International Tables for Crystallography (2006). Vol. F, Chapter 25.1, pp. 685–694.

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.

25.1.3.5. *SCALA*

The *SCALA* program (P. R. Evans, 1993, 1997) scales together multiple observations of reflections, and (optionally) merges multiple observations into an averaged intensity. Various scaling models are implemented. The scale factor is a function of the primary beam direction, either as a smooth function of φ (the rotation angle), or expressed as batch (image) number. In addition, the scale may be a function of the secondary beam direction derived from the spatial coordinates of the measured spot on the detector. In this case, the scaling is an interpolated three-dimensional function similar to that described by Kabsch (1988). The merging algorithm analyses the data for outliers and gives detailed analyses. It generates weighted means of the observations of the same reflection, after rejecting the outliers.

Location: *SCALA* is part of the *CCP*4 suite (Section 25.1.2.4). Operating systems: UNIX and VAX/VMS. Type: source code and binary. Distribution: free academic.

25.1.3.6. *STRATEGY*

STRATEGY (Ravelli *et al.*, 1997) is a program that aids in designing data-collection strategy. It is used to determine the optimal starting spindle angle, using a one-circle diffractometer with a 2D detector, in X-ray data collection from crystals of macromolecules. The input file of the program contains information, from a *DENZO* intensities x-file, about starting crystal orientation and cell parameters. The program simulates all the reflections that can occur during 360° rotation of the crystal, determines if reflections can be recorded on the detector, sorts them, provides pictures of the needed oscillation range as a function of the starting spindle angle for given degrees of completeness of the data set and produces redundancy tables for the shortest data collection possible for each desired completeness. However, neither mosaicity nor overlaps are taken into account. This program has been integrated into the *MOSFLM* package (Section 25.1.3.4).

Locations: http://www.crystal.chem.ruu.nl/distr/strategy.html; ftp://ftp.chem.uu.nl/. Operating system: UNIX. Type: binary. Distribution: free.

25.1.4. Phase determination and structure solution

25.1.4.1. *AMoRe*

AMoRe (Navaza, 1994) is a program package that carries out structure determination using molecular replacement. It reformats the data from the new crystal form, generates structure factors from the model, calculates rotation and translation functions, and applies rigid-body refinement to the solutions. *AMoRe* is part of the *CCP*4 suite (Section 25.1.2.4).

Location: http://www.dl.ac.uk/CCP/CCP4/dist/html/INDEX.html. Operating systems: UNIX, VAM/VMS and LINUX. Type: source code and binary. Distribution: free academic.

25.1.4.2. *GLRF*

GLRF (Tong & Rossmann, 1990) is a program that calculates the general locked rotation function. The self-rotation function determines noncrystallographic symmetry in a crystal. The cross-rotation function determines the orientation relationship of a structure in one unit cell to similar structures in another cell. Since the relationship between the assumed molecular symmetry axes is 'locked', the program can greatly enhance the signal peaks on the rotation function. Therefore, it is much more powerful for assemblies with high local symmetry, such as icosahedral viruses. *GLRF* is part of The Purdue University *XTAL* Program Library (*PUXTAL*: Section 25.1.2.9).

Location: http://www-structure.bio.purdue.edu/~kvz/#GLRF. Operating system: UNIX. Type: source code and binary. Distribution: free.

25.1.4.3. *HEAVY*

The *HEAVY* program package contains the *HEAVY* and *HASSP* programs. The package can carry out heavy-atom search, refinement and MIR/MAD phasing. Some of the major features of *HEAVY* include correlated phasing, Bayesian weighting and Bayesian difference refinement.

Location: http://www.iucr.org/sincris-top/logiciel/prg-heavy.html or e-mail Ncameron@lanl.gov. Operating systems: UNIX and VMS. Type: binary. Distribution: free academic.

25.1.4.4. *MADSYS*

MADSYS (Hendrickson, 1991) is a software package developed over the years in Dr Wayne Hendrickson's laboratory for determining experimental phases of macromolecular structures by multi-wavelength anomalous diffraction (MAD). The package consists of a set of programs that carry out MAD data handling, determination of anomalous-scatterer sites, refinement of MAD sites, MAD phases calculation and structure refinement.

Location: http://convex.hhmi.columbia.edu/hendw/madsys/madsys.html. Operating system: UNIX. Type: binary. Distribution: free academic.

25.1.4.5. *MLPHARE*

MLPHARE is a program for maximum-likelihood heavy-atom refinement and phase calculation. This program refines heavy-atom parameters and error estimates, then uses these refined parameters to generate phase information. The maximum number of heavy atoms that may be refined is 130 over a maximum of 20 derivatives. The program was originally written for MIR, but may also be used for phasing from MAD data, where the different wavelengths are interpreted as different 'derivatives'. *MLPHARE* is part of the *CCP*4 suite (Section 25.1.2.4).

Location: http://www.dl.ac.uk/CCP/CCP4/dist/html/mlphare.html. Operating systems: UNIX, VAX/VMS and LINUX. Type: source code and binary. Distribution: free academic.

25.1.4.6. *Shake-and-Bake*

Shake-and-Bake (*SnB*) (Weeks & Miller, 1999) is a program that uses a dual-space direct-methods phasing algorithm based on the minimal principle to determine crystal structures of macromolecules. The program requires very high resolution data to 1.2 Å or better and $|E|$ values as input. *SnB* has been used to solve structures with more than 600 atoms in the asymmetric unit. Recently, *SnB* has also been used to determine the Se sites in large selenomethionyl-substituted proteins. See Chapter 16.1 for more details.

Location: http://www.hwi.buffalo.edu/SnB/. Operating systems: UNIX, VMS and LINUX. Type: source code. Language: Fortran77. Distribution: free.

25.1.4.7. *SHARP*

SHARP (Statistical Heavy-Atom Refinement and Phasing; de La Fortelle & Bricogne, 1997) operates on reduced, merged and scaled data from SIR(AS), MIR(AS) and MAD experiments, refines the heavy-atom model, helps detect minor or disordered sites using likelihood-based residual maps, and calculates phase probability distributions for all reflections in the data set. See Chapter 16.2 for more details.

Location: http://Lagrange.mrc-lmb.cam.ac.uk/sharp/SharpHome.phtml. Operating systems: IRIX and OSF1. Type: binary. Distribution: free academic.

25.1.5. Structure refinement

Several program packages that are used for structure refinement are described in Section 25.1.2. These include *CNS*, *X-PLOR*, *BIOMOL*, *PHASES* and *PROTEIN*. See Section 25.1.2 for further information.

25.1.5.1. *ARP/wARP*

The Automated Refinement Procedure, *ARP/wARP* (Lamzin & Wilson, 1993, 1997), is a program package for automated model building and refinement of protein structures. It combines, in an iterative manner, reciprocal-space structure-factor refinement with updating of the model in real space to construct and improve protein models. *ARP/wARP* can also be used for *ab initio* structure solution of metalloproteins at high resolution. *ARP/wARP* is distributed as part of the *CCP*4 suite (Section 25.1.2.4). See Section 25.2.5 for a detailed description.

Location: http://www.embl-hamburg.de/ARP/. Operating systems: UNIX, HPUX, IRIX and LINUX. Type: source code and binary. Language: Fortran77. Distribution: free academic.

