52 02-Methods for Structure Determination and Analysis, Computing and Graphics

Many scientific fields deal with the calculation or observation of 3-dimensional data, for example electron density, Van der Waals surface and electrostatic potential calculations. Conventionally, this data is stored in grid form where a large number of data points are needed to fully describe it, which leads to problems with both storage and analysis.

An alternative approach is to approximate the grid by fitting a set of orthonormal functions to it. Once this has been carried out only the function coefficients need to be stored - i.e. a few hundred numbers instead of several hundred thousand. The 3-d information can then be regained when required by reversing the calculation.

The functions we have chosen are very similar to those used to describe atomic orbitals, they differ only in that they have been scaled to represent a whole molecule and not just the volume associated with a single atom. This technique has been applied previously for electron density display and fitting in the field of Protein Crystallography but has not been used extensively for smaller molecules.

We are particularly interested in the following applications of this technique:-
Crystallography:

The representation of molecular shape and charge for use in crystal packing studies.

The method can be used to represent voids within crystals, something that is difficult to achieve by other means.

Drug Design:

Molecular shapes are important for explaining and/or predicting the interaction of a drug with a putative binding site.

QSAR studies on the coefficients can be easily carried out. The number of parameters involved is relatively small and the nature of the functions allow the use of fast rotation algorithms to aid molecular comparisons.

The applications of the technique are limited only by the types of information that such functions can be expected to represent. If required another set of functions can be used for other cases as the choice is only limited by the need for orthonormality. Also, as the functions are able to represent a complex 3-d grid, two independent sets of data for the same molecule can be dealt with simultaneously. One is entered as though it was the real part of the grid and the other as the imaginary part; the two sets of numbers are kept completely separate during the calculations but may be displayed and analysed together if required.

OCM–02.08.06 MOLECULAR SCENE ANALYSIS: A TOPOLOGICAL APPROACH FOR THE AUTOMATED INTERPRETATION OF PROTEIN ELECTRON DENSITY MAPS

L. Leherte*, S. Fortier and J. Glasgow, Dept. of Computing and Information Science, Dept. of Chemistry, Queen's University, Kingston, Canada K7L 3N6.

As part of our project in Molecular Scene Analysis (Fortier, S., Castleden, I., Glasgow, J. I., Conklin, D., Walmsley, C., Leherte, L. & Allen, F. H. (1993). Acta Cryst. D 49, 168-178), we have been investigating methods to assist in the spatial and visual analysis of electron density maps at varying resolution. In particular, we have assessed the usefulness of the topological approach for the segmentation of medium (3 Å) resolution maps of proteins and their interpretation in terms of structural motifs. We have followed the approach implemented by Johnson (Johnson, C. K. (1977). *ORCRIT. The Oak Ridge Critical Point Network Program.* Chemistry Division, Oak Ridge National Laboratory, USA) in the program ORCRIT, which provides a global representation of the electron density distribution through the location, identification and linkage of its critical points (points where the density gradient vanishes). In the first part of our study, the topological approach was applied to ideal (calculated) maps of three proteins of small to medium size so as to develop a methodology - rules, heuristics or templates - that could then be used for analyzing maps of medium resolution. The methodology was then applied to both calculated and experimental maps of penicillopepsin at 3 Å resolution. The study shows that the networks of critical points provide a useful segmentation of the maps, tracing the protein main chains and capturing their conformation. In addition, these networks can be parsed in terms of secondary structure motifs, through a geometrical analysis of the critical points. The procedure adopted for secondary structure recognition was phrased in terms of geometry-based rules. It provides a basis for an automated implementation through the use of artificial intelligence techniques.

PS–02.08.07 THE USE OF A MODIFIED DOUBLE PATTERSON FUNCTION IN DIRECT PHASE DETERMINATION.

By Yuan-Fang Wang*, René Peschar and Henk Schenk. Laboratory for Crystallography, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands.

