Relationships", *Acc. Chem. Res.*, **2**, 232–239 (1969).
(3) Free, S. M. and Wilson, J. W., "A Mathematical Contribution to Structure-Activity Studies", *J. Med. Chem.*, **7**, 395–399 (1964).
(4) Kier, L. B., "Molecular Orbital Theory in Drug Research", Academic Press, New York, N.Y., 1971.
(5) Adamson, G. W., and Bush, J. A., "Method for Relating the Structure and Properties of Chemical Compounds", *Nature (London)*, **248**, 406–407 (1974).
(6) Adamson, G. W., and Bush, J. A., "A Comparison of the Performance of Some Similarity and Dissimilarity Measures in the Automatic Classification of Chemical Structures", *J. Chem. Inf. Comput. Sci.*, **15**, 55–58 (1975).
(7) Adamson, G. W., and Bush, J. A., "Evaluation of an Empirical Structure-Activity Relationship for Property Prediction in a Structurally Diverse Group of Local Anaesthetics", *J. Chem. Soc. Perkin Trans. 1*, in press.
(8) Cramer, R. D., Redl, G., and Berkoff, C. E., "Substructural Analysis. A Novel Approach to the Problem of Drug Design", *J. Med. Chem.*, **17**, 533–535 (1974).
(9) Kowalski, B. R., and Bender, C. F., "The Application of Pattern Recognition to Screening Prospective Anticancer Drugs. Adenocarcinoma 755 Biological Activity Test", *J. Am. Chem. Soc.*, **96**, 916–918 (1974).
(10) Smith, E. G., "The Wiswesser Line-Formula Chemical Notation", McGraw-Hill, New York, N.Y., 1968.
(11) Hansch, C., Leo, A., and Elkins, D., "Computerized Management of Structure-Activity Data. I. Multivariate Analysis of Biological Data", *J. Chem. Doc.*, **14**, 57–61 (1974).
(12) Brasie, W. C., and Liou, D. W., "Chemical Structure Coding", *Chem. Eng. Prog.*, **61**, 102–108 (1965).
(13) Osinga, M., and Verrijn Stuart, A. A., "Documentation of Chemical Reactions. II. Analysis of the Wiswesser Line Notation", *J. Chem. Doc.*, **14**, 196–198 (1974).
(14) Bird, A. E., and Marshall, A. C., "Correlation of Serum Binding of Penicillins with Partition Coefficients", *Biochem. Pharmacol.*, **16**, 2275–2290 (1967).
(15) "Statistical Analysis Mark II Applications Package", International Computers Limited Technical Publication 4301, London, ICL, 1971.
(16) Nys, G. G., and Rekker, R. F., "Statistical Analysis of a Series of Partition Coefficients with Special Reference to the Predictability of Folding of Drug Molecules. The Introduction of Hydrophobic Fragment Constants (f Values)", *Chim. Ther.*, **5**, 521–535 (1973).
(17) Saada, E., "Proposal for the Elimination of Contractions and Multipliers in the WLN Notation", *Bull. Chem. Notation Assoc.*, (1st ed), 39–41 (1974).
(18) Ash, J. E., and Hyde, E., "Chemical Information Systems", Ellis Horwood, Chichester, 1975, pp 106–108.
(19) Bush, J. A., Doctoral thesis, University of Sheffield, in preparation.
(20) Scholtan, W., "Die Bindung der Antibiotica an die Eiweisskorper des Serums", *Arzneim.-Forsch.*, **13**, 347–360 (1963).
(21) Crowe, J. E., Leggate, P., Rossiter, B. N., and Rowland, J. F. B., "The Searching of Wiswesser Line Notations by Means of a Character-Matching Serial Search", *J. Chem. Doc.*, **13**, 85–92 (1973).
(22) Adamson, G. W., Cowell, J., Lynch, M. F., McLure, A. H. W., Town, W. G., and Yapp, A. M., "Strategic Considerations in the Design of a Screening System for Substructure Searches of Chemical Structure Files", *J. Chem. Doc.*, **13**, 153–157 (1973).
(23) Heller, S. R., and Koniver, D. A., "Computer Generation of Wiswesser Line Notation. II. Polyfused, Perifused and Chained Ring Systems", *J. Chem. Doc.*, **12**, 55–59 (1972).
(24) Granito, C. E., Roberts, S., and Gibson, G. W., "The Conversion of Wiswesser Line Notations to Ring Codes. I. The Conversion to Ring Systems", *J. Chem. Doc.*, **12**, 190–196 (1972).
(25) Hansch, C., and Yoshimoto, M., "Structure-Activity Relationships in Immunochemistry. 2. Inhibition of Complement by Benzamidines", *J. Med. Chem.*, **17**, 1160–1167 (1974).
(26) Kowalski, B. R., and Bender, C. F., "Pattern Recognition. A Powerful Approach to Interpreting Chemical Data", *J. Am. Chem. Soc.*, **94**, 5632–5639 (1972).
(27) Kowalski, B. R., and Bender, C. F., "Pattern Recognition. II. Linear and Nonlinear Methods for Displaying Chemical Data", *J. Am. Chem. Soc.*, **95**, 686–693 (1973).

