Part I
Structure Determination

Modern Diffraction Methods, First Edition. Edited by Eric J. Mittemeijer and Udo Welzel.

1
Structure Determination of Single Crystals

Sander van Smaalen

1.1
Introduction

Many crystalline materials possess translational symmetry. This property implies that the positions of the atoms can be obtained from the positions of a few atoms in a small volume (the unit cell), which is periodically repeated in space. Thus, the crystal structure is completely characterized by the translational symmetry – as given by the six lattice parameters – together with the positions of the atoms within one unit cell (Figure 1.1a) [1]. The goal of the procedure of structure determination is to obtain the atomic positions in the unit cell and the lattice parameters from a diffraction experiment.

X-rays are scattered by matter (Figure 1.1b). A consequence of translational symmetry is that scattered rays are only obtained for directions **k** corresponding to scattering angles 2θ as followed from Bragg's law. Furthermore, each Bragg reflection requires a specific orientation of the crystal with respect to the directions of the primary and scattered X-ray beams. Both properties depend on the lattice parameters, while different Bragg reflections are distinguished on the basis of a unique indexing with three integral numbers $(h\,k\,l)$. The other way around, knowledge of the orientations of the crystal in conjunction with the directions of the diffracted beams for a sufficiently large number of Bragg reflections allows the determination of the lattice parameters and the indices of each Bragg reflection.

The amplitude and phase of the diffracted beam are the second unique property of each Bragg reflection. Their values depend on the structure of one unit cell of the crystal and the indices $(h\,k\,l)$ of the Bragg reflection. Knowledge of the amplitudes and phases of many Bragg reflections allow the determination of the positions of the atoms in the crystal by the simple computational procedure of Fourier transform (Section 1.2).

The diffraction experiment provides for each Bragg reflection the orientation of the single crystal and the direction and intensity of the diffracted beam. This information is sufficient to compute the lattice parameters of the crystal and

Modern Diffraction Methods, First Edition. Edited by Eric J. Mittemeijer and Udo Welzel.

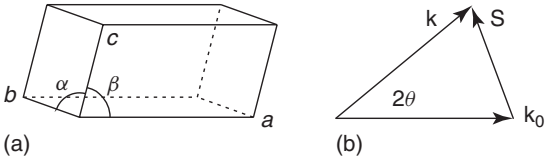

Figure 1.1 (a) A unit cell with edge lengths a, b, and c and angles between edges with α, β, and γ. (b) The scattering vector **S** defines the difference between the directions of propagation of the primary and scattered waves.

the indices of the Bragg reflections, but it does not allow computation of the crystal structure by the method of Fourier transform. The reason is that the intensity of an X-ray beam is proportional to the square of the amplitude of this electromagnetic wave, while it does not contain information about the phase. This is the crystallographic phase problem. Methods of structure solution aim at finding the crystal structure on the basis of the measured intensities $I_{obs}(h, k, l)$ of the Bragg reflections. Solving the crystal structure implies solving the phase problem, as the phases of the Bragg reflections can be computed from the structure model and the structure model is obtained by Fourier transform of the Bragg reflections, if both amplitudes and phases of the Bragg reflections are known.

Methods of structure solution depend on a few fundamental properties of matter, which is introduced in Section 1.3. Furthermore, the methods of structure solution discussed in this chapter require the availability of the diffracted intensities of a sufficient number of Bragg reflections, as they can be measured, for example, by single-crystal X-ray diffraction. On the other hand, diffraction by microcrystalline powders results in diffracted intensity as a function of the scattering angle 2θ. Different Bragg reflections with equal or nearly equal scattering angles cannot be resolved, with the consequence that intensities of individual Bragg reflections cannot be obtained by powder diffraction for at least part of the reflections. Methods of structure solution have been developed, which account for the peculiarities of powder diffraction; these are discussed in Chapter 2.

Twinning is another feature, which prevents the measurement of intensities of individual Bragg reflections. Instead, apparent Bragg reflections of a twinned crystal may be the sum of two or more different reflections. For formal reasons, the methods of structure solution discussed in this chapter do not apply to this kind of diffraction data. Nevertheless, they may often give the correct solution (e.g., for inversion twins of a noncentrosymmetric crystal) or may lead to some average structure, which then provides the essential clue for the construction of the true structure model. However, twinning may also prohibit structure solution. Special problems and solutions related to twinning are not discussed in this chapter.

Apart from the crystal structure, intensities of Bragg reflections depend on a series of other effects, such as the geometry of the diffraction experiment, the polarization of the radiation, and absorption of X-rays by the sample. Different reflections may require different correction factors accounting for these effects, but these can be computed without knowledge of the crystal structure. It is understood

that $I_{obs}(h, k, l)$ has been obtained from the real measured intensity values by the application of these correction factors. Other, trivial dependencies are the proportionality of intensities of Bragg reflections to the intensity of the primary beam, to the volume of the crystal, and to the time of the measurement. Factors such as these affect all Bragg reflections in the same way, and they go into the scale factor (Section 1.5).

This chapter concentrates on structure determination by single-crystal X-ray diffraction. However, the same or similar methods can also be successful when applied to neutron diffraction or electron diffraction data.

1.2 The Electron Density

The elastic scattering of X-rays is determined by the electron density distribution in space. The periodicity of crystal structures determines that scattering is concentrated in directions represented by scattering vectors equal to lattice vectors of the reciprocal lattice of the crystal,

$$\mathbf{H} = h\,\mathbf{a}^* + k\,\mathbf{b}^* + l\,\mathbf{c}^* \tag{1.1}$$

The integers $(h\,k\,l)$ are used for indexing of the Bragg reflections. Non-Bragg scattering is experimentally eliminated by subtracting a background measured close to each Bragg reflection from the measured Bragg intensity, eventually leading to $I_{obs}(h, k, l)$ (Section 1.1).

The amplitudes and phases of the scattered waves of the Bragg reflections are given by the Fourier coefficients of the periodic electron density $\rho(\mathbf{x})$ as

$$F(h, k, l) = \int_{cell} \rho(\mathbf{x}) \exp\left[2\pi i\,(hx + ky + lz)\right] \mathrm{d}\mathbf{x} \tag{1.2}$$

The integration is performed over one unit cell. The structure factors $F(h, k, l)$ are obtained in units of the amount of scattering by one electron, with $F(0, 0, 0)$ equal to the number of electrons in the unit cell. The electron density can be computed by inverse Fourier transform of the structure factors,

$$\rho(\mathbf{x}) = \frac{1}{V_{cell}} \sum_{hkl} F(h, k, l) \exp\left[-2\pi i\,(hx + ky + lz)\right] \tag{1.3}$$

The volume of the unit cell is V_{cell} and the summation runs over all Bragg reflections.

The electron density possesses an infinite number of Fourier coefficients $F(h, k, l)$, obtained by enumerating h, k, and l over all integers. However, the values of the magnitudes $|F(h, k, l)|$ tend to 0 for increasing length of the scattering vector (Eq. 1.1),

$$|\mathbf{H}| = 2\,\frac{\sin(\theta)}{\lambda} \tag{1.4}$$

where θ is half the scattering angle (Figure 1.1b) and λ is the wavelength of the radiation used in the diffraction experiment. In practice, it thus appears sufficient to

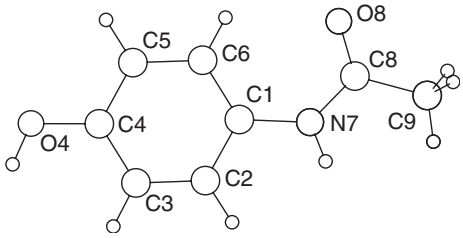

Figure 1.2 Molecular structure of paracetamol, $C_8H_9NO_2$.

include in the summation of Eq. (1.3) the structure factors of all Bragg reflections with lengths of scattering vectors less than some upper limit. The resolution of diffraction data is often described by the maximum value of $\frac{\sin(\theta)}{\lambda}$ or by the minimum value reached for

$$d_{hkl} = \frac{1}{2\sin(\theta)/\lambda} \tag{1.5}$$

Another popular designation for resolution is the maximum scattering angle, from which the actual resolution of the data can be obtained if the wavelength of the radiation is known (Eq. 1.5).

Table 1.1 compares the various measures of resolution, for values ranging from poor to excellent resolutions. Fourier maps have been computed for paracetamol (Figure 1.2), employing an extensive data set of accurately measured intensities of Bragg reflections recently published in Ref. [2]. A series of Fourier maps has been computed with the data restricted to selected resolutions as indicated in Table 1.1. Figure 1.3 shows for each map the section through the plane of phenyl ring. It

Table 1.1 Resolution of diffraction data in dependence of the maximum scattering angle for Mo–K_α and Cu–K_α radiation.

