*Acta Cryst.* (1995). D**51**, 13–20

# A Method for Processing Diffraction Data from Twinned Protein Crystals and its Application in the Structure Determination of an FAD/NADH-Binding Fragment of Nitrate Reductase

BY GUOGUANG LU, YLVA LINDQVIST AND GUNTER SCHNEIDER

*Swedish University of Agricultural Science, Department of Molecular Biology, Uppsala Biomedical Center, Box* 590, 751 24 *Uppsala, Sweden*

(*Received* 27 *April* 1994; *accepted* 25 *July* 1994)

## Abstract

A general method to deconvolute oscillation data sets from twinned protein crystals to a corresponding single crystal data set has been developed and applied to diffraction data measured from crystals of a fragment containing the FAD- and NADH-binding domains of nitrate reductase. The procedure allows straightforward processing of diffraction data from twinned crystals. Typically, $R_{merge}$ values of reduced data sets from the nitrate reductase crystals after deconvolution are about 0.06 compared to 0.13 and higher before deconvolution. Based on these deconvoluted data sets, the structure of the FAD- and NADH-binding domains of nitrate reductase could be solved successfully. The result indicates that crystal twinning does not necessarily prevent crystallographic structure determination.

## Introduction

It is not unusual that crystallization of biological macromolecules results in twin crystals. So far, crystal twinning has been considered an often insurmountable obstacle for a successful structure determination and only in a few cases the three-dimensional structure of proteins has been determined with data from twinned crystals. For instance, the structure of B-phycoerythrin (Fisher & Sweet, 1980) and muconate lactonizing enzyme (Goldman, Ollis & Steitz, 1987) was solved from merohedral twinned crystals by deconvoluting completely overlapped diffraction data sets. However, the more common type of twin crystals, where each crystal consists of two perfect lattices and reflections from these lattices do not overlap completely, is not used in structure determination. Instead, screening for untwinned crystal forms has been and is the generally chosen response. In a number of cases, in spite of large amounts of recombinant protein the search for single crystals has failed. We encountered this problem in the crystal structure determination of a C-terminal fragment of nitrate reductase, comprising the FAD/NADH-binding domains (FDNR) of the enzyme. To solve this problem, we have developed a simple method to deconvolute the diffraction data set from this type of twinned crystals by directly measuring the non-overlapping reflections on oscillation images. This procedure was necessary to successfully determine the structure of FDNR. In the following, we describe the principles of the method and its application to oscillation films and imaging-plate data.

## General description of the method

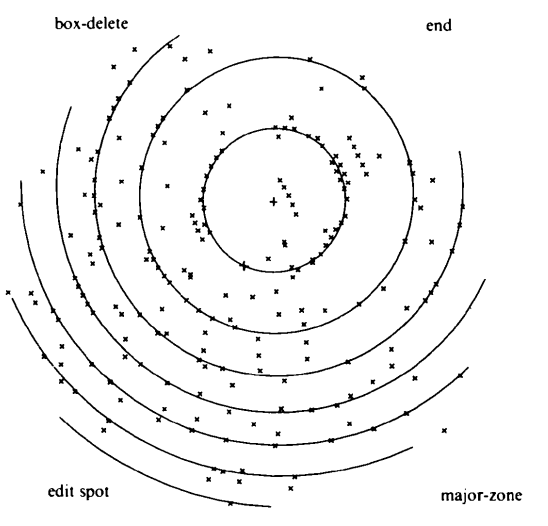
When a major zone axis of a single crystal is close to the beam direction, reciprocal lattice points on the Ewald sphere will give rise to circles in the diffraction pattern with the same center on an oscillation image. Reflections on each circle are from one reciprocal lattice plane. Therefore, on an image from a twinned crystal, the lattice of one crystal can be easily recognized by identifying its corresponding circles (Fig. 1*a*). If the spots from these circles are removed from the list of reflections, the residual spots must be from the other single crystal. An interactive graphics program has been developed to identify reflections belonging to one lattice, thus generating a spot list which only consists of spots from one of the lattices (Fig. 1*b*). This peak list can then be used to determine the orientation of this lattice using existing auto-indexing methods (Kabsch, 1988). In the same way, the orientation of the other lattice can also be determined. When both orientations have been determined all the spots on the oscillation images can be predicted and indexed (Fig. 1*c*). Overlapping reflections from the two lattices which cannot be measured reliably can be predicted and removed from the list of reflections. Overlap is defined if the distance between two spots, belonging to the two different crystal lattices is shorter than a certain distance. In principle, the overlapping reflections could also be integrated if the ratio of the two crystal lattices is known by scaling non-overlapping data from the two lattices to each other.

