SHELX for Macromolecules

George M. Sheldrick
Institut für Anorganische Chemie der Universität Göttingen
Tammannstraße 4, D-37077 Göttingen, Germany
gsheldr@shelx.uni-ac.gwdg.de

Abstract

A *new version of the complete SHELX system is being released in 1996. In addition to SHELXS-96 (structure solution) and SHELXL-96 (refinement) there are several auxiliary programs, including an interface program SHELXPRO for protein applications. Although SHELX was originally designed for small molecules, the new Patterson interpretation routine in SHELXS-96 is useful for the location of heavy atoms from* ΔF *data, and SHELXL-96 facilitates the refinement of macromolecules against high resolution data.*

1 Introduction

The first version of SHELX was written about 25 years ago. The gradual emergence of a relatively portable FORTRAN subset enabled it to be distributed (in compressed form including test data as one box of punched cards) in 1976. SHELX-76 survived unchanged - the extremely compact globally optimized code proved difficult to modify - until major advances in direct methods theory made an update of the structure solution part necessary (SHELXS-86). Rewriting and validating the least-squares refinement part proved more difficult but was finally achieved with SHELXL-93. SHELXS-86 and SHELXL-93 were as far as possible upwards compatible with SHELX-76 (for example the format of the reflection data file was unchanged) and are now employed in well over 50% of all small-molecule structure determinations. A commercial version including interative reciprocal and real space graphics is available in the form of the Siemens SHELXTL™ system.

A further release of SHELX in the current millenium was never intended, but the increased (mis)use of the programs by macromolecular crystallographers has unfortunately made it necessary. A new version of the complete package, SHELX-96, is now in the final stages of testing.

SHELXS-96 includes more powerful direct methods [1] and the use of the Patterson vector superposition method [2] - completely different to the naive Patterson interpretation algorithm used in SHELXS-86 - for the automatic location of heavy atoms. This new Patterson interpretation routine is not only effective for small structures - for some minerals it finds every atom - but also very useful for the location of heavy atom sites from isomorphous or anomalous ΔF data of macromolecules.

SHELXL-96 [3] has been extended so that it can also be used at more moderate resolution (say better than 2.8Å) and to make it easier to use for macromolecules.

A new feature in SHELX-96 is an interactive interface program SHELXPRO that is specific to protein applications; SHELXS and SHELXL are very general and in no way specific to certain types of crystal structure. SHELXPRO handles problems of communication with other widely used protein programs; for example it can convert PDB to SHELX format, adding appropriate restraints etc., and can generate sigma-A maps [4] etc. for map interpretation programs such as O [5]. SHELXPRO also displays the refinement results in the form of Postscript diagrams, and facilitates deposition of the refined structure with the PDB.

Since SHELXL-96 will be discussed in a separate talk, we shall concentrate on heavy atom location and SHELXPRO here.

2 Computing Aspects

SHELX-96 is provided in the form UNIX and VMS sources, plus precompiled versions for MSDOS, LINUX, IRIX, VMS, etc. The programs are available free to academics and for a small license fee (because it is necessary to cover all the costs) to for-profit institutions. The UNIX versions are highly portable, but sometimes it will be necessary to replace the routines that return the time, date and CPU time with local equivalents.

To run SHELXS or SHELXL, two files are almost always required: the files name.ins and name.hkl, where 'name' is a codename to identify the problem. The file name.ins contains crystal data, atomic coordinates (if required) and instructions; name.hkl contains the reflection data. The programs are always called by the program name followed by the codename; the names of all input and output files are then generated automatically by appending the appropriate suffixes. Thus:

shelxs barnase

would run the barnase Patterson interpretation test job, reading barnase.ins and barnase.hkl and writing the resutls to the listing file barnase.lst (which can be printed later). A summary of the progress of the program appears on the standard output device, and should be redirected to a file in the case of a batch job.

3 Heavy Atom Location from ΔF Data

For both the anomalous and isomorphous cases the user must prepare a .hkl file containing h,k,l,ΔF and $\sigma(\Delta F)$; the sign of ΔF is ignored. Careful scaling and suppression of outliers are essential; the CCP4 system contains suitable programs for this purpose.

The Patterson vector superposition algorithm [2,6] used in SHELXS-96 to find heavy atoms is totally different to that used in SHELXS-86; it may be summarized as follows:

1. One peak is selected from the sharpened Patterson (or input by means of a VECT instruction) to be used as a superposition vector. It must correspond to a correct heavy-atom to heavy-atom vector, otherwise the method will fail. The entire procedure may be repeated any number of times with different superposition vectors by specifying 'PATT n', with |n| > 1, or by including more than one VECT instruction in the same job.

2. The Patterson function is calculated twice, displaced from the origin by +U and –U, where U is the superposition vector. At each grid point the lower of the two values is taken, and the resulting 'superposition minimum function' is interpolated to find the peak positions. This is a much cleaner map than the original Patterson and contains only 2N (or 4N etc. if the superposition vector was multiple) peaks rather than N^2. The superposition map should ideally consist of one image of the structure and its inverse; it has an effective 'space group' of *P*-1 (or *C*-1 for a centered lattice etc.).

3. Possible origin shifts are found that place one of the images correctly with respect to the cell origin, i.e. most of the symmetry equivalents can be found in the peak-list. In this way one image is indentified and its inverse rejected.