25.1.5.2. *MULTAN88*

*MULTAN*88 (Main *et al.*, 1980) is a program that uses direct methods to determine crystal structures from single-crystal diffraction data. It can be used for very high resolution structure refinement and determination of heavy-atom positions.

Location: http://www.msc.com/. Operating systems: UNIX and VAX/VMS. Type: binary. Distribution: commercial.

25.1.5.3. *PROLSQ*

PROLSQ (Hendrickson & Konnert, 1979) is used for the restrained least-squares refinement of a protein structure. Prior to running *PROLSQ*, the program *PROTIN* must be run to analyse the protein geometry and produce an output file containing restraints information. *PROLSQ* cannot calculate structure factors. Use *SFALL* to calculate X-ray contributions to the matrix. *PROLSQ* is distributed as a unsupported program of the *CCP*4 suite (Section 25.1.2.4).

Location: http://www.dl.ac.uk/CCP/CCP4/dist/. Operating systems: UNIX, VAX/VMS and LINUX. Type: source code and binary. Distribution: free academic.

25.1.5.4. *REFMAC*

REFMAC (Murshudov *et al.*, 1997, 1999) is a macromolecular refinement program which has been integrated into the *CCP*4 suite (Section 25.1.2.4). *REFMAC* can carry out rigid-body, restrained or unrestrained refinement against X-ray data, or idealization of a macromolecular structure. It minimizes the coordinate parameters to satisfy either a maximum-likelihood or least-squares residual. There are options to use different minimization methods. If the user wishes to invoke geometric restraints, the program *PROTIN*, which analyses the protein geometry and produces an output file containing restraints information, must be run prior to running *REFMAC*. *REFMAC* also produces an MTZ output file containing weighted coefficients for *SIGMAA*-weighted mFo-DFcalc and 2mFo-DFcalc maps, where 'missing data' have been restored.

Location: http://www.dl.ac.uk/CCP/CCP4/dist/html/refmac.html. Operating systems: UNIX, SGI, SUN, DEC and LINUX. Type: source code and binary. Distribution: free.

25.1.5.5. *RSRef*

RSRef (Chapman, 1995) is a package of programs that enables an atomic model to be optimized by fitting to an electron-density map. *RSRef* uses an electron-density function that is resolution dependent, so that it accurately models a medium-resolution map. When combined with *TNT*'s (Section 25.1.5.8) *Geometry*, full stereochemical refinement is possible. *RSRef* can be used to quickly prerefine a protein structure during or after model building, or to completely refine structures with high noncrystallographic symmetry that have good electron density.

Location: http://www.sb.fsu.edu/∼rsref/. Operating systems: SGI and EVS. Type: source code and binary. Distribution: minor licence fee for academic users.

25.1.5.6. *SHELX97*

SHELX (Sheldrick & Schneider, 1997) is a set of programs for crystal structure determination from single-crystal diffraction data. Originally *SHELX* was intended only for small molecules. However, improvements in computing performance and data-collection methods have led to increased use of *SHELX* for macromolecules, especially the location of heavy atoms from isomorphous and anomalous-difference data, and the refinement of proteins against high-resolution data (2.5 Å or better). See Section 25.2.10 for a detailed description.

Location: http://shelx.uni-ac.gwdg.de/SHELX/. Operating systems: UNIX, VMS, DOS and Windows. Type: binary. Language: Fortran77. Distribution: free academic.

25.1.5.7. *SIR97*

*SIR*97 (Altomare *et al.*, 1999) is an integrated program package for the determination and refinement of small-molecule structures from single-crystal diffraction data. It is also useful in solving the heavy-atom positions in protein structure determination.

Location: http://www.ba.cnr.it/IRMEC/Sir_Waremain.html. Operating systems: UNIX, VMS, MacOS and Windows. Type: binary. Distribution: free academic.

25.1.5.8. *TNT*

TNT (Tronrud *et al.*, 1987; Tronrud, 1997) is a general-purpose program package for the structure refinement of macromolecules using single-crystal X-ray diffraction data. It is normally used to optimize a model to X-ray diffraction data while maintaining proper stereochemistry using least-squares function-minimization techniques. It can restrain a model to bond lengths, bond angles, dihedral angles, pseudo-rotation angles, planarity and non-bonded 'close' contacts (including symmetry-related contacts). A principal advantage of the *TNT* package is its great flexibility, making it ideal for restraining structures that contain cofactors, inhibitors, or nucleic acids. The package is composed of separate programs, each performing clearly defined tasks. To use the package with other forms of data you simply write programs that produce the value and first derivative of the functional term you wish to minimize. See Section 25.2.4 for a detailed description.

Location: http://www.uoxray.uoregon.edu/tnt/welcome.html. Operating systems: UNIX, VAX/VMS, DEC Alpha, EVS, AIX, SUN and SGI. Type: source code and binary. Distribution: free academic.

25.1.6. Phase improvement and density-map modification

25.1.6.1. *BUSTER*

BUSTER (Bricogne, 1997*a*,*b*) is a program for recovering missing phase information by Bayesian inference. *BUSTER* has applications

in maximum-likelihood refinement of partial structures in conjunction with the *TNT* program (Section 25.1.5.8), maximum-entropy structure completion for missing or ambiguous parts of a structure, and accurate electron-density reconstruction based on high-resolution X-ray diffraction data. *BUSTER* is related to *SHARP* (Section 25.1.4.7). See Chapter 16.2 for more details.

Location: http://lagrange.mrc-lmb.cam.ac.uk/buster/Buster Home.phtml. Operating systems: IRIX and OSF1. Type: binary. Distribution: free academic.

25.1.6.2. *DM/DMMULTI*

DM (Cowtan, 1994) is a density-modification program package. *DM* applies real-space constraints based on known features of a protein electron-density map in order to improve the approximate phasing obtained from experimental sources. Various information can be applied, including the following diverse elements: solvent flattening, histogram mapping, multi-resolution modification, NCS averaging, skeletonization and Sayre's equation. *DM* is part of the *CCP*4 suite (Section 25.1.2.4). See Chapter 15.1 and Section 25.2.2 for more details.

Operating systems: UNIX, VAX/VMS and LINUX. Type: source code and binary. Distribution: free academic.

25.1.6.3. *FINDNCS*

FINDNCS (Lu, 1999) is a program that automatically determines NCS operations from heavy-atom sites to aid in applying averaging techniques in the MIR procedure. The program outputs the NCS operations (a rotation matrix and translation vector), r.m.s. deviations, polar angles and screw distance, matching sites and other useful information for users. The program can also generate files that can be used to display NCS operations using the program *O* (Section 25.1.7.7).

Location: http://gamma.mbb.ki.se/~guoguang/findncs.html. Operating systems: UNIX, IRIX and OSF1. Type: binary. Language: Fortran77. Distribution: free academic.

25.1.6.4. *RAVE*

RAVE (Jones, 1992; Kleywegt & Jones, 1994) is a suite of programs for real-space electron-density averaging of crystallographic electron density between single and multiple domains, and between single and multiple crystal forms. It also contains tools for the detection of secondary-structure elements in macromolecular electron-density maps. See Chapter 17.1 for a detailed description.

