Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Convergent-beam method in macromolecular crystallography

Joseph X. Ho,[a]* John R. Ruble,[a] Thomas R. McInnis,[a] Daniel C. Carter,[a] Huapeng Huang[b] and Walter M. Gibson[b,c]

[a]New Century Pharmaceuticals Inc., 895 Martin Road, Huntsville, AL 35824, USA, [b]Center for X-ray Optics, University at Albany, State University of New York, Albany, NY 12222, USA, and [c]X-Ray Optical Systems Inc., 15 Tech Valley Drive, East Greenbush, NY 12061, USA

Correspondence e-mail: jho@newcenturypharm.com

Received 13 December 2001
Accepted 24 September 2002

A data-collection method for macromolecular crystals using convergent sources is described here. Because of the unique characteristics of the diffraction patterns, a software package *CBMPRO* has been developed specifically for processing data images collected with the convergent beam method (CBM). The resulting data sets from crystals with two different sets of unit-cell parameters are presented and compared. There is good agreement between data sets from the same type of crystals under slightly different experimental conditions and data sets collected and processed with CBM also agree well with those from conventional oscillation methods, marking an important step to establishing CBM as a viable alternate data-collection method for macromolecular crystals.

1. Introduction

The past decade has seen significant advances in X-ray optics. They range from single capillary to polycapillary, multilayer and doubly bent crystal devices (MacDonald & Gibson, 2001). Capillary optics are based on total reflection with wider energy bandwidths, whilst the multilayer and bent-crystal optics are based on Bragg reflections with much narrower energy bandwidths. These devices collect X-rays from a finite two-dimensional divergent source over a wide solid angle, or from a distant source over a large area, and focus them onto the sample or detector with increased flux.

Since protein and other biomacromolecular crystals are usually very weak diffracters, it is often desirable to obtain increased flux on the crystal with conventional laboratory sources. Taking advantage of new developments in X-ray optics, macromolecular crystallographers have begun using capillary or multilayer devices to increase the X-ray flux on the sample, with varying degrees of success (Owens *et al.*, 1996; Bloomer & Arndt, 1999; Li & Bi, 1998; Verman *et al.*, 2000). Conventional methods for single-crystal diffraction data collection (*i.e.* the oscillation method) are based on the assumption that the oscillating crystal sample is exposed to a highly collimated beam. With an oscillating sample, a larger convergence angle not only brings with it a higher background level, but also causes overlapping of reflections, particularly for crystals with long unit-cell axes. Hence, it is not a coincidence that all commercially available new X-ray optical devices for macromolecular crystallography have output convergence angles of $\sim$3 mrad or less.

In order to better utilize the potential of these devices for crystallographic applications, a new diffraction method – the convergent-beam method (CBM) – has been proposed (Ho *et al.*, 1998), in which a stationary crystal is exposed to a beam with a large two-dimensional convergence. The diffraction

geometry of CBM has been described. Simulated patterns of a lysozyme crystal were presented and compared with their experimental counterparts from a polycapillary lens. However, processed data in terms of integrated intensities, which are the main objective of a diffraction data-collection experiment, were not presented owing to the lack of data-processing software. Here, we present the results of data collection with convergent X-rays from a polycapillary optic on crystals with different unit-cell axis lengths. The data were processed using the recently developed software package *CBMPRO*. The resulting data sets from the same crystal system collected with different optics were compared for reproducibility. These data sets were also compared with data sets from the same crystal system but using conventional data-collection methods for equivalence. The emphasis of this article is on the validity of CBM, rather than establishing the maximum possible flux on sample for a given convergence and a given focus spot size or obtaining a source with the most uniform beam angular distribution. Although our experiments are limited to convergent X-rays from a polycapillary optic, the methods and data-processing software should be generally applicable to other convergent X-ray sources or convergent neutron sources.

In order to avoid confusion, it should be pointed out that the word 'source' is used with two different meanings in different contexts. In the context of describing our specific experimental setup, the word source means the actual microfocus X-ray tube that emits X-ray photons. When it comes to describing the CBM diffraction geometry and properties *etc.*, we are concerned only with certain characteristics of the X-ray as 'seen' by the crystal and call them the characteristics of the X-ray 'source'. In this context, for generality, we are not particularly interested in how the characteristics of the X-rays were produced, although in our specific experimental setup they are the result of the combination of a microfocus X-ray source and a polycapillary lens.

2. Diffraction properties of CBM

The essence of CBM in single-crystal data collection is the exposure of a stationary crystal to a monochromatic two-dimensionally convergent source. The diffraction geometry and special characteristics of CBM are briefly summarized here for readers' convenience. For a more detailed description, please refer to our earlier article (Ho *et al.*, 1998).

For ease of discussion, a Cartesian system is adopted where the x axis coincides with the central direct beam of a convergent source, the z axis is vertical (pointing upwards) and the y axis is perpendicular to both the x and z axes and is hence horizontal. The origin of this system is also the reciprocal-lattice origin of the crystal when we discuss the diffraction conditions. Hence, a cone around the x axis with its apex at the origin represents a convergent source. It is also convenient to construct a sphere centered on the origin with a radius of λ^{-1}. All points located on this sphere and within the convergence cone form the entirety of the convergent source, which is sometimes referred to as the source-sphere element.

2.1. Reciprocal-lattice points stimulated

The Bragg equation can be expressed as $\mathbf{s} - \mathbf{s}_0 = \mathbf{P}$, where $\mathbf{s}_0$ and $\mathbf{s}$ represent the incident and scattered beam vectors, respectively, both having a magnitude λ^{-1}, and $\mathbf{P}$ is a reciprocal-lattice point vector. The diffraction condition is illustrated in Fig. 1, where the shaded area represents the part of reciprocal space within which each reciprocal-lattice point satisfies the Bragg diffraction condition for a certain part of the source-sphere element. This condition can also be expressed as

$$\cos(\beta + \mu_{\max}) \leq \rho\lambda/2 \leq \cos(\beta - \mu_{\max}), \qquad (1)$$

where β is the angle between reciprocal-lattice vector $\mathbf{P}$ and $-x$, $\mu_{\max}$ is the maximum divergence of the source and ρ is the magnitude of vector $\mathbf{P}$.

