Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Recursive direct phasing with reference-beam diffraction

Qun Shen[a]* and Jun Wang[b]†

[a]Cornell High Energy Synchrotron Souce (CHESS) and Department of Materials Science and Engineering, Cornell University, USA, and [b]Beijing Synchrotron Radiation Facility, Institute of High Energy Physics, People's Republic of China

† Present address: Department of Chemistry and Chemical Biology, Cornell University, USA.

Correspondence e-mail: qs11@cornell.edu

Received 23 October 2002
Accepted 14 February 2003

The reference-beam diffraction technique provides a practical way to measure a large number of triplet phases in a standard oscillating-crystal diffraction experiment for protein crystals. The triplet-phase data set from such reference-beam measurements contains a unique phase-ocurrence pattern that leads to a new recursive phasing algorithm for the individual structure-factor phases. Application of the new algorithm is demonstrated for tetragonal lysozyme using 7360 triplet phases measured in a reference-beam experiment with a median phase discrepancy of 45°. An electron-density map obtained using this phasing algorithm and the measured triplet phases shows good agreement with the known protein structure.

1. Introduction

It has been demonstrated that crystallographic triplet phases among the structure factors can be obtained directly from experimental measurements of three-beam interference profiles on protein crystals (Chang *et al.*, 1991; Chang, Chao *et al.*, 1999; Weckert *et al.*, 1993, 1999; Weckert & Hummer, 1997; Shen, 1999*a*; Shen *et al.*, 2000*a*,*b*, 2001; Mo *et al.*, 2002). Because no anomalous diffraction signals are necessary, this evolving method promises to provide the phase information needed to solve a biological crystal structure without the requirement for heavy atoms to be incorporated into a native protein structure.

Given the fact that conventional heavy-atom-based methods for phasing protein structures generally require the measurement of many tens of thousands of Bragg reflection intensities, it is worthwhile to ask the following questions regarding the three-beam diffraction approach. (*a*) What is the best systematic approach to make use of the measured triplet phases? (*b*) What is the minimum number of triplet phases that need to be measured in order to solve a protein structure? (*c*) How do the measurement errors affect the possibility of a structural solution? These questions are closely related to each other and ultimately their answers depend on the measurement and phasing methods used for triplet-phase data collection and analysis.

Several authors have discussed these questions recently in the framework of the conventional ψ-scan three-beam measurement technique (Weckert & Hummer, 1997), in which the interference profiles are obtained one at a time. Hölzer *et al.* (2000) have shown that if a sufficient number of triplet phases are measured with substantial overlaps among the individual structure-factor phases in the measured triplet relations, it is then possible to deduce the individual phases using a phasing tree, much like the traditional convergence-map technique employed in the direct-methods approach. Because of the considerable number of unmeasured triplet

phases in the phasing tree, it is necessary to adopt a multi-solution entropy-maximization algorithm (Hölzer *et al.*, 2000) to deduce all the individual phases and obtain an electron-density map. Recently, Wang *et al.* (2001) employed a similar maximum-entropy method to solve a small-molecule structure based on several dozens of measured triplet phases. Measured triplet phases can also be applied in a traditional or a *Shake-and-Bake* direct-methods approach to replace the mathematical estimates with the measured values, leading to structural solutions of small proteins with lower than atomic resolution intensity data (Mo *et al.*, 1996; Weeks *et al.*, 2000).

In this article, we discuss a phasing strategy in the context of a recently developed reference-beam diffraction (RBD) experimental geometry for triplet-phase measurements (Shen, 1998, 1999*a*; Chang, Chao *et al.*, 1999). We show that the new geometry offers a significant advantage in initial phasing compared with the conventional ψ-scan technique (Weckert & Hummer, 1997) owing to a systematic occurrence pattern of reflection triplets that is unique to the reference-beam arrangement. The unique pattern leads to a new recursive phasing algorithm that allows a straightforward determination of all individual structure-factor phases from as few as four starting single phases. This would be particularly useful as a tool in the solution of novel structures.

2. Data-collection method

Fig. 1 illustrates the triplet-phase data-collection geometry using the reference-beam diffraction technique, which is a slight modification (Shen, 1998, 1999*a*) of the conventional oscillation camera arrangement commonly used in protein crystallography. Instead of being perpendicular to the incident X-ray beam $\mathbf{k}_0$, the oscillation axis in the RBD geometry is tilted by the Bragg angle θ_G of a strong reference reflection **G**, which is aligned parallel to the oscillation axis ψ. In this way, reflection **G** can be kept fully excited throughout the crystal oscillation, creating a reference beam $\mathbf{k}_G$ for the oscillation diffraction image. In principle, the intensities of all Bragg reflections recorded on an area detector during such an oscillation can be influenced by the interference with the **G**-reflected reference wave and thus are sensitive to the relative phases of the reflections involved.

