Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

A new method for predetermining the diffraction quality of protein crystals: using SOAP as a selection tool

Robin Leslie Owen and Elspeth Garman*

Laboratory of Molecular Biophysics, Department of Biochemistry, Oxford University, Rex Richards Building, South Parks Road, Oxford OX1 3QU, England

Correspondence e-mail: elspeth@biop.ox.ac.uk

Received 14 October 2004
Accepted 14 November 2004

A microscope for quantitative analysis of the birefringence properties of samples is introduced. The microscope is used to measure variations in the slow optical axis position (SOAP) across hen egg-white lysozyme, glucose isomerase and fibronectin crystals. By comparing these variations with indicators of diffraction quality, it is shown that the optical properties of a protein crystal provide a non-invasive method of determining crystal diffraction quality before any X-ray data collection is attempted.

1. Introduction

There is currently no method of predetermining how well a protein crystal will diffract prior to X-ray analysis. As the majority of laboratories are sited some distance from synchrotron sources and most samples are cryocooled in the home laboratory and transported, the successful development of such a method would prove invaluable when samples are selected for X-ray data collection. The ability to preselect those crystals likely to diffract best could also play an important role in choosing crystals to test in an X-ray beam, and in providing a ranking system for experiments where there are many crystals and not enough time to test them all. The optical properties of lysozyme have been investigated previously; Cervelle *et al.* (1974) measured the reflectance in order to evaluate the refractive indices and hence the linear birefringence of both tetragonal and orthorhombic lysozyme crystals, while Kobayashi *et al.* (1998) calculated the optically rotatory power of tetragonal lysozyme crystals. However, in the work reported here it is the variations in the optical properties across the crystal that are of primary interest rather than the absolute values of the linear birefringence or optical activity. A birefringence imaging microscope is constructed from a standard transmission microscope with additional optical components which allow measurement of the birefringence properties of a sample. Rotation of a polarizer allows qualitative and quantitative measurements of the absorption, the anisotropy and the orientation of the sample. Here, we report the use of such a microscope, the Metripol, to measure the optical properties of hen egg-white lysozyme (HEWL), glucose isomerase and fibronectin crystals at room temperature and at 100 K, with a view to using these properties as prediction metrics for the X-ray diffraction quality of the crystals. Since there is no single parameter that characterizes the quality of diffraction data produced by a protein crystal, several indicators of data quality were used in order to investigate possible correlations between diffraction quality and the optical properties of each crystal.

2. Birefringence imaging

The Metripol birefringence imaging microscope (Oxford Cryosystems, Long Hanborough, Oxfordshire) is a commercially available transmission microscope with additional optical components. These additional components allow both qualitative and quantitative measurement of the transmitted intensity (I), the optical orientation (φ) and the optical anisotropy ($|\sin\delta|$) of a sample. The intensity of light, I, transmitted through a circular-polarizer/sample/rotating linear-analyser (CP-S-RLA) system is given by

$$I = \tfrac{1}{2}I_0[1 + \sin 2(\theta - \varphi)\sin\delta], \quad (1)$$

where I_0 is the incident light intensity, θ describes the orientation of the transmission axis of the rotating polariser and δ is the phase difference introduced between the two polarizations transmitted by the protein crystal (Glazer *et al.*, 1996). A schematic view of the passage of light through the microscope is shown in Fig. 1. It should be noted that this is a simplification, as the microscope uses circularly polarized light to allow an arbitary initial sample orientation. Linearly polarized light (shown in red) is incident on the crystal. Birefringent media transmit two mutually perpendicular components (blue and green), the orientation of which define the optical axis. These components travel at different velocities through the crystal and are termed the 'fast' and 'slow' components. They are recombined at the analyser and the intensity of the resultant (silver) is recorded as a function of the analyser position by a CCD camera.

Hence, the result of a CP-S-RLA system is a light wave whose intensity varies as a sine wave modulated by the term $\sin\delta$. Using stepped polarizer rotations, the values of φ and $|\sin\delta|$ can be obtained *via* Fourier analysis by the Metripol software. In principle, only three measurements are required at a given wavelength in order to determine I, φ and $|\sin\delta|$. However, by making more measurements the camera resolution is increased and the random error on each pixel is decreased. In the experiments reported here, 50 polarizer positions were used to calculate each image.

The optical anisotropy, $\sin\delta$, is dependent on the thickness of the crystal, as $\delta = (2\pi/\lambda)\Delta nL$, where Δn is the linear birefringence, λ is the wavelength of the incident light and L is the thickness of the crystal. This dependence on L makes the quantity δ unsuitable for use in comparing the diffractive properties of the crystals. The linear birefringence, Δn, is the difference between n_s and n_f, the refractive indices in the direction of the slow and fast optical axes, respectively; *i.e.* $\Delta n = n_s - n_f$. These indices are shown in Fig. 2 by the major axis (n_s) and minor axis (n_f) of the indicatrix. It is worth noting here that the shape of the indicatrix is defined by the symmetry of the crystal (Nye, 1984).

The optical orientation φ, however, is an intrinsic property of the crystal; this angle is illustrated in Fig. 2. The orientation of the slow optical axis is defined by the angle between the x axis and the direction of greatest refractive index (n_s). Variations in φ across each crystal can be obtained and the magnitude of these variations used as a measure of the optical quality of the crystal.

In the work reported here, further modifications were made to the microscope to allow analysis of loop-mounted protein crystals at both room and cryogenic temperatures. The stage was removed and replaced with a goniometer-head mount which could be rotated in the plane perpendicular to the light path between the polarizer and the analyser. A 600 series Cryostream (Oxford Cryosystems) was also positioned so that crystals could be examined under the microscope while being held at 100 K.

Fig. 3 shows two sets of images produced by the microscope. The first (left-hand) set depicts a 'good-quality' HEWL crystal, while the second (right-hand) set shows a 'poor-quality' glucose isomerase crystal. It can be seen that the

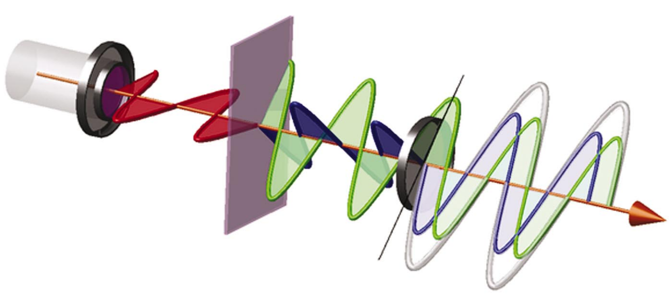

Figure 1
A schematic view of light passing through the Metripol microscope. Plane-polarized monochromatic light (shown in red) is incident on the crystal. In the case of a birefringent crystal, two perpendicular components (green and blue) are transmitted. These components are recombined at the analyser and the intensity of the resultant (silver) wave is measured using a CCD camera.