## Deconvolution of oscillation film

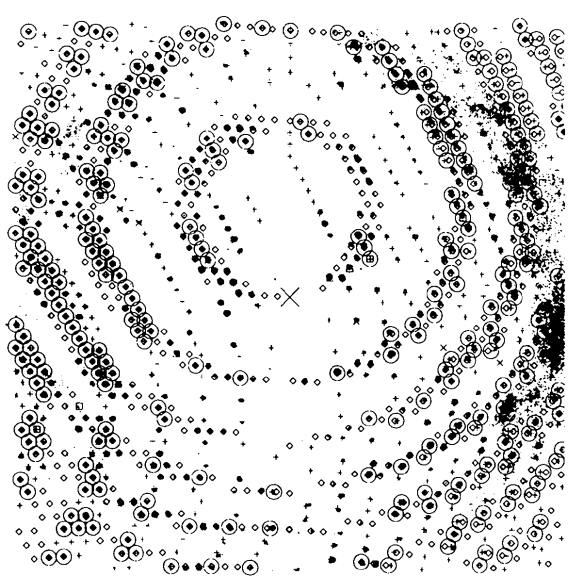
The program package *MOSCO/MOSFLM* version 3.6 (Leslie, 1991) in combination with programs developed by us to solve the twinning problem according to the method described above was used for data processing of oscillation films. A flow scheme of the procedure is shown in Fig. 2 and is as follows: *STILL* is used to search the peaks from the oscillation images. An

interactive program *EDITSPOT* enables the user to flag spots of one lattice as described above. The program also generates circles for calculated major zones in order to help users recognizing the spots from the different lattices (Fig. 1*b*). The flagged spots are used to determine the orientation of this lattice with the auto-indexing program *REFIX* (Kabsch, 1988). When one orientation is determined, the program *OSCGEN* is used to predict the diffraction pattern from this lattice. The program *WASH* removes the predicted peaks from one lattice in the peak list and the residual spots can be used to predict the second orientation with *REFIX*. Iteration of this procedure might be necessary if the initial orientations are not good enough for data evaluation in later steps of the procedure. After the two orientations have been obtained and refined, the spot coordinates of oscillation films are predicted by the program *OSCGEN*. The program *DETWIN* searches for overlapping spots by checking whether the distance between two spots from the two crystals is less than a certain distance. In this work we chose the cutoff distance 0.9–1.1 mm (twice the spot size). If the distances between spots are found to be too short, both spots are removed from the predicted spot list. When the reflections from one lattice are integrated, a large mosaicity (we used more than twice the real mosaicity) is used for predicting the reflections from the other lattice in order to compensate for small orientation errors thus making sure that all the overlaps are removed. The intensities of the predicted spots are then measured with the program *MOSFLM*. Postrefinement of both orientations with *POSTCHK* is based on the information from partial reflections between adjacent oscillation films (Jones, Bartels & Schwager, 1977). The refined orientations are used to repeat the procedure until the $R_{\text{merge}}$ value converges.

(*a*)




(*b*)



(*c*)

Fig. 1. Editing diffraction images from a twinned crystal. The spots can be sorted into two lattices when a major zone axis from one lattice is roughly parallel to the beam. (*a*) An oscillation image of a typical twinned crystal of FDNR. The major zone of one lattice can be easily recognized. (*b*) The program *EDITSPOT* enables removal of the spots belonging to one lattice in a diffraction image. Crosses represent peak positions in a still or oscillation image. When major zones from one lattice are identified, the program generates simulated circles according to the given cell dimensions and the zone center in order to remove all peaks which possibly belong to this lattice. The remaining spots can be used for auto-indexing. (*c*) Oscillation image [same as in (*a*)] superimposed with the prediction for both lattices. Boxes represent reflections from one lattice and crosses from the second lattice. Circles identify those reflections for which the separation distances of the spots from the two lattices are too short for reliable measurements.

In FDNR, it was found that the relation between the two lattices in the twinned crystals always was the same. The program *TWINED* was included which allows to calculate the orientation of the second lattice from the orientation of the first lattice and the known relation between the two lattices. This was particularly helpful when reflections from one of the lattices were weak and its orientation was difficult to determine.

The orientations of the two lattices can be determined by the auto-indexing programs *REFIX* or *AUTOINDEX*, after the peaks from the two lattices were separated with the help of program *EDITSPOT* on still or non-adjacent oscillation images, but they were usually quite inaccurate. Redetermination and refinement of one orientation after removing the predicted spots from the other lattice using the *WASH* program was necessary because the interactive sorting often was not able to completely remove all the spots from one lattice. After this step, the orientations were usually quite accurate and postrefinement was not really necessary provided still images were recorded. In the case where stills were not available, many cycles of postrefinement were needed to obtain sufficiently accurate orientations and, consequently, good data quality.

### Deconvolution of imaging-plate data

The MSC commercial processing software (Sato, Yamamoto, Imada & Katsube, 1992) was modified to include the deconvolution programs described above. The whole procedure follows the same flow scheme as the processing of oscillation films (Fig. 2).

