

- (2) J. Valls, "Reaction Documentation", in "Computer Representation and Manipulation of Chemical Information", W. T. Wipke, S. R. Heller, R. J. Feldmann, and E. Hyde, Ed., Wiley, New York, N.Y., 1974.
- (3) G. E. Vleduts, "Concerning One System of Classification and Codification of Organic Reactions", *Inf. Storage Retr.*, **1**, 117-146 (1963).
- (4) J. E. Armitage, and M. F. Lynch, "Automatic Detection of Structural Similarities among Chemical Compounds", *J. Chem. Soc. C*, 521-528 (1967).
- (5) J. E. Armitage, J. E. Crowe, P. N. Evans, M. F. Lynch, and J. A. McGuirk, "Documentation of Chemical Reactions by Computer Analysis of Structural Changes", *J. Chem. Doc.*, **7**, 209-215 (1967).
- (6) G. E. Vleduts, "Development of a Combined WLN/CTR Multilevel Approach to the Algorithmical Analysis of Chemical Reactions in View of Their Automatic Indexing", British Library, Research and Development Department, Report No. 5399, 1977.
- (7) M. F. Lynch and P. Willett, "The Automatic Detection of Chemical Reaction Sites", following paper in this issue.
- (8) J. M. Harrison and M. F. Lynch, "Computer Analysis of Chemical Reactions for Storage and Retrieval", *J. Chem. Soc. C*, 2082-2087 (1970).
- (9) R. Clinging and M. F. Lynch, "Production of Printed Indexes of Chemical Reactions. I. Analysis of Functional Group Interconversions", *J. Chem. Doc.*, **13**, 98-102 (1973); "II. Analysis of Reactions Involving Ring Formation, Cleavage, and Interconversion", *J. Chem. Doc.*, **14**, 69-71 (1974).
- (10) M. F. Lynch, P. R. Nunn, and J. Radcliffe, Final Report to the British Library, Research and Development Department on the Project "Development and Assessment of an Automatic System for Analysing Chemical Reactions", BLR&D Report 5236, 1975.
- (11) M. F. Lynch, "Screening Large Chemical Files" in ref 1, pp 177-194.
- (12) G. W. Adamson, J. Cowell, M. F. Lynch, A. H. W. McLure, W. G. Town, and A. M. Yapp, "Strategic Considerations in the Design of a Screen System for Substructure Searches of Chemical Structure Files", *J. Chem. Doc.*, **13**, 153-157 (1973).
- (13) L. Hodes, "Selection of Descriptors According to Discrimination and Redundancy. Application to Chemical Structure Searching", *J. Chem. Inf. Comput. Sci.*, **16**, 88-93 (1976).
- (14) C. E. Granito, S. Roberts, and G. W. Gibson, "The Conversion of Wiswesser Line Notations to Ring Codes. I. The Conversion of Ring Systems", *J. Chem. Doc.*, **12**, 190-196 (1972).
- (15) S. D. Bedrosian, and M. B. Milne, "Graphical Representation for Automated Retrieval of a Class of Fused Six-Rings", *J. Chem. Inf. Comput. Sci.*, **17**, 47-49 (1977).
- (16) D. Lefkowitz and A. R. Gennaro, "A Utility Analysis for the MCC Topological Screen System", *J. Chem. Doc.*, **10**, 86-94 (1970).
- (17) E. Hyde, F. W. Matthews, L. H. Thomson, and W. J. Wiswesser, "Conversion of Wiswesser Notation to a Connectivity Matrix for Organic Compounds", *J. Chem. Doc.*, **7**, 200-204 (1967).
- (18) J. B. Hendrickson, "Systematic Synthesis Design. IV. Numerical Codification of Construction Reactions", *J. Am. Chem. Soc.*, **97**, 5784-5800 (1975).
- (19) A. Leo, D. Elkins, and C. Hansch, "Computerized Management of Structure-Activity Data. III. Computerized Decoding and Manipulation of Ring Structures Coded in WLN", *J. Chem. Doc.*, **14**, 65-69 (1974).
- (20) J. E. Ash, "Connection Tables and Their Role in a System", in ref 1, pp 156-176.
- (21) M. F. Lynch, "The Microstructure of Chemical Data-Bases and the Choice of Representation for Retrieval", in ref 2.
- (22) E. Garagnani, and J. C. J. Bart, "Organic Reaction Schemes and General Reaction-Matrix Types. III. A Quantitative Analysis", *Z. Naturforsch., Teil B*, **32**, 465-468 (1977).
- (23) J. E. Crowe, P. Leggate, B. N. Rossiter, and J. F. B. Rowland, "The Searching of Wiswesser Line Notation by Means of a Character-Matching Serial Search", *J. Chem. Doc.*, **13**, 85-92 (1973).
- (24) C. E. Granito, G. T. Becker, S. Roberts, W. J. Wiswesser, and K. J. Windlix, "Computer-Generated Substructure Code (Bit Screens)", *J. Chem. Doc.*, **11**, 100-110 (1971).

## The Automatic Detection of Chemical Reaction Sites

MICHAEL F. LYNCH\* and PETER WILLETT

Postgraduate School of Librarianship and Information Science, University of Sheffield, Western Bank,  
Sheffield, S10 2TN, United Kingdom

Received December 5, 1977

An approximate structure-matching algorithm is described which rapidly identifies substructures common to the reactants and products of a chemical reaction. The deletion of these features results in the identification of those parts of the reacting molecules that have been changed in the course of the reaction; at the same time it is possible to locate the reaction sites within their parent molecules so that substructural searches could be performed both for reacting and nonreacting features. The procedure has been tested upon a sample file of 340 reactions, and intuitively reasonable analyses were obtained for 315 of them (92.6%); a detailed failure analysis is given. Potential applications to computer-aided synthesis design and to the production of large files of chemical reactions are discussed.

