Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

A two-dimensional histogram-matching method for protein phase refinement and extension

Yeu-Perng Nieh and Kam Y. J. Zhang*

Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N, Seattle, Washington 98109, USA

Correspondence e-mail: kzhang@fhcrc.org

Received 16 June 1999
Accepted 19 August 1999

A new method has been developed for protein phase improvement using the joint distribution of the electron density and its gradient (two-dimensional histogram) as a constraint in a density-modification procedure. Matching the two-dimensional (2D) histogram of a given map to that of an ideal 2D histogram was achieved through alternating applications of one-dimensional (1D) histogram matching on electron density and on density gradient. The 2D histogram-matching method was compared with the 1D density histogram-matching method for phase refinement and extension starting from either medium-resolution or high-resolution data on three different types of phases. These included phase refinement and extension using MIR phases of T_6 insulin and phases with randomly generated errors. The test results demonstrated significant improvement of the phases and the overall map quality using the 2D histogram-matching method compared with the 1D density histogram-matching method in each of the three test cases.

1. Introduction

One of the bottlenecks in macromolecular structure determination by X-ray diffraction is the phase problem. Phases are generally estimated experimentally using either the multiple isomorphous replacement method (MIR; Perutz, 1956) or the multiple anomolous dispersion (MAD; Hendrickson *et al.*, 1990) method. Estimated phases generally suffer from inaccuracies owing to experimental error. Phases measured using the MIR method sometimes suffer from lack of isomorphism and rarely extend to the full resolution of the native data. The electron-density maps calculated from these estimated phases might not be of sufficient quality to build an atomic model. Various density-modification techniques have been proposed to improve the quality of the electron-density maps and thereby the phases by imposing some known physical constraints on the electron density (Cowtan *et al.*, 1999; Podjarny *et al.*, 1996; Zhang *et al.*, 1997, 1999).

The most commonly used density-modification method is solvent flattening (Wang, 1985), which exploits the observation that the solvent region of the electron-density map is featureless at medium resolution owing to the high thermal motion and disorder of the solvent molecules. Flattening of the solvent region suppresses noise in the electron-density map and thereby improves phases. A complementary method to solvent flattening is histogram matching (Zhang & Main, 1990*a*), which modifies the protein region of the map by systematically adjusting the electron-density values so that the electron-density distribution conforms to an ideal distribution.

The electron-density distribution (one-dimensional histogram) specifies not only the permitted values of the electron density but also their frequencies of occurrence. It provides a global description of the appearance of the map and all spatial information is discarded. The ideal histogram for an unknown structure is predictable since it is independent of conformation. Proteins generally have very similar atomic compositions and atomic bonding patterns. What makes each protein structure different is the sequence of amino-acid residues and the dihedral angles adopted by each residue. The density histogram discards the spatial information and is therefore independent of the factors which make each structure unique. The density histogram captures the commonality between different structures: the similar atomic composition and the characteristic distances between atoms. These common features distinguish correct from incorrect structures. Therefore, the ideal density histogram can be used to improve an electron-density map (Harrison, 1988; Lunin, 1988; Zhang & Main, 1990*a*) or to select a correct phase set among many randomly generated phase sets in *ab initio* phasing (Lunin *et al.*, 1990).

The density histogram is degenerate in encoding structural information. While having an ideal density distribution is a necessary condition for being a correct structure, it is not a sufficient condition. Many incorrect structures may also have ideal density histograms. Moreover, the density histogram does not capture all the common features found in protein structures. The electron-density gradient and the Laplacian (second derivative) of electron density have been investigated in order to reduce the degeneracy in electron-density histograms (Zhang, 1989). The inclusion of electron-density gradient and the Laplacian in a multi-dimensional histogram was shown to be more sensitive to phase error than the one-dimensional (1D) histogram and suggested to be a more effective constraint for phase improvement (Xiang & Carter, 1996). The local environments of the electron-density map, such as the local minimum, maximum and variance of the density, were used to decouple the correlation of electron-density order in the 1D histogram-matching method (Refaat *et al.*, 1996).

Here, we introduce a two-dimensional (2D) histogram-matching method which employs the joint probability distribution of the electron-density value and its gradient as a constraint in a density-modification procedure for phase improvement. The density gradient reflects the change of the density value within a local region around a given point and therefore provides a description of the local environment. The information from the gradient distribution is complementary to the density distribution, which only accounts for the value at a given point and disregards its neighboring environment. The addition of gradient in the 2D histogram reduces the degeneracy in the 1D density histogram. We have shown previously that the 2D histogram is independent of the structural conformation and is sensitive to phase error, which suggests that it could be used as a constraint for phase improvement (Goldstein & Zhang, 1998). Here, we tested the effectiveness of this new constraint for phase improvement and compared it with solvent flattening and the 1D histogram-matching method.