## Application to FDNR

*Crystallization and data collection*

Crystals of FDNR with space group $R3$ and cell dimensions $a = b = 145.2$, $c = 47.5$ Å, $\alpha = \beta = 90$, $\gamma = 120°$ were obtained with sodium citrate as precipitant (Lu, Campbell, Lindqvist & Schneider, 1992). Crystals sufficiently large for data collection are severely twinned. Attempts to define conditions for growth of sufficiently large single crystals have failed and the structure determination was carried out with these twinned crystals. Diffraction data was collected in three ways. A native data set was collected at station 9.6, Synchrotron Radiation Source (SRS) at Daresbury, England, on photographic films. The oscillation angle of each film was 1.7°. The crystals diffracted to about 2.5 Å resolution.

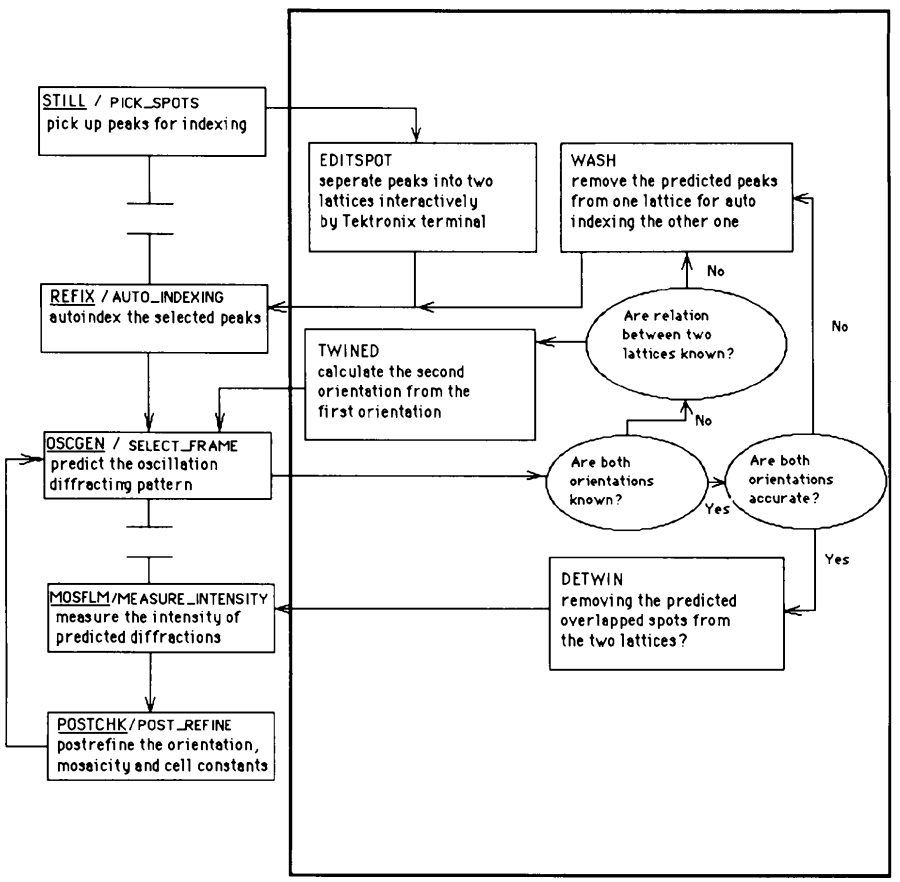



Fig. 2. A flow diagram for the deconvolution and evaluation software system for diffraction data from twinned crystals. The program works with both oscillation films and R-axis imaging-plate images. The underlined text represents the program name from *MOSCO* software while the text after the slashes represents that from MSC commercial software. Programs contained in the large box on the right-hand side of the figure represent the part of the procedure which is different from single-crystal data evaluation.

Table 1. *Statistics of diffraction data sets of FDNR*

| Name | $R_{merge}$ Lattice 1 | $R_{merge}$ Lattice 2 | $R_{merge}$ All | $R_{merge}$ without deconvolution | Average multiplicity | Completeness (%) | Resolution (Å) |
|---|---|---|---|---|---|---|---|
| Native | 0.075 | 0.079 | 0.072 | 0.167 | 2.7 | 82 | 2.8 |
| Native (SRS)* | | | 0.059 | 0.136 | 4.5 | 60 | 2.5 |
| Merged native | | | 0.063 | | 5.5 | 81 | 2.5 |
| ADP† | 0.068 | 0.071 | 0.071 | 0.158 | 2.5 | 68 | 3.0 |
| $KAu(CN)_2$ | 0.067 | 0.073 | 0.066 | 0.152 | 2.5 | 65 | 2.7 |

* This data set was collected with oscillation films at the Synchrotron Radiation Source in Daresbury and processed by the *MOSCO* software. Three crystals with a total of six lattices were processed with internal $R_{merge}$ values of 0.041, 0.044, 0.051, 0.053, 0.063 and 0.064. The other data sets were collected on an *R*-axis II image-plate area detector and processed by *R*-axis II commercial software (Sato *et al.*, 1992). One crystal with two lattices was used in each case. $R_{merge}$ values of each lattice are included in the table. The $R_{merge}$ values of lattice 1 and 2 were calculated as $\sum_{h,i}|F(h)_i - F(h)_{mean}|/\sum|F(h)_{mean}|$ by *R*-axis II commercial software, where *i* represents the individual reflections. The $R_{merge}$ values of all the lattices were calculated as $\sum_{h,i}|I(h)_i - I(h)_{mean}|/\sum|I(h)_{mean}|$ by the program *AGROVATA* in the *CCP*4 package, where *i* represents the lattices or sources.

