

Phase Improvement by Multi-Start Simulated Annealing Refinement and Structure-Factor Averaging

LUKE M. RICE,^a YOUSIF SHAMOO^a AND AXEL T. BRÜNGER^{a,b,*}

^a*Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA, and* ^b*The Howard Hughes Medical Institute, Yale University, New Haven, CT 06520, USA.*

E-mail: brunger@laplace.csb.yale.edu

(Received 3 March 1998; accepted 6 May 1998)

Abstract

Macromolecular crystallography increasingly pushes the limits of the size and complexity of the molecules and assemblies under study. Conformational variability in these assemblies frequently results in limited diffraction or in poorer phase information than that typically obtained from smaller macromolecules. Current methods for solving and refining crystal structures often perform poorly under these adverse conditions and thus new methods must be developed. Atomic-model refinement is particularly sensitive to the low information content of limited diffraction data. Here a multi-start procedure is presented for simulated annealing refinement. After multi-start refinement, poorly fitted regions of the model often display increased variability compared with correctly fitted regions. Structure-factor averaging over the resulting models improves the quality of phases derived from atomic models and reduces model bias. Averaging can be performed even at a minimum Bragg spacing of 3.5 Å, taking into account variability of atomic positions due to errors or intrinsic flexibility even when individual *B* factors cannot be refined. The average structure factor is thus closer to the true structure factor and should provide a better starting point for estimations of σ_A values for electron-density-map calculations. Test cases show increased phase quality of the average structure factor. The method is most useful when the initial model is poor or when only moderate-resolution diffraction data are available and may allow meaningful phase improvement through atomic-model refinement where it was not previously possible.

1. Introduction

A typical single-wavelength diffraction experiment cannot provide a direct image of the electron density in the crystal. Experimental phasing methods or molecular replacement must therefore be used to obtain an image of the electron density. In favorable cases, especially with the availability of multi-wavelength anomalous dispersion (MAD) phasing, these phasing methods can

provide accurate electron-density maps into which an atomic model can easily be built (Hendrickson, 1991). However, there are still many instances where MAD phasing is difficult or unavailable, or where disorder, the low resolution of the diffraction data, or intrinsic flexibility of the crystallized macromolecules produce phases such that initial electron-density maps are of only moderate-to-poor quality. When molecular replacement phasing is used, it is frequently the case that the initial phases are biased towards the search model and give rise to features in the resulting electron-density maps that do not correspond to actual features of the crystallized molecule.

Several types of error can occur when interpreting electron-density maps calculated at moderate resolution and/or with phases of poor quality (Branden & Jones, 1990). Discontinuities in the electron density and spurious connectivities commonly occur, and side-chain density can be misinterpreted as backbone density. These phase errors often result in gross model errors such as misplacement of loops, turns and side-chains. These types of model error are common to structures solved using poor MIR (multiple isomorphous replacement) or MAD phases and to molecular replacement solutions suffering from model bias. Density-modification methods can help alleviate these problems (especially if noncrystallographic symmetry is present), but require delineation of the molecular boundary, a process which again depends on the quality of the initial electron-density map. Thus, in MIR, MAD and molecular replacement, there may be many regions of the initial model that contain significant errors and for which the electron density becomes interpretable only during the course of refinement when model phases are used to improve the initial ones. Phase improvement through atomic-model refinement thus plays a pivotal role in the crystallographic structure-solution process.

Maximum-likelihood approaches (Bricogne, 1984, 1997) have been shown to improve refinement methods (Pannu & Read, 1996; Murshudov *et al.*, 1997), especially in combination with simulated annealing (Adams *et al.*, 1997). They also improve electron-density-map calculations (Read, 1986, 1990). Even these maximum-

likelihood formulations of refinement and electron-density-map calculations are prone to model bias, however. Cross-validation (Brünger, 1992) can reduce this problem and annealed omit maps can further reduce model bias (Hodel *et al.*, 1992). One fundamental problem remains: the model for coordinate error assumes that the error is uniformly distributed among all atoms (Luzzati, 1952; Read, 1986). Individual B -factor refinement can only partially compensate for this assumption, especially when disorder is present (Kuriyan *et al.*, 1986) or when some of the atoms are incorrectly placed (Brünger, 1988).

We propose a method to improve phases derived from atomic models and to account empirically for coordinate error and disorder. The method is possible at any resolution. Multi-start refinements that make use of simulated annealing (Brünger *et al.*, 1987; Rice & Brünger, 1994), starting with the same model but with different initial velocities, generate a number of different refined models (Brünger, 1988), the structure factors of which are averaged (Rice & Brünger, 1994)

$$\mathbf{F}_{\text{average}} = \sum_{\text{models}} (1/N_{\text{models}}) \mathbf{F}_c \quad (\text{model}) \quad (1)$$

and used in place of \mathbf{F}_c for σ_A -value estimation and for electron-density-map calculations. We show that this type of structure-factor averaging provides another layer of improvement when combined with currently available maximum-likelihood methods for refinement and electron-density-map calculations. Chemically meaningful models are used in the averaging process and significant phase improvement was achieved even at moderate-to-low resolution ($d_{\min} = 3.5 \text{ \AA}$). This is in contrast to a random-atom-based averaging technique, wARP (Perrakis *et al.*, 1997; Asselt *et al.*, 1998), which requires a minimum Bragg spacing significantly better than $d_{\min} = 2.5 \text{ \AA}$. The method of time averaging (Gros *et al.*, 1990) is also related to the structure-factor averaging presented here; however, it requires high-resolution data for successful application (Schiffer *et al.*, 1995).