### 1. INTRODUCTION

Work in this department<sup>1,2</sup> has led to the development of automatic methods for the detection of the overall structural changes occurring in organic reactions. This has been achieved by breaking the reacting species down into sets of fragments, eliminating the duplicate fragments, those parts of the molecules that ostensibly remain unchanged, and rebuilding the remaining features to synthesize a reaction site. As a fragmentation process is involved, a degree of ambiguity is present, and it has not proved generally possible to determine the exact location of the reaction sites within their parent molecules.

Vleduts has recently suggested a procedure whereby this might be achieved.<sup>3</sup> His approach consists of an atom-by-atom mapping of one reacting molecule onto another so as to identify the largest common substructures and, by subtraction, the differences engendered by the reaction. As no fragmentation

is involved, a much more specific localization of the reaction site may be obtained, and it should be possible to identify the bonds which have been broken or formed in the course of the reaction. The algorithm involves the identification of maximal subgraphs common to the two sides of the equation; in contrast to the problem of graph isomorphism,<sup>4-7</sup> maximal subgraph isomorphism has been little studied due to the greater complexity of the problem.<sup>8,30</sup>

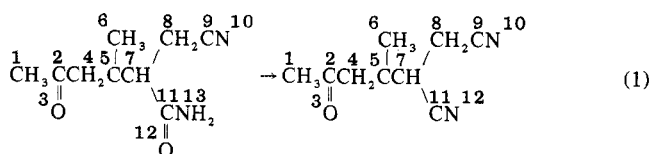
It is well known that isomorphism can be determined by a simple enumeration process;<sup>7</sup> in the present case a possible procedure would consist of generating all possible subgraphs (partial structures) from one graph (reacting molecule) and matching them against all possible subgraphs from the other.<sup>9</sup> The computation required may be substantially reduced if properties of such subgraphs which are invariant under isomorphism are taken into account; thus a reactant atom may not be mapped onto a product atom if the two atom types are different. Such "set reduction" techniques, initially described

by Unger<sup>10</sup> and Sussenguth,<sup>11</sup> form the basis of the methods used for registration and substructure search in computer-based chemical information systems. Although structure search systems have been developed to a substantial degree of efficiency,<sup>12,13</sup> they still require an excessive amount of computer time if large numbers of structures have to be matched against the query. The time is decreased by the use of complex screening systems<sup>14-16</sup> which reduce the size of the file that must be searched by 90%; moreover, the set reduction techniques are often applied by themselves without a subsequent iterative search, false drops being removed by manual inspection.<sup>13</sup> Such relatively unsophisticated structure-matching procedures are sufficient since the graphs to be manipulated are both labeled and relatively simple so that a query may be compared with many possible matches in a relatively short time.

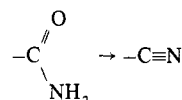
An algorithm to identify maximal common subgraphs will be of greatest practical utility if the process of subgraph matching is simple and efficient; indeed, Vleduts suggests that such an algorithm, implemented upon today's machines, will probably be limited to structures not exceeding 10-15 atoms.<sup>3</sup> He accordingly described a procedure whereby a comparison of the WLN symbol strings of the reacting molecules would be used to provide "guiding information", i.e., reactant-product atom equivalences, to reduce the amount of iterative mapping that must be performed. The work described in this report was initially undertaken to provide an alternative means of obtaining this guiding information; however, we now feel that it may, in itself, be used to process large files of chemical reactions and to provide machine-readable representations which could then be interrogated using currently available structure-search techniques. The procedure consists of identifying one or more large areas common to both sides of the reaction equation; taken together, these may correspond to the maximal common subgraph but this will not be generally so. Accordingly, it is not possible to delineate explicitly the bonds changed by the reaction, but this limitation is offset by the very much larger number of reactions that may be processed in the same amount of time.

## 2. ALGORITHM FOR AUTOMATIC REACTION SITE DETECTION

We consider chemical structures as represented by graphs whose nodes are the atoms and edges the bonds. The graphs R and P are the reactant(s) and product(s) of a chemical reaction and their nodes are denoted by  $r_i$  and  $p_i$ , or generally,  $a_i$ . The very simple reaction of eq 1 will illustrate the basic concepts of the procedure.



We wish to detect the change



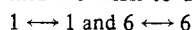
and to note that reactant atoms 11-13 have been transformed to product atoms 11-12. We have not made any attempt to specify, for example, that atom 11 in the reactant reaction site corresponds to atom 11 in the product reaction site; such mappings may only be made if assumptions are made as to the mechanism of the reaction. As advocated by Hendrickson,<sup>17</sup> we are only concerned with the overall structural changes.