$2\theta_{max}$ (degree)		d_{min}	$\frac{\sin(\theta)}{\lambda}$	Paracetamol (Figures 1.3 and 1.12)				
					Fourier map		Charge flipping	
Mo–K_α	Cu–K_α	(Å)	(Å)$^{-1}$	n_{ref}	ρ_{min}	ρ_{max}	ρ_{min}	ρ_{max}
20	44.3	2.05	0.244	93	−0.59	3.30	–	–
30	68.3	1.37	0.364	295	−1.30	6.90	−1.45	6.06
40	95.8	1.04	0.481	654	−0.56	11.91	−0.97	10.99
52	144.0	0.81	0.617	1336	−0.73	17.04	–	–
55	180.0	0.77	0.650	1554	–	–	–	–
60	–	0.71	0.704	1940	−0.84	19.97	−1.03	18.67
70	–	0.62	0.807	2827	–	–	–	–
93	–	0.49	1.021[a]	4947	−0.46	32.09	−0.72	28.73

The number of reflections (n_{ref}) and the maximum and minimum values of the electron density (electrons/Å^3) are given for paracetamol. X-ray diffraction data from [2].
[a]This is the minimum resolution required for charge density studies.

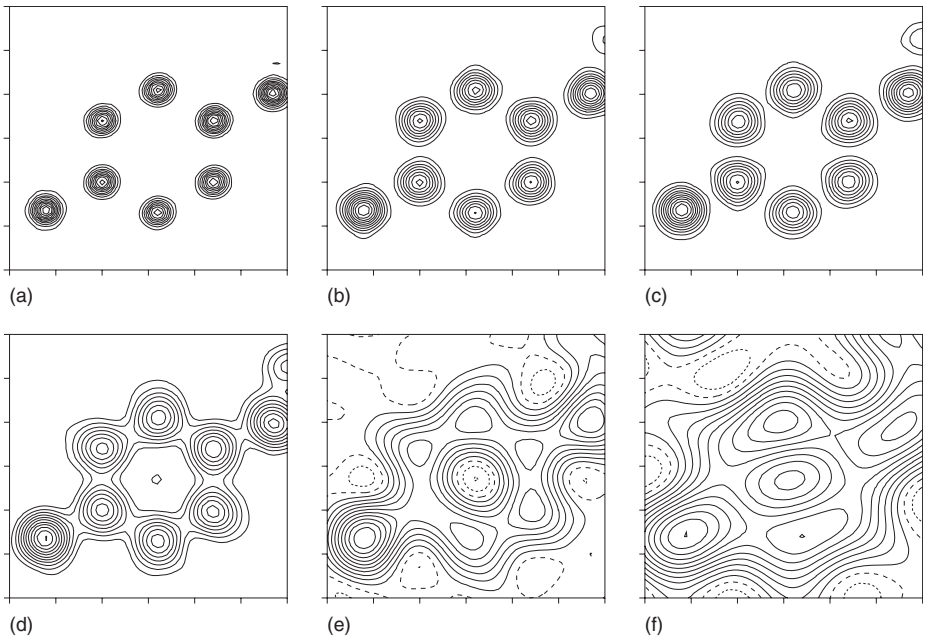

Figure 1.3 Fourier map of paracetamol in dependence on the resolution of the diffraction data. Sections are shown of 6×6 Å^2 through the phenyl ring. Contour lines are at $\frac{1}{11}$ of the respective maximum densities, with dashed lines for negative values and long-dashed lines for the value 0. (a) 1.020 Å^{-1}, (b) 0.704 Å^{-1}, (c) 0.617 Å^{-1}, (d) 0.481 Å^{-1}, (e) 0.364 Å^{-1}, and (f) 0.244 Å^{-1}. See Table 1.1 for details.

illustrates several properties of Fourier maps. Primarily, the locations of the atoms in the structure correspond to local maxima in the Fourier map. This property can be used to obtain a structure model once phases of reflections have been determined. The widths of the local maxima are determined by the combination of thermal motion and atomic size. However, the widths increase on decreasing resolution of the data. For typical resolutions of data sets of organic and inorganic compounds ($[\sin(\theta)/\lambda]_{max} = 0.6{-}0.8$ Å^{-1}), a major contribution to the widths of the local maxima comes from these series termination effects rather than from thermal motion (Figure 1.3a-c), and Fourier maps are not the proper tool for the estimation of atomic displacement parameters (ADPs).

Even at a resolution of $[\sin(\theta)/\lambda]_{max} = 0.48$ Å^{-1} ($d_{min} = 1$ Å), series termination effects do not hamper the development of a structure model on the basis of the Fourier map. At still lower resolutions, an increasing number of atomic maxima is replaced by broad features that have their maximum values at positions where there are no atoms. Thus, low-resolution Fourier maps cannot be used to directly generate a structure model, even if the phases of the reflections are correct, as it is the case for the Fourier maps given in Figure 1.3. Resolutions between 1 and 2 Å are typical for well-diffracting protein crystals. Figure 1.1(e), (f) indicates that these resolutions are insufficient for an ab initio structure determination solely on

the basis of the diffraction data. However, if we know the approximate location of a phenyl group in the structure, Fourier maps of poor resolutions – such as that of Figure 1.3f – can be used to determine the most probable orientation and the most probable position of this group within the unit cell. A variety of methods have been developed, which incorporate additional structural information – in particular the structure formula of the compound – in the process of structure determination (Sections 1.4 and 1.7).

Electron densities are strictly positive maps because they give the number of electrons present at each point in the unit cell. On the other hand, the example of paracetamol shows that Fourier maps contain regions of negative "density" (Table 1.1 and Figure 1.3). These regions occur as the result of series termination effects. Taking the minimum value of the density as a measure for the magnitude of the noise in the map, it appears that the latter is not a simple function of resolution but that it is rather the ratio between ρ_{max} and $|\rho_{min}|$ that increases on increasing resolution (Table 1.1). Better resolutions thus improve the likeliness to find atoms by way of local maxima in the map, but they still suffer from noise in the low-density regions. Strictly positive Fourier maps (Eq. 1.3) are obtained for resolutions d_{min} better than approximately 0.1 Å or with $\left[\sin(\theta)/\lambda\right]_{max}$ larger than 5 $Å^{-1}$ [3].

1.3 Diffraction and the Phase Problem

Structure factors, $F(h,k,l)$, are in general complex numbers, that can be represented by real and imaginary parts or by amplitudes and phases, according to Eq. (1.2)

$$F(h,k,l) = A(h,k,l) + i\,B(h,k,l) = |F(h,k,l)|\exp[i\,\phi(h,k,l)] \tag{1.6}$$

The two notations of complex numbers are related by

$$\begin{cases} A = |F|\cos(\phi) \\ B = |F|\sin(\phi) \end{cases} \qquad \begin{cases} |F| = \sqrt{A^2 + B^2} \\ \tan(\phi) = \frac{B}{A} \end{cases} \tag{1.7}$$

The diffraction experiment measures the intensities of the Bragg reflections. The intensity of radiation is proportional to the square of the amplitude. Observed structure factor amplitudes are thus defined as

$$|F_{obs}(h,k,l)| = \sqrt{I_{obs}(h,k,l)} \tag{1.8}$$

The diffraction experiment does not contain information on the phases $\phi(h,k,l)$. This is the crystallographic phase problem. It prevents calculation of the electron density of a crystal from its measured diffraction by Fourier inversion (Eq. 1.3).

Methods of structure solution aim at finding the phases of the structure factors from the knowledge of the measured diffraction intensities (Eq. 1.8). Since the amplitude does not contain information about the phase of a complex number, additional information must be taken into account for any method to work. This information is the following two properties of the electron density:

1) the electron density is everywhere positive;
2) the electron density has atomic character.

Randomly chosen values for the phases of the reflections lead to equal probabilities for any point of the Fourier map (Eq. 1.3) being positive or negative. Requiring a positive map thus provides a severe restriction on the acceptable combinations of phases of reflections. However, by itself it does not provide a solution to the phase problem, and finding the correct phases is the subject of the various methods of structure solution.

Atoms are at the basis of our understanding of matter. Translated to electron densities, it means that the latter can be described as the sum of atomic densities, $\rho_\mu(\mathbf{x})$, each centered on a different position in space: the position $\mathbf{x}_\mu$ of atom μ. This is the independent atom model (IAM) for the crystal structure, and the electron density

$$\rho(\mathbf{x}) = \sum_{\mu=1}^{N} \rho_\mu(\mathbf{x} - \mathbf{x}_\mu) \tag{1.9}$$

is an excellent approximation to the true density, requiring only small corrections because of the effects of chemical bonding [4]. The summation extends over all atoms N in the unit cell. The Fourier transform of the IAM can be evaluated term by term, resulting in the calculated structure factors, $F_{\text{cal}}(h,k,l)$, of the structure model (Eqs. (1.2) and (1.9))

$$F_{\text{cal}}(h,k,l) = \sum_{\mu=1}^{N} f_\mu(|\mathbf{H}|) \exp\left[-\frac{1}{4} B_\mu |\mathbf{H}|^2\right] \exp\left[2\pi i(hx_\mu + ky_\mu + lz_\mu)\right] \tag{1.10}$$

where $f_\mu(|\mathbf{H}|)$ is the atomic form factor. The isotropic temperature parameter B_μ is a positive quantity and should be replaced by the tensor of anisotropic ADPs in accurate models. Equation (1.10) shows that the intensities and phases of Bragg reflections are determined by the positions and ADPs of the atoms in the unit cell. Different atomic form factors are required for different chemical elements, but they do not depend on the positions of the atoms.