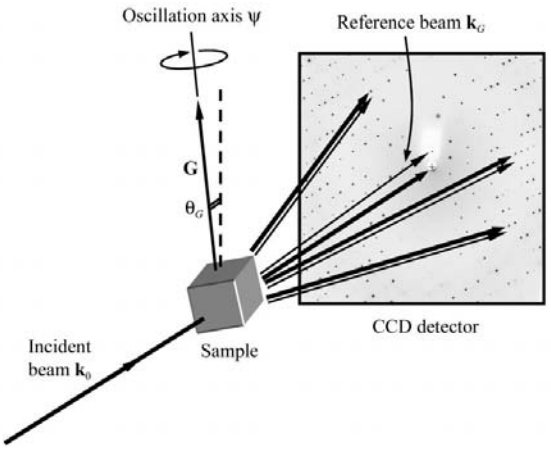

Figure 1
Schematic illustration of the reference-beam diffraction data-collection method for direct measurements of triplet phases in an oscillation camera crystallography experiment.

Specifically, two sets of diffraction images, one (thick lines in Fig. 1) created by the direct beam $\mathbf{k}_0$ and the other (thin lines in Fig. 1) by the reference beam $\mathbf{k}_G$, are superimposed and interfere with each other. For any Bragg reflection **H**, its direct-beam-excited diffracted wave is $\mathbf{k}_H = \mathbf{k}_0 + \mathbf{H}$ and the reference-beam-excited diffracted wave is $\mathbf{k}_G + (\mathbf{H} - \mathbf{G})$, which is parallel to $\mathbf{k}_0 + \mathbf{H}$ since $\mathbf{k}_G = \mathbf{k}_0 + \mathbf{G}$. It is straightforward to show that the phase difference between the two diffracted waves is the triplet phase

$$\delta = \alpha_G + \alpha_{H-G} - \alpha_H, \qquad (1)$$

where α_H is the structure-factor phase of reflection **H**. It is this phase difference δ that is measured in the reference-beam diffraction experiment.

In order to avoid the necessity of making absolute intensity measurements, complete reference-beam interference profiles are usually obtained by taking multiple exposures at several angular settings θ_i–θ_G on the **G**-reflection rocking curve. In this procedure, the interference profile for each Bragg reflection **H**, as illustrated in Fig. 2, is composed of the integrated intensities $I_H(\theta_i)$ from the oscillation image taken at θ_i. This data-collection procedure closely resembles the multiple-wavelength anomalous diffraction method, with the **G**-reflection rocking curve playing the role of an atomic absorption edge without the actual need for heavy atoms.

3. Triplet-phase data set

A 90° reference-beam diffraction data set has been collected from a $P4_32_12$ lysozyme crystal at the C1 bending-magnet station of the Cornell High Energy Synchrotron Source (CHESS) using the experimental procedure described in the previous section, with X-ray wavelength λ = 1.097 Å. A recently implemented special five-circle κ-diffractometer (Pringle & Shen, 2003) with a 1K × 1K CCD area detector is used in the RBD experiment, along with a control algorithm to align the reference reflection **G** = (111) and to obtain RBD oscillation images (Shen *et al.*, 2001). The complete data set, taken at room temperature, consists of 45 series of $\Delta\psi = 2°$ oscillation images at 19 θ-angle positions across the **G**-reflection rocking curve, with 15 s exposure time for each image. The diffraction resolution is ∼2.5 Å, which is limited by the size of the detector, with 86% completeness for the whole data set. A typical θ range is about 0.05– 0.1°, depending on the crystal mosaicity.

The integrated intensities of all recorded Bragg reflections in the data set are deduced using *MOSFLM* and *SCALA* from the *CCP*4 package (Collaborative Computational Project, Number 4, 1994). These intensities are then sorted according to the θ angle at which the original image is taken to form a set of RBD profiles for all 14 914 reflections in the data set. The total data-collection time for all these profiles is about 12 h, with only ∼6 h of exposure of the specimen to X-rays.

