small D values). One can also see that a single difference in value of path numbers by two introduces the same dissimilarity as four changes (increase of decrease) by a value of 1. The *metrics*, however, require that D rather than D^2 be used, since D satisfies the triangular rule: $D_{ab} \geq D_{ac} + D_{cd}$, a rule not necessarily obeyed by D^2.

(19) T. K. Devon and A. I. Scott, "Handbook of Naturally Occurring Compounds", Vol. II, "Terpenes", Academic Press, New York, 1972, pp 5-6.
(20) P. C. Jurs and T. L. Isenhour, "Chemical Applications of Pattern Recognition", Wiley, New York, 1975.

A Linked-Path Connection Table with Substructural Atom-Ordering

R. GEOFF. DROMEY*
Research School of Chemistry, Australian National University, Canberra, A.C.T. 2600, Australia

Received June 15, 1977

A new connection table formalism is introduced. The procedure for producing a canonical atom-ordering preserves molecular structural connectivity in a more explicit and direct way than existing systems. Consequently, substructure searching is made a much simpler process. The method focusses on atom connectivity and so judgments about bond characterization are avoided.

INTRODUCTION

Connection tables[1,2] achieve probably the simplest and most explicit topological description of molecular structure. They can be derived for all structures that are described in terms of specific atoms and bonds. In most connection table representations some form of atom-ordering procedure[3] is used to obtain a unique representation for each molecule. These simple algorithmic atom-ordering procedures invariably tend to breakup and inherent molecular "connectedness" that is conveyed so explicitly in the two-dimensional structure diagrams; that is, connectedness (in an atom adjacency sense) is sacrificed for a canonical representation. This means that substructure searching must always be on a strictly atom-by-atom basis. This complicates substructure searching of these systems.

In contrast, linear notations[4-7] tend to preserve the relationship between atoms and functionalities and so they make substructures much easier to identify. Consequently advanced pattern matching techniques that avoid the need to examine all characters in a representation may be employed.[8-10] Unfortunately, their rules for encoding have been far too complex to be used as a basis for canonicalization of connection tables. The valence-oriented rules used in the tree-structured linear notation described previously[7] depart from the formalism of earlier systems. They are sufficiently simple and algorithmic in nature to be used as a basis for producing a canonical connection table representation that to a large degree preserves molecular connectedness in much the same way as linear notations. The reason for adopting such an approach has been to produce a connection table representation that is usually more amenable to advanced pattern matching searches (in a way similar to line notations) than existing systems. It is also suitable for rapid visual interpretation. The unique numbering of atoms in the linear notation can provide the framework for the corresponding connection table ordering. In fact, on an atom-by-atom basis the two representations can be made directly compatible.

The four fundamental atom-ordering rules for the tree-structured linear notation are as follows.[7]

GENERAL RULES OF PRECEDENCE

At each stage in encoding a structure always choose to encode first

(A) that connected path with an atom of smallest valence attached earliest;
(B) where minimum valence does not resolve the path, choose to encode first the path with the atom of smallest atomic weight attached earliest;
(C) if resolution still has not been made encode along the path that contains an atom with the least number of atoms attached earliest;
(D) finally, if none of the other constraints resolves the path choose to encode along the path that has an atom with the least number of hydrogens attached earliest.

This set of rules is applied first in the sequence (A) to (D) wherever precedence must decide which atom or ring is to be encoded next.

A discussion of how the tree-structured formalism is used to obtain a unique atom numbering for a molecular structure is described in detail in the paper on the linear notation[7] and so will not be pursued further here. The present work will focus on the actual connection table representation assuming that a unique atom-numbering has been derived previously by applying the tree-structured precedence rules.

