16. DIRECT METHODS

16.1. *Ab initio* phasing

By G. M. Sheldrick, H. A. Hauptman, C. M. Weeks, R. Miller and I. Usón

16.1.1. Introduction

Ab initio methods for solving the crystallographic phase problem rely on diffraction amplitudes alone and do not require prior knowledge of any atomic positions. General features that are not specific to the structure in question (*e.g.* the presence of disulfide bridges or solvent regions) can, however, be utilized. For the last three decades, most small-molecule structures have been routinely solved by *direct methods*, a class of *ab initio* methods in which probabilistic phase relations are used to derive reflection phases from the measured amplitudes. Direct methods, implemented in widely used highly automated computer programs such as *MULTAN* (Main *et al.*, 1980), *SHELXS* (Sheldrick, 1990), *SAYTAN* (Debaerdemaeker *et al.*, 1985) and *SIR* (Burla *et al.*, 1989), provide computationaly efficient solutions for structures containing fewer than approximately 100 independent non-H atoms. However, larger structures are not consistently amenable to these programs and, in fact, few unknown structures with more than 200 independent equal atoms have ever been solved using these programs.

Successful applications to native data for structures that could legitimately be regarded as small macromolecules awaited the development of a direct-methods procedure (Weeks *et al.*, 1993) that has come to be known as *Shake-and-Bake*. The distinctive feature of this procedure is the repeated and unconditional alternation of reciprocal-space phase refinement (*Shaking*) with a complementary real-space process that seeks to improve phases by applying constraints (*Baking*). Consequently, it yields a computer-intensive algorithm, requiring two Fourier transformations during each cycle, which has been made feasible in recent years due to the tremendous increases in computer speed. The first previously unknown structures determined by *Shake-and-Bake* were two forms of the 100-atom peptide ternatin (Miller *et al.*, 1993). Subsequent applications of the *Shake-and-Bake* algorithm have involved structures containing as many as 2000 independent non-H atoms (Frazão *et al.*, 1999) provided that accurate diffraction data have been measured to a resolution of 1.2 Å or better.

The basic theory underlying direct methods has been summarized in an excellent chapter (Giacovazzo, 2001) in *IT* B to which the reader is referred for details. The present chapter focuses on those aspects of direct methods that have proven useful for larger molecules (more than 250 independent non-H atoms) or are unique to the macromolecular field. These include direct-methods applications that utilize anomalous-dispersion measurements or multiple diffraction patterns [*i.e.*, single isomorphous replacement (SIR), single anomalous scattering (SAS) and multiple-wavelength data]. The easiest way to combine isomorphous or anomalous-scattering information with direct methods is to first compute difference structure factors and then to apply direct methods to the difference data. Using this approach, the dual-space *Shake-and-Bake* procedure has been used to solve the anomalously scattering substructure of the selenomethionine derivative of an epimerase enzyme that has 70 selenium sites (Deacon & Ealick, 1999). Substructure applications require only the 2.5–3.0 Å data normally included in multiple wavelength anomalous dispersion (MAD) measurements, and data sets truncated even to 5 Å have led to solutions.

A formal integration of the probabilistic machinery of direct methods with isomorphous replacement and anomalous dispersion was initiated in 1982 (Hauptman, 1982*a*,*b*). Although practical applications of this and subsequent related theory have been limited so far, such applications are likely to have greater importance in the future, and progress is described in Sections 16.1.9.1 and 16.1.9.2. Similarly, the combination of direct methods with multiple-beam diffraction is still in its infancy. However, preliminary studies indicate that the information gleaned from multiple-beam data will greatly strengthen existing techniques (Weckert *et al.*, 1993). Progress in this area is summarized in Section 16.1.9.3.

16.1.2. Normalized structure-factor magnitudes

For purposes of direct-methods computations, the usual structure factors, $F_{\mathbf{H}}$, are replaced by the *normalized structure factors* (Hauptman & Karle, 1953),

$$E_{\mathbf{H}} = |E_{\mathbf{H}}| \exp(i\varphi_{\mathbf{H}}),$$

$$|E_{\mathbf{H}}| = \frac{|F_{\mathbf{H}}|}{\left\langle |F_{\mathbf{H}}|^2 \right\rangle^{1/2}} = \frac{k\left\langle \exp\left[-B_{\mathrm{iso}}(\sin\theta)^2/\lambda^2\right]\right\rangle^{-1}|F_{\mathbf{H}}|_{\mathrm{meas}}}{\left(\varepsilon_{\mathbf{H}} \sum_{j=1}^{N} f_j^2\right)^{1/2}},$$

(16.1.2.1)

where the angle brackets indicate probabilistic or statistical expectation values, the $|E_{\mathbf{H}}|$ and $|F_{\mathbf{H}}|$ are structure-factor magnitudes, the $\varphi_{\mathbf{H}}$ are the corresponding phases, k is the absolute scaling factor for the measured magnitudes, B_{iso} is an overall isotropic atomic mean-square displacement parameter, the f_j are the atomic scattering factors for the N atoms in the unit cell, and the $\varepsilon_{\mathbf{H}} \geq 1$ are factors that account for multiple enhancement of the average intensities for certain special reflection classes due to space-group symmetry (Shmueli & Wilson, 2001). The condition $\langle |E|^2 \rangle = 1$ is always imposed. Unlike $\langle |F_{\mathbf{H}}| \rangle$, which decreases as $\sin(\theta)/\lambda$ increases, the values of $\langle |E_{\mathbf{H}}| \rangle$ are constant for concentric resolution shells. Thus, the normalization process places all reflections on a common basis, and this is a great advantage with regard to the probability distributions that form the foundation for direct methods. Normalizing a set of reflections by means of equation (16.1.2.1) does not require any information about atomic positions. However, if some structural information, such as the configuration, orientation, or position of certain atomic groupings, is available, then this information can be applied to obtain a better model for the expected intensity distribution (Main, 1976). The distribution of $|E|$ values is, in principle and often in practice, independent of the unit-cell size and contents, but it does depend on whether a centre of symmetry is present, as shown in Table 16.1.2.1.

Direct-methods applications having the objective of locating SIR or SAS substructures require the computation of normalized *difference* structure-factor magnitudes, $|E_\Delta|$. This can, for example, be accomplished with the following series of programs from Blessing's data-reduction and error-analysis routines (*DREAR*): *LEVY* and *EVAL* for structure-factor normalization as specified by equation (16.1.2.1) (Blessing *et al.*, 1996), *LOCSCL* for local scaling of the SIR and SAS magnitudes (Matthews & Czerwinski, 1975; Blessing, 1997), and *DIFFE* for computing the actual difference magnitudes (Blessing & Smith, 1999). The *SnB* program