The correct structure model should give amplitudes of the calculated structure factors, $|F_{\text{cal}}(h,k,l)|$ equal to the corresponding observed structure factors. Because the latter are proportional to the primary intensity of the diffraction experiment, equality is only obtained after the application of a, yet unknown, scale factor K (Eqs. (1.8) and (1.10))

$$\sqrt{I_{\text{obs}}(h,k,l)} = |F_{\text{obs}}(h,k,l)| = K\,|F_{\text{cal}}(h,k,l)| \tag{1.11}$$

Various methods exist, which solve the phase problem. Different methods rely in different ways on the two properties of the electron density as mentioned above. Here, we discuss the fundamental principles that are at the foundations of these methods. It should give an idea about the possibilities and pitfalls of each method and should provide a guide for selecting the most promising method for each problem.

The Patterson function can be computed from the diffracted intensities and thus is directly accessible from the experiment (Section 1.6). Structure models on the basis of the Patterson function can be determined by Search methods, whereby the orientations and positions of known fragments of molecules are determined on the basis of their matching with the experimental Patterson function (Section 1.7). Direct methods directly determine the phases of the reflections on the basis of the measured intensities (Section 1.8). They rely on the statistical properties of the structure factors, as these follow from the properties of positivity and atomic character of the density (Section 1.5). Charge flipping and low-density elimination are modern methods that involve manipulations of the density (Section 1.9), simultaneously solving for the reflection phases and the electron density. Other methods aim at finding a structure model (positions of the atoms in the unit cell) without solving the phase problem first. They are particularly important for solving crystal structures from powder diffraction and in the case of low-resolution data sets as are usually the only available data for protein crystals.

1.4 Fourier Cycling and Difference Fourier Maps

Methods of structure determination often result in approximate values for the phases of reflections. Alternatively, they may provide a partial structure model. The phases of the calculated structure factors of the partial model (Eq. 1.2) can then serve as an approximation to the phases of the true structure factors. Fourier maps based on measured amplitudes $|F_{obs}(h, k, l)|$ and approximate phases of the reflections often allow an interpretation that leads to a structure model that is better than the (partial) structure model used for generating the phases. The new model can then be used to compute better phases (Eq. 1.8) and an improved Fourier map (Eq. 1.3). Repeating this procedure until convergence may eventually lead to a complete structure model of sufficient accuracy to initiate a successful refinement of the crystal structure. Structure refinements can often improve the atomic coordinates and ADPs of a partial structure model. A fruitful procedure for model completion can thus involve the alternate application of Fourier cycling and structure refinement.

Fourier maps may fail to disclose the positions of light atoms in the presence of heavy atoms. This failure may be due to the fact that the expected local maximum of the light atom is obliterated by the noisy features of the density of the heavy atom. In other cases, light atoms close to heavy atoms may not constitute local maxima in the density, and thus cannot be identified in Fourier maps by principle. This particularly applies to hydrogen atoms covalently bonded to atoms such as carbon, nitrogen, or oxygen [5]. The Fourier map of paracetamol is a good example of the latter feature, where the hydrogen atoms are apparently invisible at all resolutions (Figure 1.3).

The absence of local maxima for light atoms in Fourier maps does not imply that the diffraction data would not contain information on the location of these atoms;

it only demonstrates that Fourier maps are not the proper method to visualize light atoms next to heavy atoms. A solution to this problem is provided by the difference Fourier map. The latter is calculated with the difference between $F_{obs}(h,k,l)$ and $F_{cal}^{part}(h,k,l)$ of a partial structure model replacing $F(h,k,l)$ in Eq. (1.3):

$$\Delta\rho(\mathbf{x}) = \frac{1}{V_{cell}} \sum_{hkl} \left[F_{obs}(h,k,l) - K\, F_{cal}^{part}(h,k,l) \right] \exp\left[-2\pi i\,(hx + ky + lz)\right] \qquad (1.12)$$

Unlike the electron density, the electron difference density $\Delta\rho(\mathbf{x})$ should contain regions of both positive and negative values. Positive values indicate regions of density missing in the model and negative values indicate regions of too much density in the model. The latter usually occur because of inaccuracies in the positions of the atoms or yet incorrectly assigned values of ADPs.

As an example, we consider the structure of paracetamol (Section 1.2). Structure refinements have been performed against data at several resolutions of a partial structure model incorporating all nonhydrogen atoms. With the exception of the lowest resolution, Fourier maps are indistinguishable from those in Figure 1.3, and they have not been further analyzed. Minor differences have been found for the lowest resolution, which are due to variations of the structure model and the scale factor in the refinements when using a small selection of the data. These results show that a partial structure model can give phases of Bragg reflections, which are good approximations to the true phases.

A difference Fourier map was calculated for each resolution according to Eq. 1.12. Hydrogen atoms appear as local maxima in $\Delta\rho(\mathbf{x})$ for resolutions down to $[\sin(\theta)/\lambda]_{max} = 0.48$ Å^{-1} ($d_{min} = 1$ Å) (Figure 1.4a – d). These difference Fourier maps thus clearly show that the diffraction data contain information about the hydrogen atoms and that $\Delta\rho(\mathbf{x})$ can be used to locate these atoms. For data of resolution $[\sin(\theta)/\lambda]_{max} = 0.36$ Å^{-1}, local maxima in $\Delta\rho(\mathbf{x})$ suggest the positions of the hydrogen atoms, but it might already be difficult to distinguish these positions among the noisy features in the map, when the locations of the light atoms would not have been a priori known (Figure 4e,f). The hydrogen atoms cannot be located in the difference Fourier map with data of resolution $[\sin(\theta)/\lambda]_{max} = 0.24$ Å^{-1} ($d_{min} = 2.0$ Å).

The difference Fourier maps of resolutions better than $[\sin(\theta)/\lambda]_{max} = 0.70$ Å^{-1} exhibit local maxima at the midpoints of the C–C bonds of the phenyl ring (Figure 1.4a,b). This represents the effect of chemical bonding on the electron density as compared to the density of the IAM. These features can be incorporated into the model through the so-called multipole model, [4] or they can be described by the maximum entropy method (MEM) [5].

1.5 Statistical Properties of Diffracted Intensities

The atomic character of matter allows for a statistical analysis of the structure factors, eventually arriving at probability distribution functions (pdfs) for several

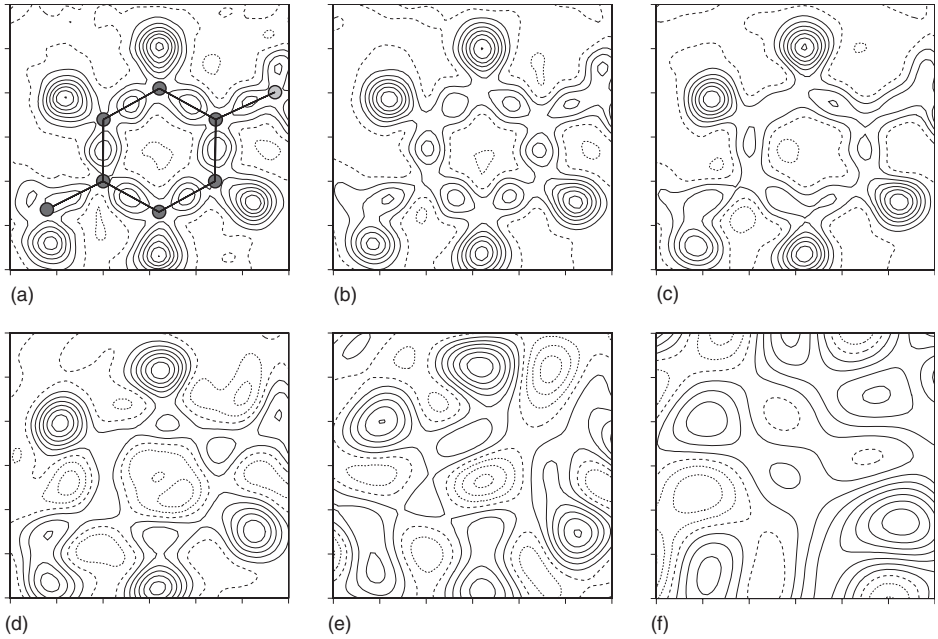

Figure 1.4 Difference Fourier map of paracetamol in dependence on the resolution of the diffraction data, obtained after refinement of a model containing nonhydrogen atoms only. Sections are shown of 6×6 Å^2 through the phenyl ring. Contour lines are at $\frac{1}{7}$ of the respective maximum densities, with dashed lines for negative values and long-dashed lines for the value 0. (a) 1.020 Å^{-1}, (b) 0.704 Å^{-1}, (c) 0.617 Å^{-1}, (d) 0.481 Å^{-1}, (e) 0.364 Å^{-1}, and (f) 0.244 Å^{-1}. In panel (a), the structure model is superimposed as dots and sticks. Compare with Figure 1.3.

quantities, such as the scattered intensity or the sum of phases of three matching reflections. These probability distributions form the foundation of direct methods (Section 1.8).

