In Main's approach, the cosine law is used to introduce the cosine of the phase error, which is converted into a figure of merit by taking expected values. Some manipulations allow us to solve for the figure-of-merit weighted map coefficient, which is approximated as a linear combination of the true structure factor and the model structure factor (Main, 1979; Read, 1986). Finally, we can solve for an approximation to the true structure factor, giving map coefficients from which the systematic model bias component has been removed.

$$m|\mathbf{F}_O|\exp(i\alpha_C) = F/2 + D\mathbf{F}_C/2 + \text{ noise terms},$$
$$F \simeq (2m|\mathbf{F}_O| - D|\mathbf{F}_C|)\exp(i\alpha_C).$$

A similar analysis for centric structure factors shows that there is no systematic model bias in figure-of-merit weighted map coefficients, so no bias correction is needed in the centric case.

15.2.5.2. *Model bias in combined phase maps*

When model phase information is combined with, for instance, multiple isomorphous replacement (MIR) phase information, there will still be model bias in the acentric map coefficients, to the extent that the model influences the final phases. However, it is inappropriate to continue using the same map coefficients to reduce model bias, because some phases could be determined almost completely by the MIR phase information. It makes much more sense to have map coefficients that reduce to the coefficients appropriate for either model or MIR phases, in extreme cases where there is only one source of phase information, and that vary smoothly between those extremes.

Map coefficients that satisfy these criteria (even if they are not rigorously derived) are implemented in the program *SIGMAA*. The resulting maps are reasonably successful in reducing model bias. Two assumptions are made: (1) the model bias component in the figure-of-merit weighted map coefficient, $m_{\rm com}|\mathbf{F}_O|\exp(i\alpha_{\rm com})$, is proportional to the influence that the model phase has had on the combined phase; and (2) the relative influence of a source of phase information can be measured by the information content, H (Guiasu, 1977), of the phase probability distribution. The first assumption corresponds to the idea that the figure-of-merit weighted map coefficient is a linear combination of the MIR and model phase cases.

MIR:	$m_{\rm MIR}\|\mathbf{F}_O\|\exp(i\alpha_{\rm MIR})$	$\simeq \mathbf{F}$
Model:	$m_C\|\mathbf{F}_O\|\exp(i\alpha_C)$	$\simeq \mathbf{F}/2 + D\mathbf{F}_C/2$
Combined:	$m_{\rm com}\|\mathbf{F}_O\|\exp(i\alpha_{\rm com})$	$\simeq [1-(w/2)]\mathbf{F} + (w/2)D\mathbf{F}_C,$

where

$$w = H_C/(H_C + H_{\rm MIR})$$

and

$$H = \int_0^{2\pi} p(\alpha)\ln\frac{p(\alpha)}{p_0(\alpha)}\,{\rm d}\alpha; \quad p_0(\alpha) = \frac{1}{2\pi}.$$

Solving for an approximation to the true $\mathbf{F}$ gives the following expression, which can be seen to reduce appropriately when w is 0 (no model influence) or 1 (no MIR influence):

$$\mathbf{F} \simeq \frac{2m|\mathbf{F}_O|\exp(i\alpha_{\rm com}) - wD\mathbf{F}_C}{2-w}.$$

15.2.6. Estimation of overall coordinate error

In principle, since the distribution of observed and calculated amplitudes is determined largely by the coordinate errors of the model, one can determine whether a particular coordinate-error distribution is consistent with the amplitudes. Unfortunately, it turns out that the coordinate errors cannot be deduced unambiguously, because many distributions of coordinate errors are consistent with a particular distribution of amplitudes (Read, 1990).

If the simplifying assumption is made that all the atoms are subject to a single error distribution, then the parameter D (and thus the related parameter σ_A) varies with resolution as the Fourier transform of the error distribution, as discussed above. Two related methods to estimate overall coordinate error are based on the even more specific assumption that the coordinate-error distribution is Gaussian: the Luzzati plot (Luzzati, 1952) and the σ_A plot (Read, 1986). Unfortunately, the central assumption is not justified; atoms that scatter more strongly (heavier atoms or atoms with lower B factors) tend to have smaller coordinate errors than weakly scattering atoms. The proportion of the structure factor contributed by well ordered atoms increases at high resolution, so that the structure factors agree better at high resolution than if there were a single error distribution.

It is often stated, optimistically, that the Luzzati plot provides an upper bound to the coordinate error, because the observation errors in $|\mathbf{F}_O|$ have been ignored. This is misleading, because there are other effects that cause the Luzzati and σ_A plots to give underestimates (Read, 1990). Chief among these are the correlation of errors and scattering power and the overfitting of the amplitudes in structure refinement (discussed below). These estimates of overall coordinate error should not be interpreted too literally; at best, they provide a comparative measure.

15.2.7. Difference-map coefficients

The computer program *SIGMAA* (Read, 1986) has been developed to implement the results described here. Apart from the two types of map coefficient discussed above, two types of difference-map coefficient can also be produced:

(1) Model-phased difference map: $(m|\mathbf{F}_O| - D|\mathbf{F}_C|)\exp(i\alpha_C)$;

(2) General difference map: $m_{\rm com}|\mathbf{F}_O|\exp(i\alpha_{\rm com}) - D\mathbf{F}_C$.

The general difference map, it should be noted, uses a vector difference between the figure-of-merit weighted combined phase coefficient (the 'best' estimate of the true structure factor) and the calculated structure factor. When additional phase information is available, it should provide a clearer picture of the errors in the model.

15.2.8. Refinement bias

The structure-factor probabilities discussed above depend on the atoms having independent errors (or at least a sufficient number of groups of atoms having independent errors). Unfortunately, this assumption breaks down when a structure is refined against the observed diffraction data. Few protein crystals diffract to sufficiently high resolution to provide a large number of observations for every refinable parameter. The refinement problem is, therefore, not sufficiently overdetermined, so it is possible to overfit the data. If there is an error in the model that is outside the range of convergence of the refinement method, it is possible to introduce compensating errors in the rest of the structure to give a better, and misleading, agreement in the amplitudes. As a result, the phase accuracy (hence the weighting factors m and D) is overestimated, and model bias is poorly removed. Because simulated annealing is a more effective minimizer than gradient methods (Brünger *et al.*, 1987), it is also more effective at locating local minima, so structures refined by simulated annealing probably tend to suffer more severely from refinement bias.

There is another interpretation to the problem of refinement bias. As Silva & Rossmann (1985) point out, minimizing the r.m.s. difference between the amplitudes $|\mathbf{F}_O|$ and $|\mathbf{F}_C|$ is equivalent (by

Copyright © 2006 International Union of Crystallography

references