† Crystal of nitrate reductase cocrystallized with ADP.

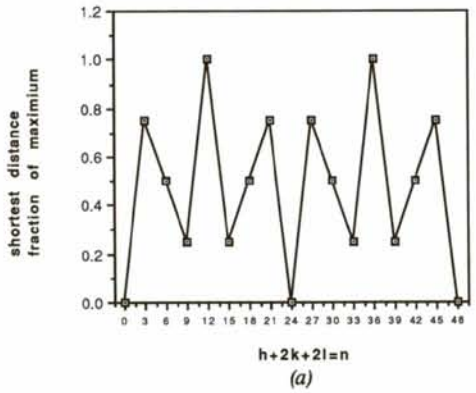
Data sets of both native and heavy-atom derivative crystals (Table 1) were also collected using an R-axis II imaging plate mounted on a Rigaku rotating-anode X-ray source. Both still and oscillation images were collected in this case. In all data-collection runs, the crystals were orientated in such a way that the *c* axis is approximately along the beam at the start and about perpendicular to the beam in the end of data collection. Fig. 1(*a*) shows the diffraction pattern of a twinned crystal from one oscillation image, in which a major zone from one lattice can be recognized.

Initially, diffraction data sets were also collected on a Xentronix area detector. The data-processing software *Buddha* (Blum, Metcalf, Harrison & Wiley, 1987) was able to index one of the lattices in the twinned crystals and process the data, yet without removing the overlaps. Typically, this resulted in $R_{merge}$ values of 8.0–11.0% for one crystal and of 12–13% between crystals with about 10% of the data rejected to 3.2 Å resolution.

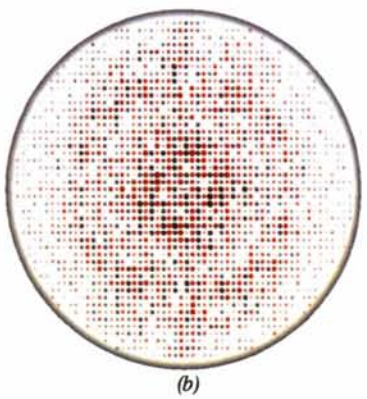
*Data processing*

The $R_{merge}$ values for the different data sets and data-processing routines are summarized in Table 1. After deconvolution of the data, the $R_{merge}$ values of a native crystal drops from 8–12% to 4–6% and the $R_{merge}$ values between crystals dropped from 13 to 5.9%, while the rejection rate decreased from 10 to 1–2%. The $R_{merge}$ value between the native data sets measured at SRS and on an R-axis imaging plate is 0.063 without rejection at 2.8 Å. This is to be compared to the $R_{merge}$ value between non-deconvoluted data sets from different crystals which is around 13% or higher. These results indicate that the deconvolution procedure significantly improved the data quality from the twinned FDNR crystals.

