International Tables for Crystallography (2012). Vol. F, Section 18.4.1, pp. 485–487.

Chapter 18.4. Refinement at atomic resolution

Z. Dauter, G. N. Murshudov and K. S. Wilson

18.4.1. The atomic model and a definition of atomic resolution

18.4.1.1. The atomic model

X-rays are diffracted by the electrons that are distributed around the atomic nuclei, and the result of an X-ray crystallographic study is the derived three-dimensional electron-density distribution in the unit cell of the crystal. The elegant simplicity and power of X-ray crystallography arise from the fact that molecular structures are composed of discrete atoms that are treated as spherically symmetric in the usual approximation. This property places such strong restraints on the Fourier transform of the crystal structures of small molecules that the phase problem can be solved by knowledge of the amplitudes alone.

Each atom or ion can be described by up to 11 parameters (Table 18.4.1.1).

The first parameter is the scattering-factor amplitude for the chemical nature of the atom in question, and has been computed and tabulated for all atom types [*International Tables for Crystallography*, Volume C (2004)]. Once the chemical identity of the atom is established, this parameter is fixed.

The next three parameters relate to the positional coordinates of the atom with respect to the origin of the unit cell.

If the resolution is high enough, then the number of observed reflections is sufficient to allow six anisotropic atomic displacement parameters to be used to describe the distribution of the atom positions in different unit cells (Fig. 18.4.1.1). Atomic displacement parameters (ADPs) reflect both the thermal vibration of atoms about the mean position as a function of time (dynamic disorder) and the variation of positions between different unit cells of the crystal arising from its imperfection (static disorder). Contributors to the apparent ADP ($U_{\rm atom}$) can be thought of as follows (Murshudov *et al.*, 1999):

$$U_{\rm atom} = U_{\rm crystal} + U_{\rm TLS} + U_{\rm torsion} + U_{\rm bond}, \qquad (18.4.1.1)$$

where $U_{\rm crystal}$ represents the fact that a crystal itself is in general an anisotropic field that will result in the intensity falling off in an anisotropic manner, $U_{\rm TLS}$ represents a translation/libration/screw (TLS), *i.e.* the overall motion of molecules or domains (Schomaker & Trueblood, 1968; Winn *et al.*, 2001), $U_{\rm torsion}$ is the oscillation along torsion angles and $U_{\rm bond}$ is the oscillation along and across bonds. In principle, all these contributors are highly correlated and it is difficult to separate them from one another. Nevertheless, an understanding of how $U_{\rm atom}$ is a sum of these different components makes it possible to apply atomic anisotropy parameters at different resolutions in a different manner. For example, $U_{\rm crystal} + U_{\rm TLS}$ can be applied at any resolution, as their refinement increases the number of parameters by at most five for $U_{\rm crystal}$ and 20 per independent moiety for $U_{\rm TLS}$. In contrast, refinement of the third contributor does pose a problem, as there is strong correlation between different torsion angles. As an alternative, ADPs along the internal degrees of freedom could in principle be refined. The fourth and final contributor, $U_{\rm bond}$, can only be refined at very high resolution. In real applications, $U_{\rm crystal}$ and $U_{\rm TLS}$ are separated for convenient description of the system, but in practice their effects are indistinguishable.

In the special case when the tensor $U_{\rm atom}$ is isotropic, *i.e.*, all non-diagonal elements are equal to zero and all diagonal terms are equal to each other, then the atom itself appears to be isotropic and its ADP can be described using only one parameter, $U_{\rm iso}$.

Thus, for a full description of a crystal structure in which all atoms only occupy a single site, nine parameters per atom must be determined: three positional parameters and six anisotropic ADPs. This assumes that the spherical-atom approximation applies and ignores the so-called deformation density resulting from the non-spherical nature of the outer atomic and molecular orbitals involved in the chemical interactions between the atoms (Coppens, 1997).

For disordered regions or features, where atoms can be distributed over two or more identifiable sites, the occupancy introduces a tenth variable for each atom. In many cases, the fractional occupancies are not all independent, but are rather constant for sets of covalently or hydrogen-bonded atoms or for those in non-overlapping solvent networks. This would apply, for example, to partially occupied ligands or side chains with two conformations.

