Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The finer things in X-ray diffraction data collection

J. W. Pflugrath

Molecular Structure Corporation, 9009 New Trails Drive, The Woodlands, TX 77381, USA

Correspondence e-mail: jwp@msc.com

Received 6 May 1999
Accepted 5 July 1999

X-ray diffraction images from two-dimensional position-sensitive detectors can be characterized as thick or thin, depending on whether the rotation-angle increment per image is greater than or less than the crystal mosaicity, respectively. The expectations and consequences of the processing of thick and thin images in terms of spatial overlap, saturated pixels, X-ray background and $I/\sigma(I)$ are discussed. The *d*TREK* software suite for processing diffraction images is briefly introduced, and results from *d*TREK* are compared with those from another popular package.

1. Introduction

Two-dimensional position-sensitive detectors have been used for many years in X-ray diffraction data collection. In particular, data from crystals of macromolecules such as proteins, oligonucleotides and their complexes are almost always acquired with an area detector such as film (now virtually obsolete), a multi-wire system, an imaging plate or the recently commercialized charge-coupled device (CCD) coupled to a phosphor-coated fiber-optic taper. With all these detectors, the crystal, centered in the X-ray beam, is rotated or oscillated around a single axis through a small angle of 0.1 to $\sim$2.0°, while counts from diffracted photons are accumulated for a specified time. At the end of the small rotation, the detector is read out and the counts are stored as an image: a two-dimensional array with each array element (pixel) related to a distinct position on the detector and the number of photons which impinged on that area during the exposure. For our purposes here, a data set consists of one or more scans of a series of individual yet contiguous images created while the crystal is rotated through a larger total angular range, so that the experiment records diffraction from a large volume of reciprocal space.

The characteristics or properties of the images which are acquired include those controlled by the experimenter such as exposure time, the rotation-angle increment and the rotation start, as well as those inherited from the experimental hardware such as the detector (physical size, position, number of pixels *etc.*), the rotation axis and the X-ray source (direction, wavelength, polarization, divergence). This manuscript considers in detail only exposure time and rotation-angle increment and develops some guidelines so that the best images can be collected and processed under any given circumstance. Special emphasis is placed on data sets which contain only partially recorded Bragg reflections on the individual images. That is, the rotation-angle increment is less than or equal to the effective mosaic spread of the crystal, so that the Bragg reflection is finely sampled in three dimensions. Two dimensions result naturally from the two dimensions of the

area detector, while the third dimension is a consequence of the detector read-out (*i.e.* image) for each rotation-angle increment of the crystal.

2. Thick *versus* thin

Those data sets consisting of relatively large rotation-angle increments, where each image contains a number of fully recorded reflections, will be called *thick*; this is defined for our purposes here as an increment greater than the effective mosaic spread of the crystal. Those with relatively small rotation-angle increments, where each image contains predominantly partially recorded reflections, will be called *thin*, finely sliced or fine φ sliced; this is defined here as an increment less than the crystal moscaicity or less than or equal to 0.5°. Therefore, the actual rotation-angle increments which characterize thick and thin images depend on the effective mosaic spread of the crystal. Thus, a data set with images of 1° rotation would be a thick data set if the crystal mosaicity was 0.5°, but would be a thin data set if the crystal mosaicity was 1.5°.

Thick and thin images have a number of intrinsic characteristics which lead to different choices during their processing. A data set of thick images usually has more fully recorded reflections, fewer partially recorded reflections, more spatial overlaps, higher X-ray background, more saturated pixels and a lower total number of images. In general, a two-dimensional integration of reflections in each image is performed with software such as *MOSFLM* or *HKL* (Leslie, 1999; Otwinowski & Minor, 1997). Intensity estimates of reflections recorded partially in adjacent images are routinely summed after the integration step in a post-refinement step. The processing is relatively quick, since a single image is reduced to a list of reflections and only later are the lists from individual images combined together. Because of the thick nature of the images, fewer images are required to cover the overall rotation range.

A thin data set has no fully recorded reflections, fewer spatial overlaps, lower X-ray background, fewer saturated pixels, more time consumed during data collection reading out the detector and a larger total number of images – often three to five times more. A three-dimensional integration of reflections (each reflection spans several adjacent images) is performed with software such as *XDS*, *XENGEN*, *BUDDHA*, *MADNES* or *d*TREK* (Kabsch, 1988; Howard *et al.*, 1987; Blum *et al.*, 1987; Messerschmidt & Pflugrath, 1987). In these packages, individual scale factors are not applied to each portion of a reflection on separate images. Instead, pixels from adjacent images are integrated to create full reflections which are then divided into scaling batches based on consecutive rotation ranges of a few images each. A separate scale factor is used for each batch of reflections. For a different treatment of scaling portions of partially recorded reflections, see Rossmann & Van Beek (1999). It follows that more computational resources (disks, memory, CPU cycles) are required to process thin images than thick ones.