2.2. Active source for a particular reflection

For a reciprocal vector $\mathbf{P}$ stimulated by the convergent source, if one spins the isosceles triangle formed by $\mathbf{s}$, $\mathbf{s}_0$ and $\mathbf{P}$ around vector $\mathbf{P}$, the trace of the apex of the triangle is a planar circle, which indicates the locations where a potential source, if it exists, could participate in the diffraction of reflection $\mathbf{P}$.

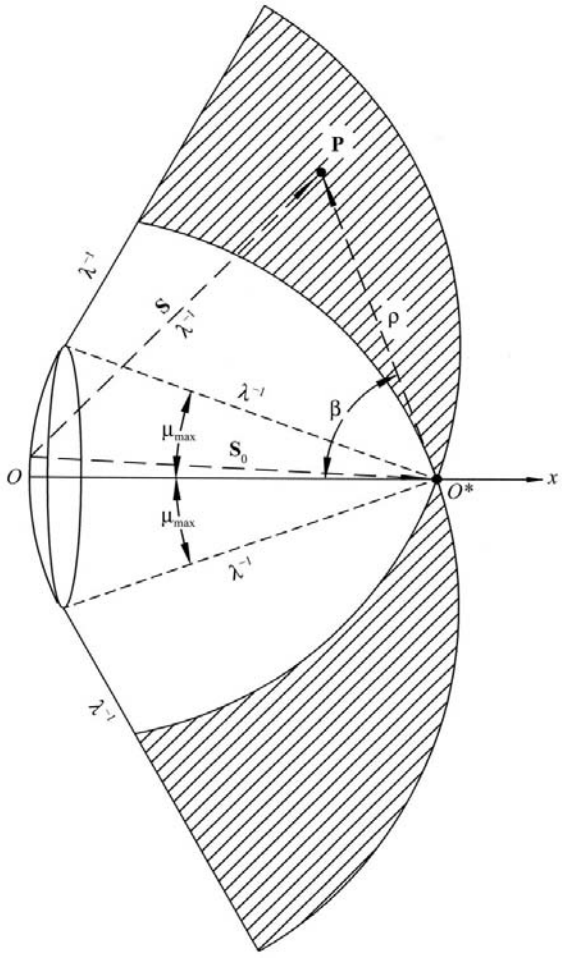

Figure 1
Region of reciprocal space involved in diffraction (shaded area). This region is rotationally symmetrical around the x axis.

However, only a small part of this circle is within the source-convergence cone and is located on the sphere element of the convergent source, which is the set of physically existent source points. Let us call this part of the source the 'active source' or 'active-source line' for reflection **P**. All points $S(\varphi_s, \mu_s)$ on this active-source line should satisfy the equation

$$\cos\Delta\varphi = \cos(\varphi_s - \varphi_p) = (\rho^2 + 2x_p\lambda^{-1}\cos\mu_s)/2(R_p\lambda^{-1}\sin\mu_s). \quad (2)$$

Important points to note: at the mid-point of this active-source line, we have

$$\mu_s = \mu_{\min} = |\arccos(\rho\lambda/2) - \beta|,$$
$$\varphi_s = \varphi_p \text{ (or } \varphi_p \pm \pi),$$

whilst at the two ends of the active-source line, we have

$$\mu_s = \mu_{\max}$$
$$\varphi_s = \varphi_p \pm \Delta\varphi \ (\mu_s = \mu_{\max}).$$

Though the mathematical expression for the active source is a little complicated, the angle $\mu_{\max}$ is small enough (around 1°) that for all practical purposes it suffices to treat the active source as a straight line connecting points $[\mu_{\max}, \varphi_p - \Delta\varphi(\mu_{\max})]$ and $[\mu_{\max}, \varphi_p + \Delta\varphi(\mu_{\max})]$ on the two-dimensional plot of the source, as schematically shown in Fig. 2.

2.3. Tangential elongation of diffraction spots

Every reflection stimulated by the convergent source has its own corresponding active source and hence its own corresponding source plot similar to that in Fig. 2. The length of the active-source line indicates the angular width of the active source and therefore also the angular width of the particular reflection. The angular width for a particular reflection is dependent on the orientation of the crystal. For a flat area detector perpendicular to the central direct beam, this leads to tangential elongation of a diffraction spot. The tangential width of a diffraction spot is proportional to the angular width and also to the distance from the crystal to the spot on the detector.

2.4. Source factor

For moving-crystal methods (*e.g.* the oscillation method), there is a Lorentz factor which calibrates the relative difference in time for different reflections. Here, we have a source factor describing the difference in the active source of different reflections. The source factor is a product of two parts. The first is $(\sin 2\theta)^{-1}$, which addresses the rate of magnitude of the diffracted beam vector **s** departing from λ^{-1} while the source points deviate laterally away from the active-source line. We will call this the pseudo-Lorentz factor L. The second addresses the differences in length of the active-source line and in the non-uniform source angular distribution. This term can be calculated as a line integral of the source angular intensity distribution $I(\mu, \varphi)$ along the active source line C, *i.e.*

$$S = \int_C I(\mu, \varphi)\,\mathrm{d}C, \quad (3)$$

where C is the active-source line, all points on which satisfy (2). We will call this component of the source factor S. The difficulty in obtaining an accurate measurement of the source angular distribution led us to use a model function with adjustable parameters which can be optimized during the intensity-scaling process by minimizing intensity differences in symmetry-related reflections.

