International Tables for Crystallography (2006). Vol. F, Section 25.2.10.3, pp. 735–736.

25.2.10.2. *Program organization and philosophy*

SHELX is written in a simple subset of Fortran77 that has proved to be extremely portable. The programs *SHELXS* (structure solution) and *SHELXL* (refinement) both require only two input files: a reflection file (*name.hkl*) and a file (*name.ins*) that contains crystal data, atoms (if any) and instructions in the form of keywords followed by free-format numbers *etc.* These programs write a file, *name.res*, that can be renamed or edited to *name.ins* for the next refinement and can output details of the calculations to *name.lst*. Although originally designed for punched cards, this arrangement is still quite convenient and has retained upwards compatibility for the last 30 years. The common first part of the filename is read from the command line by typing, *e.g.*, 'SHELXL *name*'. The programs are executed independently without the use of any hidden files, environment variables *etc.*

The programs are general for all space groups in conventional settings or otherwise and make extensive use of default settings to keep user input and confusion to a minimum. Particular care has been taken to test the programs thoroughly on as many computer systems and crystallographic problems as possible before they were released, a process that often required several years!

25.2.10.3. *Heavy-atom location using SHELXS and SHELXD*

One might expect that a small-molecule direct-methods program, such as *SHELXS* (Sheldrick, 1990), that routinely solves structures with 20–100 unique atoms in a few minutes or even seconds of computer time would have no difficulty in locating a handful of heavy-atom sites from isomorphous or anomalous ΔF data. However, such data can be very noisy, and a single seriously aberrant reflection can invalidate a large number of probabilistic phase relations. The most important direct-methods formula is still the tangent formula of Karle & Hauptman (1956); most modern direct-methods programs (*e.g.* Busetta *et al.*, 1980; Debaerdemaeker *et al.*, 1985; Sheldrick, 1990) use versions of the tangent formula that have been modified to incorporate information from weak reflections as well as strong reflections, which helps to avoid pseudo-solutions with translationally displaced molecules or a single dominant peak (the so-called uranium-atom solution). Isomorphous and anomalous ΔF values represent lower limits on the structure factors for the heavy-atom substructure and so do not give reliable estimates of weak reflections; thus, the improvements introduced into direct methods by the introduction of the weak reflections are largely irrelevant when they are applied to ΔF data. This does not apply when F_A values are derived from a MAD experiment, since these are true estimates of the heavy-atom structure factors; however, aberrant large and small F_A estimates are difficult to avoid and often upset the phase-determination process. A further problem in applying direct methods to ΔF data is that it is not always clear what the effective number of atoms in the cell should be for use in the probability formulae, especially when it is not known in advance how many heavy-atom sites are present.

25.2.10.3.1. *The Patterson map interpretation algorithm in SHELXS*

Space-group-general automatic Patterson map interpretation was introduced in the program *SHELXS*86 (Sheldrick, 1985); completely different algorithms are employed in the current version of *SHELXS*, based on the Patterson superposition minimum function (Buerger, 1959, 1964; Richardson & Jacobson, 1987; Sheldrick, 1991, 1998*a*; Sheldrick *et al.*, 1993). The algorithm used in *SHELXS* is as follows:

(1) A single Patterson peak, $\mathbf{v}$, is selected automatically (or input by the user) and used as a superposition vector. A sharpened Patterson map [with coefficients $(E^3F)^{1/2}$ instead of F^2, where E is a normalized structure factor] is calculated twice, once with the origin shifted to $-\mathbf{v}/2$ and once with the origin shifted to $+\mathbf{v}/2$. At each grid point, the minimum of the two Patterson function values is stored, and this *superposition minimum function* is searched for peaks. If a true single-weight heavy atom-to-heavy atom vector has been chosen as the superposition vector, this function should consist ideally of one image of the heavy-atom structure and one inverted image, with two atoms (the ones corresponding to the superposition vector) in common. There are thus about $2N$ peaks in the map, compared with N^2 in the original Patterson map, a considerable simplification. The only symmetry element of the superposition function is the inversion centre at the origin relating the two images.

(2) Possible origin shifts are found so that the full space-group symmetry is obeyed by one of the two images, *i.e.*, for about half the peaks, most of the symmetry equivalents are present in the map. This enables the peaks belonging to the other image to be eliminated and, in principle, solves the heavy-atom substructure. In the space group $P1$, the double image cannot be resolved in this way.