CRYSRC: A Generalized Chemical Information System Applied to a Structural Data File

JOSE VILLARREAL, JR.,[†] EDGAR F. MEYER, JR.,[*,†] ROGER W. ELLIOTT,[**] and CARL MORIMOTO[†]

Texas A&M University, College Station, Texas 77843

Received June 23, 1975

An interactive retrieval system, CRYSRC, based upon the modular design of input, query, and output routines has been implemented and tested on a well defined data base, the Cambridge Crystal Data Centre files. Use has been made of three-dimensional display facilities in this laboratory to create models of retrieved molecules. All routines have been implemented on a laboratory mini-computer, the PDP11/40, in Fortran IV. They are available for distribution and should be upwards expandable to a variety of computer systems.

INTRODUCTION

Chemistry, like other disciplines, has been caught up in the information explosion. Automated methods of data collection have contributed to the generation of numerous files of chemical information by both private and public sources that are becoming increasingly available in machine-readable form. Chemists have expressed interest in accessing information in these files to serve as an additional information source. Yet, existing files are rarely compatible in terms of their data structures.

The lack of compatibility between files is related to the mode of the representation of chemical structural data that is often included in available files. The most widely used representations of chemical structural data appear to be some type of linear notation or form of connectivity. The form of representation presents special problems to the design of chemical retrieval systems.

Although there is interest in accessing currently avail-

[†] Department of Biochemistry and Biophysics, Texas Agricultural Experment Station, Texas A&M University.
* To whom correspondence should be addressed.
** Computer Science Division, Department of Industrial Engineering, Texas A&M University.

able chemical information files, there are major problems in designing systems capable of handling various types of structural data for the retrieval and dissemination of chemical information.

THE INCOMPATIBILITY PROBLEM

The incompatibility of data bases or retrieval systems can be solved through the development of a suitable standardized chemical structural notation, interconversion between notations, or generalized chemical information systems.

A standardized notation acceptable for manipulation by similar systems at various installations might be one solution to the compatibility problem. A standardized notation, however, would probably be too late in developing to gain wide acceptance. The IUPAC notation discussed by Lynch[1] and Dammers[2] was an attempt at standardization that was slow in developing and did not quite meet the demands of the industry.

Another alternative might be the use of interconversion programs. In order to take advantage of the characteristics of some notations or to make efficient use of externally available public and private structural files, some chemical users have already found it convenient to design algorithms for the conversion of one chemical notation to another. The popularity of the WLN (Wiswesser Linear Notation) has led to the development of several conversion algorithms involving this notation. Feldmann,[3] Lynch,[4] and Miller[5] give some insight into this development, and Granito[6] discusses work on a system that will attempt to provide facilities for interconversion between various types of notations. Interconversion, however, has intrinsic problems in that conversion may not always be possible or the conversion process may not always provide an exact translation because of implicit notations. Interconversion from one notation to another, though possible, may not always be the best solution for making use of external files.

A third alternative might be the development of generalized chemical information from various types of data files. Much of the work on information systems, including generalized systems, has been done in the classical document/abstract environment. In this case, input/output specifications for the system are relatively standard. Outside of this environment, however, there are often requirements for different, if not peculiar, input/output specifications. The need for standardization in representing chemical structural information, the growth in the availability of public and private files, the varied needs of users in different operating environments, and the need for substructure searching make generalized chemical systems a practical alternative in the market for storage and retrieval of chemical structural information.

GENERALIZATION

The concept of generalization can be subjected to multiple interpretations. Generalization may be used to describe a system's operating environment capabilities. McGee[7,8] defines a generalized information system as a set of programs performing various functions which can be adapted to specific applications with less time and effort involved than developing equivalent programs from scratch. A simple, but adequate definition that describes the concept of generalization is one given by Morenoff.[9] Three requirements must be satisfied before a system can be considered to be general purpose. The requirements are that the system must (1) be independent of both the data and the structure of the data which may be represented within it; (2) be applications-independent (data base is usable to various kinds of users); and (3) possess some degree of computer hardware independence. A retrieval system should possess some degree of data independence, applications-independence, and hardware independence to be classified as a general system under this definition.