2.5. Polarization factor

In order to estimate the polarization factor, one should first analyze the polarization characteristics of the beams impinging on the diffracting crystal. Here, we consider only the case of a 'perfect' polycapillary lens. In this case, the least convergent (the 'straight-through') beams ($\mu = 0$) are circularly polarized, while the most convergent beams ($\mu = \mu_{\max}$), originating from the most divergent beams at the input side of the optic and having undergone the highest number of bounces on the capillary wall, are the most polarized, with the radial and tangential directions as their polarization directions; according to classical electromagnetic theory, the intensity ratio of the two polarized directions is

$$I_r/I_t = \cos^2(\mu_C + \mu_{\max}), \quad (4)$$

where μ_C is the (one-sided) capture angle of the optic. Any beam with intermediate value of μ, the degree of polarization also takes an intermediate value,

$$I_r/I_t = \cos^2[(\mu_C + \mu_{\max})(\mu/\mu_{\max})]. \quad (5)$$

For a particular reflection, the polarization factor for a point (μ, φ) on its active-source line can be estimated (Azároff, 1955),

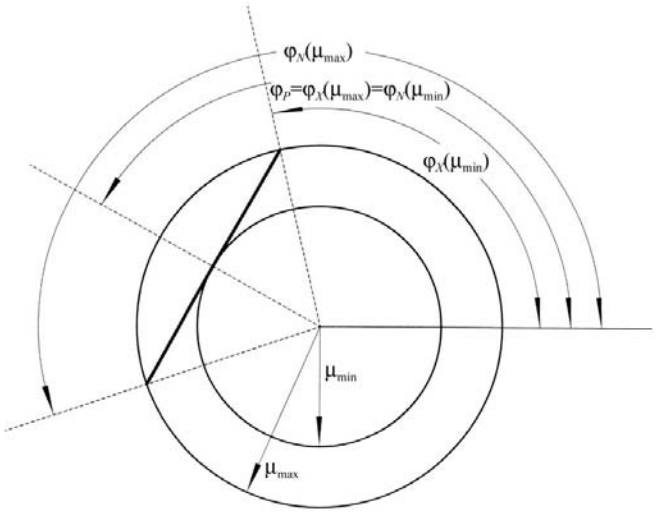

Figure 2
φ–μ polar plot for a reflection. The dark solid (nearly) straight line represents the active-source line for this particular reflection. $\mu_{\max}$ is the maximum convergence angle of the optic and $\mu_{\min}$ is the minimum convergence angle for this particular reflection. For every μ between $\mu_{\min}$ and $\mu_{\max}$ there generally are two points $\varphi_N(\mu)$ and $\varphi_X(\mu)$ on the active-source line which satisfy the diffraction condition for this particular reflection, where the subscripts 'N' and 'X' designate 'entrance' and 'exit' owing to their similarity to a screenless precession experiment (Xuong & Freer, 1971).

$$P(\mu, \varphi) = \frac{[(\cos^2 2\theta \cos^2 \varepsilon + \sin^2 \varepsilon)\cos^2 \delta + \cos^2 2\theta \sin^2 \varepsilon + \cos^2 \varepsilon]}{(1 + \cos^2 \delta)}, \quad (6)$$

where $\delta = (\mu_C + \mu_{\max})(\mu/\mu_{\max})$ and ε is the angle between the radius direction at source point (μ, φ) and the diffraction vector **P**. Because the angle ε varies along the active-source line, one soon realises that strictly speaking the polarization factor and source factor become inseparable and that they have to be included in the same line integral as (3), *i.e.* the joint polarization and source factor PS,

$$\mathrm{PS} = \int_C I(\mu, \varphi) P(\mu, \varphi)\, \mathrm{d}C. \quad (7)$$

If we consider the fact that the intensity falls off with the increase of μ, then the polarization factor is dominated by the term at the middle of the active-source line; if as an approximation, we further assume the beam-polarization condition along the active-source line all follow that of the mid-point, then one can still estimate the polarization factor separately from the source factor, *i.e.*

$$P = (\cos^2 2\theta \cos^2 \delta_m + 1)/(1 + \cos^2 \delta_m), \quad (8)$$

where $\delta_m = (\mu_C + \mu_{\max})(\mu_{\min}/\mu_{\max})$. For the optics used in this study, the upper-bound value for $(\mu_C + \mu_{\max})$ is less than 8.5° and reflections with $\mu_{\min}$ approaching $\mu_{\max}$ are associated with very small source factor and are poorly measured, and are hence often excluded from the final scaling. Therefore, the value of $\cos^2\delta_m$ generally deviates from unity by less than two percentage points. Hence, the polarization factor deviates from that for unpolarized incident beam, $P = (\cos^2 2\theta + 1)/2$, also by no more than 2%. The results shown in this article are based on the simpler formula for an unpolarized incident beam.

3. Experimental setup and alignment

3.1. Experimental setup

The experimental setup is schematically illustrated in Fig. 3. An Oxford UltraBright Micro X-ray Source (*U*) is mounted on a motorized *xyz* translation stage (*T*), which is in turn mounted on a base plate (*B*) with three height-adjustable feet. The feet of the base plate rest on the workbench. Coarse horizontal positioning with respect to the MSC/Rigaku R-AXIS detector (*D*) is achieved using individual translating devices (not shown) mounted on the workbench by each foot, in a way similar to that used to align the R-AXIS detector system. The source *xyz* translation stage is driven by stepping motors which provide incremental movements as fine as 1 µm of the X-ray source with respect to the optic/shutter/collimator (OSC) block mounted on the base plate. This allows precise lateral (*y* and *z* directions) and distance (*x* direction) adjustment of the source with respect to the optic, in order to place the micro X-ray source at the focal point of the optic. For initial alignment steps of 100 µm are adequate, while for final alignment steps of 5 µm are required. The OSC block has a fixed mount for the optic, a rotary shutter and a collimator mount with two pairs of micrometers for horizontal and vertical alignment of the collimator with respect to the optic/shutter. The collimator is of a sleeve-in-sleeve expandable design with replaceable front (close to sample) and rear (close to optic/source) apertures. Various pinhole combinations and collimator lengths can be selected for different convergence angles and focal lengths. In this regard, it is not a true collimator, but serves as a beam limiter. A piece of 12 µm Ni foil is later inserted behind the collimator for filtering $K\beta$ and shorter wavelength white radiation. A MSC/Rigaku R-AXIS IV imaging-plate detector with a single-axis goniostat is mounted on the same workbench for diffraction measurements.