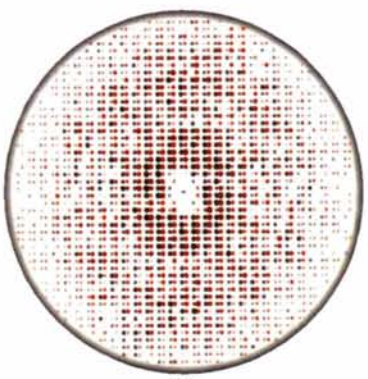
In FDNR, the orientation between the two lattices can be described as a 180° rotation about the real-space vector (122) *i.e.* the $\mathbf{a}+2\mathbf{b}+2\mathbf{c}$ axis in a hexagonal lattice system [or its threefold symmetric axis $(\bar{2}\bar{1}2)$ or $(1\bar{1}2)$ depending how the crystals are indexed]. This axis is perpendicular to the reciprocal lattice sections $h+2k+2l = n$ (where $n = 0, \pm3, \pm6 \ldots$). Therefore, all the shortest distances between each point to the points of the second overlapping lattice, the 'separation distance', are identical in each section *n* (see Fig. 3*b*). This means that this separation distance is a function of the section number n. Analysis of the diffraction data with a precession-simulating program *PATTERN* (Lu, unpublished work) showed that, due to the special cell dimensions of these crystals, *i.e.* $(c^*/a^*)^2 = 7.0$, the reciprocal points (0, ±21, ±3) in sections $h+2k+2l = \pm48$ are almost exactly on the twinning axis in reciprocal space. So are the points (0, ±42, ±6) in the sections $h+2k+2l = \pm96$ and so on. According to how lattice points relate to the twinning axis, this twinning phenomenon makes the two overlapping lattices form a 'super lattice' which has a positional repetition of all lattice points with a period of 16 sections of $h+2k+2l$. For example, the overlapping reciprocal lattice patterns are exactly the same from sections $h+2k+2l = n$ when $n = 0, \pm48, \pm96 \ldots$. So also when $n = 3, 51, 99 \ldots$, and so on. (Due to one of the reflection conditions in this space group, it must be that $n = 3N$ where $N = 0, \pm1, \pm2, \ldots$, so the real section period is 16 and not 48). The separation distance of each point, therefore, also has a period according to the section number $h+2k+2l = n$ (see Fig. 3*a*), however this period is only half of the one above, *i.e.* eight sections ($n = 24$). (See Figs. 3*b*–3*h* for detailed explanation). For example, all the shortest distances of reciprocal points of one lattice to the other are the same on sections $h+2k+2l = n$ when $n = 0, \pm24, \pm48 \ldots$, the same is true for $n = 3, 27, 51 \ldots$, and so on. As shown in Figs. 3(*a*) and 3(*b*), these distances can be sorted into five groups, according to the fraction of the maximum separation distance of all the sections, which is equal to $1/89.9\ Å^{-1}$ in this crystal form, the 'separation fraction'. Within a section period of $n < 24$ the 'separation fraction' can only be 0, *i.e* completely overlap, when $n = 0$. It will be 0.25 when $n = 9$ and $n = 15$, 0.5 when $n = 6$ and $n = 18$, 0.75 when $n = 3$ and $n = 21$, or 1.0, *i.e.* maximum separation, when $n = 12$.



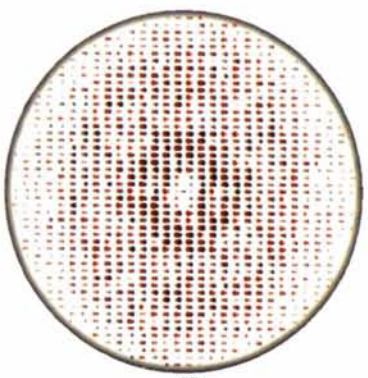

(a)



(b)

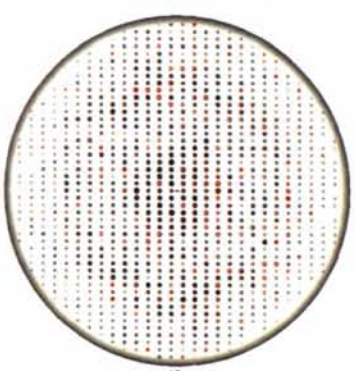


(c)

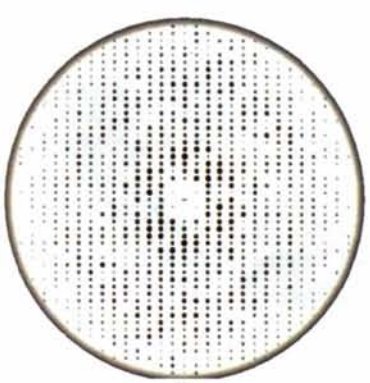


(d)

(e)



(f)



(g)

(h)

Fig. 3. An analysis of twinned crystals of FDNR (*a*) The shortest distance between spots from the two lattices is shown as a function of reciprocal lattice section number $n$, where $n = h + 2k + 2l$. It has a repetition of period $n = 24$. Data with a separation fraction less than 0.25 were regarded as overlap so that three out of eight collected layers were not used. (*b*)–(*g*) Reciprocal patterns of the two overlapped lattices from some sections of $h + 2k + 2l = n$ which represent different separation fractions (abbreviated as sf here). The twinning twofold axis is perpendicular to these sections and represented by a cross. Reflections from the two lattices are shown in black and red, respectively. The pictures were generated by the program *PATTERN* (Lu, unpublished results). (*b*) $h + 2k + 2l = 12$ with sf = 1.0, the spots have maximum separation in this section. (*c*) $h + 2k + 2l = 3$ with sf = 0.75. (*d*) $h + 2k + 2l = 6$ with sf = 0.5; (*e*) $h + 2k + 2l = 9$ with sf = 0.25; (*f*) $h + 2k + 2l = 24$ with sf = 0, all the points are completely overlapped in this section. (*g*) $h + 2k + 2l = 0$ with sf = 0. (*h*) $h + 2k + 2l = 48$ with sf = 0. The reflection (0,21,3) which is represented by the cross is exactly on the twinning axis. This section has a positional repetition with sections $n = 0$, $n = 96$ and so on which indicates that the section repetition period is $n = 48$ (16 sections). It does not have a positional repetition with section $n = 24$ but they both have sf = 0. The sf repetition period is $n = 24$ (eight sections).