Location: http://xray.bmc.uu.se/usf/menu.html#sof; ftp://xray.bmc.uu.se/. Operating systems: UNIX, SGI and DEC Alpha/OSF1. Type: binary. Distribution: free.

25.1.6.5. *SOLOMON*

SOLOMON (Abrahams & Leslie, 1996) is a program that modifies electron-density maps by averaging, solvent flattening and protein truncation. It can also remove overlapped parts of a mask between itself and its symmetry equivalents. *SOLOMON* is part of the *CCP*4 suite (Section 25.1.2.4).

Location: http://www.dl.ac.uk/CCP/CCP4/dist/html/solomon.html. Operating systems: UNIX, VAX/VMS and LINUX. Type: source code and binary. Distribution: free academic.

25.1.6.6. *SQUASH*

The *SQUASH* program (Zhang & Main, 1990*a*,*b*) provides a tool for phase refinement and extension of macromolecular structures. The starting point is a set of native structure factors to some resolution, and estimated phases and figures of merit for some subset of the phases. The result is a set of improved phases and figures of merit for the whole data set. The program combines Sayre's equation with density modification by histogram matching, solvent flattening and noncrystallographic symmetry averaging. The real-space formulation enables any electron-density constraint to be applied easily, *e.g.* solvent flattening with (eventually) known regions of density. The least-squares solution of a large system of nonlinear equations is achieved by Newton–Raphson iteration that converts the system of nonlinear equations into linear ones. The system of linear equations is solved by the conjugate-gradient method using FFTs.

Location: http://www.msc.com/brochures/software/squash.html. Operating system: UNIX. Type: binary. Distribution: commercial.

25.1.7. Graphics and model building

25.1.7.1. *AMBER*

AMBER (Assisted Model Building with Energy Refinement; Cornell *et al.*, 1995) is a molecular-dynamics and energy-minimization program. *AMBER* refers to two things: a molecular-mechanical force field for the simulation of biomolecules (which is in general use in a variety of simulation programs) and a package of molecular-simulation programs which includes source code and demonstrations.

Location: http://www.amber.ucsf.edu/amber/amber.html. Operating systems: UNIX, SGI, SUN *etc.* Type: source code and binary. Languages: Fortran and C. Distribution: commercial.

25.1.7.2. *CHARMM*

CHARMM (Chemistry at HARvard Molecular Mechanics; Brooks *et al.*, 1983; MacKerell *et al.*, 1998) is a program for macromolecular simulations, including energy minimization, molecular dynamics and Monte Carlo simulations.

Location: http://yuri.harvard.edu/. Operating systems: UNIX, SGI, SUN *etc.* Type: source code. Language: C. Distribution: minor licence fee for academic users.

25.1.7.3. *Insight II*

Insight II is a 3D graphical environment for molecular modelling. *Insight* II creates, modifies, manipulates, displays and analyses molecular systems and related data, and provides the core requirements for all *Insight* II software modules. Its powerful user interface enables the seamless flow of data between a wide range of scientific applications. The *Insight* II environment integrates builder modules, development tools, force fields, simulation and visualization tools with tools specifically developed for applications in the life and materials sciences.

Location: http://www.msi.com/life/products/insight/index.html. Operating systems: SGI and IBM UNIX systems. Type: binary. Distribution: commercial.

25.1.7.4. *MidasPlus*

MidasPlus (formerly *Midas*) (Ferrin *et al.*, 1988) is an advanced molecular-modelling system developed by the Computer Graphics Laboratory (CGL) at the University of California, San Francisco. The system can be used for display and manipulation of macromolecules such as proteins and nucleic acids. Ancillary programs allow for such features as computation of molecular surfaces and electrostatic potentials and generation of publication-quality space-filling images with multiple light sources and shadows. To address the needs of the structure-based drug-design community, *MidasPlus* has been developed with an emphasis on the interactive

selection, manipulation and docking of drugs and receptors. Although quite powerful in this application, the system is also somewhat specialized in this respect: it requires three-dimensional atomic coordinate data for the structures being displayed and expects the primary structure to be based on linear chains of subunits such as amino acids or nucleic acids. Using *MidasPlus* for complex inorganic compounds or large polymers with many cross-links is another application.

Location: http://www.cgl.ucsf.edu/Outreach/midasplus/. Operating systems: SGI, DEC Alpha, NeXT, IBM RS6000 and LINUX. Type: source code. Distribution: minor licence fee for academic users.

25.1.7.5. *MODELLER*

MODELLER (Sali & Blundell, 1993) is most frequently used for homology or comparative modelling of protein three-dimensional structure. The user provides an alignment of a sequence to be modelled with known related structures and *MODELLER* will automatically calculate a full-atom model. More generally, *MODELLER* models protein 3D structure by satisfaction of spatial restraints. In principle, the restraints can be derived from a number of different sources. These include homologous structures (comparative modelling), NMR experiments (NMR refinement), rules of secondary-structure packing (combinatorial modelling), cross-linking experiments, fluorescence spectroscopy, image reconstruction in electron microscopy, site-directed mutagenesis, intuition, residue–residue and atom–atom potentials of mean force, *etc*. The output of *MODELLER* is a 3D structure of a protein that satisfies these restraints as well as possible. The optimization is carried out by the variable-target function procedure employing methods of conjugate gradients and molecular dynamics with simulated annealing. *MODELLER* can also do several other tasks, including multiple comparison of protein sequences and/or structures, clustering, and searching of sequence databases. *MODELLER* is also available as part of *QUANTA* (Section 25.1.7.8), *Insight* II (Section 25.1.7.3) and *Weblab GeneExplorer*.

Location: http://guitar.rockefeller.edu/modeller/modeller.html; ftp://guitar.rockefeller.edu/pub/modeller/. Operating system: UNIX. Type: source code and binary. Language: Fortran. Distribution: free academic.

25.1.7.6. *MOLMOL*

MOLMOL is a molecular graphics program for display, analysis and manipulation of three-dimensional structures of biological macromolecules, with special emphasis on nuclear magnetic resonance (NMR) solution structures of proteins and nucleic acids. *MOLMOL* has a graphical user interface with menus, dialogue boxes and online help. The display possibilities include conventional presentations, as well as novel schematic drawings, with the option of displaying different presentations in one view. Covalent molecular structures can be modified by addition or removal of individual atoms and bonds. The three-dimensional structure can be manipulated by interactive rotation about individual dihedral angles. Special efforts were made to allow for appropriate display and analysis of sets of (typically 20–40) conformers that are conventionally used to represent the result of an NMR structure determination, using functions for superimposing sets of conformers, calculation of root-mean-square-distance (r.m.s.d.) values, identification of hydrogen bonds, checking and displaying violations of NMR constraints, and identification and listing of short distances between pairs of hydrogen atoms.

Location: http://www.mol.biol.ethz.ch/wuthrich/software/ molmol/. Operating systems: UNIX, IRIX, AIX, OSF1 and LINUX. Type: source code and binary. Distribution: free.