GENERALIZED SYSTEMS

The most noteworthy surveys of the many generalized systems developed are those by Fry[10] and the CODASYL Systems Committee.[11] Another interesting survey is one by Canning Publications, Inc.[12,13] A generalized system that is usable with chemically oriented data files has been designed and implemented at the IIT Research Institute, as described by Williams.[14]

Most of the generalized systems that have been developed are quite large and intended for use on relatively large computer systems. Most are written in machine or assembly language and are not truly general with respect to hardware independence. They are typically designed for use in the classical document/abstract environment and usually employ the sequential, indexed-sequential, or random-access methods of file organization. They are also mainly intended for use in multi-user environments and have limited facilities for making use of computer driven graphic display peripherals that are becoming increasingly useful and important in three-dimensional (3-d) studies of chemical structures.[15]

SYSTEM DESIGN

We report here a modular system for storage and retrieval of chemical information, including structural data. The system possesses a certain degree of hardware-independence, data-independence, and applications-independence, but certain modules that make up the system must be tailored to fit a particular application or accommodate peculiar or unusual input/output requirements, like 3-d graphic displays.

The design of the CRYSRC system takes advantage of certain basic activities that are common to most retrieval systems. Retrieval systems, for example, normally require some facilities for recognizing input, building directories to enhance retrievals, and storing information on appropriate storage devices. They also employ methods for handling queries, searching the directory and retrieving, and some facilities for organizing and transmitting output. Finally, they provide program interfaces that allow the user control over the program. Generally, retrieval systems can be said to consist of a logical organization of activities that perform the following: program control, input acquisition, indexing, information storage, query interpretation, search and retrieval, and output organization.

The CRYSRC system makes use of this separation of activities by incorporating each as a separate module. These modules can then be integrated in an organized fashion such that each has its own input and output characteristics within the overall system. Certain modules can be general to all applications; others, usually input/output related, must be tailored to specific applications.

The system, consisting of nine modules, is diagrammed in Figure 1. It incorporates the basic activities of an information storage and retrieval system. The program control function has been supplemented by modules implementing activities such as security control, error analysis, and statistics collection. The input/output modules can be tailored (as noted by the dashed lines) to fit the application at hand.

The hierarchy of modules that make up this system is related to activities concerned with management, file generation, and information retrieval (Figure 2). The program control module at the top of the hierarchy serves as the interface between users and system components.

Most of the modules are general in that they accomplish

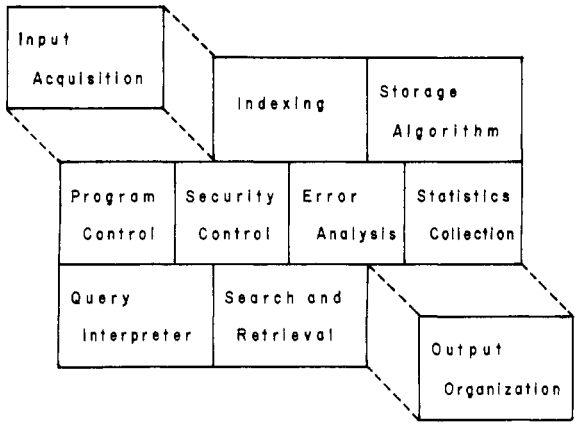

Figure 1. CRYSRC system modules.

activities common to all applications. Input/output related modules can be tailored to accommodate various kinds of chemical structural information or the use of special graphic display peripherals. The modular design of the system makes it possible to tailor or modify particular modules to meet changing requirements with a minimum amount of concern for its effect on the total system.

INITIALIZATION

The two major tasks of any information system are file generation and information retrieval. The CRYSRC system employs the inverted list method of file organization which requires a directory. The directory consists of ordered tables containing pointers to locations of entries in the data file. Search of the directory during the retrieval process is analogous to a search of a library card catalog. File generation generally represents building the data file and/or creating a directory by which to access the data file.

The pointers in the directory are ordered by particular "elements of information" which reveal special characteristics about entries in the data file. These elements of information are commonly referred to as "keys" and they often have "qualifiers" associated with them. For example, a *key/qualifier* pair might be *Name/Doe* or *Occupation/Chemist*. It is possible to reduce qualifiers into smaller subsets that might involve secondary as well as primary qualifiers. Such triplets might include the following: *Name/Doe/John* or *Occupation/Chemist/Organic*. Such triplets facilitate the retrieval of entries in the data file involving *John Does* or *Organic Chemists*. This design permits the use of keys organized as *key/qualifier* pairs or *key/primary qualifier/secondary qualifier* triplets.