CONVENTIONAL CONNECTION TABLES

In order to allow for an adequate comparison to be made between the newly proposed formalism and existing systems, the conventions for the latter will be briefly reviewed. The most common set of descriptors used in connection table representations are element descriptor, "atom-connected-to", and order of bond forming the connection. Atoms are usually represented by their conventional atomic symbols while bonds are designated 1, 2, or 3 depending upon whether they are single, double, or triple bonds. To avoid citing each bond twice, only connections to lower numbered atoms are encoded. For ring systems this requires that ring-closure rank must be separately specified by taking into account atom numbers and bond codes and doing an ascending order sort.[3]

The conventional connection table representations for an acyclic and a ring compound are given in Tables Ia and IIa. Atom-ordering has been derived using Morgan's algorithm. The accompanying linked-path connection table representations for these structures will be compared and discussed in detail in the next section. Although structural features are represented explicitly in these conventional connection tables (Tables Ia and IIa), they suffer from a scattering effect imposed by the network formalism. The result is that a search for even simple functional groups becomes a tedious multistep procedure.[11]

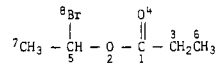
* Address correspondence to author at Department of Computing Science, University of Wollongong, Wollongong, N.S.W. 2500, Australia.

0095-2338/79/1619-0037$01.00/0

Table I

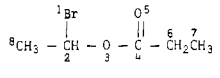
CONVENTIONAL CONNECTION TABLE				LINKED-PATH CONNECTION TABLE				
(1a) Atom No.	Atom	Connection	Bond	(1b) Atom No.	Atom	Connection Transfer	Atoms Attached	
1	C	-	-	8	C	6	4	Branch
2	O	1	1	7	C	∅	4	Branch
3	C	1	1	6	C	2	4	
4	O	1	2	5	O	∅	1	Chain
5	C	2	1	4	C	1	3	
6	C	3	1	3	O	1	2	
7	C	5	1	2	C	1	4	
8	Br	5	1	1	Br	∅	1	

Morgan's algorithm numbering



1

Linked-Path atom numbering

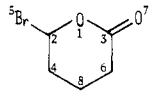


1

Table II

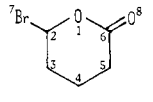
CONVENTIONAL CONNECTION TABLE				LINKED PATH CONNECTION TABLE				
Atom No.	Atom	Connection	Bond	Atom No.	Atom	Connection Transfer	Atoms Attached	
1	O	-	-	8	O	2	1	Substituent 2
2	C	1	1	7	Br	5	1	Substituent 1
3	C	1	1	6	C	5	3	
4	C	2	1	5	C	1	4	
5	Br	2	1	4	C	1	4	Ring
6	C	3	1	3	C	1	4	
7	O	3	2	2	C	1	4	
8	C	4	1	1	O	1	2	
	Ring-Closure	6-8	1					

Morgan's algorithm numbering



2

Linked-Path atom numbering



2

LINKED-PATH CONNECTION TABLE

The proposed linked-path connection table represents an attempt to construct a formalism that to a large extent preserves molecular connectivity in a more direct manner than existing systems. As a consequence substructure searching is simplified considerably in that advanced pattern matching algorithms may be employed in many instances.[8-10] The method focusses on atom connectivity as opposed to bond interconnections. This eliminates the need for bond description (e.g., single, double, aromatic, etc.). Ring closures are generally not needed and connections are specified differentially to minimize variations in the representation of substructures. These connections are specified as connection transfers of the current linked-path. They will be defined in detail in the next section. In addition, for each atom in the table a parameter is specified indicating the number of atoms directly bonded to that atom. Cyclic structures containing atoms with more than four bonds (e.g., boron compounds) and some bridging structures need ring closures.

Because the numbering procedure for atoms in a structure is based on a tree-like algorithm, the atoms deepest (latest numbered) in the tree are specified first. Connections are, in general, only made to atoms at shallower levels of the tree. Connection transfers (CT's) together with a set of conventions for establishing the bond relationship (if any) to immediate neighbors in the table are used to map out the confines of individual linked-paths and to define interconnections between linked-paths.

The crucial step in deriving the linked-path connection table description of a molecular structure (that is, presuming the atom numbering has been previously made) is the assignment of the connection transfer column of the table. In general the following rule applies.

