it is the intention to create and hold a separate display file using a chemical typewriter for input.

The effectiveness of the matrix record for generating fragment codes and structural display has been established. The consistent choice of symbols and use of logical rules in the Wiswesser notation provide an excellent beginning to a computer program. This notation compacts the chemical and topological data required to represent structures in a way which makes it highly effective in information handling.

LITERATURE CITED

(1) Smith, E. G., *et al.*, "Revised Wiswesser Notation," McGraw-Hill, in press.
(2) Feldman, A., D. B. Holland, D. P. Jacobus, "Automatic Encoding of Chemical Structures," J. CHEM. DOC. **3**, 187, (1963).
(3) Mullen, J. M., "Atom-by-atom Typewriter Input for Computerized Storage and Retrieval of Chemical Structures," *Ibid.*, **7**, 88 (1967).

Documentation of Chemical Reactions by Computer Analysis of Structural Changes*

J. E. ARMITAGE, J. E. CROWE, P. N. EVANS, M. F. LYNCH, and J. A. McGUIRK
Postgraduate School of Librarianship and Information Science, University of Sheffield, Sheffield 10, England

Received May 19, 1967

A method for detection of structural similarities among chemical compounds by computer is described. The method involves an iterative process whereby fragments (of any complexity) common to a pair of structures are generated from their topological descriptions, starting with the atoms common to both and increasing the fragment size one atom at a time until the largest connected set of atoms and bonds common to the pair of structures is determined. Computer programs to perform such an analysis on pairs of acyclic structures are described, and application of the method for identification of structural changes in the reactions of acyclic compounds is discussed.

The provision of easy and adequate access to information on chemical reactions is of fundamental importance to the advancement of chemistry. This area of chemical documentation, however, continues to pose a considerable problem that has not yet been solved even by the application of computers to handling chemical structural information. While the development of systematic nomenclatures and notations, and, more recently, of computer algorithms, has made it possible to identify individual chemical compounds uniquely, no comparable success has been attained with reaction data.

The device most widely used in representing chemical reactions is undoubtedly the reaction scheme or equation, in which the reactants are displayed on one side of the equation, and the products on the other. Such a scheme allows the chemist to deduce the nature of the changes which the molecules undergo, and to see the structural factors which influence the changes. However, the organization of this data, either manually or by computer, is far from simple (*1*). Access through the structures of the compounds which take part in a particular reaction is relatively ineffective, because it gives little indication of the type of reaction involved, nor has any uniform nomenclature of reaction types been developed which is widely accepted and used. Indeed, it is revealing that one of the more effective ways of indexing reactions is by trivial name (*2*)—i.e., by the name of the chemist associated with the discovery of the reaction. The terms "Beckmann rearrangement," "Claisen reaction," "Clemmensen reduction," and "Diels-Alder reaction" are meaningful concepts to every chemist. Descriptive word indexing is less useful, because a simple reaction can be described in many different ways. Thus the rearrangement of cyclohexanone oxime into ε-Caprolactam (Figure 1) can be described variously as an oxime rearrangement, amide formation, lactam formation, or ring enlargement, yet none of these specifies the process uniquely, nor is as adequate as the description "Beckmann rearrangement of cyclohexanone oxime," where the term "Beckmann rearrangement" signifies a pattern of bond rearrangements which is common to the reactions of a large number of compounds.

It seemed imperative that an approach to the documentation of chemical reactions using computer

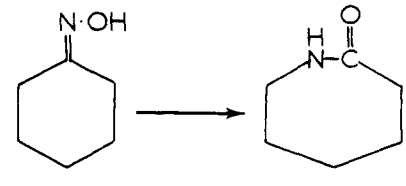


Figure 1.

*Presented before the Division of Chemical Literature, 153rd Meeting, ACS, Miami Beach, Fla., April 1967.

methods should represent the reaction equation, and that the computer should be used to derive the site and nature of the changes by analysis of the structures of reactants and products. It should also be used to store the information derived in this way to provide access by reaction type as well as through the structures or substructures of the individual molecules. Moreover, an approach such as this demands little additional effort in chemical information systems in which the structures of compounds are already stored in full detail, as it requires only that the compounds which take part in a particular reaction be identified as reactants or products.

