

Automatic crystal centring procedure at the SSRF macromolecular crystallography beamline

Zhijun Wang, Qiangyan Pan, Lifeng Yang, Huan Zhou, Chunyan Xu, Feng Yu, Qisheng Wang, Sheng Huang and Jianhua He*

Shanghai Institute of Applied Physics, Chinese Academy of Sciences, Shanghai 201204, People's Republic of China.
*Correspondence e-mail: hejh@sinap.ac.cn

Received 22 March 2016

Accepted 13 September 2016

Edited by Y. Amemiya, University of Tokyo, Japan

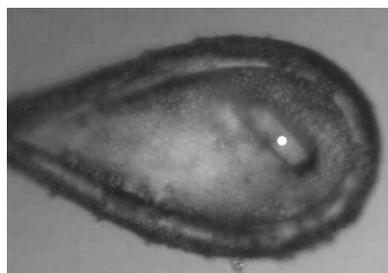
Keywords: crystal centre; algorithm; recognition; crystallography; loop centre.

X-ray diffraction is a common technique for determining crystal structures. The average time needed for the solution of a protein structure has been drastically reduced by a number of recent experimental and theoretical developments. Since high-throughput protein crystallography benefits from full automation of all steps that are carried out on a synchrotron beamline, an automatic crystal centring procedure is important for crystallographic beamlines. Fully automatic crystal alignment involves the application of optical methods to identify the crystal and move it onto the rotation axis and into the X-ray beam. Crystal recognition has complex dependencies on the illumination, crystal size and viewing angles due to effects such as local shading, inter-reflections and the presence of antifreezing elements. Here, a rapid procedure for crystal centring with multiple cameras using region segment thresholding is reported. Firstly, a simple illumination-invariant loop recognition and classification model is used by slicing a low-magnification loop image into small region segments, then classifying the loop into different types and aligning it to the beam position using feature vectors of the region segments. Secondly, an edge detection algorithm is used to find the crystal sample in a high-magnification image using region segment thresholding. Results show that this crystal centring method is extremely successful under fluctuating light states as well as for poorly frozen and opaque samples. Moreover, this crystal centring procedure is successfully integrated into the enhanced *Blu-Ice* data collection system at beamline BL17U1 at the Shanghai Synchrotron Radiation Facility as a routine method for an automatic crystal screening procedure.

1. Introduction

With the development of enhanced radiation light sources as well as faster detectors, the speed of crystallographic data collection has been greatly increased in recent years (Svensson *et al.*, 2015; Winter & McAuley, 2011; Okazaki *et al.*, 2008). Programmed sample changers can additionally decrease the time required for screening (Hiraki *et al.*, 2013; le Maire *et al.*, 2011; Bowler *et al.*, 2015; Cipriani *et al.*, 2006). With significant developments in crystal sample preparation using high-throughput crystallization facilities, there is a need for faster screening systems (Karain *et al.*, 2002; Cipriani *et al.*, 2006). However, accurate location of the crystal sample is a time-consuming step, and mainly carried out manually or semi-automatically prior to the data collection at many crystallography beamlines (Girard *et al.*, 2006; Karain *et al.*, 2002).

Algorithms to identify the loop or crystal based on histogram analysis and align the centre of loop or crystal to the X-ray beam position are reasonably reliable when the crystal size is comparable with that of the loop. Several programs for automatic crystal centring or loop centring have been reported



in order to help in faster data collection (Khajepour *et al.*, 2013; Andrey *et al.*, 2004; Dera & Katrusiak, 1999; Pauluhn *et al.*, 2011; Pothineni *et al.*, 2006; Lavault *et al.*, 2006; Cherezov *et al.*, 2009). However, these centring programs would fail under fluctuating light or when the crystal only presents faint edges. Unfavourable fluctuating light state, poorly frozen and opaque samples make crystal recognition quite difficult (Song *et al.*, 2007; Jain & Stojanoff, 2007).

Several approaches have been developed to confront the problem of automatic crystal centring. Automatically aligning the crystal to the X-ray beam using raster scans has been reported for finding small crystals (Song *et al.*, 2007; Aishima *et al.*, 2010). However, the time required for raster scans can vary from a few minutes to one hour per sample depending on the number of grid points and the readout speed of the detector (Hilgart *et al.*, 2011). Ultraviolet light has been used to recognize crystals using the fluorescence emission that the majority of proteins exhibit, when exposed to light of wavelengths between 250 nm and 290 nm, for proteins containing aromatic ring structures, for example tryptophan, tyrosine and phenylalanine (Pohl *et al.*, 2004; Chavas *et al.*, 2011; Madden *et al.*, 2013). However, this method needs accurate control of the exposure time in order to avoid UV-induced damage (Celie *et al.*, 2009). Nonlinear optical (NLO) instrumentation has been integrated with synchrotron X-ray diffraction (XRD) for crystal centring (Madden *et al.*, 2013; Kestur *et al.*, 2012; Kissick *et al.*, 2013); however, this method requires purchasing expensive NLO instrumentation.

At the Shanghai Synchrotron Radiation Facility (SSRF), the BL17U1 beamline is currently in operation for macromolecular crystallography. BL17U1 uses an in-vacuum undulator as photon source. A high-precision goniometer (Crystal Logical Inc.), an ACTOR robot (Rigaku Inc.) and a Q315 detector (ADSC Inc.) are used for data collection (Wang *et al.*, 2015; He & Gao, 2015). An upgrade to a fully automated screening procedure has been achieved at BL17U1. Here, we report a method that is used at SSRF macromolecular crystallography beamline BL17U1 to decrease the time spent on the data collection by assisting in crystal centring, after the crystal sample is mounted by the ACTOR robot arm. Firstly, an illumination-invariant loop centring and classification model is used by slicing the loop image into small region segments, then a feature vector of region segments is calculated, and the loop is classified into different loop types. Secondly, an edge detection algorithm is used to find the crystal sample using a region segment thresholding method. Results show that our crystal centring procedure is extremely

effective under fluctuating light states as well as for poorly frozen and opaque samples.

This paper is organized as follows: §2 describes the crystal centring procedure, §3 describes the integration of the crystal centring procedure into the beamline control system and the results, and the paper is concluded in §4.

2. Crystal centring procedure

The viewing system at beamline BL17U1 comprises complex optical microscopes and cameras. Fig. 1 shows the hardware setup of the automated centring system at BL17U1 at SSRF. It includes complex on-axis view video systems as inline cameras (inline camera B and inline camera C). In order to decrease lighting effects from the collimator, another video system is installed above sample point I as the sample camera (sample camera A), which is used for automated crystal centring. The field of view of low-magnification inline camera C is 4.6 mm × 3.1 mm (H × V), and for high-magnification camera A is 0.26 mm × 0.17 mm (H × V). On-axis view cameras B and C are back-illuminated by a film E mounted on the movable beamstop. Sample camera A is back-illuminated by a condenser lamp light source D, which is extremely stable. Therefore it is not necessary to remove lighting effects for sample camera A. Another LED light source F is used for lighting nearby the sample position. The crystal centring procedure is separated into automated loop centring and automated crystal centring parts. We describe these in §§2.1 and 2.2, respectively.