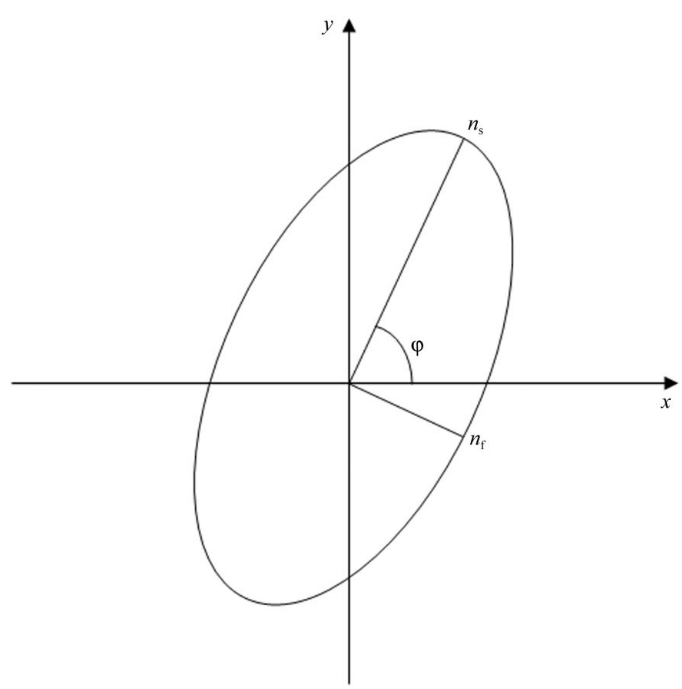

Figure 2
The indicatrix illustrates the variation in refractive index for different incident light-polarization vectors in a crystal. The orientation or extinction angle φ is the angle between the x axis and the slow optical axis of the crystal, n_s.

orientation φ (Fig. 3*c*) stays approximately constant over the face of the HEWL crystal despite variations in $\sin\delta$ (Fig. 3*e*) and striations visible in the intensity image (Fig. 3*a*). However, in the case of the glucose isomerase crystal, visible striations (Fig. 3*b*) correspond to larger variations in φ (Fig. 3*d*).

Only the magnitude of $\sin\delta$ can be determined, not its sign. If the sign of $\sin\delta$ changes in (1) then it can be seen from Fig. 4 that the waveform of $\sin 2(\theta - \varphi)$ appears to shift through $\pi/2$ radians, resulting in a change in the measured value. This can be observed in the top left corner of the crystal in Fig. 5(*b*). It should be noted that it is possible to obtain an absolute value of δ by combining measurements of $|\sin\delta|$ at different wavelengths (Geday *et al.*, 2000).

In Fig. 6(*b*), a typical area selection for analysis of the optical properties of a crystal is shown. The area was chosen to maximize the crystal area while minimizing the solvent area enclosed. Also shown are the peaks resulting from other areas of the image. Note that no peak results from the background of the image since all points in this area have $|\sin\delta| \leq 0.1$. The method for fitting these peaks is described below.

Setting a threshold on $\sin\delta$ allows non-crystalline and birefringent areas where $\sin\delta$ is small to be eliminated from the analysis. When $\sin\delta$ is small it becomes difficult to determine accurately a value for φ, as the $\sin 2(\theta - \varphi)\sin\delta$ term in (1) tends to zero. Accordingly, when analysing all the optical data, a threshold of $|\sin\delta| \geq 0.1$ was used.

The distribution of the slow optical axis within a crystal was quantified by fitting the histogram peaks in the orientation data (Fig. 5) to Gaussian functions of the form

$$P(\varphi) = \frac{k}{\sigma(2\pi)^{1/2}} \frac{1}{\exp[-(\varphi - m)^2/2\sigma^2]}, \qquad (2)$$

where k is the total area under the curve, σ is the variance and m represents the mean. Peaks were fitted using *Origin* v.6.1 software.

The optical width is defined here as being equal to 2σ of the fitted peaks and is approximately 0.849 the width of the peaks at half-maximum height. The variations in the slow optical axis position (SOAP) were then used to characterize the optical quality or 'SOAPiness' of each crystal.