Nowadays in X-ray crystallography most single crystal structures are solved by direct methods, which are based on probability theory. From this theory the values of structure invariants and seminvariants are estimated, one of them being the triplet phase sum ψ_t. If interatomic triangles ($\mathbf{r}_K$-$\mathbf{r}_\mu$, $\mathbf{r}_\lambda$-$\mathbf{r}_\mu$) can be obtained, $<\cos(\psi_t)>$ and $<\sin(\psi_t)>$ can be expressed in more accurate formulas. (Kronenburg, *Thesis, Univ. of Amsterdam*, 1992, 67-75). It is commonly known that the Patterson function consists of interatomic vectors. Two arbitrary Patterson vectors, however, do not necessary share a common atom, so the construction of triangles ($\mathbf{r}_K$-$\mathbf{r}_\mu$, $\mathbf{r}_\lambda$-$\mathbf{r}_\nu$) from Patterson vectors is a complicated task. In this respect the double Patterson function (Vaughan, *Acta Cryst.*, 1958, **11**, 111-115):

$$P(\mathbf{u},\mathbf{v}) = V^2 \int_V \rho(\mathbf{r})\ \rho(\mathbf{r}+\mathbf{u})\ \rho(\mathbf{r}+\mathbf{v})\ d\mathbf{r}$$

is more interesting for two related reasons:

1) $P(\mathbf{u},\mathbf{v})$ is the Fourier transform of the triplet phase sum;
2) non-zero $P(\mathbf{u},\mathbf{v})$ are possible only if $\mathbf{u}$ and $\mathbf{v}$ form an interatomic triangle.

A modification of the double Patterson function is proposed. Calculations for both model and real structures are presented which show that interatomic triangles can be constructed more safely with the double Patterson function than with the normal Patterson function.

PS–02.08.08 A PACKAGE OF FAST FOURIER TRANSFORM ROUTINES FOR MACROMOLECULAR CRYSTALLOGRAPHY.

E. Prince*[1], M. An[2], C. Lu[3] and R. Tolimieri[2].

1. Reactor Radiation Division, National Institute of Standards and Technology, Gaithersburg, MD 20899, U. S. A.
2. Aware, Inc., 1 Memorial Drive, 4th Floor, Cambridge, MA 02142, U. S. A.
3. Department of Computer and Information Sciences, Towson State University, Towson, MD 21204, U. S. A.

02-Methods for Structure Determination and Analysis, Computing and Graphics

The extensive use of Fourier transform methods and the large sizes of unit cell and data sets in macromolecular crystallography make it highly desirable to use efficient methods of computation. Fast Fourier transform (FFT) methods are very useful for solution of these problems. With a few exceptions, the libraries of FFT subroutines that have been available in the past have been restricted to sample sizes (numbers of grid points per period) that were products of small, prime numbers, which results in rather large gaps between numbers that may be used for large unit cells and moderately high resolution. Efficient FFT routines for different types of numbers – prime numbers, powers of prime numbers, or products of distinct prime numbers and their powers – follow many different procedures, so it is not possible to write general routines that can handle all cases. Also, when many transforms are required for sample sequences all of the same size, a routine that is optimized for that particular size is desirable.

The crystallographic problem allows exploitation of certain special conditions to make computation still more efficient. Electron density is a real function, so its Fourier transform has Hermitian symmetry. Three dimensional transforms may be performed as sequences of one dimensional transforms, and the data in the intermediate steps may have a symmetric structure. For example, a 2_1 screw axis gives a data structure in which values separated by half of a lattice translation are complex conjugates, and centered lattices produce rows that are alternately symmetric and antisymmetric. Techniques for exploiting space group symmetry for space groups that have no rotation axes of order higher than two have been described by Ten Eyck (*Acta Cryst.*, 1973, A**29**, 183-191), but space group symmetry with three-, four- and six-fold axes may also be used to reduce the computational load.