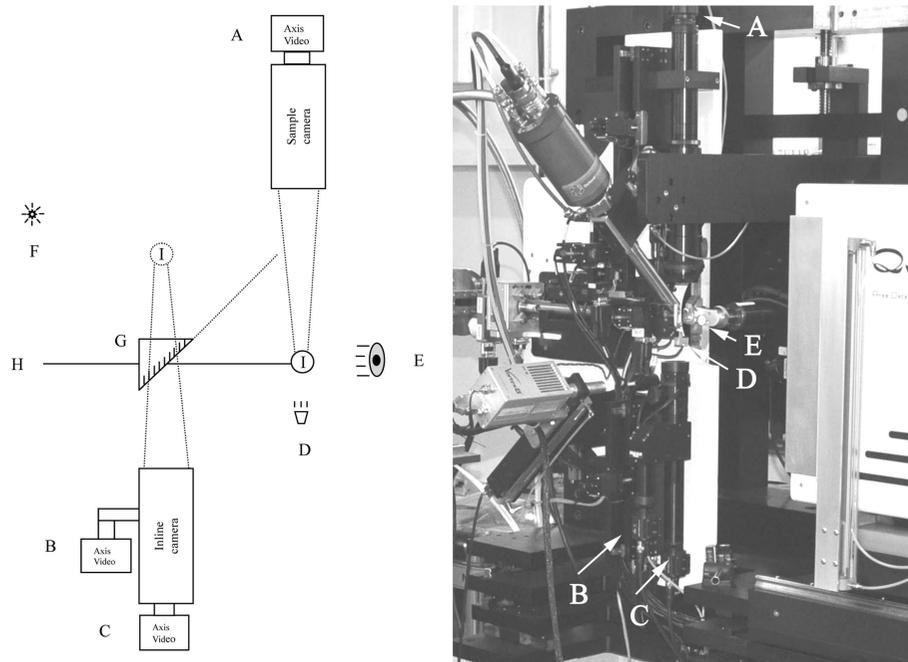


Figure 1 A multi-camera system at BL17U1, SSRF, for automatic crystal centring. A: axis video server used as sample camera at high-magnification level. B: axis video server used as inline camera at high-magnification level. C: axis video server used as inline camera at low-magnification level. D: backlight used for sample camera. E: backlight (reflection film attached to moveable beamstop) for inline camera. F: front light used for lighting sample. G: mirror used for inline camera to reflect crystal sample. H: X-ray beam to sample. I: crystal sample.

2.1. Automatic loop centring

The outline for automatic loop centring using low-magnification inline camera C is shown in Fig. 2(a). The automatic loop centring procedure contains the following steps: (1) obtain image from low-magnification inline camera C; (2) obtain position and geometric information of the loop; (3) accurately move loop to beam position; (4) rotate crystal by 90°; (5) retrieve images from low-magnification inline camera C, and repeat steps (2)–(4). In order to have a large field of view for loop centring, the loop centring procedure is first performed using the low-magnification inline camera C. Subsequently, it is repeated using the high-magnification sample camera A and then the loop is centred to the beam position.

It is very important to obtain position and geometric information of the loop from the low-magnification inline camera C, since this is highly associated with the success of the loop centring procedure.

An illuminate-invariant loop recognition algorithm is important for inline camera C because there are several beamline components (for example, collimator, slit and beamstop) that can significantly affect the lighting of inline camera C; additionally, optical distortions produced by the

cryostream can introduce speckle noise, and slightly shift light conditions at different times.

Here an illuminate-invariant loop recognition and classification algorithm is developed based on feature extraction from region segments. Then the loop is aligned to the beam position and classified into different types for further crystal recognition.

2.1.1. Loop feature extraction and classification. A novel loop feature extraction and classification model is described in this section. It is used for loop alignment and classification according to information extraction from low-magnification inline camera C. Loops are classified as either nylon cryoloops, polymer microloops or icy loops.

Given a low-magnification background image X from inline camera C, the low-magnification image X is sliced into several important region segments. Let

$$X = (X_1, X_2, \dots, X_i, \dots, X_m), \quad i = 1, \dots, m, \quad (1)$$

represent region segments of X_1, X_2, \dots, X_m from inline camera C. Fig. 3 shows a typical map of region segments of low-magnification image X from camera C. In the map, parts A and B are not included in loop feature extraction and classification due to problems of field of view from inline camera C and shadow of the beamline components. After parts A and B are removed from the image, and the remaining part is sliced into 588 region segments (20×20 pixels for each) for loop tip identification, part C is sliced into 650 region segments (2×2 pixels for each), which are used for loop feature extraction and classification.

Given a low-magnification image Y with loop from inline camera C, the low-magnification image Y is sliced into the same region segments as image X . Let

$$Y = (Y_1, Y_2, \dots, Y_i, \dots, Y_m), \quad i = 1, \dots, m, \quad (2)$$

represent a sliced loop image Y_1, Y_2, \dots, Y_m from inline camera C. The intensity of every pixel in the region segment of X_i or Y_i is expressed as $X_i(x, y)$ or $Y_i(x, y)$. $X_i(x, y)$ or $Y_i(x, y)$

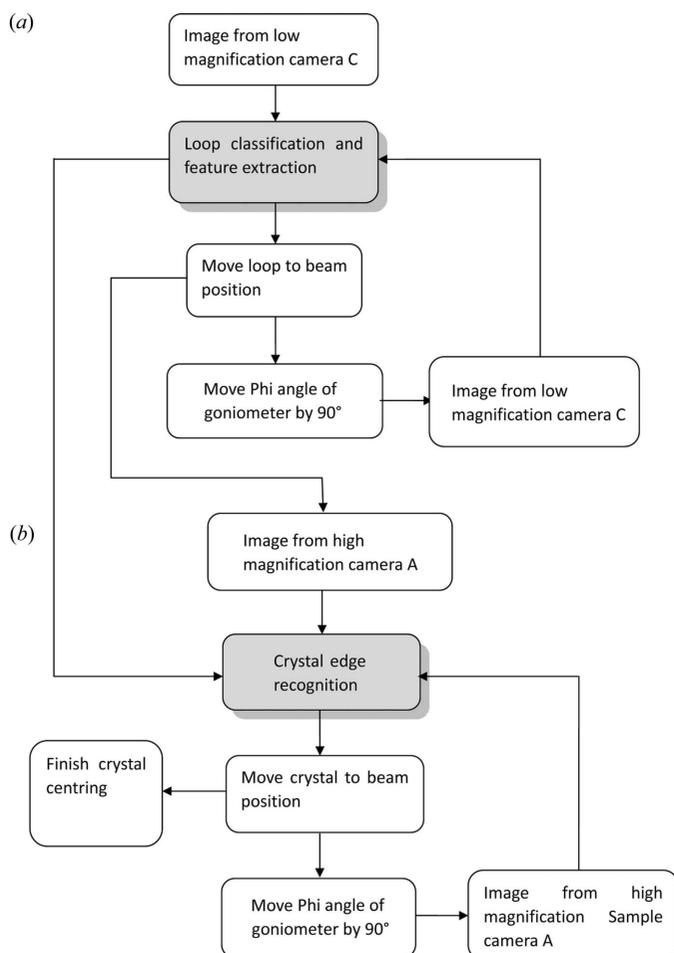


Figure 2 Automatic loop and crystal centring procedure.