25.1.7.7. *O*

O (Jones *et al.*, 1991) is a general-purpose macromolecular-modelling package. The program is aimed at scientists with a need to model, build and display macromolecules. Unlike other molecular-modelling programs, such as *FRODO* (Section 25.1.7.10), *O* is a graphical display program built on top of a versatile database system. All molecular data are kept in this database, in a predefined data structure. However, any data can be stored in the database. Data produced by associated stand-alone programs can be stored very easily in the database and used by the program, for example for colouring of atoms. The powerful macro facility of *O* enables the user to customize the use of the program to satisfy his or her specific needs. The current version of *O* is mainly aimed at the field of protein crystallography, bringing into use several new tools which ease the building of models into electron density, allowing it to be done faster and more correctly. Notably, some new auto-build options greatly enhance the speed of building and rebuilding molecular models. See Chapter 17.1 for a detailed description.

Locations: http://kaktus/imsb.au.dk/∼mok/o/; ftp://xray.bmc.uu.se/. Operating systems: UNIX and ESV. Type: binary. Distribution: free academic.

25.1.7.8. *QUANTA*

QUANTA is an extensive library of crystallographic software programs that streamline and accelerate protein structure solution. *QUANTA* provides a powerful and comprehensive modelling environment for 2D and 3D modelling, simulation and analysis of macromolecules and small organic compounds.

Location: http://www.msi.com/life/products/quanta/index.html. Operating system: SGI. Type: binary. Distribution: commercial.

25.1.7.9. *SYBYL*

SYBYL is a comprehensive computational tool kit for molecular design and analysis, with a special focus on the creation of new chemical entities. *SYBYL* provides essential construction and analysis tools for both organic and inorganic molecular structures. It is especially useful in building the structures of ligands, substrates and inhibitors.

Location: http://www.tripos.com/software/sybyl.html. Operating system: UNIX. Type: binary. Distribution: commercial.

25.1.7.10. *Turbo FRODO*

FRODO (Jones, 1978) is a general-purpose molecular-modelling program which can be used to model *de novo* macromolecules, polypeptides and nucleic acids from experimental 3D data obtained from X-ray crystallography and NMR, and to display the resulting models using various representations including van der Waals and Connolly molecular dot surfaces, as well as spline surfaces. *Turbo FRODO* is designed for ligand fitting and protein stacking. The user can interactively mutate a protein or chemically modify it, and evaluate the resulting conformational changes. There are several versions of *FRODO* around the scientific community. For LINUX and HPUX use the *Turbo FRODO* X version.

Location: http://afmb.cnrs-mrs.fr/TURBO_FRODO/turbo.html. Operating systems: HPUX, IRIX and LINUX. Type: binary. Distribution: commercial.

25.1.8. Structure analysis and verification

25.1.8.1. *DSSP*

The *DSSP* program was designed by Kabsch & Sander (1983) to standardize secondary-structure assignment. The DSSP database is a

database of secondary-structure assignments (and much more) for all protein entries in the Protein Data Bank.

Location: http://www.sander.embl-heidelberg.de/dssp/. Operating systems: SGI, SUN, DEC and LINUX. Type: source code and binary. Language: C. Distribution: free.

25.1.8.2. *HBPLUS*

HBPLUS (McDonald & Thornton, 1994) is a hydrogen-bond calculation program. It can calculate the geometries of all hydrogen bonds; optionally list neighbour interactions; calculate hydrogen-atom positions; deal with hydrogen atoms that can occupy more than one position; optionally include amino–aromatic hydrogen bonds; support full customization, such as hydrogen-bond criteria and donor- and acceptor-atom types; analyse hydrogen bonding near Asn, Gln and His side chains and suggest optimal conformations, *etc*.

Location: http://www.biochem.ucl.ac.uk/∼mcdonald/hbplus/home.html. Operating systems: UNIX and VAX/VMS. Type: binary. Language: C. Distribution: free academic.

25.1.8.3. *Molecular Surface*

The *Molecular Surface* program package (Connolly, 1983) comprises four individual programs. *MSRoll* reads Protein Data Bank format atomic coordinate files, computes a dot surface, computes piecewise quartic molecular surfaces, identifies and characterizes interior cavities, computes molecular areas and volumes, and computes a polyhedral surface. *MSDraw* renders molecular surfaces and chemical models, generates polyhedral surface plots with hidden-line elimination, and generates contours on a polyhedral molecular surface. *MSForm* measures the curvature of a polyhedral molecular surface, computes a solvent-excluded density and computes an interfacial surface between two densities. *MSTran* converts vet files and density files to SGI Inventor files, and converts a density file to an SGI Inventor file. See Section 22.1.2 for more details.

Location: http://www.biohedron.com/msp.html. Operating system: UNIX. Type: source code. Language: C. Distribution: commercial.

25.1.8.4. *MSMS*

The *MSMS* program (Sanner *et al.*, 1995) is designed to compute the triangulation of solvent-excluded surfaces very efficiently. This program can be used as a stand-alone program or as an AVS (Advanced Visualization System) module.

Location: http://www.scripps.edu/pub/olson-web/people/sanner/html/msms_home.html. Operating systems: SGI, SUN, DEC Alpha and HP 9000. Type: binary. Language: C. Distribution: free.

25.1.8.5. *NACCESS*

NACCESS is a stand-alone program that calculates the accessible area of a molecule from a PDB (Protein Data Bank) format file. It can calculate the atomic and residue accessibility for both proteins and nucleic acids. The program uses the Lee & Richards (1971) method, whereby a probe of given radius is rolled around the surface of the molecule, and the path traced out by its centre is the accessible surface. Typically, the probe has the same radius as water (1.4 Å) and hence the surface described is often referred to as the solvent-accessible surface. The calculation makes successive thin slices through the 3D molecular volume to calculate the accessible surface of individual atoms. The output from the program can also be read in by *HBPLUS* (Section 25.1.8.2) and *LIGPLOT* (Section 25.1.9.2).

Location: http://sjh.bi.umist.ac.uk/naccess.html. Operating systems: UNIX, SGI, Sun, HP, DEC and LINUX. Type: source code. Language: Fortran77. Distribution: free academic.

25.1.8.6. *NAOMI*

NAOMI (Brocklehurst & Perham, 1993) is a computer program that is aimed at both specialist and non-specialist researchers who make use of three-dimensional structures of proteins in their work. Some applications of the program include: automatic 'key' residue identification; automatic hydrophobic core/packing analysis; automatic hydrogen-bond calculations; high-quality energy calculations; automatic secondary-structure (α-helix, β-strand and turn) classification using fuzzy logic; automatic super-secondary-structure classification (β-hairpin loops); conformational parameter calculations; solvent-accessibility calculations; automatic identification of disulfide bonds, salt bridges and chain breaks; side-chain modelling and manipulation applying symmetry operators; automatic structure repair (building in missing atoms); NMR structure-refinement module; and interfaces to graphics programs [*MOLSCRIPT* (Section 25.1.9.3), *Raster3D* (Section 25.1.9.7), *Insight* (Section 25.1.7.3), *QUANTA* (Section 25.1.7.8)].

Location: http://www.psynix.co.uk/products/naomi/index.html. Operating system: IRIX. Type: binary. Distribution: free.