The exact manner in which individual keys are employed by the system is determined during an initialization phase prior to file generation. The number and type of keys are specified during this initialization phase. The third approach of key organization used by the system permits "screens" to accompany prime key qualifiers. This screen approach gives the system a substructure-search capability. Screens can be represented as a series of switches that may be on or off depending on the particular structural characteristics that are being represented.

The three key organizational designs just described are the extent of the options permitted by the system. Keys to be used by the system can thus be initialized for use as *key/qualifier* pairs, *key/primary qualifier/secondary qualifier* triplets, or *key/qualifier* pairs with accompanying screens.

The keys initialized for use with the crystallographic chemical file adapted to the system permit retrievals on (1) sequence accession numbers, (2) compound names, (3) molecular formulas, (4) author names, (5) journal entries, and (6) substructure fragments.

Sequence accession numbers (1) form the basis for the first key and were considered important for the purposes of retrieving previously accessed entries without having to repeat the search and selection process. Accession numbers are assigned to each entry during file generation.

A second key permits retrievals on compound names. Developing qualifiers for this key required some study into the nature of chemical names. A search of a sample file and comparisons to terms in a chemical dictionary yielded a list of 400 six-character words and word stems. These were used as qualifiers in the compound name key list.

Qualifiers for molecular formula retrievals (3), however, were generated from the chemical elements appearing in the formulas for the various entries. The system is initialized to take into consideration the natural order of occurrence of chemical elements for this particular key and can thus respond more quickly to queries specifying the most frequently occurring elements. The key defined for the retrieval by molecular formulas uses chemical elements as primary qualifiers and the number of atoms per element as secondary qualifiers.

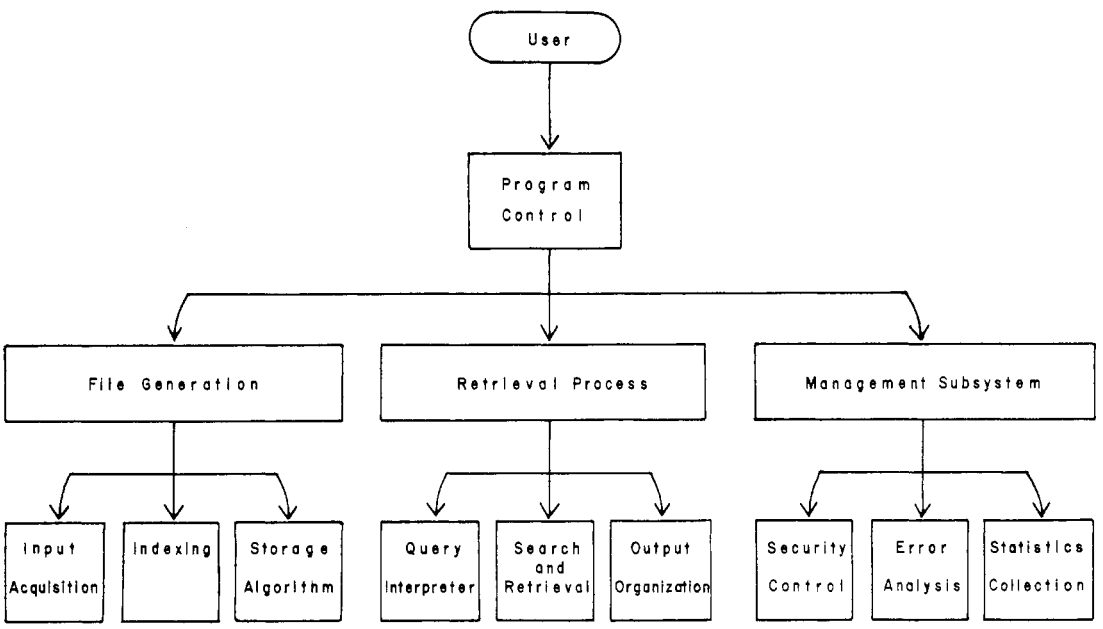

Figure 2. Flow of control.

A fourth key that permits retrieval of entries by author names has also been initialized in the adaptation of the crystallographic chemical file. The surnames are represented as codes and serve as primary qualifiers and the first name initial of the various authors serve as secondary qualifiers.