Table 18.4.1.1
The parameters of an atomic model

Parameter type	Number	Variable or fixed
Atom type	1	Fixed after identification
Positional (x, y, z)	3	Variable, subject to restraints
ADPs:		
isotropic	1	Variable beyond about 2.5 Å
anisotropic	6	Variable beyond about 1.5 Å
Occupancy	1	Variable for visible disorder

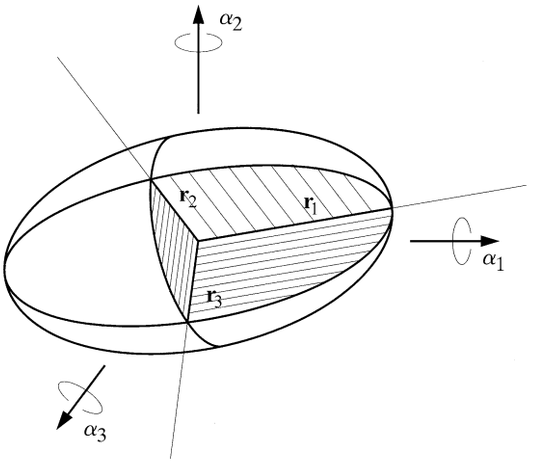

Figure 18.4.1.1
The thermal-ellipsoid model used to represent anisotropic atomic displacement, with major axes indicated. The ellipsoid is drawn with a specified probability of finding an atom inside its contour. Six parameters are necessary to describe the ellipsoid: three represent the dimensions of the major axes and three the orientation of these axes. These six parameters are expressed in terms of a symmetric U tensor and contribute to atomic scattering through the term $\exp[-2\pi^2(U_{11}h^2a^{*2} + U_{22}k^2b^{*2} + U_{33}l^2c^{*2} + 2U_{12}hka^*b^*\cos\gamma^* + 2U_{13}hla^*c^*\cos\beta^* + 2U_{23}klb^*c^*\cos\alpha^*)]$.

18.4.1.2. What is 'atomic resolution'?

Atomicity is the great simplifying feature of crystallography in terms of structure solution and refinement. For a small-molecule structure, accurate X-ray data usually extend to 0.8 Å, and this has three important implications for crystallography.

(1) *Ab initio phasing using direct methods*. Automatic *ab initio* solution of the phase problem depends on the assumption of positivity and atomicity of the electron density. The fact that current *ab initio* methods in the absence of heavy atoms are only effective when meaningful data extend beyond 1.2 Å reinforces the idea that this is a reasonable working criterion for its definition as atomic resolution. In addition, approaches such as solvent flattening and automated map interpretation benefit enormously from such data.

(2) *Resolved atomic peaks in the Fourier maps*. Although some individual peaks can be seen at resolutions beyond ~2.0 Å, they become more fully resolved at around 1.2 Å.

(3) *Refinement of a full anisotropic model*. The number of reflections is sufficient for the minimization of the discrepancy between the experimentally determined amplitudes or intensities of the Bragg reflections and those calculated from the atomic model with up to ten (usually nine) independent parameters per atom. This has been classically achieved by least-squares refinement as described in *International Tables for Crystallography* Volume C, Chapter 8.1 (Prince & Boggs, 2004) or more recently by maximum-likelihood procedures (Bricogne & Irwin, 1996; Pannu & Read, 1996; Murshudov *et al.*, 1997). For small-molecule structures, accurate amplitude data are normally available to around 0.8 Å, giving an observation-to-parameter ratio of about seven for non-centrosymmetric crystals, which allows positional parameters to be determined with an accuracy approaching 0.001 Å. This reflects the high degree of order of such crystals, in which the molecules in the lattice are in a close-packed array. In addition the X-ray data are of high quality, with a high $I/\sigma(I)$ ratio (and hence low merging R value) even in the outer resolution shells.

It is now necessary to define what constitutes 'atomic resolution'. A pragmatic approach has been that data extending to 1.2 Å or better with at least 50% of the intensities in the outer shell being higher than 2σ is the acceptable limit (Sheldrick, 1990; Sheldrick & Schneider, 1997), which means that the statistical problem of refinement is overdetermined. This appears to remain a good working definition for refinement applications and indeed has been put on a more solid theoretical basis (Morris & Bricogne, 2003; Morris *et al.*, 2004). However, for application of direct phasing methods it is advantageous to record even a small fraction of significant reflections beyond this cutoff. These outer shells should be included in the refinement procedure with correct maximum-likelihood weights, but they will not significantly improve the effective resolution.