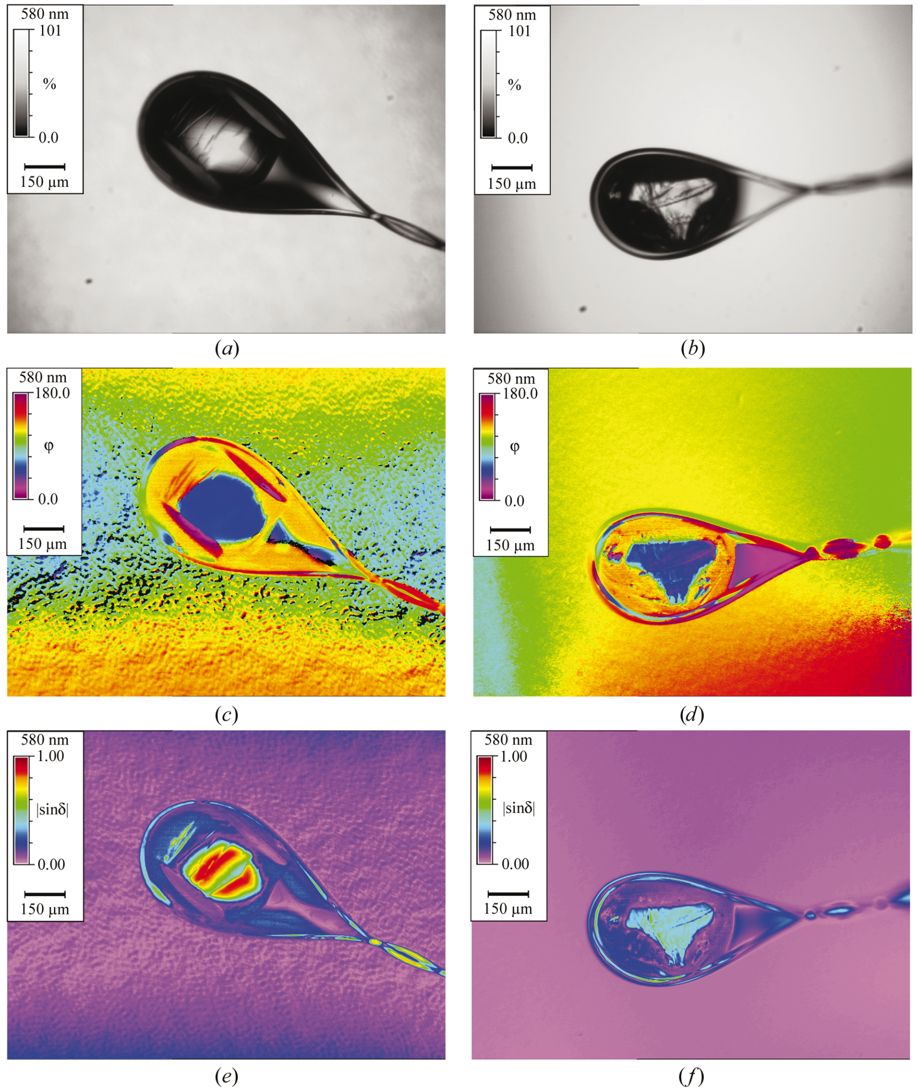

Figure 3
The three images produced by the microscope: transmitted intensity (I) (top), orientation (φ) (centre) and $\sin\delta$ (bottom). (*a*), (*c*) and (*e*) show a room-temperature HEWL crystal. The variation in orientation seen in the background of the central image arises from reflection at the surface of the glass capillary within which the crystal was mounted. (*b*), (*d*) and (*f*) show a cryocooled (100 K) glucose isomerase crystal. Note that the orientation φ (*c* and *d*) stays constant across the face of the crystals, while $\sin\delta$ (*e* and *f*) varies as the thickness of the crystal changes.

3. Diffraction quality

There is no single parameter that characterizes how well a protein crystal diffracts; accordingly, several parameters have been used to assess the diffraction quality.

3.1. *R* values

All crystal lattices possess a point of inversion at the origin. This imposes the requirement (Freidel's law) that the intensity of symmetry-related reflections is the same. This fact can be used to give an indication of the data quality by calculating the R values.

The quantity R_{sym} compares the intensity of symmetry-related reflections. If n measurements are made of a reflection intensity and its symmetry-related partners, then R_{sym} is given by

$$R_{\rm sym} = \frac{\sum_h |\hat{I}_h - I_{h,i}|}{\sum_h \sum_i^{n_h} I_{h,i}}, \qquad (3)$$

where the mean intensity $\hat{I}_h = (1/n_h)\sum_i^{n_h} I_{h,i}$, $I_{h,i}$ is the measured intensity of a reflection and n_h is the multiplicity of the contributing reflections h. A value of around 5% for $R_{\rm sym}$ is taken to represent good data quality for protein crystals.

However, $R_{\rm sym}$ is flawed in that it has an implicit dependence on the redundancy of the data. $R_{\rm sym}$ will increase if a reflection is measured many times even though this increases the accuracy of the data and thus their reliability. To overcome this problem two alternate indicators have been used, $R_{\rm meas}$ and the pooled coefficient of variation (PCV), in which the contribution of each individual reflection is weighted according to its multiplicity. $R_{\rm meas}$ differs from PCV by a factor of $\pi^{1/2}/2$ (Diederichs & Karplus, 1997),

$$R_{\rm meas} = \frac{\sum_h [n_h/(n_h-1)]^{1/2} \sum_i^{n_h} |\hat{I}_h - I_{h,i}|}{\sum_h \sum_i^{n_h} I_{h,i}}, \qquad (4)$$

$$\mathrm{PCV} = \frac{\sum_h \left\{ [1/(n_h-1)] \sum_i^{n_h} (I_{h,i} - \hat{I}_h)^2 \right\}^{1/2}}{\sum_h \hat{I}_h}. \qquad (5)$$

$R_{\rm meas}$ and PCV are relative to I^+ or I^-; $R_{\rm meas}0$ and PCV0 referred to later in this paper are relative to the overall mean of I^+ and I^- reflections.

3.2. $I/\sigma(I)$

The value of the ratio of the mean intensity I to the variation in the intensity $\sigma(I)$ gives an indication of the signal-to-noise ratio, the data quality and the diffracting power of the crystal. Values of $I/\sigma(I)$ were extracted for both the whole image, $I/\sigma(I)_{\rm all}$, and for the outermost resolution shell, $I/\sigma(I)_{\rm outer}$. As a rough guide, more than 50% of the data in the highest resolution shell should have $I/\sigma(I) > 2$ (Dauter, 1999).

3.3. Mosaicity

The mosaicity of a crystal is measured by determining the average rocking width of the reflections (Darwin, 1922). The nature of this measurement means that the mosaicity is not just a measure of the crystal quality but also of the wavelength dispersion of the beam ($\lambda/\delta\lambda$) and the beam divergence. For an in-house rotating-anode source with multi-layer optics, the mosaicity is dominated by the beam characteristics (Dauter, 1999).

Table 1
Crystallization conditions for HEWL.

Crystal No.	MDL screen	Crystallization conditions	Cryoprotectant agent added
H1	Screen II No. 1	0.1 *M* NaCl	Yes
H3, H7		0.1 *M* Bicine pH 9.0	Yes
H10	Screen I No. 17	30%(*w*/*v*) PEG monomethyl-ether 550	No
H2, H4		0.1 *M* imidazole pH 6.5 1.0 *M* NaCl trihydrate	Yes
H5	n/a	200 m*M* sodium acetate pH 4.7, 6%(*w*/*v*) NaCl	Yes
H6	Screen II No. 10	0.01 *M* $NiCl.6H_2O$, 0.1 *M* Tris pH 8.5 20%(*w*/*v*) PEG mono-methylether 2000	Yes
H8	Screen I No. 16	0.1 *M* Na HEPES pH 7.5 0.8 *M* K/Na tartrate tetrahydrate	Yes
H9	Screen II No 12	0.1 *M* HEPES pH 7.5 10%(*w*/*v*) PEG 6000, 5%(*v*/*v*) MPD	Yes

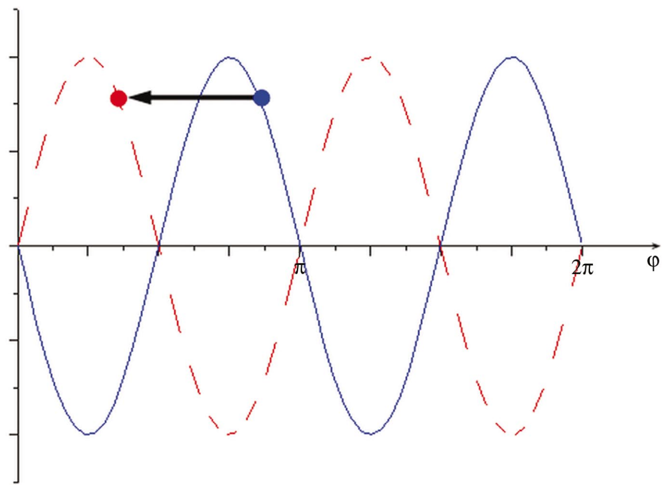

Figure 4
$\sin 2(\theta - \varphi)$ plotted as a function of φ. A change of $\pi/2$ in the orientation appears to occur when the sign of $\sin\delta$ changes. $\sin 2(\theta - \varphi)$ is shown as a solid blue line, while $-\sin 2(\theta - \varphi)$ is shown as a dashed red line.