We have developed a package of FORTRAN routines for one-dimensional FFTs for all odd number sample sizes from 11 to 99 and all even number sample sizes from 20 to 200 that have no prime factors larger than 17. The package also includes transforms from real sequences to Hermitian sequences and from Hermitian sequences to real sequences, as well as between sequences that have translational conjugate symmetry and sequences that have alternating Hermitian symmetry and Hermitian antisymmetry. These routines are utilized by transform routines for a set of space groups that contain more than 80% of reported protein crystals.

PS–02.08.09 STATISTICAL ANALYSIS AND REACTION PATHWAYS OF MOLECULES. By K.Huml* , F.Soldán and W.Hummel. Institute of Macromolecular Chemistry, Academy of Sciences of the Czech Republic, Heyrovského nám.2, 162 06 Praha 6, Czech Republic

A hypothesis concerning possible minimum energy conformations and conformational interconversions can be supported by statistical analysis of similar molecular fragments, potential energy, calculation, topology and other methods (Bürgi H.-B. and Dunitz J.D., Acc.Chem.Res., 1983, 16, 153–161; Mezey P.G. Potential Energy Hypersurfaces, Amsterdam: Elsevier, 1987). Molecular symmetry of a fragment under study gives additional limiting conditions (Frei H., Bauder A. and Günthard H.H., Top.Curr.Chem., 1979, 81, 1–97). It is believed that properly chosen model of a static and/or dynamic behaviour of the molecular fragment leads to a good mutual agreement among results reached by the methods mentioned above. Potential energy minima correspond to chemical entities which represent the most frequently occurring conformations (molecular templates). On the other hand, sporadically scattered conformations bridging stable conformations map the probable reaction pathways during the conformational interconversions.

1,2-dihydroxybenzene (catechol) and 1,2-dimethoxybenzene molecules were exploited to elucidate the procedure. Structural information was retrieved from the Cambridge Structural Databank and two-dimensional potential energy maps were calculated using the molecular mechanics method.

PS–02.08.10 SHELXL-92 A NEW LEAST-SQUARES REFINEMENT PROGRAM FOR USE WITH SINGLE CRYSTAL DIFFRACTION DATA

George M. Sheldrick, Institut für Anorganische Chemie, der Universität, Göttingen, Germany, and Ward T. Robinson*, Department of Chemistry, University of Canterbury, Christchurch, New Zealand.

SHELXL-92 is a FORTRAN-77 program for the refinement of crystal structures from X-ray or neutron diffraction data, and is primarily designed for single crystal data from small structures (1-1000 unique atoms) at atomic resolution. It is intended to be easy to install and use on a wide variety of computers, and replaces the structure-refining part of SHELX-76.

SHELXL-92 is general and efficient for all space groups in all settings and there are no arbitrary limits to the size of problems which can be handled, except for the total memory available to the program. All instructions are in machine independent free format, with extensive use of default settings to minimize the amount of input required from the user. Instructions and data are taken from two standard (ASCII) text files, so that input files can easily be transferred between different computers. SHELXL-92 is a PUBLIC DOMAIN progam; it is provided in source form, also as a precompiled version which has been optimized for Pc's. There are no restrictions on its use or distribution anywhere in the world for non-commercial purposes.

The program produces all the information required for efficient development of a complete structural model from a partially correct one derived using separate direct or vector methods programs such as SHELXS86. It contains many options for fully automatic handling of constraints on positional and thermal parameters and for placement of hydrogen atoms. Structural models can also be restrained to conform with chemically reasonable expectations for interatomic distances and thermal displacement parameters. These similarity restraints can be particularly useful in refinements of macromolecular structure. The atom designation rules, which accommodate standard practices, help minimize the instructions required to invoke constraints and restraints. All tables necessary for electronic mail submission to both Acta Cryst. (C) and the CSD are provided in CIF format and a companion program CIFTAB can produce hard copy in formats suitable for other journals.

PS–02.08.11 'CRYSTALS' in control

D.J.Watkin* & Keith Prout
Chemical Crystallography Laboratory,
9, Parks Road, Oxford. OX1 3PD