This is rarely achieved for crystals of macromolecules: as of October 2009 around 1250 out of 52 000 crystal structures in the Protein Data Bank (PDB) had a resolution higher than 1.2 Å compared to 157 out of 13 000 in March 2000. Firstly, the large unit-cell volume leads to an enormous number of reflections for which the average intensity is weak compared to those for small molecules (see Table 9.1.1.1 in Chapter 9.1). Secondly, the intrinsic disorder of the crystals further reduces the intensities at high Bragg angles and usually gives a resolution cutoff which is much less than atomic. Thirdly, the large solvent content leads to substantial decay of crystal quality under exposure to the X-ray beam at room temperature. While the secondary damage (resulting from the migration of ions and radicals produced by the primary absorption event) is largely avoided by vitrification of such crystals, the effect of primary damage has become significant on high intensity beamlines (see Section 9.1.12). The upper resolution limit of the data affects all stages of a crystallographic analysis, but especially restricts the features of the model that can be independently refined (Table 18.4.1.2). Solutions to the problem of refining macromolecular structures with a paucity of experimental data evolved during the 1970s and 1980s with the use of either constraints or restraints on the stereochemistry, based on that of known small molecules. With constraints, the structure is simplified as a set of rigid chemical units (Diamond, 1971; Herzberg & Sussman, 1983), whereas using restraints, the observation-to-parameter ratio is increased by introduction of prior chemical knowledge of bond lengths and angles (Konnert & Hendrickson, 1980).

Table 18.4.1.2
Features which can be seen in the electron density at different resolutions

Disordered regions will not necessarily be visible even at these limiting values. Some features should be included even at lower resolutions, *e.g.* hydrogen atoms at their riding positions can be incorporated at 2.0 Å, but their positions will not be verifiable from the density. The contents of this table should not be taken as dogmatic rules, but as approximate guidelines.

Resolution (Å)	Feature
0.8	Deformation density, *i.e.* deviation from the spherical-atom model
1.0	Hydrogen atoms
1.5	Anisotropic atomic displacement
2.0	Multiple conformations
2.5	Individual isotropic atomic displacement
3.5	Overall temperature factor
4.0	α-Helices and β-sheets
6.0	Domain envelopes

As expected, atoms with different ADPs contribute differently to the diffraction intensities, as discussed by Cruickshank (1999*a*,*b*). The relative contribution of the different atoms to a given reflection depends on the difference between their ADPs $\{\exp[-(B_1 - B_2)s^2]$, where $s = \sin\theta/\lambda\}$. Clearly, if the average ADP of a molecule is small, then the spread will also be narrow, and most atoms will contribute to diffraction over the whole range of resolution. When the mean ADP is large, then the spread of the ADPs will be wide, and fewer atoms will contribute to the high-resolution intensities (Fig. 18.4.1.2).

Three advances in experimental techniques have combined effectively to overcome these problems for an increasing number of well ordered macromolecular crystals, namely the use of high-intensity synchrotron radiation (SR), efficient two-dimensional detectors and cryogenic freezing (discussed in Parts 8, 7 and 10, respectively). These advances mean that there is no longer a sharp division between small-molecule and macromolecular crystallography, but rather a continuum from small through medium-sized structures, such as cyclodextrins and other supramolecules, to proteins. The inherent disorder in the crystal generally increases with the size of the structure, due in part to the increasing solvent content. Thus, it has become tractable to refine a significant number of protein structures at atomic resolution with a full anisotropic model (Dauter, Lamzin & Wilson, 1997; Dauter, 2003). This work of course benefits tremendously from the experience and algorithms of small-molecule crystallography, but does pose special problems of its own. The tech-

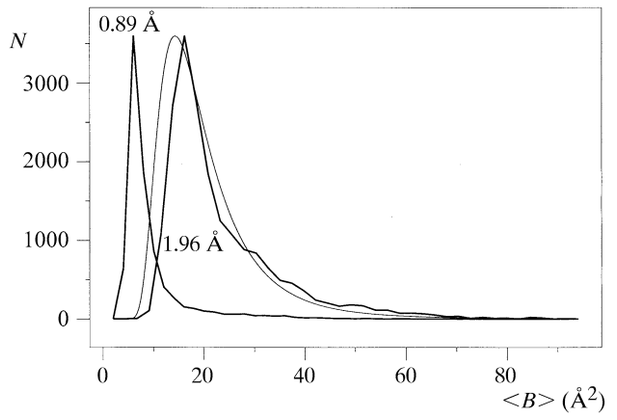

Figure 18.4.1.2
Histograms of B values for a protein structure, *Micrococcus lysodecticus* catalase (Murshudov *et al.*, 1999), for two different crystals which diffracted to different limiting resolutions. For both crystals, the resolution cutoff reflects the real diffraction limit from the sample, and hence its level of order. At 0.89 Å, the mean B value is 8.3 Å^2 and the width of the distribution is small. In contrast, at 1.96 Å, the mean B value is 25.5 Å^2 and the spread is correspondingly large. Thus, for the 0.89 Å crystal, most atoms contribute to the high-resolution terms, whereas for the 1.96 Å crystal, only the atoms with lower B values do so. The thin line shows the theoretical inverse gamma distribution $\mathrm{IG}(B) = (b/2)^{d/2}/\Gamma(d/2)B^{-(d+2)/2}\exp[-b(2B)]$, where b and d are the parameters of the distribution, and Γ is the gamma function. For this figure, the values $b = 2$ and $d = 10$ were chosen, which correspond to a mean B value of 20 Å^2 and σ_B of 11 Å^2. In the gamma distribution, the abscissa was multiplied by $8\pi^2$ to make it comparable with the measured B values. All three histograms were normalized to the same scale.