2. Methods

2.1. Theory

The starting point for maximum-likelihood formulations is the conditional probability distribution of the ‘true’ structure factors \mathbf{F} , given a partial model with structure factors \mathbf{F}_c and estimates of its error (Sim, 1959; Srinivasan, 1966; Read, 1986, 1990) (for simplicity we will only discuss the case of acentric reflections)

$$P_a(\mathbf{F}; \mathbf{F}_c) = [1/(\pi\varepsilon\sigma_\Delta^2)] \exp \left[-\frac{(\mathbf{F} - D\mathbf{F}_c)^2}{\varepsilon\sigma_\Delta^2} \right], \quad (2)$$

where σ_Δ is related to the fraction of the asymmetric unit that is missing from the model. Assuming a perfect partial structure and that the missing fraction is

approximately described by a random distribution of p atoms, σ_Δ is given by (Read, 1990)

$$\sigma_\Delta = \left(\sum_p |\mathbf{f}_p|^2 \right)^{1/2} \quad (3)$$

where \mathbf{f}_p are the structure factors of the missing atoms. In general, σ_Δ incorporates the effect of the atoms missing from the model as well as the effect of errors in the partial model. It cannot be directly calculated and is instead estimated (see below). D is a factor that takes into account model error: it is unity in the limiting case of an error-free model and zero if no model is available (Luzzati, 1952; Read, 1986). It can be shown that

$$D = [(\langle |\mathbf{F}_o|^2 \rangle - \sigma_\Delta^2) / \langle |\mathbf{F}_c|^2 \rangle]^{1/2} \quad (4)$$

where the angular brackets denote statistical averages over narrow-resolution shells (see below). For a complete and error-free model, $\mathbf{F}_c = \mathbf{F}_o$, $\sigma_\Delta = 0$ and $D = 1$, so that the probability distribution $P_a(\mathbf{F}; \mathbf{F}_c)$ is infinitely sharp.

Taking into account measurement errors requires multiplication of (2) with an appropriate probability distribution (usually a normal distribution with standard deviation σ_o) of the observed structure-factor amplitudes around the ‘true’ structure-factor amplitudes

$$P_{\text{meas}}(|\mathbf{F}_o|; |\mathbf{F}|) \sim \exp \left[-\frac{(|\mathbf{F}_o| - |\mathbf{F}|)^2}{\sigma_o^2} \right]. \quad (5)$$

Prior knowledge of the phases of the structure factors can be incorporated by multiplication of (2) with a phase probability distribution $P_{\text{phase}}(\varphi)$ and rewriting (2) in terms of the structure-factor moduli and amplitudes of $\mathbf{F} = |\mathbf{F}| \exp(i\varphi)$.

The unknown variables $|\mathbf{F}|$ and φ have to be eliminated by integration in order to obtain the conditional probability distribution of the observed structure-factor amplitudes, given a partial model with errors, the amplitude measurement errors and phase information

$$P_a(|\mathbf{F}_o|; \mathbf{F}_c) = [1/(\pi\varepsilon\sigma_\Delta^2)] \int |\mathbf{F}| d\varphi d|\mathbf{F}| P_{\text{meas}}(|\mathbf{F}_o|; |\mathbf{F}|) \times P_{\text{phase}}(\varphi) \exp \left\{ -\frac{[|\mathbf{F}| \exp(i\varphi) - D\mathbf{F}_c]^2}{\varepsilon\sigma_\Delta^2} \right\}. \quad (6)$$

The likelihood \mathcal{L} of the model is defined as the joint probability distribution of the structure factors of all reflections. Assuming uncorrelated structure factors, \mathcal{L} is simply given by multiplication of the distributions in (6) for all reflections. Instead of maximizing the likelihood it is more common to minimize the negative logarithm of the likelihood,

$$\mathcal{L} = - \sum_{hkl} \log[P_a(|\mathbf{F}_o|; \mathbf{F}_c)]. \quad (7)$$

This function represents part of the refinement target in maximum-likelihood approaches. It is augmented by

chemical restraints (Hendrickson, 1985). Thus, the total target function is given by

$$E = E_{\text{chem}} + w_{\text{xray}}\mathcal{L} \quad (8)$$

where w_{xray} is a weight appropriately chosen to balance the gradients (with respect to atomic parameters) arising from the two terms.

Given the observed diffraction data and an atomic model, σ_{Δ} cannot be obtained on a reflection-by-reflection basis. It is instead assumed to be approximately constant in thin-resolution shells and determined by minimization of \mathcal{L} while keeping the atomic model fixed (Read, 1986, 1997). D is then calculated from σ_{Δ} using the same resolution shells as in (4). In order to avoid model bias and to achieve an improvement over the least-squares residual, cross-validation was found to be essential (Pannu & Read, 1996; Adams *et al.*, 1997) for the computation of σ_{Δ} and D . Occasional recomputation of these values is required as the model improves because refinement methods that improve the model structure factors \mathbf{F}_c have a beneficial effect on both σ_{Δ} and D . Better estimates of these values then enhance the next refinement cycle. Thus, powerful optimization methods and maximum-likelihood targets are expected to interact in a synergistic fashion. Structure-factor averaging of multi-start refinement models can provide another layer of improvement by producing a better description of \mathbf{F}_c if the model shows significant variability due to errors or intrinsic flexibility (see below).

