International Tables for Crystallography (2006). Vol. F, Section 25.1.2, pp. 685–687.

25. MACROMOLECULAR CRYSTALLOGRAPHY PROGRAMS

25.1. Survey of programs for crystal structure determination and analysis of macromolecules

BY J. DING AND E. ARNOLD

25.1.1. Introduction

Since the pioneering work of Max Perutz and John Kendrew that yielded the structures of haemoglobin and myoglobin roughly forty years ago, macromolecular crystallography has become a cutting-edge research area of modern molecular biology. There has been a dramatic increase in the number of structures of biological macromolecules determined by X-ray crystallography in the past two decades. The number of new structures of proteins, nucleic acids and their complexes with substrates and/or inhibitors deposited with the Protein Data Bank has been expanding dramatically in the last few years. The complexity of these new structures is also increasing. Knowledge derived from these structural studies is growing at a continually accelerating pace, as is their applicability to diverse problems in science and medicine.

These increases have resulted in part from major advances in the instrumentation, analytical methods and recombinant expression techniques that support macromolecular crystallography, including the utilization of more brilliant light sources (synchrotron radiation), charge-coupled-device (CCD) detectors, cryocrystallography, multi-wavelength anomalous dispersion (MAD) phasing analysis and selenomethionyl proteins. Dramatic enhancement of all aspects of structure determination, including the introduction of powerful computer hardware of increasing capacity and sophisticated computational software, has markedly reduced the time and resources required to determine new structures while increasing the quality and accuracy of the results. Program development has benefited not only from technological advances but also from the development of new theories and algorithms in macromolecular X-ray crystallography.

The burgeoning field of structural genomics is presenting additional opportunities, as well as challenges, for structural biologists. In the near future, the complete map of the human genome will be known, representing a milestone in our ability to describe the natural world. The opportunities provided by knowing the complete human genetic blueprint are myriad across many fields, including biology, chemistry, materials science and medicine. Scientists are seeking answers to a growing number of challenging biological questions and ultimately would like to have access to the complete catalogue of protein structures in living systems, as well as to comprehend protein-folding space. Although it is not currently feasible to determine the structure of every protein, it has been suggested that structure determination of about 10 000 properly chosen proteins should permit reliable modelling of three-dimensional structures for hundreds of thousands of other proteins. X-ray crystallography is likely to produce the majority of structures required to achieve such a goal. More powerful, high-throughput methods are needed to facilitate determination and analysis of the hoard of new structures that will emerge from this initiative.

This article presents a survey of the computational software used most frequently by protein X-ray crystallographers in the structure determination of proteins and nucleic acids. This is not intended to provide complete or comprehensive information about every program on each aspect of protein crystallography, nor is it intended to present a complete compilation of available programs (apologies to those whose programs were not included – this is not meant as a slight!). Also, in cases where programs or program systems are described in articles elsewhere in this volume, only minimal descriptions are given here. Brief annotations on some of the most popular or frequently used programs in the crystallographic community are provided. We have liberally pirated program descriptions from the program authors where possible.

We anticipate that parts of this Chapter will become outdated rapidly, owing to the ceaseless evolution of new methods and proliferation of new programs. Among the most volatile information may be addresses for locating the programs on the internet; judicious use of search engines should facilitate the task of finding updated locations. The reader is also referred to http://www.iucr. org/sincris-top/logiciel/, which contains a compilation of a broad range of programs and software systems in crystallography, structural biology and molecular biology.

The program summaries are grouped somewhat arbitrarily into the following categories:

(1) multipurpose crystallographic program systems (Section 25.1.2);
(2) data collection and processing (Section 25.1.3);
(3) phase determination and structure solution (Section 25.1.4);
(4) structure refinement (Section 25.1.5);
(5) phase improvement and density-map modification (Section 25.1.6);
(6) graphics and model building (Section 25.1.7);
(7) structure analysis and verification (Section 25.1.8); and
(8) structure presentation (Section 25.1.9).

25.1.2. Multipurpose crystallographic program systems

25.1.2.1. *Biological software from the EBI*

The European Bioinformatics Institute (EBI) is a centre for research and services in bioinformatics. The EBI manages databases of biological data including nucleic acid sequences, protein sequences and macromolecular structures. The EBI also maintains an archive for a large collection of free software for molecular biologists, including crystallographic applications.

Location: http://www.ebi.ac.uk/. Operating systems: UNIX, VAX/VMS, MS-DOS and Macintosh. Type: source code and binary. Distribution: free.