2. Methods

2.1. The calculation of electron-density gradients

Given the electron density ρ at position (x, y, z) as a Fourier transform of structure factors $F(hkl)$,

$$\rho(x, y, z) = (1/V)\sum_{hkl} F(hkl)\exp[-2\pi i(hx + ky + lz)], \quad (1)$$

where (x, y, z) are the fractional coordinates along the crystal axes (a, b, c), (hkl) are the structure-factor indices and V is the unit-cell volume, the gradients along each of the three crystal axes (a, b, c) are calculated as

$$\begin{aligned}
\frac{\partial\rho(xyz)}{\partial x} &= -\frac{2\pi i}{V}\sum_{hkl} hF(hkl)\exp[-2\pi i(hx + ky + lz)] = g_x,\\
\frac{\partial\rho(xyz)}{\partial y} &= -\frac{2\pi i}{V}\sum_{hkl} kF(hkl)\exp[-2\pi i(hx + ky + lz)] = g_y,\\
\frac{\partial\rho(xyz)}{\partial z} &= -\frac{2\pi i}{V}\sum_{hkl} lF(hkl)\exp[-2\pi i(hx + ky + lz)] = g_z. \quad (2)
\end{aligned}$$

The three gradient maps g_x, g_y and g_z can be efficiently calculated by FFT using the modified Fourier coefficients hF, kF and lF, respectively. The three gradient maps g_x, g_y and g_z are then transformed from the crystal axes system (x, y, z) to the orthogonal axes system (u, v, w) before the accumulation of 2D histograms.

$$\begin{pmatrix} g_u & g_v & g_w \end{pmatrix} = \begin{bmatrix} a & b\cos\gamma & c\cos\beta \\ 0 & b\sin\gamma & c\left(\frac{\cos\alpha - \cos\beta\cos\gamma}{\sin\gamma}\right) \\ 0 & 0 & V/(ab\sin\gamma) \end{bmatrix} \begin{pmatrix} g_x \\ g_y \\ g_z \end{pmatrix}. \quad (3)$$

Here, a, b, c, α, β and γ are the unit-cell dimensions and angles, and g_u, g_v and g_w are the gradient maps along the orthogonal axes.

Given an electron-density map and its gradient maps, the 2D histogram of the joint probability distribution of the electron density and its gradients can be obtained as previously described (Goldstein & Zhang, 1998).

2.2. The calculation of structure factors from modified gradient maps

The gradient maps can be modified by 1D histogram matching on gradients in an analogous way to the 1D histogram matching on density. The modified gradient maps are transformed back into the crystal axes system from the orthogonal axes system. Although the histogram matching is carried out on the density gradients, it is necessary to transform the modified gradient maps back to structure factors in order to carry out the phase combination with the observed structure factors. Based on equation (2), we have found that the structure factors can be calculated through inverse Fourier transform with either of the three density gradients g_x, g_y and g_z as the Fourier coefficients,

$$hF(hkl) = -(V/2\pi iN)\sum_{xyz} g_x \exp[2\pi i(hx + ky + lz)],$$
$$kF(hkl) = -(V/2\pi iN)\sum_{xyz} g_y \exp[2\pi i(hx + ky + lz)],$$
$$lF(hkl) = -(V/2\pi iN)\sum_{xyz} g_z \exp[2\pi i(hx + ky + lz)]. \quad (4)$$

Therefore, after the histogram matching on g_x, g_y and g_z followed by the inverse FFT on each of the three modified gradient maps, three structure-factor sets are generated. A single structure-factor data set is obtained by averaging the equivalent reflections in the three structure-factor sets.

2.3. The 2D histogram-matching procedure

The 2D histogram matching on density and its gradients is achieved through two alternating steps of 1D histogram matching on density and 1D histogram matching on gradients. The 1D histogram matching on density follows the method described by Zhang & Main (1990*a*). In this method, the new electron-density value is derived from the old electron-density value through a linear transform such that the cumulative distribution of the new density value equals the cumulative distribution of the ideal histogram. The histogram matching on gradients also follows a similar protocol, in which the density value was replaced by the gradients. The modified gradient maps were converted to the modified structure factors by the fast Fourier transform method, as shown in (4).

The implementation of 2D histogram matching has been tested in two modes, the parallel mode and the sequential mode. For the parallel mode, the histogram matching on density and gradients is applied in parallel using the same initial structure-factor set. After matching, the two new structure-factor sets are combined by vector summation. For the sequential mode, the structure-factor set calculated after density histogram matching is used as input for the gradient histogram matching and *vice versa*. Test results showed that the histogram matching in sequential mode gave better phase improvements and converged closer to the ideal 2D histogram in fewer matching cycles compared with the parallel mode. Therefore, the test cases shown in §3 are all based on the sequential mode, which is illustrated in Fig. 1. The effectiveness of the 2D histogram-matching procedure in bringing the initial distribution to the desired distribution is shown in Fig. 2. The new 2D histogram-matching technique was incorporated into the density-modification program *SQUASH* (Zhang, 1993; Zhang & Main, 1990*b*), which is available from our www site, http://squash.fhcrc.org.

The consensus histograms used in the test cases were generated from the atomic coordinates of fibroblast growth factor (4fgf; Eriksson *et al.*, 1993) after removing the overall temperature factor from the electron-density map. Since 2D histograms are resolution dependent, several ideal histograms were generated for a range of resolutions from 4.0 to 1.5 Å using the method described by Goldstein & Zhang (1998). During phase extension to the higher resolution, the histogram matching was carried out with a range of ideal histograms at different resolutions.