Table 1
Thin and thick data sets.

Method	Thin	Thick
Rotation-angle increment per image (°)	0.25	1.5
Overall rotation (°)	0–45	0–45
Exposure time per image (s)	120	720
Exposure rate (min $^{\circ -1}$)	8	8
Total exposure time (h)	6	6
Number of images	180	30

3. Expectations

Consider thin and thick data sets from a frozen *Serratia marcescens* endonuclease crystal (Miller & Krause, 1996) with a mosaicity of about 0.4° collected with Cu $K\alpha$ radiation (λ = 1.5418 Å) and the same experimental hardware in two different rotation increments: 0.25° per image and 1.5° per image, as shown in Table 1.

These two similar, yet different, data-collection schemes should result in the exact same total number of diffracted X-ray photons, although the detector read noise manifests itself differently in the two experiments. Let us examine some expectations and see if these are validated later by data from an actual experiment.

First, diffraction behaves as a Poissonian process. Thus, if we are counting X-ray photons and not arbitrary analog-to-digital units (ADUs), the variance of an observed quantity of photons Q is simply the quantity of photons: $\mathrm{var}(Q) = Q$. The standard deviation, $\sigma(Q)$, is the square root of the variance:

$$Q = \text{quantity of photons}, \tag{1}$$

$$\mathrm{var}(Q) = Q, \tag{2}$$

$$\sigma(Q) = Q^{1/2}. \tag{3}$$

It follows for a relatively simple integration of a Bragg peak containing diffracted X-ray photons that

$$I = \text{observed intensity}, \tag{4}$$

$$I = \sum(\mathrm{Peak}_i - \mathrm{Background}_i), \tag{5}$$

$$\mathrm{var}(I) = \sum[\mathrm{var}(\mathrm{Peak}_i) + \mathrm{var}(\mathrm{Background}_i)]. \tag{6}$$

That is, the observed intensity is the sum of each pixel in the peak region minus the estimated background for the corresponding peak pixel (5). The background of the peak pixels is estimated from nearby pixels which are not within the peak. Standard error-propagation rules apply, so that the variance of the observed intensity is a function of the variance of the peak pixels and background estimate. Now we introduce an R_{err} ('relative error') term which is simply the standard deviation of a quantity divided by the quantity. This is very similar to the R_{merge} values popular in scaling programs. For the observed intensity,

$$R_{\mathrm{err}} = \sigma(I)/I, \tag{7}$$

$$R_{\mathrm{err}} \cong I^{-1/2} \text{ if } \mathrm{var}(\mathrm{Background}) = 0. \tag{8}$$

In Table 2, the relative error is shown as a function of observed intensity and background values. For background values of

Table 2
Relative errors (%) for hypothetical integrated intensity and background values.

Example: Intensity = 100, Background = 500, therefore Peak = 600 and $R_{\rm err} = \sigma(I)/I = (600 + 500)^{1/2}/100 = 33.2\%$.

Integrated intensity (photons)	Background (photons)		
	0	500	1000
100	10.0	33.2	45.8
500	4.5	7.7	10.0
1000	3.3	4.5	5.6
10000	1.0	1.0	1.1

500 and 1000, (6) and (2) are used to estimate $\sigma(I)$ and not (8), since in real experiments the variance of the background is *not* 0.

From the table, it is clear that lower X-ray backgrounds and higher peak intensities give the lowest relative errors. If the crystal described in Table 1 has a mosaicity of about 0.4°, the X-ray background superimposed directly on a Bragg peak contained within two images of the thin-sliced data set (column 1 in Table 1) will result from 4 min of X-ray exposure. The same Bragg peak collected in thick-sliced mode (column 2) will have background from 12 min (Bragg peak contained in one image, *i.e.* full) or even 24 min (Bragg peak partial on two adjacent images) of X-ray exposure. Thus, the X-ray background is three to six times higher for the Bragg reflections in the thickly sliced data set. One could conclude that thinly sliced images, with their inherently lower X-ray background, should yield better data. This is indeed the case, but only if there are no additional sources of error when going from a thick rotation increment to a thin one. To make the point, once again consider the data sets described above in Table 1. Now assume a single reflection has an integrated intensity of 1000 photons, a size on the detector of about 5 × 5 pixels and a rocking curve or mosaicity of about 1° (that is, larger than the previous example). Therefore, for the thick data set the reflection fits on one image and for the thin data set it fits on four images. Suppose further that the X-ray background is about 10 photons per pixel in the thin case and thus 60 photons per pixel in the thick case. Table 3 tabulates how other noise (perhaps arising from the detector itself) affects the relative error in the thick and thin cases (see also Blessing, 1987, for more on error analysis).