Pannu & Read (1996) have developed an efficient Gaussian approximation (7) for the case of no prior phase information, termed the ‘MLF’ function. In the limit of a perfect model (*i.e.* $\sigma_{\Delta} = 0$ and $D = 1$), MLF reduces to the traditional residual with $1/\sigma_o^2$ weighting. In the event that prior phase information is available, the integration over the phase angles can be carried out numerically in (6), resulting in the ‘MLHL’ function (Pannu *et al.*, 1998). A maximum-likelihood function which expresses (6) in terms of observed intensities has also been developed, termed ‘MLI’ (Pannu & Read, 1996).

The σ_{Δ} and D functions also provide the starting point for optimal electron-density-map calculations. It can be shown that

$$(2m|\mathbf{F}_o| - D|\mathbf{F}_c|) \exp(i\varphi_c) \quad (9)$$

is an approximation of the true structure factor up to second order (Read, 1986). φ_c and m are, respectively, the centroid and figure of merit of the model phase probability distribution obtained from (6)

$$m = [I_1(X)/I_0(X)] \quad (10)$$

where X is given by

$$X = 2\sigma_A E_o E_c / (1 - \sigma_A^2) \quad (11)$$

and E_o and E_c are the normalized structure-factor amplitudes corresponding to \mathbf{F}_o and \mathbf{F}_c , respectively, and

$$\sigma_A = [1 - (\sigma_{\Delta}^2 / \langle |\mathbf{F}_o|^2 \rangle^{1/2})]^{1/2} \quad (12)$$

2.2. Refinement protocol

The refinements employed torsion-angle-dynamics simulated annealing (Rice & Brünger, 1994) against the target E [equation (8)] using the MLF approximation as described in Adams *et al.* (1997). Briefly, σ_{Δ} values were calculated from the starting models and used through the first 200 steps of conjugate-gradient minimization, at which point estimates of σ_{Δ} and the weight (w_{xray}) were updated. The simulated annealing refinement started from a temperature of 5000 K that was decreased to 300 K, at which point 100 steps of Cartesian molecular dynamics at 300 K (Brünger *et al.*, 1987) were carried out. Then 100 steps of conjugate-gradient minimization were performed, σ_{Δ} and w_{xray} values were updated, and a final 100 steps of conjugate-gradient minimization were performed. All observed diffraction data were used to the specified minimum Bragg spacing. A bulk solvent correction (Jiang & Brünger, 1994) and an overall anisotropic B factor were applied. All calculations were carried out with the program *Crystallography and NMR System (CNS)* (Brünger *et al.*, 1998).

2.3. Multi-start refinements

Refinement algorithms differ in the manner by which they optimize E [equation (8)] by generating changes in the atomic model. Gradient-driven methods such as conjugate gradient (Jack & Levitt, 1978) or least-squares (Hendrickson, 1985; Tronrud *et al.*, 1987) are typically unable to make large changes in an atomic model. Molecular-dynamics-based simulated annealing (Brünger *et al.*, 1987; Rice & Brünger, 1994; Gros *et al.*, 1989; Brünger *et al.*, 1997) algorithms are able to make larger changes in the model and are therefore the preferred method when faced with poor initial models.

Another advantage of simulated annealing refinement is that it provides a way of enhancing sampling of conformational space by a multi-start procedure (Brünger, 1988; Rice & Brünger, 1994). Multiple independent refinements can be performed starting from the same initial structure but using different randomly assigned initial velocities for the molecular-dynamics-based simulated annealing. Molecular-dynamics-based simulated annealing is very sensitive to the initial conditions and different starting velocities will result in very different ‘trajectories’. Thus, it is not necessary to modify the initial coordinates of the model in order to obtain good sampling of conformational space: each refinement with different starting velocities will produce a different model even though the same initial model

was used. Independent refinements of poor initial models can produce models which differ considerably in regions of intrinsic flexibility in the molecule (Burling & Brünger, 1994), or, more importantly in the context of this work, in regions of the model that are incorrect (Brünger, 1988) (Fig. 2, see below).

2.4. Structure-factor averaging

The multi-start refinement method described above allows one to perform structure-factor averaging (Rice & Brünger, 1994). Each individual model after a multi-start refinement may have only a subset of the initially incorrectly placed atoms correctly placed and thus errors will remain. As long as these errors differ from model to model, structure-factor averaging over the multiple models should improve phases. The average structure factor is calculated according to (1) and substituted into (2), (4), (6), (7) and (9) in place of F_c . This results in a likelihood function derived from the average structure factor and denoted \mathcal{L}^{ave} , which is minimized to obtain $\sigma_{\Delta}^{\text{ave}}$. The quantities D^{ave} , σ_A^{ave} and m^{ave} are then obtained from $\sigma_{\Delta}^{\text{ave}}$ according to (4), (12) and (10).