The analysis of structural changes is a procedure for which computer techniques have not previously been developed. Existing programs are concerned primarily with the search for identity (*19*), or with substructure search (*3, 4*). While it is true that the substructure search method would be sufficient in most cases in which either the product or the reactant is wholly contained within the other, the cases to which this applies are infrequent. Rather, a novel approach has been adopted, which simulates some of the mental processes which a chemist employs when he examines a reaction diagram and deduces the nature of the changes taking place. The chemist, in scanning an equation, identifies the common features on either side of the diagram as a preliminary to pinpointing the differences, and identifying the site and nature of the reaction. Thus a search for similarities is the first step in the search for differences.

The comparison of molecules related by reaction is, of course, only one instance of a more general procedure in which numbers of structures, related perhaps by common nonstructural attributes, are scanned, and their common features identified. The advancement of structural chemistry, indeed, is based largely on this procedure, for it is by this means that correlations between structure and various forms of activity are deduced.

A project to develop computer methods to determine similarities among sets of chemical structures and to apply these techniques to a variety of problems related to the processing of information on chemical structures is at present in progress at the Postgraduate School of Librarianship and Information Science at Sheffield University. The project is supported by a grant from the Office of Scientific and Technical Information, London. This progress report describes the early stages of development of these techniques, and discusses their application to the documentation of chemical reactions.

DOCUMENTATION OF CHEMICAL REACTIONS BY COMPUTER

Vleduts (*5*) has considered the problem of manual and mechanized organization of chemical reaction data at length, and, indeed, has posed the problem of automatic comparison of the structures of initial and final molecules. The objective of this comparison is the identification of the parts of the molecules which remain unchanged, and, therefrom, the determination of the site and nature of the structural changes. The changes would then be described as a skeleton reaction scheme, in which the bonds formed or broken, and their immediate environment, would be represented and stored for search in a variety of ways. More recently, Vleduts has described a computer program which derives "empirical formulas of bonds," or atom-bond-atom pairs, for reactants and products, and has evaluated the use of differences in these formulas as an index of reaction types (*6*). There are many types of chemical reactions, particularly those involving rearrangements of multiple bonds, in which no difference can be detected in the sum of the types of pairs present in the structures before and after reaction.

Dyson (*7*) has outlined a classification scheme for organic reactions in which a hierarchical code, which takes account of reaction conditions, is derived manually for search by machine.

A technique for representing chemical reactions by superimposing the structures of reactants and products when recorded on a grid for input to a computer by optical scanning has been described by Meyer (*8*), while approaches using notations and fragmentation codes have been variously described by Gelberg (*9*), Fugmann (*10*), Vleduts and Mishchenko (*11*), and Shevyakova and Stoyanovich (*12*).

Two of us (*13*) have recently described a general algorithm which extends the range of computer manipulations of chemical graphs to include the automatic detection of similarities among chemical structures. Similarity has been defined for this purpose as the largest connected set of atoms and bonds common to a pair of compounds—i.e., the maximum overlap of the graphs. The algorithm is general in nature, and could be applied equally well in other situations in which the identification of similarities among pairs of graphs is meaningful and useful. Indeed, the initial stimulus for the development of the technique arose from the need to identify synonymous expressions in variant forms of indexing phrases (*14*). The practicability of the general method depends both on the sizes of the graphs and on the range of values that the lines and nodes assume—that is, the variety of bonds and atoms. If the range of both of these is very limited, excessive computing times are likely to result (*15*).

An algorithm which has a similar objective is at present being developed by A. R. Meetham (*16*) of the National Physical Laboratory, London.

COMPUTER ANALYSIS OF PAIRS OF STRUCTURES

The essence of the problem is the comparison of pairs of structures to determine what parts of the structures remain unchanged during the reaction. The reactions may be as simple as the formation of ethyl ether from ethanol, or as complex as the transformation of squalene into lanosterol (Figure 2).

$$2C_2H_5OH \rightarrow C_2H_5OC_2H_5$$

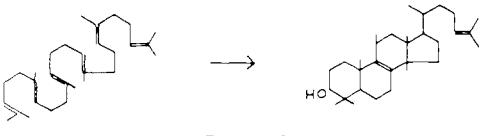
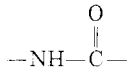


Figure 2.