Consideration of the expression for the structure factor in Eq. 1.10 shows that each atom contributes one term that is the product of three factors: the atomic form factor, the Debije–Waller factor, and the phase factor. Both the atomic form factor and the Debije–Waller factor are positive real-valued functions, which gradually decrease on increasing length of the scattering vector. Because the phase factor is a complex number of magnitude 1, these properties imply that the magnitudes of structure factors are – on the average – smaller at high values of $\sin(\theta)/\lambda$ than at low values. This property of structure factors has been formalized in the so-called Wilson plot that describes the averaged scattered intensity as a function of the length of the scattering vector. Realizing that the magnitudes of the observed and calculated structure factors are related by the scale factor K (Eq. 1.11), the Wilson plot is [6]

$$\ln[G] = \ln\left[\frac{1}{\mathcal{S}_2}\left\langle \tfrac{1}{p_{\mathbf{H}}}|F_{\mathrm{obs}}(\mathbf{H})|^2\right\rangle_H\right] = \ln\left[K^2\right] - \frac{1}{2}BH^2 \tag{1.13}$$

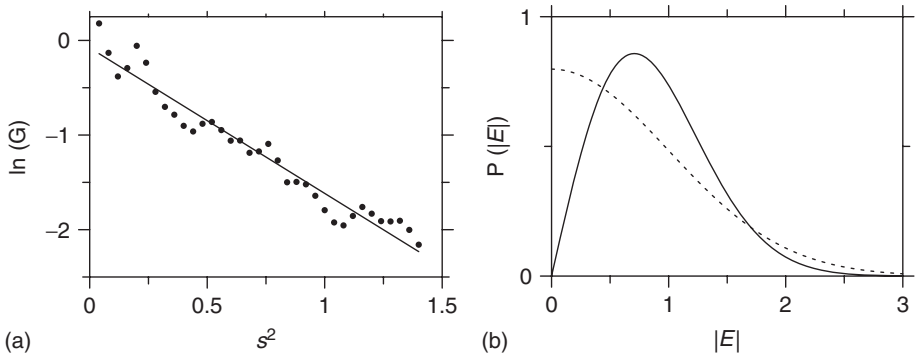

Figure 1.5 (a) Wilson plot of paracetamol with data from [2] averaged in intervals of 0.04 Å^{-2} in $s^2 = [\sin(\theta)/\lambda]^2$. The line represents a fit of Eq. (1.13) to the data points. (b) The Probability distribution function of the amplitudes of normalized structure factors for acentric (solid curve; Eq. (1.16)) and centrosymmetric (dashed curve; Eq. (1.17)) crystals.

where B is the average or overall ADP. The symmetry enhancement factor $p_{\mathbf{H}}$ accounts for different multiplicities of different reflections. The combined scattering power of the atoms in the unit cell is given by

$$\mathcal{S}_2 = \sum_{\mu=1}^{N} \left[f_\mu H)\right]^2 \tag{1.14}$$

This expectation value of the intensity of reflection $\mathbf{H}$ is $\left\langle |F(\mathbf{H})|^2 \right\rangle_H$. The expectation value depends on the length of the scattering vector as indicated by the subscript H. An experimental estimate for this quantity is obtained by the average value of the magnitude $|F_{\text{obs}}(\mathbf{H})|^2$, where the average is taken for all reflections with a scattering vector of length H. For any given real number H, usually none, but at most a few reflections will exist that have a scattering vector with exactly this length. Therefore, the average intensity is computed as an average over all reflections with lengths of their scattering vectors falling within a suitably chosen interval around H. A linear fit to the values of the left-hand side of Eq. (1.13) as a function of $[\sin(\theta)/\lambda]^2 = \frac{1}{4}H^2$ then gives the scale factor and the average ADP. The Wilson plot thus allows the scale of the experimental data and the overall ADP of the structure to be determined from the knowledge of the contents of the unit cell (Figure 1.5a). The Wilson plot has been extended toward incommensurate crystals, then involving additional parameters for the average modulation amplitudes [7].

Intensities of individual reflections are not expected to follow the linear dependence of Eq. (1.13) and Figure 1.5a. Instead, one expects that they cover a wide range about the average value. This property is used for the definition of normalized structure factors, $E(\mathbf{H})$, which are given by the corresponding structure factor divided by the square root of the average intensity according to (Eq. 1.13)

$$E_{\text{obs}}(\mathbf{H}) = \frac{F_{\text{obs}}(\mathbf{H})}{\sqrt{p_{\mathbf{H}}\mathcal{S}_2}\, K\, \exp\left[-\frac{1}{4}BH^2\right]} \tag{1.15}$$

Normalized structure factors bring the scattering of every crystalline material on the same scale. Unified probability relations exist for this quantity, which do not depend on the crystal structure. The chemical composition enters indirectly through the factor $\mathcal{S}_2$ in the definition of $E(\mathbf{H})$ (Eqs. (1.14) and (1.15)).

Disregarding prior information about the crystal structure, all points in the unit cell have the same probability to be the site of an atom. For a sufficiently large structure, the positions of the atoms can be considered as being independent from each other. This feature has been used in statistical analyses of the diffraction, resulting in pdfs for the amplitudes and the phases of structure factors [6]. The pdf for the amplitudes $|E|$ of normalized structure factors is defined as the probability of $|E|$ to have values between $|E|$ and $|E| + \mathrm{d}|E|$. For an acentric structure (space group $P1$) this pdf is [6]

$$ {}_1P_{|E|}(|E|)\ \mathrm{d}|E| = \ 2|E|\ \exp\left[-|E|^2\right]\ \mathrm{d}|E| \tag{1.16} $$

The distribution of normalized structure factors is found to be independent of the structural parameters and of the filling of the unit cell (Figure 1.5b). The integral of ${}_1P_{|E|}(|E|)$ over all possible values of its argument, that is, for $|E| : 0 \to \infty$, gives 1, because a reflection will have a value of its normalized structure factor with certainty.

Deviations from the acentric pdf will occur when correlations between structural parameters exist. An important type of correlation is the space group symmetry. In particular, the presence of an inversion center determines that the phase of each structure factor is restricted to one out of two possible values, instead of being a continuous variable in the acentric case. The centric pdf (space group $P\overline{1}$) for amplitudes of normalized structure factors is (Figure 1.5b) [6],

$$ {}_{\overline{1}}P_{|E|}(|E|)\ \mathrm{d}|E| = \ \sqrt{\frac{2}{\pi}}\ \exp\left[-\frac{|E|^2}{2}\right]\ \mathrm{d}|E| \tag{1.17} $$

For other nontrivial symmetries, the form of the pdf will depend on the space group [8], even if the symmetry enhancement factor is taken into account.

The simple forms of the statistical functions (Eqs. (1.13), (1.16), and (1.17)) have been derived under the assumption of a random filling of the unit cell with equal atoms. Any violation of this assumption will lead to deviations of pdfs from those given above. Space group symmetry is one important source of correlations between the atoms. Other deviations from random filling may pertain to the presence in the structure of both heavy and light atoms, or to the presence of molecular groups, such as the phenyl group.

The second fundamental assumption of the statistical analyses of diffraction data is that expectation values can be computed as averages over many reflections. Accordingly, proper statistics require narrow intervals of $[\sin(\theta)/\lambda]^2$ or $|E|$, each containing many reflections. These are contradictory requirements and their violation – as it is intrinsic to structures with small and intermediately sized unit cells – is the source of important deviations from the distributions in Eqs. (1.13), (1.16), and (1.17).

Violations of both assumptions explain the deviations from linearity of the Wilson plot for paracetamol (Figure 1.5a). In general, the discrepancy between the measured diffraction data and their expected statistical properties is one reason why direct methods may fail to solve the phase problem, as it happens to be the case for a small percentage of compounds.

1.6 The Patterson Function

The crystallographic phase problem prevents the calculation of the electron density by the inverse Fourier transform (Section 1.3). Nevertheless, the inverse Fourier transform of the intensities of Bragg reflections can be computed as (Eq. (1.3))

$$P(\mathbf{u}) = \frac{1}{V_{\text{cell}}} \sum_{hkl} |F(h,k,l)|^2 \exp\left[-2\pi i\,(hu + kv + lw)\right] \tag{1.18}$$

where positions in space are now described by the vector $\mathbf{u} = (u, v, w)$. $P(\mathbf{u})$ is the Patterson function [8]. The importance of the Patterson function is that it can be computed from the experimental data without further assumptions. This calculation does not solve the phase problem, but the interpretation of features of the Patterson function can be used to develop a structure model, from which reflection phases can be obtained through $F_{\text{cal}}(h,k,l)$ (Eq. 1.10).