Rule for Connection Transfer Assignment: At each stage in encoding where a connection transfer of the linked-path is necessary, it must be made to the earliest numbered atom that is directly attached to the atom currently being entered into the table. The latest numbered atom must be entered into the table first.

It follows that

connection transfer* = number of current atom − number of earliest atom to which current atom is attached

*See convention (c).

There are several conventions for making connection transfer assignments. The first of these requires the definition of a terminal atom. A *terminal atom* is an atom bonded to only one other nonhydrogen atom.

(a) Connection Transfers Equal to Zero. Terminal atoms at the end of a linked-path (a consecutively numbered subchain or branch containing more than one atom) are assigned a connection transfer of zero (∅). This by convention is taken to mean that the corresponding atom in the table is connected to the atom immediately preceding it *but not* to the atom immediately following it. For example, in Table Ib atom 5 is connected to atom 4 but not to atom 6. Atom 8 in Table Ib is also a terminal atom, but since it is not at the end of a linked-path (as defined by the atom numbering), it is not assigned a connection transfer of zero.

(b) Connection Transfers Equal to One. A connection transfer of one (1) indicates that the associated atom is directly bonded to the atoms immediately preceding and following it in the table. In Table Ib atom 3 is taken to be connected to atoms 2 and 4.

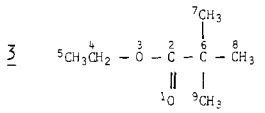
(c) Connection Tranfers Greater than One. A connection transfer of N ($N > 1$) means that the current atom is connected to an atom encoded N atoms before the current atom. Atom 6 in Table Ib has a CT of 2 which means that it is connected to atom 4 (6 − 2). A connection transfer greater than 1 does not contain any information as to how the atom is related to atoms immediately preceding and following it (except when equal to the current atom number; see below) in the table. That is, a CT of 2 for atom 6 in Table Ib does not indicate how atom 6 is related to atoms 5 and 7. However, the CT's for atoms 7 and 5 indicate that atom 7 is bonded to atom 6 but atom 5 is not bonded to atom 6. The differential character of connection transfers tends to standardize the representation for a given substructure irrespective of where it occurs in the table. For example, the methyl groups in $-C(CH_3)_3$ of Table III will always be encoded as

element	CT	atoms attached
C	3	4
C	2	4
C	0	4

if none of them begins the numbering for the structure.

Table III

Atom No.	Element	C.T.	Atoms Attached	
9	C	3	4	Branch
8	C	2	4	Branch
7	C	ø	4	Branch
6	C	4	4	
5	C	ø	4	Chain
4	C	1	4	
3	O	1	2	
2	C	1	3	
1	O	ø	1	



(d) Connection Transfers Equal to the Current Atom Number. A connection transfer equal to the current atom number is used to indicate that the current atom is not connected to lower numbered atoms, but it is connected to the atom immediately above. In effect it is not a connection transfer. This situation is encountered quite often in interconnected ring systems. Atom 15 in Table V is such an example. It is connected to atoms 16 and 20 but not to any lower numbered atoms.

(e) Connection Transfers Greater than the Current Atom Number. A connection transfer greater than the current atom number is defined as a secondary connection transfer. It is used only rarely in complex bridge structures where the primary connection transfers are insufficient to establish all the atom interconnections. This situation is encountered in Tables IX to XII. There are two other descriptors that must be considered for the linked-path representation. They are the element descriptor and the "atoms-attached" count associated with each atom. The latter is just a numeral associated with each atom which indicates the number of atoms (hydrogens included) attached to that atom. For the element descriptor two choices are possible; the conventional atom symbols can be used, or alternatively the 4-bit valence-atomic weight formalism of the tree-structured linear notation can be employed. The former approach is used here for convenience.