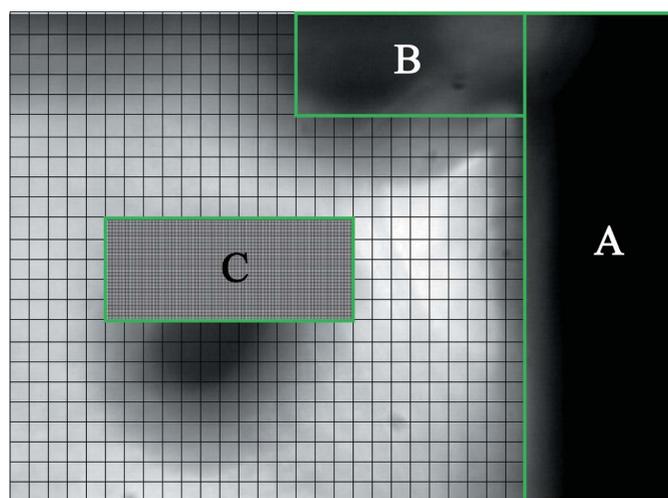


Figure 3 A typical map of region segments of an image obtained using the low-magnification inline camera.

is dependent on the intrinsic reflectivity of the loop $L_i(x, y)$ or $R_i(x, y)$, and the lighting conditions $I_i(x, y)$. According to Land's retinex algorithm (Land & McCann, 1971), $X_i(x, y)$ and $Y_i(x, y)$ can be expressed as

$$X_i(x, y) \approx L_i(x, y) I_i(x, y), \quad (3)$$

$$Y_i(x, y) \approx R_i(x, y) I_i(x, y). \quad (4)$$

It has been observed that the illumination information $I_i(x, y)$ is present at low spatial frequencies while the intrinsic reflectance information $L_i(x, y)$ or $R_i(x, y)$ is present at higher spatial frequencies. Using this observation we can assume that, over a small region segment, illumination $I_i(x, y)$ is approximately constant. We propose an illumination model which separates the effects of illumination over a small region segment of the loop image into two components.

Let φ represent a small region segment in a given loop image. The pixel intensity $X_i(x, y)$ and $Y_i(x, y)$ of region segment φ can be expressed as

$$X_i(x, y) \approx k_1 L_i(x, y) + c_1, \quad (5)$$

$$Y_i(x, y) \approx k_2 R_i(x, y) + c_2. \quad (6)$$

The region segment of an image is separated into two components. The first component $k_1 L_i(x, y)$ or $k_2 R_i(x, y)$ is multiplicative, modelling the change of the variance of the pixel intensity in the region segment. The second component c_1 or c_2 is additive, modelling the change of the mean pixel intensity in the region segment, which varies with illumination.

c_1 or c_2 can be removed by equations (7) and (8),

$$X_i(x, y) - \bar{X}_i(x, y) = k_1 L_i(x, y), \quad (7)$$

$$Y_i(x, y) - \bar{Y}_i(x, y) = k_2 R_i(x, y), \quad (8)$$

where $\bar{X}_i(x, y)$ and $\bar{Y}_i(x, y)$ represent the mean pixel intensity of region segments X_i and Y_i , respectively. Then $L_i(x, y)$ and $R_i(x, y)$ can be expressed as

$$L_i(x, y) = k_\varphi [X_i(x, y) - \bar{X}_i(x, y)], \quad (9)$$

$$R_i(x, y) = k_l [Y_i(x, y) - \bar{Y}_i(x, y)]. \quad (10)$$

In this paper, our loop identification can be viewed as finding the cosine correlation coefficient vector of intrinsic reflectance information between background and loop images, which is approximated by the collection of the cosine similarities between their intrinsic reflectivity pixel intensities of region segments. The cosine correlation coefficient vector can be represented as

$$V_{C(x,y)} = \begin{Bmatrix} C[L_1(x, y), R_1(x, y)] \\ \vdots \\ C[L_i(x, y), R_i(x, y)] \\ \vdots \\ C[L_m(x, y), R_m(x, y)] \end{Bmatrix}. \quad (11)$$

$V_{C(x,y)}$ is the feature vector of the vertical concatenation of the cosine correlation coefficient of region segments in each

column of the image obtained from camera C. With equations (9)–(10), $V_{C(x,y)}$ can be expressed as

$$V_{C(x,y)} = \begin{Bmatrix} C\{k_\varphi[X_1(x, y) - \bar{X}_1(x, y)], k_l[Y_1(x, y) - \bar{Y}_1(x, y)]\} \\ \vdots \\ C\{k_\varphi[X_i(x, y) - \bar{X}_i(x, y)], k_l[Y_i(x, y) - \bar{Y}_i(x, y)]\} \\ \vdots \\ C\{k_\varphi[X_m(x, y) - \bar{X}_m(x, y)], k_l[Y_m(x, y) - \bar{Y}_m(x, y)]\} \end{Bmatrix}. \quad (12)$$

Apply a 2D Fourier transform to equations (9) and (10). Employ the magnitude of the 2D Fourier transformation of intrinsic reflectance information $L_i(x, y)$ or $R_i(x, y)$ and ignore the zeroth Fourier coefficient $\mathcal{L}(0, 0)$, which is equivalent to subtracting the mean; thus we obtain a model for the magnitude Fourier features of $L_i(x, y)$ and $R_i(x, y)$,

$$\mathcal{F}\{k_\varphi[X_i(x, y) - \bar{X}_i(x, y)]\} \Leftrightarrow k_\varphi[\mathcal{L}_i(u, v) - \mathcal{L}_i(0, 0)], \quad (13)$$

$$\mathcal{F}\{k_l[Y_i(x, y) - \bar{Y}_i(x, y)]\} \Leftrightarrow k_l[\mathcal{R}_i(u, v) - \mathcal{R}_i(0, 0)], \quad (14)$$

where $\mathcal{L}_i(u, v)$ and $\mathcal{R}_i(u, v)$ are the Fourier transforms of $X_i(x, y)$ and $Y_i(x, y)$, respectively. Using the magnitude of the Fourier image representation, equation (12) can be rewritten as

$$V_{C(x,y)} = \begin{Bmatrix} C[k_\varphi \mathcal{L}_1(u, v), k_l \mathcal{R}_1(u, v)] \\ \vdots \\ C[k_\varphi \mathcal{L}_i(u, v), k_l \mathcal{R}_i(u, v)] \\ \vdots \\ C[k_\varphi \mathcal{L}_m(u, v), k_l \mathcal{R}_m(u, v)] \end{Bmatrix}, \quad (u, v \neq 0, 0), \quad (15)$$

where $\mathcal{L}_i(u, v)$ is the magnitude of the Fourier representation of region segment X_i with the mean removed and $\mathcal{R}_i(u, v)$ is the magnitude of the Fourier representation of region segment Y_i with the mean removed.