The Patterson function can alternatively be defined on the basis of the electron density,

$$P(\mathbf{u}) = \frac{1}{V_{\text{cell}}} \int_{V_{\text{cell}}} \rho(\mathbf{r})\,\rho(\mathbf{r}+\mathbf{u})\,\mathrm{d}\mathbf{r} \tag{1.19}$$

This second definition has the interpretation of a ''sum'' (as approximation to the integral) over the unit cell of the product of the electron density at one point with the electron density at a point displaced over $\mathbf{u}$. Large contributions to the integral occur if both $\rho(\mathbf{r})$ and $\rho(\mathbf{r}+\mathbf{u})$ have large values. Large values of the electron density are concentrated at the positions of the atoms, while most of the unit cell has only a small density. As a result, the Patterson function possesses local maxima for $\mathbf{u}$ equal to interatomic vectors, that is, for $\mathbf{u}$ connecting one atom with another (Figure 1.6a). Equation (1.19) shows that the Patterson function is periodic with the same lattice as the crystal structure. In all analyses, it is therefore sufficient to consider $P(\mathbf{u})$ over one unit cell.

The number of peaks in the Patterson function is $N(N-1)+1$, if the unit cell contains N atoms. $N(N-1)$ local maxima occur at nonzero $\mathbf{u}$. One additional peak is present at $\mathbf{u} = (0,0,0)$. The origin peak has contributions from all atoms and thus represents the absolute maximum of the Patterson function, with a value that is generally much higher than any of the local maxima. Knowledge of the locations of all peaks uniquely defines the structure model. However, overlap of peaks is an intrinsic property of Patterson functions. Positions can thus be determined for only part of the peaks, and this prevents the use of the

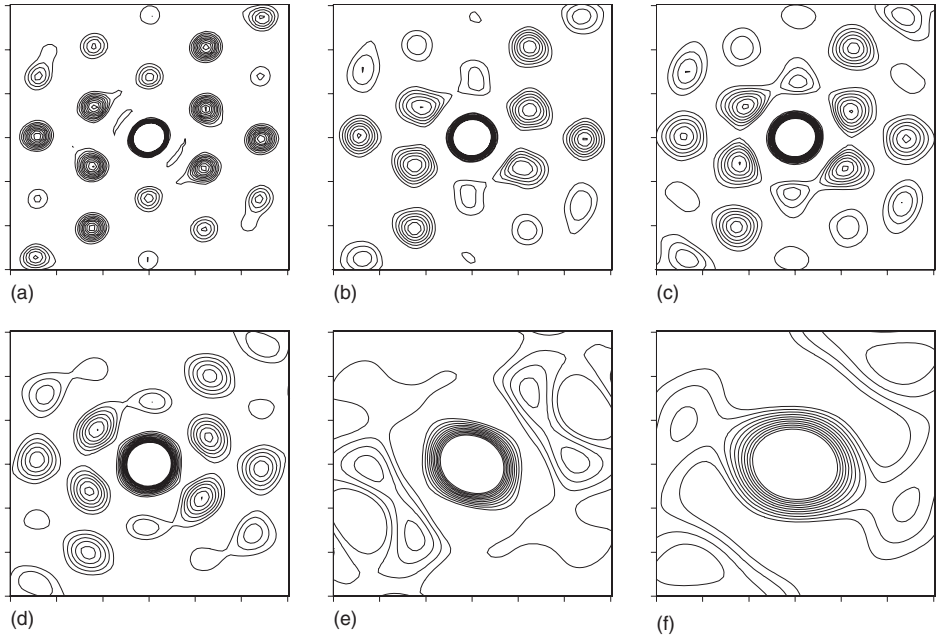

Figure 1.6 Patterson function of paracetamol in dependence on the resolution of the diffraction data (Eq. 1.18) and calculated with $|F_{obs}|^2$ from [2]. Sections are shown of 6×6 Å^2 parallel to the plane of phenyl ring and centered on the origin. (a) 1.020 Å^{-1}, (b) 0.704 Å^{-1}, (c) 0.617 Å^{-1}, (d) 0.481 Å^{-1}, (e) 0.364 Å^{-1}, and (f) 0.244 Å^{-1}. Contour lines are at $\frac{1}{11}$ of the maximum densities except the origin peak.

Patterson function as an ab initio method of structure determination in many cases.

These properties are illustrated by the Patterson function of paracetamol. Figure 1.6a displays a section of the Patterson function centered on the origin and parallel to the plane of the phenyl ring. Next to the origin peak, 16 peaks are observed. A phenyl ring contains 6 atoms and at least $6 \times 5 = 30$ peaks would have been expected in this section. Comparison of Figures 1.2 and 1.3 shows that C–C bonds of the phenyl ring are pairwise parallel, thus leading to perfect overlap of the corresponding peaks in the Patterson function. Furthermore, the O and N atoms bonded to the phenyl ring provide C–O and C–N bonds that are parallel to and of nearly equal length as one pair of C–C bonds of the phenyl ring. This leads to further overlaps of peaks in the Patterson function, thus explaining the many ''missing'' peaks in this section and the different heights of the observed local maxima, the latter being roughly proportional to the number of overlapping peaks. Figure 1.6 furthermore shows that, on decreasing resolution of the data, the Patterson function loses its resolving power much sooner than Fourier maps do (compare to Figure 1.3). A reduction in the number of local maxima is observed at a resolution of $[\sin(\theta)/\lambda]_{max} = 0.48$ Å^{-1}, while Patterson maps at lower resolutions do not even suggest the presence of a phenyl group. The resolving power

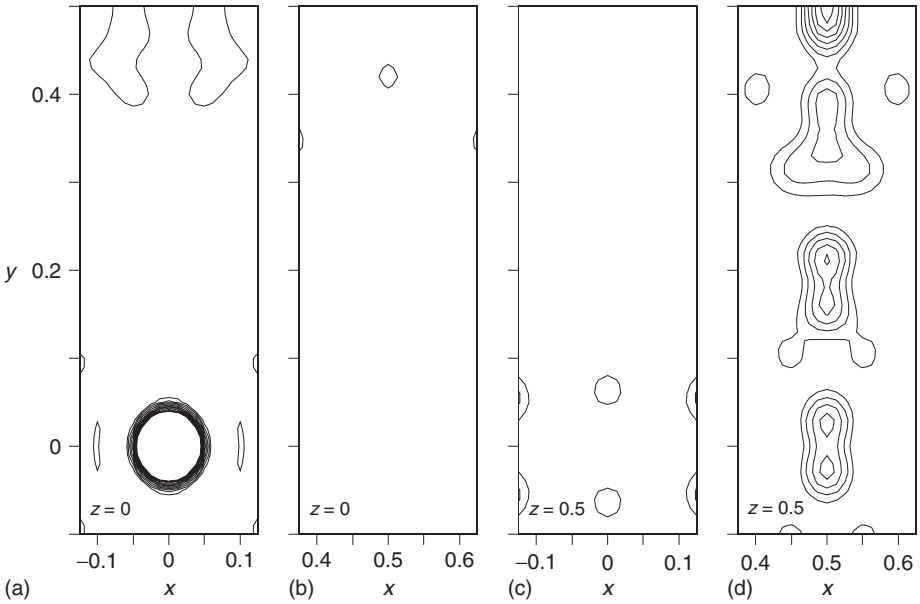

Figure 1.7 Patterson function of paracetamol obtained with diffraction data of a resolution of $[\sin(\theta)/\lambda]_{\max} = 1.020$ Å^{-1}. Sections parallel to the x, y-plane are centered on four different points as indicated. Panel (a) contains the origin peak. Contour lines are at $\frac{1}{11}$ of the highest peak except the origin peak.

is improved for the so-called sharpened Patterson maps, which exhibit narrower peaks than the Patterson map itself. They are computed through the replacement of $|F_{\text{obs}}|^2$ by corresponding E-values or through multiplication $|F_{\text{obs}}|^2$ by an "inverse" Debije–Waller factor containing an overall ADP [9].

