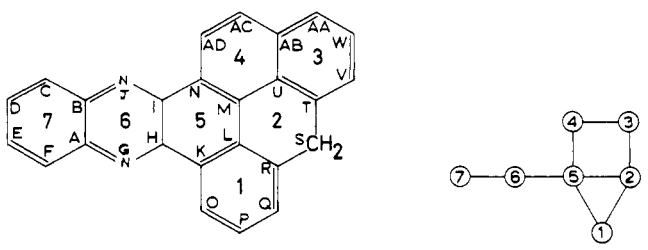
C 12

From each spanning tree, form a set of elementary trees, by deleting from it those nodes which constitute cycles in the reduced graph; for each such set of elementary trees form the tessellation–knit code for the original chemical graph and its mirror image. Use rule C 9 to choose a unique code.

C 13

If the reduced graph of the perifused ring system does not have any branch, or, in other words, the reduced graph itself consists of only circuits, then choose that node in the outermost circuit (or periphery), with a maximum degree as the starting point. If there are many such nodes of equal degree, then among them choose the one satisfying C 6a. Otherwise all the nodes on the periphery are valid starting points and hence all the corresponding codes are formed using C 11 and C 12 and a unique code is chosen using C 9.

Example 26



ALWIN : T<6ϕ7>A,H,K,T : R,M,AB,N=ϕGNϕJNϕSHJ

In example 26 the reduced graph has a branch 5–6–7, and hence 7 is chosen as the first ring (rule C 10). Its spanning tree is 7–6–5–1–2–3–4 (rule C 11), and the elementary trees are 7–6–5–1 and 2–3, while the connectivities are the edges 1–2 and 3–4. The tessellation code formed by clockwise labeling is AHMS while that by anticlockwise labeling is AHKT. Since the weighted sum of the latter (40) is smaller than that of the former (41), ALWIN is written as

T6666666A, H, K, T:R, M, AB, N = ϕGNϕJNϕSHJ

This is contracted to

T⟨6ϕ7⟩A, H, K, T:R, M, AB, N = ϕGNϕJNϕSHJ

Note: To distinguish between the locants having a single letter or more in cyclic structures having more than 23 nodes, a comma is placed between the locants in the tessellation–knit code.

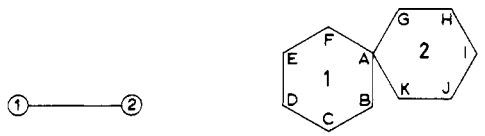
b. Spiro Rings. The information format for the spiro ring system is the same as that for the fused ring system. While forming the reduced graph of the spiro rings of the first kind, each sharing node is represented by an edge. For spiro rings of the II and III kind, however, since the edge relationships take precedence over the node relationships, only the former are represented as edges in the reduced graph.

Spiro Rings of the First Kind. If two rings are fused by a spiro atom, as in example 27, the knit code will be written as 66:AA, since the second ring can be thought of as a ring segment having A as the initial and final locant.

C 14

For the sake of indexing, it is necessary to identify a spiro ring; hence, in general, the knit code :$\alpha\alpha$ is denoted as αX, indicating the presence of a spiro atom X at locant α (X also refers to a carbon atom connected to four nonhydrogen atoms). This is considered as a part of the tessellation code.

Example 27



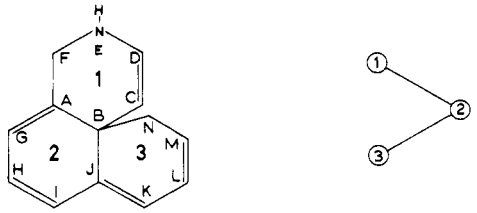
ALWIN : LT66AX=J

Spiro Rings of the Second Kind

C 15

If any spiro atoms present in the structure are not explicitly indicated in the tessellation code, then they are cited separately with their respective locants followed by X in zone X (zone of ring atom description).

