Map-likelihood phasing

Thomas C. Terwilliger

Bioscience Division, Mail Stop M888,
Los Alamos National Laboratory, Los Alamos,
NM 87545, USA

Correspondence e-mail: terwilliger@lanl.gov

The recently developed technique of maximum-likelihood density modification [Terwilliger (2000), *Acta Cryst.* D**56**, 965–972] allows a calculation of phase probabilities based on the likelihood of the electron-density map to be carried out separately from the calculation of any prior phase probabilities. Here, it is shown that phase-probability distributions calculated from the map-likelihood function alone can be highly accurate and that they show minimal bias towards the phases used to initiate the calculation. Map-likelihood phase probabilities depend upon expected characteristics of the electron-density map, such as a defined solvent region and expected electron-density distributions within the solvent region and the region occupied by a macromolecule. In the simplest case, map-likelihood phase-probability distributions are largely based on the flatness of the solvent region. Though map-likelihood phases can be calculated without prior phase information, they are greatly enhanced by high-quality starting phases. This leads to the technique of prime-and-switch phasing for removing model bias. In prime-and-switch phasing, biased phases such as those from a model are used to prime or initiate map-likelihood phasing, then final phases are obtained from map-likelihood phasing alone. Map-likelihood phasing can be applied in cases with solvent content as low as 30%. Potential applications of map-likelihood phasing include unbiased phase calculation from molecular-replacement models, iterative model building, unbiased electron-density maps for cases where $2F_o - F_c$ or σ_A-weighted maps would currently be used, structure validation and *ab initio* phase determination from solvent masks, non-crystallographic symmetry or other knowledge about expected electron density.

1. Introduction

Density-modification techniques are a firmly established and important tool for macromolecular structure determination. These methods include such powerful approaches as solvent flattening, non-crystallographic symmetry averaging, histogram matching, phase extension, molecular replacement, entropy maximization and iterative model building (Abrahams, 1996; Béran & Szöke, 1995; Bricogne, 1984, 1988; Cowtan & Main, 1993, 1996; Giacovazzo & Siliqi, 1997; Goldstein & Zhang, 1998; Gu *et al.*, 1997; Lunin, 1993; Perrakis *et al.*, 1997; Podjarny *et al.*, 1987; Prince *et al.*, 1988; Refaat *et al.*, 1996; Roberts & Brünger, 1995; Rossmann & Arnold, 1993; Szöke, 1993; Szöke *et al.*, 1997; Terwilliger, 2000; van der Plas & Millane, 2000; Vellieux *et al.*, 1995; Wang, 1985; Wilson & Agard, 1993; Xiang *et al.*, 1993; Zhang & Main, 1990; Zhang, 1993; Zhang *et al.*, 1997). The central basis of these

approaches is that prior knowledge about expected values of the electron density in part or all of the unit cell can be a very strong constraint on the crystallographic structure factors. For example, prior knowledge about electron density often consists of the identification of a region where the electron density is flat owing to the presence of disordered solvent (Wang, 1985). Real-space information of this kind has generally been used to improve the quality of crystallographic phases obtained by other means, such as multiple isomorphous replacement or multiwavelength experiments, but phase information from such real-space constraints can sometimes be so powerful as to be useful in *ab initio* phase determination (Béran & Szöke, 1995; van der Plas & Millane, 2000; Wang *et al.*, 1998; Roversi *et al.*, 2000).

2. Maximum-likelihood density modification

We recently developed maximum-likelihood density modification, a method for carrying out density modification in which the phasing information coming from various sources is explicitly kept separate (Terwilliger, 1999, 2000). This separation of phasing information allowed a statistical formulation for density modification that was very straightforward and avoided major existing difficulties with density modification. In maximum-likelihood density modification, the total likelihood of a set of structure factors $\{F_\mathbf{h}\}$ is defined in terms of three quantities: (i) any prior knowledge from other sources about these structure factors, (ii) the likelihood of measuring the observed set of structure factors $\{F_\mathbf{h}^{\text{OBS}}\}$ if this set of structure factors were correct and (iii) the likelihood that the map resulting from this set of structure factors $\{F_\mathbf{h}\}$ is consistent with our prior knowledge about this and other macromolecular structures. This can be written as

$$\text{LL}(\{F_\mathbf{h}\}) = \text{LL}^o(\{F_\mathbf{h}\}) + \text{LL}^{\text{OBS}}(\{F_\mathbf{h}\}) + \text{LL}^{\text{MAP}}(\{F_\mathbf{h}\}), \quad (1)$$

where $\text{LL}(\{F_\mathbf{h}\})$ is the log-likelihood of a possible set of crystallographic structure factors $F_\mathbf{h}$, $\text{LL}^o(\{F_\mathbf{h}\})$ is the log-likelihood of these structure factors based on any information that is known in advance, such as the distribution of intensities of structure factors (Wilson, 1949), $\text{LL}^{\text{OBS}}(\{F_\mathbf{h}\})$ is the log-likelihood of these phases given the experimental data alone and $\text{LL}^{\text{MAP}}(\{F_\mathbf{h}\})$ is the log-likelihood of the electron-density map resulting from these phases. In this formulation, density modification consists of maximizing the total likelihood given by (1).

We showed previously that the total likelihood in (1) could be maximized efficiently by an iterative procedure in which a probability distribution for each phase is calculated independently of those for all other phases in each cycle (Terwilliger, 1999, 2000). In one cycle of optimization, an electron density map is calculated using current estimates of the structure factors. Each structure factor is then considered separately from the others and a phase-probability distribution for that structure factor is calculated from the variation of the total likelihood in (1) with the phase (or phase and amplitude) of that structure factor.

3. Map-likelihood phasing

In previous applications of the maximum-likelihood density-modification approach, phase information was derived from a combination of experimental probabilities and from the characteristics of the map (Terwilliger, 1999, 2000). In principle, however, experimentally derived or other prior phase information need not necessarily be included in the maximum-likelihood density-modification procedure. Instead, phase information could be derived from the agreement of the map with expectations alone.

The overall procedure for one cycle of map-likelihood phasing has five basic steps which are based on the methods used for maximum-likelihood density modification (Terwilliger, 2000). First, a starting set of phases is used to calculate a figure-of-merit-weighted electron-density map. This map is important because a comparison of this map with expected electron-density distributions in the unit cell will form the basis for the determination of phase probabilities. Next, the expectations about the electron-density distributions in this map are evaluated. As described in more detail below, this can consist of probability distributions for electron density in the protein and solvent regions along with probability estimates of whether each point in the map is within the protein or solvent region, for example. These probability distributions are crucial for defining the prior expectations about the electron-density map and therefore the log-likelihood of the map. Third, the log-likelihood of this map and the first and second derivatives of this log-likelihood with respect to electron density at each point in the map are calculated. These derivatives will be used to predict how the log-likelihood of the map will change as the electron density in the map is changed. Fourth, using the chain rule and an FFT-based algorithm, the first and second derivatives of the log-likelihood of the map with respect to structure factors are calculated. Fifth, for each reflection k the variation of the log-likelihood of the map with the phase (or phase and amplitude) of the reflection is estimated from these derivatives. This is the key step in map-likelihood phasing. Through the use of the derivatives of the log-likelihood of the map with respect to the structure factor k, map-likelihood phasing allows relative probabilities to be assigned for each possible value of the phase of reflection k. These phase probabilities are used to estimate a new weighted mean estimate of the structure factor k.

In this calculation of the phase-probability distribution for reflection k, ordinarily the measured amplitude is kept fixed and the allowed phases for this reflection are sampled at regular intervals (typically increments of $5°$ for acentric reflections). The log-likelihood of the map is approximated in terms of a Taylor's series based on the derivatives with respect to structure factors as described previously (Terwilliger, 2000), with the addition of a cross-term in the Taylor's series as suggested by Cowtan (2000). To the extent that this approximation is accurate (that is, that higher order terms do not contribute substantially), this phase-probability calculation estimates how the log-likelihood of the map will vary with the phase of reflection k without regard to the value of the phase

that was used to calculate the original electron-density map. Once all five steps in map-likelihood phasing are carried out, it is possible to calculate a new figure-of-merit-weighted electron-density map using the newly estimated phase-probability distributions. These phases can then be used to initiate a new cycle of map-likelihood phasing. As the phases are modified in this fashion, it is useful to update the analysis of the probability estimates for whether each point in the map is in the protein or solvent region and any other analyses based on the map.

The effect of each cycle in this procedure is to obtain a probability distribution for each phase independently of all the others, based on the agreement of the electron-density map with expectations. In the phase-probability calculations, all possible values of the phases are considered without any preference for the values used in the previous cycle.

The iteration of phasing and analysis of the map is ordinarily repeated until phase changes are minimal. As described below, however, in some cases where there is relatively little phase information available from the map-likelihood function it is useful to end the iterative process before complete convergence. Also, in some cases iterations of this procedure lead to some oscillations in which the changes in the weighted mean estimate of structure factor k are strongly anticorrelated from one cycle to the next. In such cases a damping factor (typically 0.5) can be applied to the changes from one cycle to the next to reduce the oscillations.

Map-likelihood phasing is related to the methods of Béran & Szöke (1995) and to the application of NCS phase refinement starting from a solvent mask (Braik *et al.*, 1994; Kleywegt & Read, 1997; van der Plas & Millane, 2000; Wang *et al.*, 1998) in which crystallographic phases are obtained by matching the electron density in a part of the unit cell to a target value. The method of Béran & Szöke (1995), which employs simulated annealing to find a set of phases consistent with constraints on electron density, was shown to be capable of *ab initio* phase determination using a solvent mask. Similarly, high-order non-crystallographic symmetry has been shown to be sufficient to determine crystallographic phases starting just from a solvent mask (Braik *et al.*, 1994; Kleywegt & Read, 1997; van der Plas & Millane, 2000; Wang *et al.*, 1998). The maximum-likelihood approach described here and in Terwilliger (2000) differs from the methods of Béran & Szöke (1995), van der Plas & Millane (2000) and Wang *et al.* (1998) in that probabilistic descriptions of the expected electron density are used, allowing a calculation of phase-probability distributions, rather than searching for a set of phases that is consistent with constraints.