niques of solving and refining macromolecular structures thus also overlap with those conventionally used for small molecules; a prime example is the use of *SHELXL* (Sheldrick, 2008), which was developed for small structures and has now been extended to treat macromolecules.

18.4.1.3. A theoretical approach to 'atomic resolution'

An alternative and stricter definition of atomic resolution comes from using a measure of the information content of the data. There are a variety of definitions of the information in the data about the postulated model (see, for example, O'Hagan, 1994). A suitable one is the Bayesian definition for quadratic information measure:

$$I_Q(p, F) = \mathrm{tr}(A\{\mathrm{var}(p) - E[\mathrm{var}(p, F)]\}), \qquad (18.4.1.2)$$

where I_Q is the quadratic information measure, p is the vector of parameters, F is the experimental data, var(p) is the variance matrix corresponding to prior knowledge, var(p, F) is the variance matrix corresponding to the posterior distribution (which includes prior knowledge and likelihood), E is the expectation, tr is the trace operator (*i.e.* the sum of the diagonal terms of the matrix) and A is the matrix through which the relative importance of different parameters or combinations of parameters is introduced. For example, if A is the identity matrix, then the information measure is unitary and all parameters are assigned the same weight. If A is the identity matrix for positional parameters and zero for ADPs, then only the information about positional parameters is included. By appropriate choice of A, the information about selected key features, such as the active site, can be estimated.

Equation (18.4.1.2) shows how much the experiment reduces the uncertainty in given parameters. Prior knowledge is usually taken to be information about bond lengths, bond angles and other chemical features of the molecule, known before the experiment has been carried out. In the case of an experiment designed to provide information about the ligated protein or mutant, when information about differences between two (or more) different states is needed, the prior knowledge can be thought of in a different way – as knowledge about the native protein.

Unfortunately, there are problems in applying equation (18.4.1.2). Firstly, careful analysis of the prior knowledge and its variance is essential. The target values used at present, or more properly the distributions for these values, need to be re-evaluated. Another problem concerns the integration required to compute the expectation value (E). Nevertheless, the equation provides some idea of how much information about a postulated model can be extracted from a given experiment.

This alternative definition of atomic resolution assumes that the second term of equation (18.4.1.2) for positional parameters is sufficiently close to zero for most atoms to be resolved from all their neighbours. Defining atomic resolution using this information measure reflects the importance of both the quality and quantity of the data [through the posterior var(p, F)]. In addition, data may come from more than one crystal, in which case the information will be correspondingly increased. There may be additional data from mutant and/or complexed protein crystals, where, again, the information measure will be increased and, moreover, the differences between different states can be analysed. The effect of redundancy of different crystals of the same molecule(s) in different space groups is to reduce the limit of data necessary for achieving atomic resolution, which is equivalent to the advantage of noncrystallographic averaging.

Thus, in practice, while it would be ideal to develop the strict application of equation (18.4.1.2), for the present it is necessary to rely on the pragmatic approach in Section 18.4.1.2.

18.4.2. Data

The quality of the refined model relies finally on that of the available experimental data. Data collection has been covered extensively in Chapter 9.1 and will not be discussed here.

18.4.2.1. Data quality

As can be seen from equation (18.4.1.2), the measure of information about all or part of the crystal contents depends strongly on the quality and quantity of the data. Of course, before the experiment is carried out some questions should be answered. Firstly, what is the aim of the experiment? Secondly, what is the cost of the experiment and what are the available resources? With modern techniques, if SR is used with an efficient detector, the cost of the experiment for different resolutions does not vary greatly (provided that a suitable quality crystal is available). In practice, the apparent increase in cost to attain high-resolution data will generally make solving the phase problem both easier and faster. A full analysis at atomic resolution provides a wealth of additional structural detail which may shed light on the subtleties of the protein's chemistry not seen at lower resolution. However, this may require some considerable time and effort, and is an area where development of more automated approaches would be beneficial. In contrast, low-resolution data can make it difficult to answer not only the question currently being asked, but can also necessitate further experiments to address other problems that arise.