Consider the methyl groups present in the reacting molecules

above; the possible mappings are (in an obvious notation)



and we wish to detect the equivalences



Equivalent atoms within a single molecule may be detected by application of the Morgan algorithm.<sup>18</sup> This partitions the atoms present by considering the number of their attachments, the first-order connectivity; as connectivity values rarely exceed four or five, further refinement is obtained by consideration of higher order connectivities. The  $n$ th-order connectivity of an atom is calculated by summing the  $(n-1)$ th connectivity values of all adjacent atoms; thus, the two reactant methyl groups of eq 1 may be differentiated by their third-order connectivities since their sets of adjacent atoms have different bond patterns. The discriminatory power of the procedure may be further increased by the use of additional properties, such as atom type and the surrounding bond pattern, in conjunction with the connectivity;<sup>19</sup> at the same time the  $n$ th-order property (connectivity) value of an atom  $a_i$ ,  $V_{a_i}^n$ , more accurately represents a circular substructure of radius  $(n-1)$  bonds centered upon  $a_i$ . We may consider the number  $V_{a_i}^n$  to be a "hash" of this substructure which may be obtained without a detailed atom-by-atom investigation of the feature it describes.

The matching procedure is based upon principles:

(i) The Morgan algorithm, with a different stopping rule (see below), is applied simultaneously to both of the reacting molecules so that inter-, rather than intra-, molecular equivalences are detected. An account of an analogous procedure is given by Figueras.<sup>13</sup>

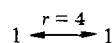
(ii) We assume that the  $n$ th-order property value  $V_{a_i}^n$  is a unique representation of an  $(n-1)$ -bond radius, circular substructure centered upon atom  $a_i$ . Hence if  $V_{r_i}^n = V_{p_j}^n$ , the reactant and product atoms  $r_i$  and  $p_j$  may be considered to be at the center of identical substructures, and these areas may be assumed to be the same without a detailed examination of the atoms contained therein. The assumption may, under certain circumstances, be incorrect, but we have found that for reacting molecules it is true in a very high percentage of cases (see below).

Wiswesser Line Notations were used as the input structure representations. These were converted to CROSSBOW connection tables using software kindly provided by ICI Pharmaceuticals Division Ltd. The initial property value of each atom,  $V_{a_i}^1$ , is an integer derived from the atomic "dot-plot" symbol which uniquely describes the type and the bond pattern of the atom.<sup>20</sup> Higher order property values are obtained from the equation

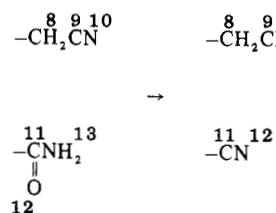
$$V_{a_i}^n = 2V_{a_i}^{n-1} + \sum V_{a_j}^{n-1}$$

where the summation is over all atoms adjacent to  $a_i$ . The function is similar to that used by Shelley and Munk to identify intramolecular atomic equivalences.<sup>19</sup> Higher order property values are calculated for all reactant and product atoms until there are no remaining pairs of atoms for which  $V_{r_i}^n = V_{p_j}^n$ . We then note the pair(s) of atoms for which  $V_{r_i}^{n-1} = V_{p_j}^{n-1}$ , i.e., those reactant-product atom pairs which are at the center of identical circular substructures. As the  $V_{a_i}^1$  values correspond to the initial atoms, the substructures are of radius  $(n-2)$  bonds, the match radius  $r$ , and we accordingly delete all atoms contained therein from the reactant and product adjacency matrices.

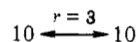
To return to the reaction of eq 1, possible matches  $r_i \leftrightarrow p_j$  are obtained until  $V_{a_i}^6$  values have been calculated; we hence obtain the equivalence



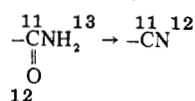
which, after the elimination of all atoms within the match radius, results in the reaction diagram



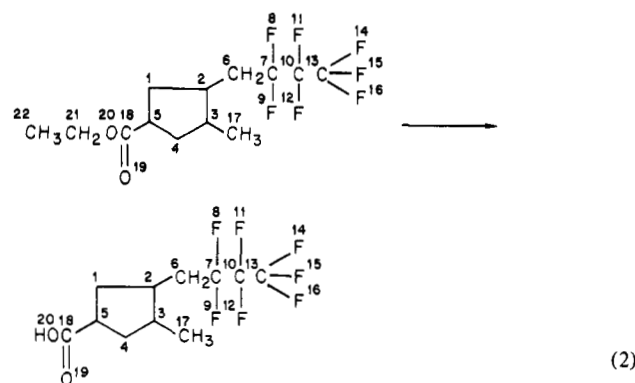
We may now repeat the process, considering only these partial structures, to obtain the equivalence



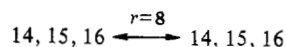
No further mappings can be found, so after updating the reactant and product adjacency matrices the procedure terminates to yield the reaction scheme



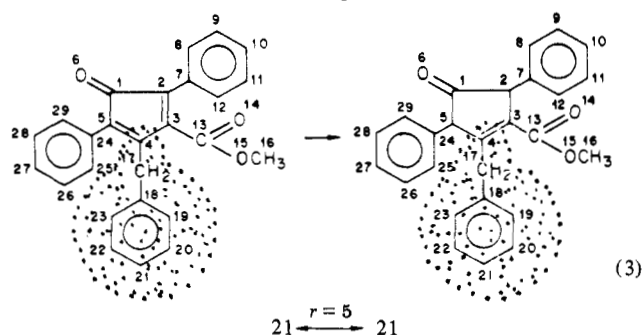
The actual implementation has three additional features which should be mentioned. Firstly, we must allow for multiple equivalences as illustrated by the reaction shown in eq 2 for