The phase information coming from the map-likelihood function $\mathrm{LL}^{\mathrm{MAP}}(\{F_{\mathbf{h}}\})$ comes from the agreement of the electron-density map with prior expectations about that map. This agreement depends on the phase of each reflection, in the context of the phases of all other reflections. In the implementation used in maximum-likelihood density modification, the probability (based on the map likelihood) for a particular structure factor that the phase has a value φ is given by the relative likelihood of the map obtained with this value of the phase. For example, a simple map-likelihood function might be based on defined regions containing the macromolecule and containing disordered solvent. A value of the phase for a particular reflection k that leads to a map with a relatively flat solvent is more likely to be correct than a phase that does not.

In a more general case, a map-likelihood function can be defined that describes solvent and 'protein' regions of the electron-density map and probability distributions for electron density in each such region. The probability of a particular phase for a particular reflection can then be estimated from how well the resulting map matches these expected characteristics. The concept can also be extended further to include non-crystallographic symmetry. A map-likelihood function could be constructed that reflects the extent to which symmetry-related density in the map is indeed similar, for example.

A formulation of the map log-likelihood function $\mathrm{LL}^{\mathrm{MAP}}(\{F_{\mathbf{h}}\})$ that follows this approach (Terwilliger, 2000) can be written as the integral over the map of a local log-likelihood of electron density, $\mathrm{LL}[\rho(\mathbf{x}, \{F_{\mathbf{h}}\})]$,

$$\mathrm{LL}^{\mathrm{MAP}}(\{F_{\mathbf{h}}\}) \simeq \frac{N_{\mathrm{REF}}}{V} \int_{V} \mathrm{LL}[\rho(\mathbf{x}, \{F_{\mathbf{h}}\})] \, \mathrm{d}^{3}\mathbf{x}. \qquad (2)$$

where this local log-likelihood of electron density describes the plausibility of the map at each point.

The local log-likelihood function, in turn, can be expressed in terms of whether the point is in the solvent or protein regions and the expected electron-density distributions in each case. As it is often uncertain whether a particular point $\mathbf{x}$ is in the protein or solvent region, it is useful to write the local map-likelihood function as the sum of conditional probabilities dependent on which environment the point is located in,

$$\mathrm{LL}[\rho(\mathbf{x}, \{F_{\mathbf{h}}\})] = \ln\{p[\rho(\mathbf{x})|\mathrm{PROT}]p_{\mathrm{PROT}}(\mathbf{x}) \\ + p[\rho(\mathbf{x})|\mathrm{SOLV}]p_{\mathrm{SOLV}}(\mathbf{x})\}, \qquad (3)$$

where $p_{\mathrm{PROT}}(\mathbf{x})$ is the probability that $\mathbf{x}$ is in the protein region, $p[\rho(\mathbf{x})|\mathrm{PROT}]$ is the conditional probability for $\rho(\mathbf{x})$ given that $\mathbf{x}$ is in the protein region and $p_{\mathrm{SOLV}}(\mathbf{x})$ and $p[\rho(\mathbf{x})|\mathrm{SOLV}]$ are the corresponding quantities for the solvent region. The probability that $\mathbf{x}$ is in the protein or solvent regions can estimated by a modification of the methods of Wang (1985) and Leslie (1987) as described earlier (Terwilliger, 1999) or by other probability-based methods (Roversi *et al.*, 2000).

The probability distributions for electron density given that a point is in the protein or solvent regions are central to map-based phasing. They define the expectations about electron density in the map. These expectations about electron-density distributions in the map are not derived from 'perfect' maps, but rather from the current electron-density map. There are several reasons for doing this (Terwilliger, 2000). The key reason is that it is unreasonable to expect any value of the phase for a particular reflection to lead to a map matching expectations of a perfect map because the map has large errors from all the other reflections. In particular, the correct value of the phase for reflection k can only be expected to

reduce the variation in the solvent region only slightly, not to make it perfectly flat. The amount by which the solvent can be expected to be flattened by adjusting just one reflection is dependent on the overall noise in the map. In effect, the expectations about the electron-density map include not just the features of a perfect map, but also effects of the errors in all of the structure factors other than the one under consideration. Consequently, for a starting phase set with large phase errors, the target probability distribution of electron density in the solvent region is very broad, while for a starting phase set that is very accurate this distribution can be very narrow.

Because the targeted features of the electron-density map are only weakly defined for poor starting phase sets but are more precisely defined for accurate ones, the phase information coming from the map-likelihood function becomes stronger as the phases improve. In essence, the more accurate the starting phases and the less noise in the map, the more precisely the phase of a particular reflection can be expected to lead to a map that matches the characteristics of a perfect map, and the more precisely the values of each phase can be determined.

4. Bias and map-likelihood phases

Somewhat paradoxically, although the quality of the starting phase set is an important factor in determining the phase information that comes from the map, the phase probability for a reflection obtained from map-likelihood phasing is completely unbiased with respect to the prior probabilities for that phase. On the other hand, the map-likelihood phase probability for a reflection can be biased by a model used to calculate all starting phases.

To see how the map-likelihood phase for a reflection can be unbiased with respect to prior probabilities for that phase, consider using map-likelihood phasing to obtain a probability distribution for the phase of reflection k. In order to make the situation clear, the procedure described will be a little simpler than the one used in practice. First, calculate an electron-density map using all reflections other than k. This map clearly has no bias towards the prior value for reflection k, as reflection k was not even used to obtain the map. Now examine all possible phases of the reflection k in question. For each phase, add to the map the electron density that would result from reflection k with this phase. Then compare the characteristics of the resulting electron-density map with the ones that we expect, given the location of solvent and macromolecule and given the expected distributions of electron density in solvent and protein regions. Some values of the phase of reflection k will generally lead to more plausible maps than others. This defines our probability distribution for the phase of reflection k and the process has made no use whatsoever of any prior information about this reflection. Consequently, the resulting phases are completely unbiased with respect to any prior information about reflection k. In practice, this cross-validation procedure is carried out with all the reflections at once employing an approximation and an FFT-based method (Terwilliger, 2000). The resulting phase-

probability distributions are essentially the same as the ones described above, however.

Each individual phase-probability distribution obtained with map-likelihood phasing is independent of the prior phase-probability distribution for that reflection. Nevertheless, there are kinds of bias that can affect map-likelihood phasing. If the set of phases used to initiate map-likelihood phasing has been adjusted as a whole in a way that leads to a relatively flat solvent region, for example, then the first few cycles of map-likelihood phasing will tend to find these starting phases to be probable ones (because they lead to a flat solvent when combined with all the other starting phases) even if these starting phases are incorrect. This situation can occur for example if a model has been used to calculate the starting phases, as the solvent region will tend to be relatively flat even if the model is not entirely correct. It can also occur if the phases have been refined in order to flatten the solvent region. Fortunately, as described below, this type of model bias is generally removed by iterative application of map-likelihood phasing.

As described above, other approaches to using expectations about electron-density distributions in a map for determining crystallographic phases without including phase-probability distributions from other sources have been demonstrated (Wang *et al.*, 1998; van der Plas & Millane, 2000; Béran & Szöke, 1995). Each of these approaches begins with no prior phase information and is designed to result in an *ab initio* phase determination. These approaches could be modified to begin with a starting phase set as described here for map-likelihood phasing; however, the probability-based approach described here is more general and can include a variety of expectations about the map. Additionally, map-likelihood phasing leads to phase-probability distributions rather than phases consistent with expectations, so that optimally weighted maps can be calculated.

5. Prime-and-switch phasing to remove model bias

Model bias is a very serious problem in macromolecular crystallography (Adams *et al.*, 1999; Hodel *et al.*, 1992; Kleywegt, 2000). A bias in phases that leads to electron-density patterns that are incorrect, yet look like features of a macromolecule, is very difficult to detect. Such a bias is much more serious than an equivalent amount of noise in a map that is distributed in a random fashion in the unit cell. Bias of this kind commonly occurs when crystallographic phases are calculated based on a model that contains atoms that are incorrectly placed. Maps that are based on these phases tend to show peaks at the positions of these atoms even if the correct electron density would not.

Model bias in an electron-density map does not necessarily mean that the phases are very inaccurate. Relatively accurate phases, calculated from a largely correct model with some atoms in incorrect locations, can still lead to peaks at the coordinates of the incorrectly placed atoms. In this sense, the phases are biased. In the sense that the phases are close to the true phases, the phases are still relatively accurate. As in many

situations, there is an important trade-off between accuracy and bias in the calculation of electron-density maps. In the crystallographic case, this trade-off is fundamentally related to the difference between errors in electron density that are randomly distributed about the unit cell and those that are focused on certain locations in the cell. In many situations, the most accurate map (the one with the minimum expected mean-square error, for example) will be one that is based on all available information. Unfortunately, in the crystallographic situation the errors in such a map can be highly non-random. As mentioned above, high peaks can be obtained at the specific positions of atoms in a model used to calculate phases, even if those atoms are incorrectly placed. Such a map, though accurate, can be highly misleading. A map that is less accurate, but that does not suffer from this bias, could in many cases be far more informative.