In contrast to the random-atom-based wARP technique (Perrakis *et al.*, 1997; Asselt *et al.*, 1998), which uses a weighting scheme that weights each reflection separately, no weighting scheme is required for averaging over models produced by multi-start simulated annealing refinement. The positions of atoms that are not well determined will show more variability than those that are well determined and thus poorly determined atoms will generally contribute less to the average structure factor. Atoms that are correctly placed show less variation in multi-start refinements and thus contribute more to the average structure factor.

Over the course of refinement in certain space groups, the model may undergo a small translation along an axis with continuously permissible origin shifts. These translations (if present) must be corrected for before averaging the structure factors.

2.5. Test cases

To obtain a controlled test case with which to evaluate the performance of various refinement strategies, errors were systematically introduced into the crystal structure coordinates of penicillopepsin (James & Sielecki, 1983) ($d_{\text{min}} = 1.8 \text{ \AA}$, space group $C2$, unit-cell dimensions $a = 97.37$, $b = 46.64$, $c = 65.47 \text{ \AA}$, $\beta = 115.4^\circ$) by performing molecular dynamics simulations without reference to E_{xray} as described in Rice & Brünger (1994). A series of six models with good geometry but with backbone-atom coordinate error [root mean square (r.m.s.) difference to the crystal structure] ranging from 0 to more than 2.0 \AA were refined against the penicillopepsin diffraction amplitudes truncated to a

minimum Bragg spacing of $d_{\text{min}} = 2.8$ or 3.5 \AA . Water molecules were not included in these test calculations.

The model errors obtained by this molecular dynamics ‘scrambling’ procedure are potentially representative of those occurring when energy-minimized homologous structures are used as molecular replacement search models. Indeed, the models are realistic in that when they are used as search models for molecular replacement using the experimental penicillopepsin diffraction data, the correct orientation and position of the molecule in the unit cell are correctly identified as the top peaks of the rotation (1 standard deviation above the first noise peak for the worst model) and translation (1.6 standard deviations above the first noise peak for the worst model) functions. Test refinements using actual homologous structures of penicillopepsin yield similar results (Adams *et al.*, 1998).

The second test case presented below is the refinement of the two RNA-binding domains of heterogeneous ribonucleoprotein A1 (hnRNP A1, referred to as A1) using an intermediate data-set with $d_{\text{min}} = 2.4 \text{ \AA}$ resolution (Shamoo *et al.*, 1997). Unit-cell dimensions in space group $P2_1$ are $a = 38.1$, $b = 44.0$, $c = 56.1 \text{ \AA}$ and $\beta = 94.8^\circ$. The A1 test case typifies some of the problems encountered when working with only low-to-moderate-resolution data. Phases for the initial MIR electron-density map used for model building were obtained from two heavy-atom derivative data-sets, one of which only diffracted to a minimum Bragg spacing of 4 \AA . The overall figure of merit for the MIR phases was 0.49 (Shamoo *et al.*, 1998). Model building relied heavily on the coordinates of a homologous RNA-binding domain (U1A; Nagai *et al.*, 1990), resulting in a significant model bias. Thus, although native diffraction amplitudes to higher resolution were measured later, the initial model was obtained by tracing a poor 4 \AA resolution electron-density map and contained several substantial errors.

The performance of the multi-start structure-factor-averaging method was assessed by comparison with the phases and electron-density maps calculated from the refined high-resolution crystal structures (1.8 \AA penicillopepsin and 1.78 \AA for A1). Phase differences were computed with unit weighting and map correlation coefficients were computed over the protein region.

3. Results

The variability of the free R value (Brünger, 1992) and phase accuracy is shown in Fig. 1 (horizontal bars) for multi-start refinements of the penicillopepsin test case. This variability is also manifest in the atomic coordinates. For multi-start refinements of the original 1.8 \AA crystal structure coordinates using diffraction amplitudes truncated to $d_{\text{min}} = 2.8 \text{ \AA}$, the average backbone-atom pairwise r.m.s. deviation between the resulting ten models was 0.08 \AA . For multi-start refinements of the most scrambled starting model against the same

diffraction data, this deviation increased to 0.50 Å. Similar variability was observed for multiple refinements of A1. For example, for a multi-start refinement of the initial model of A1, the average pairwise backbone-atom r.m.s. deviation between models was 0.71 Å (Fig. 2a). For a multi-start refinement of a later, improved model, this measure was much lower, 0.25 Å for the same atoms as for the early model (Fig. 2b). In general, as the starting models improved, the variability produced by multi-start refinement decreased. This decrease in model variability generally provides a convenient measure of the convergence of refinement.

Structure-factor averaging results in a significant improvement in the free R value, the phase accuracy and σ_A -weighted electron-density maps for the penicillopepsin test case (Fig. 1, Table 1). Phases from the

average structure factor are better (typically by several degrees) than phases from any single model. The improvements in the free R value, the phase accuracy and the map correlation coefficient are most pronounced when the initial model is very poor. These improvements are also particularly significant at lower resolution ($d_{\min} = 3.5$ Å). The ability to improve phases from refinements of poor initial models or from refinements against limited data is an important property of structure-factor averaging in combination with multi-start refinements.