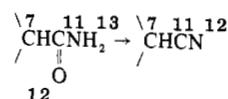
which we obtain the mapping



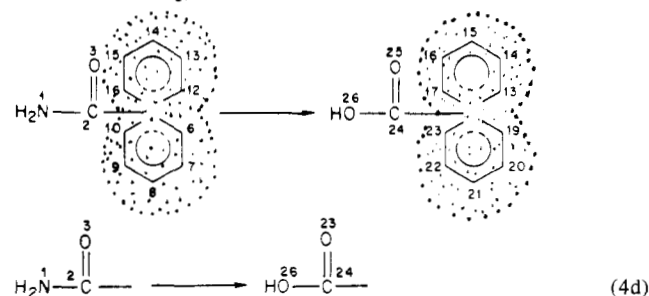
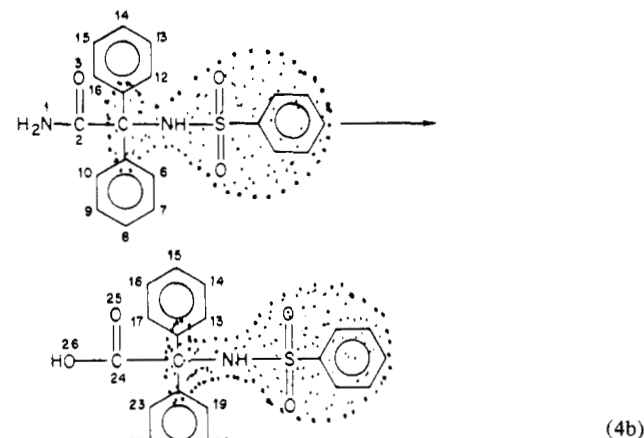
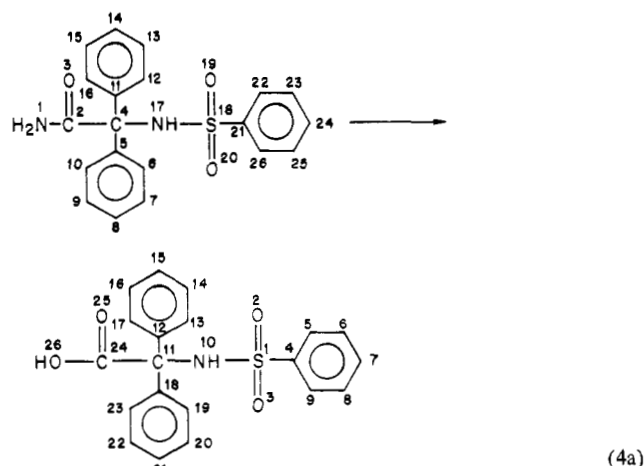
The atoms, and the substructures centered upon them, may be deleted only if all three reactant atoms have the same set of three possible matches, i.e., if the members of the reactant set are equivalent one to another as well as to the product set. If this is found to be so, an arbitrary assignment is made for each member of the reactant set and the deletion process takes place as normal. Secondly, we have defined a minimal match radius of two bonds; for the reactions tested so far we find that if a radius of one is used, corresponding to an augmented atom,<sup>21</sup> there is a noticeable increase in the number of incorrect mappings. Finally, for a match radius  $r$ , only the atoms within  $(r-1)$  bonds are deleted; this step is taken to guard against cases such as the reaction of eq 3 where the bonds attached



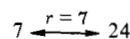
to the outermost atoms ( $r_4, p_4$ ) are differently oriented in the two structures (the matched substructures are shaded). These latter two restrictions tend to reduce slightly the number of atoms eliminated; thus the reaction of eq 1 is now analyzed as



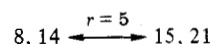
To further illustrate the working of the algorithm, consider the simple functional-group change shown in eq 4a. Possible



reactant-product matches are obtained until the ninth-order property values have been calculated. At this point we obtain the mapping



and thence delete all the atoms within a radius of six bonds; the removed substructures are shaded in eq 4b. Consideration of the remaining atoms yields the mappings



As both reactant atoms have the same pair of possible product

matches, a symmetry is presumed to exist and a further elimination takes place as shown in eq 4c (which contains only the partial structures remaining after the initial match). No further mappings are possible so the algorithm terminates to yield the reaction scheme shown in eq 4d.

## RESULTS AND DISCUSSION

The file of one-reactant, one-product reactions described previously<sup>2</sup> was used. A 1 in 25 sample was used to generate redundant adjacency matrices for the reacting molecules, and these matrices were used as input to the analysis program (written in Algol-68). This program required 43 seconds (excluding input and output) to process the sample file of 340 reactions, and the results obtained are shown in Table I. The output from the analysis consists of the adjacency matrices together with a bit vector denoting whether a given atom is to be considered as belonging to the reaction site.