(3) For each plausible origin shift, the potential atoms are displayed as a triangular table that gives the minimum distance and the Patterson superposition minimum function value for all vectors linking each pair of atoms, taking all symmetry equivalents into account. This table enables spurious atoms to be eliminated and occupancies to be estimated, and also in some cases reveals the presence of noncrystallographic symmetry.

(4) The whole procedure is then repeated for further superposition vectors as required. The program gives preference to general vectors (multiple vectors will lead to multiple images), and it is advisable to specify a minimum distance of (say) 8 Å for the superposition vector (3.5 Å for selenomethionine MAD data) to increase the chance of finding a true heavy atom-to-heavy atom vector.

25.2.10.3.2. *Integrated Patterson and direct methods: SHELXD*

The program *SHELXD* (Sheldrick & Gould, 1995; Sheldrick, 1997, 1998*b*) is now part of the *SHELX* system. It is designed both for the *ab initio* solution of macromolecular structures from atomic resolution native data alone and for the location of heavy-atom sites from ΔF or F_A values at much lower resolution, in particular for the location of larger numbers of anomalous scatterers from MAD data. The dual-space approach of *SHELXD* was inspired by the *Shake and Bake* philosophy of Miller *et al.* (1993, 1994) but differs in many details, in particular in the extensive use it makes of the Patterson function that proves very effective in applications involving ΔF or F_A data. The *ab initio* applications of *SHELXD* have been described in Chapter 16.1, so only the location of heavy atoms will be described here. An advantage of the Patterson function is that it provides a good noise filter for the ΔF or F_A data: negative regions of the Patterson function can simply be ignored. On the other hand, the direct-methods approach is efficient at handling a large number of sites, whereas the number of Patterson peaks to analyse increases with the square of the number of atoms. Thus, for reasons of efficiency, the Patterson function is employed at two stages in *SHELXD*: at the beginning to obtain starting atom positions (otherwise random starting atoms would be employed) and at the end, in the form of the triangular table described above, to recognize which atoms are correct. In between, several cycles of real/reciprocal space alternation are employed as in the *ab initio* structure solution, alternating between tangent refinement, *E*-map calculation and peak search, and possibly *random omit maps*, in which a specified fraction of the potential atoms are left out at random.

25.2.10.3.3. *Practical considerations*

Since the input files for the direct and Patterson methods in *SHELXS* and the integrated method in *SHELXD* are almost identical (usually only one instruction needs to be changed), it is easy to try all three methods for difficult problems. The Patterson map interpretation in *SHELXS* is a good choice if the heavy atoms have variable occupancies and it is not known how many heavy-atom sites need to be found; the direct-methods approaches work best with equal atoms. In general, the conventional direct methods in *SHELXS* will tend to perform best in a non-polar space group that does not possess special positions; however, for more than about a dozen sites, only the integrated approach in *SHELXD* is likely to prove effective; the *SHELXD* algorithm works best when the number of sites is known. Especially for the MAD method, the quality of the data is decisive; it is essential to collect data with a high redundancy to optimize the signal-to-noise ratio and eliminate outliers. In general, a resolution of 3.5 Å is adequate for the location of heavy-atom sites. At the time of writing, *SHELXD* does not include facilities for the further calculations necessary to obtain maps. Experience indicates that it is only necessary to refine the *B* values of the heavy atoms using other programs; their coordinates are already rather precise.

Excellent accounts of the theory of direct and Patterson methods with extensive literature references have been presented in *IT* B Chapter 2.2 by Giacovazzo (2001) and Chapter 2.3 by Rossmann & Arnold (2001).

25.2.10.4. *Macromolecular refinement using SHELXL*

SHELXL is a very general refinement program that is equally suitable for the refinement of minerals, organometallic structures, oligonucleotides, or proteins (or any mixture thereof) against X-ray or neutron single- (or twinned!) crystal data. It has even been used with diffraction data from powders, fibres and two-dimensional crystals. For refinement against Laue data, it is possible to specify a different wavelength and hence dispersion terms for each reflection. The price of this generality is that it is somewhat slower than programs specifically written only for protein structure refinement. Any protein- (or DNA-)specific information must be input to *SHELXL* by the user in the form of refinement restraints *etc.* Refinement of macromolecules using *SHELXL* has been discussed by Sheldrick & Schneider (1997).