25.1.8.7. *PASS*

PASS (Putative Active Sites with Spheres; Brady & Stouten, 2000) is a simple computational tool that uses geometry to characterize regions of buried volume in proteins and to identify positions likely to represent binding sites based upon the size, shape and burial extent of these volumes. *PASS*'s utility as a predictive tool for binding-site identification was tested by predicting known binding sites of proteins in the PDB using both complexed macromolecules and their corresponding apoprotein structures. The results indicated that *PASS* can serve as a front-end to fast docking. The main utility of *PASS* lies in the fact that it can analyse a moderate-size protein (∼30 kDa) in under 20 s, which makes it suitable for interactive molecular modelling, protein-database analysis and aggressive virtual screening efforts. As a modelling tool, *PASS*: (1) rapidly identifies favourable regions of the protein surface; (2) simplifies visualization of residues modulating binding in these regions; and (3) provides a means of directly visualizing buried volume, which is often inferred indirectly from curvature in a surface representation. *PASS* produces output in the form of standard PDB files, which are suitable for any modelling package, and provides script files to simplify visualization in *Cerius*2, *Insight* II (Section 25.1.7.3), *MOE*, *QUANTA* (Section 25.1.7.8), *RasMol* (Section 25.1.9.6) and *SYBYL* (Section 25.1.7.9).

Location: http://www.delanet.com/∼bradygp/pass/. Operating systems: SGI, SUN and LINUX. Type: binary. Distribution: free.

25.1.8.8. *PROCHECK*

PROCHECK is a widely used program for checking the stereochemical quality of a protein structure. The aim of *PROCHECK* is to assess how normal, or, conversely, how unusual, the geometry of the residues in a given protein structure is, as compared with stereochemical parameters derived from well refined high-resolution structures. *PROCHECK* is part of the *CCP*4 suite (Section 25.1.2.4). See Section 25.2.6 for a detailed description.

Location: http://www.biochem.ucl.ac.uk/∼roman/procheck/procheck.html. Operating systems: UNIX, VAX/VMS and Windows. Type: source code. Distribution: free.

25.1.8.9. *ProFit*

ProFit is designed to be the ultimate least-squares fitting program and is written to be as easily portable between systems as possible. It performs the basic function of fitting one protein structure to another, but allows as much flexibility as possible in this procedure. Thus one can specify subsets of atoms to be considered or specify zones to be fitted by number, sequence, or by sequence alignment. The program

will output an r.m.s. deviation and, optionally, the fitted coordinates. R.m.s. deviations may also be calculated without actually performing a fit. Zones for calculating the r.m.s. deviation can be different from those used for fitting.

Location: http://www.biochem.ucl.ac.uk/~martin/programs/index.html. Operating system: SGI. Type: binary. Language: C. Distribution: free.

25.1.8.10. *PROSA*

The *PROSA* (PROtein Structure Analysis) program is a useful tool in protein structure research. *PROSA* supports and guides your studies aimed at the determination of a protein's native fold. It is helpful for experimental structure determinations and modelling studies.

Locations: http://www.iucr.org/sincris-top/logiciel/prg-prosa.html, ftp://Gundi.came.sbg.ac.at/pub/Prosa/. Operating systems: SGI and DEC Alpha. Type: binary. Distribution: free academic.

25.1.8.11. *SARF*

SARF (Spatial ARangement of backbone Fragments; Alexandrov, 1996) can perform a search for similar structural motifs in a list of structures or analyse a new structure that is not in the PDB. Comparison of the protein backbone can provide more information than classification of protein structures, because it can reveal unexpected local similarities important for protein function.

Location: http://www-lmmb.ncifcrf.gov/~nicka/prerun.html. Operating systems: SGI and DEC. Type: binary. Distribution: free.

25.1.8.12. *SQUID*

The program *SQUID* (Oldfield, 1992) was developed for the graphical display of information and the analysis of data. Major applications of the program are the analysis of protein structures and molecular-dynamics simulations.

Location: http://www.ysbl.york.ac.uk/~oldfield/squid/. Operating systems: UNIX, SGI, SUN, VAX, DEC and DOS. Type: source code and binary. Language: Fortran. Distribution: free.

25.1.8.13. *STAMP*

The *STAMP* program package comprises 15 programs for alignment and analysis of three-dimensional structures of protein molecules. The program package has the following applications: (1) fast alignment and superimposition of two or more protein structures; (2) generation and display of superimposed 3D structures of protein molecules, as well as sequence alignments; (3) comparison of a protein 3D structure to a database of other protein structures; (4) direct interface to *MOLSCRIPT* (Section 25.1.9.3) and *ALSCRIPT* drawing programs; and (5) a clear method for assigning which regions within a family of proteins are structurally equivalent, without the need for graphical intervention.

Location: http://www.iucr.org/sincris-top/logiciel/prg-stamp.html or e-mail gjb@bioch.ox.ac.uk. Operating system: UNIX. Type: binary. Distribution: free academic.

25.1.8.14. *SURFNET*

SURFNET (Laskowski, 1995) is a program that generates molecular surfaces, cavities and intermolecular interactions from coordinate data files in PDB format. These molecular surfaces and void regions can be visualized graphically.

Location: http://www.biochem.ucl.ac.uk/~roman/surfnet/surfnet.html; ftp://ftp.biochem.ucl.ac.uk. Operating system: UNIX. Type: source code and binary. Distribution: free academic.

25.1.8.15. *WHAT CHECK*

WHAT CHECK (Rodriguez *et al.*, 1998) is a free subset of protein verification programs from the *WHAT IF* package (Section 25.1.8.16).

Location: http://www.sander.embl-heidelberg.de/whatcheck/. Operating systems: SGI and OSF1. Type: source code. Distribution: free.

25.1.8.16. *WHAT IF*

WHAT IF (Vriend, 1990) is a versatile protein structure analysis program that can be used for mutant prediction, structure verification, molecular graphics *etc*. The program makes extensive use of structural databases, permitting diverse query possibilities in structural analysis.

Location: http://www.cmbi.kun.nl/whatif/index.html. Operating systems: UNIX and Windows. Type: binary. Distribution: minor licence fee for academic users.

25.1.9. Structure presentation

25.1.9.1. *GRASP*

GRASP (Nicholls *et al.*, 1991) is a molecular visualization and analysis program. It is particularly useful for the display and manipulation of the surfaces of molecules and their electrostatic properties. Its particular strength compared to other such programs is its facility for surfaces and electrostatics. The program contains extremely rapid algorithms for the construction of rendered molecular surfaces and for solving the Poisson–Boltzmann equation. *GRASP*'s surface can be molecular or accessible and can be colour-coded by electrostatic potential derived from its internal Poisson–Boltzmann solver or external programs such as *DelPhi*. This representation has become a standard tool in assessing electrostatic character of large, typically protein, molecules. Surfaces can also be coloured by other properties, such as any of those of the underlying atoms (*e.g.* hydrophobicity) or by its own intrinsic properties, such as local curvature. The program also contains several other unique data-representation forms in addition to standard ones such as ball-and-stick for atoms and bonds, and backbone splines, or 'worms', to indicate secondary structure. See Chapter 22.3 for more details.

Location: http://honiglab.cpmc.columbia.edu/grasp/. Operating system: IRIX. Type: binary. Distribution: commercial.

25.1.9.2. *LIGPLOT*

The *LIGPLOT* program (Wallace *et al.*, 1995) automatically generates schematic diagrams of protein–ligand interactions for a given PDB file. The interactions shown are those mediated by hydrogen bonds and by hydrophobic contacts.