Successful analyses (92.6% of the sample file) are those judged to be an adequate representation of the change; examples are shown in eq 5a-d. In each case the equivalences

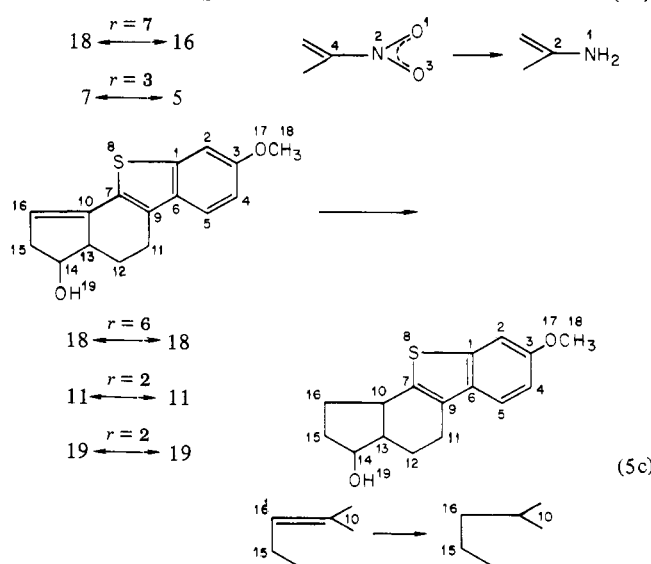
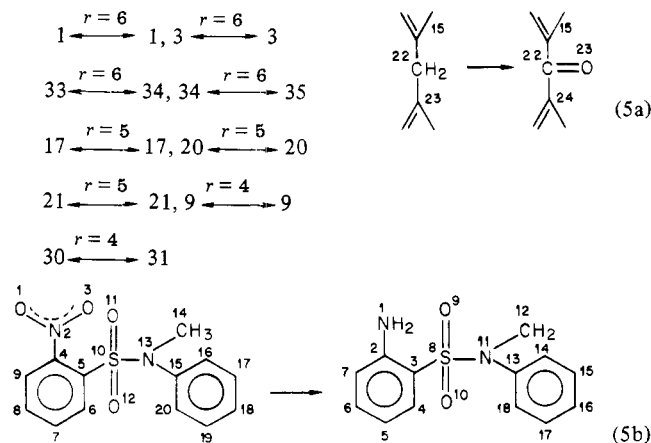
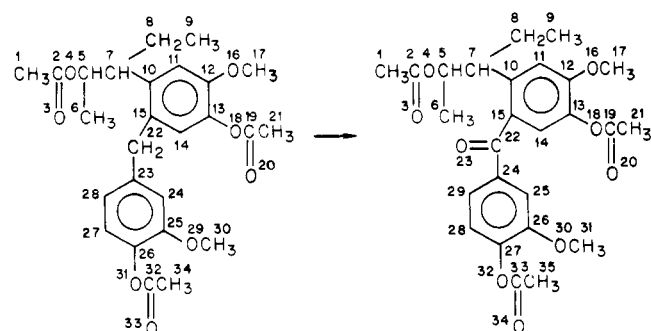
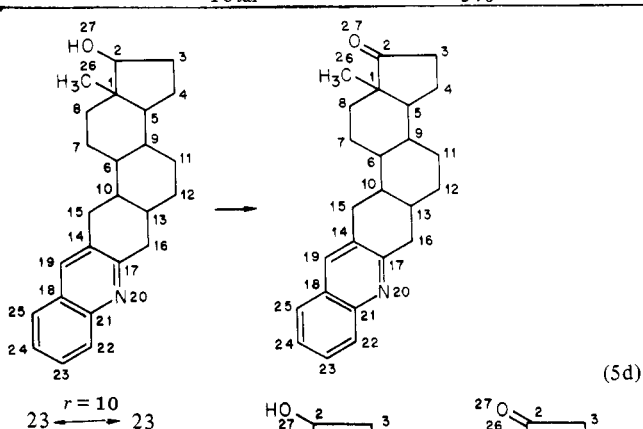


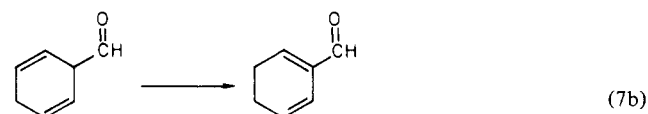
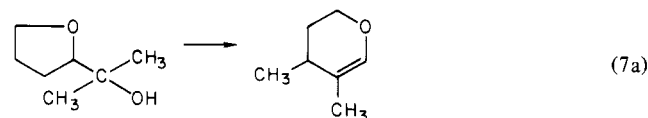
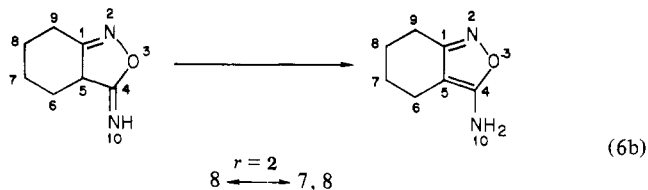
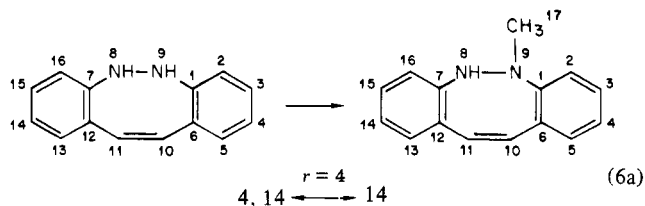
Table I. Results for Analysis of Sample File of 340 Reactions