During data collection, the crystal to camera distance and oscillation angle was chosen in such a way that the reflections with a separation fraction 0.5, 0.75 or 1.0 could be processed without any overlap. The reflection spots with a separation fraction 0.25 or 0 were flagged as overlapping data and removed during processing. Therefore, as shown in Fig. 3(*a*), data from three of every eight sections (37.5%) were not used. Since there are two lattices on the oscillation images, redundant measurements compensated for data loss due to overlap. As shown in Table 1, there is, therefore, no problem in the completeness of the diffraction data in spite of the removal of overlapping reflections.

*Phasing and electron-density maps*

The major binding sites of the heavy atoms could be found from difference Patterson maps and difference Fourier maps calculated by the program *PROTEIN* (Steigemann, 1974) based on data both with and without data deconvolution. In all derivatives, the peak/$\sigma$ ratio was about two times higher in the maps after data deconvolution. However, minor sites were only found from difference Patterson maps and from difference Fourier maps calculated with deconvoluted data while they were completely drowned in noise in the maps from data without deconvolution.

MIR phases were calculated from data both with and without deconvolution of the twinned crystals with the program *MLPHARE* (Otwinowski, 1991) contained in the *CCP*4 program suite (SERC Daresbury Laboratory, 1979). The phasing statistics did not show significant differences between the two data sets except for an improvement of the figure of merit (Table 2). However, the quality of the MIR map calculated from the deconvoluted data sets was dramatically improved. As can be seen in Fig. 4 the boundary of the molecule in the map without data deconvolution was not defined, whereas in the map based on deconvoluted data the molecular envelope is clearly visible. As shown in Fig. 5, the electron-density map calculated with data without deconvolution still contained recognizable features of secondary structure, but it was seriously disturbed. The main chain often showed breaks in electron density and chain tracing was not possible. The MIR map after data deconvolution was clear and after solvent flattening (Wang, 1985; Leslie, 1987) the polypeptide chain could