Example 28



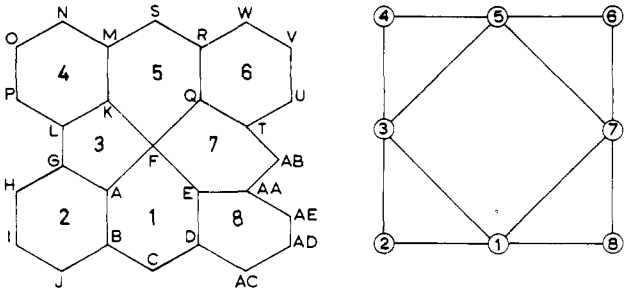
ALWIN : T666AJ=ϕBXϕEMJ

Spiro Rings of the Third Kind. The spiro rings of the third kind are encoded just as the spiro rings of the second kind. Here also, the spiro atoms are explicitly cited in the notation as defined in rule C 15.

Table II. Tessellation–Knit Code for Example 29

No.	Spanning tree	Sense of labeling	Tessellation locant sum
		Clockwise	
1	1-2-3-4-5-6-7-8	A,K,R:J,C,P,C,W,D,AB,E	30
2	1-8-7-6-5-4-3-2	A,K,R:F,G,F,N,E,U,D,AB	30
3	5-6-7-8-1-2-3-4	A,L,S:J,C,Q,C,AA,D,AB,E	32
4	5-4-3-2-1-8-7-6	A,K,Q:F,G,F,M,E,T,D,AA	29
5	7-8-1-2-3-4-5-6	A,L,R:J,C,Q,C,W,D,AB,E	31
6	7-6-5-4-3-2-1-8	A,K,R:F,G,F,N,E,T,D,AA	30
		Anticlockwise	
7	1-2-3-4-5-6-7-8	A,K,Q:F,G,F,M,E,T,D,AA	29
8	1-8-7-6-5-4-3-2	A,L,S:J,C,Q,C,AA,D,AB,E	32
9	5-6-7-8-1-2-3-4	A,K,R:F,G,F,N,E,U,D,AB	30
10	5-4-3-2-1-8-7-6	A,K,R:J,C,P,C,W,D,AB,E	30
11	7-8-1-2-3-4-5-6	A,K,R:F,G,F,N,E,T,D,AA	30
12	7-6-5-4-3-2-1-8	A,L,R:J,C,Q,C,W,D,AB,E	31

Example 29



ALWIN : LT665<6ϕ5>A,K,Q : F,G,F,M,E,T,D,AA = ϕFXJ

In example 29, among the nodes in the periphery of the reduced graph, the nodes 1, 3, 5, and 7 have the maximum degree 4. Since 3 is a five-membered ring, only the rings 1, 5, and 7 are valid starting points (rule C 13) and the tessellation codes are listed in Table II. It is evident that the codes formed with the choice of ring 1 corresponding to the spanning tree 1–2–3–4–5–6–7–8 as the first ring with an anticlockwise labeling, as well as the choice of ring 5 corresponding to the spanning tree 5–4–3–2–1–8–7–6 as the first ring with a clockwise labeling give the least sum. However, we note that both these codes are identical. (Example 29 illustrates the labeling corresponding to ring 1 as the first ring.)

POLY-EDGE SHARED RING SYSTEMS

If in a system of rings at least two rings share two or more edges, then it is called a poly-edge shared ring system. We further classify such systems into bridged ring systems and ring of rings systems.

a. Bridged Ring System. The bridged ring system comprises a main ring system containing one or more bridges (the connected edges which are commonly shared), each bridge being an acyclic structure.

The most complex ring is chosen as the main ring system (using rule C 19); the main ring system and the bridge are encoded separately. These are then cited in the information format as follows.

C 16a

Cite the initial locant from where the bridge starts (immediately after the delimiter colon—zone XI); then cite the ALWIN notation for the bridges. At the end, cite the locant of the main ring system to which this bridge is connected.

If there are two or more bridges, then the highest order locants are given to the bridgeheads according to the increasing order of complexity of the respective bridges.

C 16b

The acyclic structures are assigned the following order of increasing complexity:

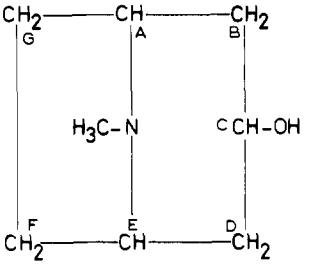
(1) Linear carbo-chain
(2) Linear hetero-chain
(3) Branched carbo-chain
(4) Branched hetero-chain

Whenever these structures belong to the same category, they are resolved by choosing the lower order symbols.

