16. DIRECT METHODS

$$G = 2(\sigma^3/\sigma_2^{3/2})_p E_{p\mathbf{H}} E_{p\mathbf{K}} E_{p\mathbf{H}+\mathbf{K}} + 2q(\sigma^3/\sigma_2^{3/2})_{\mathbf{H}} \Delta_{\mathbf{H}} \Delta_{\mathbf{K}} \Delta_{\mathbf{H}+\mathbf{K}}$$

$$\Delta = (F_d - F_p)/\Sigma_H^{1/2}. \tag{16.1.12.10}$$

The factor q takes into account lack of isomorphism and measurement errors, and the Δ parameters are isomorphous differences normalized with respect to the heavy-atom structure. Φ_p is expected to be close to 0 or π according to whether G is positive or negative. A starting set of phases is generated by a random process, a weighted tangent formula is applied to these phases and various trials are produced among which the correct solution may be found by a suitable figure of merit. If multiple derivatives are available, the program is able to estimate automatically, for each derivative, the scattering power of the heavy-atom structure and also to suggest which is the best derivative. The *IL MILIONE* package can be obtained from http://www.ic.cnr.it.

16.1.12.3. SHELX

The *SHELX* family of programs is widely used for small- to medium-sized structure solution and refinement. The family also contains three programs that are extensively used in macromolecular crystallography: *SHELXC*, *SHELXD* and *SHELXE*. For an overview of the *SHELX* system, see Sheldrick (2008). *SHELXC* is a housekeeping program designed to prepare the necessary files for *SHELXD* and *SHELXE*. *SHELXD* (Sheldrick, 1998; Schneider & Sheldrick, 2002; Usón & Sheldrick, 1999) is employed both for substructure solution and for *ab initio* direct methods for atomic resolution data as described elsewhere in this chapter. Fu *et al.* (2007) have shown how to adapt it to multiple CPU systems. *SHELXE* (Sheldrick, 2002, 2008, 2010) improves experimental phases from SAD, SIRAS or MAD data or starting phases from molecular replacement by iterative density modification and autotracing. *SHELX* can be obtained at http://shelx.uni-ac.gwdg.de/SHELX/.

16.1.12.4. SnB and BnP

SnB was the first program to solve small macromolecules *ab initio*, using a global cost function [equation (16.1.4.2)] that reflects how well the calculated phases fit the expected distribution of the triplets. It is fully described elsewhere in this chapter and is an effective tool in structure and substructure determination. Versions are available for multiple CPU systems (Rappleye *et al.*, 2002) and computational grids (Miller *et al.*, 2007). It is also available as part of the *BnP* package (Weeks *et al.*, 2002) that was produced in collaboration with the Biocrystallography Laboratory at the University of Pittsburgh for the experimental phasing of macromolecules. *SnB* is available from http://www.hwi.buffalo.edu/SnB/ and *BnP* from http://www.hwi.buffalo.edu/BnP/.

16.1.12.5. HySS

The substructure solution program *HySS* (Grosse-Kunstleve & Adams, 2003), which is part of the *PHENIX* package (Adams *et al.*, 2007), is closely modelled on *SHELXD* but was implemented using the cctbx libraries (Grosse-Kunstleve *et al.*, 2002). The main differences to *SHELXD* are (1) the translational search for two-atom fragments is performed by Fourier methods followed by a peak search rather than a random search, (2) the use of the tangent formula in reciprocal space is replaced by squaring the density in real space, and (3) several termination criteria are implemented so that the program can stop when the structure appears to be solved. The *PHENIX* package can be obtained at http://www.phenix-online.org/.

16.1.12.6. SUPERFLIP: charge flipping

Charge flipping is a disturbingly simple dual-space algorithm (Oszlányi & Sütő, 2004, 2005, 2008). It uses as input only the cell parameters of the structure, the reflection indices and the intensities. The intensities can be corrected for thermal motion *via* an overall temperature factor if required, and this is often beneficial. Neither chemical information nor the symmetry is explicitly used in the structure solution process. The electron density is sampled on a discrete rectangular grid of pixels with values ρ_i, $i = 1, N_{\mathrm{pix}}$. The algorithm proceeds iteratively. To begin the process, a starting set of structure factors is created by combining the experimental structure-factor amplitudes with random phases. Each iteration or cycle (numbered n) involves four steps:

(1) A trial electron density $\rho^{(n)}$ is obtained by inverse Fourier transform of the structure factors in the usual way.

(2) A modified density $g^{(n)}$ is obtained from $\rho^{(n)}$ by reversing the sign (charge flipping) of all density pixels with density below a certain positive threshold δ as follows:

$$\begin{aligned} g_i^{(n)} &= \rho_i^{(n)} \quad \text{if} \quad \rho_i^{(n)} > \delta, \\ g_i^{(n)} &= -\rho_i^{(n)} \quad \text{if} \quad \rho_i^{(n)} \le \delta. \end{aligned} \tag{16.1.12.11}$$

(3) Modified structure factors are obtained by Fourier transform of $g^{(n)}$,

$$G_{\mathbf{H}}^{(n)} = FT[g^{(n)}]. \tag{16.1.12.12}$$

(4) The structure factors $F_{\mathbf{H}}^{(n+1)}$ are obtained from $F_{\mathbf{H}}^{(n+1)}$ and $G_{\mathbf{H}}^{(n+1)} = |G_{\mathbf{H}}^{(n+1)}| \exp(2\pi i \varphi_{\mathbf{H}}^G)$ as follows:

(a) $F_{\mathbf{H}}^{(n+1)} = |F_{\mathbf{H}}| \exp(2\pi i \varphi_{\mathbf{H}}^G)$ for $|F_{\mathbf{H}}|$ observed and strong.

(b) $F_{\mathbf{H}}^{(n+1)} = |G_{\mathbf{H}}^{(n)}| \exp[2\pi i(\varphi_{\mathbf{H}}^G + 0.25)]$ for $|F_{\mathbf{H}}|$ observed and weak. In other words, for these reflections, calculated moduli are accepted unchanged and calculated phases are shifted by a constant $\Delta\varphi = \pi/2$. This means that the observed data of weak reflections are not used actively in the process, except for the knowledge that they are indeed weak. Experience shows that about 20–40% of all reflections can be considered weak. This use of the phase shifting of the weak reflections significantly improves the performance of the algorithm in cases of more complex structures (Oszlányi & Sütő, 2005); in some cases the success rate is increased by a factor of ten, in other cases a previously unsolvable structure becomes solvable by the modified algorithm.

(c) $F_{\mathbf{H}}^{(n+1)} = 0$ for $|F_{\mathbf{H}}|$ unobserved.

(d) $F_{\mathbf{H}}^{(n+1)} = G_{\mathbf{H}}^{(n)}$ for $\mathbf{H} = 0$. In other words, the value of F_{000} is not fixed.

The new set of structure factors enters the next cycle or iteration. The cycles are repeated until the calculation converges. Progress is monitored by a conventional R factor where a small change in R signals convergence. The parameter δ is the main variable of the iteration, and its value can be critical. It must often be determined by trial and error, but this search can be automated. The second variable parameter of the algorithm is the proportion of the reflections considered weak in each cycle.