Multi-start refinements were also performed for A1 and the average structure factor was used to compute a σ_A -weighted (Read, 1990) $2F_o - F_c$ electron-density map. For the initial model of A1, free R values computed from average structure factors tended to be about 1%

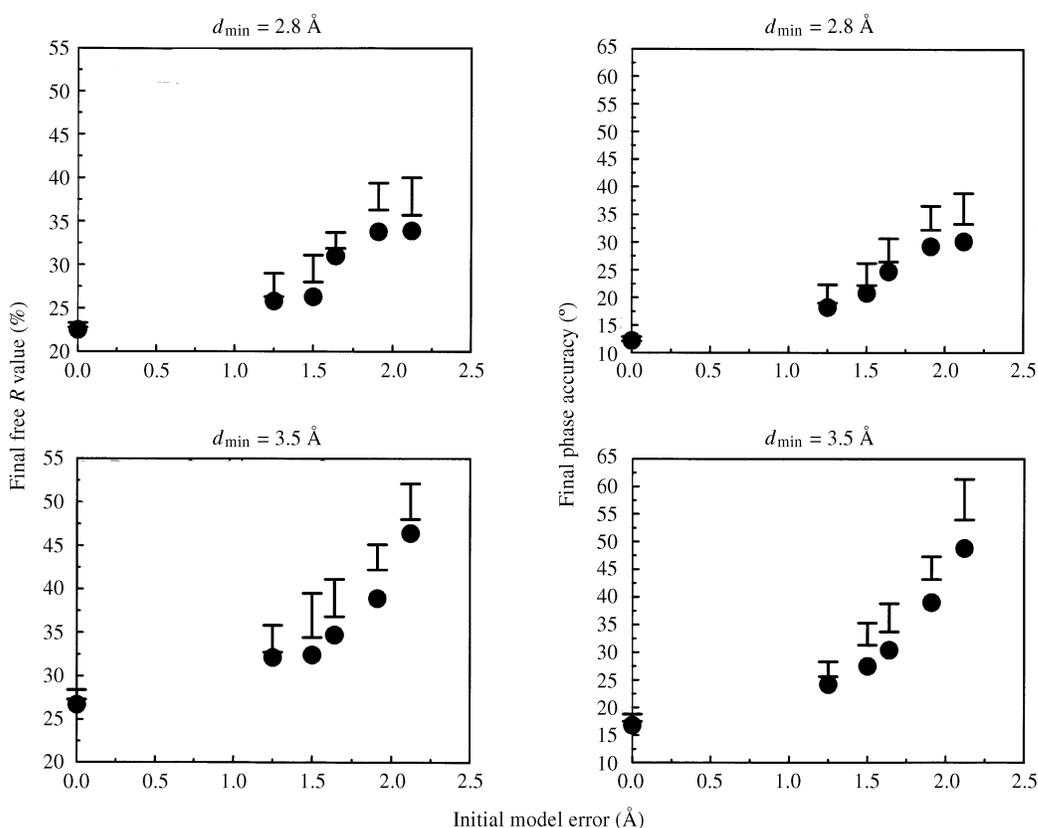


Fig. 1. Convergence, variability and effect of structure-factor averaging on the free R values and phase accuracy of scrambled models of penicillopepsin refined against observed structure-factor amplitudes at two different minimum Bragg spacings. R values and phase accuracy are plotted against initial model error (backbone-atom r.m.s. difference to the 1.8 Å resolution crystal structure coordinates). The top row shows the results from multi-start refinements (ten starts) using diffraction data truncated to $d_{\min} = 2.8$ Å resolution, the bottom row using data truncated to $d_{\min} = 3.5$ Å resolution. In the two left panels, the performance of refinements is measured using the free R value (horizontal bars indicate the highest and lowest free R value from each set of ten refinements). The solid circles show the free R value obtained from the average structure factor. In the two right panels, the performance of refinements is measured using the phase accuracy, *i.e.* the unweighted mean difference between phases calculated from a particular refined model and those calculated from the refined 1.8 Å crystal structure coordinates. Horizontal bars again indicate the highest and lowest values, and solid circles show the phase accuracy of the average structure factor. Refinements were performed as described in §2. No water molecules were included. The average structure factor produces better phases than any single model. The most significant improvements are obtained when the starting model is very poor or when the diffraction data were truncated to $d_{\min} = 3.5$ Å.

