Acta Cryst. (1996). D**52**, 826–828

xdlMAPMAN and *xdlDATAMAN* – Programs for Reformatting, Analysis and Manipulation of Biomacromolecular Electron-Density Maps and Reflection Data Sets

GERARD J. KLEYWEGT* AND T. ALWYN JONES

Department of Molecular Biology, Biomedical Centre, Uppsala University, Box 590, S-751 24 Uppsala, Sweden. E-mail: gerard@xray.bmc.uu.se

(Received 24 July 1995; accepted 31 October 1995)

Abstract

Two user-friendly computer programs are described for use in macromolecular X-ray crystallography. *xdlMAPMAN* provides an interface for electron-density map exchange between some of the most commonly used phase refinement, structure refinement and model-building programs. In addition, it contains several options to analyse and abstract such maps. *xdlDATAMAN* provides similar functionality for the analysis and manipulation of macromolecular reflection data sets. Both programs have a simple graphical user interface, and their source code has been put into the public domain.

1. The crystallographic problem

(I) To provide a user-friendly electron-density-exchange interface between most commonly used phase and structure-refinement packages in macromolecular crystallography and (*a*) the programs in the *CCP*4 package (Collaborative Computational Project, Number 4, 1994), and (*b*) the crystallographic model-building program *O* (Jones, Zou, Cowan & Kjeldgaard, 1991). Format-conversion programs are often necessary, but they tend to be jiffy programs which usually do not excel in user-friendliness. Programs to prepare map files suitable for display in *O* (so-called DSN6 files) and to skeletonise electron density for manipulation in *O* suffer from the same defect.

(II) To provide a user-friendly reflection-exchange interface between most commonly used phase and structure refinement packages in macromolecular crystallography, and to provide a user-friendly mechanism for manipulating and analysing reflection data sets.

2. Method of solution

(I) *xdlMAPMAN* is an interactive program with a simple point-and-click X-windows interface, based on the XDL_VIEW toolkit routines (Campbell, 1995). Its major features are as follows.

(*a*) Reading electron-density maps from a variety of sources, supporting *CCP*4, *X-PLOR*, *TNT*, *PHASES* and *O* formats. In addition, some types of molecular masks can be read, as well as so-called '3D-matrix' files from *X-PLOR* rotation and translation functions.

(*b*) Writing *CCP*4 maps. This enables users of other phase-refinement packages to convert their maps into *CCP*4 format and to use *CCP*4 programs.

(*c*) Writing maps suitable for display in *O* (so-called 'mappage' function).

(*d*) Altering the scale of maps (*e.g.*, multiplying, dividing or adding to all electron-density values, automatic scaling to a certain range of density values, and map normalization which transforms the map so that it has an average density of zero and a sigma level of one).

(*e*) Skeletonising the electron density (Greer, 1974) and producing so-called *BONES* files suitable for *ab initio* building of new structures with *O* (S. Thirup & TAJ, unpublished results).

(*f*) Peak-picking of maps (*e.g.*, difference Fourier maps), using an algorithm developed previously to pick peaks in three-dimensional NMR spectra (Kleywegt *et al.*, 1993).

The fact that many different types of map can be read with this program means that only one program for producing *O* maps and *BONES* needs to be maintained.

(II) *xdlDATAMAN* is an interactive program with a simple point-and-click X-windows interface, based on the XDL_VIEW toolkit routines (Campbell, 1995). Some of its features are as follows.

(*a*) Reading and writing ASCII reflection data sets in a variety of formats. For every reflection, the Miller indices and the observed structure-factor amplitude (*F*) are read/written; for most file types, the standard deviation (σ) associated with each *F* is also read/written. In addition, flags indicating a partitioning of the data in test and work reflections for free-*R*-value calculations (Brünger, 1992) are also supported. ASCII reflection files can be read in free or user-defined format and written in default of user-defined formats. In addition, several fixed-format file types are supported (*e.g.*, *X-PLOR*, *TNT*, *SHELX*). CIF-formatted reflection files [containing the Miller indices, *F*, $\sigma(F)$ and free-

R-value test flags] can also be generated with *xdlDATAMAN*.

(*b*) Calculation of the 'effective resolution' of a data set as defined by Bart Hazes (personal communication, 1994): the effective resolution is that resolution at which the actual number of observed reflections in the data set would have constituted a 100% complete data set. In other words, the number of reflections N is proportional to dC^3, where d is the nominal resolution and C is the fractional completeness of the data. The effective resolution, $d_{\rm eff}$, is the resolution for which N reflections corresponds to a complete data set ($C = 1$).

(*c*) Calculating the value of $R_{\rm sym}$ for pairs of reflections *hkl* and *khl*. For some space groups ($P3_x$, $P4_x$, $I4_x$ and $P6_x$), a small value for this R factor may well indicate missed symmetry, for example space group $I422$ instead of $I4$ (Kleywegt, Hoier & Jones, 1996).

(*d*) Deleting unwanted reflections. Reflections can be deleted using cut-off values for F, σ, F/σ or the resolution; alternatively, individual reflections can be deleted by providing their Miller indices. Also, all reflections with either h, k or l odd or even can be deleted.