The RBD profiles are then analysed using a curve-fitting procedure (Shen *et al.*, 2001) with an interference function based on a phase-sensitive diffraction theory in a distorted-wave approximation (Shen, 1999*b*, 2000; Shen & Huang, 2001). The fitting function,

$$\frac{I_H(\theta)}{I_0} = 1 - p\sin\delta\frac{\sin^2\Delta\theta}{\Delta\theta^2} + \frac{p}{\Delta\theta}\left(\cos\delta + \frac{p}{2\Delta\theta}\right)\left[1 - \frac{\sin(2\Delta\theta)}{2\Delta\theta}\right], \qquad (2)$$

where $\Delta\theta = (\theta - \theta_G)/w$, with w being the rocking-curve width, involves four adjustable parameters: background intensity I_0, center θ_G of the **G** reflection, amplitude p of the RBD interference and the triplet phase δ. Examples of the fits to the experimental data are shown in Fig. 2, along with the corresponding inverse-beam $\overline{\mathbf{H}}/\overline{\mathbf{G}}$ measurements that allow enantiomorph specification and more accurate triplet-phase determinations (Chang, Chao *et al.*, 1999; Shen *et al.*, 2000*a*).

As pointed out by several authors (Weckert *et al.*, 1993; Weckert & Hummer, 1997; Chang *et al.*, 1999), the inverse three-beam measurement provides a good way to separate the phase-sensitive interference effect from a phase-independent intensity contribution arising from the overall energy-flow balance in a mosaic crystal. In practice, both qualitative (Weckert & Hummer, 1997) and quantitative (Chang, Stetsko *et al.*, 1999) analytical methods have been used to extract triplet phases from a pair of inverse-beam related three-beam cases. In our work, we have adopted the following simple quantitative curve-fitting procedure to take into account both the direct-beam **H**/**G** and the inverse-beam $\overline{\mathbf{H}}/\overline{\mathbf{G}}$ cases. This procedure is based on Weckert & Hummer (1997), but is developed in the framework of an automated curve-fitting routine using (2). As illustrated in Fig. 2, we first perform the same curve-fitting for both cases in each pair, yielding amplitudes p and $\overline{p}$ and triplet-phases δ and $\overline{\delta}$ for each case. We note that although in most cases both δ and $\overline{\delta}$ are consistent with each other, *i.e.* $\delta \simeq -\overline{\delta}$, occasionally this is not true because of the phase-independent intensity contribution. The problem is easily solved by choosing the triplet phase corresponding to the largest amplitude case in the inverse-beam pair and then assigning the phase of the other case to the negative of the chosen one. For example, if $p > \overline{p}$, then $\delta = \delta$ and $\overline{\delta} = -\delta$, but if $\overline{p} > p$, then $\overline{\delta} = \overline{\delta}$ and $\delta = -\overline{\delta}$. We find that this procedure works very well, as shown in Fig. 2, where the chosen phase is indicated by the boxed text in each pair of the **H**/**G** and $\overline{\mathbf{H}}/\overline{\mathbf{G}}$ cases.

Although we treat the entire data set as if every recorded reflection yields an RBD interference effect, this is only one step in our overall procedure to obtain the measured triplet phases in an automated fashion. After the curve-fitting is completed, we use an acceptance criterion based on the goodness-of-fit parameters such as χ^2 and standard deviation σ_δ to select a subset of triplet phases that are more reliable (Shen *et al.*, 2000*b*). This procedure is illustrated in Fig. 3, where the error histogram of the measured triplet phases in the data set is shown compared with the