25.2.10.4.1. *Constraints and restraints*

In refining macromolecular structures, it is almost always necessary to supplement the diffraction data with chemical information in the form of *restraints*. A typical restraint is the condition that a bond length should approximate to a target value with a given estimated standard deviation; restraints are treated as extra experimental data items. Even if the crystal diffracts to 1.0 Å, there may well be poorly defined disordered regions for which restraints are essential to obtain a chemically sensible model (the same can be true of small molecules too!). *SHELXL* is generally not suitable for refinements at resolutions lower than about 2.5 Å because it cannot handle general potential-energy functions, *e.g.* for torsion angles or hydrogen bonds; if noncrystallographic symmetry restraints can be employed, this limit can be relaxed a little.

For some purposes (*e.g.* riding hydrogen atoms, rigid-group refinement, or occupancies of atoms in disordered side chains), *constraints*, exact conditions that lead to a reduction in the number of variable parameters, may be more appropriate than restraints; *SHELXL* allows such constraints and restraints to be mixed freely. Riding hydrogen atoms are defined such that the C—H vector remains constant in magnitude and direction, but the carbon atom is free to move; the same shifts are applied to both atoms, and both atoms contribute to the least-squares derivative sums. This model may be combined with anti-bumping restraints that involve hydrogen atoms, which helps to avoid unfavourable side-chain conformations. *SHELXL* also provides, *e.g.*, methyl groups that can rotate about their local threefold axes; the initial torsion angle may be found using a difference-electron-density synthesis calculated around the circle of possible hydrogen-atom positions.

25.2.10.4.2. *Least-squares refinement algebra*

The original *SHELX* refinement algorithms were modelled closely on those described by Cruickshank (1970). For macromolecular refinement, an alternative to (blocked) full-matrix refinement is provided by the conjugate-gradient solution of the least-squares normal equations as described by Hendrickson & Konnert (1980), including preconditioning of the normal matrix that enables positional and displacement parameters to be refined in the same cycle. The structure-factor derivatives contribute only to the diagonal elements of the normal matrix, but all restraints contribute fully to both the diagonal and non-diagonal elements, although neither the Jacobian nor the normal matrix itself are ever generated by *SHELXL*. The parameter shifts are modified by comparison with those in the previous cycle to accelerate convergence whilst reducing oscillations. Thus, a larger shift is applied to a parameter when the current shift is similar to the previous shift, and a smaller shift is applied when the current and previous shifts have opposite signs.

SHELXL refines against F^2 rather than *F*, which enables all data to be used in the refinement with weights that include contributions from the experimental uncertainties, rather than having to reject *F* values below a preset threshold; there is a choice of appropriate weighting schemes. Provided that reasonable estimates of $\sigma(F^2)$ are available, this enables more experimental information to be employed in the refinement; it also allows refinement against data from twinned crystals.

25.2.10.4.3. *Full-matrix estimates of standard uncertainties*

Inversion of the full normal matrix (or of large matrix blocks, *e.g.* for all positional parameters) enables the precision of individual parameters to be estimated (Rollett, 1970), either with or without the inclusion of the restraints in the matrix. The standard uncertainties in dependent quantities (*e.g.* torsion angles or distances from mean planes) are calculated in *SHELXL* using the full least-squares correlation matrix. These standard uncertainties reflect the data-to-parameter ratio, *i.e.* the resolution and completeness of the data and the percentage of solvent, and the quality of the agreement between the observed and calculated F^2 values (and the agreement of restrained quantities with their target values when restraints are included).

Full-matrix refinement is also useful when domains are refined as rigid groups in the early stages of refinement (*e.g.* after structure solution by molecular replacement), since the total number of parameters is small and the correlation between parameters may be large.

25.2.10.4.4. *Refinement of anisotropic displacement parameters*

The motion of macromolecules is clearly anisotropic, but the data-to-parameter ratio rarely permits the refinement of the six independent anisotropic displacement parameters (ADPs) per atom; even for small molecules and data to atomic resolution, the

references