Taking again the example of the Beckmann rearrangement of cyclohexanone oxime (Figure 3), the problem resolves itself into determining that the pentamethylene chain remains unchanged, and that the

configuration has been converted into



Substructure search is useless here, unless one searches for every substructure of one compound in the structure of the second. Instead, we have approached the problem by considering the stepwise generation of fragments from each structure in turn. The smallest fragments are the atoms of the structure, the next are the atom-bond-atom pairs, etc. At each stage, the fragments of one structure are compared with those from the second, and *vice versa*, and only those which are common to both are considered for further growth. This process is continued, the fragments growing in size at each step, until no larger common fragments can be formed.

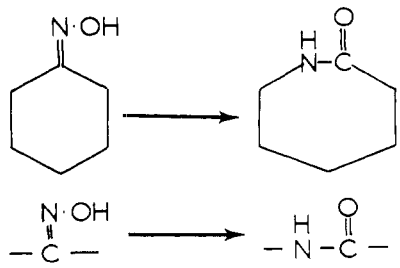


Figure 3.

The comparison of fragments at each stage in the procedure must be performed many times during the analysis; it is essential, therefore, that it be carried out as efficiently as possible. This comparison is in fact a search for identity. A general program for a search for identity makes considerable demands in terms of computing times. For this reason, we have chosen to limit the fragments generated to simple structures for which canonical forms can be generated quickly and efficiently to compare the fragments in canonical form, and subsequently to synthesize the actual common fragments, which may be complex and highly branched, from the simple units.

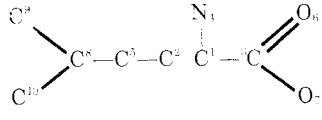
COMPUTER PROGRAMS

We have concentrated initially on acyclic structures, and therefore on reactions in which both reactants and products are acyclic, since the acyclic case presents fewer complexities than that in which rings are also present. At the present time we have completed and successfully tested programs which analyze pairs of structures in terms of simple fragments, and print out the results of the analysis.

Programs to complete the detection of similarity are at an advanced stage of development. In the meantime, however, we have used manual methods to simulate the computer procedures, and have evaluated the effectiveness of the analysis on a sample of chemical reactions.

A simple list-processing language has been used to great advantage in this work. This language is ALP, Autocode List-Processing Language, developed for the Ferranti Mercury Computer by Cooper and Whitfield (*17*), but also available on Atlas. Because Mercury has very limited core space—480 words in list-processing mode—the programs were developed on Mercury and tested more extensively on the Science Research Council Atlas at Harwell.

The procedures use a topological description of the structures of the molecules. The record used is a variant of the compact list suggested by Gluck (*18*) and used at Chemical Abstracts Service (*19*). The use of the variant was dictated by language and machine considerations, which also necessitated the use of atomic numbers rather than atomic symbols. A typical record is shown in Figure 4.



Node No.	Node value	Connection	Bond value
1	6	0	0
2	6	1	1
3	6	1	1
4	7	1	1
5	6	2	1
6	8	3	2
7	8	3	1
8	6	5	1
9	6	8	1
10	6	8	1

Figure 4.

The program operates on a pair of these records, develops fragments of increasing size from them, and compares these fragments at each stage, excluding them if not common to both. This procedure is continued until the largest common fragments have been determined.

The simple fragments which the program generates are the atoms of each structure at the first step, the atom-bond-atom pairs, or chains of length 2, at the second, and thereafter, simple unbranched chains of increasing length. The choice of straight chains as simple fragments was dictated by the fact that an unbranched chain can be described by listing the atoms and bonds from either end, and a choice between these paths can be made by ordering the descriptions lexicographically, and arbitrarily selecting one of them. This is therefore a simple and highly efficient means of generating a canonical description of a fragment, in which form it can then be compared with fragments from the other structure. This concept of chain generation as a means of describing fragments of structures has already been adopted by Hyde (*20*).

Although reference has been made only to pairs of structures, reactions in which more than one reactant or product molecule take part can also be handled by this technique by considering the sets of compounds on each side of the equation as disconnected graphs, and numbering the nodes with successive integers.

The program is divided into four chapters, the functions of which are:

Chapter I:

(a) To create for each of the two structures a list of node numbers, node values, connections, and bond values.

(b) To eliminate from these lists any atoms which are not common to both structures.

(c) To generate, for each structure, a *List of Pairs*, each entry of which contains the node numbers and atomic numbers of atoms which are bonded to one another, and the value of the bond between them.

(d) To eliminate from the *List of Pairs* all pairs of atoms which are not common to the two structures in terms of the element and bond values.