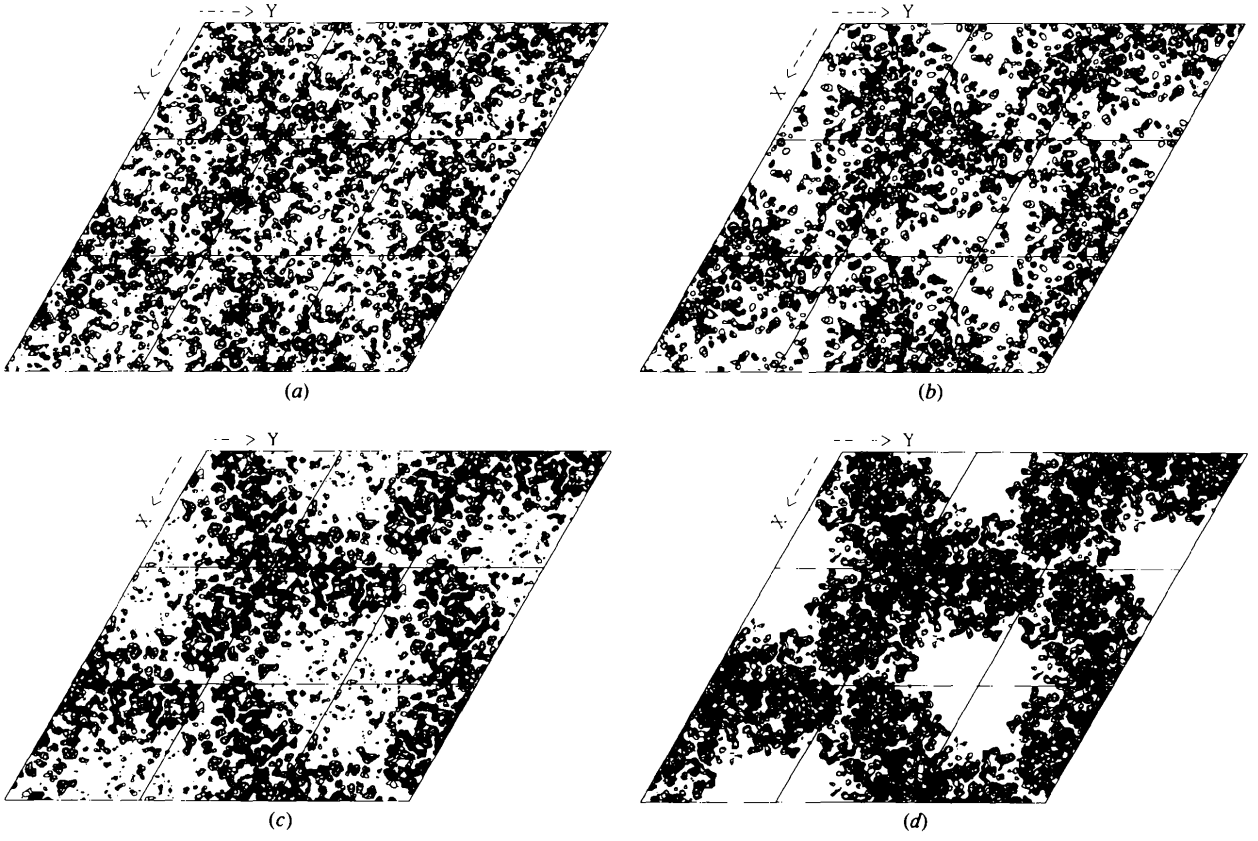



Fig. 4. Electron-density maps at 3.0 Å resolution, plotted as projections along the *c* axis in the whole cell. The box in the figure outlines one asymmetric unit. (*a*) MIR map and (*b*) solvent-flattened map calculated from data without deconvolution. (*c*) MIR map, (*d*) solvent-flattened map from data after deconvolution. All the maps were contoured at 2.0 $\sigma$.

be traced. Model building and refinement of the structure of FDNR has resulted in an $R$-factor value of 0.193 and will be published elsewhere (Lu, Campbell, Schneider & Lindqvist, unpublished results). The packing of the molecules looking down the twofold twinning axis in the crystals is shown in Fig. 6. The orientation of the second axis can be obtained by rotating 180° about this twinning axis. However the position of this axis cannot be determined. Analysis of the crystal packing suggests a plausible way for formation of twinned crystals during growth. The nitrate reductase molecules, shown in red in Fig. 6, could form dimers related by the twofold twinning axis, thus creating nucleation sites for the second lattice at the surface of the crystal. This lattice will grow onto the surface of the original lattice as shown in Fig. 6.

Table 2. *Statistics of phase calculation based on diffraction data before and after deconvolution*

| | After data deconvolution | | Before data deconvolution |
|---|---|---|---|
| Resolution | 2.8 Å | 3.2 Å | 3.2 Å |
| Phasing power | | | |
| $KAu(CN)_2$ | 1.7 | 1.7 | 1.9 |
| PCMB | 1.6 | 1.7 | 2.0 |
| $Pt(NH_3)_2Cl_2$ | 1.5 | 1.5 | 1.0 |
| $(NH_4)_2PtCl_6$* | 0.7 | 0.7 | 1.4 |
| $K_2PtCl_4$ | 0.8 | 0.9 | 1.1 |
| Figure of merit | 0.46 | 0.52 | 0.44 |

* Data to only 3.5 Å available.

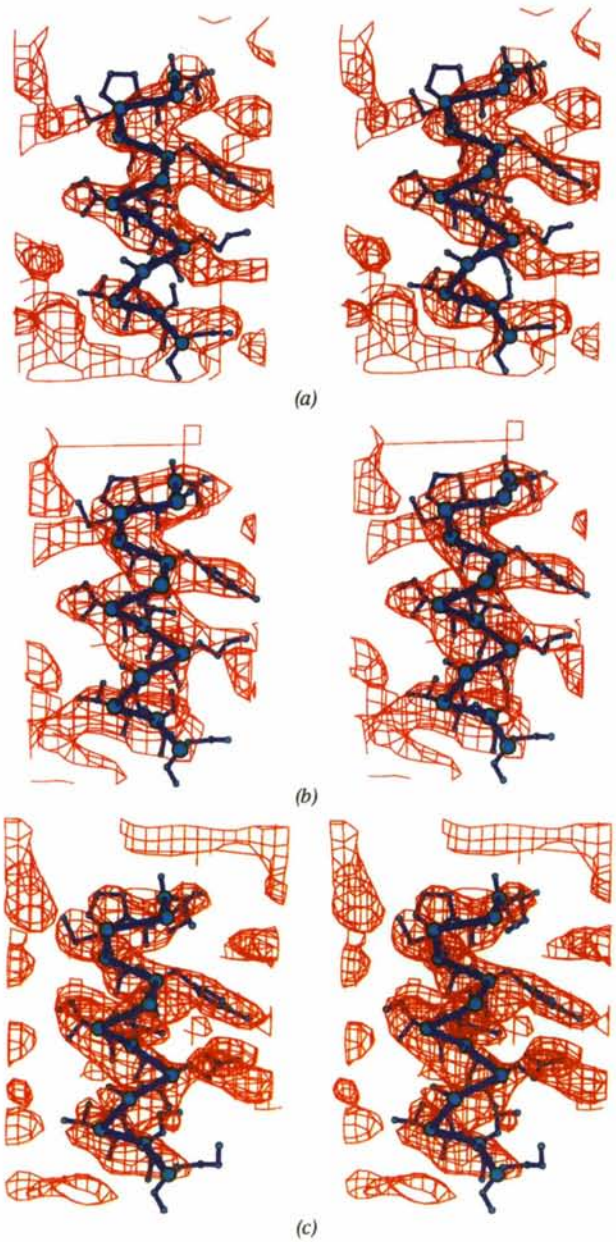



Fig. 5. Stereo picture of parts of the electron-density map (red) of FDNR. The refined model of FDNR (blue) is superimposed. The spheres represent the atomic positions of the model. Thin cylinders represent bonds between atoms and thick cylinders which link atoms, represent the helical main-chain trace. (*a*) MIR from data before deconvolution. (*b*) MIR map after data deconvolution. (*c*) $2F_o - F_c$ map calculated with combined phases from model and MIR phases. All the maps were contoured at 1.1 $\sigma$.

