15.2. Model phases: probabilities, bias and maps

BY R. J. READ

15.2.1. Introduction

The intensities of X-ray diffraction spots measured from a crystal give us only the amplitudes of the diffracted waves. To reconstruct a map of the electron density in the crystal, the unmeasured phase information is also required. In fact, the phases are much more important to the appearance of the map than the measured amplitudes. When phases are supplied by an atomic model, therefore, some degree of model bias is inevitable.

The optimal use of model phase information requires an estimate of its reliability, specifically the probability that various values of the phase angle are true. Such a probability distribution can be derived, starting first with the relationship between the structure factor (amplitude and phase) of the model and that of the true crystal structure. The phase probability distribution can then be obtained from this and used, for instance, to provide a figure-of-merit weighting that minimizes the r.m.s. error from the true electron density.

Even with figure-of-merit weighting, model-phased electron density is biased towards the model. The systematic bias component of model-phased map coefficients can be predicted, allowing the derivation of map coefficients that give electron-density maps with reduced model bias. With the help of a few simple assumptions, a correction for bias can also be made when different sources of phase information are combined.

Finally, the refinement of a model against the observed amplitudes allows a certain amount of overfitting of the data, which leads to an extra 'refinement bias'. Fortunately, the use of appropriate refinement strategies, including maximum-likelihood targets, can reduce the severity of this problem.

15.2.2. Model bias: importance of phase

Dramatic illustrations of the importance of the phase have been published. For instance, Ramachandran & Srinivasan (1961) calculated an electron-density map using phases from one structure and amplitudes from another. In this map there are peaks at the positions of the atoms in the structure that contributed the phase information, but not in the structure that contributed the amplitudes. Similar calculations with two-dimensional Fourier transforms of photographs (Oppenheim & Lim, 1981; Read, 1997) show that the phases of one completely overwhelm the amplitudes of the other.

These examples, though dramatic, are not completely representative of the normal situation, where the structure contributing the phases is partially or even nearly correct. Nonetheless, model phases always contribute bias, so that the resulting map tends to bear too close a resemblance to the model.

15.2.2.1. *Parseval's theorem*

The importance of the phase can be understood most easily in terms of Parseval's theorem, a result that is important to the understanding of many aspects of the Fourier transform and its use in crystallography. Parseval's theorem states that the mean-square value of the variable on one side of a Fourier transform is proportional to the mean-square value of the variable on the other side. Since the Fourier transform is additive, Parseval's theorem also applies to sums or differences.

If ρ_1 and ρ_2 are, for instance, the true electron density and the electron density of the model, respectively, Parseval's theorem tells us that the r.m.s. error in the electron density is proportional to the r.m.s. error in the structure factor. (The structure-factor error is a vector error in the complex plane.)

$$\langle \rho^2 \rangle = (1/V^2) \sum_{\text{all } \mathbf{h}} |\mathbf{F}(\mathbf{h})|^2,$$

$$\left\langle (\rho_1 - \rho_2)^2 \right\rangle = (1/V^2) \sum_{\text{all } \mathbf{h}} |\mathbf{F}_1(\mathbf{h}) - \mathbf{F}_2(\mathbf{h})|^2.$$

This understanding of error in electron-density maps explains why the phase is much more important than the amplitude in determining the appearance of an electron-density map. As illustrated in Fig. 15.2.2.1, a random choice of phase (from a uniform distribution of all possible phases) will generally give a larger error in the complex plane than a random choice of amplitude [from a Wilson (1949) distribution of amplitudes].

15.2.3. Structure-factor probability relationships

To use model phase information optimally, the probability distribution for the true phase (or, equivalently, the distribution of the error in the model phase) needs to be known. Such a distribution can be derived by first working out the probability distribution for the true structure factor (or the distribution of the vector difference between the model and true structure factors). Then the phase probability distribution is obtained by fixing the known value of the structure-factor amplitude and renormalizing.

A number of related structure-factor distributions have been derived, differing in the amount of information available about the structure and in the assumed form of errors in the model. These range from the Wilson distribution, which applies when none of the atomic positions is known, to a distribution that applies when there are a variety of sources of error in an atomic model.

15.2.3.1. *Wilson and Sim structure-factor distributions in P1*

For the Wilson distribution (Wilson, 1949), it is assumed that the atoms in a crystal structure in space group $P1$ are scattered randomly and independently through the unit cell. In fact, it is sufficient to make the much less restrictive assumption that the

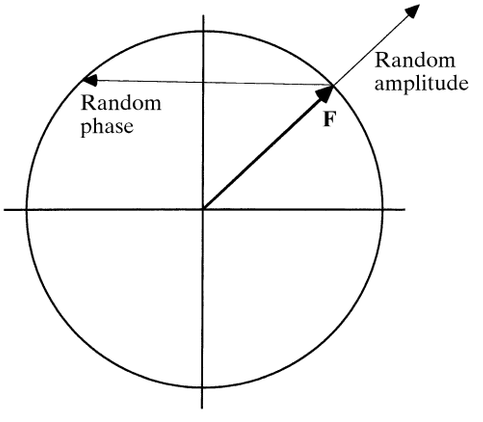

Fig. 15.2.2.1. Schematic illustration of the relative errors introduced by a random choice of phase or a random choice of amplitude. The example has been constructed to represent the r.m.s. errors introduced by randomization (computed by averages over the Wilson distribution). Phase randomization will introduce r.m.s. errors of $(2)^{1/2}$ ($\simeq 1.41$) times the r.m.s. structure-factor amplitude $|\mathbf{F}|$. By comparison, map coefficients weighted by figures of merit of zero would have r.m.s. errors equal to the r.m.s. $|\mathbf{F}|$, so a featureless map would be more accurate than a random-phase map. Amplitude randomization will introduce r.m.s. errors of $[(4-\pi)/2]^{1/2}$ ($\simeq 0.66$) times the r.m.s. $|\mathbf{F}|$, so a map computed with random amplitudes will be closer to the true map than a featureless map.

references