25.1.2.2. *BIOMOL*

The *BIOMOL* software suite comprises a set of programs developed by the crystallography group at the University of Groningen, The Netherlands. The program package covers applications for many aspects of the structure determination of macromolecules, including post processing of diffraction data, data merging and scaling, calculation of Fourier and Patterson maps, FFT map inversion, vector search, heavy-atom refinement, solvent flattening, molecular replacement, atomic model refinement, data plotting *etc.*

Location: http://www.xray.chem.rug.nl/Biomol.htm, ftp://rugcbc.chem.rug.nl/. Operating systems: VAX, SGI, Convex, HP, DEC Alpha and LINUX. Type: binary. Distribution: free.

25.1.2.3. *BLANC*

The *BLANC* program suite (Vagin *et al.*, 1998) is a collection of programs used for structure determination of macromolecules by X-ray crystallography. The suite is designed to provide experienced crystallographers and students with a number of simple tools. It contains 'super-programs' that consist of several small programs and utilizes the 'black-box principle' that requires minimum input or intervention from a user.

Location: ftp://ftp.yorvic.york.ac.uk/pub/alexei/blanc/. Operating systems: UNIX, VMS and Windows. Type: source code. Languages: Fortran77. Distribution: free.

25.1.2.4. *CCP4 program suite*

The *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994) is the program package most widely used by X-ray crystallographers in structure determination and analysis of macromolecules. The *CCP*4 suite is an integrated set of programs for protein crystallography developed by close collaboration of crystallographers under an initiative by the UK Biotechnology and Biological Sciences Research Council (formerly the SERC). Some software developed elsewhere is also included. The *CCP*4 suite contains programs for all aspects of protein crystallography, including data processing, data scaling, Patterson search and refinement, isomorphous and molecular replacement, structure refinement, phase improvement and density modification, and presentation of results. Individual program documentation is available, together with a PostScript version of the *CCP*4 manual with content distinct from the program documentation (about 1.5 Mbyte). Runnable example files are also distributed with the suite.

The *CCP*4 program suite is distributed in the source form (mostly Fortran), supported for VMS and various UNIX platforms. The suite is available free to academic institutions, subject to a completed licence form being returned to the *CCP*4 secretary. A charge is made to commercial users, who should contact the *CCP*4 secretary to make arrangements. All charges for the suite are used for *CCP*4 activities.

*CCP*4 holds two-day study weekends on selected topics. There have been several meetings to date; some copies of the proceedings to these meetings are available. *CCP*4 also publishes an occasional newsletter; some recent issues are available by anonymous ftp. Starting from June 1996, newsletters are available in html format.

There is a *CCP*4 listserv at ccp4@dl.ac.uk, which provides a forum for users to discuss problems, report bugs and ask for help. A frequently asked questions (FAQ) list has also been set up. If you have problems either compiling or running *CCP*4 programs then have a look at the problem page, which contains various fixes since the latest release.

Locations: http://www.dl.ac.uk/CCP/CCP4/main.html; http://www.sdsc.edu/Xtal/Xtal.html; ftp://ccp4a.dl.ac.uk/pub/ccp4/; ftp://ftp.sdsc.edu/pub/sdsc/xtal/CCP4/ and ftp://ftp2.protein.osaka-u.ac.jp/mirror/ccp4/ccp4. Operating systems: UNIX, VAX/VMS and LINUX. Type: source code. Languages: Fortran and C. Distribution: free academic.

25.1.2.5. *CNS*

Crystallography & NMR System (*CNS*) (Brünger *et al.*, 1998) is a new program suite for structure determination of macromolecules by X-ray crystallography or solution nuclear magnetic resonance (NMR) spectroscopy. The program has been designed to provide a flexible multi-level hierarchical approach for the most commonly used algorithms in macromolecular structure determination. Highlights include heavy-atom searching, experimental phasing (including MAD and MIR), density modification, crystallographic refinement with maximum-likelihood targets, and NMR structure calculation using NOEs, *J* coupling, chemical shift and dipolar coupling data. *CNS* is the result of an international collaborative effort among several research groups. See Chapter 18.2 and Section 25.2.3 for more details.

Location: http://cns.csb.yale.edu/v1.0/. Operating systems: UNIX, SGI, SUN, DEC Alpha, HP, LINUX and Windows-NT. Type: source code. Languages: Fortran77 and C. Distribution: free academic.