(e) To create, from the remaining *Lists of Pairs*, representations of each pair, to be known as *Chains*, and to store the location of each chain in a *List of Chains*.

Chapter II:

(a) To extend each chain, if possible, by adding at each end of each chain further atoms to which the terminal atom is connected.

Chapter III:

(a) To eliminate any duplicate chains (in terms of identical node numbers) that may have been generated.

(b) To generate a canonical form of the description of each chain.

(c) To eliminate any chains not common to both structures.

(d) To return control to Chapter II for further extension, if possible.

Chapter IV:

(a) To print the chains common to both structures after all possible chains have been generated and examined.

EXECUTION

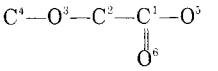
Chapter I:

The program is designed for list processing and is written in Autocode List-Processing Language. ALP divides each word in the available storage space into four fields, of which three are available for items of information, and one is used to store the address of the next cell in the list. It is thus possible to process each cell in turn by means of subroutines until the end of the list is reached.

(a) As there are four items of information for each node of a structure, and each cell can store only three items, each node requires a cell proper and a sub-cell. The data are stored in these in the following order:

Cell Proper	Node number	/	Atomic number	/	Subcell address
Subcell:	Number of attached node	/	Bond value	/	0

Thus a description of



would appear as:

Cell proper	Subcell
1/6/Address	0/0/0
2/6/Address	1/1/0
3/8/Address	2/1/0
4/6/Address	3/1/0
5/8/Address	1/1/0
6/8/Address	1/2/0

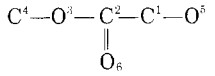
(b) When the lists of data have been compiled for each of the structures, the atomic numbers of the first are compared with those of the second, and any which do not appear are erased, along with their subcells. This procedure is then repeated for the second structure.

(c) To create the List of Pairs, the program adds to each row, the atomic number of the connected atom. The data are then re-arranged in a new list in the following order:

Cell proper:	Atomic number	/	Atomic No. of attached node	/	Subcell address
Subcell:	Node number	/	Number of attached node	/	Bond value

The nodes are ordered by atomic number, with the lower first, to give the unique form of description of each pair.

The list of pairs for the structure



is thus:

Cell proper	Subcell
6/6/Address	2/1/1
6/8/Address	2/3/1
6/8/Address	4/3/1
6/8/Address	1/5/1
6/8/Address	2/6/2

(d) The same procedure is followed for the second structure, and the pairs thus generated are compared by atomic numbers and bond value. Those not common to both structures are erased.

(e) The lists of pairs are rearranged so that the atomic numbers appear in the subcells and the node numbers in the cells proper. At the same time, a chain is created for each pair on each list, taking the following form:

Cell 1:	Atomic No.	/	Node No.	/	0
Cell 2:	0	/	0	/	Bond value
Cell 3:	Atomic No.	/	Node No.	/	0

e.g. $C_4 = O_3$ is represented as: 6/4/0, 0/0/2, 8/3/0

This method of storage is adopted in order to facilitate inversion of the chains.

Because each chain is a separate list, new lists are created to contain the addresses of the chains for each structure. Each cell in these lists stores the address of a chain and its length.

Chapter II:

Chain extension is performed on each chain of current length in turn. The first node of the chain is found in the List of Pairs for its structure. The attached node of the pair is compared with the second node of the chain to ensure that a different pair is in fact being examined. If so, extension can be made by adding a new node and bond. Thus, to extend the chain given above, node number 4 is found in the list of pairs. It is attached to node number 3, which is the second node of the chain. As 4 does not appear elsewhere, the chain cannot be extended in this direction, and must therefore be inverted for extension from the other end. After inversion, the topmost node, now 3, is treated in similar fashion. The chain can now be extended to include node 2:

6/2/0
0/0/1
8/3/0
0/0/1
6/4/0

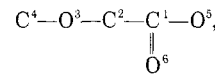
After extension, each chain is copied in inverted form; the topmost nodes of the original and inverted chains are compared. That chain with the lower value is added to the list of chains; it is now in unique form with regard to node numbers, and in this form duplicate chains can be located and discarded.