From Table 3, it is clear that any additional error or noise per pixel is less deleterious to thickly sliced images than to thinly sliced images. In the latter case, more pixels are required to integrate the observed peak pixels. Table 3 shows some hypothetical values, since modern detectors do not contribute 50 photons per pixel of noise. For example, for the R-AXIS IV detector read-out noise is less than 1 photon per pixel. However, detector hardware does contribute read-out noise and other errors. For example, a CCD-based detector records 'zingers', radioactive-decay events in the fiber-optic taper and phosphor which may affect Bragg reflections. Furthermore, commercial CCD-based detectors subtract a dark-current image and then correct for non-uniformity of

Table 3
Relative error for different levels of noise in thin and thick cases.

	Thin	Thick
Observed intensity	1000	1000
Images per spot	4	1
Spot volume (pixels)	100	25
Background per pixel	10	60
Average peak pixel value	20	100
Other noise	$R_{\rm err}$ (%)	
None	5.5	6.3
10 photons per pixel noise	6.3	6.5
20 photons per pixel noise	7.1	6.7
50 photons per pixel noise	8.9	7.2

response and spatial distortion before the image is available to the experimenter. There are definitely errors and noise associated with these manipulations which should be propagated to the standard uncertainties of the observed intensities. That is,

$$\begin{aligned}\mathrm{var}(I) = {} & \mathrm{var(Peak)} + \mathrm{var(Background)} + \mathrm{var(Dark)} \\ & + \mathrm{var(Nonunf)} + \mathrm{var(SpatDis)} + \text{read noise} \\ & + \mathrm{var(Lorentz)} + \mathrm{var(Polarization)} + \ldots \end{aligned} \quad (9)$$

We can conclude from the analysis in the previous paragraph that not only are a low X-ray background, high source intensity and a low-noise detector important, but that achieving spots as small as possible is also advantageous. In other words, a lower crystal mosaicity and smaller more brilliant more parallel source can help improve $I/\sigma(I)$.

Since the X-ray background falls off as the square of the distance, the expectation is that a larger crystal-to-detector distance is better for the reduction of the X-ray background. While this is true, it is also possible that the physical size of a reflection increases with the distance. For example, with a non-parallel or slightly divergent source, a reflection which is 5 × 5 pixels in size at 100 mm might be 10 × 10 pixels at 200 mm. In this case, there is no reduction in the variance of the observed intensity owing to decreased background per pixel. Indeed, there may be an increase in the variance because of noise introduced by the detector!

Other contributions to the variance arise from the shutter and goniometer hardware. In general, area-detector diffractometers are designed so that shutter opening and closing are synchronized with the start and end of the crystal goniometer rotation. Even so, mechanical shutter operations are not perfectly reproducible, but include some extra 'jitter' of plus or minus a few milliseconds. This jitter will introduce extremely small underscan or overscan errors which are more likely to affect thinly sliced data sets than thickly sliced data sets, as shown in Figs. 1 and 2.

In Fig. 2, the reflection appears on two thin-sliced images or on a single thick-sliced image. If the shutter closes slightly before the end of each small rotation, then a small portion of the reflection is missing ('3 under' in Fig. 2) from each image.

If the rotation goes beyond the desired ending angle, then a small portion of the reflection is exposed twice ('3 over' in Fig. 2). It is evident from Fig. 2 that only the partially recorded reflections are affected, while the fully recorded reflections have only minor errors introduced. Thus, fully recorded reflections from an experiment with moderate hardware problems can still be used for subsequent calculations. In other words, one could still use the full reflections from thick-sliced images, while all of the thin-sliced images would be unusable.

Analogous errors can be introduced if the incident-beam intensity varies considerably from exposure to exposure. These errors can be reduced if an independent measure of the source intensity is available, so that the images can be scaled. Many scaling algorithms can correct for variation in the source intensity. Some of the algorithms rely on fully recorded reflections, so that thickly sliced images have an advantage over thinly sliced images in these cases. More recently, Rossmann & Van Beek (1999) have proposed a method to separately scale each image based on the information available from the reflections partially recorded in the image. For experimental setups with a variable source intensity, unreliable shutter and goniometer control and significant crystal radiation damage, per-image scale factors will be important for good results. If possible, however, the experimental setup should be corrected to help avoid these complications. For example, the source intensity can be measured and used for scaling, the shutter and goniometer designed and fabricated better and crystal radiation damage reduced or avoided by collecting at cryogenic temperatures.

The above paragraphs make the point that thin-sliced area-detector data collection is more demanding of the instrument alignment and stability. Historically, thin-sliced images were collected on diffractometers equipped with area detectors. Software packages with three-dimensional profile analysis were developed at the same time, such as *XDS*, *BUDDHA*, *XENGEN* and *MADNES* (Kabsch, 1988; Blum *et al.*, 1987; Howard *et al.*, 1987; Messerschmidt & Pflugrath, 1987). *d*TREK* is a new software package which also performs three-dimensional profile analysis.

