

# CRYSTALLOGRAPHIC DATA STORAGE, RETRIEVAL AND DISSEMINATION: NATIONAL INFORMATION CENTRE FOR CRYSTALLOGRAPHY [NICRYS] ACTIVITY AT MADRAS

R. SRINIVASAN and E. SUBRAMANIAN

*Department of Crystallography and Biophysics, University of Madras, Guindy Campus,  
Madras 600 025, India.*

**C**RYSTALLOGRAPHY is concerned with the investigation of the atomic and molecular structure of matter. It provides the structural parameters that can be used to describe the three-dimensional atomic arrangement of matter in the solid state. The techniques and results of crystallography are often combined with those of other scientific disciplines, resulting in significant advances in these fields. Crystallography is thus an interdisciplinary science that interacts with physical, chemical and biological sciences, contributing to fundamental progress in all these fields.

Crystallography as a discipline generates large amounts of data. Every substance that forms a crystal is a potential candidate for structural investigation. For each crystal studied, one has the unit cell parameters, symmetry information, cartesian coordinates and isotropic or anisotropic thermal vibration amplitudes for each atom of the molecule constituting the crystal etc. In addition there are the diffraction intensities themselves, numbering from several hundreds to several thousands. These parameters can be used to deduce molecular packing arrangements, electron distributions in bonds, modes of vibrations, types of inter-molecular associations and various other structural features which constitute a major source of information for building a conceptual framework around which the diverse structural principles can be explained.

The last few decades have witnessed an unprecedented increase in the volume of crystallographic literature published annually. It is a stupendous task for scientists to keep abreast of all the latest information and progress in crystallographic research. Moreover, the data appear in a selective form that only provides emphasis and interpretation based on a particular author's chosen field of specialisation as well as on his own point-of-view. The same data may find other uses

in related and inter-disciplinary areas of science, but the data are often scattered over innumerable journals published all around the world making literature survey a really time-consuming task. A new situation has arisen in recent times in that many research institutions, faced with static budgets and escalating journal costs, have cut down on journal subscriptions, making it even more difficult for research workers to have normal access to the literature most vital to their research activities. This often results in unnecessary duplication and repetition of efforts by scientists who need to get at the same data in order to enable them to make scientific inferences or extrapolations using previously available data on related matters.

All these, point to the imperative need for the services of a centralised agency for the codification and continuous up-dating of the data in selected areas of science, as and when such data become available, and computerised data banks offer the best solution to the problems discussed above. In the field of Crystallography, the staff of the Chemistry Department at Cambridge University, U.K., under the leadership of Dr Olga Kennard, have compiled and continue to maintain a data base relating to the structure of organics, organo-metallics and metal complexes as determined by x-ray and neutron diffraction. The Cambridge Centre continues this compilation on global literature basis and makes them available (periodically updated also) in the form of magnetic tapes to other centers all over the world which serve as focal points for local dissemination of data. With financial assistance from DST, a National Crystal Data Centre was established in 1978 at the Department of Crystallography and Biophysics, University of Madras. Currently this centre has been rechristened as NICRYS (National Information Centre for

Crystallography) and forms one of five sectoral centres of NISSAT (National Information System for Science & Technology). This centre has been procuring the data files from Cambridge on an annual basis for use by Indian scientists. The Madras Centre has thus been functioning as an affiliate of the Cambridge Data Centre, for dissemination of Crystallographic data in India. The data files are updated periodically and modified for use with the local computer, and the information on these files is made available to all Indian scientists through computer searches appropriate to their specific needs and requests. It may be mentioned that the Madras Centre is in a way unique in the sense that it deals with 'hard-data' and is probably only one of its kind in the south Asian region.

The Cambridge Crystallographic Database currently contains bibliographic, chemical connectivity and numeric structural data for some 37,367 organics, organometallics and metal complexes. The total data base, together with programmes for search, retrieval, geometric analysis and visual display of stored information is made available to world-wide scientific community *via* Affiliated Data Centres in about 20 countries. These data present scientists with an opportunity for such studies as (i) comparisons with M.O. calculations, (ii) calculation of 'ideal' molecular geometries, (iii) effect of substituents on molecular geometry, (iv) crystal packing, (v) Hydrogen bonding, (vi) Inter-molecular association, (vii) conformational analysis in such areas as 'drug design' and a score of other related aspects.

To illustrate the uses of the database, one could start by asking the question as to how does a newcomer find the information he wants, from the large amount of computerized crystallographic data on tapes. The investigator may be a physicist, chemist or biologist and each will have a different point of view as to what is of interest to him. This brings us to the problem of search and retrieval. The general strategy for search and retrieval of the required information consists in deciding what parameters are likely to be of interest. For any given parameter, a sub-file consisting only of answers to this question is

created. For instance, if one is interested only in nucleosides, one could create a subfile which contains only those compounds which are nucleosides. Such a subfile is much smaller than the primary file and hence subsequent file manipulations can be simplified for such queries regarding the nucleosides as to how many are purine nucleosides, etc.

The database files are divided into three categories: (a) Bibliographic, (b) Connectivity and (c) Data. The Bibliographic files contain compound name, chemical formula, chemical class, Journal where the data had been published, Author, etc. The connectivity file contains the chemical connectivity of the molecules in terms of atoms, bond types etc. The Data file contains unit cell parameters, symmetry information coordinates of non-hydrogen atoms etc.

Copies of these files are at present made available by the Madras Centre to other Indian Institutions which have ready access to computers. Presently, Bhabha Atomic Research Centre at Bombay and the Indian Institute of Science, Bangalore have been supplied with these data files. It may be mentioned in this context that so far 150 scientists from about 40 institutions from all over India have made use of these data facilities and the total jobs cleared so far run to 1000.

Lastly, at the time of this report a national level Workshop on "X-ray Crystallographic Data Storage, Retrieval and Dissemination" is being planned during 28-30 November, 1983. This is jointly sponsored by the CODATA Committee in India (under INSA) and the NISSAT in Department of Science & Technology, Delhi. The idea here is to bring together current and potential users of the data base system for effective interaction and with a view to possible improved data service in future and the possibility of enlargement of circle of potential users. Particular mention may be made of the high scope of its utilisation by chemists, biologists, material scientists besides the brand of 'crystallographers'. It is strongly felt that there is a need to increase the awareness among the former class of scientists. It is hoped that the Workshop will fulfil this objective.