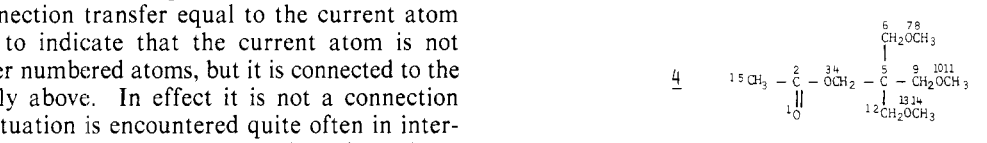
The application of the rules and conventions relating to connection transfers is best seen by example. Several acyclic, ring, and bridging examples will now be discussed.

REPRESENTING STRUCTURES WITH THE LINKED-PATH CONNECTION TABLE

(a) Acyclic Structures. Establishing the connection transfers for all acyclics is straightforward. In Table Ib, for example, atom 8 has a connection transfer of 6 to atom 2. Atom 7 is at the upper end of a linked-path ($-C^6H_2C^7H_3$) and so is assigned a connection transfer of zero. Atom 6 is connected to atom 4 via a connection tranfer of 2. Atom 5 is at the end of a chain. The other atoms are part of the chain. *By convention atom 1 in all purely acyclic compounds is assigned a connection transfer of zero to differentiate acyclics from ring-containing compounds.* The zero CT for atom 1 is not meant to imply that atoms 1 and 2 are unconnected. Two other acyclic examples are given in Tables III and IV. This convention is not essential to the connection table description and may be discarded if absolute consistency is demanded for the representation. The motivation for adopting the convention has been to avoid unnecessary searches.

Table IV

Atom No.	Element	C.T.	Atoms Attached	
15	C	13	4	Branch
14	C	ø	4	Branch
13	O	1	2	
12	C	7	4	
11	C	ø	4	Branch
10	O	1	2	
9	C	4	4	
8	C	ø	4	Chain
7	O	1	2	
6	C	1	4	
5	C	1	4	
4	C	1	4	
3	O	1	2	
2	C	1	3	
1	O	ø	1	



(b) Cyclic Structures. In the linked-path formalism (and the tree-structured linear notation) ring paths take precedence over substituents. This is clearly seen in Tables IIb and V. Substituents for a ring are specified only after that ring has been completely specified. For the example in Table IIb the encoding is begun at atom 8, the atom deepest in the tree. Atom 8 is connected to atom 6 via CT of 2. A CT of 5 connects atom 7 to atom 2. The double bond between atoms 6 and 8 is reflected in the corresponding "atoms-attached" counts. Atom 7 and 8 cannot be connected because they both involve connection transfers greater than one. The linked-path for the ring starts at atom 6. Atom 6 is connected to atoms 1 and 5. The encoding rule determines that a CT of 5 to atom 1 is made. Atom 6 is seen to be connected to atom 5 because the latter has a CT of one. *Atom 1 for ring systems is always assigned a CT of 1* (except for complex bridging systems) *to make the distinction with acyclic systems.* The 1 indicates that atom 1 is connected to atom 2 but to no lower numbered atom (convention d). In comparing the conventional table (Table IIa) with Table IIb, it can be seen that the ring and substituents are clearly distinguished in the latter system. A ring closure term is also avoided in Table IIb. Some slightly more complex structures are given in Tables V, VI, and VIII. In Table V once again the rings, substituents, and ring interconnections are all clearly separated. Atom 15 is not connected to an earlier numbered atom and so it is assigned a connection transfer equal to its atom number (convention d). Table VII demonstrates that spiro systems do not cause any encoding problems.

In fused systems the structure is built up ring by ring. Table VIII is a typical fused structure. Each new fused ring to be added to the existing set is added as a linked set of ring atoms with connection transfers at each end of the chain. This makes the identification of each ring segment straightforward.

(c) Bridging Structures. Compounds with bridges represent an increase in topological complexity over other ring systems. They are handled in essentially the same way as other structures except that when primary connection transfers do not completely define the structure, secondary connection transfers must be introduced.