The Patterson function exhibits the full symmetry of the crystal structure. The point symmetry is the same as that of the diffraction pattern; the Patterson function is thus centrosymmetric. Nontrivial point symmetry leads to the concentration of Patterson peaks in planes or on lines. Consider, for example, a mirror plane perpendicular to **b**. The N atoms in the unit cell can then be divided into $N/2$ pairs of atoms with coordinates (x, y, z) $(x, -y, z)$, defining $N/2$ Patterson peaks on the line $\mathbf{u} = (0, 2y, 0)$. With glide operators replacing the mirror, the lines containing $N/2$ Patterson peaks are $\left(\frac{1}{2}, 2y, 0\right)$, $\left(0, 2y, \frac{1}{2}\right)$, and $\left(\frac{1}{2}, 2y, \frac{1}{2}\right)$ for an *a*-glide, *b*-glide, and *n*-glide, respectively. These lines are called *Harker lines*. Sections of the Patterson function of paracetamol containing these four lines clearly show high values on only the line $\left(\frac{1}{2}, 2y, \frac{1}{2}\right)$ (Figure 1.7). This implies the presence of an *n*-glide, in agreement with the space group of paracetamol. The positions of these $N/2$ Patterson peaks would immediately give the y-coordinates of the atoms. However, these coordinates cannot be determined for paracetamol because the $44/2 = 22$ peaks are too crowded on the line $\left(\frac{1}{2}, v, \frac{1}{2}\right)$ and individual peaks cannot be resolved (Figure 1.7d).

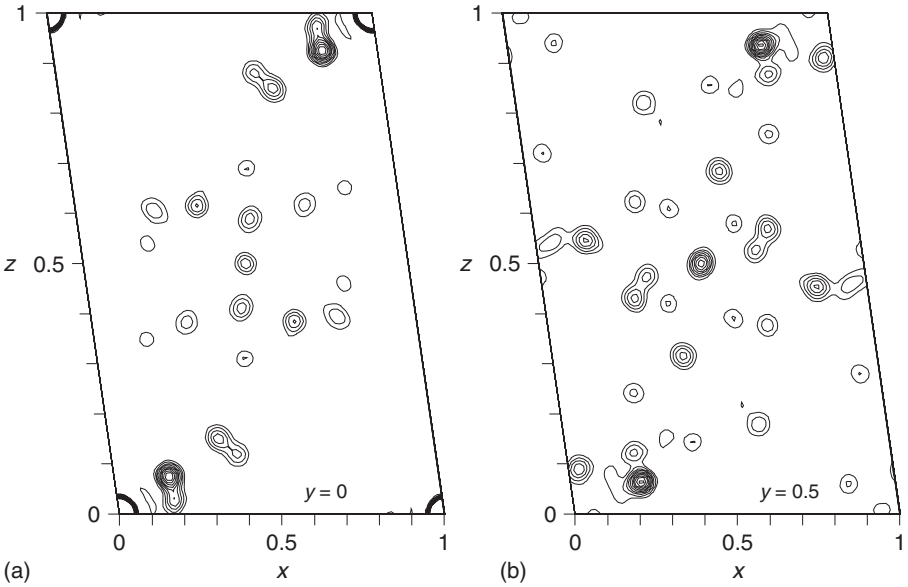

Figure 1.8 Patterson function of paracetamol obtained with diffraction data of a resolution of $[\sin(\theta)/\lambda]_{max} = 1.020$ Å^{-1}. (a) Section $y = 0$ and (b) section $y = 0.5$. Contours as in Figure 1.7.

A rotation or screw axis leads to the concentration of Patterson peaks in Harker planes perpendicular to the direction of the axis. For paracetamol, the larger number of Patterson peaks in the plane $(2x, \frac{1}{2}, 2z)$ than in $(2x, 0, 2z)$ is in agreement with the presence of a 2_1 screw axis in its space group (Figure 1.8). Harker planes would allow the determination of the x- and z-coordinates of all atoms, if all peaks could be resolved. This is not the case for paracetamol, and the use of the Patterson function as ab initio method of structure solution fails again. Even if Patterson peaks are not resolved, Harker lines and Harker planes allow the determination of the point symmetry of the structure and the intrinsic translations of the symmetry elements. Since the Patterson function is based on all diffraction information, this is a much more robust method than the analysis of reflection conditions for the determination of the intrinsic translations.

1.7 Patterson Search Methods

Once partial information is available about the structure, Patterson search methods can successfully be used for completing the structure. One type of information is the molecular formula of the compound. For the example of paracetamol, this includes the phenyl ring, which together with the N and O atoms, forms a planar rigid group of which the structure is well known from the crystal structures of other

compounds containing phenyl rings. While the Patterson function of this group suffers from many overlaps of peaks (Section 1.6), it also displays a typical pattern of local maxima close to the origin and arranged within a plane that is parallel to the plane of the phenyl ring (Figure 1.6a). The determination of the plane of this typical pattern within the Patterson function then provides the orientation of the phenyl ring in the crystal structure.

An automated search for the orientation of a typical pattern in the Patterson function can be achieved by the rotation function [10]. The theoretical Patterson pattern of the rigid group, $P_{rg}(\mathbf{u})$ is rotated by the rotation R until the integral

$$RI = \int P(\mathbf{u})\, P_{rg}(R\mathbf{u})\, d\mathbf{u} \tag{1.20}$$

attains its maximum value. The integration is restricted to a sphere that just covers all peaks of $P_{rg}(\mathbf{u})$. A maximum value of RI is obtained if all local maxima of $P_{rg}(\mathbf{u})$ are matched with local maxima of the Patterson function. The rotation bringing this coincidence then defines the sought orientation of the rigid group. The translation function can be used in a similar way to determine the position of a rigid group within the unit cell, after its orientation has been found through the rotation function [10].

Of more general applicability is the superposition minimum function, SMF($\mathbf{u}$), defined as [11]

$$\mathrm{SMF}(\mathbf{u}) = \mathrm{Min}\{P(\mathbf{u} - \mathbf{u}_1), P(\mathbf{u} - R\mathbf{u})\} \tag{1.21}$$

At each point in space the minimum value is taken from the values of the shifted Patterson function, $P(\mathbf{u} - \mathbf{u}_1)$, and a Harker line or Harker section, $P(\mathbf{u} - R\mathbf{u})$, where R is one of the rotational symmetry operators of the crystal. More than one term of each type can be included in the definition of the SMF, and the SMF can be used to compare different functions. The integral over $\mathbf{u}$ of the SMF of $P(\mathbf{u})$ and $P_{rg}(R\mathbf{u})$ can be used instead of the rotation function (Eq. 1.20) in a rotation search of a known fragment. For $\mathbf{u}_1$ equal to an interatomic vector between heavy atoms, the SMF can be used as a method of phase determination.

Modern computer programs for ab initio phase determination include both Patterson search methods and Direct methods (Section 1.8), as it is the case, for example, for SHELXS and SHELXD [9], SIR [11], and SnB [12].

1.8 Direct Methods

Direct methods aim at the determination of phases of structure factors from the knowledge of their (measured) amplitudes. The first problem is the feature that phases of structure factors are not a given quantity but depend on the choice of the origin of the coordinate system. For example, the rumor that reflection phases are equal to 0 or π for centrosymmetric crystals is only true if the origin has been chosen on an inversion center. Therefore, care should be taken that an admissible origin is chosen, which conforms the space group symmetry [1].

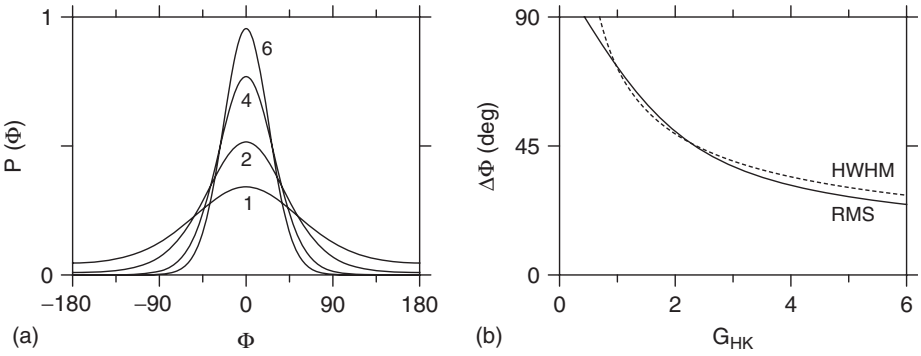

Figure 1.9 (a) The Cochran distribution describing the probability distribution for the value of the triplet phase $\Phi_{\mathbf{HK}}$ (in degree) for several values of $G_{\mathbf{HK}}$, as indicated (Eqs. (1.23) and (1.24)). (b) Half-width-at-half-maximum (HWHM; dashed line) and the square root of the second moment (RMS; full line) of the Cochran distribution in dependence on $G_{\mathbf{HK}}$.