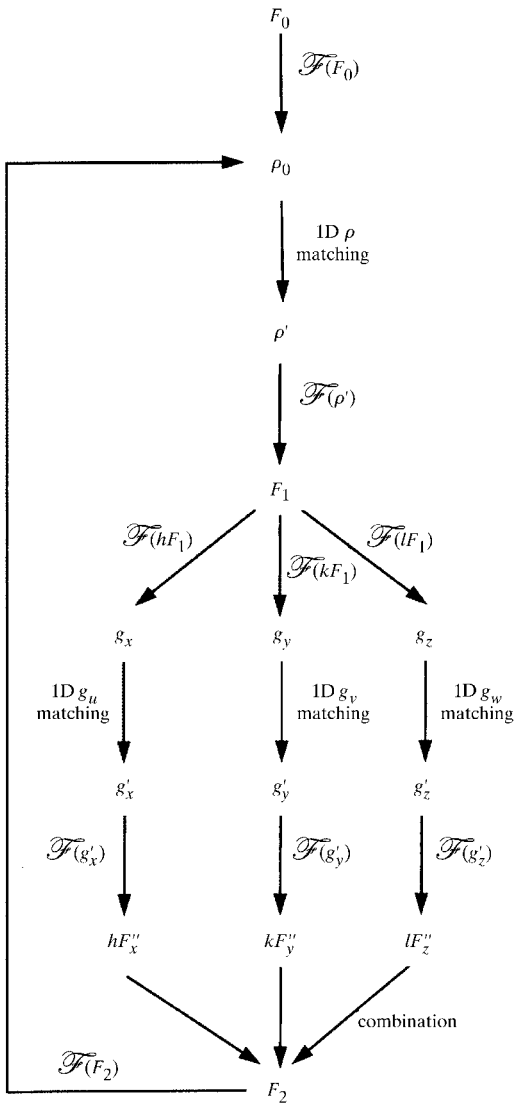

Figure 1
The 2D histogram-matching procedure. 2D histogram matching is achieved through alternating application of 1D histogram matching on electron density and 1D histogram matching on density gradients in a sequential mode. The ideal 1D density histogram and 1D gradient histogram are obtained from the projection of the ideal 2D histogram along the gradient and density, respectively. Starting from an initial structure-factor set F_0, an electron-density map ρ_0 is calculated by a fast Fourier transform $\mathcal{F}$. The initial map is modified by 1D histogram matching to produce a new map ρ', the density histogram of which conforms to the ideal density histogram. The modified map ρ' is Fourier transformed to give a new structure-factor set F_1, from which three gradient maps, g_x, g_y and g_z, are calculated along each crystal axis. The three gradient maps are transformed from the crystal axes system to the orthogonal axes system. The transformed gradient maps, g_u, g_v and g_w, are modified by 1D histogram matching on gradients similar to that on density to produce new gradient maps, the gradient distribution of which matches the ideal gradient distribution. The new gradient maps are then transformed from the orthogonal axes to the fractional axes. Three sets of structure factors, F''_x, F''_y and F''_z, are calculated from each of the three new gradient maps, g'_x, g'_y and g'_z by fast Fourier transforms. The three sets of structure factors, F''_x, F''_y and F''_z, are combined to produce a new set of structure factors, F_2, from which a new map ρ_0 is calculated. The process is iterated until the 2D histogram of the modified map matches that of the ideal 2D histogram.

3. Results

3.1. Test structure and phase set selection

The 2D histogram matching was tested with both MIR phases and calculated phases with random errors. The MIR phases contain both random and systematic errors. The contribution of these two components to MIR phase errors varies from structure to structure. While random errors can all be modeled statistically and are therefore easier to eliminate, systematic errors vary from case to case and are difficult to model and therefore harder to eliminate. Phase refinement and extension results from phases with random errors will demonstrate the upper limit of phase improvement when no systematic errors are present in the MIR phases.

The structure of T_6 insulin (Baker *et al.*, 1988) was used to test the effectiveness of this new approach. The native structure-factor amplitudes were collected to 1.5 Å and MIR phases were reliable to 1.9 Å. Phases with random errors were created by adding randomly generated errors to the calculated phases from the atomic structure. The random error distribution followed that of the MIR phases both as a function of resolution and structure-factor amplitude. Two schemes were used to estimate the figure of merit (FOM) for phases with random errors. The first scheme assigns the estimated FOM to all reflections in a resolution bin based on the mean phase error within that resolution bin (hereafter referred to as the 'mean FOM scheme'). The second scheme estimates the FOM based on the phase error for each reflection (hereafter referred as the 'individual FOM scheme'); this is an optimum case where a perfect weighting scheme is available for each reflection.

The 2D histogram-matching procedure was monitored using phase errors and map correlation coefficients. The phase error is a non-weighted average of the phase differences between the phases being monitored and the correct phases calculated from the atomic model (Baker *et al.*, 1988). The map correlation coefficient is the correlation between the current map and the map calculated from the atomic model at the target resolution. The phases and maps calculated from the atomic model are only biased approximations to the correct phases and true maps owing to the incompleteness of the model (Sayre, 1974). However, the phases and maps calcu-