25.1.2.6. *MAIN*

MAIN (Turk, 1995) is an interactively driven suite of programs for molecular modelling, density modification, model refinement and structure analysis.

Locations: ftp://stef.ijs.si/dist/ and http://stef.ijs.si/doc/index.html. Operating system: UNIX. Type: source code. Distribution: minor licence fee for academic users.

25.1.2.7. *PHASES*

PHASES (Furey & Swaminathan, 1997) is a general-purpose package of computer programs. The package contains programs used in all steps of the structure determination of macromolecules using single-crystal diffraction data, including data manipulation, phasing, density modification and averaging, structure refinement *etc.* See Section 25.2.1 for a detailed description.

Location: http://www.imsb.au.dk/~mok/phases/phases.html. Operating systems: SGI, Sun, IBM R6000, ESV and DEC Alpha. Languages: Fortran77 and C. Distribution: free.

25.1.2.8. *PROTEIN*

The *PROTEIN* program package (Steigemann, 1991) is an integrated collection of crystallographic programs designed for the structure determination and analysis of macromolecules. Its applications include: (1) generation and expansion of data files with reflection data; (2) scaling of reflection data from different crystals or films onto a common scale; (3) averaging of the reflection data and elimination of inaccurate or obviously wrong measurements; (4) calculation of Patterson, difference Patterson, Fourier and difference Fourier maps by normal or FFT algorithms; (5) MIR and heavy-atom parameter refinement; (6) listing, contouring and peak searching of 3D maps in all directions of the crystal axes; (7) fast calculation of structure factors from atomic coordinates; (8) statistical supplements, *e.g.* calculation of the distribution of figure of merit, significance of anomalous-dispersion data, crystallographic *R* factor *etc.*; and (9) real-space search methods, *e.g.* self-rotation, cross-rotation and translation functions using Patterson and Fourier maps, rotation of Fourier maps, vector verification as an aid in the interpretation of difference Patterson maps *etc.* The *PROTEIN* program system intentionally does not contain programs for structure refinement or interactive graphics modelling programs.

Location: http://www.biochem.mpg.de/PROTEIN/. Operating systems: UNIX, VAX/VMS, SUN, SGI, EVS and CONVEX. Type: binary. Distribution: free academic.

25.1.2.9. *The Purdue University XTAL Program Library*

The Purdue University *XTAL* Program Library (*PUXTAL*) was developed as part of the macromolecular structure research efforts at Purdue. Since the 1960s, a series of crystallographic computing techniques have been developed at Purdue, and many of the *XTAL* programs have been used extensively in laboratories around the world. These programs cover all aspects of macromolecular crystallography, including data processing, MIR, molecular replace-

ment, electron-density modification, structure refinement and structure comparison, and include many utility programs.

Location: http://www-structure.bio.purdue.edu/∼kvz/ (also includes references for individual programs). Operating systems: IBM RS/6000 and UNIX. Type: source code and binary. Distribution: free.

25.1.2.10. *SOLVE*

SOLVE (Terwilliger & Berendzen, 1999) is a complete program package designed for automated crystallographic structure solution for MIR and MAD. *SOLVE* can carry out all steps of macromolecular structure determination automatically using MIR and MAD methods, ranging from scaling data to calculation of an electron-density map. It scales data, solves Patterson functions, calculates difference Fourier maps, searches native Fourier maps for distinct solvent and protein regions, and scores partial MAD and MIR solutions to build up a complete solution.

Locations: http://www.solve.lanl.gov/, ftp://solve.lanl.gov/pub/solve. Operating systems: SGI, SUN, HP, DEC and LINUX. Type: binary. Distribution: minor licence fee for academic users.

25.1.2.11. *USF*

The Uppsala Software Factory (*USF*) comprises a large collection of programs written by Dr Gerard Kleywegt at Uppsala University. These programs have applications in many aspects of structure determination and analysis, including electron-density modification, multiple crystal forms and protein domain averaging, structure validation, error detection and recognition of spatial motifs in protein structures, and includes many utility programs and interface programs for program *O*. See Chapter 17.1 for more details.

Location: http://alpha2.bmc.uu.se/∼gerard/usf/. Operating systems: UNIX and VAX/VMS. Type: binary. Distribution: free academic.