C 16c

Among the pair of bridge locants, the higher order locant is cited as the initial locant.

Example 30



ALWIN : T7T : ϕANA-ϕEJϕCQ

In example 30 there are three choices of the main ring system; the corresponding ALWIN are

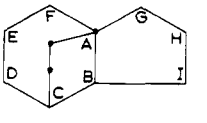
(i) T7T:ϕANA-ϕEJϕCQ
(ii) T6TN:ϕB2-ϕFJϕAAϕDQ
(iii) T5TN:ϕB1YQ1-ϕEJϕAA

Here i is chosen by rule C 19.

C 17

If a bridge starts from a spiro junction (or ends in a spiro junction), which means the corresponding bridgehead is a spiro atom, and if this spiro atom is not already explicitly cited either in the tessellation code (zone IV) or in the zone of ring atom description (zone X), then X is cited immediately after the bridgehead.

Example 31



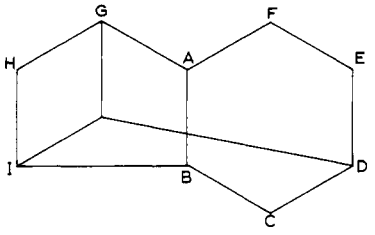
ALWIN : LT65 : ϕAX2-ϕCJ

Here the bridgehead is at locant A which is a spiro atom; since it is not explicitly cited earlier in the notation, we cite X after the initial bridge locant A.

C 18

If in a bridged ring system two or more bridges are interconnected, then these interconnections are cited appropriately within parentheses.

Example 32



ALWIN : LT65 : ϕDY(-ϕG) -ϕ1J

In example 32 the carbon atom in the bridge is connected to two other ring atoms at locants G and I. The connection to the atom at G is enclosed within parentheses.

b. Ring of Rings Systems. Type I System. A system comprising a ring system which is connected back to itself through one or several other substituent ring systems is called a type I ring of rings system. The ring system which is cited first in the notation of the ring of rings is called the main ring system. Among the ring systems which constitute the ring of rings system, the most complex ring system in C 19 is chosen as the main ring system.

C 19

The following is the increasing order of complexity among the ring systems:

(1) Benzene rings
(2) Carbo-monocycles
(3) Hetero-monocycles
(4) Carbo-bicyclic fused ring systems
(5) Hetero-bicyclic fused ring systems
(6) Carbo-polycyclic fused ring systems
(7) Hetero-polycyclic fused ring systems
(8) Carbo-perifused ring systems
(9) Hetero-perifused ring systems
(10) Carbo-spiro rings of the first kind
(11) Hetero-spiro rings of the first kind
(12) Carbo-spiro rings of the second kind
(13) Hetero-spiro rings of the second kind
(14) Carbo-spiro rings of the third kind.
(15) Hetero-spiro rings of the third kind
(16) Carbo-bridged ring systems
(17) Hetero-bridged ring systems
(18) Carbo-ring of rings systems—type I
(19) Hetero-ring of rings systems—type I
(20) Carbo-ring of rings systems—type II
(21) Hetero-ring of rings systems—type II.

If the choice of the main ring is still unresolved, then select that ring whose notation has the lowest symbol order.

C 20

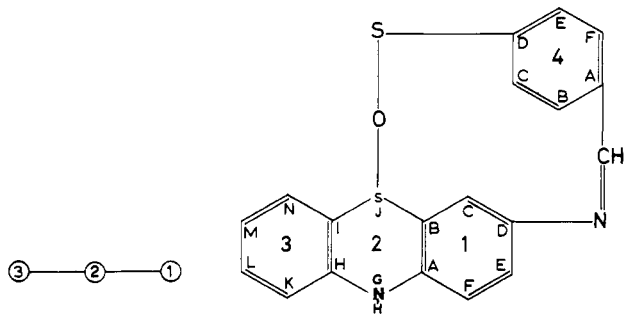
The interconnections between the different systems are denoted as follows.