Many methods for reducing model bias in electron density maps have been developed. One of the most widely used approaches is the σ_A method of Read (1986), in which the weighting and amplitudes of structure factors (but not the phases) are optimized for minimizing effects of model bias. As the phases remain based on the model, σ_A weighting retains some model bias (Hodel et al., 1992). Another important method is the use of omit maps, in which all atoms in a region of the unit cell in the model are removed before using the model to calculate phases. This method reduces model bias, but leads to electron-density maps that are intrinsically much noisier than those calculated with all atoms present. Omit maps can still contain some model bias despite the omission of atoms in a region of space, as refinement can adjust the parameters describing all the other atoms in such a way as to leave a 'memory' of the coordinates of the omitted atoms (Hodel et al., 1992). This memory in omit maps corresponds to the model bias described above that can occur in the first few cycles of map-likelihood phasing. The residual bias in omit maps can be reduced by simulated annealing of the partial model (Hodel et al., 1992), if the resolution of the data and the accuracy of the starting model allows atomic refinement. In general terms, this corresponds to the iterative application of map-likelihood phasing to remove residual bias. Maximum-likelihood refinement of the model structure can also be used to reduce model bias even in cases where σ_A-weighted electron-density maps are not interpretable (Adams et al., 1999).

The technique of prime-and-switch phasing takes advantage of the lack of bias in map-likelihood phasing and the strong dependence of the accuracy of map-likelihood phases on the quality of the phases used to initiate the process discussed above. In this technique, all available phase information, including any coming from a model, is used to initiate map-likelihood phasing. The model phases are then set aside and not used further. As discussed above, model-based phases can be relatively accurate and biased at the same time. Owing to their accuracy, they can be useful in initiating map-likelihood phasing. Owing to their bias, setting them aside during map-likelihood phasing can reduce the bias in the final phases. As prime-and-switch phasing does not use all the phase infor-

mation available, the final phases could be less accurate than a set that could be obtained using all this information. As shown below, in most cases any loss of accuracy is compensated for by a corresponding decrease in bias.

Map-likelihood phasing has the potential for producing electron-density maps that have little or no bias, as the phase probabilities for each reflection are independent of the prior phases for that reflection. However, as described above, it is possible for map-likelihood phasing to be biased by a starting phase set that has a systematic bias, for example by a starting set of incorrect phases that has a relatively flat solvent region. The iteration of cycles of map-likelihood phasing is a useful tool in reducing or eliminating this bias. The reason for expecting that an iterative application of map-likelihood phasing would remove the bias present in a single cycle is that the bias for an individual reflection comes from the set of starting phases as a whole. Once many of the phases in the set are substantially changed, the bias might be greatly reduced.

6. Convergence of map-likelihood phasing

There are two general cases that could arise in carrying out iterative cycles of map-likelihood phasing. If the solvent content or non-crystallographic symmetry are high, then the phases are likely to be well determined and simple iterative map-likelihood phasing would be effective. If the solvent content is low and non-crystallographic symmetry is lacking, however, the phases might not be entirely determined by the map-likelihood function. In this case, it might be necessary to trade off a small bias towards the starting phase set in order to obtain a well defined set of phases. One way to carry out such a trade-off is simply to end the iteration of map-likelihood phasing before complete convergence. This is generally the preferable alternative in practice because, as shown below, map-likelihood phases often change very rapidly during early cycles then only very gradually after that.

An alternate procedure is to introduce a small weighting towards the model-based prior phases. This introduction of some model-based phase information has several effects. One is to reintroduce some bias into the final phases. A second is to stabilize the phasing process. A third is (at least potentially) to increase the overall accuracy of the phases. The degree of bias towards the starting phase set in map-likelihood phasing can be adjusted using a weight on the prior phase probabilities. In cases where the phase information in the map is insufficient to fully define the phases (such as substantially less than 50% solvent content with no non-crystallographic symmetry), it is sometimes useful to trade off a small amount of bias in order to increase the stability of the iterative phasing process. This can typically be accomplished with a weighting of a few percent on the prior phase-probability distribution.

7. The bias ratio α_B

A useful measure of the degree of bias towards model-based or other prior phases used to initiate prime-and-switch

phasing or used as a source of combined phase information is the bias ratio α_B, defined as

$$\alpha_B = \frac{\langle\cos(\varphi_{\mathrm{PRIOR}} - \varphi_{\mathrm{MAP}})\rangle}{\langle m_{\mathrm{PRIOR}}\rangle\langle m_{\mathrm{MAP}}\rangle}, \qquad (4)$$

where φ_{PRIOR} is the centroid phase based on prior phase information, φ_{MAP} is that based on the map-likelihood function, $\langle m_{\mathrm{PRIOR}}\rangle$ and $\langle m_{\mathrm{MAP}}\rangle$ are estimates of $\langle\cos(\varphi_{\mathrm{PRIOR}} - \varphi_{\mathrm{TRUE}})\rangle$ and $\langle\cos(_{\mathrm{MAP}} - \varphi_{\mathrm{TRUE}})\rangle$, the mean cosines of the phase differences between the true phase φ_{TRUE} and φ_{PRIOR} or φ_{MAP}, respectively, and the averages are taken over all reflections.

If the estimates of $\langle m_{\mathrm{PRIOR}}\rangle$ and $\langle m_{\mathrm{MAP}}\rangle$ are reasonably accurate, the bias ratio α_B can be a useful measure of the extent of correlation between the prior phases φ_{PRIOR} and the map-likelihood phases φ_{MAP}, compared with the correlation expected for two independent sources of phasing. The numerator $\langle\cos(\varphi_{\mathrm{PRIOR}} - \varphi_{\mathrm{MAP}})\rangle$ is a measure of the actual correlation between prior and map-likelihood phases. If these sources of phase information are independent, then the errors in each are independent and we can write that

$$\langle\cos(\varphi_{\mathrm{PRIOR}} - \varphi_{\mathrm{MAP}})\rangle \simeq \langle\cos(\varphi_{\mathrm{PRIOR}} - \varphi_{\mathrm{TRUE}})\rangle$$
$$\times \langle\cos(\varphi_{\mathrm{MAP}} - \varphi_{\mathrm{TRUE}})\rangle \qquad (5)$$

or

$$\langle\cos(\varphi_{\mathrm{PRIOR}} - \varphi_{\mathrm{MAP}})\rangle \simeq \langle m_{\mathrm{PRIOR}}\rangle\langle m_{\mathrm{MAP}}\rangle, \qquad (6)$$

leading to a bias ratio α_B of about unity. In contrast, if the sources of phase information are not independent, such as might occur if the bias in the model-based phases was not completely removed by iterative map-likelihood phasing, then the correlation between prior and map-likelihood phases will typically be greater than when they are independent,

$$\langle\cos(\varphi_{\mathrm{PRIOR}} - \varphi_{\mathrm{MAP}})\rangle > \langle\cos(\varphi_{\mathrm{PRIOR}} - \varphi_{\mathrm{TRUE}})\rangle$$
$$\times \langle\cos(\varphi_{\mathrm{MAP}} - \varphi_{\mathrm{TRUE}})\rangle, \qquad (7)$$

and the bias ratio α_B will generally be greater than unity.

The utility of the bias ratio is dependent on having reasonable estimates of the figures of merit of phasing $\langle m_{\mathrm{PRIOR}}\rangle$ and $\langle m_{\mathrm{MAP}}\rangle$ for the prior and map-likelihood phases. If these are overestimated, for example, then the bias ratio will be underestimated. In an extreme case, a bias ratio of substantially less than unity can be used as a diagnostic for overestimated values of the figures of merit of one or both of these sources of phasing. The bias ratio can potentially be used in a third approach for handling situations where the map-likelihood phase information is insufficient in itself to fully define the phases. In this approach, iterations of map-likelihood phasing are carried out until the bias ratio reaches a value of approximately unity, indicating that much of the bias from the prior model-based phases has been removed.

8. Examples of map-likelihood phasing

8.1. Separation of experimental and map-likelihood phase information

Fig. 1 illustrates how the phase information in maximum-likelihood density modification can be separated into experimentally derived phase information and map-likelihood phase information. Fig. 1(a) shows an experimental electron-density map based on MAD phasing of initiation factor 5a (Peat *et al.*, 1998). Fig. 1(b) shows an electron-density map calculated from the map-likelihood phase probabilities obtained on the first cycle of maximum-likelihood density modification using the experimental map in Fig. 1(a) as a starting point. The crystals of initiation factor 5a contain about 60% solvent, so the phasing information that can be obtained from the map likelihood is very substantial. The map-likelihood phased map in Fig. 1(b) is clearly of equivalent or higher quality than the experimental map in Fig. 1(a). This is quite remarkable when it is recognized that the phase probabilities for the map in Fig. 1(b) are obtained simply by matching calculated and expected electron-density distributions in the solvent and protein regions. Fig. 1(c) illustrates that the accuracy of the starting phase set used in map-likelihood phasing has a substantial effect on the final phase-probability distributions. In this panel, the starting phases were those obtained after maximum-likelihood density modification with the program *RESOLVE* (Terwilliger, 2000) was applied to the data used in Fig. 1(a). This electron-density map is of even higher quality than those in Fig. 1(a) or Fig. 1(b).

### 8.2. Convergence and phase improvement in map-likelihood phasing

In order to evaluate the range of applicability of map-likelihood phasing and the utility of iterative phase improvement with this technique, several tests were carried out with model data, where the quality of phasing could readily be assessed. Figs. 2 and 3 illustrate the convergence properties of map-likelihood phasing as a function of percentage of the asymmetric unit that is occupied by disordered solvent. Model data sets were constructed based on the refined structure of dehalogenase enzyme from *Rhodococcus* (Newman *et al.*, 1999) to a resolution of 3 Å as described in Terwilliger (2000). To simulate varying amounts of solvent, varying numbers of water molecules and C-terminal residues were left out of the phase calculations. This led to models with disordered solvent content ranging from 31% (as in the actual crystals) to 73%. Starting phase sets with simulated errors were constructed and used along with the model amplitudes in map-likelihood phasing. In these simulations, a mask defining the solvent and protein regions was calculated from the atomic coordinates in the model, defining all points within 2.5 Å of an atom as being within the protein region. In each test, 20 cycles of phase calculation followed by figure-of-merit weighted map calculation were carried out. For each cycle, the mean true figure of merit, given by the cosine of the phase error $\langle\cos\Delta\varphi\rangle$ is plotted.