Chapter III:

(a) Duplicate chains are eliminated by taking each chain of current length, and searching in the list below it for a chain of the same length with identical node numbers; if found, the lower chain is discarded. The procedure is repeated for the second structure.
(b) The canonical form, with regard to atomic numbers and bond values, of each chain of current length, is now generated, again by copying in inverted form and comparing atomic numbers and bond values for each. The chain which has a lower value of either at the first point of difference is selected, and the other erased. In the case of symmetric chains the original is retained.
(c) The chains of current length of one structure are compared with those of the same length in the other in regard to atomic numbers and bond values. All chains for which exact equivalents are not found are discarded. When all chains of length 2 have been extended, they are discarded.
(d) Control returns to Chapter II in order to generate further extensions, if possible. Chapters II and III are repeated until no further growth of chains is possible. Control then passes to Chapter IV.

Chapter IV:

The chains that have been found common to both structures are now printed. Atomic symbols are substituted for atomic numbers, while single, double, and triple bonds are represented by —, =, and ≠ respectively. Node numbers are printed after the appropriate atomic symbol. The description of each chain is preceded by its length. Thus for the structure



presuming similar fragments to be present in a second structure, the chains would be-

CHAINS FOR PAIR NUMBER 1.
STRUCTURE NUMBER 1

5	C4—O3—C2—C1—O5
5	C4—O3—C2—C1=O6
4	C1—C2—O3—C4
4	O3—C2—C1—O5
4	O3—C2—C1=O6
3	C4—O3—C2
3	C1—C2—O3
3	C2—C1—O5
3	C2—C1=O6
3	O5—C1=O6

This program has been tested with a number of types of pairs of structures, including disconnected graphs, and has been found to perform satisfactorily. Nonetheless, there is a considerable degree of redundancy in the description of the fragments, and the number of chains generated is large for pairs of compounds which are similar in structure. As our initial aim was to detect the largest fragment common to pairs of structures, we introduced a routine to eliminate shorter chains which could not add any significant information in the synthesis of complex common fragments. The largest fragment that chains of length 3, alone, can give rise to, presuming a maximum ligancy of four for any atom, would contain five atoms. For chains of lengths 4 and 5, the corresponding numbers are 8 and 17. Thus, once common chains of length 6 have been identified, chains of length 3 can be discarded without loss of information about the largest fragment; and likewise chains of lengths 4 and 5 when chains of 9 and 18 have been reached. This substantially reduced the number of chains printed out in each comparison.

SYNTHESIS OF COMPLEX FRAGMENTS

Procedures were next developed to synthesize branched fragments actually common to both structures from the information available in the straight chains. For this purpose, a small sample of 22 chemical reactions involving only covalent acyclic compounds was selected at random from recent issues of *Index Chemicus*. The structures of the compounds were input in the form of compact lists; when more than one reactant or product was present, the set of structures was treated as a disconnected graph. The program described above was applied to the 22 sets of compounds, and the lists of chains examined.

A procedure was adopted in which a chain of greatest length in one structure was considered first. This was termed the *Reference Chain*. An equivalent for this, in terms of the atom and bond values, was located among the chains of the second structure, called the *Reference Equivalent*. The node numbers of the Reference Chain were used to locate another chain of the first structure with which it has a partial overlap. This chain was called the *Candidate Chain*. Finally, an equivalent for the Can-

REFERENCE CHAIN ←A→ REFERENCE EQUIVALENT
↑ B ↓ ↑ B ↓
CANDIDATE CHAIN ←A→ CANDIDATE EQUIVALENT

A: Equivalence in regard to atoms and bonds.
B: Overlap of node numbers.

Figure 5.