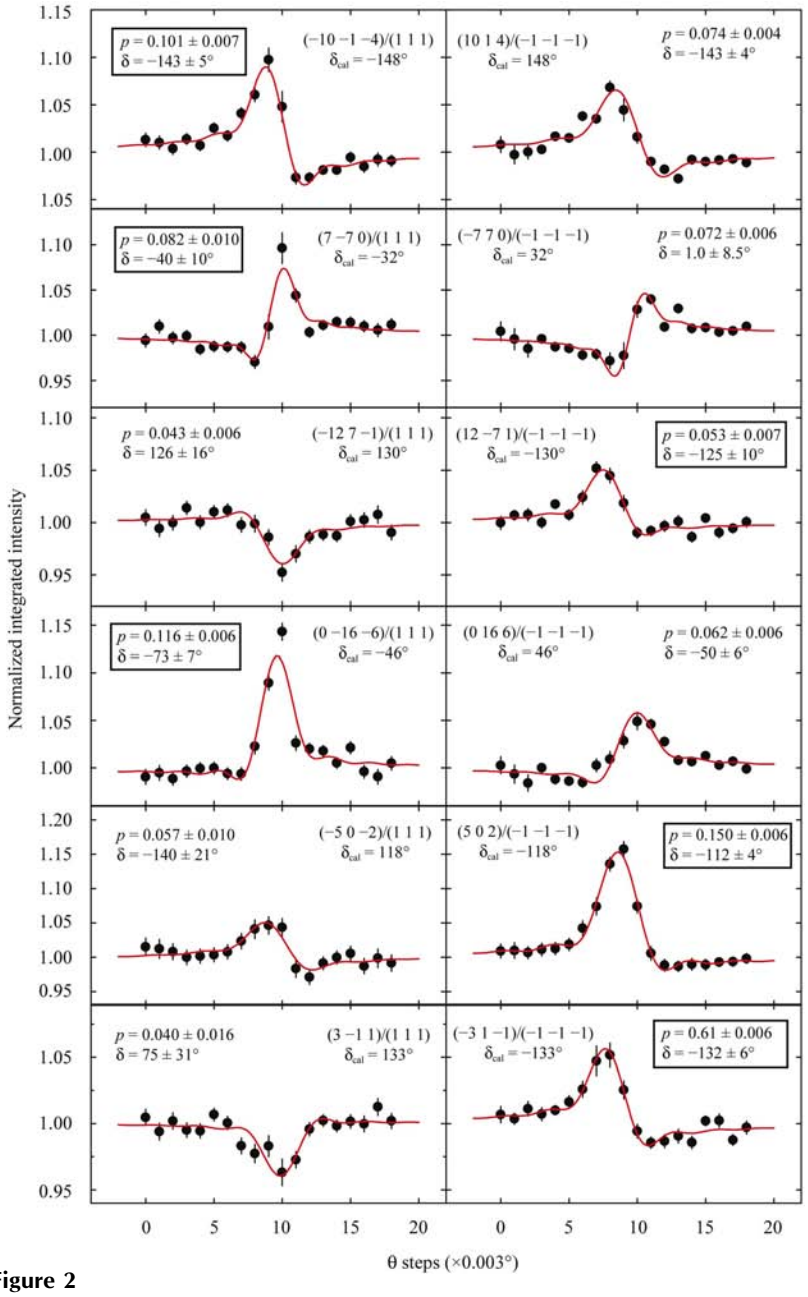

Figure 2
Examples of reference-beam diffraction interference profiles obtained from a tetragonal lysozyme crystal. Each row shows a pair of inverse-beam related three-beam cases, **H**/**G** (left) and $\overline{\mathbf{H}}/\overline{\mathbf{G}}$ (right). The solid curves are fits to the data using (2), with four adjustable parameters. The final value for the triplet phases $\pm\delta$ of each pair is determined by the fit result with the larger amplitude p. See text for more details.

calculated values based on the Protein Data Bank entry 193l (Vaney *et al.*, 1996). When all N = 14 914 reflections are included, the median triplet-phase discrepancy is $\Delta\delta_m = 61^\circ$. This discrepancy is reduced substantially to $\Delta\delta_m = 45^\circ$ if a subset of $N = 7360$ reflections are selected with goodness-of-fit parameters $\chi^2 < 7$, $\sigma_\delta < 55^\circ$ and $\theta_G \neq$ limits.

The relatively large median phase discrepancy in our RBD measurements may be a consequence of several factors. Firstly, compared with the conventional ψ-scan method, intensity statistics may be poorer for weaker reflections because of the same exposure time being used for both strong and weak reflections as set in the oscillation data collection. Secondly, there may be accidental multiple-beam effects arising from quartets or higher orders since all measurements are made at a fixed X-ray wavelength. Finally, some of the phase discrepancies observed may be real since the phases of low-order reflections such as (111) may be affected by solvent which is not taken into account in the calculated phases.

Nevertheless, as demonstrated by Weeks *et al.* (2000), a mean triplet-phase error of 40–50° may be sufficient for solving small protein structures when used in conjunction with a direct-methods based algorithm such as *SnB*. Thus, the data set we have obtained may be good enough to lead to a structural solution. The key next step is to deduce the individual structure-factor phases α_H involved in all triplet relations (1) from the measured triplet-phase data set.