Equation (15) can be rewritten as

$$V_{C(x,y)} = \begin{Bmatrix} C[\mathcal{L}_1(u, v), \mathcal{R}_1(u, v)] \\ \vdots \\ C[\mathcal{L}_i(u, v), \mathcal{R}_i(u, v)] \\ \vdots \\ C[\mathcal{L}_m(u, v), \mathcal{R}_m(u, v)] \end{Bmatrix}, \quad (u, v \neq 0, 0), \quad (16)$$

because the cosine correlation coefficient is invariant to a scale factor.

Then $V_{C(x,y)}$, as shown in equation (16), represents the feature vector of the loop image from low-magnification camera C. Subsequently, the loop tip is determined according to loop feature vector $V_{C(x,y)}$, and then the loop tip is aligned to the beam position.

The benefit of the feature vector is that our method is not only effective under fluctuating light states but also effective in centring oblique loops to the beam position.

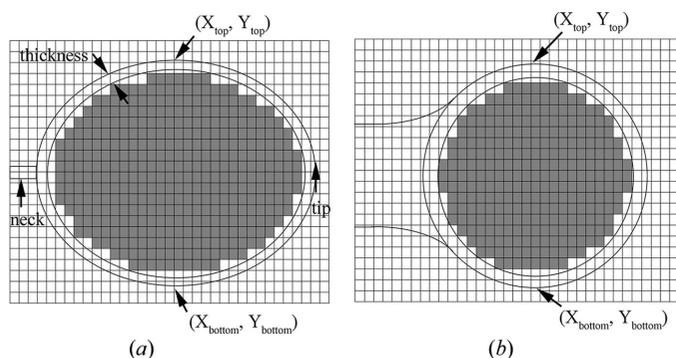


Figure 4
Geometric definition for different loop types. (a) Nylon cryoloop. (b) Polymer microloop. Geometric parameter vector ω is the horizontal concatenation of the value of each column of region segments in the image: the value is 1 for each shadow region segment, and 0 for each white region segment. $(X_{\text{top}}, Y_{\text{top}})$ and $(X_{\text{bottom}}, Y_{\text{bottom}})$ are the top and bottom pixels of the loop margin for loop centre identification.

In order to assist the crystal centring in the following, the loop is classified into different types according to $V_{C(X,Y)}$ from low-magnification camera C. Loop classification is represented as computing the Euclidean distance metric D [equation (17)], and then comparing it with a threshold, τ ,

$$\arg \min_k D[V_{C(X,Y)}^{(k)}, V_{C(X,Y)}], \quad k = 1, \dots, K. \quad (17)$$

$V_{C(X,Y)}^{(k)}$ is the feature vector derived from equation (16) for different loop images, which represents class k , and $V_{C(X,Y)}$ is the feature vector of the input loop image.

After the loop tip is aligned to the beam position with low-magnification camera C, the loop tip position is also aligned to the beam position with high-magnification camera A, because the centres of camera C and camera A are pre-aligned to the beam position.

A typical loop centre position is determined as shown in Fig. 4 with camera A. The top edge pixel position $(X_{\text{top}}, Y_{\text{top}})$ and bottom edge pixel position $(X_{\text{bottom}}, Y_{\text{bottom}})$ are determined by directly searching the pixel position of the image from camera A in the horizontal and vertical direction. Then the loop centre position C_{loop} can be calculated as

$$C_{\text{loop}} = \left(\frac{X_{\text{bottom}} - X_{\text{top}}}{2}, \frac{Y_{\text{bottom}} - Y_{\text{top}}}{2} \right). \quad (18)$$

2.2. Automatic crystal centring

It is important to further centre the crystal using the high-magnification sample camera A, which includes acquisition images from high-magnification sample camera A and recognition of crystal edges. After crystal edge information is determined, the crystal is accurately moved to the beam position. Crystal edge detection from the acquired image is a fundamental problem in crystal recognition and centring.

Edge detection can be performed based on the gradient using different thresholding methods (Yazid & Arof, 2013). There are many variations of the thresholding approach (Zhang *et al.*, 2013; Patankar *et al.*, 2013). The simplest method

is to select the edge pixels with their gradient magnitude using the threshold. Edge pixels with a gradient magnitude below the lower threshold or above the upper threshold are removed or kept, respectively.

However, it is very difficult to choose a good threshold which can filter out most of the noise while keeping most of the useful crystal edges. Some crystals have very faint edges, and therefore crystal recognition from the loop is not enough *via* a threshold.

The outline for crystal centring using sample camera A is shown in Fig. 2(b). The backlight system D for sample camera A is extremely stable. Therefore it is not necessary to subtract a background image for camera A.

Here loop type information from the loop centring and geometry parameter vector for different classified loop types are used for crystal edge detection with region segment thresholding. Our procedure, which is used for extracting the centre position of the crystal, is described as follows.

2.2.1. Region segment thresholding. The absolute value of the gradient magnitude is not the only criterion for evaluating the saliency of a region segment for crystal edge identification. Its spatial distribution is also an important clue as to whether an edge is a crystal boundary or due to noise or the texture of the loop material.

In order to improve the robustness of crystal edge detection, a crystal region segment thresholding based on saliency value instead of edge pixels is proposed, *i.e.* whether to remove or to keep the whole clustered region segments based on their saliency values. This also makes our approach more robust for detecting crystal edges with weak boundaries. In our method, edges are formed by region segments and thresholding is based on saliency values. Some individual edge pixels may not be sufficient to distinguish a crystal from noise, but by grouping the pixels into a region segment the distinction becomes more reliable.

2.2.2. Saliency value of individual region segment. The saliency value of an individual region segment is determined as follows. For a region segment R_p with gradient magnitude V_g , we search along the gradient direction to find the closest region segment R_{p1} with a larger gradient magnitude. The distance between two region segments is represented as V_{d1} , and the gradient magnitude of R_{p1} is represented as V_{p1} . We then search in the opposite direction to find the closest region segment R_{p2} with a gradient magnitude larger than V_g . The distance between R_{p2} and R_p is represented as V_{d2} , and the gradient magnitude of R_{p2} is represented as V_{p2} .

Because the larger gradient magnitude side of a crystal boundary is often very cluttered, the region segment with larger gradient magnitude between R_{p1} and R_{p2} is selected according to the saliency range V_r , which can be expressed as

$$V_r = \max(V_{p1}, V_{p2}). \quad (19)$$

If $V_{p1} > V_{p2}$, the average gradient magnitude of the region segments on the line R_{pp1} is computed as $V_{\bar{g}}$, and the saliency range between R_p and R_{p1} is represented as $V_{\bar{d}}$, with $V_{\bar{d}} = V_{d1}$. If $V_{p1} \leq V_{p2}$, the average gradient magnitude of the region

segments on the line R_{pp2} is computed as $V_{\bar{g}}$, and the saliency range between R_p and R_{p1} is represented as $V_{\bar{a}}$, with $V_{\bar{a}} = V_{d2}$.