Suppose the αth locant of a ring system is connected to the βth locant of a previously described ring system (say the nth cited system); this interconnection is cited as $\flat\alpha$-$\flat n\beta$.

If the αth locant of the ring is connected to the βth locant of the nth system through an acyclic chain, then its appropriate description in ALWIN is inserted between $\flat\alpha$ and $\flat n\beta$ (see example 33).

Here the rings 1, 2, 3 form a polycyclic fused ring system,

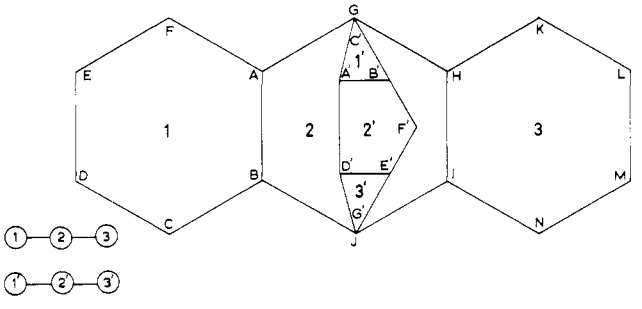
Example 33



ALWIN : &1T666AH = ϕGMϕJSJϕDNU@RϕDSO-ϕ1J

while 4 is a monocycle (benzene). Among the terminal rings 1 and 3 of the former, ring 3 is chosen as the main ring by rule C 10. Here the Dth locant of the second ring system is connected back to the Jth locant of the first ring system. Also note that the notation starts with the symbols &1 to indicate that this example corresponds to ring of rings of type I.

Example 34



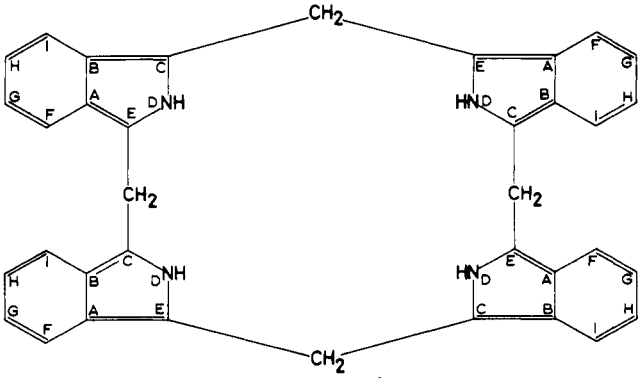
ALWIN : &1LT666AH = JϕGX-ϕCLT353AD = JϕGX-ϕ1J

In example 34, the main ring has the reduced graph 1–2–3 and the substituent ring has 1′–2′–3′ as its reduced graph.

Example 35

ALWIN:

&1T56ϕDMJϕC1—ϕET56ϕDMJϕC1—ϕET56ϕDMJ ϕC1—ϕET56ϕDMJϕC1—ϕ1E

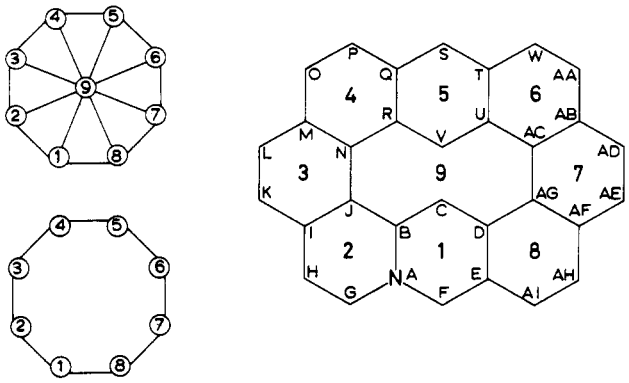


Contracted as ALWIN: &1<T56ϕDMJϕC1-*<ϕE-ϕ3>ϕ1E>

Type II Ring System. Another type of ring system is a fused ring of rings system, where a set of fused rings form another ring which shares two or more edges with the other rings. These are called type II ring of rings systems. The

rules for encoding perifused structures are still applicable here. Since there is a closure of the ring systems, the code will contain the multi-knit part in the notation.