Another key has been designated to specify journals, which are represented as three-letter numeric codes that have been assigned by the Crystal Data Centre.[15] The secondary qualifier is the year of publication. The journal key may be useful in reducing the number of entries in the output list when broad keyword queries have been specified.

The final key (6) initialized in the adaptation of the Cambridge data file gives the system a substructure-search capability. Predefined substructure fragments are associated with a label and a screen. Inverted lists for labels, used as qualifiers, point to entries containing particular fragments. Retrievals on individual substructure fragments or a logical combination of fragments are thus possible without having to do substructure matching during the retrieval process. The result is relatively fast (<1 min) retrievals on preselected fragments at the expense of an exhaustive search for random fragments.

The cost of an exhaustive search is somewhat reduced by the system's screen matching retrieval mechanism. Since each preselected substructure fragment is associated with a screen, it is possible to match an input query screen with those associated with the various fragment labels in locating fragments that come closest to matching the structural characteristics of the input query substructure. After building a list of possible candidates, it is possible to browse through and edit the list, display particular structures, or seek more specific pattern-matching on a subset of the file through the use of tailored output searching and editing routines.

The exact nature of the structural characteristics represented by the screens used by the system was determined by an analysis of a sample file created from the original crystallographic chemical file. The screens represent frequently occurring elements and strings of elements and to some degree reflect the particular chemical interests of the current users of the chemical system.

The key that has been initialized for substructure searching permits direct retrievals on a number of prechosen fragments. The use of the screen retrieval mechanism permits searches for fragments that have structural characteristics in common with the input query substructure. The nature of the screen type retrieval mechanism may lead to the retrieval of irrelevant material, and the retrieval of all relevant material from the file is not necessarily assured. The screen searching capability of the system is a compromise between generating indices or lists for every possible substructure fragment and an exhaustive search of the entire file for a random fragment.

The adaptation of the crystallographic data base to the system led to the initialization of six keys which make use of the three forms of key organization design available on the system. The acronyms assigned to the various keys during initialization form the basis for specifying particular keys during the implementation of the retrieval process.

IMPLEMENTATION

The CRYSRC system permits (1) accession number, (2) browsing, (3) keyword, and (4) screen searching retrievals. The search and retrieval mechanism for the four alternative methods of retrieval involves building an output list of relevant materials which will later be acted upon by available output organization routines.

The first and simplest approach to retrieval provided by the system is accession number retrieval. To initiate this approach, the user specifies the partition number (group number) and the entry number of the required entry. The approach assumes that the user has had prior access to the system and knowledge of the required numbering information. The method provides easy access to information from previously accessed entries.

The second retrieval approach gives users the opportunity to browse through the individual records that make up entries in the data file. The browsing mode can be entered at any starting location in the data file. Once initiated, users interact with the system to control the display of material, skipping of entries, or saving of entry information. The user can exit from the browsing mode at his discretion.

The third approach to retrieval available on the system relies heavily on keys defined during initialization of the system. It is a two-step process that involves definition of retrieval keys and declaration of a logical combination of keys.

During the definition process, users specify the exact nature of all retrieval keys to include the key acronym, primary qualifiers, and/or secondary qualifiers. Some examples of key definitions might be as follows:

K1 = NMBR:GROUP40; 23
K2 = CMPD:PORPHYRIN;NICKEL;ETHYL
K3 = FRML:Ni;C36;N4;H44
K4 = ATHR: CULLEN,D;MEYER,E
K5 = JRNL: 004,1974
K6 = SSNM: C-N-C;Ni

The system permits the definition of up to eight keys on any particular retrieval.

The next step in the keyword retrieval process is to construct a logical combination of keys to be used during the search process. The system can accommodate ths use of a conjunctive (intersection) or disjunctive (union) combination of keys. The key combination must also consist of a statement in "disjunctive normal form". Statement variables may contain no symbols other than those for conjunction and disjunction and conjunction symbols must occur only between single variables. For example, if one lets an "asterisk" represent a conjunction (logical *and*) and a "plus sign" represent a disjunction (logical *or*), the following would be legal statements.

K1 * K2 * K3
K1 + K2 + K3
K1 + (K2 * K3)
(K1*K2) + (K3*K4*K5)

This latter restriction simplifies the programming problem and helps reduce confusion in the user formulation of keyword retrieval queries.

The final approach to retrievals available on the system employs a screen searching mechanism. To make use of this facility, some key on the system must have been initialized to carry accompanying screens. The approach also requires the use of a user-provided screen generating routine for use during retrievals to generate an input query screen from data-dependent information supplied to the user. In the crystallographic chemical file application, for example, input to the screen generating routine is in the form of the connectivity for the query substructure (appendix). The screen that is generated is the only link between the system and the data-dependent screen generating routine. The query input screen is matched against those of the keytype specified, and the system interacts with the user in the selection process. The screen searching approach is thus very much dependent upon the nature of the screen building process and the quality and range of characteristics represented by the screens in the qualifier lists.