A second result of the above observation is that general relations do not exist concerning the phases of single reflections. Instead, properties need to be considered, which are independent from the choice of the origin. The most important of these relations exists for groups of three reflections, called triplets, related by

$$\mathbf{H}_1 + \mathbf{H}_2 + \mathbf{H}_3 = 0 \tag{1.22}$$

The sum of phases of these three reflections is a structure invariant with a value independent from the choice of the origin. The condition (Eq. 1.22) is fulfilled for any $\mathbf{K}$, if we choose $\mathbf{H}_1 = \mathbf{H}$, $\mathbf{H}_2 = -\mathbf{K}$, and $\mathbf{H}_3 = -(\mathbf{H} - \mathbf{K})$. The triplet phase becomes

$$\Phi_{\mathbf{HK}} = \phi_{\mathbf{H}} - \phi_{\mathbf{K}} - \phi_{\mathbf{H}-\mathbf{K}} \tag{1.23}$$

While the value of $\Phi_{\mathbf{HK}}$ does not depend on the origin, it is still not known. Here, probability considerations enter, which state that $\Phi_{\mathbf{HK}}$ is close to 0, with a probability that increases for increasing values of the amplitudes of the normalized structure factors contributing to the triplet under consideration. More precisely, the probability distribution for $\Phi_{\mathbf{HK}}$ of noncentrosymmetric structures has a maximum for $\Phi_{\mathbf{HK}} = 0$ and a width that decreases with increasing $G_{\mathbf{HK}}$,

$$G_{\mathbf{HK}} = \frac{2}{\sqrt{N}} |E_{\mathbf{H}}||E_{\mathbf{K}}||E_{\mathbf{H}-\mathbf{K}}| \tag{1.24}$$

The Cochran distribution is illustrated in Figure 1.9 for several values of $G_{\mathbf{HK}}$. If more than one triplet contributes to a reflection $\mathbf{H}$, the phase of its structure factor can be estimated with the tangent formula [6].

Phases for up to three reflections must be given numerical values, which then determine the origin. Direct methods involve the multiple application of Eq. (1.23) or the tangent formula. Useful triplets are those for which $G_{\mathbf{HK}}$ is sufficiently large, like possessing values larger than a lower bound between 1 and 2. Further

information is included through quartets and other structure invariants. Similar but different equations apply if the structure is built from more than one element, and in case of centrosymmetric symmetries [6]. Further extensions include entropy maximization for improved estimates of probabilities [13].

Equation (1.24) shows that $G_{\mathbf{HK}}$ decreases for all triplets with increasing size of the unit cell, even if $|E|$-values are large. Direct methods will eventually fail, because all $G_{\mathbf{HK}}$ are too small, as it is the case for protein crystals. The largest structures solved by direct methods incorporate about 1000 atoms in the unit cell [9].

1.9 Charge Flipping and Low-Density Elimination

Density modifications techniques can be considered as a special case of Fourier cycling (Section 1.4). Each iteration includes the modification of the electron density on a pixel by pixel basis, instead of creating a modified structure model, as it is done in the classical method of Fourier cycling. Basis of these methods thus is a discrete representation of the electron density on a sufficiently fine grid over the unit cell, for example, with a pixel size better than 0.1 Å.

Each cycle of the iterative procedure starts with the observed structure factor amplitudes and a set of phases that follow from the previous cycle (Figure 1.10). Iterations are initiated with approximate phases, as they have been obtained by any of the methods discussed in this chapter. An inverse Fourier transform results in the nth estimate of the density, which may include negative regions due to erroneous phases, series termination effects, and errors in the observed structure factor amplitudes. The essential step is the modification of this density according to some recipe (discussed in the following text). Calculated structure factors for the modified density then provide phases that can be combined with the observed structure factor amplitudes in the $(n + 1)$th cycle of the procedure [14–16]. Density

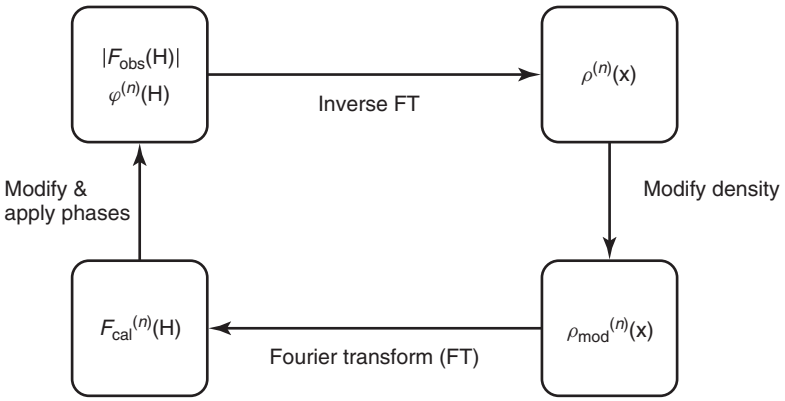

Figure 1.10 The iterative procedure of density modification.

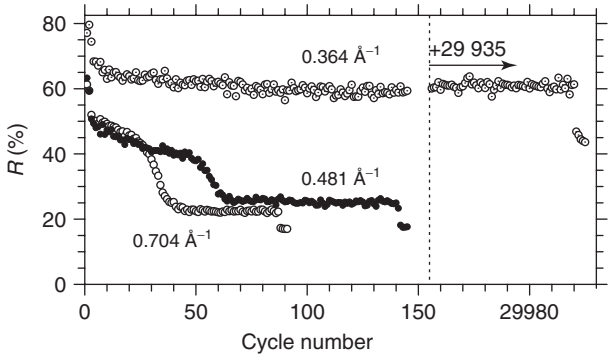

Figure 1.11 *R* index as a function of the iteration cycle for runs of charge flipping with diffraction data of paracetamol of three different resolutions, as indicated. Compare to Table 1.1.

modification techniques simultaneously solve for the reflection phases and the electron density.

Low-density elimination replaces the density by zero for all pixels with a negative density or with a density less than some positive threshold δ. It is not very powerful as ab initio method of phase determination, but it has been successfully used for the improvement of the density, especially after solving the phase problem by charge flipping.

The method of charge flipping starts with a set of random phases. In each cycle the density is replaced by its negative value for all pixels with a density less than some positive threshold δ. That is, negative densities become positive and small positive densities become negative. The threshold δ is chosen as a fraction of the value of the density in the local maxima of light atoms, but it should be above the noise level as followed from the inverse Fourier transform of the observed data. Charge flipping converges to a bistable state, where the density in the low-density regions alternates between positive and negative values in consecutive cycles. A converged solution is obtained if charge flipping is followed by a few cycles of low-density elimination. The performance of charge flipping has been found to improve considerably, if – in each cycle – the calculated phases of some of the reflections are modified, for example, a phase shift of 90° is applied to the phases of weak reflections.

Progress of the iterations can be monitored by a conventional *R* index on the observed structure factor amplitudes. After a sharp decrease in the first few cycles, *R* stays at a high value for many cycles. Convergence is indicated by a rapid drop of *R* toward a lower value (Figure 1.11). This "low" value is typically between 20% and 30%. A further drop of the *R* index occurs after the few cycles of low-density elimination.

The charge flipping procedure is applied to the diffraction data of paracetamol at several different resolutions, employing the computer program superflip [15]. For resolutions down to $\left[\sin(\theta)/\lambda\right]_{\max} = 0.481$ Å^{-1}, charge flipping converges and

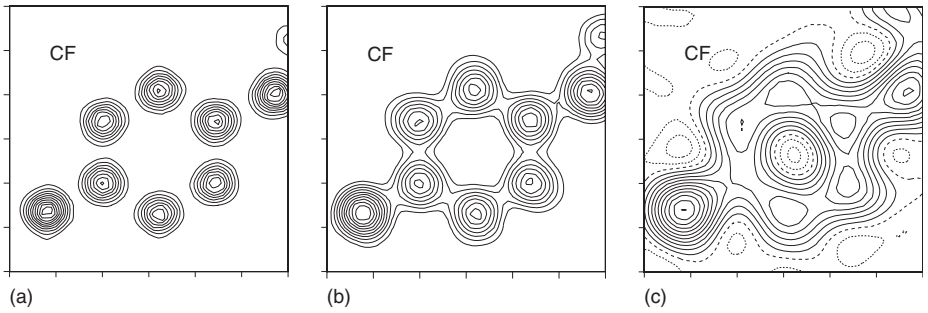

Figure 1.12 Density maps after charge flipping performed with diffraction data of resolution $[\sin(\theta)/\lambda]_{max}$ equal to (a) 0.704 Å^{-1} CF, (b) 0.481 Å^{-1} CF, and (c) 0.334 Å^{-1} CF. Sections are shown of 6 × 6 Å^2 through the phenyl ring. Contour lines are at $\frac{1}{11}$ of the respective maximum densities, with dashed lines for negative values and long-dashed lines for the value 0. See Table 1.1 for details.

the density produced by charge flipping is very similar to the density obtained by Fourier transform of the corresponding set of diffraction data (compare Figures 1.3 and 1.12 and see Table 1.1). Local maxima in the density can be identified for all nonhydrogen atoms, and the structure can be considered as solved.