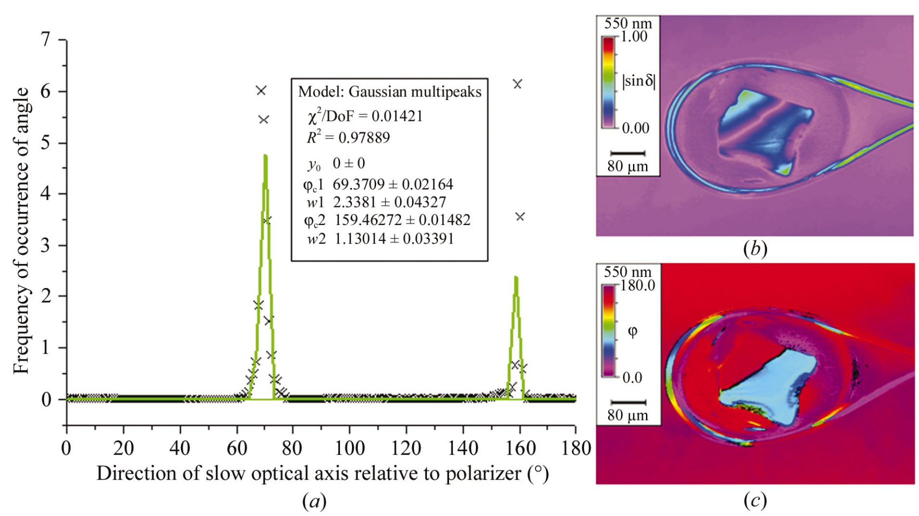

Figure 5
(*a*) Variation in the relative orientation of the slow optical axis across a protein crystal. Two peaks 90° apart can be seen; this arises from a change in the sign of $\sin\delta$ as explained in the main text. (*b*) shows $\sin\delta$ changing across the crystal, while (*c*) shows the change of φ.

4. Crystallization and data collection

4.1. Crystallization

Hen egg-white lysozyme was obtained from Sigma and used with no further purification. 50 mg ml^{-1} HEWL was crystallized with an equal volume of crystallization solution from Molecular Dimensions Screens I and II by hanging-drop vapour diffusion (Jancarik & Kim, 1991). The crystals grew under various conditions (Table 1) and were tetragonal, belonging to space group $P4_32_12$ with unit-cell parameters $a = b = 78.9$, $c = 37.8$ Å. Crystals were soaked in a solution of 20% glycerol and mother liquor for approximately 30 s before flash-cooling to 100 K. The cryoprotectant was prepared so that the mother liquor was not diluted by glycerol, but rather water in the mother liquor was replaced by glycerol.

Streptomyces rubiginosus glucose isomerase was obtained in solution from Hampton Research and dialysed twice against 0.01 *M* HEPES pH 7.0 before concentration of the protein solution to 70 mg ml^{-1}. The protein solution was then centrifuged at 13 000 rev min^{-1} prior to crystallization using the hanging-drop method with a well solution consisting of 3.4 *M* 1,6-hexanediol, 0.2 *M* $MgCl_2$ and 0.1 *M* Tris pH 8.5. The crystals produced were orthorhombic, belonging to space group $I222$ with unit-cell parameters $a = 93$, $b = 98$, $c = 102$ Å. Crystals were cryoprotected by soaking them for between 4 and 40 min in a solution consisting of 30% MPD, 0.2 *M* $MgCl_2$ and 0.1 *M* Tris pH 8.5.

The fibronectin domains ^{2}F1 and ^{3}F1 were crystallized using the hanging-drop method with a well solution consisting of 0.1 *M* acetate buffer pH 4.5, 0.2 *M* ammonium sulfate and 5% PEG 550 (Rudiño-Piñera *et al.*, 2004). The crystals produced belonged to space group $P4_12_12$, with unit-cell parameters $a = b = 37.8$, $c = 107.9$ Å. Crystals were cryoprotected by increasing the PEG 550 concentration to 40% in one step and soaking for ~60 s.

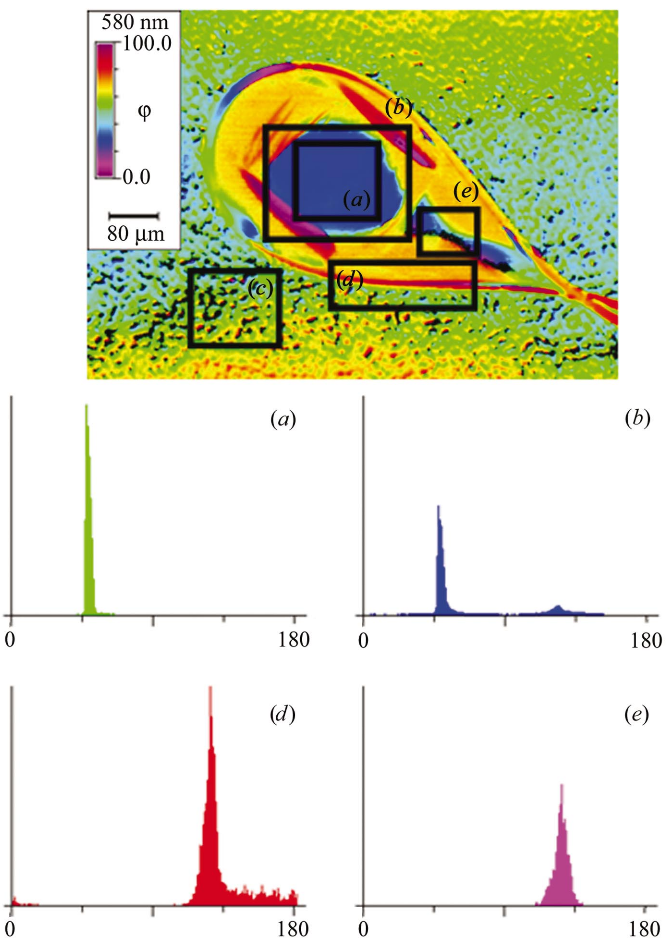

Figure 6
Illustration of the different peaks obtained from a typical orientation (φ) image. A threshold of $\sin\delta \geq 0.1$ was set for all the peak analyses. It can be seen that the ordered region of the crystal results in a sharp peak as in (*a*) and (*b*). After a threshold of 0.1 was set no data remained in region (*c*), so it is not plotted. The loops in which the crystals are mounted are birefringent, so selection of region (*d*) also results in a sharp peak. The area used for analysis was chosen such that the loop did not contribute to the peak fitted to the crystal data. Disordered regions such as the solvent produce broader peaks (*e*).