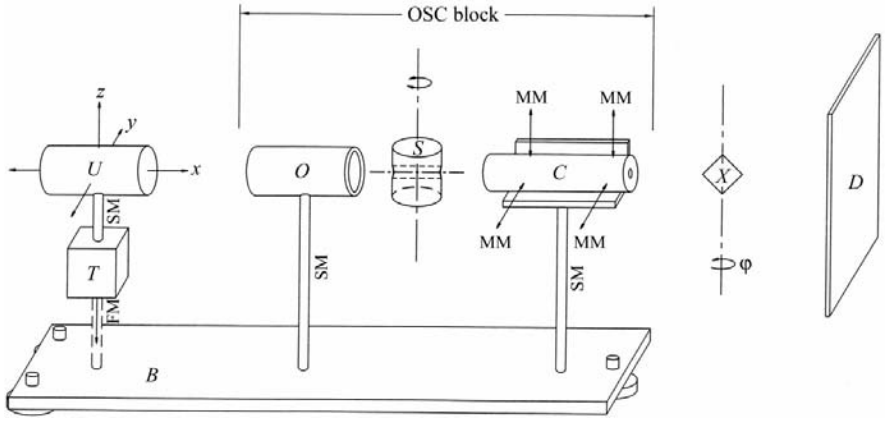

Figure 3
A schematic diagram for the experimental setup. Notation: *U*, ultrabright micro X-ray source; *T*, motorized *xyz* translational stage; *B*, base plate; *O*, polycapillary optic; *S*, rotary shutter; *C*, collimator; *X*, crystal; *D*, detector. Note also that MM indicates micrometer, SM indicates a solid mount and FM a flexible mount.

3.2. Alignment process

For the initial alignment, all apertures are removed from the collimator and a pin diode is placed at the end of the collimator. A *yz* search in steps of 100 µm is performed to locate the initial beam. Centering of the X-ray focal spot (with respect to the polycapillary lens) is then optimized by maximizing the pin-diode reading with reduced *yz* increments. Once this is complete, a search is conducted in the *x* direction using steps of 100 µm to optimize the focal distance. These steps are repeated until no further improvement is obtained. Next, the collimator is aligned to the maximum intensity using the four micrometers for collimator alignment. At the beginning of this procedure a large aperture combination was used; small apertures were then used to achieve precise alignment of the collimator.

Finally, the source and base plate need to be aligned with the detector. This is achieved using similar techniques described for alignment of the R-AXIS IV to any source. Coarse alignment adjustments are made to the feet on the base plate using the table-mounted translators. Fine alignment is achieved by adjusting the detector and the φ-stage.

4. Data-collection and processing procedures

4.1. The first diffraction image and its indexing

An important feature of diffraction with a convergent source is that it is necessary to select a set of orientations of the crystal sample for X-ray exposures where the overlapping effect is minuscule, whilst avoiding orientations where the overlapping effect is very serious. Severely overlapped diffraction patterns can rarely be indexed. To avoid the necessity of taking several trial exposures at different orientations for indexing purposes, one can take a single diffraction image with limited convergence. This can easily be achieved by reducing the size of the back aperture of the collimator. From this first image, the orientation matrix can be determined.

In the processes of initial peak-picking, conversion of peaks into reciprocal-space vectors and orientation-matrix refinement, special algorithms are designed to accommodate the unique nature of diffraction-spot location and shape of CBM, *i.e.* the above-mentioned (in §2.3) tangential elongation for a flat detector perpendicular to the central direct beam. The spot-indexing procedure generally follows the algorithm of Rossmann & van Beek (1999).

4.2. Selecting orientations for exposure: data-collection strategy

With the orientation determined from the first image, the software can select from a set of orientations, *e.g.* 120 ω angles ranging from 0 to 180° in 1.5° increments (increment size selectable), an optimized subset of orientations which would produce a data set with a user-specified completeness for a targeted resolution. The optimization procedure uses the algorithm of Nikonov & Chirgadze (1985) and is currently implemented only for single-axis (ω) crystal rotations. It should be straightforward to implement this for additional axes, *e.g.* of κ-geometry or Eulerian geometry. This procedure totally disregards the order of the ω values. This is understandable if one recognizes the fact that under CBM there is no connectiveness of consecutive orientations as in OM, where the reciprocal space is explored consecutively from one frame (crystal orientation) to the next.

4.3. Intensity integration

Once the data frames are collected, the diffraction spots are integrated by simple summation. The detector surface is divided into nine zones according to the distance from the detector center. An 'elastic shoebox' is assigned to each of these zones, which maps the reflection peak- and background-defining areas. The two directions of the shoebox represent the tangential and radial directions of a diffraction spot. The tangential dimension contracts or expands according to the tangential widths of individual reflections. The general procedure for intensity calculation for each frame is as follows.

4.3.1. Parameter and orientation-matrix refinement. Firstly, the exact locations of all reflections and their deviations from prediction from the initial orientation matrix are found. The unit-cell parameters, camera parameters and crystal mis-setting angles, and hence the orientation matrix, can then be refined according to a protocol set by the user through a least-squares iteration procedure on a selected subset of the reflections.

4.3.2. Marking peak areas on detector. With the refined orientation matrix, all the reflections falling on the detector frame are calculated. The peak area of all reflections are marked onto detector pixels according to the shoeboxes. If overlapping occurs (*i.e.* a given pixel marked as part of the peak areas of more than one reflection), it is recorded as such. Within the shoebox of a particular diffraction spot, the background pixels (pixels designated for background estimation) which are also marked as peak areas of neighboring spots are excluded from the background calculation for the spot in question.