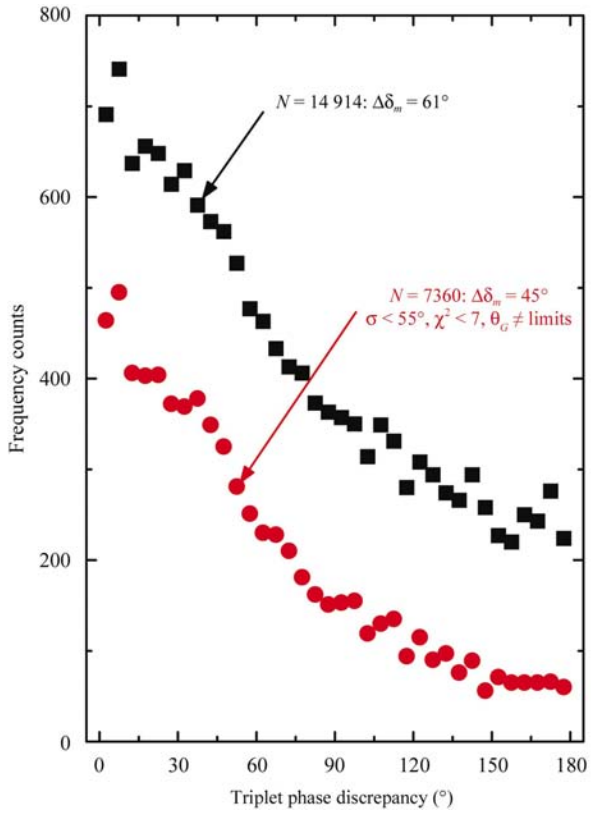

Figure 3
Error histogram of the measured triplet phases compared with the calculated values based on PDB entry 193l. It is shown by selecting a subset of phases from the whole data set using goodness-of-fit criteria that a random background in the error distribution can be minimized and the median phase error $\Delta\delta_m$ can be reduced from 61° for the whole data set to 45° for the subset.

4. Recursive phasing

In addition to direct-methods (Mo *et al.*, 1996; Weeks *et al.*, 2000) and maximum-entropy (Hölzer *et al.*, 2000; Wang *et al.*, 2001) based approaches, other strategies of how to proceed from measured triplet phases to individual structure-factor phases are being explored, as certain specific experimental features may be taken into account in these alternative approaches.

One possibility is to take advantage of a triplet-occurrence pattern that is unique to the RBD geometry. As illustrated in Fig. 4, for any reflection **H** recorded on an RBD image, **H** + **G** is its adjacent reflection next to **H** along direction **G**. It is obvious that reflections **G**, **H** and **H** + **G** form a triplet. Thus, if triplet phases along a single column **H** + n**G** (n = ..., −2, −1, 0, 1, 2, ...) are all measured, then a simple recursive method can be devised to deduce all individual structure-factor phases from the measured triplets, once a single phase in that column is known. In other words, if m triplet phases are measured in a single column, there are only $m + 1$ individual phases plus the G-reflection phase α_G (which is common to all columns in the data set) associated with all m triplet phases. Thus, each additional triplet along a column adds only one unknown individual phase. This situation is dramatically different from the conventional ψ-scanning three-beam technique where each additional three-beam case would generally introduce two unknown individual phases.

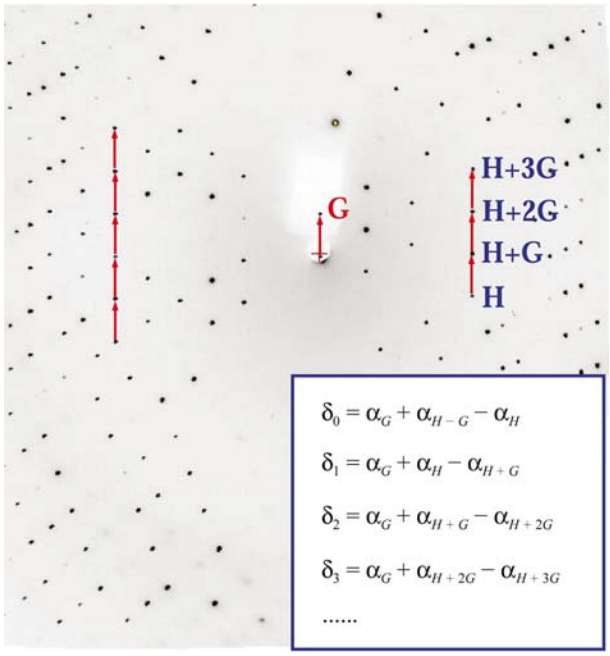

Figure 4
Illustration of the triplet-occurrence pattern that is unique in RBD geometry, *i.e.* the three-beam condition is satisfied for every adjacent reflections along any column in the direction of the reference reflection **G**.

We have developed a new recursive-phasing algorithm based on the unique triplet-occurrence pattern mentioned above. The new algorithm has been implemented in the program *RBD_phasing* and its flow chart is shown in Fig. 5. A preliminary test of the new program has been performed using the measured triplet-phase data set obtained from the tetragonal lysozyme crystal and a set of initial known phases. The initial phases are chosen to be the 191 (*hk*0) reflections in the data set since the phases of these reflections are restricted to either 0 or π owing to the symmetry requirements of the space group $P4_32_12$.