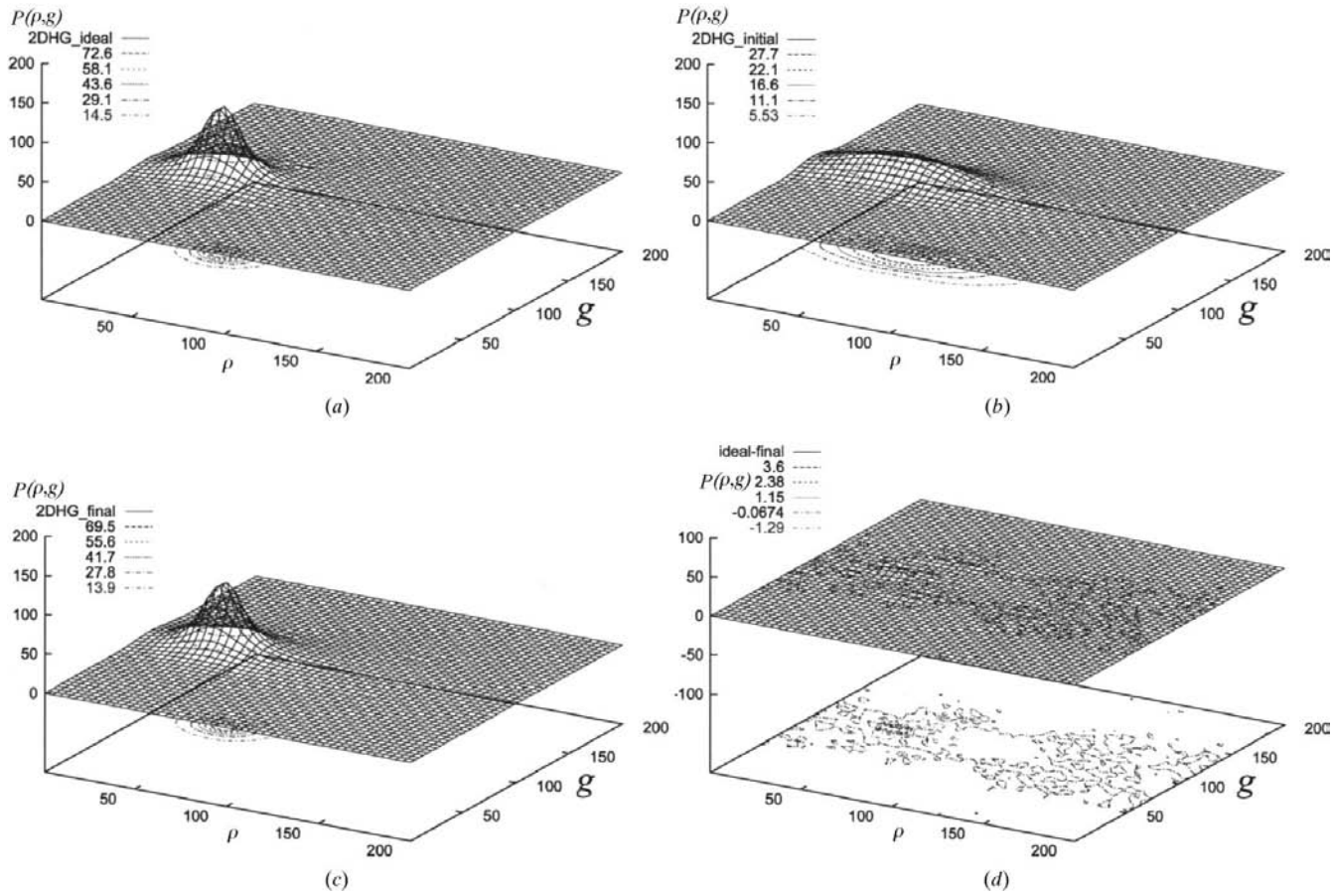

Figure 2
An example of 2D histogram matching. The initial 2D histogram (*b*) is transformed to the final 2D histogram (*c*) after three cycles of iterative application of 1D histogram matching on density and 1D histogram matching on gradients to match that of the target ideal 2D histogram (*a*). The final 2D histogram is very similar to the target ideal 2D histogram. The difference between the final 2D histogram and the ideal 2D histogram is shown in (*d*), which contains minimum residual. The density ρ and gradient g are in grid-point units. Here g represents either g_u, g_v or g_w. The 2D histogram $P(\rho, g)$ is in relative units.

Table 1
Phase and map improvement statistics.

Three different density-modification techniques, solvent flattening and 1D and 2D histogram matching, were compared in terms of the improvement of mean phase errors ($\Delta\varphi$) and map correlation coefficients (MCC). Three phase sets, MIR phases and two sets of phases with randomly generated phase errors, were used as test cases. Phase refinement and extension were tested at 3.0 Å with extension to 2.0 Å, and refinement at 1.9 Å with extension to 1.5 Å, respectively.

	Initial†		SF†		SF + 1D†		SF + 2D†	
	$\Delta\varphi$ (°)‡	MCC§	$\Delta\varphi$ (°)	MCC	$\Delta\varphi$ (°)	MCC	$\Delta\varphi$ (°)	MCC
Phase refinement at 3.0 Å and extension to 2.0 Å								
MIR, 3.0 Å	46.8		44.6		42.7		42.4	
3.0–2.0 Å	90.0		90.4		75.7		67.0	
2.0 Å	76.3	0.581	76.2	0.592	65.5	0.635	59.3	0.658
Random I¶, 3.0 Å	46.1		43.9		39.8		39.9	
3.0–2.0 Å	90.0		90.6		75.1		68.5	
2.0 Å	76.4	0.571	76.1	0.590	64.1	0.658	59.6	0.661
Random II¶, 3.0 Å	46.1		40.4		36.3		35.2	
3.0–2.0 Å	90.0		89.7		73.8		63.5	
2.0 Å	76.4	0.602	74.3	0.615	62.2	0.681	54.7	0.704
Phase refinement at 1.9 Å and extension to 1.5 Å								
MIR, 1.9 Å	58.9		56.6		51.0		47.0	
1.9–1.5 Å	90.0		90.0		73.3		58.5	
1.5 Å	75.1	0.627	74.0	0.634	62.6	0.716	53.0	0.763
Random I, 1.9 Å	58.8		55.5		48.1		41.7	
1.9–1.5 Å	90.0		89.3		75.2		52.6	
1.5 Å	75.0	0.600	73.0	0.640	62.2	0.732	47.3	0.793
Random II, 1.9 Å	58.8		44.8		36.0		27.9	
1.9–1.5 Å	90.0		88.9		73.6		39.4	
1.5 Å	75.0	0.672	67.8	0.711	55.5	0.807	33.9	0.875