The four methods of retrieval available on the system give users several alternatives for retrieval. Very narrow or very broad queries can be specified. Regardless of the retrieval method employed and the size of the query, the user has the option of processing query results through a flexible set of output processing routines.

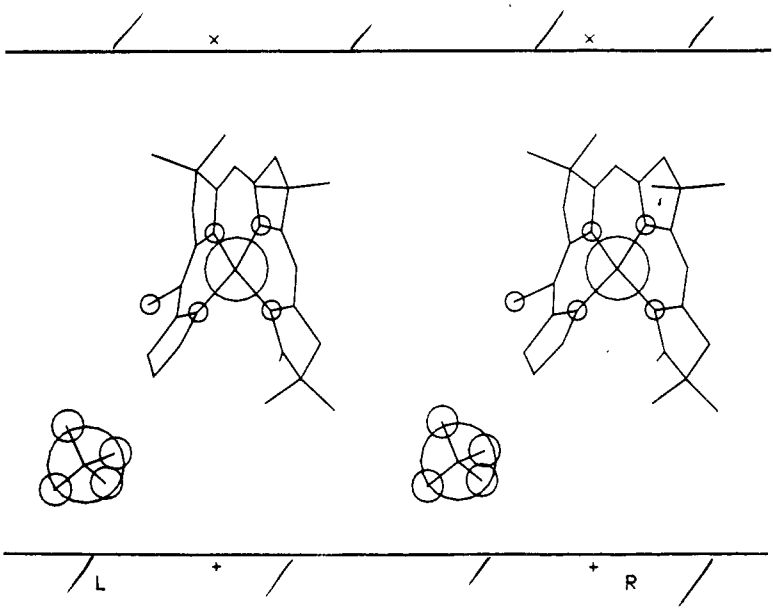

Figure 3. Stereoview of the molecule[17] retrieved by the browse option (cf. Appendix). Viewing may be aided by stereo spectacles or by placing a sheet of paper between the images. Complex images on the CRT are viewed in addition with the aid of a half-silvered mirror and polarizing filters or by continuous rotation, or both.

OUTPUT PROCESSING

The output processing mechanism adds an additional dimension to the system. Submodules of the output processing module can be added and/or modified to meet changing data or hardware requirements without affecting the rest of the system. The system currently has facilities for editing and/or browsing through retrieved lists, creating disk or tape files of relevant material, and processing structural data for display on an available three-dimensional graphics display (Figure 3).[16] The nature of the design of the output processing mechanism makes it possible to tailor output to specific applications or hardware.

Interactive graphics makes complex structural models available to the user and helps overcome a limitation of textual data to describe spatial relationships in a holistic form. The user is then free to ask general geometric questions of a molecular model "suspended" before him in 3-d.

CONCLUSION

This paper has discussed a generalized system for storage and retrieval that has been designed, developed, and implemented on a laboratory minicomputer, the PDP11/40, with 28K words of core memory, 1.2 megaword disk, and an interactive graphics display. It is a modularly designed system that has been adapted to a crystallographic chemical data base. Interest in a system of this nature grew from the desire to retrieve chemically related data, with emphasis on structural data from various kinds of chemical data bases.

The form of chemical structural representation presents special problems to the design of generalized chemical information systems. Factors within the industry have contributed to the evolution of various methods and notations for representing the topology of chemical structures. Use of externally available structural files almost always involves notational or conversion problems. The use of generalized chemical retrieval systems capable of handling structural information from various kinds of data bases is an alternative capable of solving chemical problems arising from diverse notations.

The CRYSRC system has been an experiment in the use of a generalized approach in an application departing from the classical document/abstract environment. It possesses, to some degree, characteristics sometimes used in the definition of generalized systems. It can, for example, be characterized as hardware-independent by virtue of its implementation in a high-level language on a small computer. The code used follows closely the guidelines set by the American National Standards Institute for ANS FORTRAN.[18]

The system's data and application independence can be related to its facilities for handling input and output. Applications departing from the classical document/abstract environment can be adapted to the system by building an input processing module. Peculiar output requirements are handled in a similar fashion. In this way the major portion of the system remains independent of the input data. Implementation of any specific application would only require the design and coding of appropriate input and output processing modules. In view of the growth of computerization, this does not appear to be an unreasonable request in most applications. The facility for handling input and output is especially suited to solving intersystem notational and computer driven display problems in chemical applications.