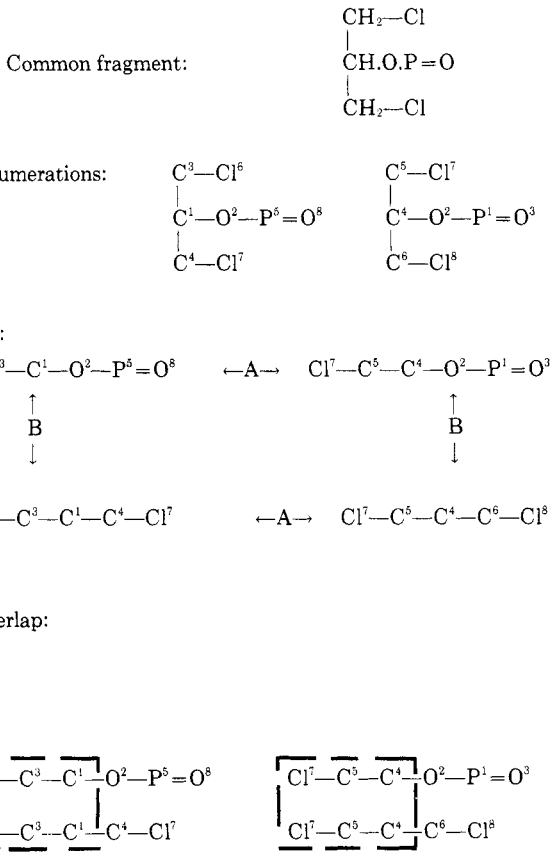


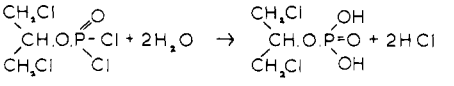
Figure 6.

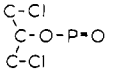
didate Chain was located in the second structure and was termed the *Candidate Equivalent*. A check was performed to ensure that the numbers of nodes common to Reference and Candidate, and Reference Equivalent and Candidate Equivalent chains were the same, and secondly, that the relative positions of these common nodes in each of the chains were the same. These steps are illustrated in Figures 5 and 6. If the conditions were satisfied, the equivalence was accepted, and the node numbers in each structure noted. If the tests were not satisfied, further equivalents or candidates were sought. After an equivalence was noted, further candidates were considered for superposition on the Reference Chain. With these, a further test was introduced—namely, that the number of new nodes that the Candidate and the Candidate Equivalent added to the nodes in the fragments already synthesized was the same. When no further candidates were found, the node numbers of the fragments were noted, a new reference chain taken, and the process repeated. The numbers of nodes in each fragment synthesized in this manner were counted, and the largest selected. While this procedure does not necessarily discover all possible fragments, since these may not contain a chain of greatest length, this condition did not arise in the sample examined.

Programs to carry out the synthesis of branched common fragments, which will be free from the above limitation, are at present under development.

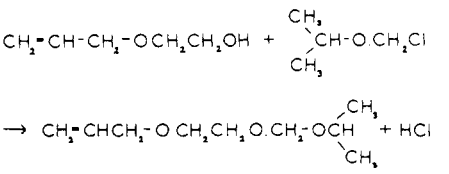
EVALUATION WITH CHEMICAL REACTIONS

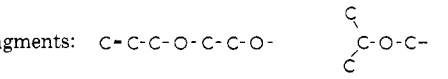
In almost half of the examples of chemical reactions studied, the manual procedure outlined above enabled a correct skeleton reaction scheme to be derived. In each case, the common fragments were subtracted from the original structure, and the bonds in reactants and products which had undergone change were identified. The technique is illustrated in the following examples:

A) 

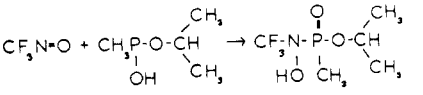
Common fragment: 

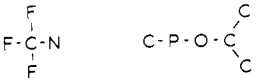
Bonds broken P—Cl, P—Cl. Bonds formed P—O, P—O.

B) 

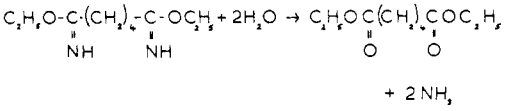
Common fragments: 

Bond broken C—Cl. Bond formed O — C.

C) 

Common fragments: 

Bonds broken: N=O, P—O. Bonds formed: N—O, P=O, N—P.

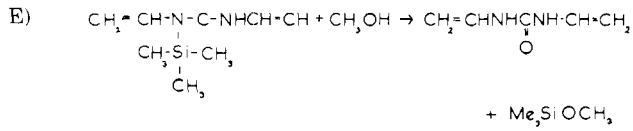
D) 

Common fragment:

Bonds broken: C=N, C=N; Bonds formed: C=O, C=O.

In each of the above cases, as in several other instances, the identification of the site and nature of reaction, and its description in terms of bonds broken or formed, was possible.

In another case, however, the routine which discarded shorter chains caused the loss of some valuable information. In example E, the fact that the trimethylsilyl group was common to both was not apparent.

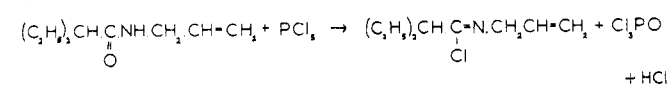
E) 

Common fragments

Obviously we shall need to retain shorter chains if we are to extract complete information about the nature of the reactions.