Successful	315
Overflow	2
No atoms detected	15
Failure	8
Total	340



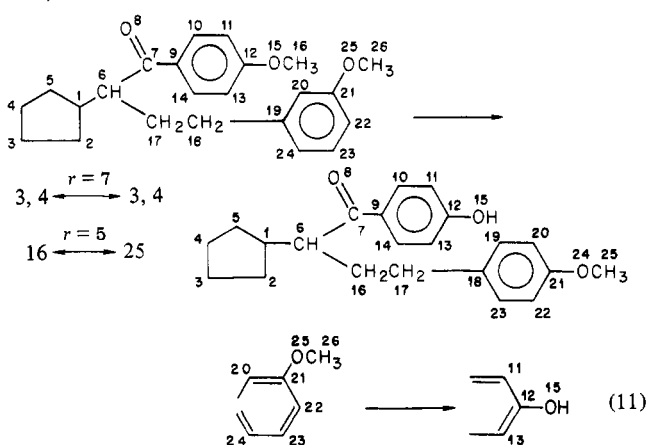
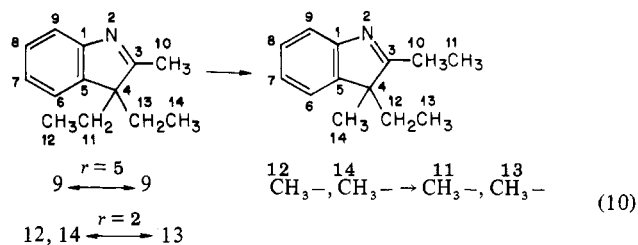
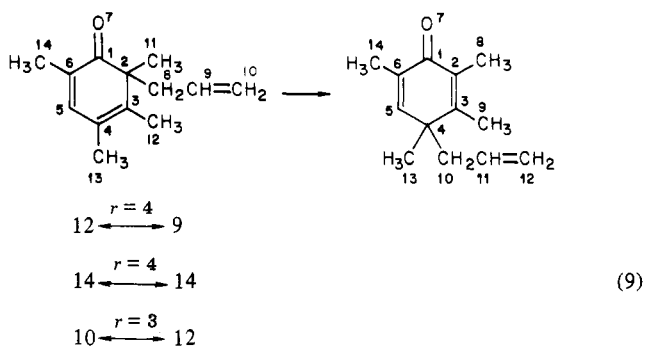
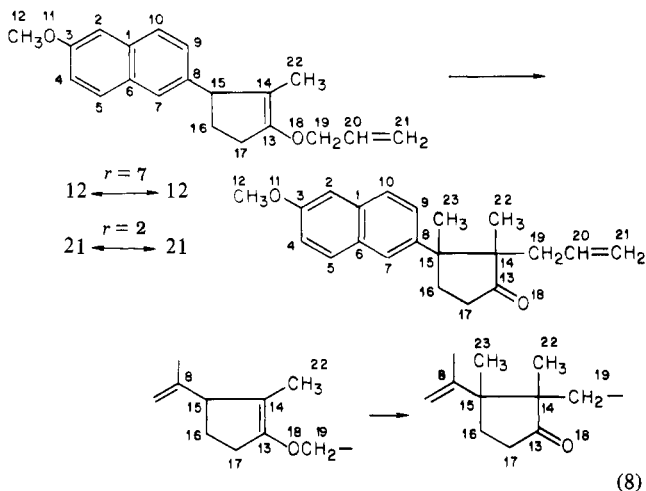
are shown in the lower left-hand part of the figure and the derived reaction site in the lower right-hand part. For large match radii, the property values may become quite high and the two reactions described as "overflow" in Table I correspond to cases where the values become too great for the computer word reserved for them. In both cases the match radius, if calculated, would have been over 20.

Reactions where no atoms have been deleted are shown in eq 6 and 7. Reactions 6a and 6b are not processed since an

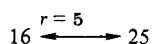


ambiguous mapping is detected. For symmetric molecules an arbitrary assignment procedure could be used; Ming and Tauber<sup>12</sup> give a detailed description of assignment procedures for structure matching but these are not generally applicable since incorrect assignments could not be detected subsequently. We will also not obtain any mappings if the reaction has occurred in a fairly small molecule, e.g., reactions 7a and 7b where no pairs of atoms have a match radius > 1.