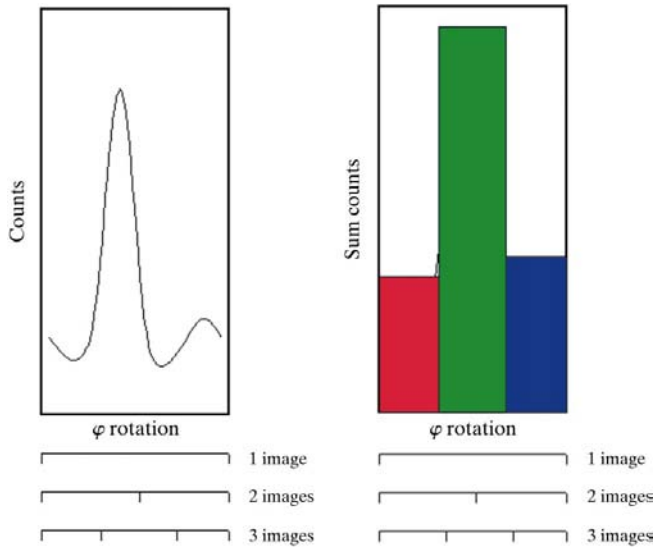

Figure 1
Reflection rocking curve: counts *versus* φ angle for a single detector pixel. If the detector is read out three times for the rotation, then the three images (red, green and blue) shown in the right-hand panel result.

4. *d*TREK*

The *d*TREK* software suite has been developed over the last few years at Molecular Structure Corporation for the processing of diffraction images from position-sensitive detectors such as CCDs and imaging plates. The package can read and process images from almost all area detectors available today. *d*TREK* has modules for displaying images, finding spots, three-dimensional Fourier autoindexing (Bricogne, 1986), refining crystal, detector and source properties, predicting reflections, calculating the best rotation start

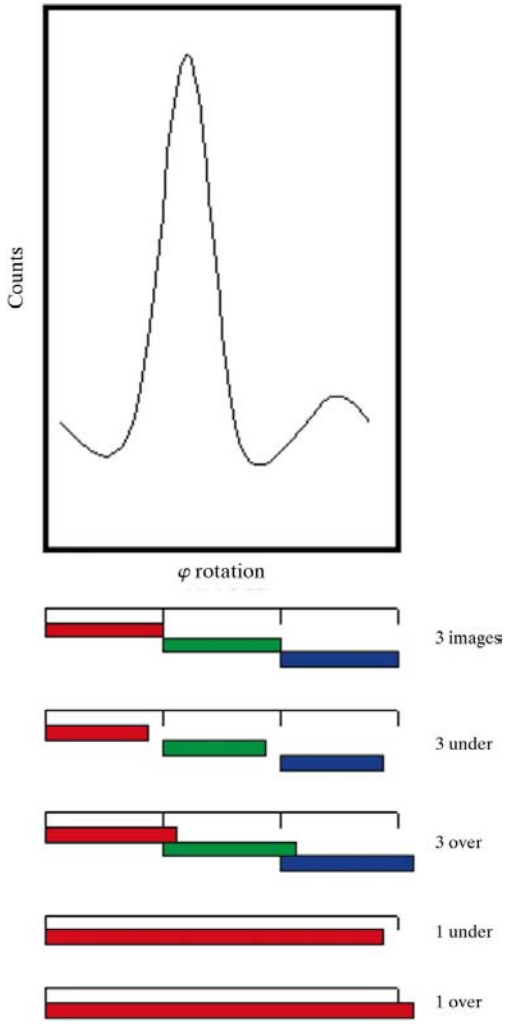

Figure 2
Underscan and overscan goniometer errors. A single reflection recorded completely within an image is affected neither by underscan nor by overscan errors. A single reflection sliced into three separate image *is* affected by these instrument errors.

and end for data collection, integrating the images, merging reflections, correcting empirically for absorption, scaling and averaging reflections. All of the modules are accessed through a graphical user interface developed with X Window and OSF/Motif toolkits.

5. Two-dimensional *versus* three-dimensional integration

Some diffraction analysis suites, such as *HKL* and *MOSFLM*, consider each image separately and integrate reflections in two dimensions first, then sum the integrated contributions from individual images in a subsequent step. On the other hand, the three-dimensional packages mentioned above work with multiple images. For instance, *d*TREK* treats all reflections as filling a three-dimensional volume. Two dimensions are from the area detector itself, while the third dimension arises from the rotation angle of the crystal. One consequence of two-dimensional *versus* three-dimensional integration is that the reflection centroid in the rotation-angle direction at the time of integration is known inaccurately in the two-dimensional case, but very accurately in the three-dimensional case. In the two-dimensional case, a subsequent step known as post-refinement determines accurate rotation angles for reflections. Accuracy is important when predicting reflection positions and fitting reflection profiles.