4.3.3. Set up of two-dimensional arrays by interpolation and integration of intensities by simple summation. For every non-overlapping reflection, a two-dimensional orthogonal array is set up which represents detector readings along tangential and radial directions. The tangential width varies from reflection to reflection. The value of each array element is calculated from interpolation of nearest detector pixels. The 'elastic' shoebox of the zone is used to map the elements for peak or background. The background is then calculated by least-squares and the intensity is calculated by simple summation.

4.4. Data-scaling procedure

In the process of the 'post-refinement' styled scaling procedure, there are two types of parameters to be refined in the least-squares procedure. These are 'local' parameters pertaining to individual frames and 'global' parameters pertaining to all of the data. The local parameters include one inverse scale factor, one temperature factor and three mis-setting angles, a maximum of five variables per frame. The global parameters include cell dimensions (a maximum of six independent variables) and parameters defining the source-factor model function. Currently, the model source-factor function S is defined as a product of two functions, $R(|\mu|)$ and $A(\varphi)$, *i.e.*

$$S = R(|\mu|) \cdot A(\varphi), \tag{9}$$

where

$$R(|\mu|) = c_1 + c_2 \exp(-c_3|\mu|^2) + c_4 \exp(-c_5|\mu|^3),$$
$$A(\varphi) = 1 + \sum_{j=1,4} (a_j \cos j\varphi + b_j \sin j\varphi).$$

There are 13 independent variables describing the source factor. In principle, it is possible to calculate the source factor

Table 1
Experimental conditions and statistics of data sets using CBM.

Data set	LCBM1	LCBM2	HCBM
Internal reference	ns789ly2	ms291ly	ns680hmg
Protein	Chicken egg-white lysozyme	Chicken egg-white lysozyme	Hemoglobin (oxidized)
Optic used	No. 789	No. 1291	No. 680
Size of source (estimated, μm)	70	35	70
Generator setting (kV, mA)	38, 0.66	38, 0.53	38, 0.66
Convergence (°)	1.00	0.85	0.80
a (Å)	79.21	79.02	54.1
b (Å)	79.21	79.02	54.1
c (Å)	38.01	37.84	195.7
Laue symmetry	4/*mmm*	4/*mmm*	4/*mmm*
Detector distance (mm)	150	150	200
No. frames collected	32	42	34
Exposure (min)	40	30	90
Resolution (Å)	2.0	2.0	2.5
Data-processing results			
No. of observations	47109	61631	45377
Unique reflections	8435	8401	10526
Minimum σ	0.00	0.00	0.00
Minimum source factor†	0.10 S_{max}	0.25 S_{max}	0.20 S_{max}
Observations used in scaling	32519	40055	21938
Reflections used in scaling	7357	7773	6844
R_{merge}(RMS)‡ (%)	6.88	7.58	7.13
R_{merge}(ABS)‡ (%)	9.18	8.88	12.08
No. observations output	33404	40592	24637
No. reflections output	8249	8296	9544
Completeness (%)	96.1	98.1	89.6
Last-shell completeness (%)	89.2	89.2	85.3

† S_{max} is the maximum source-factor value for the source. ‡ $R_{merge}(\mathrm{RMS}) = [\sum w(I - \langle I\rangle)^2 / \sum w\langle I\rangle^2]^{1/2}$ and $R_{merge}(\mathrm{ABS}) = \sum |I - \langle I\rangle| / \sum \langle I\rangle$, where I are individual integrated intensities with frame scale and source factor applied, $\langle I\rangle$ is the weighted average over symmetry-related reflections, w is the weight and the summations are over all observations.

for individual reflections from the orientation matrix and the source angular distribution through a linear integration (Ho *et al.*, 1998). It is the difficulties involved in deconvoluting the source angular distribution from the direct-beam image that motivated us to use a mathematical model function to describe the source factor and optimize the defining parameters in the final scaling procedure. The above model function, though extremely restrictive, has been used for several data sets collected with polycapillary lenses with reasonable success. Owing to the highly correlated nature of these parameters, very small damping factors and many cycles are required in the scaling process, even though an eigenvector-filtering procedure is used in the least-squares.

5. Results

Three data sets were collected using CBM and processed with the *CBMPRO* program package. The experimental conditions are listed in Table 1. Two different microfocus X-ray sources under development at Oxford Scientific Inc. were used. All three data sets were from different crystals and used different capillary optic lenses. Two of them were data sets collected on the tetragonal form of chicken egg-white lysozyme and are referred to as LCBM1 and LCBM2 for convenience. The third set, from a hemoglobin crystal, was selected for its relatively long unit-cell axis (approaching 200 Å) and is referred to as HCBM. Fig. 4 shows a typical diffraction image from a lysozyme crystal, while Fig. 5 shows that of a hemoglobin crystal of longer unit-cell axis. The diffraction spots show tangential

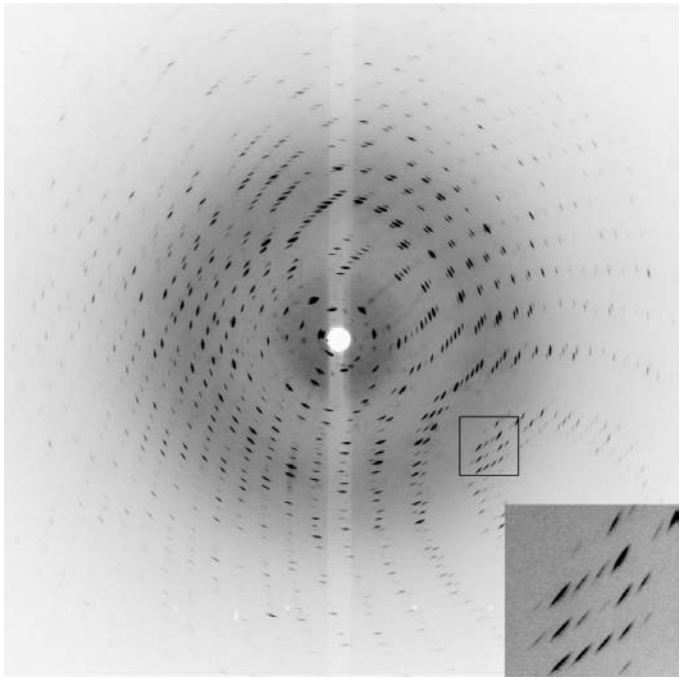

Figure 4
A typical diffraction pattern of a tetragonal lysozyme crystal with a convergent source. The source was running at 25 W (38 kV, 0.66 mA) coupled with XOS lens No. 789. The crystal-to-detector distance was 200 mm and the exposure time was 40 min. The inset at the lower right is a threefold enlargement of the area marked by the square border.