† Initial = initial phases or maps before density modification, SF = solvent flattening, 1D = histogram matching on density, 2D = histogram matching on density and gradient. ‡ Phases errors are shown for the original phases, the extended phases and the overall phases. The phase errors were calculated by comparing with the correct phases derived from the atomic coordinates. The error for the extended phases in the initial data set is also calculated for comparison. It equals 90°, since there is no phase information available. § The map correlation coefficients are shown for maps calculated from all the phases to the extended resolution. It is compared with the correct map derived from the atomic coordinates. ¶ 'Random I' = the phases with randomly generated errors and the mean FOM scheme, 'random II' = the phases with randomly generated errors and the individual FOM scheme.

lated from the T_6 insulin model are fairly accurate because it was refined at high resolution (1.5 Å) and contains all water molecules.

Phase refinement and phase extension for all three test cases have been performed on 1.9 Å (extended to 1.5 Å) and 3.0 Å (extended to 2.0 Å) data, respectively, using solvent flattening, 1D and 2D histogram-matching methods as implemented in the *SQUASH* program (Zhang, 1993). The results are summarized in Table 1.

3.2. Phase refinement at 3.0 Å and extension to 2.0 Å

For the MIR case, the original phases (∞ to 3.0 Å) were improved slightly after applying 2D histogram matching compared with 1D histogram matching, whereas the extended phases (3.0–2.0 Å) were improved on average by 8.7°. The 2D histogram-matching method decreased the overall phase error by an additional 6.2° over the improvement observed with 1D histogram matching. For phases with randomly generated errors using the individual FOM scheme, the mean phase error between ∞ and 3.0 Å was slightly improved with 2D histogram matching compared with 1D histogram matching, whereas the extended phases were improved by 10.2°. This resulted in an overall phase improvement of 7.5°. As expected, the randomly generated phases with both FOM schemes can be refined and extended more accurately compared with MIR phases. In general, there is a slight improvement in the original phases with 2D histogram matching *versus* 1D histogram matching. However, there is a significant improvement in the extended phases, resulting in a significant improvement of the overall phases and map quality.

3.3. Phase refinement at 1.9 Å and extension to 1.5 Å

When applied to higher resolution data (1.9 Å), the 2D histogram-matching method improved the phases more than at medium resolutions. For the MIR case, the original phases between ∞ and 1.9 Å were improved by 4.0° after applying 2D histogram matching compared with 1D histogram matching, whereas the extended phases (1.9–1.5 Å) were improved by 14.8°. The improvement of the overall quality of the map after 2D histogram matching is shown in Fig. 3, where a residue-by-residue map correlation coefficient was calculated for both the initial map and the map after 2D histogram matching. The map quality for each residue has been uniformly improved.

The improvements are much greater for phases with randomly generated errors. For the individual FOM scheme case (random II in Table 1) in particular, the mean phase error between ∞ and 1.9 Å was improved by an additional 8.1° after applying 2D histogram matching compared with 1D histogram matching. Moreover, the extended phases improved by an additional 34.2°. The breakdown of phase error *versus* resolution and the comparison of phase improvement by solvent flattening and 1D and 2D histogram-matching methods are shown in Fig. 4. Solvent flattening improved the original 1.9 Å phases significantly; however, the extended phases between 1.9–1.5 Å were almost random. The addition of 1D histogram matching reduced the error in the original phases and significantly improved the extended phases. The application of 2D histogram matching not only further improved the original phases but also greatly reduced the error in the extended phases over solvent flattening alone or solvent flattening followed by 1D histogram matching.

A region of the electron density around the asparagine residue 21 in the C chain is shown in Fig. 5(*a*). In the initial

map, there is no electron density for the C^{α} atom and most of the side chain except for the $O^{\delta1}$ and $N^{\delta2}$ atoms. After solvent flattening, there is some density corresponding to the C^{α} atom. The addition of 1D histogram matching further improved the side-chain density around the C^{α} and $O^{\delta1}$ and $N^{\delta2}$ atoms; however, there are still gaps in the side-chain density. The 2D histogram matching brought out the density corresponding to the entire side chain of the asparagine residue.

Another region of the electron density around an anti-parallel β-strand in the dimer interface is shown in Fig. 5(*b*). There are many places where significant improvements in the map were made through the successive application of solvent flattening, 1D and 2D histogram matching. For example, the main-chain break between Gly123 and Phe124 in the initial map was closed after solvent flattening and 2D histogram matching, whereas the gap was still present in the map after solvent flattening and 1D histogram matching. The false connection between Phe85 and Phe124 in the adjacent β-strands was gradually broken. The missing side-chain density for Ile2 in the initial map appeared gradually after the successive application of solvent flattening, 1D and 2D histogram matching. The density corresponding to the carbonyl atoms was much better defined after solvent flattening and 2D histogram matching than in the initial map. The main-chain connectivity and atomic boundaries were more clearly defined after applying 2D matching. These improvements would have greatly facilitated chain tracing and model building.