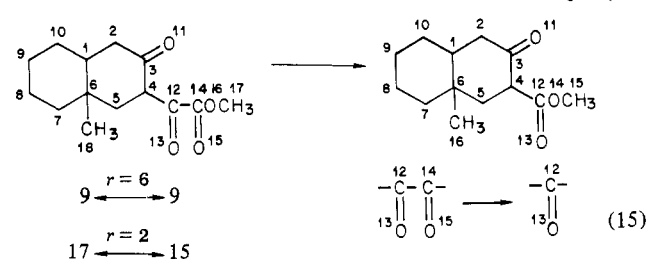
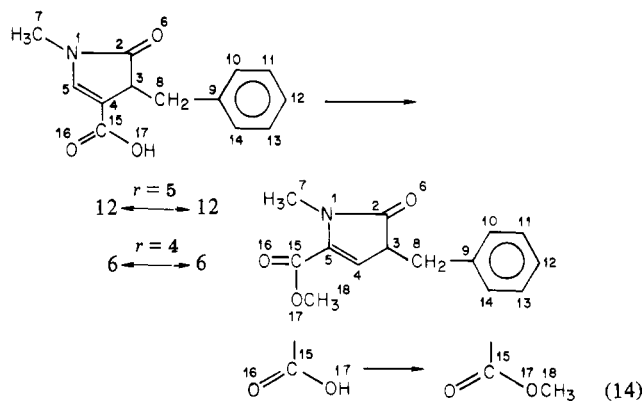
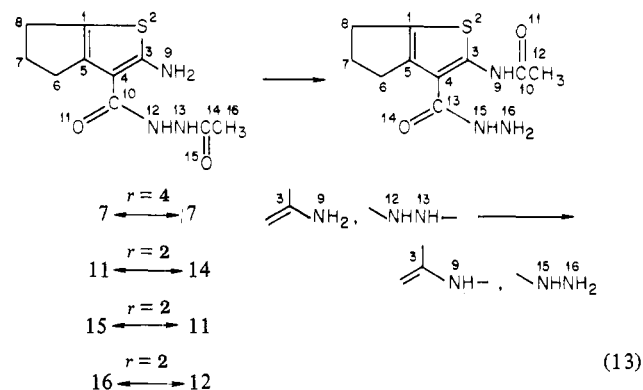
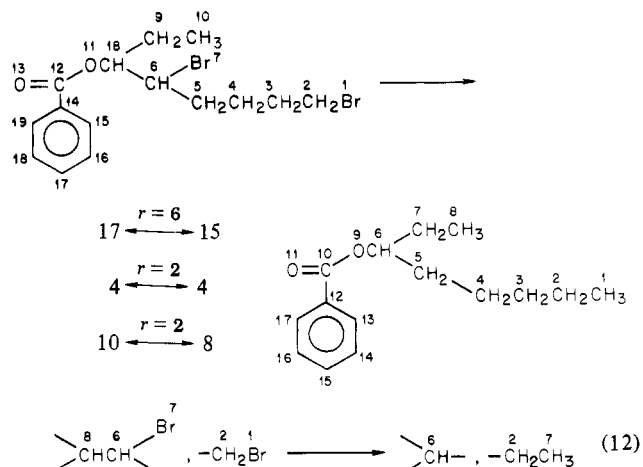
The failures are shown in eq 8-15 and it will be seen that they arise for a variety of reasons. Problems arise in the first example (eq 8) because of the shift of the allyl group. Re-



actions 9 and 10 are the only ones which contradict the assumption that equal property values correspond to identical substructures; both cases involve functional-group shifts not detected by the matching algorithm. Incorrect mappings will be obtained if an atom involved in the change is matched with a nonreacting atom; thus in eq 11, the invalid equivalence below is obtained.



The remaining failures are shown in eq 12–15, the last of these involving another group shift. From the small sample file tested, it is not yet possible to suggest how the analysis



procedure might be improved to obviate some of its current limitations, but clearer guidelines may present themselves when a larger file has been processed.

We have recently described<sup>2</sup> how reactions may be indexed by an automatic comparison of the WLN symbol strings of the reacting molecules. As was pointed out in that report, such an approach will be most applicable where the reacting molecules are similar in WLN terms as well as structurally; in cases where this is not so, the methods described here would provide an alternative means of access to the data. This will also be so for those reactions in which no common fragments could be detected and for those in which a more detailed

consideration of individual atoms and bonds is required. We would expect to be able to provide three levels of access to a reaction file, comparable to the facilities provided by current structure-search systems.<sup>16</sup> The initial approach would be via an algorithmically generated, fragment bit-screen; next would be a string search upon the WLN fragments and reaction sites obtained by the methods of ref 2, and the final search would be of the truncated structures produced as above. We have found that connection-table-derived analyses are inadequate for the retrieval of reactions in which rings are broken or formed.<sup>22,23</sup> This could probably be overcome by some form of ring perception algorithm,<sup>24,25</sup> but such reactions are more easily described by the WLN analyses since WLN explicitly delineates a subset of the rings present which permits the rapid identification of any ring changes that have occurred.<sup>2</sup> This being so, the bit screen would represent features generated from both types of analysis.

Such large reaction files would form a natural complement to the rapidly growing area of computer-aided synthesis design.<sup>26</sup> Programs in this field analyze the target molecule for the presence of a range of substructural features which are then searched in a limited database of, perhaps, 300 reactions of proven synthetic value.<sup>27-29</sup> The extraction of all possible substructures from a molecule is extremely time-consuming, and so only relatively simple attributes are detected in the target molecule by a synthesis program; an alternative procedure would be to select certain features, of whatever complexity, to be searched in a more substantial database. For very large files the operations probably could not be carried out on-line via a terminal, but this drawback would be more than counterbalanced by the much greater range of reaction types that could be brought to bear upon the synthesis.

### CONCLUSIONS

We have described an approximate structure-matching procedure by means of which areas of molecules unaffected by a chemical reaction can be identified; the deletion of these common substructures permits the rapid detection of the reaction site while retaining a full description of the surrounding parts of the reacting molecule. A computer program based upon this algorithm produced intuitively reasonable analyses for 92.6% of a sample file of 340 one-reactant, one-product reactions.

### ACKNOWLEDGMENT

We thank the Institute for Scientific information for providing WLN tapes from which the sample reaction file was created and ICI Pharmaceuticals Division Ltd. for software. P.W. thanks the Department of Education and Science for an Information Science Research Studentship. We thank the referees for their suggestions as to the presentation of this work and for a recent reference.

