Towards an Automatic, Three Dimensional Display of Structural Data*

EDGAR F. MEYER, JR.
Department of Biochemistry and Biophysics, Texas A&M University, College Station, Tex. 77843

Received November 7, 1969

A prototype information retrieval system for searching and displaying models of molecular structures is described. Computer programs to minimize the need for human data handling have been written. A three dimensional graphical model is presented on a color television monitor using atomic coordinates determined by X-ray crystallography. A series of options permit real time manipulation of the model by the operator.

A vast amount of structural information exists in the scientific literature, with more being added each year. If this information is to be readily accessible, it must be stored and retrieved automatically, with the aid of digital computer techniques. The storing problem of organic and inorganic structural information is being solved by the Crystal Data Centre,[1] Cambridge, England, under the direction of Mrs. Olga Kennard. Publications such as "Crystal Data"[2] and "Structure Reports"[3] are bringing together structural data in a coherent form, but publishing three-dimensional representations of molecular structures is still difficult to achieve on a large scale, even though an excellent computer program[4] is finding wide use on an individual basis in the scientific literature.

Computer driven graphical displays such as the *X-Y* plotter, the cathode ray tube (CRT), or the television (TV) monitor are becoming increasingly common as devices able to produce rapid, realistic graphical representations of molecules. As such, they have a vital place in a chemical information retrieval system.

A computer program, DISPLAY,[5] has been written to display graphical (ball and stick) representations of small (200 atoms or less) molecules on a color TV monitor using the BRAD[6] (Brookhaven RAster Display) technique. Each of two stereo views of the model is drawn[7] in a separate color (red or green) on a color TV monitor and viewed through an appropriate filter. Each eye sees only one image, but the mind of the viewer recombines the two views into a three-dimensional whole in the same way that we normally perceive depth visually. The display is flicker-free and does not require a dedicated computer to refresh the TV images. Such a display presents itself as a facile terminal in an information retrieval system.

A computer program, LINK, was written to convert automatically (i.e., with a minimum of human interaction) information stored in the files of the Crystal Data Centre into a form that can be displayed on the color TV monitor at Brookhaven National Laboratory. An eight-character name is assigned to each compound in the library; the user types on the console teletype the library name and reference name of the desired compound. After the operator selects a sequence of options, the desired graphical model is displayed in a matter of seconds.

For the prototype system under discussion here, the author was supplied with data on approximately 50 compounds from the files of the Crystal Data Centre. Program LINK, written for the CDC 6600 and IBM 360/65 computers, interprets the information stored on the *compnd, formul, author, jrnl, class, cdfrml, remark, toler, bondla, cell, symtry, atom* and *bond* cards.[3] (Each card in the structures library has a title at the beginning of the card for identification purposes.) It generates output in a form that can be read by program PICLIB and called by program DISPLAY.

The first 40 characters of the title stored on the *compnd* card are retained for each structure, but other comment cards (*formul ... bondla*) are currently not retained for display. Crystallographic data such as unit cell dimensions (*cell*), space group symbol or symmetry (*symtry*), and atomic name and coordinates (*atom*) are stored for display. The atomic coordinates are orthogonalized and scaled to Angstrom units. With the aid of the atomic symbol in the atom name, a radius is assigned to each atom. Subroutine UMZUG assumes a random order of atoms, resorts the order of the atoms according to chemical bonding, and establishes connectivity tables for the new atomic sequence.

These tables, consisting of "FRØM" and "RING" closure lists, are similar to the connectivity tables[8,9] derived by *Chemical Abstracts*. They should be reducible to the *unique*[9] table of a specific compound stored by *Chemical Abstracts*.

This information, plus the symmetry operators in numeric form, is currently punched on cards to be read by program PICLIB on the SDS Sigma 7 (Σ7) computer in the Bubble Chamber Group of the Physics Department at Brookhaven National Laboratory. This same computer drives the color TV raster display.

Currently, the library is written by PICLIB on the high-speed random access disk of the Σ7 computer for retrieval under operator control. A direct link is being installed so that the Σ7 and the 6600 computer will be in direct communication. Searching of a larger library could then be handled by either machine or by both in conjunction.

*Presented before the Eight International Congress of Crystallography, Stony Brook, N. Y., August 15, 1969. Published in *Acta Cryst.* A25, S3, p. S63.