A profile describes a peak boundary or shape, as well as the intensity distribution within that shape. Profile fitting is the standard technique for integrating reflection peaks (Rossmann, 1979; Diamond, 1969). Two main assumptions are made in profile fitting. The first is that reflections within a region of reciprocal space or detector space have the same profile. The second is that the positions of peak profiles can be accurately predicted. If either of these assumptions fail, then profile fitting is not a valid technique to use for spot integration.

Profile fitting has been detailed in the references cited above. In short, reference profiles are created from the observed profiles of a collection of reflections. Next, each observed profile is fitted with a least-squares procedure to the reference profile and the estimated intensity derived from this minimization. How do the assumptions affect the results? The first assumption (all reflections have the same profile) can be tested by examining correlations of the observed profiles to the corresponding reference profiles. Strong correlations are expected for reflections with non-weak intensities.

The second assumption affects the creation of the initial reference profiles and the least-squares minimization procedure. Superimposing observed profiles creates reference profiles. Usually, peaks are superimposed by shifting their observed or calculated center to a reference profile array with appropriate treatment of non-integral shifts. The shifted peaks are then summed into the reference array. One of the fine details in this step is knowing accurately the required shift vector which in turn requires the second assumption above to be true. Otherwise, the reference profiles will be smeared out and inaccurate. Once reference profiles are available, each observed profile is shifted again to match the corresponding reference profile or *vice versa* in order to perform the least-squares fit.

In principle, thin-sliced images give the best reflection centroids and thus the best profile fitting. Since three-dimensional peaks are strongly influenced by the