25.1.2.12. *X-PLOR*

X-PLOR (Brünger *et al.*, 1987; Brünger, 1992) is an integrated program package for structure determination of macromolecules using X-ray crystallography and NMR. The main features of *X-PLOR* related to X-ray crystallography include: (1) crystallographic refinement by the simulated-annealing method; (2) rigid-body refinement; (3) conventional positional refinement; (4) refinement of individual *B* factors, group *B* factors and overall anisotropic *B* factors; and (5) analysis of macromolecular structures. The new release, *X-PLOR*98, includes maximum-likelihood refinement as well.

Locations: for *X-PLOR*98, http://www.msi.com/; for *X-PLOR*3.851, http://xplor.csb.yale.edu/xplor-info/. Operating system: UNIX. Type: source code and binary. Distribution: commercial.

25.1.2.13. *Xtal*

The *Xtal* system (Hall *et al.*, 1999) is a comprehensive package of crystallographic software for structure determination, including applications for manipulation of diffraction data, structure solution, structure refinement, structure analysis and presentation of crystal structures. These programs are applicable to X-ray, neutron and electron diffraction analyses, including charge-density studies. The package contains a number of interactive graphics tools and is distributed as execution modules for most commonly available workstations and PCs.

Locations: http://www.crystal.uwa.edu.au/xtal/; ftp://ftp.crystal.uwa.edu.au/xtal. Operating systems: UNIX, VMS and Windows. Type: binary. Language: Fortran77. Distribution: commercial.

25.1.2.14. *XtalView*

XtalView (McRee, 1993) is a crystallographic software package for fitting electron-density maps and solving crystal structures of macromolecules by MIR and MAD methods. Applications include graphics, visualization, virtual reality, modelling and structure determination. It has a simple but comprehensive Windows-based interface. The main menu drives a suite of crystallographic modules by clicking on icons. Standard file formats are used, which facilitate communication between *XtalView* and programs such as *X-PLOR*, *TNT* and *MERLOT*.

Location: http://www.scripps.edu/pub/dem-web/toc.html. Operating systems: UNIX, SGI, SUN, DEC, IBM and LINUX. Type: source code and binary. Distribution: free academic.

25.1.3. Data collection and processing

25.1.3.1. *DPS*

The Data Processing Suite (*DPS*) (Rossmann & van Beek, 1999) is a complete package for processing X-ray diffraction data from crystals of proteins, viruses, nucleic acids and other large biological complexes. The emphasis is on diffraction data collected using synchrotron sources. Currently *DPS* consists of dps_index and dps_scale, and uses some of the programs from the *MOSFLM*/*CCP*4 suite. The dps_index program uses Fourier analysis for the automatic indexing of oscillation images. The dps_scale program uses a scaling method that does not depend on the exclusive use of full reflections. See Chapters 11.1 and 11.5 for more details.

Location: http://ultdev.chess.cornell.edu/MacCHESS/DPS. Operating systems: UNIX, SGI and LINUX. Type: binary. Distribution: free academic.

25.1.3.2. *HKL*

The *HKL* program package (Otwinowski & Minor, 1996) is a complete set of data-processing programs for the analysis of X-ray diffraction data collected from single crystals. The package comprises three components: *XDISPLAY* for graphical visualization of the diffraction image; *DENZO* for autoindexing, reduction and integration of diffraction data; and *SCALEPACK* for scaling and merging of intensities from multiple images. See Chapter 11.4 for more details.

Location: http://www.hkl-xray.com/. Operating systems: SGI, DEC Alpha, SUN and HP-UX. Type: binary. Distribution: commercial.

25.1.3.3. *LOCSCL*

LOCSCL (Blessing, 1997) is a program used to optimize statistically local scaling of single-isomorphous-replacement and single-wavelength anomalous-scattering data.

Location: e-mail blessing@hwi.buffalo.edu. Operating systems: UNIX and Windows. Type: source code. Language: Fortran77. Distribution: free.

25.1.3.4. *MOSFLM*

MOSFLM is a general-purpose data-processing package developed by Dr Andrew Leslie at the MRC, England. The programs have two main applications: (1) determination of crystal orientation, cell parameters and possible space group; and (2) autoindexing of images, generation of reflection lists and integration of diffraction spots. *MOSFLM* is distributed as part of the *CCP*4 suite and runs on multiple platforms. See Chapters 11.2 and 11.3 for more details.

Location: ftp://ftp.mrc-lmb.cam.ac.uk/. Operating systems: UNIX and VAX/VMS. Type: source code and binary. Distribution: free academic.