4. For each acceptable origin shift, atomic numbers are assigned to the potential atoms based on average peak heights, and a 'crossword table' is generated. This gives the minimum distance and Patterson minimum function for each possible pair of unique atoms, taking symmetry into account. This table should be interpreted by hand to find a subset of the atoms making chemically sensible minimum interatomic distances linked by consistently large Patterson minimum function values. The Patterson values are recalculated from the original F_o data, not from the peak-list. For high symmetry space groups the minimum function is calculated as an average of the two (or more) smallest Patterson densities.

5. For each set of potential atoms a 'correlation coefficient' is calculated as a measure of the agreement between E_o^2 and E_c^2, and expressed as a percentage. The solution with the highest value for this figure of merit is usually the best.

The test file barnase.ins, for the location of three gold sites from isomorphous ΔF data (kindly donated by Eleanor Dodson), looks like:

```
TITL Barnase Au del(F) in P3(2)
CELL 1.5418 58.97 58.97 81.58 90 90 120
ZERR 1       0.05  0.05  0.08  0  0   0
LATT -1
SYMM -Y, X-Y, .66667+Z
SYMM -X+Y, -X, .33333+Z
SFAC N AU
UNIT 200 9
PATT 2
HKLF 3
END
```

The TITL..UNIT instructions specify the cell, space group and cell contents; they would be the same for any SHELX job for this structure, except that for use with ΔF data it is necessary to fudge the cell contents (by giving the square root of the number of light atoms, specified as nitrogen, followed by the expected number of heavy atoms in the cell). This job would try 2 suitable vectors from the Patterson peaklist for superposition; in a difficult case more should be specified, and possibly the other parameters for the PATT instruction would need fine-tuning (for example n should be made negative for a more exhaustive search). SHELXS outputs a summary of all the parameter settings it has used so that they can be modified by the user.

To employ direct methods [1] instead, only one line in the .ins file needs to be changed (from PATT to

TREF). Most of the recent advances in direct methods exploit either the weak reflections or more sophisticated formulas for probability distributions, so are wasted on ΔF data. Nevertheless, direct methods will tend to perform better in space groups with (a) translation symmetry (not counting lattice centering), (b) a fixed rather than a floating origin, and (c) no special positions; thus $P2_12_12_1$ is more suitable for direct methods than is $C2$. Anomalous data are less suitable for direct methods because of the missing centric reflections, unless MAD $F_\mathbf{a}$ estimates can be used.

4 SHELXPRO - Interface for Proteins

SHELXPRO is designed to be simple to use without the need of a manual.. The program is started as usual with **shelxpro** followed by the filename stem; this enables the program to read the files created by a refinement job with SHELXL for example. A main menu is then displayed. Choosing a particular option from the main menu produces a detailed description of that option, after which the user has the choice of typing <CR> to continue or N<CR> to return to the main menu. The program then procedes by question and answer, with sufficient information for the user to decide when not to take the default answers proposed by the program. On completing each operation, the program returns to the main menu.

Typical operations covered are the preparation of map and PDB files for O and some other map interpretation programs, converting PDB files into .ins files for input to SHELXL (this replaces PDBINS that was distributed with SHELXL-93 and includes updating between refinement jobs, automatic generation of disordered groups, restraints etc.), preparation of PDB files for deposition in Brookhaven, preparing .hkl files for SHELXL, displaying refinement results in the form of Postscript plots, analysis of esds, thermal motion and non-crystallographic symmetry, R-factors and data completeness as a function of resolution, etc. The restraints incorporated into the .ins file are stored internally in SHELXPRO, so no separate dictionary files are required. It should be emphasized that SHELXPRO is by no means the last word on the subject, and suggestions from users would be particularly welcome.

5 Summary

SHELX-96 consists of the following 6 programs; no environment variables, hidden files etc. are required:

SHELXS - Structure solution by Patterson and DM.

SHELXL - Structure refinement.

SHELXA - Post-absorption corrections (like DIFABS).

CIFTAB - Tables for publication via CIF format.

SHELXPRO - Protein interface to SHELX.

SHELXWAT - Automatic water divining.

Although SHELX was originally intended for small molecules, all these programs have potential uses for macromolecules; the last two were written specifically for macromolecules.

References

[1] G. M. Sheldrick, "Phase Annealing in SHELX-90; Direct Methods for Larger Structures," Acta Crystallogr., Vol. A46, pp. 467-473, 1990.

[2] G. M. Sheldrick, Z. Dauter, K. S. Wilson, H. Hope and L. C. Sieker, "The Application of Direct Methods and Patterson Interpretation to High-Resolution Native Protein Data," Acta Crystallogr., Vol. D49, pp. 18-23, 1993.

[3] G. M. Sheldrick and T. R. Schneider, "SHELXL: High-Resolution Refinement,".Methods in Enzymology (eds. C.W. Carter & R.M. Sweet), Academic Press, 1996 (in press).

[4] R. J. Read, "Improved Fourier Coefficients for Maps Using Phases from Partial Structures with Errors," Acta Crystallogr., Vol. A42, pp. 140-149, 1986.

[5] T. A. Jones, J.-Y. Zou, S. Cowan and M. Kjeldgaard, "Improved Methods for Building Protein Models in Electron Density Maps, and the Location of Errors in these Models," Acta Crystallogr., Vol. A47, pp. 110-119, 1991.

[6] G. M. Sheldrick, "Patterson Superposition and Ab Initio Phasing,".Methods in Enzymology (eds. C.W. Carter & R.M. Sweet), Academic Press, 1996 (in press).

[7] M. Fujinaga and R. J. Read, “Experiences with a new Translation-Function Program,” J. Appl. Crystallogr., Vol. 20, pp. 517-521, 1987.