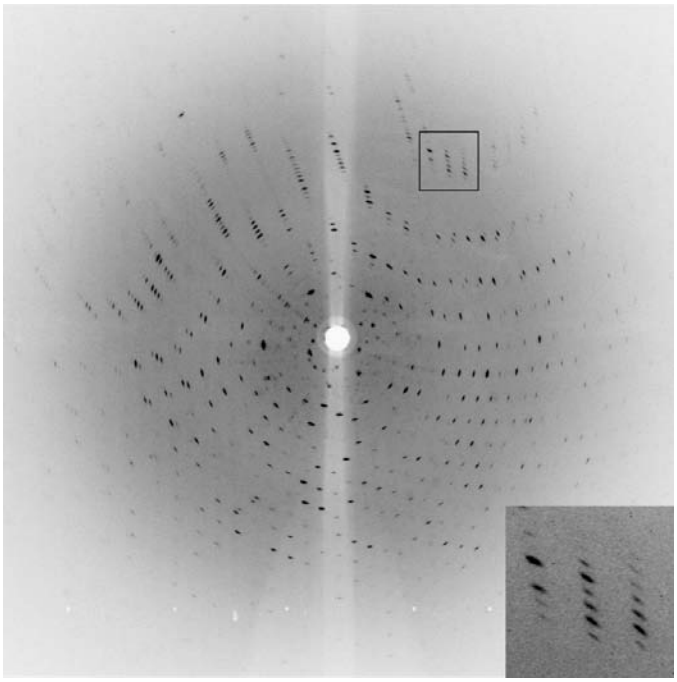

Figure 5
A typical diffraction pattern of a tetragonal hemoglobin crystal with a convergent source. The source was running at 25 W (38 kV, 0.66 mA) with XOS lens No. 680. The crystal-to-detector distance was 250 mm and the exposure time was 30 min. The inset at the lower right is a threefold enlargement of the area marked by the square border.

Table 2
Data-collection strategy for data set LCBM1.

Notes. (1) The Nikonov & Chirgadze (1985) procedure is implemented in a way best described by the following steps. (i) Set up a set number of still frames at equi-orientation space from 0 to 180°. (ii) For each frame, calculate the set of reflections and their indices that would appear in a diffraction experiment. Assess their possibility of overlapping. (iii) Among all the frames, select the frame which contains the highest number of independent reflections. Put the indices of these independent reflections in an Indices Storage. (iv) Select the next frame among those remaining frames which would add the highest number of new independent reflections (with indices other than those already stored one) to the total set. Expand the Indices Storage to include the indices of those new independent reflections. (v) Calculate the combined completeness; if it reaches the targeted completeness at the targeted resolution, stop. Otherwise, repeat steps (iv) and (v). (2) For any particular reflection rejection could happen no more than once. The rejection tests start from left to right, *i.e.* a reflection rejected because of being off the detector plate (EDGE-REJ) is not then further tested for overlapping. Likewise, a reflection rejected for overlapping (OVERLAP) will not be rejected again because its source factor falls under a set minimum (SRC-REJ). The total number of usable observations was 34 855; the total number of independent reflections was 8497; the average redundancy of the data set was 4.1.

(*a*) Minimum No. of frames selected by the program.

NO.	ω (°)	OBSVD	EDGE-REJ	OVERLAP	SRC-REJ	NET-OBS	UNIQUE	(%)	CUMULAT	(%)
1	34.0	2238	21	0	891	1347	1282	14.9	1282	14.9
2	120.0	2167	23	0	850	1317	1248	14.5	2356	27.5
3	136.0	2197	19	0	899	1298	1249	14.6	3270	38.1
4	108.0	2181	24	0	902	1279	1235	14.4	4064	47.4
5	126.0	2146	19	0	843	1303	1258	14.7	4735	55.2
6	140.0	2198	22	0	923	1275	1238	14.4	5273	61.5
7	130.0	2160	20	0	889	1271	1222	14.2	5747	67.0
8	116.0	2189	21	0	922	1267	1211	14.1	6151	71.7
9	174.0	2093	15	0	845	1248	1183	13.8	6499	75.7
10	114.0	2203	28	0	961	1242	1210	14.1	6781	79.0
11	128.0	2183	25	0	921	1262	1207	14.1	7023	81.9
12	132.0	2207	21	0	918	1289	1241	14.5	7240	84.4
13	148.0	2166	27	0	922	1244	1184	13.8	7428	86.6
14	106.0	2177	25	0	899	1278	1227	14.3	7594	88.5
15	124.0	2198	22	0	926	1272	1222	14.2	7736	90.2
16	142.0	2172	22	0	941	1231	1198	14.0	7861	91.6
17	134.0	2168	19	0	893	1275	1204	14.0	7963	92.8
18	102.0	2170	24	1031	891	723	695	8.1	8043	93.7
19	122.0	2179	31	0	954	1225	1187	13.8	8117	94.6
20	138.0	2174	22	0	916	1258	1211	14.1	8178	95.3
21	118.0	2152	20	0	903	1249	1208	14.1	8232	95.9
22	74.0	2202	24	1411	924	539	529	6.2	8284	96.6
23	112.0	2188	23	0	906	1282	1240	14.5	8327	97.1
24	80.0	2173	20	1565	899	440	433	5.0	8367	97.5
25	84.0	2171	22	1589	905	427	413	4.8	8393	97.8
26	110.0	2168	23	0	936	1232	1184	13.8	8419	98.1
27	104.0	2218	26	332	943	1096	1052	12.3	8438	98.3
28	78.0	2168	25	1525	885	459	451	5.3	8455	98.5
29	32.0	2095	21	0	879	1216	1162	13.5	8468	98.7
30	98.0	2185	23	1431	905	511	501	5.8	8479	98.8
31	144.0	2185	22	0	898	1287	1237	14.4	8489	98.9
32	24.0	2216	22	1826	932	213	210	2.4	8497	99.0

(*b*) Resolution breakdown of the simulated data set.