## Concluding remarks

Crystal twinning often occurs in the crystallization of proteins. It could be easily dealt with if data evaluation and processing software were able to automatically

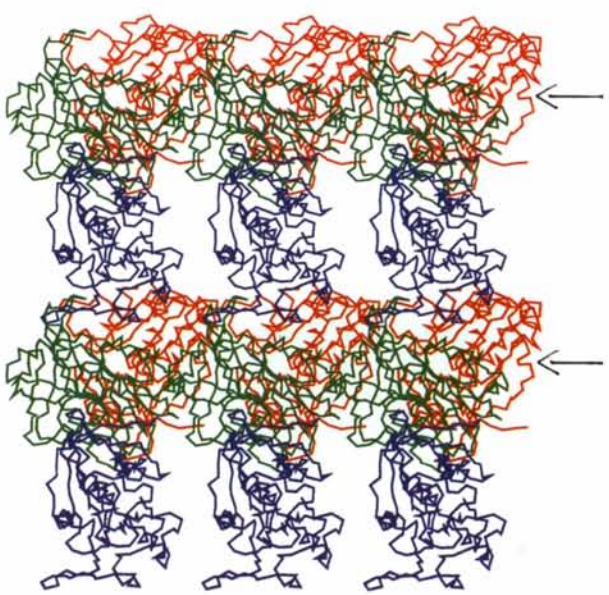


Fig. 6. Packing of the FDNR molecules in the crystals. The molecules which are related by crystallographic threefold symmetry are shown in red, green and blue, respectively. The twofold twinning axis is perpendicular to the plane of the paper. The arrows indicate possible nucleation sites for the second lattice (for details see text).

process these data. The data-convolution method described here provides some means to evaluate diffraction data from twinned crystals in a semi-automatic mode. The method was designed in a general space group and should be useful for other protein crystals with this or a similar type of twinning. Application of this method to FDNR was a necessary requirement for the success of the structure determination.

There are still two problems in this procedure. During refinement of orientation and post-refinement, the cell parameters in the two lattices were treated as independent when they should be constrained to the same value. On an image with a high percentage of overlapping reflections, the number of low-resolution non-overlapping spots sometimes was not enough to refine the film parameters (scale, twist, tilt and so on), because only non-overlapping spots from both of the two crystals should be used for the same film parameter refinement. Implementations of these features in the processing software will allow the use of twinned crystals in protein structure determination on a more routine basis.

We thank Professor W. Campbell, Houghton, for samples of the FAD/NADH-binding fragment of corn nitrate reductase. We thank the staff of the SRS (Daresbury, UK) for access to synchrotron radiation. This work was supported by a grant from the Swedish Agricultural Research Council.

## References

BLUM, M., METCALF, P., HARRISON, S. C. & WILEY, D. C. (1987). *J. Appl. Cryst.* **20**, 235–242.

FISHER, R. & SWEET, R. (1980). *Acta Cryst.* A**36**, 755–760

GOLDMAN, A., OLLIS, D. L & STEITZ, T. A. (1987). *J. Mol. Biol.* **194**, 143–153.

JONES, T. A., BARTELS, K. & SCHWAGER, P. (1977). In *The Rotation Method in Crystallography*, edited by U. W. ARNDT & A. J. WONACOTT. Amsterdam: North Holland.

KABSCH, W. (1988). *J. Appl. Cryst.* **21**, 67–71.

LESLIE, A. G. W. (1987). *Acta Cryst.* A**43**, 134–136.

LESLIE, A. G. W. (1990). *Crystallographic Computing* 5, edited by D. MORAS, A. D. PODJARNAY & J. C. THIERRY, pp. 50–61. Oxford Univ. Press.

LU, G., CAMPBELL, W., LINDQVIST, Y. & SCHNEIDER, G. (1992). *J. Mol. Biol.* **224**, 277–279

OTWINOWSKI, Z. (1991). *MLPHARE, CCP4 Proceedings*, pp. 80–88. Warrington: Daresbury Laboratory.

SATO, M., YAMAMOTO, M., IMADA, K. & KATSUBE, Y. (1992). *J. Appl. Cryst.* **25**, 348–357

SERC Daresbury Laboratory (1979). *CCP4. A Suite of Programs for Protein Crystallography*. SERC Daresbury Laboratory, Warrington WA4 4AD, England.

STEIGEMANN, W. (1974). PhD thesis, Technische Univ. München, Germany.

WANG, B. C. (1985). *Methods Enzymol.* **115**, 90–112.