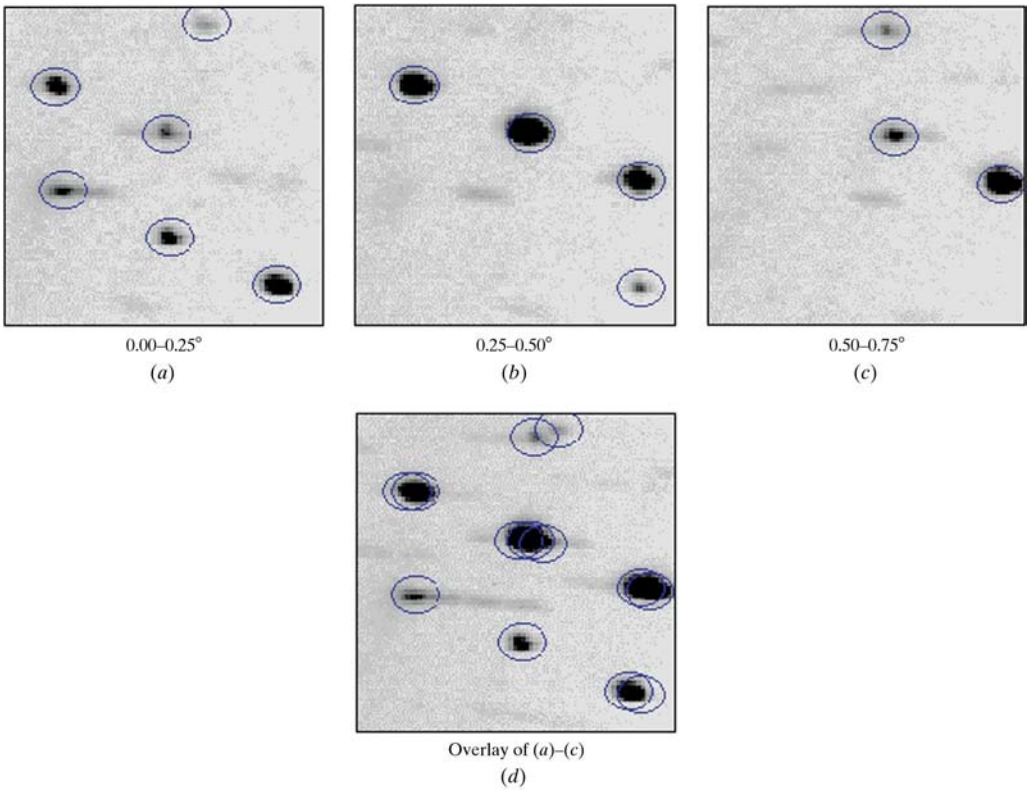

Figure 3
Portions of three adjacent images of 0.25° rotation each: (*a*) 0.00–0.25°, (*b*) 0.25–0.50°, (*c*) 0.50–0.75° and (*d*) the superposition of these three images. Blue ellipses are centered on the observed peaks in (*a*)–(*c*) and superimposed in (*d*). This figure illustrates several features of thin-sliced (0.25°) images *versus* thick-sliced (0.75°) images. Firstly, thin-sliced data reduces spatial overlaps: the unoverlapped reflections at the top of (*a*) and (*c*) are overlapped in (*d*). Secondly, thin-sliced data reduces saturation: the strong reflection near the center of (*a*)–(*c*) sums to be saturated in (*d*). Thirdly, true spot centroids require the complete reflection in three dimensions or fully recorded reflections in two dimensions: the reflection centroids of the partially recorded reflections shift slightly as the crystal is rotated, as shown by the non-superposition of ellipses in (*d*). The streaking of the spots arises from the mirror optics used.

Lorentz factor or distance from the rotation axis, care must be taken to satisfy the first assumption. Kabsch had an ingenious insight in this regard (Kabsch, 1988). He described a conversion from rotation geometry to a precession-like geometry for the area-detector pixels. If one examines a screened precession photograph, it is clear that the observed spots have the same boundary, shape and intensity distribution no matter what their position on the photograph is. That is, *they have the same profile*. The transformation devised by Kabsch assures the first assumption of profile fitting is true. Readers are strongly advised to read Kabsch's excellent manuscript for the fine details of this procedure.

Another fine detail is that the observed spot centroid of a partial reflection cannot be accurately determined from a single image. Since observed centroids are used in refinement of crystal, detector and source properties, it follows that refinement with such centroids yields inaccurate results. A concrete example is shown in Fig. 3. Three images each of 0.25° rotation are shown along with their superposition (a pseudo-0.75° rotation). A blue ellipse is drawn centered at each observed spot position. It is clear that the *xy* positions of the partial peaks shift as the crystal is rotated and that the *xy* centroids are correct only for the fully recorded reflections in the 0.75° image. Accurate three-dimensional centroids of reflections are readily calculated from the complete three-dimensional pixel array surrounding a reflection which is available with thin-sliced diffraction images.

Usually two-dimensional algorithms calculate reference profiles only from reflections within the image integrated. If there are few reflections per image, then the reference profiles may not represent the true profile. Even casual examination of images shows that reflection profiles do not change from image to image, so observed profiles from adjacent images (or adjacent regions of reciprocal space) can lead to improved reference profiles. *d*TREK* uses nearby reflections in reciprocal space to create reference profiles during profile analysis.

6. Results

The previous paragraphs suggest that better results can be achieved with thin-sliced diffraction images as opposed to thick-sliced ones, as long as potential problems are minimized. In this section, results from three different experiments all collected with Cu $K\alpha$ radiation (λ = 1.5418 Å) are presented. Firstly, thin-sliced and thick-sliced images from the same *S. marcescens* endonuclease crystal (kindly supplied by K. Krause) and the same hardware are compared. Secondly, thick-sliced 1° images from a recombinant sperm-whale myoglobin crystal are processed and anomalous Patterson maps calculated. Thirdly, a hen egg-white lysozyme crystal data collection where absorption correction is important is shown.

Fig. 4 plots $I/\sigma(I)$ against resolution for reflections integrated by both *d*TREK* and *HKL* from a 180-image data set (thin) and from a 30-image dataset (thick). The data-collection parameters are described in Table 1. The only difference between the thin and thick data sets is the rotation-angle increment per image. The crystal was frozen and kept at 110 K. The detector was an R-AXIS II equipped with MSC/Yale Total Reflection Mirrors.

As expected, $I/\sigma(I)$ increases significantly with the thin-sliced results. *d*TREK* and *HKL* gave similar results for the thin-sliced data ($R_{\rm merge}$ values of 1.3% and 1.4%, respectively), while *HKL* was better for the thick-sliced data ($R_{\rm merge}$ of 2.1% *versus* 2.9% for *d*TREK*). This is most likely to be a consequence of the relatively large ratio of rotation-angle increment (1.5°) to the effective crystal mosaicity (~0.3°). Currently, *d*TREK* still uses a three-dimensional integration algorithm even with such thick images.

In the myoglobin test case, a total of 120 images, each with a 1° rotation increment and a 75 s exposure time, were collected from a recombinant sperm-whale myoglobin crystal on an R-AXIS IV detector with X-rays of wavelength 1.5418 Å. The crystal mosaicity was about 0.21°. The images were processed with both *d*TREK* and *HKL*. The w = 0 Harker sections of the anomalous difference Patterson syntheses calculated from the processing results are shown in Fig. 5. The map clearly shows the presence of peaks arising from anomalous scattering of the Fe atoms in this experiment of only 2.5 h exposure with a rotating-anode X-ray generator. There is virtually no difference in the Harker sections calculated from the *d*TREK* and *HKL* results.