Example 36



ALWIN: &2<6ϕ8>A,I,M,Q,T,AB;AF2E,D,AG=NJ

Since ring 9 shares more than a single edge, it is omitted from the reduced graph (as shown). Then the spanning tree is formed as 1–2–3–4–5–6–7–8 and the elementary tree 1–2–3–4–5–6–7 and chord 7–8. Note that the notation starts with the symbols &2, since this example corresponds to ring of rings system—type II. (cf. with the ALWIN of Figure 3a).

ACKNOWLEDGMENT

The authors are indebted to Professor G. N. Ramachandran, Mr. M. Ranganathan, Dr. S. K. Sen, and Dr. K. S. N. Iyer for valuable discussions during the course of this work. The authors are also grateful to W. J. Wiswesser for several valuable suggestions. Also they thank the referees and Dr. H. Skolnik, for suggesting several improvements in the manuscript.

Two of the authors (P. V. S. and S. K.) thank the National Institutes of Health, U. S. A. (Grant No. 15964), and N.C.E.R.T. (India) for providing research fellowships.

LITERATURE CITED

(1) National Academy of Sciences and National Research Council (U. S.), "Survey of Chemical Notation Systems," Publication No. 1150, Washington, D. C., 1964.

(2) Rules for I.U.P.A.C. Notation for Organic Compounds, Longmans, London, 1961.

(3) Smith, E. G., "The Wiswesser Line Formula Chemical Notation," McGraw-Hill, New York, N. Y., 1968.

(4) Korfhage, R. R., "Logic and Algorithms," Wiley, New York, N. Y., 1966.

(5) Knuth, D. E., "Art of Computer Programming," Vol. 1, "Fundamental Algorithms," Addison-Wesley, Reading, Mass, 1968.

(6) Berge, C. E., "The Theory of Graphs and Its Application," Methuen, London, 1962.

(7) Sankar, P. V., "ALWIN—Algorithmic Wiswesser Notation for Organic Compounds," Ph.D. Thesis, Indian Institute of Science, Bangalore, 1973.

(8) Sankar, P. V., Krishnamurthy, E. V., and Krishnan, S., "Representation of Stereoisomers in ALWIN," *J. Chem. Doc.*, **14**, 141 (1974).

Representation of Stereoisomers in ALWIN†

P. V. SANKAR,** E. V. KRISHNAMURTHY,* and S. KRISHNAN

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India

Received May 22, 1973

The well-known Cahn–Ingold–Prelog method of specifying the stereoisomers is introduced within the framework of ALWIN—Algorithmic Wiswesser Notation. Given the structural diagram, the structural ALWIN is first formed; the specification symbols are then introduced at the appropriate places to describe the stereoisomers.

This paper describes the application of the Cahn, Ingold, and Prelog[1-4] method (CIP) for the representation of stereoisomers in ALWIN. It is well known that the stereoisomers have the same molecular and structural formula, but differ only in configuration or in spatial arrangement of their groups. Since the structural formula (or equivalently the connectivity table) remains identical for these stereoisomers, the line notations, such as WLN or ALWIN[5], remain identical. Therefore, if one wants to distinguish these stereoisomers, it is necessary to introduce stereochemical conventions in the form of additional punctuations in ALWIN to describe the spatial orientation or the configuration of the molecule. Stereoisomers are termed cis–trans isomers when they differ only in the position of the atoms relative to a specified plane where these atoms are parts of a rigid structure (we restrict our definition only to cis–trans isomerism about double bonds in general and to certain achiral cyclic structures whose ring skeleton approximates to a plane); these isomers are also known as geometrical isomers.

These stereoisomers which differ only in configuration are known as optical isomers because of their special property (called optical activity) of rotating the plane of polarized light. In such a case, the molecule lacks elements of symmetry, so that it is not identical with and hence not superimposable on its mirror image. Such molecules are also called chiral. Chirality of a molecule may arise because of

† Contribution No. 47, Molecular Biophysics Unit, Indian Institute of Science, Bangalore, India.

* Author to whom correspondence should be addressed.

** Presently at Tata Institute of Fundamental Research, Bombay 400005.