4. Discussion

4.1. The 2D histogram-matching method is more effective in phase improvement than the 1D histogram-matching method

It can be seen from the test cases that employing extra constraints based on the density gradients can further reduce the phase errors and improve the overall map quality. For phase refinement and extension to high resolution, the results showed that 2D histogram matching improves the phases more than 1D histogram matching. The phase improvement for the refinement and extension of MIR phases from 1.9 to 1.5 Å was 9.6°. However, the difference between 2D and 1D histogram-matching methods decreases at medium resolution. The phase improvement for the refinement and extension of MIR phases from 3.0 to 2.0 Å was 6.2°.

While the results using the MIR phases of T_6 insulin provided a typical example of phase refinement and extension, the results from the phases with random errors gave us an upper limit of phase improvement when no systematic errors are present in the MIR phases. When tested on phases with randomly generated errors, the 2D histogram-matching method improved the 1.9–1.5 Å phases by 34.2° compared with the 1D histogram-matching method. This result demonstrates the importance of eliminating systematic errors during the phase-estimation process.

Comparing the phase improvements of the MIR phases and the phases with random phase error and mean FOM scheme at both medium and high resolution gave us an estimate of the systematic error components in the MIR phases. The differences in phase improvement between the MIR and random I

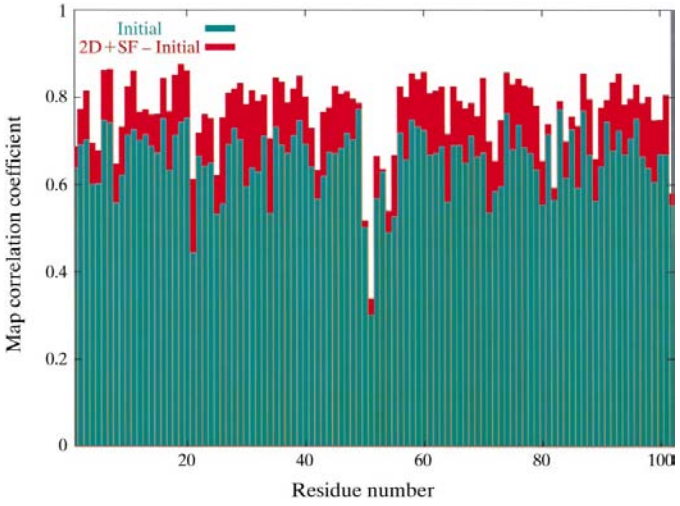

Figure 3
Map correlation coefficients as a function of residue number. The residue-by-residue map correlation coefficients were calculated between the initial MIR map at 1.9 Å and the correct map at 1.5 Å from the atomic coordinates of T_6 insulin (Baker *et al.*, 1988). The residue-by-residue map correlation coefficients were also calculated between the correct map at 1.5 Å and the density-modified map after phase refinement at 1.9 Å and extension to 1.5 Å using solvent flattening and 2D histogram matching. The cyan bars represent the residue-by-residue correlation coefficients between the initial map and the correct map. The red bars represent the improvement of residue correlation coefficients in the density-modified map over the initial map. The map quality was improved for all residues after solvent flattening and 2D histogram matching.

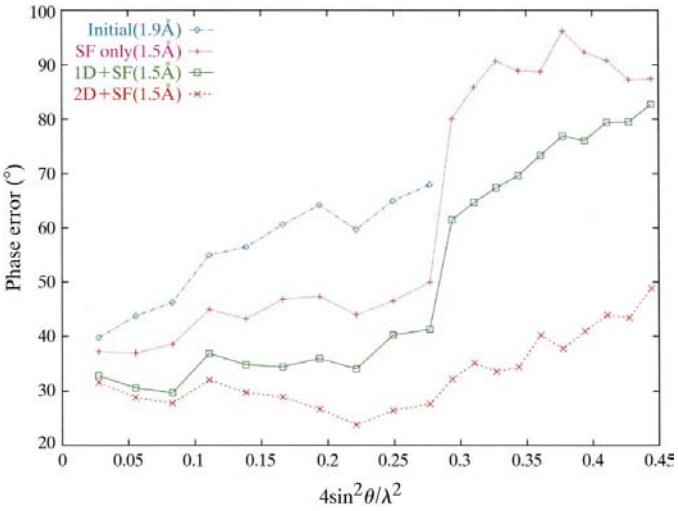

Figure 4
Phase-error distribution as a function of resolution. The 1.9 Å calculated phases with randomly generated phase error and individual FOM scheme were refined and extended to 1.5 Å by solvent flattening, solvent flattening plus 1D histogram matching and solvent flattening plus 2D histogram matching. The refined and extended phases using the above density-modification methods were compared with the correct phases and the mean phase error was calculated. The color and symbol schemes used to represent the phase error distributions are: cyan diamonds, initial phases; magenta plus signs, phases after solvent flattening; green squares, phases after solvent flattening and 1D histogram matching; red crosses, phases after solvent flattening and 2D histogram matching.