Fig. 2(*a*) shows the effect of the percentage of the cell occupied by the macromolecule and by 'solvent' (actually simply absence of atoms in these simulations) on the phases obtained from map-likelihood phasing starting with very poor initial phases. The starting mean true figure of merit in each case was 0.32 and the data extended to a resolution of 3 Å. For simulations with about 50% solvent or greater, each cycle of map-likelihood phasing resulted in phases that were at least as accurate as those in the previous cycle, with convergence essentially complete within 20 cycles. For those with 39% solvent, the phases became slightly worse with map-likelihood phasing compared with the starting phases; for the case with 31% solvent they were considerably worse. Fig. 2(*b*) illustrates the effect of high-resolution and low-resolution cutoffs on the

quality of the phasing for the simulation with 53% solvent shown in Fig. 2(*a*). When all the data from 2.8 to 20 Å are included, the final mean true figure of merit was 0.56. When data from only 2.8 to 6 Å are included, the resulting true figure of merit decreases to only 0.44 and when data from only 2.8 to 5 Å are included, to only 0.34. Conversely, when data from only 5 to 20 Å are included, the resulting true figure of merit is only 0.28 and, as high-resolution data to 2.8 Å are included, this increases to 0.56.

Fig. 3 expands on the simulation shown in Fig. 2, illustrating the stability and convergence of phasing beginning with phases with varying errors, for solvent contents of 31, 47 and 73%. In the case of 31% solvent content, for all starting phase sets the quality of phases generally decreased with each cycle of map-likelihood phasing, although when the starting true figure of merit was about 0.6 or greater, the overall phasing process was relatively stable. In contrast, for the simulation with 47% solvent the quality of phases increased slightly with each cycle. Starting from phase sets with a true figure of merit of about 0.45 or greater, all of the test simulations converged to phase sets with similar true figures of merit of about 0.6. For 73% solvent, the quality of the phases reached the same very high true figure of merit of about 0.8, regardless of the true figure of merit of the starting set of phases in the range 0.3–0.8.

Fig. 4 illustrates the effect of errors in the definition of solvent and protein regions on phasing. The simulations in this figure were carried out in the same way as those in Fig. 2, except that the mask used was based on a model that was missing about 10% of the atoms, so that about 10% of the 'protein' region was classified as 'solvent'. The quality of the


(*a*)