4.2. Optical data collection

For both the room-temperature and 100 K experiments, crystals were loop-mounted. For the former, they were sealed within a 2.0 mm diameter quartz capillary using a method developed from that of Skrzypczak-Jankun *et al.* (1996), using a standard Oxford Cryosystems tophat and Plasticine to seal the capillary to the hat; a picture of this is shown in Fig. 7. This method of mounting crystals was chosen in order to eliminate any crystal handling between data sets collected at room and cryogenic temperatures, allowing observation of the effects of the cooling process on the optical and diffractive properties.

Crystals were orientated such that the plane of the loop was coincident with the focal plane of the microscope. This avoided any contribution to the optical width from the nylon loops, which are themselves birefringent. Images calculated

Figure 7
Loop-mounted crystal enclosed in a quartz capillary. The column of mother liquor used to stop the crystal drying out can be seen at the top of the capillary.

Table 2
Comparison of the optical properties of HEWL, glucose isomerase and fibronectin crystals with data from complete diffraction data sets.

H1 and H7 and also H2 and H4 refer to data sets taken at room temperature and 100 K, respectively. The outermost resolution shells for HEWL, glucose isomerase and fibronectin data are 2.77–2.59, 2.36–2.20 and 2.51–2.39 Å, respectively.

		Diffraction data					
Data set	Optical width (°)	$I/\sigma(I)_{\rm all}$	$I/\sigma(I)_{\rm outer}$	$R_{\rm sym}$	$R_{\rm meas}0$	PCV0	Mosaicity (°)
H1	1.50	85.3	33.5	0.037	0.047	0.054	0.43
H2	1.71	111.1	39.2	0.039	0.051	0.059	0.36
H3	1.77	78.4	32.0	0.043	0.056	0.058	0.51
H4	1.83	104.1	60.8	0.034	0.044	0.051	0.66
H5	2.77	55.6	27.5	0.046	0.059	0.069	1.01
H6	3.12	96.3	44.0	0.061	0.082	0.091	0.30
H7	3.48	46.4	22.5	0.051	0.064	0.073	1.15
H8	3.63	54.4	22.2	0.061	0.077	0.088	0.49
H9	7.66	46.9	22.8	0.082	0.103	0.116	0.35
H10	9.13	34.7	9.8	0.099	0.129	0.144	0.74
G1	3.5	42.2	18.4	0.050	0.057	0.067	0.549
G2	3.7	62.0	31.1	0.044	0.051	0.059	0.536
G3	3.9	88.0	47.0	0.046	0.053	0.061	0.408
G4	5.4	67.6	34.1	0.047	0.054	0.063	0.346
G5	5.6	77.3	38.6	0.039	0.044	0.051	0.379
G6	10.1	39.0	16.7	0.052	0.060	0.071	0.479
G7	11.3	49.2	20.2	0.046	0.053	0.063	1.288
G8	15.0	38.4	14.8	0.064	0.073	0.085	1.238
F1	3.4	43.4	7.1	0.039	0.050	0.055	0.791
F2	4.5	66.3	23.8	0.033	0.048	0.051	0.714
F3	4.6	52.8	12.8	0.041	0.052	0.056	0.788
F4	6.0	61.5	23.1	0.040	0.056	0.055	0.949
F5	6.1	36.8	6.4	0.038	0.058	0.064	0.662
F6	6.2	42.1	8.0	0.042	0.053	0.058	0.741
F7	6.4	29.3	3.6	0.052	0.076	0.084	0.859
F8	7.3	70.3	27.0	0.047	0.072	0.079	1.066
F9	8.5	43.0	9.3	0.047	0.060	0.066	0.691

Table 3
Additional diffraction data for HEWL, glucose isomerase and fibronectin.

The outermost resolution shells for HEWL, glucose isomerase and fibronectin data are 2.77–2.59, 2.36–2.20 and 2.51–2.39 Å, respectively.

		Total No. of reflections			Completeness (%)	
Data set	Crystal dimensions (μm)	Observed	Unique	Multiplicity	Overall	Outer shell
H1	300 × 290 × 70	17379	3792	4.6	99.3	96.7
H2	380 × 200 × 150	16780	3632	4.6	99.4	99.0
H3	290 × 200 × 60	14418	3791	3.8	98.9	93.9
H4	380 × 200 × 150	16041	3632	4.4	99.2	96.8
H5	185 × 85 × 80	14985	3412	4.4	96.8	92.6
H6	250 × 360 × 150	9389	2882	3.3	90.4	90.9
H7	300 × 290 × 70	15237	3457	4.4	99.3	97.6
H8	380 × 360 × 90	14346	3672	3.9	99.3	97.3
H9	220 × 250 × 60	14820	3711	4.0	99.0	95.6
H10	200 × 280 × 80	12672	3564	3.6	97.7	92.3
G1	230 × 240 × 150	94784	22900	4.1	99.5	96.7
G2	260 × 210 × 250	95111	22831	4.2	99.6	97.6
G3	460 × 470 × 300	95609	22708	4.2	99.5	97.1
G4	260 × 340 × 200	95964	22874	4.2	99.6	98.0
G5	270 × 320 × 200	96796	23058	4.2	99.6	97.3
G6	210 × 270 × 170	95756	22796	4.2	99.6	97.2
G7	360 × 220 × 110	95165	22800	4.2	98.6	94.7
G8	250 × 320 × 100	93273	22768	4.1	98.8	93.7
F1	180 × 200 × 140	13543	5043	2.7	98.6	95.4
F2	180 × 210 × 160	13671	5000	2.7	97.8	95.5
F3	240 × 210 × 130	13693	5181	2.6	99.2	98.1
F4	270 × 160 × 170	12531	4754	2.6	96.0	90.0
F5	160 × 210 × 120	13764	5115	2.7	98.7	96.5
F6	180 × 180 × 100	13884	5230	2.7	98.4	96.7
F7	160 × 210 × 120	13710	5185	2.6	98.6	96.2
F8	200 × 320 × 140	13029	4998	2.6	97.4	94.6
F9	200 × 180 × 100	13408	4900	2.7	98.4	94.8

from 50 stepped positions of the polarizer could then be collected. Cryocooled crystals were transfered from the microscope by the use of either pre-cooled tongs or a removable arc (Garman & Schneider, 1997).

Initially, it was found that a recurring problem encountered with room-temperature data collection was crystal slippage arising from movement of crystals within the loop of liquid. This gave difficulty in both indexing the diffraction patterns and in scaling the data. Thus, for room-temperature analysis, loop size had to be chosen carefully when mounting crystals. In practice, it was found that the ideal loop was approximately the same size as the crystal, providing support for the crystal edges while allowing the crystal to remain enclosed by a drop of mother liquor.