In the last example, a hen egg-white lysozyme crystal was mounted by flash-freezing in liquid nitrogen in a small loop (Hampton Research, Laguna Niguel, CA). A data set of 90 images was collected, each image with a 0.5° increment and a 10 min exposure. The crystal-mounting loop had been inadvertently bent, so that in some images portions of the loop were positioned between the crystal and the area detector. This situation resulted in a large difference in absorption between reflections on the top and bottom areas of the detector. With *HKL*, which implements a Fox & Holmes

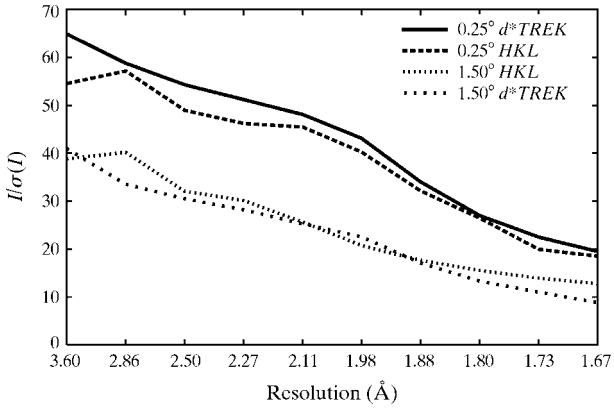

Figure 4
$I/\sigma(I)$ *versus* resolution for thick and thin images. Images from 0.25° and 1.5° diffraction data sets from the same crystal of *S. marcescens* endonuclease were processed with *HKL* and *d*TREK*. The average $I/\sigma(I)$ in ten resolution shells are plotted. The overall $R_{\rm merge}$ is 1.3% for the 0.25° *d*TREK* processing, 1.4% for the 0.25° *HKL* processing, 2.1% for the 1.5° *HKL* processing and 2.9% for the 1.5° *d*TREK* procesing. $R_{\rm merge}$ as a percentage is defined in the usual way as $\sum_h \sum_i |I_{hi} - \langle I_h \rangle| / \sum_h \sum_i |\langle I_h \rangle| \times 100$.

(1966) style scale factor for each image, R_{merge} was 5.7%. With *d*TREK* without absorption correction, R_{merge} was also 5.7%, but with the empirical absorption correction implemented by R. Jacobson, R_{merge} was 3.2%.

7. Conclusions

When exposing diffraction images for single-crystal data collection, one should select the optimal rotation start and end, rotation-angle increment per image and exposure time per image. The ultimate choices are a compromise based on a number of parameters including, but not limited to, the type of hardware, the desired precision of the results and the time available to complete the experiment. The exposure time per image (or per rotation increment) is mainly a function of the available source intensity, diffracting power of the crystal, the desired R_{merge} and the time available. The crystal symmetry and orientation dictates the overall rotation start and end. The rotation increment per image can be chosen to reduce spatial overlaps, saturated pixels and noise in the images. Thin- or fine-sliced images will generally produce fewer spatial overlaps, fewer saturated pixels and lower X-ray noise, but may suffer from increased detector noise and problems with underscans and overscans stemming from inadequate shutter and goniometer control. Thick-sliced images have more spatial overlaps, more saturated pixels and usually more noise, but also have fully recorded reflections which avoid some of the problems of inadequate shutter and goniometer control. Thin-sliced images require more computational resources to process than thick-sliced images.

There is no advantage to ever-thinner rotation increments per image. If the crystal mosaicity is less than 1°, then use 0.5° per image. Otherwise, use one-half of the crystal mosaicity. If this still results in spatial overlaps, reduce the rotation increment, but remember that for large crystal mosaicity values, reflections will be spatially inseparable no matter how small the rotation increment. If there are saturated pixels, one must choose between overexposing the strong or low-resolution reflections (which are required especially for the techniques of solvent flattening and molecular replacement) or inadequate statistics on weak or high-resolution reflections. One possibility is to collect images with multiple passes of different exposure times. For example, the detector could be swung to a higher 2θ with an exposure time of t to collect the high-resolution low-intensity reflections. The detector could then be repositioned to a 2θ of 0° or to a distance further from the crystal with a shorter exposure time of t/F (where F is an exposure-time factor) to re-collect the low-resolution high-intensity reflections. Another method would be to collect two images one after the other with the detector position unchanged, but with different exposure times, t and t/F. Saturated pixel values in the first image could be replaced with appropriately scaled pixel values from the second image.

If the exposure time per image is relatively long compared with the detector read-out time, then there are advantages to decreasing the rotation-angle increment per image. For example, collecting 0.5° images of 4 min each is better than collecting 1° images of 8 min each. These two experiments have identical exposure rates, but the thinner 0.5° images will have lower X-ray background noise, fewer spatial overlaps and fewer saturated pixels. The cost is a very slight increase in the total data-collection time owing to the additional read-outs of the detector.