(*b*)   (*c*)
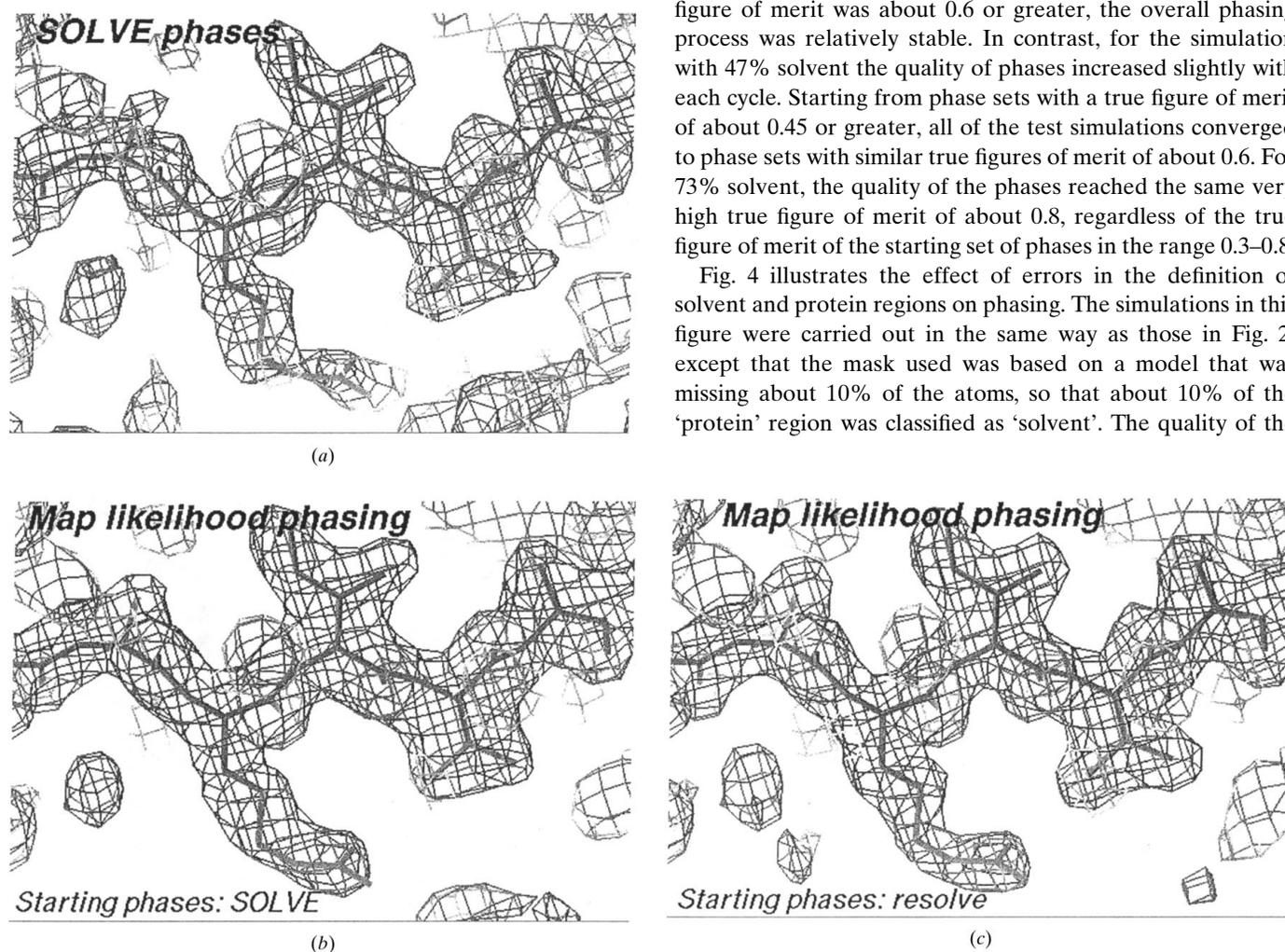
Figure 1
Map-likelihood phasing of initiation factor 5a. (*a*) Experimental (MAD) map. (*b*) Map-likelihood phased map, using experimental phases as a starting phase set. (*c*) Map-likelihood phased map, using maximum-likelihood density-modified phases as starting phase set. Experimental (MAD) phases for initiation factor 5a (Peat *et al.*, 1998) had an overall figure of merit of 0.61 to 2.2 Å. The refined model of initiation factor 5a (PDB entry 1bkb; Berman *et al.*, 2000) is overlaid on the experimental map. The map is contoured at 1.5σ. In (*b*), one cycle of map-likelihood phasing was carried out and phase probabilities were calculated from the map likelihood only. The figure of merit of map-likelihood phases was 0.37. The calculation of the solvent mask and the map-likelihood phasing was carried out as described for maximum-likelihood density modification (Terwilliger, 2000), except that experimental phases were not used in the phase-probability calculation. Experimental phases were used in determining the mask, as in maximum-likelihood density modification. In (*c*), a full set of five mask cycles of maximum-likelihood density modification, each with ten minor cycles of phase improvement, were carried out. These phases were then used to calculate a mask and to initiate one cycle of map-likelihood phasing as in (*b*). The figure of merit of map-likelihood phasing was 0.50.
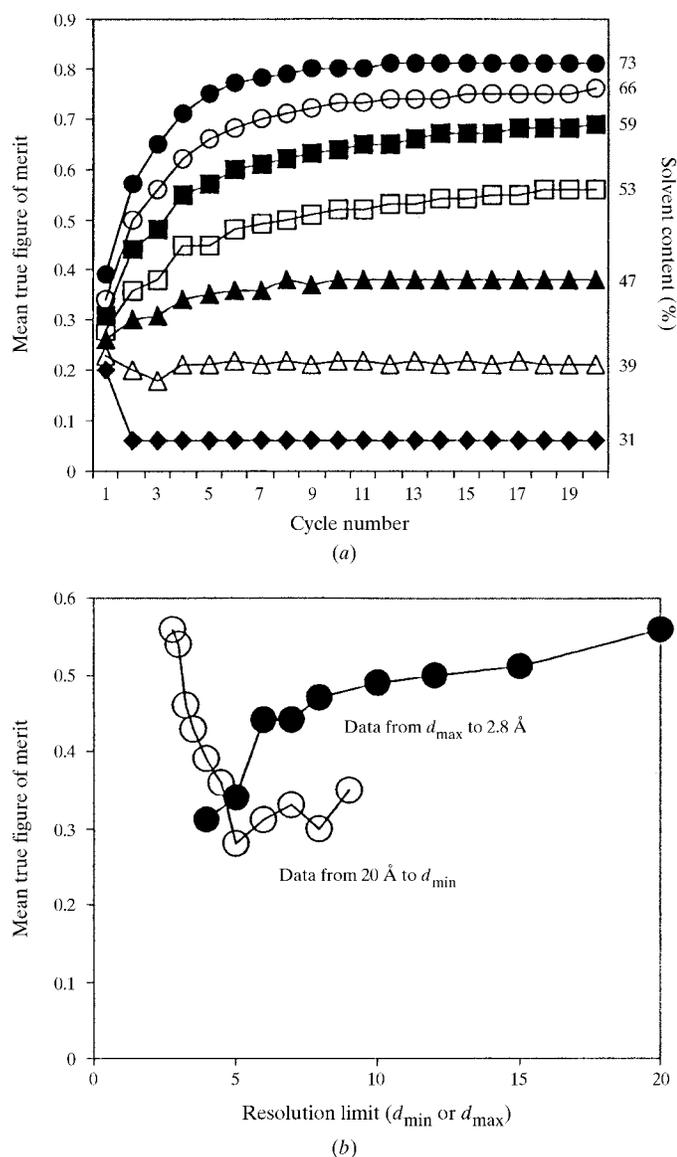
Figure 2
Effect of solvent content on map-likelihood phasing. Mean true figure of merit ($\langle\cos\Delta\varphi\rangle$) is plotted as a function of cycle number for simulations phased based on map likelihood with model cases having solvent contents of 31 (closed diamonds), 39 (open triangles), 47 (closed triangles), 53 (open squares), 59 (closed squares), 66 (open circles) and 73% (closed circles). Model phases were calculated from a model based on the refined structure of dehalogenase from *Rhodococcus* (Newman *et al.*, 1999; PDB entry 1bn7), except that varying numbers of atoms were omitted from the calculation to simulate varying solvent content. The space group is $P2_12_12$, with unit-cell parameters 93.8, 79.8 and 43.1 Å, and the resolution limits used were from 20 to 2.8 Å. The protein atoms in the simulations in Fig. 2(*a*) occupy from 27 to 69% of the unit cell, based on an algorithm in which all points within 2.5 Å of an atom in the model are considered occupied by protein. In each simulation, a starting set of phases was generated by adding Gaussian errors to the model phases so as to achieve a mean true figure of merit ($\langle\cos\Delta\varphi\rangle$) of 0.32. This starting set of phases, along with model amplitudes, was used to initiate map-likelihood phasing. Solvent masks were based on the atomic models in this simulation. In (*a*), the mean true figure of merit ($\langle\cos\Delta\varphi\rangle$) is plotted as a function of cycle number for simulations with solvent content of 31 (closed diamonds), 43 (open triangles), 47 (closed triangles), 53 (open squares), 59 (closed squares), 66 (open circles) and 73% (closed circles). In (*b*), the simulation with 53% solvent was repeated using varying high-resolution cutoffs (open circles) or low-resolution cutoff (closed circles) and the final mean true figure of merit after 20 cycles is plotted.
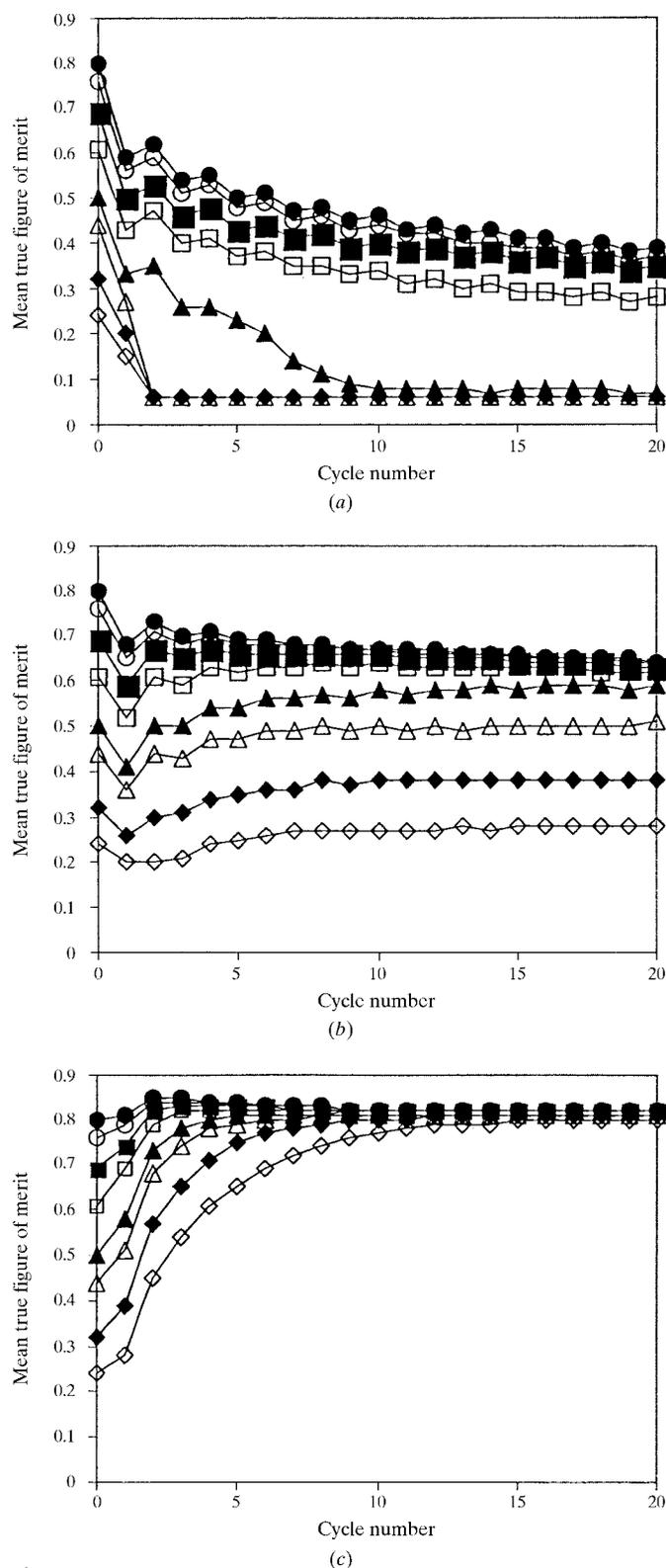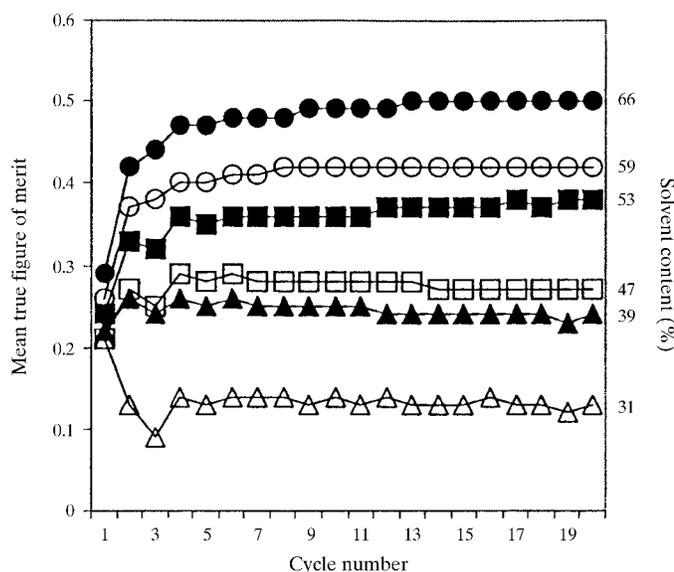

Figure 3
Effect of starting phase accuracy on map-likelihood phasing. The mean true figure of merit ($\langle\cos\Delta\varphi\rangle$) is plotted as a function of cycle number for simulations with starting mean true figure of merit of 0.24 (open diamonds), 0.32 (closed diamonds), 0.44 (open triangles), 0.50 (closed triangles), 0.61 (open squares), 0.69 (closed squares), 0.76 (open circles) and 0.80 (closed circles) and with solvent contents of (*a*) 31, (*b*) 47 and (*c*) 73%. All simulations were carried out as in Fig. 2 except for the starting phase sets.

map-likelihood phases obtained was less than that obtained with the correct mask; even so, in the cases with about 50% or greater solvent content the phase quality with map-likelihood phasing improves over the starting phase set.

8.3. *Ab initio* phase determination with map-likelihood phasing

Fig. 3(c) showed that in cases with very high solvent content (73%), map-likelihood phasing yielded very substantial phase improvements and converged to essentially the same point regardless of the starting phase set used. Fig. 5 explores this further by illustrating the phase quality obtained by map-likelihood phasing as a function of solvent content, beginning with zero phase information (a blank map), but with a perfect solvent mask calculated from the atomic model. Fig. 5 shows that in cases with 66 and 73% solvent, map-likelihood phasing is sufficient in itself to determine crystallographic phases with high accuracy. In the model cases with 59 and 53% solvent, modest phase quality was obtained. These results are similar to those obtained by Béran & Szöke (1995) using a very different approach (simulated annealing) to find phase sets for model data that are compatible with defined solvent and protein regions.

It should be noted that although the map-likelihood approach was successful in *ab initio* phasing when using model data, tests carried out so far with experimental data have not resulted in substantial phase improvement. Presumably, this is because of complications from measurement errors and from the smaller differentiation between solvent and protein

regions in real crystals compared with the model data sets examined here.

8.4. Reduction of model bias with prime-and-switch phasing

A very important feature of map-likelihood phasing is the potential for reducing or eliminating model bias in electron-density map calculations through the technique of prime-and-switch phasing. Test cases with model data were set up in order to examine how thoroughly model bias could be removed using prime-and-switch phasing and how this depended on the solvent content of the crystal. Additionally, the effect of including some prior phase information on bias and map quality for various solvent contents was examined.