The cases in which the algorithm failed to discover the site and nature of the reaction were those in which bond changes in the main skeleton of the reactant occurred. This resulted, as was anticipated, in the discovery only of smaller common fragments, in which no bond changes occurred.

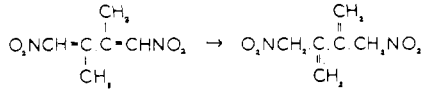
Examples of these are as follows:

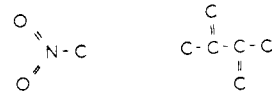
F) 

Common fragment: 

G)

Common fragments:

H) 

Common fragments 

In order to discover such shifts in multiple bonds, it will be necessary to relax the tests for identity of bond values selectively so as to discover common fragments in which the connectivity pattern remains the same. Tests on a larger sample of chemical reactions of acyclic compounds, and routines to overcome the limitations determined thus far are currently in hand, as is also the extension of the technique to determination of similarities among cyclic structures.

CONCLUSION

The concept of similarity among sets of chemical structures has far-reaching implications, not only in the analysis of chemical reactions, but in many other areas involving chemical structural information. It involves procedures which chemists use intuitively whenever they survey a set of chemical structures, and attempt to relate structure and activities of various kinds, including reactivities, physical properties, and biological properties. We believe that many of these procedures are amenable to automation, and hope to report on progress towards this aim in the future. We also believe that techniques developed from this basic structural comparison algorithm will provide chemists with powerful tools to extract significant correlations from the large masses of data which they must try to interrelate in the development of chemical theory.

ACKNOWLEDGMENT

We gratefully acknowledge a grant from the Office of Scientific and Technical Information, London, in support of this work, and thank G. M. Dyson and R. N. Maddison for valuable discussions.

LITERATURE CITED

(1) See for example, Theilheimer, W., "Synthetic Methods of Organic Chemistry," Interscience, New York (21 volumes), 1946–1967.
(2) See for example, Krauch, H. and W. Kunz, "Organic Name Reactions," Wiley, New York, 1964.
(3) Sussenguth, E. H., Jr., J. CHEM. DOC. **5**, No. 1, 36–43 (1965)
(4) Cossum, W. E., M. L. Krakiwsky, and M. F. Lynch, *Ibid.*, pp. 33–35
(5) Vleduts, G. E., *Inform. Stor. Retr.* **1**, Nos. 2/3, 117–46 (1963).
(6) Mishchenko, G. L., A. M. Shefter, and G. E. Vleduts *Nauchn.-tekhn. Inform.* **1965**, No. 10, 13–17.
(7) Dyson, G. M., and E. F. Riley, J. CHEM. DOC. **2**, No. 1, 19–22 (1962).
(8) Meyer, E., "Automation and Scientific Communication," H. P. Luhn, Ed., pp. 131–2. American Documentation Institute, Washington, D. C., 1965.
(9) Gelberg, A., J. CHEM. DOC. **6**, No. 1, 60–61 (1966).
(10) Fugmann, R., W. Braun, and W. Vaupel, *Nachr. Dok.* **14**, No. 4, 179–90 (1963).
(11) Vleduts, G. E., and G. L. Mishchenko, *Zavodsk. Lab.* **30**, No. 3, 329–31 (1964).
(12) Shevyakava, L. A., and F. M. Stoyanovich, *Nauchn.-tekhn. Inform.* **1965**, No. 12, p. 21–22.
(13) Armitage, J. E., and M. F. Lynch, *J. Chem. Soc., c*, **1967**, 521–8.
(14) Lynch, M. F., *J. Doc.* **22**, No. 3, 167–85 (1966).
(15) Unger, S. H., *Comm. ACM* **7**, No. 1, 26–34 (1964).
(16) Meetham, A. R., National Physical Laboratory, London, personal communication, 1967.
(17) Cooper, D. C., and H. Whitfield, *Computer J.* **5**, No. 1, 28–32 (1962).
(18) Gluck, D. J., J. CHEM. DOC. **5**, 225–9 (1965).
(19) Morgan, H. L., *Ibid.*, p. 107–113.
(20) Hyde, E., I.C.I. Ltd., Pharmaceutical Division, Alderley Edge, England, personal communication, 1966.