### REFERENCES AND NOTES

- (1) J. M. Harrison and M. F. Lynch, "Computer Analysis of Chemical Reactions for Storage and Retrieval", *J. Chem. Soc. C*, 2082-2087 (1970).
- (2) M. F. Lynch and P. Willett, "The Production of Machine-Readable Descriptions of Chemical Reactions Using Wiswesser Line Notations", preceding paper in this issue.
- (3) G. E. Vleduts, "Development of a Combined WLN/CTR Multilevel Approach to the Algorithmic Analyses of Chemical Reactions in View of Their Automatic Indexing", British Library, Research and Development Department Report No. 5399, 1977.
- (4) D. G. Corneil and C. C. Gotlieb, "An Efficient Algorithm for Graph Isomorphism", *J. Assoc. Comput. Mach.*, **17**, 51-64 (1970).
- (5) A. T. Bertziss, "A Backtrack Procedure for Isomorphism of Directed Graphs", *J. Assoc. Comput. Mach.*, **20**, 365-377 (1973).
- (6) G. Levi, "Graph Isomorphism: a Heuristic Edge-Partitioning-Oriented Algorithm", *Computing*, **12**, 291-313 (1974).
- (7) J. R. Ullmann, "An Algorithm for Subgraph Isomorphism", *J. Assoc. Comput. Mach.*, **23**, 31-42 (1976).
- (8) H. G. Barrow and R. M. Burstall, "Subgraph Isomerism, Matching Relational Structures and Maximal Cliques", *Inf. Proc. Lett.*, **4**(4), 83-84 (1976).
- (9) G. Levi, "A Note on the Derivation of Maximal Common Subgraphs of Two Directed or Undirected Graphs", *Calcolo*, **9**, 1-12 (1972).
- (10) S. H. Unger, "GIT—a Heuristic Program for Testing Pairs of Directed Line Graphs for Isomorphism", *Comm. Assoc. Comput. Mach.*, **7**, 26-34 (1964).
- (11) E. H. Susseguth, "A Graph-Theoretic Algorithm for Matching Chemical Structures", *J. Chem. Doc.*, **5**, 36-43 (1965).
- (12) T. K. Ming and S. J. Tauber, "Chemical Structure and Substructure Search by Set Reduction", *J. Chem. Doc.*, **11**, 47-51 (1971).
- (13) J. Figueras, "Substructure Search by Set Reduction", *J. Chem. Doc.*, **12**, 237-244 (1972).
- (14) G. W. Adamson, J. Cowell, M. F. Lynch, A. H. W. McLure, W. G. Town, and A. M. Yapp, "Strategic Considerations in the Design of a Screening System for Substructure Searches of Chemical Structure Files", *J. Chem. Doc.*, **13**, 153-157 (1973).
- (15) A. Feldman and L. Hodes, "An Efficient Design for Chemical Structure Searching. I. The Screens", *J. Chem. Inf. Comput. Sci.*, **15**, 147-152 (1975).
- (16) D. R. Eakin and E. Hyde, "Evaluation of On-Line Techniques in a Substructure Search System" in "Computer Representation and Manipulation of Chemical Information", W. T. Wipke, S. R. Heller, R. J. Feldman, and E. Hyde, Ed., Wiley, New York, N.Y., 1974.
- (17) J. B. Hendrickson, "A General Protocol for Systematic Synthesis Design", *Top. Curr. Chem.*, **62**, 49-172 (1976).
- (18) H. L. Morgan, "The Generation of a Unique Machine-Description for Chemical Structures—a Technique Developed at Chemical Abstracts Service", *J. Chem. Doc.*, **5**, 107-113 (1965).
- (19) C. A. Shelley and M. E. Munk, "Computer Perception of Topological Symmetry", *J. Chem. Inf. Comput. Sci.*, **17**, 110-3 (1977).
- (20) W. J. Wiswesser, "The 'Dot-Plot' Computer Program", in "Proceedings of the Wiswesser Line Notation Meeting of the Army Chemical Information and Data Systems", Program AD 655-397, J. P. Mitchell, Ed., 1968.
- (21) J. E. Crowe, M. F. Lynch, and W. G. Town, "Analysis of Structural Characteristics of Chemical Compounds in a Large Computer-Based File." Part II. "Atom-Centred Fragments", *J. Chem. Soc. C*, 3702-3706 (1971).
- (22) D. Bawden, "An Investigation into the Fundamental Aspects of the Analysis of Chemical Reactions for Storage and Retrieval", unpublished M.Sc. thesis, Sheffield, 1974.
- (23) M. F. Lynch, "Computer Organisation and Retrieval of Chemical Structure Information", Office of Scientific and Technical Information Report No. 5208, 1974.
- (24) A. Zamora, "An Algorithm for Finding the Smallest Set of Smallest Rings", *J. Chem. Inf. Comput. Sci.*, **16**, 40-43 (1976).
- (25) A. Esack, "A Procedure for Rapid Recognition of the Rings of a Molecule", *J. Chem. Soc., Perkin Trans. 1*, 1120-1124 (1975).
- (26) M. Bersohn and A. Esack, "Computers and Organic Synthesis", *Chem. Rev.*, **76**, 269-282 (1976).
- (27) H. W. Orf, "Computer-Assisted Synthetic Analysis", Ph.D. thesis, Harvard University, 1976.
- (28) M. Bersohn and A. Esack, "A Computer Representation of Synthetic Organic Reactions", *Comput. Chem.*, **1**, 103-107 (1976).
- (29) A. F. Sanders, "Some Applications of Graph Theory to the Design of a Heuristic Program for the Discovery of Organic Syntheses", Ph.D. thesis, State University of New York at Stony Brook, 1976.
- (30) M. M. Cone, R. Venkataraghavan, and F. W. McLafferty, "Molecular Structure Comparison Program for the Identification of Maximal Common Substructures", *J. Am. Chem. Soc.*, **99**, 7668-7671 (1977).