4.3. X-ray data collection

Full crystallographic data sets were collected for all of the investigated crystals using Cu $K\alpha$ X-rays from an in-house Rigaku RU-200H generator equipped with Osmic optics and a MAR345 imaging-plate detector. In the case of HEWL, data were collected in two segments of 30°, the orientations of which were determined using the strategy option in *MOSFLM* (Leslie, 1999). Each image had an oscillation angle $\Delta\varphi = 1°$. Diffraction data for HEWL were collected with a crystal-to-detector distance (XTD) of 250 mm; the outermost resolution shell thus had a range of 2.77–2.59 Å. Full data sets were also collected for glucose isomerase crystals; because of the space group (*I*222), a wider sweep of data had to be collected (105° in total) to ensure that a complete data set was obtained. However as the crystal form was body-centred, the oscillation angle of each image could be increased to $\Delta\varphi = 1.5°$ without losing data completeness from overlapping reflections. The glucose isomerase data were collected with a XTD of 200 mm, resulting in an outer resolution shell of 2.36–2.20 Å. Fibronectin data were collected with XTD of 220 mm, resulting in an outer shell of 2.51–2.39 Å. An oscillation angle of $\Delta\varphi = 1°$ was used and 60° of data were collected for each crystal.

For each of the protein types the same crystal to backstop distance, slit size and generator power were used throughout.

5. Results

Optical and diffraction data sets were collected for each crystal. The optical widths were obtained as described in §2. Diffraction data were integrated using *MOSFLM* and scaled using *SCALA* from the *CCP*4 package (Collaborative Computational Project, Number 4, 1994; Leslie, 1999). The optical widths could then be compared with various indicators of diffraction quality.

One HEWL data set was discarded as it was of questionable quality; with R values of more than 30%, these data were considered too poor to be used as part of a useful comparison.

Numerous attempts were made to collect room-temperature optical and diffraction data for fibronectin, using both the mounting method described above and also more traditional capillary-mounting techniques (Garman, 1999). However, fibronectin was found to be very sensitive to dehydration, resulting in the quality of diffraction being time-dependent. The results from room-temperature studies of fibronectin are therefore not reported here. All the other results are summarized in Tables 2 and 3.

It can be seen from Figs. 8, 9, 10 and 11 that $I/\sigma(I)$, R_{sym} and $R_{meas}0$ scale approximately linearly with the width of fitted optical peaks for both HEWL and glucose isomerase crystals. This suggests that variations in optical orientation across a protein crystal can be used as an indicator of the diffractive quality of that crystal. The trends for both $I/\sigma(I)$ and the different R values indicate that a crystal with smaller variations in optical orientation will be of superior diffraction quality than one with larger variations, irrespective of its volume. There is no obvious trend for the variation in the mosaicity with optical peak width (Fig. 12). There is also no correlation between variations in the SOAP and the volume of the crystal (Fig. 13).

In the case of fibronectin crystals it was found that $I/\sigma(I)$ did vary as a function of volume (Fig. 14). However, it can also be seen that R values no longer scale linearly with $I/\sigma(I)$ (Fig. 15). This suggests that when $I/\sigma(I)$ is dependent on crystal volume, $I/\sigma(I)$ becomes a less reliable indicator of crystal quality. Fig. 16 shows a lack of correlation between $I/\sigma(I)$ and the width of the fitted optical peak; however, Fig. 17 shows a linear correlation between R value and variation in SOAP.

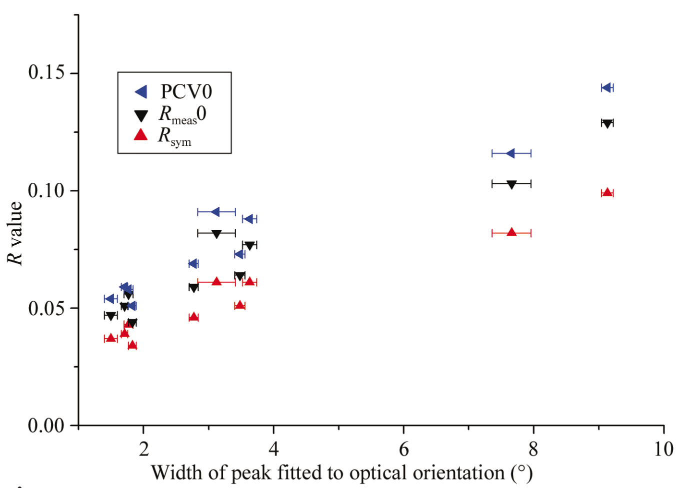

Figure 8
Plot of R_{sym}, $R_{meas}0$ and the pooled coefficient of variation (PCV0) against the width of the fitted optical peak for ten HEWL crystals.

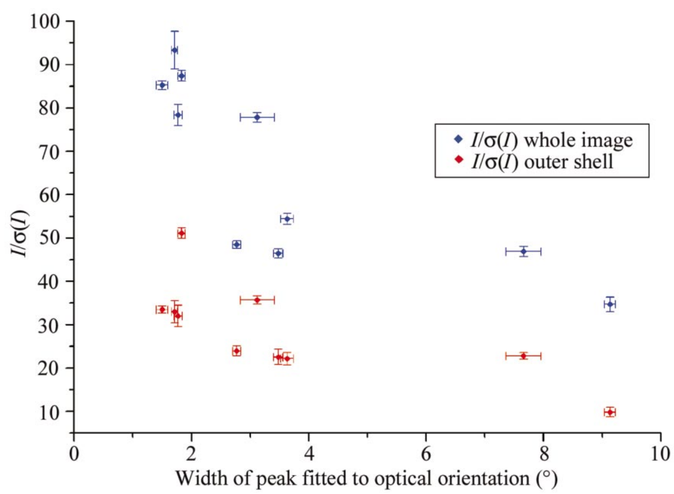

Figure 9
Plot of $I/\sigma(I)$ for both the whole image and the outer resolution shell against the optical peak width for the HEWL crystals.

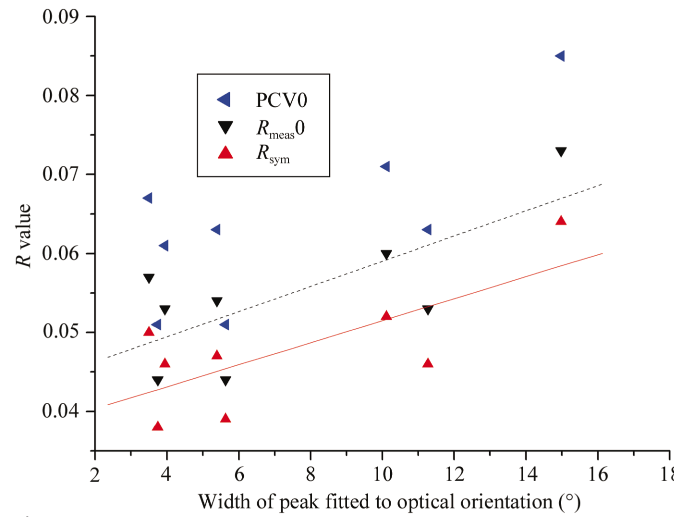

Figure 10
Plot of R_{sym}, $R_{meas}0$ and the pooled coefficient of variation (PCV0) against the width of the fitted optical peak for eight glucose isomerase crystals. The trend lines for R_{sym} (red, continuous) and $R_{meas}0$ (black, dashed) are shown.