The system incorporates retrieval facilities that make it suitable for use in a variety of applications. It permits four methods of retrieval which include facilities for accession number, browsing, and keyword retrievals. It also has a screen searching retrieval capability. Screens, representing structural characteristics, are matched between that of the input query and those of the preselected fragment lists. The facility, although dependent on the quality of screens and the type of preselected fragments, makes use of the inverted list approach of file organization. It permits the chemists to enter into the selection process and reduces sequential searching requirements at retrieval time. These four methods of retrieval available on the system permit users to specify from very narrow to very broad queries.

The retrieval mechanism employed at any particular time may result in any number of relevant entries in the output list. Once the output list has been generated, one can execute appropriate output routines to reduce the number in the list. The flexibility of the output processing mechanism makes it relatively simple to add routines for interactive graphical displays or very specific searching or editing on entries in the output list.

This experiment in storage and retrieval has addressed

itself to solving some of the problems inherent in the design of generalized chemical information systems. The result has been a modularly designed general purpose chemical information storage and retrieval system suitable for interactive on-line use. The system is relatively exportable and offers various approaches to retrieval. Trends in the chemical industry with respect to the interest in available private and public structural files, coded in various notations, make flexible systems like the one developed through this experiment a reasonable solution to retrieval problems associated with the dissemination of information in such files.

ACKNOWLEDGMENTS

Assistance from Drs. Olga Kennard, David Watson, and Richard Feldmann is noted, as is the encouragement of Dr. M. F. Lynch. This work has been supported by NSF Grant GJ33248X and the Texas Agricultural Experiment Station.

APPENDIX

The four different retrievals are illustrated by the following examples. The user responses are underlined.

1. Retrieval by accession number

```
COMMAND=*NUMBER

  ACCESSION NUMBER RETRIEVAL
PARTITION NO. =*1

ENTRY NO. =*39

RETRIEVAL OK,
ENTRIES IN OUTPUT LIST = 1
```

```
COMMAND=*NUMBER

  ACCESSION NUMBER RETRIEVAL
PARTITION NO. =*1

ENTRY NO. =*73

RETRIEVAL OK,
ENTRIES IN OUTPUT LIST = 1
```

2. Retrieval by browsing

```
COMMAND=*BROWSE

  BROWSING RETRIEVAL
STARTING POSITION?
PARTITION NO. =*1

RELATIVE POSITION=*3

NO. OF COLUMNS FOR DISPLAY=*72

NO. OF LINES=*5

TYPE 'Y' IF USING CRT=*

CDCARD      349   32Ø211Ø11 Ø Ø Ø Ø 1ØØ  Ø1Ø1 Ø 38  Ø  9  Ø Ø Ø Ø  58Y
COMPND 16/Ø5/72 1  *PLATINUM(II) 15-CYANO-2,2,7,7,12,12-HEXAMETHYL-
COMPND          2  1-METHYLENE-/AS//D-SECO-CORRIN PERCHLORATE
FORMUL        1 1  C27 H34 N5 PT1 +
FORMUL        2 1  CL1 O4 -
TYPE 'C' TO CONTINUE('S' TO SKIP)=*S
END OF ENTRY NO.    3
TYPE 'Y' TO PLACE ENTRY IN OUTPUT LIST('E' TO EXIT)=*Y

OK, ('C' OR 'E')=*C

CDCARD      449   41Ø111Ø11 Ø Ø Ø  1ØØ  Ø111 Ø 31 31  8  0 6 6 2  94Y
COMPND 27/07/72 1  *TETRABENZO-MONOAZA-PORPHINE
FORMUL        1 1  C35 H21 N5
AUTHOR          1  I.M.DAS,B.CHAUDHURI
JRNL            1  ACTA CRYST.(B)                    28   579  1972 1Ø7
TYPE 'C' TO CONTINUE('S' TO SKIP)=*S

END OF ENTRY NO.    4
TYPE 'Y' TO PLACE ENTRY IN OUTPUT LIST('E' TO EXIT)=*Y

OK, ('C' OR 'E')=*E

RETRIEVAL OK,
ENTRIES IN OUTPUT LIST = 2
```

3. Retrieval by keyword

```
COMMAND=*KEYS

  KEYS RETRIEVAL

KEY DEFINITIONS
K1=*FRML:C35

K2=*FRML:N5

K3=*CMPD:PORPHINE

K4=*CMPD:BENZENE

K5=*

KEY COMBINATION

  LIST =*(K1*K2)+(K3*K4)

RETRIEVAL IN PROGRESS
RETRIEVAL OK,
ENTRIES IN OUTPUT LIST = 1
```

