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

anisotropic refinement of disordered regions requires the use of restraints. *SHELXL* employs three types of ADP restraint (Sheldrick 1993; Sheldrick & Schneider, 1997). The *rigid bond restraint*, first suggested by Rollett (1970), assumes that the components of the ADPs of two atoms connected *via* one (or two) chemical bonds are equal within a specified standard deviation. This has been shown to hold accurately (Hirshfeld, 1976; Trueblood & Dunitz, 1983) for precise structures of small molecules, so it can be applied as a 'hard' restraint with small estimated standard deviation. The *similar ADP restraint* assumes that atoms that are spatially close (but not necessarily bonded, because they may be different components of a disordered group) have similar U^{ij} components. An *approximately isotropic restraint* is useful for isolated solvent molecules. These two restraints are only approximate and so should be applied with low weights, *i.e.* high estimated standard deviations.

The transition from isotropic to anisotropic roughly doubles the number of parameters and almost always results in an appreciable reduction in the *R* factor. However, this represents an improvement in the model only when it is accompanied by a significant reduction in the free *R* factor (Brünger, 1992*b*). Since the free *R* factor is itself subject to uncertainty because of the small sample used, a drop of at least 1% is needed to justify anisotropic refinement. There should also be a reduction in the goodness of fit, and the resulting displacement ellipsoids should make chemical sense and not be 'non-positive-definite'!

25.2.10.4.5. *Similar geometry and NCS restraints*

When there are several identical chemical moieties in the asymmetric unit, a very effective restraint is to assume that the chemically equivalent 1,2 and 1,3 distances are the same, but unknown. This technique is easy to apply using *SHELXL* and is often employed for small-molecule structures and, in particular, for oligosaccharides. Similarly, the terminal P—O bond lengths in DNA structures can be assumed to be the same (but without a target value), *i.e.* it is assumed that the whole crystal is at the same pH. For proteins, the method is less suitable because of the different abundance of the different amino acids, and, in any case, good target distances are available (Engh & Huber, 1991).

Local noncrystallographic symmetry (NCS) restraints (Usón *et al.*, 1999) may be applied to restrain corresponding 1,4 distances and isotropic displacement parameters to be the same when there are several identical macromolecular domains in the asymmetric unit; usually, the 1,2 and 1,3 distances are restrained to standard values in such cases and so do not require NCS restraints. Such *local NCS restraints* are more flexible than *global NCS constraints* and – unlike the latter – do not require the specification of a transformation matrix and mask.

25.2.10.4.6. *Modelling disorder and solvent*

There are many ways of modelling disorder using *SHELXL*, but for macromolecules the most convenient is to retain the same atom and residue names for the two or more components and assign a different 'part number' (analogous to the PDB alternative site flag) to each component. With this technique, no change is required to the input restraints *etc*. Atoms in the same component will normally have a common occupancy that is assigned to a 'free variable'. If there are only two components, the sum of their occupancies can be constrained to be unity; if there are more than two components, the sum of their free variables may be restrained to be unity. Since any linear restraint may be applied to the free variables, they are very flexible, *e.g.* for modelling complicated disorder. By restraining distances to be equal to a free variable, a standard deviation of the mean distance may be calculated rigorously using full-matrix least-squares algebra.

Babinet's principle is used to define a bulk solvent model with two refinable parameters (Moews & Kretsinger, 1975), and global anisotropic scaling (Usón *et al.*, 1999) may be applied using a parameterization proposed by Parkin *et al.* (1995). An auxiliary program, *SHELXWAT*, allows automatic water divining by iterative least-squares refinement, rejection of waters with high displacement parameters, difference-electron-density calculation, and a peak search for potential water molecules that make at least one good hydrogen bond and no bad contacts; this is a simplified version of the *ARP* procedure of Lamzin & Wilson (1993).

25.2.10.4.7. *Twinned crystals*

SHELXL provides facilities for refining against data from merohedral, pseudo-merohedral and non-merohedral twins (Herbst-Irmer & Sheldrick, 1998). Refinement against data from merohedrally twinned crystals is particularly straightforward, requiring only the twin law (a 3×3 matrix) and starting values for the volume fractions of the twin components. Failure to recognize such twinning not only results in high *R* factors and poor quality maps, it can also lead to incorrect biochemical conclusions (Luecke *et al.*, 1998). Twinning can often be detected by statistical tests (Yeates & Fam, 1999), and it is probably much more widespread in macromolecular crystals than is generally appreciated!

25.2.10.4.8. *The radius of convergence*

Least-squares refinement as implemented in *SHELXL* and other programs is appropriate for structural models that are relatively complete, but when an appreciable fraction of the structure is still to be located, maximum-likelihood refinement (Bricogne, 1991; Pannu & Read, 1996*a*; Murshudov *et al.*, 1997) is likely to be more effective, especially when experimental phase information can be incorporated (Pannu *et al.*, 1998). Within the least-squares framework, there are still several possible ways of improving the radius of convergence. *SHELXL* provides the option of gradually extending the resolution of the data during the refinement; a similar effect may be achieved by a resolution-dependent weighting scheme (Terwilliger & Berendzen, 1996). Unimodal restraints, such as target distances, are less likely to result in local minima than are multimodal restraints, such as torsion angles; multimodal functions are better used as validation criteria. It is fortunate that validation programs, such as *PROCHECK* (Laskowski *et al.*, 1993), make good use of multimodal functions such as torsion angles and hydrogen-bonding patterns that are not employed as restraints in *SHELXL* refinements.

25.2.10.5. *SHELXPRO – protein interface to SHELX*

The *SHELX* system includes several auxiliary programs, the most important of which for macromolecular users is *SHELXPRO*. *SHELXPRO* provides an interface between *SHELXS*, *SHELXL* and other programs commonly used by protein crystallographers, particularly graphics programs; for example, it can write map files for *O* (Jones *et al.*, 1991) or (*Turbo*)*Frodo* (Jones, 1978). For *XtalView* (McRee, 1992), this is not necessary, because *XtalView* can read the CIF format reflection data files written by *SHELXL* directly, and *XtalView* is generally the interactive macromolecular graphics program of choice for use with *SHELX* because it can interpret and display anisotropic displacement parameters and multiple conformations.

Often, *SHELXL* will be used only for the final stages of refinement, in which case *SHELXPRO* is used to generate the *name.ins* file from a PDB format file, inserting the necessary restraints and other instructions. The geometric restraints for