Location: http://www.biochem.ucl.ac.uk/bsm/ligplot/ligplot.html. Operating systems: UNIX, IRIX and LINUX. Type: source code. Language: C. Distribution: free academic.

25.1.9.3. *MOLSCRIPT*

MOLSCRIPT (Kraulis, 1991) is a program for creating schematic or detailed molecular-graphics images in the form of PostScript plot files from molecular 3D coordinates, usually, but not exclusively, of protein structures. Possible representations are simple wire models, CPK spheres, ball-and-stick models, text labels and Jane Richardson-type schematic drawings of proteins, based on atomic coordinates in various formats. Colour, greyscale, shading and depth cueing can be applied to the various graphical objects. See Section 25.2.7 for a detailed description.

Location: http://www.avatar.se/molscript/. Operating system: UNIX. Type: source code and binary. Language: C. Distribution: free academic.

An enhanced variant of *MOLSCRIPT*, called *BOBSCRIPT*, has been developed by Robert Esnouf (http://orval.rega.kuleuven.ac.be/~robert/Bobscript/). In addition to the functions provided by *MOLSCRIPT*, *BOBSCRIPT* can generate an input file automatically and allows for the display of electron density.

25.1.9.4. *NUCPLOT*

NUCPLOT (Luscombe *et al.*, 1997) is a program which generates schematic diagrams of protein–nucleic acid interactions. The program automatically identifies these interactions from the 3D atomic coordinates of the complex from a PDB file and generates a plot that shows them in a clear and simple manner.

Location: http://www.biochem.ucl.ac.uk/~nick/nucplot.html. Operating systems: UNIX, IRIX, LINUX and Windows. Type: source code. Language: C. Distribution: free academic.

25.1.9.5. *ORTEP*

The Oak Ridge Thermal Ellipsoid Plot (*ORTEP*, version III) program (Burnett & Johnson, 1996) is a computer program for drawing crystal-structure illustrations. Ball-and-stick type illustrations of a quality suitable for publication are generated with either spheres or thermal-motion probability ellipsoids, derived from anisotropic temperature-factor parameters, on the atomic sites. The program also produces stereoscopic pairs of illustrations that aid in the visualization of complex arrangements of atoms and their correlated thermal-motion patterns.

Location: http://www.ornl.gov/ortep/ortep.html. Operating systems: UNIX, LINUX, DOS, MacOS and Windows. Type: source code and binary. Language: Fortran77. Distribution: free.

25.1.9.6. *RasMol*

RasMol is a molecular-graphics program intended for the visualization of proteins, nucleic acids and small molecules. The program is aimed at display, teaching and generation of high-quality images for publication. It is easy to use and produces beautiful space-filling three-dimensional colour images. *RasMol* reads in molecular coordinate files in a number of formats and interactively displays the molecule on the screen in a variety of colour schemes and representations. The X Windows version of *RasMol* provides optional support for a hardware dials box and accelerated shared memory rendering (*via* the XInput and MIT-SHM extensions) if available.

Location: http://www.umass.edu/microbio/rasmol/. Operating systems: UNIX, VAX/VMS, Windows and MacOS. Type: source code. Distribution: free.

25.1.9.7. *Raster3D*

Raster3D (Bacon & Anderson, 1988; Merritt & Murphy, 1994; Merritt & Bacon, 1997) is a set of tools for generating high-quality raster images of proteins or other molecules. The core program renders spheres, triangles and cylinders with special highlighting, Phong shading and shadowing. It uses an efficient software *Z*-buffer algorithm that is independent of any graphics hardware. Ancillary programs process atomic coordinates from PDB files into rendering descriptions for pictures composed of ribbons, space-filling atoms, bonds, ball-and-stick *etc.* *Raster3D* can also be used to render pictures composed in Per Kraulis' program *MOLSCRIPT* (Section 25.1.9.3) in glorious 3D with highlights, shadowing *etc.* Output is pixel image files with 24 bits of colour information per pixel.

Location: http://www.bmsc.washington.edu/raster3d/raster3d.html. Operating systems: DEC, SGI, ESV, SUN, IBM, HP and LINUX. Type: source code and binary. Distribution: free.

25.1.9.8. *Ribbons*

Ribbons software (Carson & Bugg, 1986; Carson, 1997) interactively displays molecular models, analyses crystallographic results and creates publication-quality images. Space-filling and ball-and-stick representations, dot and triangular surfaces, electron-density-map contours, and text are also supported. Input atomic coordinates are in Protein Data Bank (PDB) format. Output may be produced in the Inventor/VRML format.

Location: http://www.cmc.uab.edu/ribbons/. Operating systems: UNIX, LINUX and PC. Type: source code and binary. Distribution: commercial.

25.1.9.9. *SETOR*

SETOR (S. V. Evans, 1993) is designed to render high-quality raster images of macromolecules that can undergo rotation and translation interactively. *SETOR* can render standard all-atom and backbone models of proteins or nucleic acids, but focuses on displaying protein molecules by highlighting elements of secondary structure. The program has a very friendly user interface that minimizes the number of input files by allowing the user to interactively edit parameters such as colours, lighting coefficients and descriptions of secondary structure *via* mouse-activated dialogue boxes. The choice of polymer-chain representation can be varied from standard vector models and van der Waals models, to a beta-spline fit of polymer backbones that yields a smooth ribbon, and to strict Cardinal splines that interpolate the smoothest curve possible that will precisely follow the polymer chain. The program provides a photograph mode, save/restore facilities, and efficient generation of symmetry-related molecules and packing diagrams. Additionally, *SETOR* is designed to accept commands and model coordinates from standard output. Ancillary programs provide a method to edit interactively hardcopy plots of all vectors and many solid models generated by *SETOR*, and to produce standard HPGL or PostScript files.

Location: http://flint.biochm.uottawa.ca/~setor_docs/. Operating system: SGI. Type: binary. Distribution: commercial.

25.1.9.10. *VMD*

VMD (Visual Molecular Dynamics) is designed for the visualization and analysis of biological systems such as proteins, nucleic acids, lipid bilayer assemblies *etc*. It may be used to view more general molecules, as *VMD* can read standard PDB files and display the structure contained in them. *VMD* provides a wide variety of methods for rendering and colouring a molecule: simple points and lines, CPK spheres and cylinders, licorice bonds, backbone tubes and ribbons, cartoon drawings, and others. *VMD* can be used to animate and analyse the trajectory of a molecular-dynamics (MD) simulation. In particular, *VMD* can act as a graphical front end for an external MD program by displaying and animating a molecule undergoing simulation on a remote computer.

Location: http://www.ks.uiuc.edu/Research/vmd/allversions. Operating systems: SGI, SUN, DEC Alpha, IBM AIX, HP-UX and LINUX. Type: binary. Distribution: free.

Acknowledgements

We are grateful to Millie Georgiadis, Kalyan Das and Deena Oren for helpful suggestions.

standard amino acids are based on those of Engh & Huber (1991). *SHELXPRO* is also used to prepare the *name.ins* file for a new refinement job based on the results of the previous refinement (possibly modified by an interactive graphics program such as *XtalView*) and to prepare data for PDB deposition. In addition, the refinement results can be summarized graphically in the form of PostScript plots.