RES (Å)	OBSVD	EDG-REJ	OVERLP	SRC-REJ	NET-OBS	INDEP	(%)	CUMUL	(%)
5.43	3631	0	272	1475	2005	478	97.0	478	97.0
4.31	3572	0	337	1421	1973	454	99.6	932	98.2
3.76	3607	0	494	1488	1834	431	98.4	1363	98.3
3.42	3523	0	486	1416	1845	434	99.5	1797	98.6
3.17	3497	0	497	1454	1790	430	99.5	2227	98.8
2.99	3600	0	494	1477	1856	441	99.1	2668	98.8
2.84	3494	0	524	1399	1797	414	99.0	3082	98.8
2.71	3446	0	536	1387	1762	419	99.1	3501	98.9
2.61	3557	0	580	1492	1737	432	99.8	3933	99.0
2.52	3443	0	565	1440	1705	415	99.5	4348	99.0
2.44	3477	0	574	1379	1781	418	99.5	4766	99.1
2.37	3422	0	544	1385	1738	423	100.0	5189	99.1
2.31	3503	0	565	1455	1736	421	99.3	5610	99.2
2.25	3448	0	588	1431	1689	411	99.5	6021	99.2
2.20	3435	0	572	1440	1701	419	99.1	6440	99.2
2.15	3438	0	608	1423	1684	421	99.1	6861	99.2
2.11	3296	0	586	1385	1582	400	100.0	7261	99.2
2.07	3477	0	637	1454	1667	426	98.8	7687	99.2
2.03	3441	128	629	1465	1638	408	97.8	8095	99.1
2.00	3380	593	622	1755	1335	402	97.3	8497	99.0

elongation with various widths. In all these experiments, the crystals used were about or slightly larger than the X-ray focal spot size (~300 μm for OM and CBM experiments).

Results of a data-collection strategy session for data set LCBM1 are partially reproduced in Table 2. In addition to space group, unit-cell parameters and mis-setting angles, the calculation requires additional information such as the source convergence, parameters defining the source-factor model function, crystal mosaicity, crystal-to-detector distance and detector dimensions *etc.* To save computing time, a simplified algorithm is used in estimating overlaps.

For each data set, a set of crystal orientations were selected in the fashion described above. The diffraction images thus collected were integrated and converted into a series of files containing reflection indices and their corresponding integrated intensities. At the beginning of a 'post-refinement' styled scaling process, all observations were sorted according to their unique reflection indices. The results from these iterative parameter refinement and scaling processes are tabulated in Table 1. The $R_{\rm merge}$(RMS) values range from 6.88 to 7.58% [$R_{\rm merge}$(ABS) values range from 8.88 to 12.08%], no statistical outliers have been rejected (a mechanism for this has not yet been implemented in the program package). The data completenesses range from 89.6 to 98.1% for the corresponding targeted resolutions, respectively.

For comparison purposes, we collected another data set from yet another crystal of lysozyme on a Rigaku RU-200 rotating-anode generator coupled with an Osmic Confocal-Blue Optic, using the conventional (oscillation) method. This data set was processed using the *DENZO* package from *HKL* and is referred to as LOM. All three sets of lysozyme data (LCBM1, LCBM2 and LOM) are compared in Table 3. The *R* factors based on intensities range from 13.8 to 14.5% for all common reflections between two data sets. Considering the fact that the *R* factors based on structure factors are usually half those based on intensities, these are fairly good agreements. It is worth noting that the *R* factor between data sets from CBM (14.5% between LCBM1 and LCBM2) is virtually identical to the *R* factors between these data sets and their counterparts from OM (14.1% between LCBM1 and LOM and 13.8% between LCBM2 and LOM, respectively), suggesting both the data reproducibility of CBM and the equivalency of data from CBM and from OM.

Similarly, we collected a data set from a different hemoglobin crystal using OM and again processed the data with the *DENZO* package for purposes of comparison. This data set is referred to as HOM. The *R* factor between HCBM and HOM is 21.20% for all 7798 common reflections. This is a little less satisfying than the results for lysozyme crystals. We believe that the higher *R* factor is partially caused by a different degree of oxidization in the two crystals.

Table 3
Comparison of three data sets of lysozyme by *R* factors.

All *R* values are unweighted and based on intensities, *i.e.* $R = |\sum I_1 - kI_2|/\sum |I_1|$.

	LCBM2	LOM
LCBM1	14.5% (7852 reflection pairs), 14.1% (after 16 pairs rejected)	14.1% (7884 reflection pairs), 11.4% (after 16 pairs rejected)
LCBM2		13.8% (7923 reflection pairs), 11.1% (after 19 pairs rejected)

6. Summary and discussion

It has been demonstrated that meaningful macromolecular crystal diffraction data can be collected using X-ray sources with high two-dimensional convergence/divergence (~1°). The resulting data sets are in good agreement for crystals of the same type and are also in good agreement with data sets from conventional oscillation methods. Though these agreements are of modest to good quality, it is an important step towards establishing CBM as a viable data-collection technology.

In CBM experiments, the integrated intensity of each reflection is weighted by a source factor which is not directly observed. Similar to fitting a λ-curve in pseudo-Laue data processing, fitting a model function of a two-dimensional source factor is an important source of errors. Since the λ-curve is a one-dimensional function which can be independently verified by experiment, our task at hand here is more difficult. As explained in §4.4, we used a simple model function to describe the two-dimensional source factor, which is a product of two one-dimensional functions (9). This implies that the source intensity varies with φ in the same way at all μ levels. Though we do not have direct observation of the source intensity angular distribution, the images of the direct beam over large distances provided us with somewhat convoluted pictures. Inspection of these direct-beam images indicated to us that this is an over-simplification. This simple form was used as a necessary compromise to avoid over-parameterization in the least-squares scaling procedure.