Table 1. Effect of structure-factor averaging on the map correlation coefficient for multi-start refinements of the most scrambled penicillopepsin model (the rightmost point in the graphs in Fig. 1)

Map correlation coefficients were computed between the σ_A -weighted electron-density map calculated using the average structure factor and a σ_A -weighted map calculated using the final high-resolution crystal structure of penicillopepsin (James & Sielecki, 1983). In the correlation coefficient calculations only regions over protein atoms were taken into account. Best and worst refined models are those with the lowest and highest free R values, respectively. The average structure factor is the one that gives the lowest free R value. σ_A -weighted maps computed from the average structure factor show higher correlation to the final crystal structure than those computed from any single model.

d_{\min} (Å)	Map correlation coefficient		
	Best model	Worst model	Average structure factor
2.8	0.71	0.68	0.74
3.5	0.34	0.22	0.39

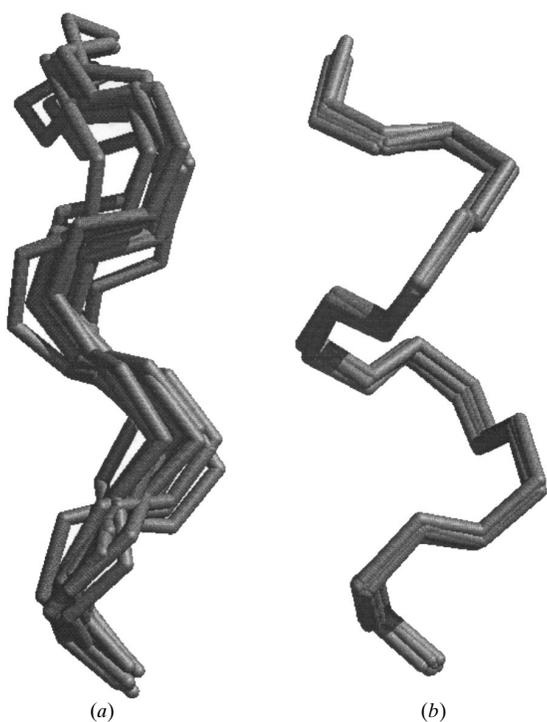


Fig. 2. Conformational variability in refined models after multi-start refinement. Ten start refinements were performed on the initial model of A1 and on a later, manually rebuilt model. The region corresponding to residues 67–74 in the crystal structure is shown; this region at first contained a register error for the α helix. The refinement protocol was as described in §2. In (a), the α -helix backbone had been built in the incorrect register in the initial model of A1 used for multi-start refinement. In (b), this error had been corrected (after refinement and manual rebuilding) in the starting model used for multi-start refinement. The refined models containing the register-shift error show considerably more variability (average backbone-atom pairwise r.m.s. difference for the α helix 0.72 Å) compared with the refined models with the correct register (average backbone-atom pairwise r.m.s. difference for the α helix 0.25 Å). Thus, the error in the model produced a variability in the atomic positions which contributes to the success of structure-factor averaging.

lower than the best free R value from any single model (Fig. 4a). The R value decreased less and thus the reduced $R_{\text{free}} - R$ difference indicates a reduction in the degree of overfitting. Other measures like the phase accuracy and the map correlation coefficient demonstrate the improvement in the quality of electron-density maps upon structure-factor averaging (Fig. 4). Structure-factor averaging resulted in more interpretable electron-density maps in difficult regions (Fig. 3). In the example shown, the initial model contained a register shift involving an α helix (encompassing residues 64–74 in the final, high-resolution structure); electron-density maps computed from the two refined models (Figs. 3a and 3b) with the lowest free R value show significant model bias. The electron-density map computed from the average structure factor of the two models is less model biased and shows more clearly the correct location of the α -helix backbone. In the later stages of refinement, multi-start refinement produced less variable models (e.g. Fig. 2), indicating that refinement had converged.

The ensemble of structures produced by multi-start refinement can have practical utility beyond merely providing better phases. During the refinement of A1, it was very useful to inspect the structures using molecular graphics before starting manual rebuilding. Regions of the model that contained significant errors such as register shifts or other backbone misplacements tended to show more variability in the ensembles than regions that were essentially correct (Fig. 2). In general, although one or several models may stand out due to lower free R values, all of them can be useful as guides for model rebuilding.

4. Summary

The success of structure-factor averaging of individually refined models and the reduction in overfitting can be easily understood. The reciprocal-space structure-factor averaging is formally equivalent to real-space electron-density-map averaging. Thus, regions of the electron density that are well predicted by several models will be more heavily weighted in the averaged map compared with poorly modeled regions. Regions of the model that are incorrect can show greater variability in the ensemble of refined structures and the electron density corresponding to these regions will be attenuated by the averaging. Estimates of σ_A (Read, 1990) are also improved because contributions from individual atoms will be weighted according to how much variability they display after a multi-start refinement (Table 1, Fig. 4). Electron-density maps computed from averaged structure factors typically display less model bias than an electron-density map computed from a single model (Fig. 3).

Structure-factor averaging using multi-start refinements, maximum-likelihood targets and torsion-angle

dynamics all interact synergistically. Compared with least-squares residuals, maximum-likelihood targets reduce model bias. Multi-start refinements sample more local minima. Structure-factor averaging of the resulting models produces more reliable, less biased electron-

density maps. Simulated annealing constrained to torsion angles increases the radius of convergence of individual refinements. The combination of all these methods produces better model phases and can thereby significantly facilitate the rebuilding process and accel-