25.2.10.6. *Distribution and support of SHELX*

The *SHELX* system is available free to academics and, for a small licence fee, to commercial users. The programs are supplied as Fortran77 sources and as precompiled versions for Linux and some other widely used operating systems. The programs, examples and extensive documentation may be downloaded by ftp or (if necessary) supplied on CD ROM. Details of new developments, answers to frequently asked questions, and information about obtaining and installing the programs are available from the SHELX homepage, http://shelx.uni-ac.gwdg.de/SHELX/. The author is always interested to receive reports of problems and suggestions for improving the programs and their documentation by e-mail (gsheldr@shelx.uni-ac.gwdg.de).

Acknowledgements

KDC (Section 25.2.2) acknowledges the support of the UK BBSRC (grant No. 87/B03785). KYJZ (Section 25.2.2) acknowledges the National Institutes of Health for grant support (GM55663).

For Section 25.2.3, support by the Howard Hughes Medical Institute and the National Science Foundation to ATB (DBI-9514819 and ASC 93-181159), the Natural Sciences and Engineering Research Council of Canada to NSP, the Howard Hughes Medical Institute and the Medical Research Council of Canada to RJR (MT11000), the Netherlands Foundation for Chemical Research (SON–NWO) to PG and the Howard Hughes Medical Institute to LMR is gratefully acknowledged.

Significant contributors to the programs in the *PROCHECK* suite (Section 25.2.6) include David K. Smith, E. Gail Hutchinson, David T. Jones, J. Antoon C. Rullmann, A. Louise Morris and Dorica Naylor. Part of the development work was funded by a grant from the EU Framework IV Biotechnology programme, contract CT96-0189.

Development of the programs described in Section 25.2.8 for research use was supported by NIH grant GM 15000 and by an educational leave from Glaxo Wellcome Inc. for J. Michael Word; development for teaching use was supported by NSF grant DUE-9980935.

References

25.1

Abrahams, J. P. & Leslie, A. G. W. (1996). *Methods used in the structure determination of bovine mitochondrial F_1 ATPase. Acta Cryst.* D**52**, 30–42.

Alexandrov, N. N. (1996). *SARFing the PDB. Protein Eng.* **9**, 727–732.

Altomare, A., Burla, M. C., Camalli, M., Cascarano, G., Giacovazzo, C., Guagliardi, A., Moliterni, A. G. G., Polidori, G. & Spagna, R. (1999). *SIR97: a new tool for crystal structure determination and refinement. J. Appl. Cryst.* **32**, 115–119.

Bacon, D. J. & Anderson, W. F. (1988). *A fast algorithm for rendering space-filling molecule pictures. J. Mol. Graphics*, **6**, 219–220.

Blessing, R. H. (1997). *LOCSCL: a program to statistically optimize local scaling of single-isomorphous-replacement and single-wavelength-anomalous-scattering data. J. Appl. Cryst.* **30**, 176–177.

Brady, G. P. Jr & Stouten, P. F. W. (2000). *Fast prediction and visualization of protein binding pockets with PASS. J. Comput.-Aided Mol. Des.* **14**, 383–401.

Bricogne, G. (1997*a*). *Ab initio macromolecular phasing: a blueprint for an expert system based on structure factor statistics with built-in stereochemistry. Methods Enzymol.* **277**, 14–19.

Bricogne, G. (1997*b*). *Efficient sampling methods for combinations of signs, phases, hyperphases, and molecular orientations. Methods Enzymol.* **276**, 424–448.

Brocklehurst, S. M. & Perham, R. N. (1993). *Prediction of the three-dimensional structures of the biotinylated domain from yeast pyruvate carboxylase and of the lipoylated H-protein from the pea leaf glycine cleavage system: a new automated method for the prediction of protein tertiary structure. Protein Sci.* **2**, 626–639.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). *CHARMM: a program for macromolecular energy minimization and dynamics calculations. J. Comput. Chem.* **4**, 187–217.

Brünger, A. T. (1992). *X-PLOR version 3.1: a system for X-ray crystallography and NMR*. New Haven: Yale University Press.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Crystallography & NMR System: a new software suite for macromolecular structure determination. Acta Cryst.* D**54**, 905–921.

Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Crystallographic R factor refinement by molecular dynamics. Science*, **235**, 458–460.

Burnett, M. N. & Johnson, C. K. (1996). *ORTEPIII: Oak Ridge thermal ellipsoid plot program for crystal structure illustrations*. Report ORNL-6895. Oak Ridge National Laboratory, Tennessee, USA.

Carson, M. (1997). *Ribbons. Methods Enzymol.* **277**, 493–505.

Carson, M. & Bugg, C. E. (1986). *Algorithm for ribbon models of proteins. J. Mol. Graphics*, **4**, 121–122.

Chapman, M. S. (1995). *Restrained real-space macromolecular atomic refinement using a new resolution-dependent electron-density function. Acta Cryst.* A**51**, 69–80.

Collaborative Computational Project, Number 4 (1994). *The CCP4 suite: programs for protein crystallography. Acta Cryst.* D**50**, 760–763.

Connolly, M. L. (1983). *Solvent-accessible surfaces of proteins and nucleic acids. Science*, **221**, 709–713.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M. Jr, Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. & Kollman, P. A. (1995). *A second generation force field for the simulation of proteins and nucleic acids. J. Am. Chem. Soc.* **117**, 5179–5197.

Cowtan, K. (1994). *DM: an automated procedure for phase improvement by density modification. CCP4 ESF-EACBM Newsl. Protein Crystallogr.* **31**, 34–38.

Evans, P. R. (1993). *Data reduction*. In *Proceedings of the CCP4 study weekend. Data collection and processing*, edited by L. Sawyer, N. W. Isaacs & S. Bailey, pp. 114–122. Warrington: Daresbury Laboratory.

Evans, P. R. (1997). *Scaling of MAD data*. In *Proceedings of the CCP4 study weekend. Recent advances in phasing*, edited by M. Winn, Vol. 33, pp. 22–24.

Evans, S. V. (1993). *SETOR: hardware-lighted three-dimensional solid model representations of macromolecules. J. Mol. Graphics*, **11**, 134–138.

Ferrin, T. E., Huang, C. C., Jarvis, L. E. & Langridge, R. (1988). *The MIDAS display system. J. Mol. Graphics*, **6**, 13–27, 36–37.

Furey, W. & Swaminathan, S. (1997). *PHASES-95: a program package for the processing and analysis of diffraction data from macromolecules. Methods Enzymol.* **277**, 590–620.

25.1 (*cont.*)

Hall, S. R., du Boulay, D. J. & Olthof-Hazekamp, R. (1999). *Xtal3.6 system*. Perth: University of Western Australia Press.

Hendrickson, W. A. (1991). *Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. Science*, **254**, 51.

Hendrickson, W. A. & Konnert, J. H. (1979). *Stereochemically restrained crystallographic least-squares refinement of macromolecule structures*. In *Biomolecular structure, conformation, function and evolution*, edited by R. Srinivasan, Vol. I, pp. 43–57. New York: Pergamon Press.

Jones, T. A. (1978). *A graphics model building and refinement system for macromolecules. J. Appl. Cryst.* **11**, 268–272.