The saliency value vector of a region segments is computed as

$$V_{S(x,y)} = (V_g - V_{\bar{g}})^\alpha (V_{\bar{a}})^\beta \omega \nabla_{C(x,y)}. \quad (20)$$

The parameters α and β control the relative importance between the gradient magnitude and the neighbour range for crystal recognition. α and β can be set at different values according to the different crystal types. For example, to enhance the crystal boundary in the image, it is better to use a larger α and β . Empirically, α and β are set to 1 for normal crystal region segments, and are set to 2 to enhance faint crystal region segments. $\nabla_{C(x,y)}$ is the feature vector of vertical concatenation of the cosine correlation coefficient of region segments in each column of the image obtained from the high-magnification sample camera A.

Because sample camera A is extremely stable, it is not necessary to consider the illumination problem for this camera, and $\nabla_{C(x,y)}$ is calculated as

$$\nabla_{C(x,y)} = \left\{ \begin{array}{c} C\{[\bar{X}_1(x,y) - \bar{X}_1(x,y)], [\bar{Y}_1(x,y) - \bar{Y}_1(x,y)]\} \\ \vdots \\ C\{[\bar{X}_i(x,y) - \bar{X}_i(x,y)], [\bar{Y}_i(x,y) - \bar{Y}_i(x,y)]\} \\ \vdots \\ C\{[\bar{X}_m(x,y) - \bar{X}_m(x,y)], [\bar{Y}_m(x,y) - \bar{Y}_m(x,y)]\} \end{array} \right\}. \quad (21)$$

$\bar{X}_i(x,y)$ and $\bar{Y}_i(x,y)$ represent pixels in the region segments of the crystal image and background image from sample camera A. $\bar{X}_i(x,y)$ and $\bar{Y}_i(x,y)$ are the average pixel intensity of the region segment obtained from sample camera A.

ω is the geometry parameter vector for different classified loop types in order to exclude the area outside of the loop, and is determined according to the distribution of $\nabla_{C(x,y)}$. For each column of region segments in the image, search from the top downwards; if $C\{[\bar{X}_i(x,y) - \bar{X}_i(x,y)], [\bar{Y}_i(x,y) - \bar{Y}_i(x,y)]\} < 0.9$, then the top region segment of that column is expressed as k ; search from the bottom upwards: if $C\{[\bar{X}_i(x,y) - \bar{X}_i(x,y)], [\bar{Y}_i(x,y) - \bar{Y}_i(x,y)]\} < 0.9$, then the bottom region segment of that column is expressed as j . The region segments between k and j are marked as the searching range for that column.

In order to reduce the effects of the loop edge, for a nylon cryoloop the loop thickness is removed from the searching range as shown in Fig. 4(a); for a polymer microloop, the loop thickness and neck are removed from the searching range as shown in Fig. 4(b). For icy samples, region segments between k and j are directly marked as the searching range for that column.

If region segments belong to the searching length of that column, then the value of the geometry parameter of the region segment is marked as 1; otherwise it is marked as 0.

ω is the horizontal concatenation of each column of geometry parameters of region segments.

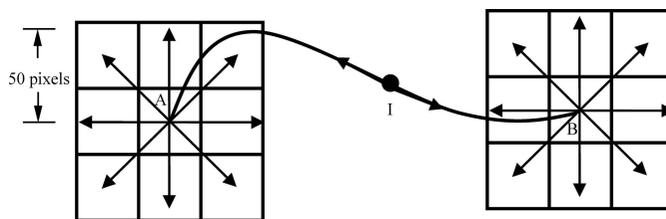


Figure 5
A typical method for forming clustered region segments. From the region segment (I) with the highest saliency value, the region segment grows by including the region segment with most significant saliency value among neighbours of two endpoints (A, B) with 16 direction (8 direction for each endpoint). In each direction the maximum searching length is 50 pixels.

2.2.3. Crystal edge detection using the saliency value.

Region segments are clustered into crystal edges using the following method. From the region segment with the highest saliency value, the region segment grows by including the region segment with the most significant saliency value among the neighbours of the two endpoints (Fig. 5), forming clustered region segments. Growing of the region segment is ended when no more region segments can be added to the clustered region segments according to the saliency value within the searching distance of 50 pixels from each direction in the two endpoints. Then a new region segment is started with the highest saliency value among the remaining region segments that have not been included in a clustered region segment. Until no more new clustered region segments can be generated, all clustered region segments are threshold by the saliency value for crystal edge detection. To improve speed, the search stops if the search arrives at the marginal of loops. The marginal of loops is determined by the geometry parameter vector ω .

For region segment thresholding, a lower threshold L and an upper threshold H are used. The clustered region segment with saliency values above H is kept; the clustered region segment with saliency values lower than 50% of L is removed.

Finally, the clustered region segment is used for crystal centre identification. The centre of the crystal is determined according to the average positions of all clustered region segments. If N crystal region segments are found, the centre position of each crystal region segment can be expressed as $C(X_i, Y_i)$, then the crystal centre position C_{crystal} is expressed as

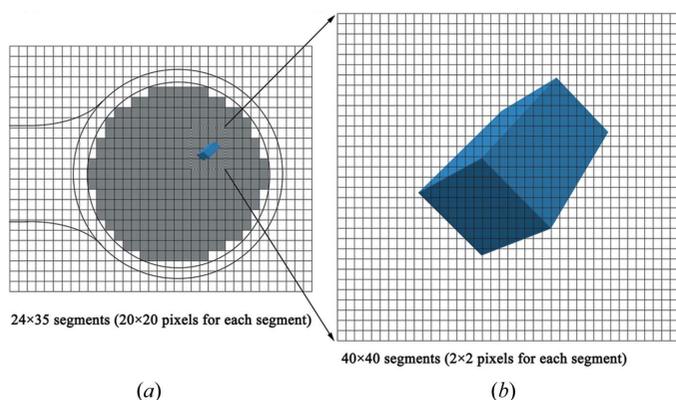
$$C_{\text{crystal}} = \left(\frac{\sum_{i=1}^N X_i}{N}, \frac{\sum_{i=1}^N Y_i}{N} \right). \quad (22)$$

To improve the speed of determination of the crystal centre, firstly the crystal centre is detected using a size of region segments of 20×20 pixels for each; subsequently, smaller region segments (2×2 pixels for each) around the crystal edge are used to search for the crystal centre again using the method as described in Fig. 6, and the clustered region segments can be selected using the threshold. Therefore, a more accurate position of the crystal centre can be obtained using equation (22).

Table 1

Average success rate and time used in loop classification and centring under different conditions with more than 1000 automatic loop centrings at BL17U1.