With the known (*hk*0) phases plus the (111) phase taken from PDB entry 193l (Vaney *et al.*, 1996), it is possible to use the algorithm *RBD_phasing* and to deduce new structure-factor phases from the measured subset of 7360 triplet phases. The median phase error for these new individual phases is 66°, which is reasonable based on the 45° median error in triplet phases. An electron-density map (Fig. 6*a*) is then calculated based on the structure-factor phases from the 7360-reflection RBD data set. For comparison, in Fig. 6(*b*) we show the same map obtained using the calculated triplet phases for the same 7360 reflections. The two maps at $z = 0$ are in reasonable agreement, as can be seen both from Fig. 6 and from the map-correlation coefficient of 0.697, showing the feasibility of the new recursive phasing algorithm. We believe that further reductions in the individual phase errors and corresponding improvements in the electron-density map are feasible using density modification and other standard crystallographic refinement techniques, but this topic is not the focus of this article.

Since a relatively large number of known phases are used in the present work, it may be more appropriate to view the new recursive phasing procedure as a phase extension. However, it is entirely possible to substantially reduce the number of initial phases that are needed in the recursive RBD algorithm. For example, if three RBD data sets are measured with non-coplanar $\mathbf{G}_1$, $\mathbf{G}_2$ and $\mathbf{G}_3$ as the reference reflections, then in principle only four initial phases, those of $\mathbf{G}_1$, $\mathbf{G}_2$ and $\mathbf{G}_3$ plus a single reflection $\mathbf{H}_0$, are needed to phase the whole structure.

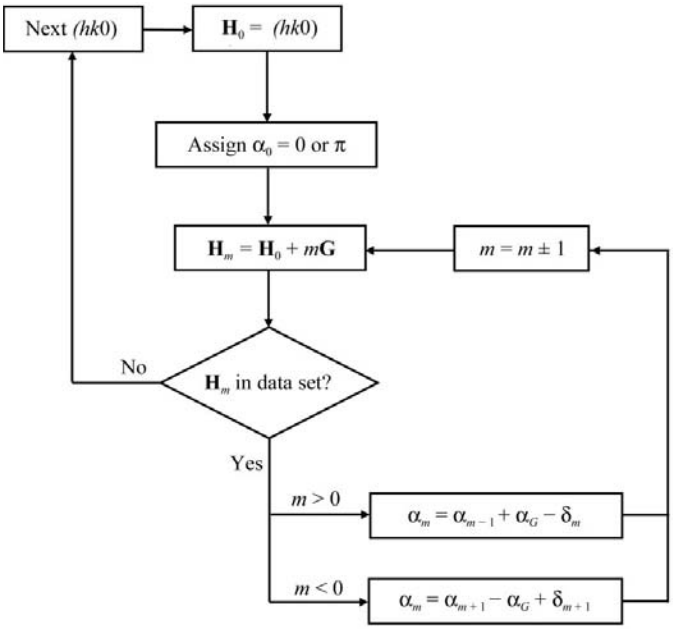

Figure 5
New recursive phasing algorithm from a measured triplet-phase data set for tetragonal lysozyme, based on the unique triplet-occurrence pattern in reference-beam diffraction geometry.