Table V

Atom No.	Element	C.T.	Atoms Attached
21	O	6	2} Substituent
20	C	5	3
19	C	1	3
18	C	1	3
17	C	1	3 } Ring 2
16	C	5	3
15	C	15	3
14	C	Ø	4
13	C	2	4 } Ring Interconnection
12	O	Ø	2
11	C	7	4
10	C	Ø	4
9	O	7	2 } Substituents for Ring 1
8	C	Ø	4
7	O	6	2
6	C	5	3
5	C	1	3
4	C	1	3
3	C	1	3 } Ring 1
2	C	1	3
1	C	1	3

5

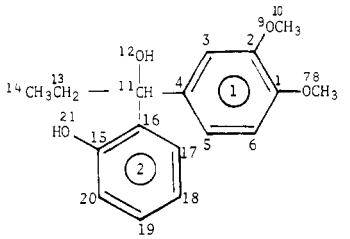
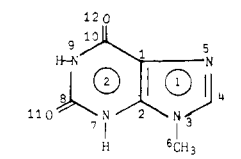


Table VI

Atom No.	Element	C.T.	Atoms Attached
12	O	4	1} Substituent
11	O	3	1} Substituent
10	C	9	3
9	N	1	3
8	C	1	3 } Ring 2
7	N	5	3
6	C	3	4} Substituent
5	N	4	3
4	C	1	3
3	N	1	2 } Ring 1
2	C	1	3
1	C	1	3

6



A primary connection transfer is always less in magnitude than the atom number with which it is associated because it

Table VII

Atom No.	Element	C.T.	Atoms Attached
10	C	9	4
9	C	1	4
8	C	1	4
7	C	1	4
6	C	5	4
5	C	4	4
4	C	1	4
3	C	1	4
2	C	1	4
1	C	1	4

7

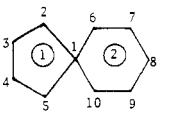
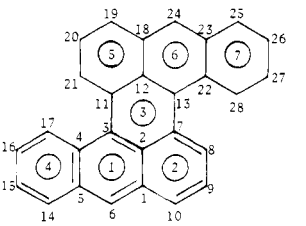


Table VIII

Atom No.	Element	C.T.	Atoms Attached
28	C	6	4
27	C	1	4 } Ring 7
26	C	1	4
25	C	2	4
24	C	6	4
23	C	1	4 } Ring 6
22	C	9	4
21	C	10	4
20	C	1	4 } Ring 5
19	C	1	4
18	C	6	4
17	C	13	3
16	C	1	3 } Ring 4
15	C	1	3
14	C	9	3
13	C	6	4
12	C	1	4 } Ring 3
11	C	8	4
10	C	9	3
9	C	1	3 } Ring 2
8	C	1	3
7	C	5	3
6	C	5	3
5	C	1	3
4	C	1	3 } Ring 1
3	C	1	3
2	C	1	3
1	C	1	3

8



represents a transfer of the linked-path to an earlier numbered atom. Secondary connection transfers indicate links to atoms occurring at later numbers in the table. They are always greater in magnitude than the atom number with which they are associated. For example, if it is necessary to represent a secondary connection transfer from atom 8 to atom 12, then the number 12 is placed in the CT slot for atom 8 (i.e., the secondary CT is equal to the atom number to which the connection must be made).

In general the following algorithm must be employed when encoding bridging structures.

Table IX

Atom No.	Element	C.T.1	C.T.2[†]	Atoms Attached	
12	C	4	4	4	Bridge
11	O	7*	11	2	Bridge
10	C	9	9	3	Ring
9	C	1	1	3	Ring
8	C	1	1	3	Ring
7	C	5	5	3	Ring
6	C	5	5	3	Ring
5	C	1	1	3	Ring
4	C	1*	11	3	Ring
3	C	1	1	3	Ring
2	C	1	1	3	Ring
1	C	1	1	3	Ring

9

[†] The highest numbered C.T. column is taken as the final representation. Earlier numbered C.T. columns are not part of the representation.

* The connection transfers that must be changed at each stage of the secondary phase are asterisked.