It is important to note that the number of successfully integrated reflections is smaller than that estimated in the data-collection strategy session. This discrepancy is partly because of the fact that a different and more stringent criterion for assessing overlapping is used in the actual integration of diffraction spots (see §4.3). Another source of this discrepancy is the aberration (mis-focusing) of the optic lenses, which affects the actual sizes of spot tangential lengths. This would cause errors in the peak and background assignment of pixels and hence cause errors in integrated intensities. In severe instances, it would even result in rejection of the reflections (negative intensities), particularly for weaker reflections.

Improvement in the quality of the convergent source will mitigate the problems described above. A convergent source with low circular variation (or high isotropy) can be easily modeled by a simple function with fewer parameters. A source with low optical aberration would also make the prediction of the shapes and locations of the diffraction spots more precise. Further improvement in the source angular distribution could also increase the data-collection efficiency. An ideal distribution is one that forms a step function radially, *i.e.* a top-hat-

shaped two-dimensional angular distribution, in which the source factor falls slowly along the radial directions, and hence more diffraction spots can be adequately measured. We believe that with continuing improvements in X-ray optics over time, the availability of quality convergent sources will make CBM more attractive.

Though it is not the main focus of this report, it is always interesting to compare the efficiencies of different experiments. It should be pointed out that it is difficult to perform a fair and controlled comparison for data sets generated from two completely different methods, particularly when no control over the crystal sample sizes and their diffraction qualities was maintained. If we use the product, PT, of the power P of the generator power settings and the total exposure time T as a general guide for X-ray usage, then the PT values for data sets LCBM1, LCBM2 and LOM were 0.53, 0.42 and 70.0 kW h (42 frames of 20 min exposure each at 5 kW for LOM), respectively. This puts CBM at an advantage by two orders of magnitude over OM. However, it has been reported (Bloomer & Arndt, 1999) that a combination of a micro X-ray source and a single-capillary optic gave a performance rather close to this scale at a low convergence angle, where a conventional oscillation method can be used. Generally speaking, a convergent beam should drastically outperform a near-collimated beam on this score. The developmental micro X-ray source tubes used in this study were extremely unstable and the focus-spot sizes were greater than expected. This no doubt affected the performance of the experimental setup tremendously. The higher collection angles of polycapillary lenses (compared with single capillary optics) also pose a difficult challenge to the manufacturing process. At present, the transmission factor of the outer fibres decreases more quickly than is desired. Reducing the fibre inner diameter and other improvements in manufacturing process could potentially increase the transmission factor at high convergent angles, which is beyond the scope of this report.

In conventional oscillation methods, increasing the crystal-to-detector distance usually improves signal-to-background ratios and reduces overlap effects. However, in CBM this only slightly improves spot separation along the radial direction, while at the same time significantly worsening spot separation tangentially. A technique widely considered to be helpful in improving signal-to-noise ratios, the profile-fitting technique, has been successfully applied in oscillation methods, although there is difficulty in properly differentiating fully recorded and varying partially recorded reflections. This difficulty is magnified many-fold in CBM because of the variation in the tangential lengths of diffraction spots. Therefore, it has not been implemented in the current software package.

The convergent-beam method takes advantage of the high flux on the crystal sample provided by a convergent source, thus increasing the data-collection efficiency, which is important for macromolecular crystallography where the samples are weak diffractors. It can also find its applications in cases such as neutron diffraction, where the source flux is a rate-determining factor, and X-ray microdiffraction, where samples are extremely small and have low diffracting power.

This work was supported by the Microgravity and Science Applications Division of the National Aeronautics and Space Administration (NASA) through Contracts NAS8-97247 and NAS8-39926. We would like to thank B. S. Wright for preparing the crystals used in this study. X-ray Optical System Inc. is acknowledged for providing capillary optic lenses used in this study. Thanks are also due to Mr R. C. Sisk for arranging a motorized three-dimensional translational stage with controller on loan from the Marshall Space Flight Center. One of the reviewers is acknowledged for bringing our attention to the polarization factors, which resulted in adding a new section (§2.5).

References

Azároff, L. V. (1955). *Acta Cryst.* **8**, 701–704.
Bloomer, A. C. & Arndt, U. W. (1999). *Acta Cryst.* D**55**, 1672–1680.
Ho, J. X., Snell, E. H., Sisk, R. C., Ruble, J. R., Carter, D. C., Owens, S. M. & Gibson, G. M. (1998). *Acta Cryst.* D**54**, 200–214.
Li, P.-W. & Bi, R.-C. (1998). *J. Appl. Cryst.* **31**, 806–811.
MacDonald, C. A. & Gibson, W. M. (2001). *Handbook of Optics*, 2nd ed., Vol. III. New York: McGraw–Hill.
Nikonov, S. V. & Chirgadze, Yu. N. (1985). *Acta Cryst.* A**41**, 195–203.
Owens, S. M., Ullrich, J. B., Ponomarev, I. Yu., Carter, D. C., Sisk, R. C., Ho, J. X. & Gibson, W. M. (1996). *Proc. SPIE*, **2859**, 200–209.
Rossmann, M. G. & van Beek, C. G. (1999). *Acta Cryst.* D**55**, 1631–1640.
Verman, B., Kim, B., Jiang, L., Carosella, J., McGill, J., Grupido, N., Wilcox, D., Smith, R., Ferarra, J., Crane, K., Klein, C., Swepston, P., Yang, C., Harada, J., Kuribayashi, M., Omote, K. & Yamano, A. (2000). Abstr. Annu. Meet. Am. Crystallogr. Assoc. Abstract P230.
Xuong, N.-H. & Freer, S. T. (1971). *Acta Cryst.* B**27**, 2380–2387.