

## 18.4. REFINEMENT AT ATOMIC RESOLUTION

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. The appropriate choice of  $A$  allows the estimation of information on selected key features, such as the active site.

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) separate states is needed, the prior knowledge can be considered instead as knowledge about the native protein.

However, 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 gives some idea about 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  $\text{var}(p, F)$ ]. 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 crystal forms is to reduce the limit of data necessary for achieving atomic resolution, which is equivalent to the advantage of noncrystallographic averaging.

18.4.1.1. *Ab initio* phasing and atomic resolution

*Ab initio* methods of phase calculation normally depend on the assumption of positivity and atomicity of the electron density. Such methods rely largely on the availability of atomic resolution data. In addition, approaches such as solvent flattening and automated map interpretation benefit enormously from such data. 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 atomic resolution.

## 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 synchrotron radiation (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 provide a saving in terms of the time spent by the investigator, since the interpretation of the resulting electron density is much easier and faster. In general, to answer the same question is much easier and cheaper if high-resolution data are available. In addition, high-resolution data mean that answers to some of the questions which may arise during analysis of the experiment will already be addressable. In contrast, low-resolution data not only make it difficult to answer the question currently being asked, but may also necessitate further experiments to address other problems that arise.

While the information content of the data appears to depend quantitatively on the nominal resolution, in fact it is dependent on the data quality throughout the resolution range, and both high- and low-resolution completeness and their statistical significance affect the information content of the data and derived model. High-intensity low-resolution terms remain important for refinement at atomic resolution, as they define the contrast in the density maps between solvent and protein, and because their omission biases the refinement, especially that of parameters such as the ADPs. The rejection of low-intensity observations will have a similar biasing effect. In particular, all the maps calculated for visual or computer inspection by Fourier transformation are diminished in quality by omission of any terms, but are especially affected by omission of strong low-resolution data. This is particularly true in the early stages of structure solution, where low-resolution data can be vital. Although most phase-improvement algorithms rely on relations between all reflections, terms involving low-resolution reflections will be large, will be involved in many relations and will play a dominant role. Hence, omission of these terms will severely degrade the power of these methods, which may indeed converge to solutions that have nothing whatsoever to do with the real structure.

## 18.4.2.2. Anisotropic scaling

The intensity data from a crystal may display anisotropy, *i.e.*, the intensity fall-off with resolution will vary with direction, and may be much higher along one crystal axis than along another. If the structure is to be refined with an isotropic atomic model (either because there are insufficient data or the programs used cannot handle anisotropic parameters), then the fall-off of the calculated  $F^2$  values will, of necessity, also be isotropic. In this situation, an improved agreement between observed and calculated  $F^2$  values can be obtained either by using anisotropic scaling during data reduction to the expected Wilson distribution of intensities, or by including a maximum of six overall anisotropic parameters during refinement. This will result in an isotropic set of  $F^2$  values. For crystals with a high degree of anisotropy in the experimental data, this can lead to a substantial drop of several per cent in  $R$  and  $R_{\text{free}}$  (Sheriff & Hendrickson, 1987; Murshudov *et al.*, 1998).

This ambiguity effectively disappears with use of an anisotropic atomic model. The individual ADPs, including contributions from both static and thermal disorder, take up relative individual displacements, but also the overall anisotropy of the experimental  $F^2$  values. The significance of the overall anisotropy is a point of some contention, and its physical meaning is not clear. It may represent asymmetric crystal imperfection or anisotropic overall displacement of molecules in the lattice related to TLS parameters. Refinement of TLS parameters, which can be performed using, for example, *RESTRAIN* (Driessen *et al.*, 1989), removes the overall crystal contribution to the ADP.