Model data sets were constructed using the refined structure of dehalogenase enzyme from *Rhodococcus* (Newman *et al.*, 1999) and leaving out varying numbers of water molecules and atoms from the C-terminus to simulate varying amounts of solvent content as in Fig. 2. These models were considered the 'correct' structures in the tests. Then, from each correct model, a 'molecular replacement' model was constructed by varying the coordinates of atoms in the correct model by an r.m.s.d. of 1.4 Å, using a function that varied sinusoidally in space so that the connectivity of the molecule remained intact. Next, all the atoms in the molecular-replacement model that were placed incorrectly were identified by noting the value of
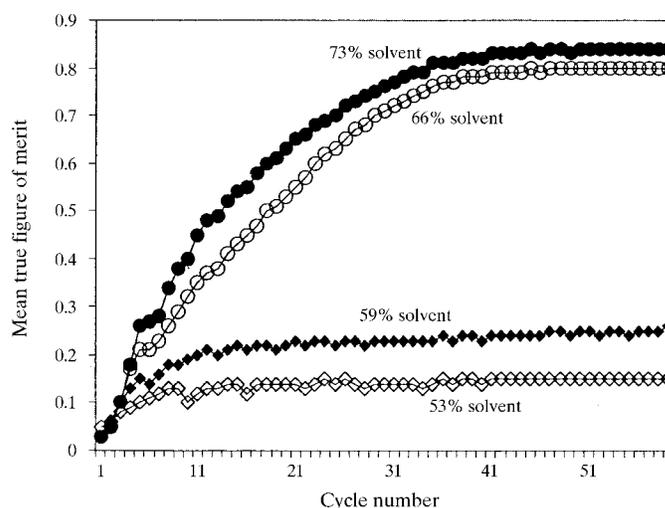
Figure 5
Effect of solvent content on map-likelihood phasing with no prior phase information. The mean true figure of merit ($\langle \cos \Delta \varphi \rangle$) is plotted as a function of cycle number for simulations with solvent contents of 53, 59, 66 and 73%. Simulations were started with zero phase information (a flat map). Starting probability distributions for electron density in the protein and solvent regions were taken from the first cycle of the simulation shown in Fig. 2(a) (for the simulation with 73% solvent). For all subsequent cycles, probability distributions were estimated by cross-validation as follows. The general procedure was to obtain an 'omit' map in which each point was derived from a density-modification cycle in which that point had not been included in the solvent mask. An 'omit' region was uniformly defined as 'protein', regardless of its actual location. Three cycles of density modification were carried out and the 'omit' region was saved. Omit regions covering the entire asymmetric unit were calculated and combined to make a complete 'omit' map of the asymmetric unit. This map was used to estimate probability distributions for density in the protein and solvent region for the next overall cycle.

Figure 4
Effect of solvent content on map-likelihood phasing with a partially incorrect mask. Mean true figure of merit ($\langle \cos \Delta \varphi \rangle$) is plotted as a function of cycle number for simulations with solvent contents of 31 (open triangles), 39 (closed triangles), 47 (open squares), 53 (closed squares), 59 (open circles) and 66% (closed circles). In each case the solvent mask used was based on the correct atomic model, except that the last approximately 10% of atoms in the structure were omitted in order to create an incorrect mask. All simulations were carried out as in Fig. 2 except for the starting phase sets.

the electron density in a 'perfect' map calculated with structure factors based on the correct model. All those atoms in the molecular-replacement model that were in density from -0.5σ to 0.5σ were considered to be incorrectly placed. From 20 to 30% of the atoms in the molecular-replacement models were incorrectly placed according to this criterion. The mean density at coordinates of these incorrectly placed atoms in the perfect electron-density maps for the simulations with various solvent content ranged from 0.03σ to 0.06σ and the mean density at the coordinates of atoms in the correct model in the perfect electron-density map ranged from 1.7σ to 2.9σ, with the higher values corresponding to higher solvent contents (in which most of the cell is solvent, so the ratio of peak height to the r.m.s. density of the map is higher even with perfect data).

In the tests of model bias, the overall accuracy of electron-density maps in these tests was assessed from the normalized mean value of electron density at the coordinates of atoms in the correct model. The model bias was assessed from the normalized mean value of electron density at coordinates of incorrectly placed atoms in the molecular-replacement model used in phasing. Fig. 6(a) shows the overall accuracy and model bias obtained by prime-and-switch phasing (with no prior phase information included in probability calculations) as a function of the solvent content in the model crystals. For comparison, the accuracy and model bias for σ_A-weighted maps based on the same data are shown. The overall accuracy of both the σ_A-weighted and prime-and-switch phased maps was quite high in all cases, with the prime-and-switch phased maps showing greater accuracy in all cases except at very low solvent content. The σ_A-weighted maps had mean values of electron density at coordinates of atoms in the correct model ranging from 0.9σ (31% solvent) to 1.8σ (73% solvent), while the prime-and-switch phased maps had mean values of electron density at coordinates of atoms in the correct model ranging from 0.9σ (31% solvent) to 2.6σ (73% solvent).

The level of bias was very different in the two methods. The σ_A-weighted maps had mean values of electron density at coordinates of incorrectly placed atoms in the molecular-replacment model ranging from 0.5σ (31% solvent) to 1.1σ (73% solvent). In contrast, the map-likelihood phased maps had values ranging from just 0.01σ (31% solvent) to 0.13σ (73% solvent), only slightly higher than the values of 0.03σ to 0.06σ found for a perfect map. Overall, the fractional bias, the ratio of the mean values of electron density at incorrectly placed to correctly placed atoms, for σ_A-weighted maps was in the range 0.5–0.6 for all values of solvent content (Fig. 6b). The fractional bias using prime-and-switch phasing was in the range 0.03–0.09 for all values of the solvent content, indicating that bias was nearly eliminated in all cases.

Fig. 7 illustrates the relationship between including model-based phase information and the resulting bias in the electron-density map. The overall quality of maps and fractional bias (as in Fig. 6) for map-likelihood phasing with 31, 47 and 73% solvent and including varying amounts of prior phase information, ranging from zero weight on prior phases to equal weighting of prior phases and map-likelihood phases, are shown. For the simulations with solvent content of 31 and

47%, the overall quality of the maps generally increases as expected with inclusion of prior phase information and then slowly decreases, with mean electron density at coordinates of atoms in the perfect model with 31% solvent increasing from 0.89 (zero prior phase information) to 1.09 (10% weight on
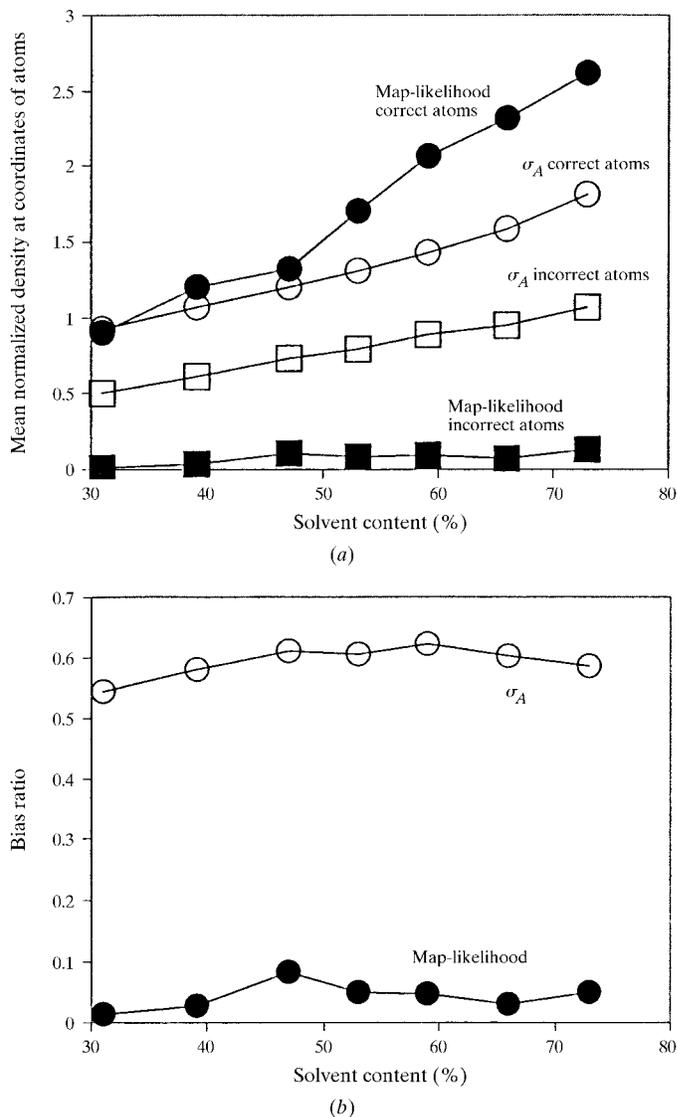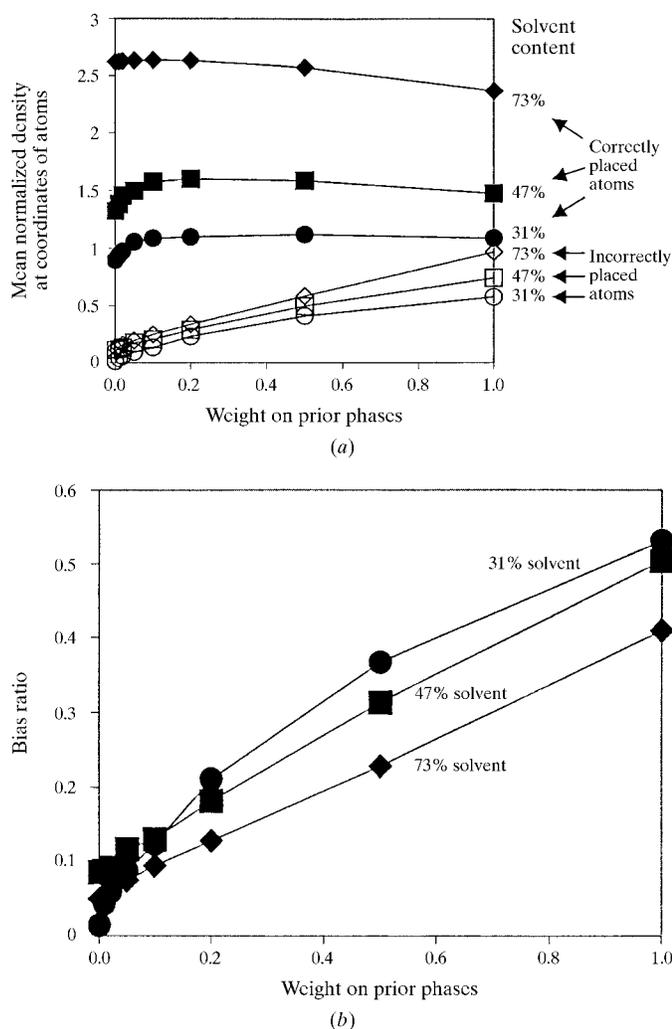