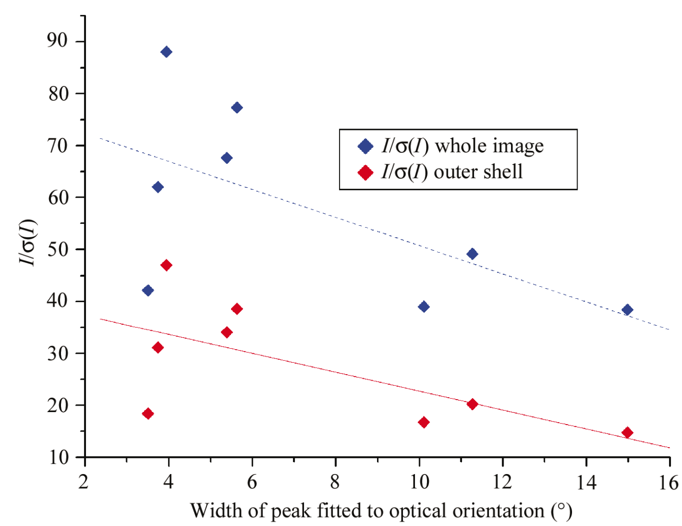

Figure 11
Plot of $I/\sigma(I)$ for both the whole image and the outer resolution shell against the optical peak width for the glucose isomerase crystals.

In comparing Figs. 8, 9, 10, 11 and 17 it can been seen that while $I/\sigma(I)$ in the case of HEWL and glucose isomerase and R values in the case of HEWL, glucose isomerase and fibronectin follow the same trends, *i.e.* decreasing $I/\sigma(I)$ with increasing variations in SOAP and increasing R value with increasing variations in SOAP, the rates of change differ. If a linear regression is fitted to R value *versus* variation in SOAP for both HEWL and glucose isomerase (linear fit shown for glucose isomerase in Fig. 11), the gradients differ by up to a factor of five in the case of $R_{\rm sym}0$ and a factor of six in the case of $R_{\rm meas}0$. These results clearly indicate that variations in SOAP across a crystal can only be used to determine the relative quality of crystals of the same protein type and not as an absolute indicator of quality for all protein crystals.

In order to further investigate the lack of a correlation of SOAP with crystal mosaicity, the mosaic spreads for HEWL data were reintegrated taking into account the beam divergence for blue Osmic optics, which was set to 0.1° (A. Leslie, personal communication). It is problematic to deconvolute the beam divergence and the crystal mosaicity, but this at least puts a lower limit on the beam divergence. Despite this reanalysis of the data, there was still no correlation between variations in SOAP and the mosaicity.

Glucose isomerase data were also integrated using *d*TREK* (Pflugrath, 1999). Reflection files output from *d*TREK* were then converted to multi-record MTZ format using *DTREK2SCALA* (G. Evans, personal communication; now available in *CCP*4 v.5.0) and then scaled in the same way as MTZ files output from *MOSFLM*. The data displayed in Fig. 18 show the results of this; the *d*TREK* mosaicities quoted are the full-width at half-maximum (FWHM) height of the modelled rocking-curve widths. A direct comparison of crystal mosaicities from different programs is difficult owing to the different formulae used to relate the rocking width of a particular reflection to the crystal mosaicity. However, Fig. 18 shows that while the specific program used results in changes in the absolute value of the mosaicity obtained, the trend of

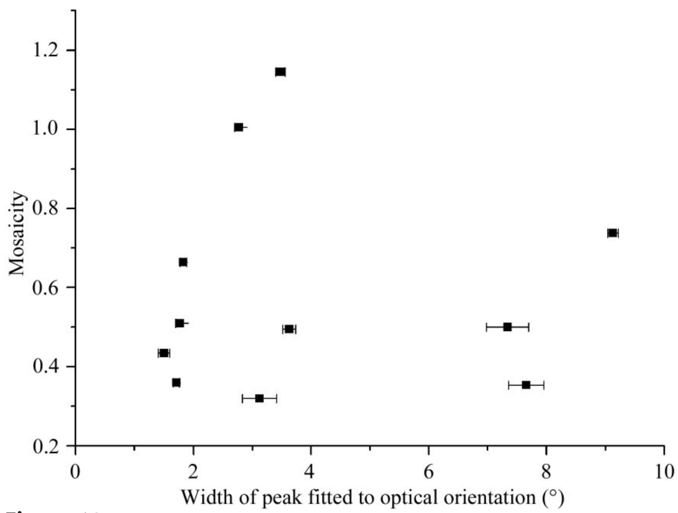

Figure 12
Plot of mosaicity against the width of the fitted optical peak for the HEWL crystals.

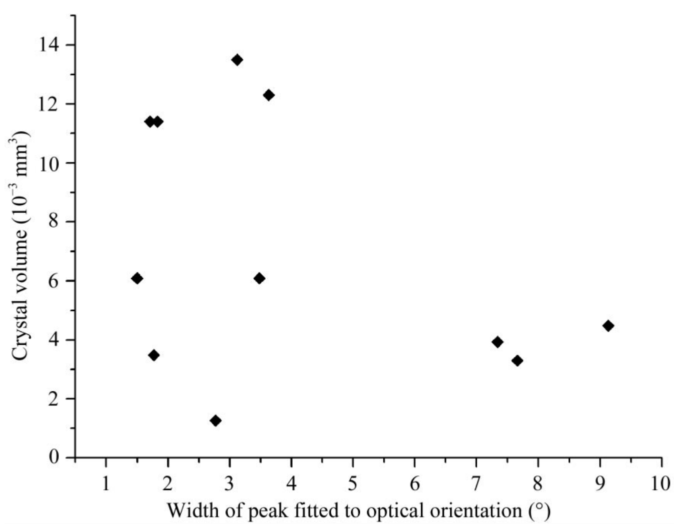

Figure 13
Plot of the crystal volume against the width of the fitted optical peak for the HEWL crystals.

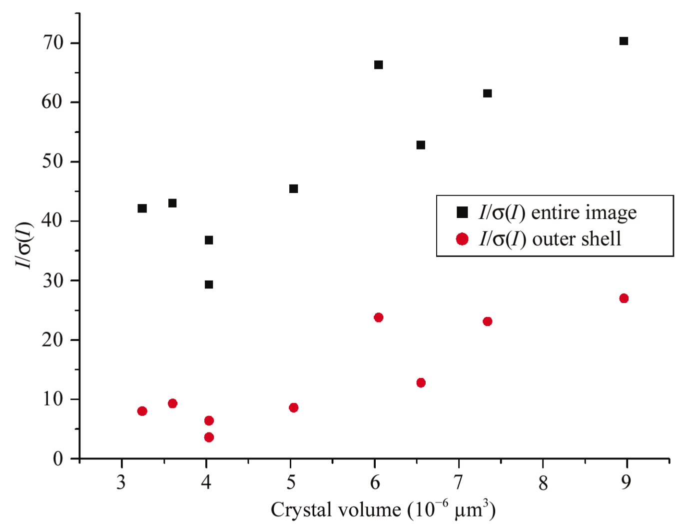

Figure 14
Plot of $I/\sigma(I)$ against the crystal volume for nine fibronectin crystals.