(*e*) Listing specific reflections. Reflections of interest can be selected in myriad ways, including using cut-off values for F, σ, F/σ or the resolution; in addition, special reflections (*e.g.*, *hhl*, or 0*k*0, *etc.*) and systematic absences can be listed in order to shed light on the nature of one or more (possible) screw axes.

(*f*) Twinning analysis. Following Stanley (1972) and Rees (1980), an analysis option is provided to investigate possible merohedral twinning by means of intensity statistics.

(*g*) Modifying structure factors. Structure factors (and σ values) can be modified by applying a resolution-dependent temperature factor, or by applying a multiplicative or additive factor. In addition, F's can be converted into I's (and *vice versa*) using the approximation $I = F^2$.

(*h*) Modifying indices. Reflections can be mapped into the standard asymmetric unit of reciprocal space appropriate for the Laue group [following *CCP*4 (1994) conventions]. They can also be re-indexed (for instance, if two data sets merge poorly due to different indexing of equivalent axes, *e.g.* in $P4_x$), and sorted by their Miller indices.

(*i*) Selecting reflections for free-*R*-value calculations (Brünger, 1992). Such reflections can be selected at random, in small spheres or in thin resolution shells. The latter two options are aimed at reducing the bias that is introduced by relationships between reflections (*G* function) in the work and test set due to bulk solvent and non-crystallographic symmetry, respectively [as far as we know, Martin Noble (personal communication, 1994) was the first person to point out the effect of such relationships on the free R value; see also Kleywegt & Jones (1995)]. The *spheres* option selects test reflections at random, and then also includes all nearby reflections inside a sphere in reciprocal space with a user-defined radius; this option is suitable for cases where there is no non-crystallographic symmetry. The *shells* option divides the data into resolution bins, and singles out a thin shell from each bin to constitute the test reflections; this method can be used in the presence of non-crystallographic symmetry. In addition, an option is provided to create multiple data sets for complete cross-validation purposes (Brünger, 1992).

3. Software environment

xdlMAPMAN and *xdlDATAMAN* were designed for use on Unix workstations running X-windows. The programs were written in Fortran-77 with some commonly used ('VAX') extensions. The XDL_VIEW subroutine library and an X-windows system are required for the interface; the *CCP*4 subroutine library is required for the reading and writing of *CCP*4 maps.

4. Hardware environment

xdlMAPMAN and *xdlDATAMAN* were developed on a Silicon Graphics Indigo/Elan workstation. An implementation for DEC Alpha/OSF1 workstations exists as well, and no serious difficulties are anticipated when porting either program to other Unix platforms.

The memory size of *xdlMAPMAN* depends mostly on the maximum size of the maps to be accommodated. All major array dimensions are defined in a separate include file which can be modified to suit local needs (for example, to cope with very large maps in the case of virus structures).

The memory size of *xdlDATAMAN* depends mostly on the maximum number of reflections to be accommodated. All major array dimensions are defined in a separate include file which can be modified to suit local needs.

5. Program specification

xdlMAPMAN was designed for use with macromolecular electron-density maps. Most operations take on the order of 1–5 CPUs to complete. Density skeletonisation is the only exception and may require several minutes of CPU time for extremely large maps (or extremely unfortunate parameter choices). The program comprises roughly 10 000 lines of Fortran code, including many comment cards.

xdlDATAMAN was designed for use with macromolecular reflection data sets. Most operations take on the order of 1–5 CPU s to complete. The program comprises roughly 7600 lines of Fortran code, including many comment cards.

6. Documentation

xdlMAPMAN and *xdlDATAMAN* both contain a help option which explains most of the available commands. In addition, there are short ASCII documents (~250 lines for *xdlMAPMAN*, and ~450 lines for *xdlDATAMAN*) which describe the use of these programs. Since both programs have been deposited with the *CCP*4, installation instructions will be given as part of the regular *CCP*4 distribution.

7. Availability

Program documentation and executables for Silicon Graphics and DEC Alpha/OSF1 workstations are available to the entire community *via* anonymous ftp from rigel.bmc.uu.se (directory pub/ccp4). In the near future, the complete source code of *xdlMAPMAN* and *xdlDATAMAN* will be available from the *CCP*4 and further distribution will take place *via* that channel.

We wish to acknowledge the many people in the crystallographic community who throughout the past decade or so have provided (modifications to) subroutines used by *xdlMAPMAN*, in particular Dr P. R. Evans (Cambridge; map I/O), Dr S. Thirup (Aarhus; skeletonisation), and Dr M. Kjeldgaard (Aarhus; brix-mappage functionality).

References

Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
Campbell, J. W. (1995). *J. Appl. Cryst.* **28**, 236–242.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 76—763.
Greer, J. (1974). *J. Mol. Biol.* **82**, 279–301.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kleywegt, G. J. & Jones, T. A. (1995). *Structure*, **3**, 535–540.
Kleywegt, G. J., Hoier, H. & Jones, T. A. (1996). *Acta Cryst.* D**52**, 858–863.
Kleywegt, G. J., Vuister, G. W., Padilla, A., Knegtel, R. M. A., Boelens, R. & Kaptein, R. (1993). *J. Magn. Reson. B*, **102**, 166–176.
Rees, D. C. (1980). *Acta Cryst.* A**36**, 578–581.
Stanley, E. (1972). *J. Appl. Cryst.* **5**, 191–194.