Figure 6
Bias in prime-and-switch phasing. Model data sets with varying solvent contents and 'molecular-replacement' models with atomic coordinates differing from the perfect models by an r.m.s.d. of 1.4 Å were constructed as described in the text. Phases were calculated with σ_A weighting and with prime-and-switch phasing. The prime-and-switch phasing was carried out beginning with the σ_A-weighted phases; ten cycles of solvent-mask identification each with 40 iterations of phasing were carried out. In all cases, essentially complete convergence was achieved within this number of cycles. In (a), the normalized electron density in the σ_A-weighted map (open circles) or prime-and-switch phased map (closed circles) at coordinates of atoms in the perfect model are shown. Additionally, the normalized electron density in the σ_A-weighted map (open squares) or prime-and-switch phased map (closed squares) at coordinates of incorrectly placed atoms in the molecular-replacement model used for phasing are shown. In (b), the ratio of the electron density at incorrectly placed atoms to correctly placed atoms is shown for the σ_A-weighted map (open circles) or prime-and-switch phased map (closed circles).

prior information). When equal weight is placed on the prior information, overall quality decreases slightly, indicating that the prior phase-probability distributions may not be quite optimal. For the simulation with 73% solvent, inclusion of prior phase information had only a small and generally negative effect on the overall accuracy of phasing. This is presumably owing to the very high amount of unbiased phase information in the map-likelihood function in this case of high solvent content.

Fig. 8 illustrates the convergence of the map-likelihood phasing procedure as a function of the solvent content of the unit cell. In an ordinary application of map-likelihood phasing about 50 cycles of iteration would be carried out. In order to examine the convergence properties in more detail, 400 cycles were carried out for each simulation, with weights on the prior
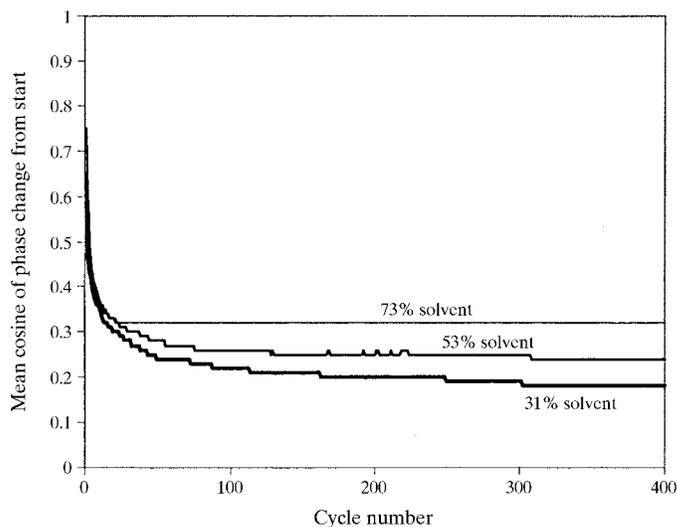

Figure 7
Effect of including prior phase information on map quality and on bias. Iterative map-likelihood phasing was carried out on the model data with 31 (circles), 47 (squares) and 73% (diamonds) solvent as in Fig. 6, except that prior σ_A-based phase probabilities were included with varying weights. The mean electron density at coordinates of atoms in the perfect model (filled symbols) and the mean density at coordinates of incorrectly placed atoms (open symbols) are shown in (a) and the ratio of electron density at incorrectly placed atoms to density at correctly placed atoms is shown in (b) (filled symbols).

phase information ranging from zero to unity. The procedure converges rapidly for the cases with 73% solvent, requiring fewer than 50 cycles for essentially complete convergence. In the cases with 53% and with 31% solvent, convergence was not fully achieved even after 400 cycles. This illustrates cases where one of the simple procedures discussed above for stopping the iterative phasing procedure before full convergence or for the introduction of a limited amount of prior phase information to stabilize the phasing process would be applicable.

8.5. Structure validation

An important application of map-likelihood phasing is likely to be structure validation (Wilson *et al.*, 1998; Kleywegt, 2000). An unbiased method of comparing a model with amplitudes of structure factors that can identify specific places in the structure that are not fully compatible with the data would be of great help in structure validation. The map-likelihood phasing method is well suited to this task as it produces phase probabilities that are essentially unbiased by the starting phase set. Fig. 9 illustrates an example of this. The structure of gene 5 protein has been determined several times, and one of the earlier structures, refined at the moderate resolution of 2.3 Å (PDB entry 2gn5; Berman *et al.*, 2000; Brayer & McPherson, 1983), differed in the loops and consequently in the register of the β-strands from structures determined at the higher resolution of 1.8 Å (PDB entry 1vqb; Skinner *et al.*, 1994) and by NMR (Folkers *et al.*, 1994).

Structure validation of PDB entry 2gn5 was carried out in two steps. The data used consisted of the atomic model 2gn5 and measured structure factors from 20 to 2.6 Å. First, the atomic model 2gn5 was used to calculate model phases and the σ_A approach of Read (1986) was used to estimate phase probability distributions for all of the structure factors. A

Figure 8
Convergence of prime-and-switch phasing as a function of solvent content. The mean cosine of the phase angle difference between the starting model phases and the iterative prime-and-switch phases obtained using the same data as in Fig. 6 is plotted as a function of cycle number for solvent contents of 31, 53 and 73%.

region of the σ_A-weighted map containing the loop at residues 64–67 of gene 5 protein is shown in Figs. 9(a) and 9(b). In Fig. 9(a), the atomic model from PDB entry 2gn5 is overlaid on the map and in Fig. 9(b) the atomic model from the higher resolution model 1vqb is overlaid on the map. Somewhat surprisingly, considering the difference in register between the models, in general this map agrees quite well with both structures. In the region from residues 64–67, neither model fits it perfectly and neither is entirely incompatible with the map. Next, the σ_A-weighted phases were used to initiate map-likelihood phasing and five cycles of solvent-mask identifica-tion, each with ten minor cycles of phase optimization, were carried out. In this map-likelihood phasing process, the σ_A-weighted starting phases were only used to initiate the first cycle of phasing and were not used in phase-probability calculations or in subsequent cycles of phasing. The crystals of gene 5 protein contain about 40% solvent. Figs. 6(c) and 6(d) illustrate the same region shown in Figs. 6(a) and 6(b), this time with the prime-and-switch based phasing. Once again, overall the map agrees relatively well with both structures, but the prime-and-switch based phasing results in a map that is clearly more consistent with the higher resolution structure 1vqb. Figs. 6(c) and 6(d) illustrate, for example, that in the region of residues 64–67, this map shows connectivity that is in excellent agreement with the higher-resolution atomic model 1vqb, even though it is derived from the model 2gn5.

9. Applications of map-likelihood and prime-and-switch phasing

The technique of map-likelihood phasing has potential applications in many situations in X-ray crystal-lography. The critical characteristics of map-likelihood phasing are (i) that it derives phase information from the agreement of features of the electron-density map with expectation and (ii) that it produces phase (or amplitude and phase) probability information that is minimally biased by the starting phase set. The phases it produces are comple-mentary to those obtained by experi-mental (e.g. MIR, MAD) approaches because the source of phase information is completely separate (e.g. solvent flat-ness versus MAD measurements). For the same reason, phases are also complementary to phases calculated from a model or partial model by σ_A-based (Read, 1986) or related approa-ches. Prime-and-switch phasing is a special case of map-likelihood phasing in which an accurate but potentially biased source of prior phase information such as might be obtained from an atomic model is used to initiate map-likelihood phasing but then is not used further in the phasing process.