	Success rate (%)	Average time (s)
Loop classification		
Nylon cryoloop	100	0.1
Polymer microloop	100	0.1
Icy loop	98	0.2
Loop centring with feature vector from equation (16)		
Nylon cryoloop	98	15
Polymer microLoop	96	20
Icy loop	90	23
Loop centring with histogram analysis of pixel intensities		
Nylon cryoloop	41	8.2
Polymer microLoop	50	8.1
Icy loop	38	6.2

**Figure 6**

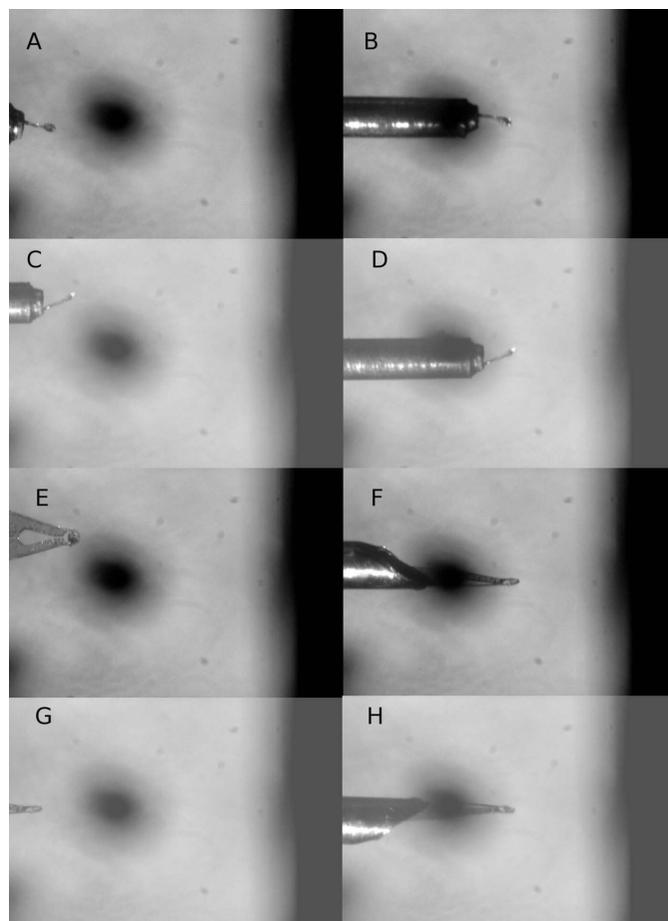
Determination of the crystal centre position with different sizes of region segments. (a) The crystal centre is detected using the size of region segments of 20×20 pixels for each. (b) Smaller region segments (2×2 pixels for each) around the crystal edge are used to search for the crystal centre.

3. Results

3.1. Automatic loop centring

Once a loop is mounted on the goniometer, automatic loop centring is started, and loop recognition is determined according to the image obtained from low-magnification inline camera C (Fig. 1). After the loop position is determined, then the loop is moved to the beam position using the EPICS (Norum, 2002) control system.

Loop recognition and classification are determined by feature vectors according to equation (16); different loop types can be recognized and well classified. The success rates of classification are almost 100% for nylon cryoloops and polymer microloops (Table 1). Our loop-centring technique is stable and quick and is not affected by a fluctuating light state; it does not lead to temporal overrun compared with manual centring. Depending on the initial location, loop centring can be accomplished within an average time of 15 s for nylon cryoloops, 20 s for polymer microloop and 23 s for icy samples. Achievement rates of loop centring are about 98% for nylon cryoloops and 96% for polymer microloops. Icy samples do

**Figure 7**

Loop centring with different loop types (nylon cryoloop, polymer microloop) under fluctuating light states. (A, B) Initial and final position of the nylon cryoloop. (C, D) Initial and final position of the nylon cryoloop under a different lighting state from (A, B). (E, F) Initial and final positions of the polymer microloop. (G, H) Initial and final positions of polymer microloop under a different lighting state from (E, F).

not significantly affect loop centring; however, they can decrease the possibility of centring, with a success rate of 90%.

A comparison of our loop-centring method with the method based on pixel intensities histograms is shown in Table 1. The success rate is only 41% for nylon cryoloops, 50% for polymer microloops and 38% for icy loops.

Fig. 7 shows the initial and final positions of two typical loop types under fluctuating light states. Our results show that the loop-centring procedure is not affected by fluctuating light states.

3.2. Automatic crystal centring

Once loop centring is finished, automatic crystal centring is started. Crystal recognition is determined according to the image obtained from high-magnification sample camera A (Fig. 1). Crystal edge detection using the saliency value with region segment thresholding is analyzed using equation (20).

Fig. 8 shows a typical result of crystal detection using saliency values. From our results, crystal edges can be easily detected using our saliency values, as shown in Fig. 8(B). An

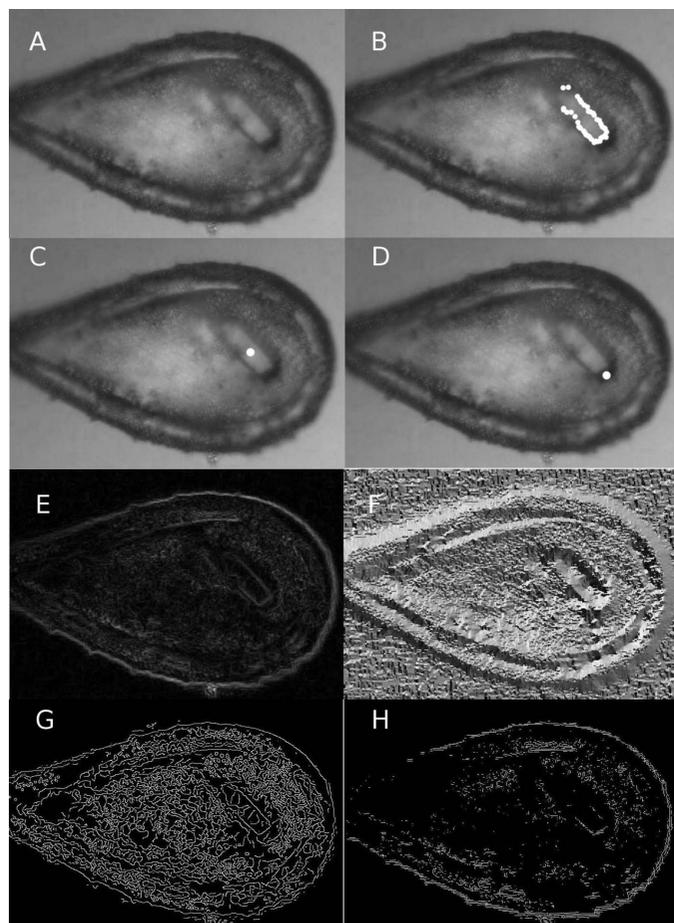


Figure 8
Crystal detection using region segment thresholding. (A) Original image for crystal detection. (B) Crystal edge (white circles) detection with clustered region segments using region segment thresholding with saliency values. (C) Crystal (white circle) detection using saliency equation (22). (D) Crystal (white circle) detection using histogram of pixel intensities. (E) Gradient magnitude of crystal. (F) Gradient direction of crystal. (G) Crystal edge detection according to the Canny algorithm. (H) Crystal edge detection according to the Sobel operator.

important advantage of the saliency value is that it is much easier to find faint crystal edges. Our approach can detect crystal boundaries (Fig. 8B) even if crystal edge pixels have low gradient magnitudes (Fig. 8E).

Our method can identify the crystal centre using equation (22) (Fig. 8C). As can be seen, there are some discrepancies between the identified crystal position (Fig. 8C) and the final recognized position using the pixel intensities histograms (Fig. 8D).