In general, previous experience with the experimental setup used and the selected crystals will influence the best choices for the data collection. One is not limited by considerations of the number of fully and partially recorded reflections, since software exists which works well in all situations. Usually, a rotation-angle increment of 0.5° per image is an excellent choice.

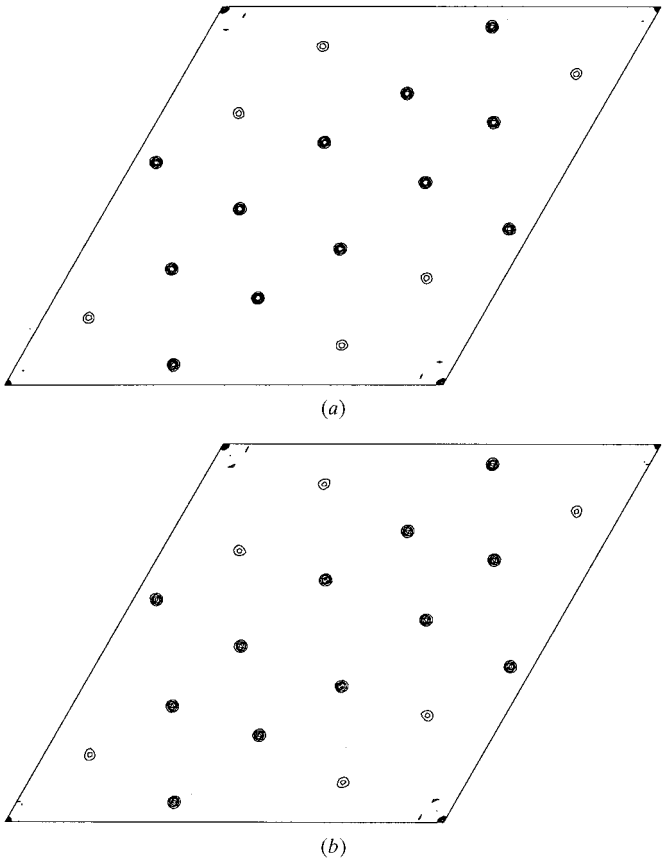

Figure 5
Harker sections ($w = 0$) of the anomalous difference Patterson maps from recombinant sperm-whale myoglobin. Images were processed with either (*a*) the *HKL* suite or (*b*) the *d*TREK* suite and the resulting reflections were processed with the *CCP*4 (Collaborative Computational Project, Number 4, 1994) package. Contours start at 4σ and are at intervals of 8σ, where σ is the standard deviation of the grid values in the complete unit cell. The R_{merge} for the *HKL*-processed data is 4.7%; for the *d*TREK*-processed data, R_{merge} is 4.9%.

I would like to thank my colleagues at Molecular Structure Corporation for the opportunity and resources to pursue this work, especially J. D. Ferrara who collected the endonuclease and myoglobin images. Thanks also to Gerard Bricogne, who organized the EEC Cooperative Programming Workshop on Position-Sensitive Detector Software in the late 1980s and first

coined the name *d*TREK*. I am indebted to E. M. Westbrook, M. Westbrook and W. Kabsch for key insights over the years. K. Krause provided the endonuclease crystal. The development of *d*TREK* was supported in part by US DOE Contract 943072401.

References

Blessing, R. (1987). *Crystallogr. Rev.* **1**, 3–58.
Blum, M., Metcalf, P., Harrison, S. C. & Wiley, D. C. (1987). *J. Appl. Cryst.* **20**, 235–242.
Bricogne, G. (1986). *Proc. EEC Coop. Workshop Position-Sensit. Detect. Softw.* **3**, 28.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Diamond, R. (1969). *Acta Cryst.* A**25**, 43–55.
Fox, G. C. & Holmes, K. C. (1966). *Acta Cryst.* **20**, 886–891.
Howard, A., Gilliland, G. L., Finzel, B. C., Poulos, T. L., Ohlendorf, D. H. & Salemme, F. R. (1987). *J. Appl. Cryst.* **20**, 383–387.
Kabsch, W. (1988). *J. Appl. Cryst.* **21**, 916–924.
Leslie, A. G. W. (1999). *Acta Cryst.* D**55**, 1696–1702.
Messerschmidt, A. & Pflugrath, J. W. (1987). *J. Appl. Cryst.* **20**, 306–315.
Miller, M. D. & Krause, K. L. (1996). *Protein Sci.* **5**, 24–33.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Rossmann, M. G. (1979). *J. Appl. Cryst.* **12**, 225–238.
Rossmann, M. G. & Van Beek, C. G. (1999). *Acta Cryst.* D**55**, 1631–1640.