Jones, T. A. (1992). *A, yaap, asap, @#*? A set of averaging programs*. In *Molecular replacement*, edited by E. J. Dodson, S. Gover & W. Wolf, pp. 91–105. Warrington: Daresbury Laboratory.

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Cryst.* A**47**, 110–119.

Kabsch, W. (1988). *Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector*. *J. Appl. Cryst.* **21**, 916–924.

Kabsch, W. & Sander, C. (1983). *Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers*, **22**, 2577–2637.

Kleywegt, G. J. & Jones, T. A. (1994). *Halloween . . . masks and bones*. In *From first map to final model*, edited by S. Bailey, R. Hubbard & D. Waller, pp. 59–66. Warrington: Daresbury Laboratory.

Kraulis, P. J. (1991). *MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Cryst.* **24**, 946–950.

La Fortelle, E. de & Bricogne, G. (1997). *Maximum-likelihood heavy-atom parameter refinement in the MIR and MAD methods. Methods Enzymol.* **276**, 472–494.

Lamzin, V. S. & Wilson, K. S. (1993). *Automated refinement of protein models. Acta Cryst.* D**49**, 129–147.

Lamzin, V. S. & Wilson, K. S. (1997). *Automated refinement for protein crystallography. Methods Enzymol.* **277**, 269–305.

Laskowski, R. A. (1995). *SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. J. Mol. Graphics*, **13**, 323–330.

Lee, B. & Richards, F. M. (1971). *The interpretation of protein structures: estimation of static accessibility. J. Mol. Biol.* **55**, 379–400.

Lu, G. (1999). *FINDNCS: a program to detect non-crystallographic symmetries in protein crystals from heavy-atom sites. J. Appl. Cryst.* **32**, 365–368.

Luscombe, N. M., Laskowski, R. A. & Thornton, J. M. (1997). *NUCPLOT: a program to generate schematic diagrams of protein–nucleic acid interactions. Nucleic Acids Res.* **25**, 4940–4945.

McDonald, I. K. & Thornton, J. M. (1994). *Satisfying hydrogen bonding potential in proteins. J. Mol. Biol.* **238**, 777–793.

MacKerell, A. D. Jr, Brooks, B., Brooks, C. L., Nilsson, L., Roux, B., Won, Y. & Karplus, M. (1998). *CHARMM: the energy function and its parameterization with an overview of the program*. In *The encyclopedia of computational chemistry*, edited by P. V. R. Schleyer, Vol. 1, pp. 271–277. Chichester: John Wiley & Sons.

McRee, D. E. (1993). *Practical protein crystallography*. San Diego: Academic Press.

Main, P., Fiske, S. J., Hull, S. E., Lessinger, L., Germain, G., Declercq, J.-P. & Woolfson, M. M. (1980). *MULTAN80. A system of computer programs for the automatic solution of crystal structures from X-ray diffraction data*. Universities of York, England, and Louvain, Belgium.

Merritt, E. A. & Bacon, D. J. (1997). *Raster3D: photorealistic molecular graphics. Methods Enzymol.* **277**, 505–524.

Merritt, E. A. & Murphy, M. E. P. (1994). *Raster3D version 2.0. A program for photorealistic molecular graphics. Acta Cryst.* D**50**, 869–873.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Refinement of macromolecular structures by the maximum-likelihood method. Acta Cryst.* D**53**, 240–255.

Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). *Efficient anisotropic refinement of macromolecular structures using FFT. Acta Cryst.* D**55**, 247–255.

Navaza, J. (1994). *AMoRe: an automated package for molecular replacement. Acta Cryst.* A**50**, 157–163.

Nicholls, A., Sharp, K. & Honig, B. (1991). *Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins*, **11**, 281–296.

Oldfield, T. J. (1992). *SQUID: a program for the analysis and display of data from crystallography and molecular dynamics. J. Mol. Graphics*, **10**, 247–252.

Otwinowski, Z. & Minor, W. (1996). *Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol.* **276**, 307–326.

Ravelli, R. B. G., Sweet, R. M., Skinner, J. M., Duisenberg, A. J. M. & Kroon, J. (1997). *STRATEGY: a program to optimize the starting spindle angle and scan range for X-ray data collection. J. Appl. Cryst.* **30**, 551–554.

Rodriguez, R., Chinea, G., Lopez, N., Pons, T. & Vriend, G. (1998). *Homology modeling, model and software evaluation: three related resources. Bioinformatics (CABIOS)*, **14**, 523–528.

Rossmann, M. G. & van Beek, C. G. (1999). *Data processing. Acta Cryst.* D**55**, 1631–1640.

Sali, A. & Blundell, T. L. (1993). *Comparative protein modeling by satisfaction of spatial restraints. J. Mol. Biol.* **234**, 779–815.

Sanner, M. F., Olson, A. J. & Spehner, J.-C. (1995). *Fast and robust computation of molecular surfaces. Proc. 11th ACM Symp. Comp. Geom.* C6–C7.

Sheldrick, G. M. & Schneider, T. R. (1997). *SHELXL: high-resolution refinement. Methods Enzymol.* **277**, 319–343.

Steigemann, W. (1991). *Recent advances in the PROTEIN program system for the X-ray structure analysis of biological macromolecules*. In *Crystallographic computing 5: from chemistry to biology*, edited by D. Moras, A. D. Podjarny and J. C. Thierry, pp. 115–125. New York: Oxford University Press.

Terwilliger, T. C. & Berendzen, J. (1999). *Automated MIR and MAD structure solution. Acta Cryst.* D**55**, 849–861.

Tong, L. & Rossmann, M. G. (1990). *The locked rotation function. Acta Cryst.* A**46**, 783–792.

Tronrud, D. E. (1997). *The TNT refinement package. Methods Enzymol.* **277**, 306–319.

Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. (1987). *An efficient general-purpose least-squares refinement program for macromolecular structures. Acta Cryst.* A**43**, 489–501.

Turk, D. (1995). *MAIN: a computer program for macromolecular crystallographers, now with utilities you always wanted*. In *Am. Crystallogr. Assoc. Annu. Meet.* Vol. 27 (ISSN 0596-4221), p. 54, Abstract 2m.6.B.

Vagin, A. A., Murshudov, G. N. & Strokopytov, B. V. (1998). *BLANC: the program suite for protein crystallography. J. Appl. Cryst.* **31**, 98–102.

Vriend, G. (1990). *WHAT IF: a molecular modeling and drug design program. J. Mol. Graphics*, **8**, 52–56.

Wallace, A. C., Laskowski, R. A. & Thornton, J. M. (1995). *LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions. Protein Eng.* **8**, 127–134.

Weeks, C. M. & Miller, R. (1999). *The design and implementation of SnB version 2.0. J. Appl. Cryst.* **32**, 120–124.

Zhang, K. Y. J. & Main, P. (1990*a*). *Histogram matching as a new density modification technique for phase refinement and extension of protein molecules. Acta Cryst.* A**46**, 41–46.

Zhang, K. Y. J. & Main, P. (1990*b*). *The use of Sayre's equation with solvent flattening and histogram matching for phase extension and refinement of protein structures. Acta Cryst.* A**46**, 377–381.

25.2

Abrahams, J. P. (1993). *Compression of X-ray images. Jt CCP4 ESF–EACBM Newsl. Protein Crystallogr.* **28**, 3–4.