Crystal edge detection with the saliency value has also been compared with the Canny edge detector (Canny, 1986). Fig. 8(G) shows that it is very difficult to find the crystal edge using the Canny edge detector. Our crystal detection method is also compared with the Sobel operator; it is also very difficult to find the crystal edge using the Sobel operator (Fig. 8H).

From Fig. 8 we see that our crystal recognition using the saliency value is extremely effective for finding faint crystal edges with region segment thresholding, because our saliency

value is adjusted by geometric parameters and feature vectors of the loop; therefore our saliency value can effectively enhance faint crystal edges.

After the crystal position is determined, then the crystal is moved to the beam position using the EPICS control system.

3.3. Integration of crystal centring procedure into fully automatic screening system

Our crystal recognition algorithm is integrated into the enhanced *Blu-Ice* beamline control system (McPhillips *et al.*, 2002). Fig. 9 shows the operation interface of the crystal centring service in the enhanced *Blu-Ice* control system.

Loop centring and crystal centring services are deployed into standalone applications. Once a user clicks the ‘Center Loop’ or ‘Center Crystal’ button, *Blu-Ice* sends out commands to the centring server to execute the respective standalone program.

At the beginning of loop centring the image is checked for information content, and if no loop is detected the user is requested to check the loop and the program is interrupted. Loop centring is automatically stopped if the motor movement exceeds a distance larger than 2 mm, which is designed to prevent moving the sample out of the cryostream.

If a user clicks the ‘Center Crystal’ button, loop type and feature vectors are determined, and the loop will be aligned to the beam position. Subsequently, the crystal position is determined according to the saliency value with region segment thresholding, and then the crystal will be aligned to the beam position. If a user clicks the ‘Center Loop’ button, only the loop will be aligned to the beam position. If the crystal is prepared within lipidic cubic phases, a raster scan can

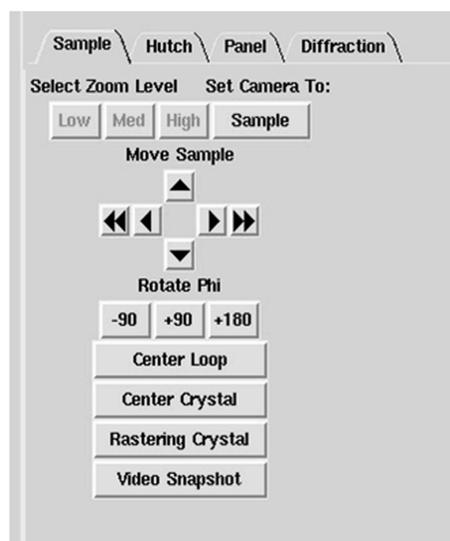


Figure 9
Interface of the loop centring and crystal centring services in the enhanced *Blu-Ice* control system at BL17U1. Our loop centring and crystal centring algorithms have been integrated into *Blu-Ice* as standalone applications through a centring server. Moreover, raster scanning of a crystal prepared within lipidic cubic phases is used to complement our automatic crystal centring method.

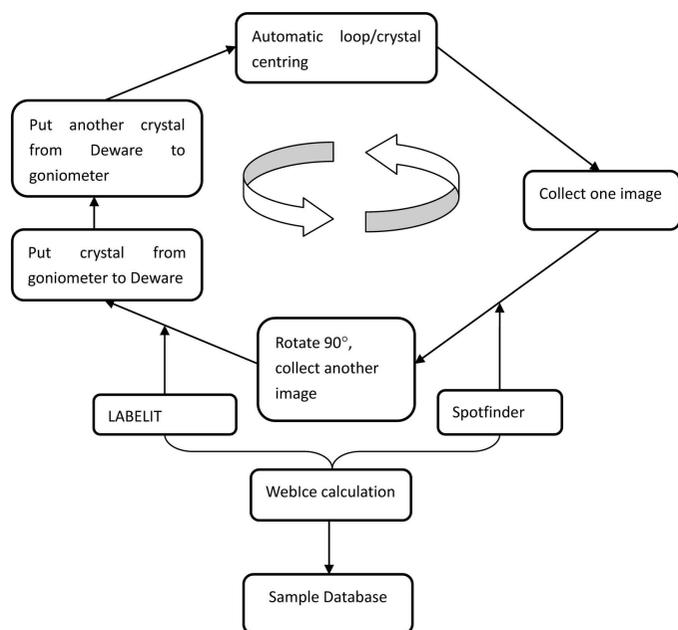


Figure 10

Fully automatic crystal screening procedure at macromolecular crystallography beamline BL17U1. Graphical overview of the automated characterization of screening on BL17U1. All steps in the process are shown. Sample mounting/dismounting and crystal centring, data collection are performed in sequence. Crystal centring starts once a sample has been mounted and then centred in the X-ray beam. Information about the samples, dewar tracking and the calculation of results are automatically saved in the crystal sample database.

be used as a complement to crystal centring through clicking the ‘Rastering Crystal’ button (Fig. 9).

Moreover, our crystal centring procedure has been integrated with the programmed sample changer ACTOR robot to be a fully automatic screening system (Fig. 10). The automatic screening procedure is described as follows: mount crystal on goniometer, centre crystal to beam position, collect one image, rotate Φ angle of goniometer by 90° , collect another image and then dismount the screened crystal and mount another crystal. Data analysis is performed using the *WebIce* server (González *et al.*, 2008), the spot number in the diffraction image is analyzed with *Spotfinder*, and the index of the crystal is calculated using *Labelit* (Sauter & Zwart, 2009). Screening results are saved into the crystal sample database, from where a user can download them. Our screening system is extremely effective in ligand screening for drug discovery.

4. Conclusion

We have set up an effective crystal centring procedure for automatic screening at SSRF macromolecular crystallography beamline BL17U1. The time required for automatic crystal centring is as fast as manually centring a crystal. Our method can be easily adapted into other macromolecular crystallography beamlines. However, our approaches will fail to centre microcrystals prepared within turbid matrices; an

additional raster centring should be employed for turbid matrices.

Acknowledgements

This work was supported by the Major State Basic Research Development Project in China (No. 2011CB911102), Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDB08030101) and National Natural Science Foundation of China (No. Y615141061). The authors are grateful for helpful discussions with Professor Thomas Earnest.