For a resolution of $\left[\sin(\theta)/\lambda\right]_{max} = 0.364$ Å^{-1}, charge flipping indicates the failure of convergence, even after 300 000 iteration cycles (Figure 1.11). The final five cycles of low-density elimination reduce R from ∼60% to ∼44%, more than twice as high as the R value after the successful runs with data of higher resolutions. Nevertheless, the final density is nearly indistinguishable from the Fourier map obtained at the same resolution (Figures 1.3 and 1.12 and Table 1.1; see also a second section of the density in Figure 1.13). This result indicates that charge flipping has successfully solved the phase problem for this particular data set of low resolution. However, it did not solve the crystal structure, because the Fourier map fails to show local maxima for all atoms, but instead gives unresolved, broad features near the positions of the atoms.

Charge flipping has several advantages over other methods of structure solution. In particular, it does not require advance knowledge of the chemical content of the unit cell. Instead, the integrated charges around the local maxima can be used for estimating the chemical element to be assigned to each peak. Second, charge flipping does not require prior knowledge about the symmetry. The algorithm works in space group $P1$. Possible symmetry elements are determined through the analysis of the resulting density map [15]. Charge flipping has successfully been applied to the solution of a series of compounds, and it has been used for the solution of the structures of aperiodic crystals [7, 15]. Similar to direct methods and Patterson-based methods, charge flipping fails for data from twinned crystals (Section 1.1).

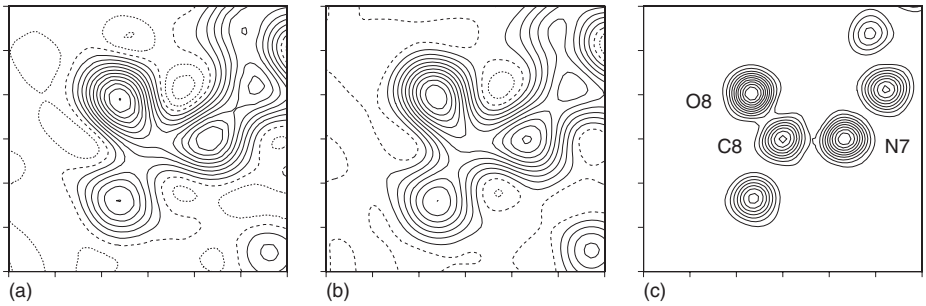

Figure 1.13 (a) Density map after charge flipping performed with diffraction data of resolution $[\sin(\theta)/\lambda]_{\max} = 0.334\,\text{Å}^{-1}$. Fourier maps of resolutions of (b) $0.334\,\text{Å}^{-1}$ and (c) $0.704\,\text{Å}^{-1}$. Sections are shown of $6 \times 6\ \text{Å}^2$ through the atoms N7–C8–O8. Contour lines are at $\frac{1}{11}$ of the respective maximum densities, with dashed lines for negative values and long-dashed lines for the value 0. See Table 1.1 for details.

1.10 Outlook and Summary

Methods of structure determination from single-crystal diffraction data have seen a continuous development since the original proposal of the Patterson function in 1934 and of Direct methods in the 1950s. These developments are closely associated with the increase in available computational power over the years. The modern versions of direct methods and Patterson search techniques and, in particular, methods based on density modification have become feasible only with the computational power, as it has become available since about 1990. Nowadays, most problems can be solved on a personal computer.

Organic, organometallic, and most inorganic crystal structures can be solved automatically by any of the above mentioned methods. As computational power becomes less of an issue, Charge flipping in combination with low-density elimination is likely to become the method of choice [15], because it is least biased by possibly incorrect assumptions about the symmetry or the content of the unit cell. Failure of all of these methods can be expected if light and heavy atoms occur in one structure (for example, in tungsten oxides), and the use of difference Fourier maps might be necessary for finding the positions of the light atoms.

A fundamental problem for all methods is large structures, that is, more than several hundred atoms in the unit cell. The phasing power decreases on increasing size of the unit cell, as it is most clearly demonstrated by the pdf of the triplet phase, which becomes nearly flat for large numbers of atoms in the unit cell (Eq. 1.24 and Figure 1.9). This is the situation for protein crystals. Alternative methods of structure solution are required, such as the heavy atom method, the isomorphous replacement, and the use of multiple wavelength anomalous dispersion (MAD) methods [17].

Apart from the typical problems encountered in the field of protein crystallography, the major cause for failure of methods of structure solution is data quality. Missing information or incorrect values for the observed structure factor

amplitudes of the few strongest, low-order reflections may hamper direct methods and charge flipping. In any case, they lead to noisy features in Fourier and Patterson maps, eventually preventing atomic maxima and artifacts to be distinguished from each other.

A poor crystal quality or restrictions imposed by nonambient sample environments can be responsible for a limited resolution of the diffraction data (Table 1.1). Series termination effects then cause the disappearance of atomic maxima in Fourier maps of resolutions considerably below about $[\sin(\theta)/\lambda]_{max} = 0.48$ Å^{-1} (Figure 1.3). For paracetamol, it has been shown here that charge flipping can solve the phase problem for data sets of limited resolutions where they fail to solve the crystal structure (Section 1.9). This failure has been identified as exactly the series termination effects of the Fourier map with limited data, while reflection phases appear to be mostly correct. Data sets of limited resolutions have recently been described as a cause for the failure of direct methods to solve crystal structures. It was demonstrated that data set extension methods repair at least part of these problems [18]. Therefore, it is likely that direct methods still solve the phase problem for data of limited resolutions but that series termination effects in Fourier maps cause the disappearance of atomic maxima and thus prevent the structure to be determined. Data extension methods precisely aim at repairing this problem [18]. Alternatively, the MEM could be used for this purpose [7].

The most hideous problem is twinning, especially merohedral of pseudomerohedral twinning. Diffraction data from a twinned crystal do not give the amplitudes of individual structure factors but provide sums of intensities of different reflections instead. Methods of structure determination discussed in this chapter could fail. However, they still might give a solution that then is incorrect. Twinning is often not obvious from the diffraction experiment. Problems in finding a reasonable crystal structure from the diffraction data are often the first clue that the crystal is actually twinned. A general method does not exist for structure solution of twinned crystals, except that powder diffraction by principle does not suffer from twinning. Methods developed for structure solution from powder diffraction data might thus be adapted for twinned single-crystal diffraction data. If the size of the unit cell is not too large, powder diffraction itself might be used in these cases.

References

1. Giacovazzo, C. (ed.) (2002) *Fundamentals of Crystallography*, 2nd edn, Oxford University Press, Oxford.
2. Bouhmaida, N., Bonhomme, F., Guillot, B., Jelsch, Ch., and Ghermani, N.E. (2009) *Acta Crystallogr., Sect. B*, **65**, 363–374.
3. de Vries, R.Y., Briels, W.J., and Feil, D. (1996) *Phys. Rev. Lett.*, **77**, 1719–1722.
4. Coppens, P. (1997) *X-ray Charge Densities and Chemical Bonding*, Oxford University Press, Oxford.
5. Hofmann, A., Netzel, J., and van Smaalen, S. (2007) *Acta Crystallogr., Sect. B*, **63**, 285–295.
6. Giacovazzo, C. (1998) *Direct Phasing in Crystallography*, Oxford University Press, Oxford.

7. van Smaalen, S. (2007) *Incommensurate Crystallography*, Oxford University Press, Oxford.
8. Shmueli, U. (2007) *Theories and Techniques of Crystal Structure Determination*, Oxford University Press, Oxford.
9. Sheldrick, G.M. (2008) *Acta Crystallogr., Sect. A*, **64**, 112–122.
10. Rossmann, M.G. and Arnold, E. (2006) Patterson and molecular-replacement techniques, in *International Tables for Crystallography*, **vol. B**, (ed. U. Shmueli), Kluwer Academic Publishers, Dordrecht, pp. 235–263.
11. Burla, M.C., Caliandro, R., Camalli, M., Carrozzini, B., Cascarano, G.L., De Caro, L., Giacovazzo, C., Polidori, G., and Spagna, R. (2005) *J. Appl. Crystallogr.*, **38**, 381–388.
12. Miller, R., Gallo, S.M., Khalak, H.G., and Weeks, C.M. (1994) *J. Appl. Crystallogr.*, **27**, 613–621.
13. Bricogne, G. (1997) *Methods Enzymol.*, **276**, 361–423.
14. Oszlanyi, G. and Suto, A. (2008) *Acta Crystallogr., Sect. A*, **64**, 123–134.
15. Palatinus, L. and Chapuis, G. (2007) *J. Appl. Crystallogr.*, **40**, 786–790.
16. Zhang, K.Y.J., Cowtan, K.D., and Main, P. (2006) Phase improvement by iterative density modification, in *International Tables for Crystallography*, **vol. F**, (eds M.G. Rossmann and E. Arnold), Kluwer Academic Publishers, Dordrecht, pp. 311–324.
17. Rupp, B. (2009) *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*, Garland Science, London.
18. Caliandro, R., Carrozzini, B., Cascarano, G.L., Giacovazzo, C., Mazzone, A., and Siliqi, D. (2009) *J. Appl. Crystallogr.*, **42**, 302–307.