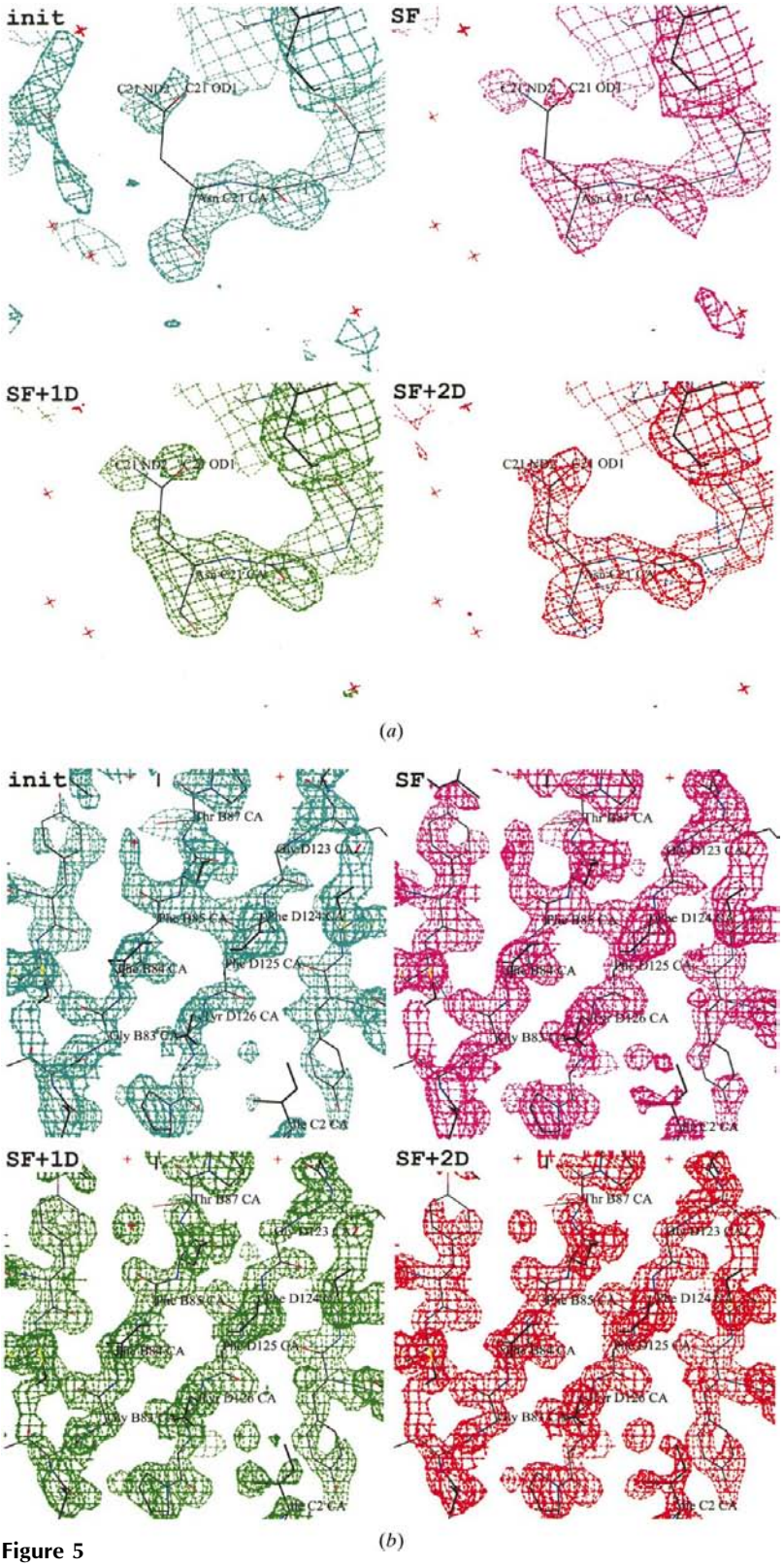

Figure 5
Comparison of electron-density maps in two regions, (*a*) around the asparagine residue 21 in the C chain and (*b*) near the dimer interface between two β-strands, residues 83–87 in the *B* chain and residues 123–126 in the *D* chain. Four maps were represented in each panel: INIT, initial map (cyan); SF, map after solvent flattening (magenta); SF+1D, map after solvent flattening and 1D histogram matching (green); SF+2D, map after solvent flattening and 2D histogram matching (red). The initial map was calculated using phases with randomly generated errors and individual FOM scheme. The maps were all contoured at 1σ.

phases were much smaller at medium resolution than at high resolution. This suggests that the medium-resolution MIR phases contain less systematic error than the high-resolution MIR phases. There could be potentially even greater phase improvement using the 2D histogram-matching method on MIR phases at high resolution if the systematic components of the phase error were reduced.

The results from the two FOM schemes used in the randomly generated phase cases showed the importance of accurate phase-error estimation in the density-modification techniques. There were significant improvements for the original and extended phases at both medium and high resolutions when better FOMs were available. Most current density-modification methods tend to overestimate the FOM as the number of density-modification cycles increase (Cowtan & Main, 1996). The overestimated FOMs tend to weaken the impact of density-modification techniques. Therefore, good phase combination algorithms which accurately estimate the phase errors are as important as searching for new useful constraints (Abrahams, 1997).

4.2. The 2D histogram encodes more stereochemical information of the correct structures than the 1D histogram

The 2D histogram specifies not only the probability of the electron-density value for a given grid point in the map but also the local environment around that grid point, as reflected by the density gradients. The constraint in the 2D histogram is more specific and therefore has higher phasing power than the 1D histogram. Xiang & Carter (1996) have also found the increased sensitivity to phase error in the multi-dimensional histogram of the density value, gradient magnitude and Laplacian of the density compared with the 1D histogram. They suggested that the multi-dimensional histogram could be an improved target for density modification. Our result has demonstrated that the increased sensitivity to phase error, at least in the 2D histogram of the density and gradient, can be translated into an improved density-modification method. The potential use of higher dimensional histograms for phase improvement remains to be explored.