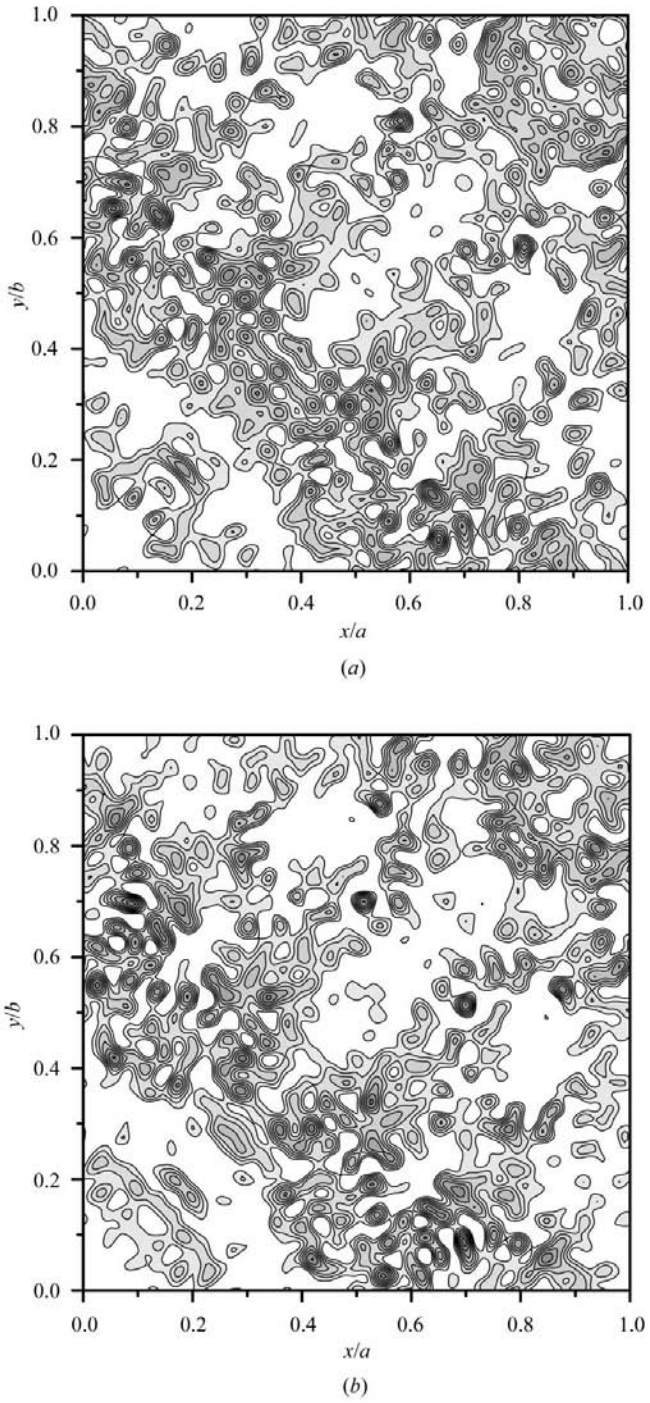

Figure 6
(*a*) Electron-density map of the $z = 0$ basal plane of tetragonal lysozyme using the 7360 measured triplet phases and the new recursive phasing algorithm. (*b*) The same map using all calculated phases from PDB for the 7360 reflections in the data set. Contour levels are from 0.1 to 0.8 in both maps, with a linear increment of 0.1.

This idea is schematically shown in Fig. 7, where the phases of all reflections in reciprocal space are progressively determined from a single point $\mathbf{H}_0$ to a line of nodes through $\mathbf{G}_1$, then to a plane of nodes through $\mathbf{G}_2$ and finally to a volume of nodes through $\mathbf{G}_3$ data sets. It is worth noting that the $\mathbf{G}_1$ and the $\mathbf{G}_2$ data sets do not need to be complete for the overall method to work. Further reduction of the number of initial phases may also be possible if origin-defining and symmetry-related reflections are taken into account in a systematic manner.

5. Concluding remarks

In summary, we have developed a new recursive phasing algorithm to deduce the individual structure-factor phases from a triplet-phase data set measured using the reference-beam diffraction technique. A preliminary test of the algorithm on a 7360 triplet-phase data set from tetragonal lysozyme has yielded a reasonable electron-density map that is in good agreement with the map based on the calculated phases. The new algorithm makes use of a triplet-occurrence pattern that is unique to the reference-beam geometry and is not present in conventional three-beam experiments. The unique triplet pattern offers a substantial advantage in providing a systematic and definitive way to obtain overlaps among the individual structure-factor phases within a triplet-phase measurement data set.

Future work in this area is likely to focus on reduction of the number of initial phases that are necessary to start the recursive process in the phasing algorithm and on proper treatment of error propagation in the recursion owing to poorly or inaccurately measured triplet phases in the data set. It may be possible to include the unique triplet-occurrence pattern in direct-methods or maximum-entropy based programs to increase the likelihood of structural solutions with a smaller number of measured triplet phases. It is hoped that our work will also stimulate more discussions in the future on the optimal strategies of three-beam experiments in protein crystallography.