Table X

Atom No.	Element	C.T.1	C.T.2	C.T.3[†]	Atoms Attached	
12	C	7	7	7	4	
11	C	1	1	1	4	Ring 3
10	O	2	2	2	2	Ring 3
9	C	8	8	8	4	
8	C	1	1	1	4	Ring 2
7	C	1*	12	12	4	Ring 2
6	C	4	4*	6	4	Ring 2
5	C	4	4	4	4	
4	C	1	1	1	4	
3	O	1	1	1	2	Ring 1
2	C	1	1*	6	4	Ring 1
1	C	1	1	1	4	Ring 1

10

Table XI

Atom No.	Element	C.T.1	C.T.2	C.T.3[†]	Atoms Attached
13	C	7	7	7	4
12	C	8	8	Ø	4
11	C	4*	13	13	4
10	C	9	9	9	4
9	C	1	1	1	4
8	C	1	1	1	4
7	C	5*	11	11	4
6	C	5	5	5	4
5	C	1	1	1	4
4	C	1	1	12	4
3	O	1	1	1	2
2	C	1*	7	7	4
1	C	1	1	1	4

11

Algorithm for Handling Bridging Compounds:

Step 1 First make all connection transfers according to the basic encoding rule (i.e., make the largest magnitude transfers for each atom).

Step 2 Check if all the necessary ring and bridging interconnections have been made. The procedure usually stops at this stage. In the event that some connections have not been established go to step 3.

Step 3 Find the latest numbered atom L for which connections have not been completed.

Step 4 Test the CT of the latest earlier numbered atom E whose attachment to L has not been defined. If the CT of E is less than the atom number value for E, then go to step 5, else test the CT of L itself. If the latter is less than L make the CT change to L that establishes the connection to E and also make any other accompanying changes. Otherwise, add a special connection transfer at the end of the table below atom 1. Then go back to step 2.

Step 5 Establish the connection between E and L by replacing the CT for E with the secondary CT for connection to L (i.e., with the atom-order number for L).

Step 6 Check if all the connections to E are still intact. If so then find the next lastest L and repeat the whole procedure by going to step 2, else treat E as though it were L and go to step 4. Repeat the whole procedure until there is termination as step 2.

The need for this procedure will become apparent in the examples which follow.

Examination of Table IX shows that after all primary CT's have been made (CT1) there is no implied connection between atoms 11 and 12. The secondary transfer algorithm implies that CT for atom 11 must be changed to remedy the situation. Giving atom 11 a CT of 11 connects it to atom 12 but disconnects it from atom 4. The induced disconnection is corrected by assigning atom 4 a secondary connection transfer of 11. The final CT representation is given by CT2. Some slightly more complex examples are given in Tables X–XII.

In Table XII it is necessary to introduce an additional connection transfer at the bottom of the table to establish the connection between atom 6 and atom 5. Such instances are uncommon. The first four atoms are assumed connected because atom 1 has a connection transfer of 4 which encloses the first four atoms.

FAST SEARCHING OF LINEARIZED CHEMICAL STRUCTURE PATTERNS

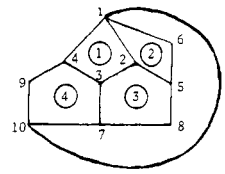
It was mentioned earlier that linearized molecular structure representations (e.g., line formula notations and the linked-path connection table) were more suitable than conventional connection table representations for fast pattern searches.