4.3. The decoupling of the electron-density order by the 2D histogram-matching procedure

2D histogram matching reduces the density correlation which is inherent in the 1D histogram-matching process, where the order of the elec-

tron-density values is retained. Two grid points with the same electron-density value will have the same density value after histogram matching. Therefore, the pattern of peaks and troughs in the modified map is similar to that in the original map. This is the result of the implementation of the histogram-matching procedure, where the electron-density order is maintained (Zhang & Main, 1990*a*). However, this feature is undesired, especially when spurious electron densities need to be removed and new electron densities corresponding to missing atoms need to be generated. The coupling of original and modified maps can be broken during the phase-combination step or with the application of additional constraints.

In order to decouple the correlation of electron-density order in the 1D histogram-matching method, Refaat *et al.* (1996) reported a density-modification method based on density histograms which take into account the local environments of the electron-density map, such as the local minimum, maximum and variance of the density. They created ten sub-histograms categorized based on the local environments and carried out histogram matching on each of the ten sub-histograms. Their tests showed significant improvement over the 1D density histogram-matching method when the local minimum, maximum and variance of the density were used.

The 2D histogram-matching method provides an alternative method for decoupling the electron-density order between the modified and original maps through the incorporation of the electron-density gradient distribution. The matching of density gradient in addition to the matching of density values can break up the density correlation that is inherent in 1D density histogram matching. This effect can be corroborated by the fact that the sequential application of 1D histogram matching on density and on gradient produced better results in phase improvement than the parallel application. This is because the sequential application is more effective in breaking up the density correlation than the parallel application.

4.4. Computational efficiency of the 2D histogram-matching procedure and its potential extension to higher dimension histograms

The method used to achieve 2D histogram matching is computationally efficient. Our approach makes direct use of the density gradients. The density-gradient maps, the corresponding structure factors and the density maps can be efficiently transformed back and forth through fast Fourier transform (FFT) techniques. More importantly, the 2D histogram matching can be efficiently achieved by applying 1D matching on density and density gradients sequentially and iteratively, as shown in Fig. 2 (see also §2).

In our earlier work, the 2D histogram of electron density and the gradient modulus was examined and was found to be independent of structural conformation and sensitive to phase error (Goldstein & Zhang, 1998). During the implementation of the 2D histogram-matching algorithm, we found that it is computationally much more efficient to use the three Cartesian components of the electron-density gradient instead of its modulus. This is because the three Cartesian components of the density gradient can be calculated from and converted to the electron density using fast Fourier transforms. We found that the 2D histogram of the electron density and one of the three Cartesian components of the density gradient is also independent of structural conformation and sensitive to phase error.

The strategy of using alternating 1D histogram matching to achieve 2D histogram matching can be generalized to the matching of higher dimensional histograms. By exploring histograms of density values and its higher order derivatives, it may be possible to obtain a density-modification method with further increased phase power and enhanced effectiveness in phase refinement and extension.

We would like to thank Drs Peter Main, Kevin Cowtan, Barry Stoddard and Roland Strong for stimulating discussions and critical reading of the manuscript. This work was supported by a grant from the National Institutes of Health (R29GM55663) to KYJZ.

References

Abrahams, J. P. (1997). *Acta Cryst.* D**53**, 371–376.
Baker, E. N., Blundell, T. L., Cutfield, J. F., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Hodgkin, D. M., Hubbard, R. E., Isaacs, N. W., Reynolds, C. D., Sakabe, N. & Vijayan, M. (1988). *Philos. Trans. R. Soc. London Ser. B*, **319**(1195), 369–456.
Cowtan, K. D. & Main, P. (1996). *Acta Cryst.* D**52**, 43–48.
Cowtan, K. D., Zhang, K. Y. J. & Main, P. (1999). *Progress in Biophysics and Molecular Biology*, edited by T. Blundell. In the press. London: Elsevier Science.
Eriksson, A. E., Cousens, L. S. & Matthews, B. W. (1993). *Protein Sci.* **2**(8), 1274–1284.
Goldstein, A. & Zhang, K. Y. J. (1998). *Acta Cryst.* D**54**, 1230–1244.
Harrison, R. W. (1988). *J. Appl. Cryst.* **21**, 949–952.
Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.
Lunin, V. Y. (1988). *Acta Cryst.* A**44**, 144–150.
Lunin, V. Y., Urzhumtsev, A. G. & Skovoroda, T. P. (1990). *Acta Cryst.* A**46**, 540–544.
Perutz, M. F. (1956). *Acta Cryst.* **9**, 867–873.
Podjarny, A. D., Rees, B. & Urzhumtsev, A. G. (1996). *Methods Mol. Biol.* **56**, 205–226.
Refaat, L. S., Tate, C. & Woolfson, M. M. (1996). *Acta Cryst.* D**52**, 252–256.
Sayre, D. (1974). *Acta Cryst.* A**30**, 180–184.
Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–113.
Xiang, S. & Carter, C. W. (1996). *Acta Cryst.* D**52**, 49–56.
Zhang, K. Y. J. (1989). PhD thesis, University of York, UK.
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., Cowtan, K. D. & Main, P. (1999). *International Tables for Crystallography*, edited by M. G. Rossmann & E. Arnold, Vol. F. Dordrecht, Boston, London: Kluwer. In the press.
Zhang, K. Y. J. & Main, P. (1990*a*). *Acta Cryst.* A**46**, 41–46.
Zhang, K. Y. J. & Main, P. (1990*b*). *Acta Cryst.* A**46**, 377–381.