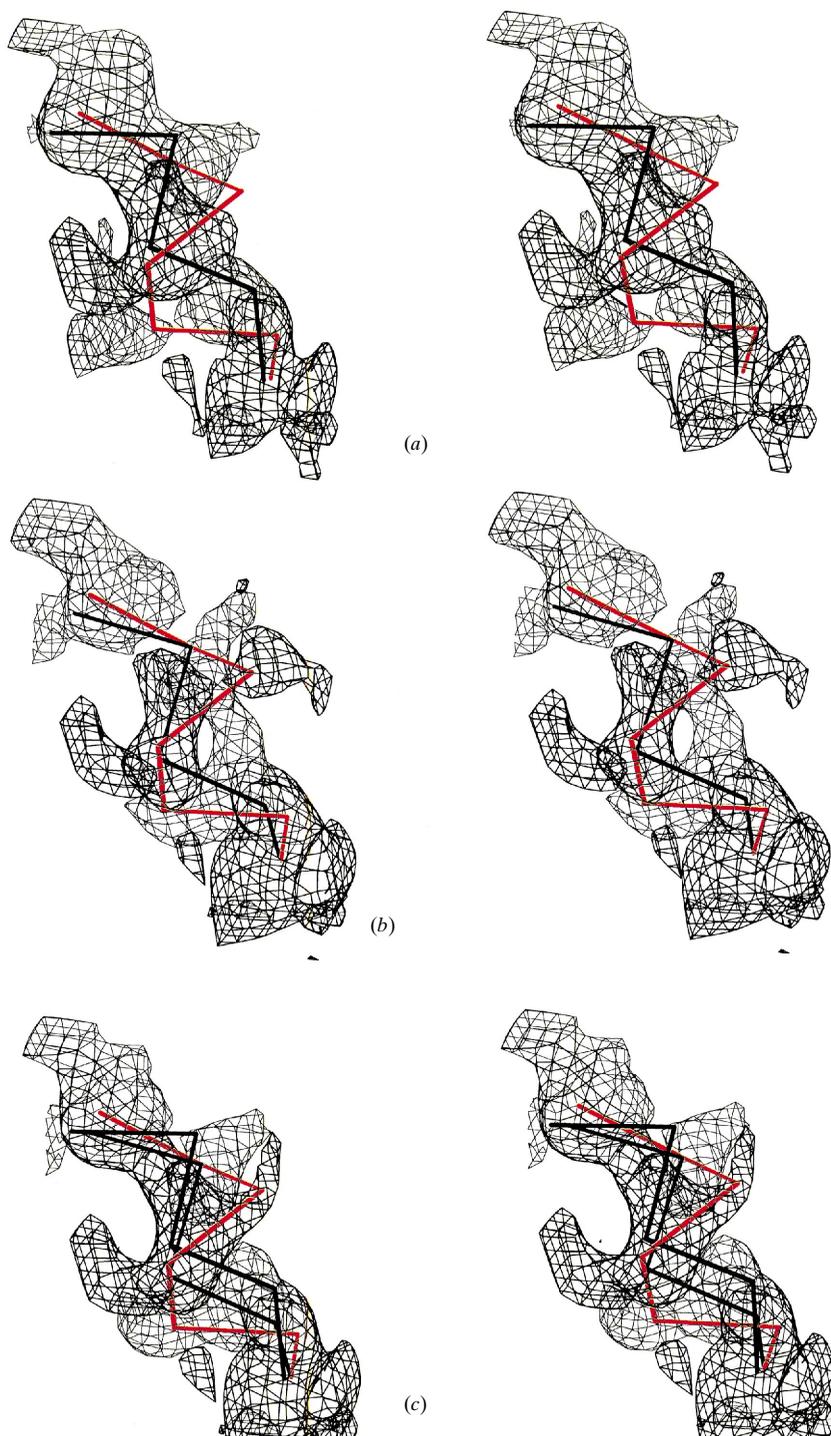


Fig. 3. Improved electron-density maps of A1 in the region of residues 68–74 obtained from multi-start refinements followed by structure-factor averaging. In this case, the best average structure factor (as judged by the free R value) came from averaging over two models. All electron-density maps are a σ_A -weighted $2F_o - F_c$ electron-density map calculated at 2.4 Å resolution and contoured at 1σ . The map calculated from the average structure factor using σ_A^{ave} shows less model bias than maps calculated from single models. (a) Electron-density map calculated from the single model with the lowest free R value (49.45%). In black is shown the C_α trace of the model, in red the C_α trace of the refined crystal structure coordinates. (b) Electron-density map calculated from the single model with the second lowest free R value (49.50%). In black is shown the C_α trace of the model, in the red C_α trace of the refined crystal structure coordinates. (c) Electron-density map calculated from the average structure factor obtained from the two best models. The free R value for this average structure factor is 48.36%. In black is shown the C_α traces of the two models, in red the C_α trace of the final high-resolution crystal structure coordinates of A1 (Shamoo *et al.*, 1997).