4. Retrieval by screen search

```
COMMAND=*USERIN

  USER INPUT RETRIEVAL
KEY TYPE=*SSNM

SCREEN GENERATION ROUTINE
INPUT FROM CR OR KB?=*KB

TYPE ATOM FOLLOWED BY CONNECTIVITY
ATOM  1=*N,2,Ø

Atom  2=*C,3,7,Ø

Atom  3=*C,4,Ø

Atom  4=*C,5,Ø

Atom  5=*C,6,Ø

Atom  6=*C,7,Ø

Atom  7=*C,Ø

Atom  8=*

RETURN TO SYSTEM
RETRIEVAL IN PROGRESS
RESULTS OF SCREEN COMPARES:
   MIN. NO. OF HITS = 1
   MAX. NO. OF HITS = 7
SPECIFY MIN. FOR RETRIEVAL =*7

SPECIFY MAX. FOR RETRIEVAL =*7

RETRIEVAL OK,
ENTRIES IN OUTPUT LIST = 29
```

LITERATURE CITED

(1) Lynch, M. F., Harrison, J. M., Town, W. G., and Ash, J. E., "Computer Handling of Chemical Structure Information", American Elsevier, New York, N.Y., 1971.

(2) Dammers, H. F., and Polton, D. J., "Use of the IUPAC Notation in Computer Processing of Information of Chemical Structures", *J. Chem. Doc.*, **4**, 186–190 (1964).

(3) Feldmann, R. J., and Koniver, D. A., "Interactive Searching of Chemical Files and Structural Diagram Generation from Wiswesser Line Notation", *J. Chem. Doc.*, **11**, 154–9 (1971).

(4) Lynch, M. F., "Conversion of Connection Table Description of Chemical Compounds into a Form of Wiswesser Notation", *J. Chem. Doc.*, **8**, 130–3 (1968).

(5) Miller, G. A., "Encoding and Decoding WLN", *J. Chem. Doc.*, **12**, 60–7 (1972).

(6) Granito, C. E., "CHEMTRAN and the Interconversion of Chemical Substructure Systems", *J. Chem. Doc.*, **13**, 72–4 (1973).

(7) McGee, W. C., "File Structures for Generalized Data Management", *Proc. IFIP Congr. 1968*, 1233–9 (1969).

(8) McGee, W. C., "Generalized File Processing", *Annu. Rev. Automat. Program*, **5**, 77–149 (1969).

(9) Morenoff, E., and McLean, J. B., "On the Design of a General Purpose Data Management System", *Proc. 4th Annu. Natl. Colloq. Inform. Retrieval*, 19–29 (1967).

(10) Fry, J. P., Bramson, S., Fried, D. C., Grabowsky, W. P., Jeffries, J. Jr., Mahle, S. B., and Sternick, H. J., "Data Management Systems Survey", AD 684–707, The Mitre Corp., Washington, D. C., Jan 1969.

(11) CODASYL Systems Committee, "Feature Analysis of Generalized Data Base Management Systems", *Assoc. Comp. Mach.*, New York, N.Y., May 1971.

(12) Canning Publications, Inc., "Trends in Data Management. Part 1", *EDP Analyzer*, **9**, 1–14 (May 1971).

(13) Canning Publications, Inc., "Trends in Data Management. Part 2", *EDP Analyzer*, **9**, 1–13 (June 1971).

(14) Williams, M. E., and Schipma, P. B., "Design and Operation of a Computer Search Center for Chemical Information", *J. Chem. Doc.*, **10**, 158–162 (1970).

(15) Allen, F. H., Kennard, O., Motherwell, W. D. S., Town, W. G., and Watson, D. G., "Cambridge Crystallographic Data Centre. II. Structural Data File", *J. Chem. Doc.*, **13**, 119–123 (1973).

(16) Morimoto, C. N., and Meyer, E. F., "Information Retrieval, Computer Graphics, and Remote Computing", in "International Summer School on Crystallographic Computing—1975", F. R. Ahmed, Ed., Munksgaard, Ltd., Copenhagen, in press.

(17) Currie, M., and Dunite, J. D., "Structural Similarities and Differences between A/D-seco-Corrinoid Complexes of Divalent Nickel, Palladium, and Platinum", *Helv. Chim. Acta*, **54**, 98–112 (1971).

(18) A copy is being prepared for distribution via the Digital Equipment Corp. User's group, DECUS.