The characteristics of map-likelihood phasing make it suited for a diverse set of applications, including minimally biased phase calculations from search models in the method of molecular replacement (Rossmann, 1990, 1995),
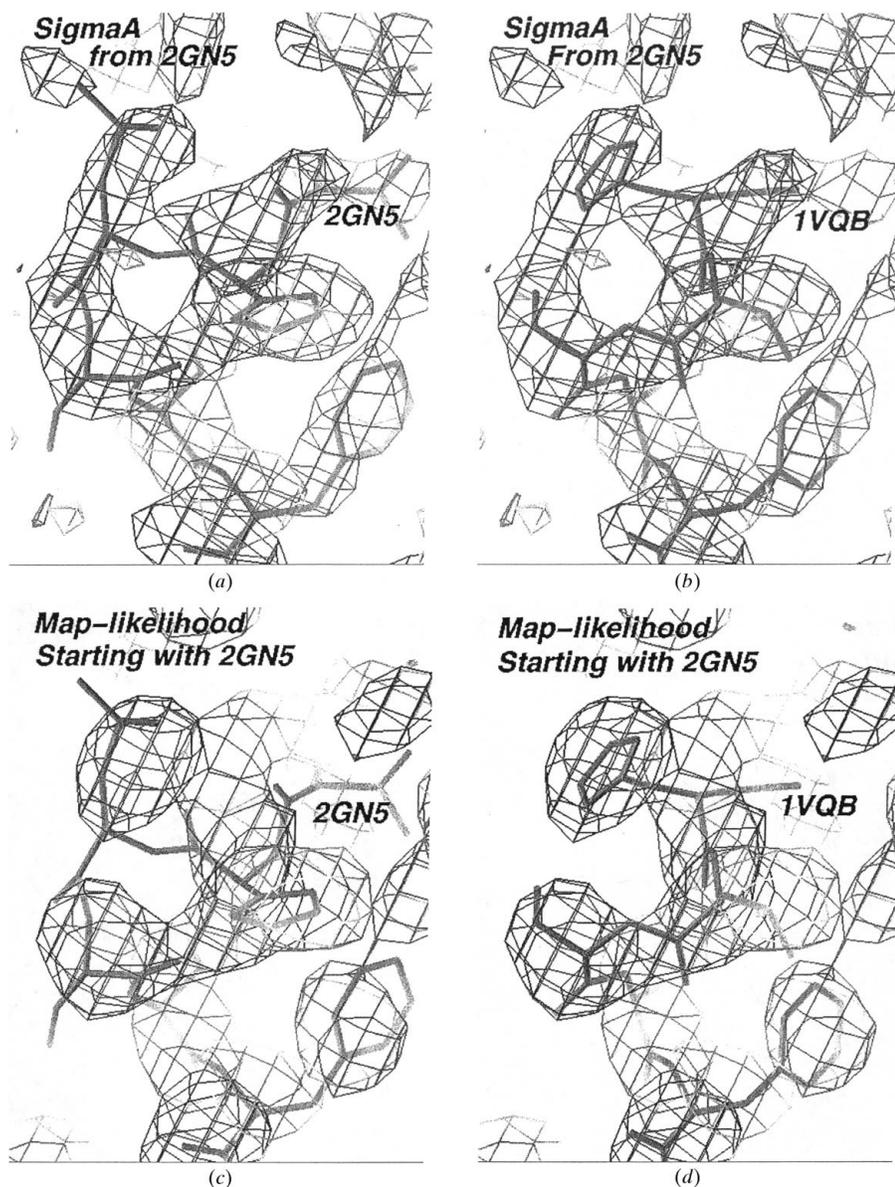
Figure 9
Structure validation with map-likelihood phasing. The atomic model in PDB entry 2gn5 was used with amplitudes of structure factors from 20 to 2.6 Å. Structure-factor amplitudes were from selenomethionine-containing gene 5 protein at λ = 0.9794; they consisted of the FP values output by *SOLVE* (Terwilliger & Berendzen, 1999) when run on the gene 5 protein MAD data (Skinner *et al.*, 1994). (a) and (b) are σ_A-weighted maps. (c) and (d) are from map-likelihood phasing as described in the text. The atomic coordinates of residues 60–70 of PDB entry 2gn5 are overlaid on the map in (a) and (c); those of the same residues from 1vqb are overlaid in (b) and (d). The maps are all contoured at 1σ.

iterative model-building (Perrakis *et al.*, 1999), structure validation (Wilson *et al.*, 1998) and *ab initio* phase determination from solvent masks or non-crystallographic symmetry (Béran & Szöke, 1995; Rossmann, 1995; van der Plas & Millane, 2000; Wang *et al.*, 1998).

The approach is applicable to any situation in which phase probabilities unbiased by a starting phase set are desirable and in which some characteristics of the electron-density map can be anticipated in advance. It is most readily applied to cases where a starting set of phases exists although, as shown above, this is not required.

The accuracy of the phases obtained using map-likelihood phasing can be expected to depend largely on two factors. One is the extent of constraints that are known in advance about the electron-density map. If the structure contains a very large amount of solvent, for example, then much phase information can be obtained because electron density in the solvent region is very highly constrained. The other is the quality of the starting phase information. In an extreme case, if the phases of all reflections with significant intensities except one were known perfectly, then the phase of the final reflection could be determined perfectly because only the perfect phase would lead to a perfectly flat solvent region. In general, the higher the quality of starting phase information, the better defined the resulting probability distributions.

References

Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* D**52**, 30–32.
Adams, P. D., Pannu, N. S., Read, R. J. & Brunger, A. T. (1999). *Acta Cryst.* D**55**, 181–190.
Béran, P. & Szöke, A. (1995). *Acta Cryst.* A**51**, 20–27.
Berman, H. M., Westbrook, J., Feng., Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Braik, K., Otwinowski, Z., Hegde, R., Boisvert, D. D., Joachimiak, A., Horwich, A. L. & Sigler, P. B. (1994). *Nature (London)*, **371**, 578–586.
Brayer, G. D. & McPherson, A. (1983). *J. Mol. Biol.* **169**, 565–596.
Bricogne, G. (1984). *Acta Cryst.* A**40**, 410–445.
Bricogne, G. (1988). *Acta Cryst.* A**44**, 517–545.
Cowtan, K. D. (2000). *Acta Cryst.* D**56**, 1612–1621.
Cowtan, K. D. & Main, P. (1993). *Acta Cryst.* D**49**, 148–157.
Cowtan, K. D. & Main, P. (1996). *Acta Cryst.* D**52**, 43–48.
Folkers, P. J. M., Nilges, M., Folmer, R. H. A., Konings, R. N. H. & Hilbers, C. W. (1994). *J. Mol. Biol.* **236**, 229–246.
Giacovazzo, C. & Siliqi, D. (1997). *Acta Cryst.* A**53**, 789–798.
Goldstein, A. & Zhang, K. Y. J. (1998). *Acta Cryst.* D**54**, 1230–1244.
Gu, Y., Zheng, C., Zhao, Y., Ke, H. & Fan, H. (1997). *Acta Cryst.* D**53**, 792–794.
Hodel, A., Kim, S.-H. & Brünger, A. T. (1992). *Acta Cryst.* A**48** 851–858.
Kleywegt, G. J. (2000). *Acta Cryst.* D**56**, 249–265.
Kleywegt, G. J. & Read, R. J. (1997). *Structure*, **5**, 1557–1569.
Leslie, A. G. W. (1987). *Proceedings of the CCP4 Study Weekend*, pp. 25–31. Warrington: Daresbury Laboratory.
Lunin, V. Y. (1993). *Acta Cryst.* D**49**, 90–99.
Newman, J., Peat, T. S., Richard, R., Kan, L., Swanson, P. E., Affholter, J. A., Holmes, I. H., Schindler, J. F., Unkefer, C. J. & Terwilliger, T. C. (1999). *Biochemistry*, **38**, 16105–16114.
Peat, T. S., Newman, J., Waldo, G. S. Berendzen, J. & Terwilliger, T. C. (1998). *Structure*, **15**, 1207–1214.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Perrakis, A., Sixma, T. K., Wilson, K. S. & Lamzin, V. S. (1997). *Acta Cryst.* D**53**, 448–455.
Plas, J. L. van der & Millane, R. P. (2000). *Proc. SPIE*, **4123**, 249–260.
Podjarny, A. D., Bhat, T. N. & Zwick, M (1987). *Annu. Rev. Biophys. Biophys. Chem.* **16**, 351–373.
Prince, E., Sjolin, L. & Alenljung, R. (1988). *Acta Cryst.* A**44**, 216–222.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Refaat, L. S., Tate, C. & Woolfson, M. M. (1996). *Acta Cryst.* D**52**, 252–256.
Roberts, A. L. U. & Brünger, A. T. (1995). *Acta Cryst.* D**51**, 990–1002.
Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 73–82.
Rossmann, M. G. (1995). *Curr. Opin. Struct. Biol.* **5**, 650–655.
Rossmann, M. G. & Arnold, E. (1993). *International Tables for Crystallography*, Vol. B. edited by U. Shmueli, pp. 230–258. Dordrecht: Kluwer Academic Publishers.
Roversi, P., Blanc, E., Vonrhein, C., Evans, G. & Bricogne, G. (2000). *Acta Cryst.* D**56**, 1316–1323.
Skinner, M. M., Zhang, H., Leschnitzer, D. H., Guan, Y., Bellamy, H., Sweet, R. M., Gray, C. W., Konings, R. N. H., Wang, A. H.-J. & Terwilliger, T. C. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 2071–2075.
Szöke, A. (1993). *Acta Cryst.* A**49**, 853–866.
Szöke, A., Szöke, H. & Somoza, J. R. (1997). *Acta Cryst.* A**53**, 291–313.
Terwilliger, T. C. (1999). *Acta Cryst.* D**55**, 1863–1871.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Vellieux, F. M. D. A. P., Hunt, J. F., Roy, S. & Read, R. J. (1995). *J. Appl. Cryst.* **28**, 347–351.
Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
Wang, J. M., Hartling, J. A. & Flanagan, J. M. (1998). *J. Struct. Biol.* **124**, 151–163.
Wilson, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.
Wilson, C. & Agard, D. A. (1993). *Acta Cryst.* A**49**, 97–104.
Wilson, K. S., Butterworth, S., Dauter, Z., Lamzin, V. S., Walsh, M., Wodak, S., Pontius, J., Richelle, J., Vaguine, A., Sander, C., Hooft, R. W. W. & Vriend, G. (1998). *J. Mol. Biol.* **276**, 417–436.
Xiang, S., Carter, C. W. Jr, Bricogne, G. & Gilmore, C. J. (1993). *Acta Cryst.* D**49**, 193–212.
Zhang, K. Y. J. (1993). *Acta Cryst.* D**49**, 213–222.
Zhang, K. Y. J., Cowtan, K. D. & Main, P. (1997). *Methods Enzymol.* **277**, 53–64.
Zhang, K. Y. J. & Main, P. (1990). *Acta Cryst.* A**46**, 41–46.