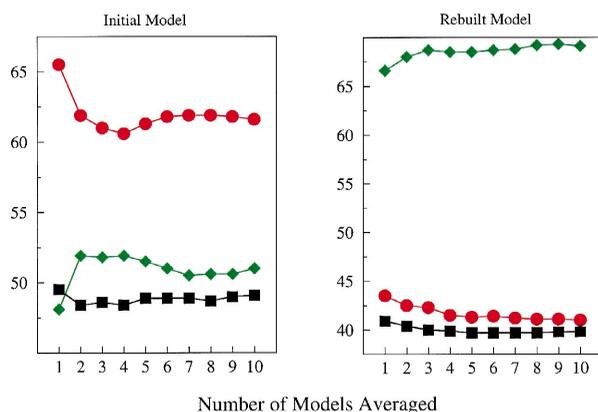


Fig. 4. Effect of structure-factor averaging on the free R value (black symbol) (%), the phase accuracy (red symbols) ($^{\circ}$) and the map correlation coefficient (green symbols) (%). For the initial and a later, rebuilt model of A1, refinement was performed as described in §2. All observed diffraction amplitudes to $d_{\min} = 2.4 \text{ \AA}$ were used. Phase accuracy is defined as the unweighted difference between model phases and those computed from the final, high-resolution crystal structure (Shamoo *et al.*, 1997). Map correlation coefficients were calculated over the protein region. Ten starts were performed for multi-start refinement and the resulting models were sorted in order of increasing free R value. Average structure factors were calculated from the top two, the top three, up to all ten models. For each average, the free R value, phase accuracy and map correlation coefficient are plotted.

erate the process of model phase improvement. Perhaps more importantly, the multi-start averaging strategy allows productive phase improvement through refinement where other methods fail, notably in the moderate-to-low-resolution regime. This method should become increasingly important for crystallographic studies of very large macromolecules or of large macromolecular assemblies.

We thank Paul Adams for advice and stimulating discussions. LMR is an HHMI predoctoral fellow. This work was supported in part by a grant from the National Science Foundation to ATB (BIR 9514819).

References

Adams, P. D., Pannu, N. S., Read, R. J. & Brünger, A. T. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 5018–5023.
 Adams, P. D., Pannu, N. S., Read, R. J. & Brünger, A. T. (1998). In preparation.

Asselt, E. J. van., Perrakis, A., Kalk, K. H., Lamzin, V. S. & Dijkstra, B. W. (1998). *Acta Cryst. D***54**, 58–73.
 Branden, C. I. & Jones, A. T. (1990). *Nature (London)*, **343**, 687–689.
 Bricogne, G. (1984). *Acta Cryst. A***40**, 410–445.
 Bricogne, G. (1997). *Methods Enzymol.* **276**, 361–423.
 Brünger, A. T. (1988). *J. Mol. Biol.* **203**, 803–816.
 Brünger, A. T. (1992). *Nature (London)*, **355**, 472–474.
 Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst. D***54**, 905–921.
 Brünger, A. T., Adams, P. D. & Rice, L. M. (1997). *Structure*, **5**, 325–336.
 Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Science*, **235**, 458–460.
 Burling, F. T. & Brünger, A. T. (1994). *Isr. J. Chem.* **34**, 165–175.
 Gros, P., Betzel, C., Dauter, Z., Wilson, K. S. & Hol, W. G. J. (1989). *J. Mol. Biol.* **210**, 347–367.
 Gros, P., van Gunsteren, W. F. & Hol, W. G. J. (1990). *Science*, **249**, 1149–1152.
 Hendrickson, W. A. (1985). *Methods Enzymol.* **115**, 252–270.
 Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
 Hodel, A., Kim, S.-H. & Brünger, A. T. (1992). *Acta Cryst. A***48**, 851–859.
 Jack, A. & Levitt, M. (1978). *Acta Cryst. A***34**, 931–935.
 James, M. N. G. & Sielecki, A. R. (1983). *J. Mol. Biol.* **163**, 299–361.
 Jiang, J.-S. & Brünger, A. T. (1994). *J. Mol. Biol.* **243**, 100–115.
 Kuriyan, J., Petsko, G. A., Levy, R. M. & Karplus, M. (1986). *J. Mol. Biol.* **190**, 227–254.
 Luzzati, V. (1952). *Acta Cryst.* **5**, 802–810.
 Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst. D***53**, 240–255.
 Nagai, K., Oubridge, C., Jessen, T. H., Li, J. & Evans, P. R. (1990). *Nature (London)*, **348**, 515–520.
 Pannu, N. S., Murshudov, G. N., Dodson, E. J. & Read, R. J. (1998). *Acta Cryst. D*. In the press.
 Pannu, N. S. & Read, R. J. (1996). *Acta Cryst. A***52**, 659–668.
 Perrakis, A., Sixma, T. K., Wilson, K. S. & Lamzin, V. S. (1997). *Acta Cryst. D***53**, 448–455.
 Read, R. J. (1986). *Acta Cryst. A***42**, 140–149.
 Read, R. J. (1990). *Acta Cryst. A***46**, 900–912.
 Read, R. J. (1997). *Methods Enzymol.* **278**, 110–128.
 Rice, L. M. & Brünger, A. T. (1994). *Proteins*, **19**, 277–290.
 Schiffer, C. A., Gros, P. & van Gunsteren, W. F. (1995). *Acta Cryst. D***51**, 85–92.
 Shamoo, Y., Krueger, U., Rice, L. M., Williams, K. R. & Steitz, T. A. (1997). *Nat. Struct. Biol.* **3**, 215–222.
 Sim, G. A. (1959). *Acta Cryst.* **12**, 813–815.
 Srinivasan, R. (1966). *Acta Cryst.* **19**, 143–144.
 Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. (1987). *Acta Cryst. A***43**, 489–500.