The following example of searching the WLN notation for the pyridine group (T6NJ) illustrates how these techniques

Table XII

Atom No.	Element	C.T.1	C.T.2	C.T.3	C.T.4[†]	Atoms Attached
10	C	4	4	4	4	4
9	C	5	5*	9	9	4
8	C	3	3	3*	Ø	4
7	C	4*	10	10	1Ø	4
6	C	5	5	5	5	4
5	C	3	3	3*	8	4
4	C	3	3*	9	9	4
3	C	1*	7	7	7	4
2	C	1	1	1*	5	4
1	C	1	1*	4	4	4
Ø	6	–	–	–	1	–

12



are applied to searching linear patterns such as the linked-path connection table

```
   1   2   3  4765
T66 BNJ EQ D- CT6NJ EVQ  {pattern notation to be searched}
T6NJ  {pattern sought}
```

The numbers illustrate where successive characters are examined. The pattern T6NJ is 4 characters long so the first position it could occur would be such that the character J was at position 4 [probe (1)]. Examining the 4th character first we find a space which is not even in the pattern we are seeking. We can therefore skip and examine the 8th character next. It is also a space and so we can skip to the 12th character. There we encounter a D which again is not present in T6NJ and so we can skip to the 16th character where we encounter a T which *is* in the pattern we are looking for. In this case we make a jump of only three to the position where we expect the position of the last character relative to T. The principle of this strategy is always to jump as far as possible without risk of bypassing a possible match. In this search to find the pattern we have needed to examine only 7 rather than 19 characters. A detailed discussion of the use of this algorithm for text searching is described elsewhere.[8,9] Its adaptation to the linked-path connection table formalism for substructure searches is straightforward.

CONCLUSIONS

The linked-path connection table has several favorable attributes. It tend to hold atoms and functionalities together in a way that is much more convenient than existing systems. Bond characterization and ring closures are generally avoided. In addition, because the connection table representation parallels the tree-structured linear notation, automated bidirectional conversion between the two systems should be realizable. The atom ordering formalism is based on very fundamental chemical precepts and consequently it maps out structures in a way that is quite natural for the chemist to work with. The less difficult task of going from the present connection table representation to other connection tables should also be possible although the task has not been fully evaluated.

To automate the derivation of the linked-path connection table, it would be necessary to set up a table that indicated all the atoms attached to a given atom. The following properties would also need to be listed for each atom: valence, atomic weight, atoms attached, hydrogens attached, and nonhydrogens attached (identifies terminal atoms).

ACKNOWLEDGMENT

The author would like to thank Dr. J. K. MacLeod for discussions on the manuscript. Greta Pribyl and Carol Jacobs are also to be thanked for their part in preparing the manuscript and tables.

REFERENCES AND NOTES

(1) D. J. Gluck, "A Chemical Structure Storage and Search System Developed at Du Pont", *J. Chem. Doc.*, **1**, 43–51 (1965).
(2) L. C. Ray and R. A. Kirsch, "Finding Chemical Records by Digital Computers", *Science*, **126**, 814–819 (1957).
(3) H. L. Morgan, "The Generation of a Unique Machine Description for Chemical Structures—A Technique Developed at Chemical Abstracts Service", *J. Chem. Doc.*, **5**, 107 (1965).
(4) E. J. Smith, "W. J. Wiswesser's Line Formula Chemical Notation", McGraw-Hill, New York, N.Y., 1968.
(5) H. W. Hayward, Patent Office Research and Development Report, No. 21, Patent Office, Washington, D.C., 1964.
(6) G. M. Dyson, M. F. Lynch, and H. L. Morgan, *Inf. Storage Retr.*, **4**, 27 (1968).
(7) R. G. Dromey, "A Simple Tree-Structured Line-Formula Notation for Representing Molecular Topology", *J. Chem. Inf. Comput. Sci.*, **18**, 225 (1978).
(8) R. G. Dromey, "QUICKSEARCH—An Algorithm for Fast Keyword Searching in Text", *Commun. ACM*, submitted for publication.
(9) R. S. Boyer and J. R. Moore, "A Fast String Searching Algorithm", *Commun. ACM*, **20**, No. 10, 752–772 (1977).
(10) M. C. Harrison, "Implementation of the Substring Test by Hashing", *Commun. ACM*, **14**, No. 12, 777–779 (1971).
(11) D. Gould, E. B. Gasser, and J. F. Ryan, "A Chemical Oriented Information Storage and Retrieval System", *J. Chem. Doc.*, **5**, 24 (1965).