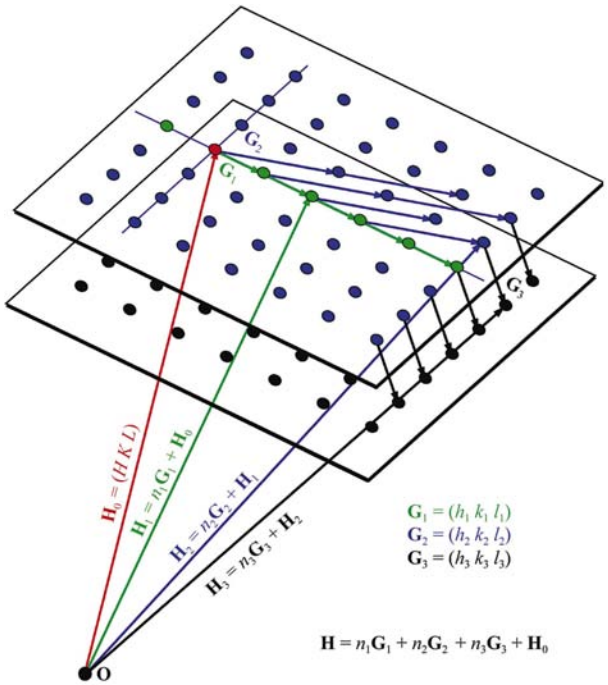

Figure 7
Illustration in reciprocal space to show that three sets of reference-beam measurements with $\mathbf{G}_1$, $\mathbf{G}_2$, $\mathbf{G}_3$ as the non-coplanar reference reflections would in principle allow a systematic deduction of all individual phases from a single starting phase $\mathbf{H}_0 = (HKL)$ and the three reference reflection phases that define the origin.

We would like to thank Rob Thorne, Steve Ealick, Sol Gruner, Quan Hao, Ken Finkelstein, Marian Szebenyi, Chris Heaton, Bill Miller and others at Cornell, and Herb Hauptman, Chuck Weeks, Bob Blessing, Jimmy Xu and George DeTitta at Hauptman–Woodward Institute (HWI) for their useful discussions and help during the various stages of this work. QS acknowledges the warm hospitality he received during his stay at Beijing Synchrotron Radiation Facility, where part of the algorithmic development was performed. This work has been supported by NSF Grant DMR 97-13424 through CHESS, by NIH Grant GM-46733 through HWI and by a grant from the Wang Kuancheng Education Foundation of the Chinese Academy of Sciences in China.

References

Chang, S. L., Chao, C. H., Huang, Y. S., Jean, Y. C., Sheu, H. S., Liang, F. J., Chien, H. C., Chen, C. K. & Yuan, H. S. (1999). *Acta Cryst.* A**55**, 933–938.
Chang, S. L., King, H. E., Huang, M. T. & Gao, Y. (1991). *Phys. Rev. Lett.* **67**, 3113–3116.
Chang, S. L., Stetsko, Y. P., Huang, Y. S., Chao, C. H., Liang, F. J. & Chen, C. K. (1999). *Phys. Lett. A*, **264**, 328–333.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Hölzer, K., Weckert, E. & Schroer, K. (2000). *Acta Cryst.* D**56**, 322–327.
Mo, F., Mathiesen, R. H., Alzari, P. M., Lescar, J. & Rasmussen, B. (2002). *Acta Cryst.* D**58**, 1780–1786.
Mo, F., Mathiesen, R. H., Hauback, B. C. & Adman, E. T. (1996). *Acta Cryst.* D**52**, 893–900.
Pringle, D. & Shen, Q. (2003). *J. Appl. Cryst.* **36**, 29–33.
Shen, Q. (1998). *Phys. Rev. Lett.* **80**, 3268–3271.
Shen, Q. (1999*a*). *Phys. Rev. B*, **59**, 11109–11112.
Shen, Q. (1999*b*). *Phys. Rev. Lett.* **83**, 4784–4787.
Shen, Q. (2000). *Phys. Rev. B*, **61**, 8593–8597.
Shen, Q. & Huang, X. R. (2001). *Phys. Rev. B*, **63**, 174102.
Shen, Q., Kycia, S. & Dobrianov, I. (2000*a*). *Acta Cryst.* A**56**, 264–267.
Shen, Q., Kycia, S. & Dobrianov, I. (2000*b*). *Acta Cryst.* A**56**, 268–279.
Shen, Q., Pringle, D., Szebenyi, M. & Wang, J. (2001). *Rev. Sci. Instrum.* **73**, 1646–1648.
Vaney, M. C., Maignan, S., Riès-Kautt, M. & Ducruix, A. (1996). *Acta Cryst.* D**52**, 505–517.
Wang, C.-M., Chao, C.-H. & Chang, S.-L. (2001). *Acta Cryst.* A**57**, 420–428.
Weckert, E., Hölzer, K., Schroer, K., Zellner, J. & Hummer, K. (1999). *Acta Cryst.* D**55**, 1320–1328.
Weckert, E. & Hummer, K. (1997). *Acta Cryst.* A**53**, 108–143.
Weckert, E., Schwegle, W. & Hummer, K. (1993). *Proc. R. Soc. London Ser. A*, **442**, 33–46.
Weeks, C. M., Xu, H., Hauptman, H. A. & Shen, Q. (2000). *Acta Cryst.* A**56**, 280–283.