References

- Aishima, J., Owen, R. L., Axford, D., Shepherd, E., Winter, G., Levik, K., Gibbons, P., Ashton, A. & Evans, G. (2010). *Acta Cryst.* **D66**, 1032–1035.
- Andrey, P., Lavault, B., Cipriani, F. & Maurin, Y. (2004). *J. Appl. Cryst.* **37**, 265–269.
- Bowler, M. W., Nurizzo, D., Barrett, R., Beteva, A., Bodin, M., Caserotto, H., Delagenière, S., Dobias, F., Flot, D., Giraud, T., Guichard, N., Guijarro, M., Lentini, M., Leonard, G. A., McSweeney, S., Oskarsson, M., Schmidt, W., Snigirev, A., von Stetten, D., Surr, J., Svensson, O., Theveneau, P. & Mueller-Dieckmann, C. (2015). *J. Synchrotron Rad.* **22**, 1540–1547.
- Canny, J. (1986). *IEEE Trans. Pattern Anal. Mach. Intell.* **8**, 679–698.
- Celie, P. H. N., Toebes, M., Rodenko, B., Ova, H., Perrakis, A. & Schumacher, T. N. M. (2009). *J. Am. Chem. Soc.* **131**, 12298–12304.
- Chavas, L. M. G., Yamada, Y., Hiraki, M., Igarashi, N., Matsugaki, N. & Wakatsuki, S. (2011). *J. Synchrotron Rad.* **18**, 11–15.
- Cherezov, V., Hanson, M. A., Griffith, M. T., Hilgart, M. C., Sanishvili, R., Nagarajan, V., Stepanov, S., Fischetti, R. F., Kuhn, P. & Stevens, R. C. (2009). *J. R. Soc. Interface*, **6**, S587–S597.
- Cipriani, F., Felisaz, F., Launer, L., Aksoy, J.-S., Caserotto, H., Cusack, S., Dallery, M., di-Chiaro, F., Guijarro, M., Huet, J., Larsen, S., Lentini, M., McCarthy, J., McSweeney, S., Ravelli, R., Renier, M., Taffut, C., Thompson, A., Leonard, G. A. & Walsh, M. A. (2006). *Acta Cryst.* **D62**, 1251–1259.
- Dera, P. & Katrusiak, A. (1999). *J. Appl. Cryst.* **32**, 510–515.
- Girard, E., Legrand, P., Roudenko, O., Roussier, L., Gourhant, P., Gibelin, J., Dalle, D., Ounsy, M., Thompson, A. W., Svensson, O., Cordier, M.-O., Robin, S., Quiniou, R. & Steyer, J.-P. (2006). *Acta Cryst.* **D62**, 12–18.
- González, A., Moorhead, P., McPhillips, S. E., Song, J., Sharp, K., Taylor, J. R., Adams, P. D., Sauter, N. K. & Soltis, S. M. (2008). *J. Appl. Cryst.* **41**, 176–184.
- He, J. H. & Gao, X. Y. (2015). *Eur. Phys. J. Plus*, **130**, 32.
- Heidari Khajepour, M. Y. H., Lebrette, H., Vernede, X., Rogues, P. & Ferrer, J.-L. (2013). *J. Appl. Cryst.* **46**, 740–745.
- Hilgart, M. C., Sanishvili, R., Ogata, C. M., Becker, M., Venugopalan, N., Stepanov, S., Makarov, O., Smith, J. L. & Fischetti, R. F. (2011). *J. Synchrotron Rad.* **18**, 717–722.
- Hiraki, M., Yamada, Y., Chavas, L. M. G., Wakatsuki, S. & Matsugaki, N. (2013). *J. Synchrotron Rad.* **20**, 890–893.
- Jain, A. & Stojanoff, V. (2007). *J. Synchrotron Rad.* **14**, 355–360.
- Karain, W. I., Bourenkov, G. P., Blume, H. & Bartunik, H. D. (2002). *Acta Cryst.* **D58**, 1519–1522.
- Kestur, U. S., Wanapun, D., Toth, S. J., Wegiel, L. A., Simpson, G. J. & Taylor, L. S. (2012). *J. Pharm. Sci.* **101**, 4201–4213.
- Kissick, D. J., Dettmar, C. M., Becker, M., Mulichak, A. M., Cherezov, V., Ginell, S. L., Battaile, K. P., Keefe, L. J., Fischetti, R. F. & Simpson, G. J. (2013). *Acta Cryst.* **D69**, 843–851.
- Land, E. H. & McCann, J. J. (1971). *J. Opt. Soc. Am.* **61**, 1–11.
- Lavault, B., Ravelli, R. B. G. & Cipriani, F. (2006). *Acta Cryst.* **D62**, 1348–1357.

- McPhillips, T. M., McPhillips, S. E., Chiu, H.-J., Cohen, A. E., Deacon, A. M., Ellis, P. J., Garman, E., Gonzalez, A., Sauter, N. K., Phizackerley, R. P., Soltis, S. M. & Kuhn, P. (2002). *J. Synchrotron Rad.* **9**, 401–406.
- Madden, J. T., Toth, S. J., Dettmar, C. M., Newman, J. A., Oglesbee, R. A., Hedderich, H. G., Everly, R. M., Becker, M., Ronau, J. A., Buchanan, S. K., Cherezov, V., Morrow, M. E., Xu, S., Ferguson, D., Makarov, O., Das, C., Fischetti, R. & Simpson, G. J. (2013). *J. Synchrotron Rad.* **20**, 531–540.
- Maire, A. le, Gelin, M., Pochet, S., Hoh, F., Pirocchi, M., Guichou, J.-F., Ferrer, J.-L. & Labesse, G. (2011). *Acta Cryst.* **D67**, 747–755.
- Norum, W. E. (2002). *Rev. Sci. Instrum.* **73**, 1560–1562.
- Okazaki, N., Hasegawa, K., Ueno, G., Murakami, H., Kumasaka, T. & Yamamoto, M. (2008). *J. Synchrotron Rad.* **15**, 288–291.
- Patankar, S. S., Mone, A. R. & Kulkarni, J. V. (2013). *2013 IEEE International Conference on Computational Intelligence and Computing Research*, pp. 7–11.
- Pauluhn, A., Pradervand, C., Rossetti, D., Salathe, M. & Schulze-Briese, C. (2011). *J. Synchrotron Rad.* **18**, 595–600.
- Pohl, E., Ristau, U., Gehrmann, T., Jahn, D., Robrahn, B., Malthan, D., Dobler, H. & Hermes, C. (2004). *J. Synchrotron Rad.* **11**, 372–377.
- Pothineni, S. B., Strutz, T. & Lamzin, V. S. (2006). *Acta Cryst.* **D62**, 1358–1368.
- Sauter, N. K. & Zwart, P. H. (2009). *Acta Cryst.* **D65**, 553–559.
- Song, J., Mathew, D., Jacob, S. A., Corbett, L., Moorhead, P. & Soltis, S. M. (2007). *J. Synchrotron Rad.* **14**, 191–195.
- Svensson, O., Malbet-Monaco, S., Popov, A., Nurizzo, D. & Bowler, M. W. (2015). *Acta Cryst.* **D71**, 1757–1767.
- Wang, Q. S., Yu, F., Huang, S., Sun, B., Zhang, K. H., Liu, K., Wang, Z. J., Xu, C. Y., Wang, S. S., Yang, L. F., Pan, Q. Y., Li, L., Zhou, H., Cui, Y., Xu, Q., Earnest, T. & He, J. H. (2015). *Nucl. Sci. Technol.* **26**, 12–17.
- Winter, G. & McAuley, K. E. (2011). *Methods*, **55**, 81–93.
- Yazid, H. & Arof, H. (2013). *J. Vis. Commun. Image Representation*, **24**, 926–936.
- Zhang, X. B., Feng, X. C., Wang, W. W., Zhang, S. L. & Dong, Q. F. (2013). *Comput. Electr. Eng.* **39**, 934–944.