Atom type is designated by displaying the atom name or drawing circles of specified radii for the different types of atoms. The former is useful for viewing an individual molecule or a portion thereof but is not very practical when the full contents of a crystallographic unit cell are displayed, due to confusion resulting from a high degree of overlap of atomic symbols and bonds. Circles show atom type equally well. The size of the circles is an input option, so that ionic, covalent, Van der Waals, or hydrogen bonding radii may be chosen to help characterize the structure of interest.

The user may choose to display an individual molecule (the crystallographic asymmetric unit) or the full contents of a crystallographic unit cell, the other molecules being related by symmetry operations to the asymmetric unit. This latter option facilitates the study of intermolecular interactions in the solid state. (A limited number of colored pictures and viewing glasses is available from the author.)

For more complicated molecules the picture sometimes becomes complicated and therefore a "zoom" feature may be invoked to display only the contents of a desired volume element, enlarged to fill the screen. With the present version of the program DISPLAY the user may ask for output of bond lengths; a future version will include extended geometric routines for extraction of bond angles, torsion angles, and best-fit planes of specific atoms. These features will be initiated by means of a track ball (X-Y) pointer that allows the operator to indicate the desired atoms individually.

As the title implies, an attempt has been made to produce three-dimensional structural and chemical information in a visual, meaningful manner with a minimum of human interaction in the steps between the library and the console. Scaling up by one or more orders of magnitude will surely present problems that will necessitate changes in strategy and program logic. At this point the essential facts are that a library exists, a real-time terminal with 3-D graphical display exists, and an automatic link has been established between them.

All programs have been written in Fortran IV. Generous support has been given by the College of Agriculture, Texas A&M University, and by the U.S. Atomic Energy Commission, through the hospitality of Brookhaven National Laboratory. The support and encouragement of Walter Hamilton and Olga Kennard is gratefully acknowledged.

LITERATURE CITED

(1) Kennard, O., and D. G. Watson, "Crystal Structure Library Manual," University Chemical Laboratory, Cambridge, U.K. (1969).
(2) Donnay, J. D. H., G. Donnay, E. G. Cox, O. Kennard, and M. V. King, "Crystal Data, Determinative Tables, 2nd Edition" (1963), Monograph 5, American Crystallographic Association.
(3) Pearson, W. B., Ed., "Structure Reports," N. V. A. Oosthoeks Uitgevers Mij., Utrecht.
(4) Johnson, C. K., "ORTEP: A Fortran Thermal-Ellipsoid Plot Program for Crystal Structure Illustrations" (1965). ORNL, 3794.
(5) Meyer, E. F., "Three Dimensional Graphical Models of Molecules and a Time-Slicing Computer," *J. Appl. Crystallog.*, in press.
(6) Ophir, D., S. Rankowitz, B. J. Shepherd, and R. J. Spinrad, *Comm. ACM*, Vol. 11, No. 6, p. 415 (1968).
(7) Ophir, D., B. J. Shepherd, and R. J. Spinrad, *Comm. ACM*, Vol. 12, No. 6, p. 309 (1969).
(8) Whittingham, D. J., F. R. Wetsel, and H. L. Morgan, J. CHEM. DOC. 6, 230 (1966).
(9) Morgan, H. L., J. CHEM. DOC. 5, 107 (1965), cf. p. 110.

A Utility Analysis for the MCC Topological Screen System*†

DAVID LEFKOVITZ**
The Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, Pa. 19104

ALFONSO R. GENNARO
The Philadelphia College of Pharmacy and Science, Philadelphia, Pa.

Received August 18, 1969

This paper is the fourth in a series that has attempted to develop a systematic approach to the problem of searching chemical substructures in a large-scale automated system. The three preceding papers in the series were:

"Use of a Nonunique Notation in a Large-Scale Chemical Information System"[1]

"A Chemical Notation and Code for Computer Manipulation"[2]

"Substructure Search in the MCC System"[3]

The first of these discussed a systems concept which includes the five fundamental requisites of a chemical information storage and retrieval system—viz, structure input, registry, search file generation and storage, substructure search, and structural display. A conclusion of this paper was that a computer-oriented chemical code was a desirable component of the system in order to represent the compound as accurately as a connection table and as concisely as a notation; in addition, this code should be readily convertible to a connection table for the purposes of further computer manipulation, such as screen assignment, atom by atom search and display.

The second paper specified such a language and indicated how it met the requirements of the former paper. This language was called the Mechanical Chemical Code (MCC).

* Presented in part at the 4th Middle Atlantic Regional Meeting, ACS, Washington D. C., February 1969.
† This study was supported by the National Science Foundation under Contract No. NSF-C547.
** Present addres, COMSERV, 4025 Chestnut St., Philadelphia, Pa. 19104.