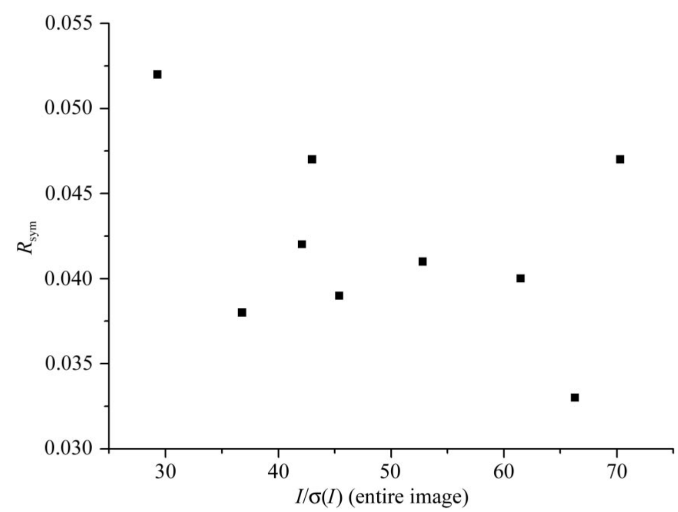

Figure 15
Plot of $I/\sigma(I)$ against R value for the fibronectin crystals.

variation in mosaicity with the width of the fitted optical axis peak remains unchanged.

The subsequent scaling of data in *SCALA* meant that trends in R values for data could be compared with the trends found for data integrated in *MOSFLM*. The trends proved to be the same (not shown).

Importantly, there is no correlation between the volume of a crystal and variations in SOAP, as shown in Fig. 13. Thus, the observed trends in R values and $I/\sigma(I)$ do not arise from the effect of differing crystal size on diffraction quality when using an in-house X-ray source.

6. Discussion

The plots presented above show a clear correlation between variations in the slow optical axis position and the diffractive properties of protein crystals. To interpret these results, we need to examine the different possible contributions to the optical properties of the crystals.

The optical anisotropy of a protein crystal can be related directly to the optical anisotropy of the protein molecules and their internal structure (Ruiz & Oldenbourg, 1988). The observed birefringence is believed to be a combination of the intrinsic birefringence of the protein molecules themselves and the effect of immersing molecules of a high refractive index, n_2, in a solvent of a lower refractive index, n_1. This problem was first tackled by Wiener (1912), although the work summarized here is that of Bragg & Pippard (1953), who used more modern notation.

In the case of a regular array of protein molecules, occupying a fraction f of the unit cell, the general form of the birefringence is

$$\Delta n^2 = f(n_2^2 - n_1^2)\left(\frac{1}{1+kL_\alpha} - \frac{1}{1+kL_\beta}\right), \qquad (6)$$

where L_α and L_β are depolarizing coefficients dependent on the shape of the protein molecules and $k = (1-f)[(n_2^2/n_1^2) - 1]$. This expression was obtained by considering the polarizability of ellipsoids suspended in a solution when an electric field was applied. The above expression can be simplified if we consider the shape of the protein molecule in two limiting cases. If the molecules are plate-like, then

$$\Delta n = -f(n_2 - n_1)\frac{9k}{(3+k)^2}. \qquad (7)$$

Rod-like molecules result in a birefringence of the form

$$\Delta n = \frac{1}{2}(n_2 - n_1)\frac{9k}{(3+k)^2}. \qquad (8)$$

It should be noted that at high concentrations of protein molecules we expect the form birefringence to be reduced, as the last two terms in (6) both tend to $\frac{1}{2}$ and therefore Δn^2 tends to zero. Physically, the protein molecules can no longer be

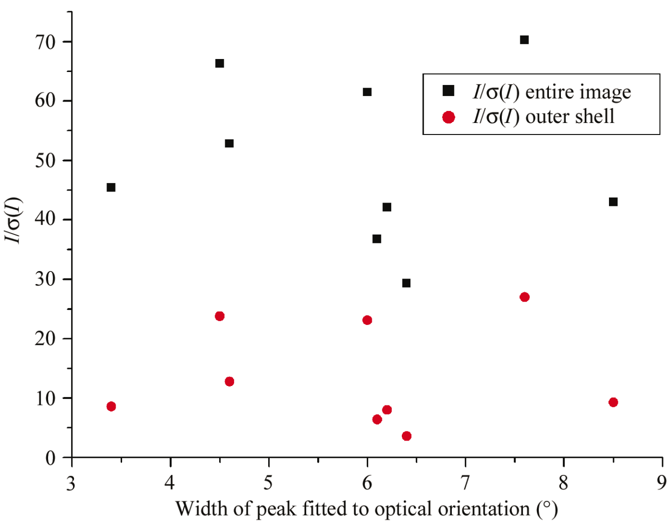

Figure 16
Plot of $I/\sigma(I)$ for both the whole image and the outer resolution shell against the optical peak width for the fibronectin crystals.

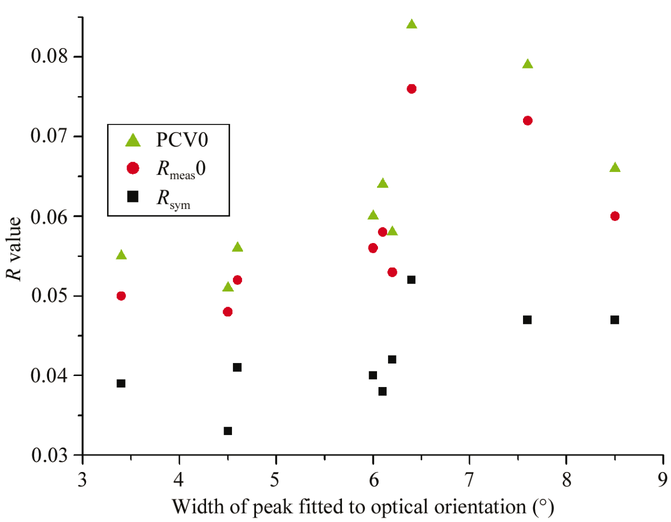

Figure 17
Plot of R_{sym}, $R_{meas}0$ and the pooled coefficient of variation (PCV0) against the width of the fitted optical peak for the fibronectin crystals.

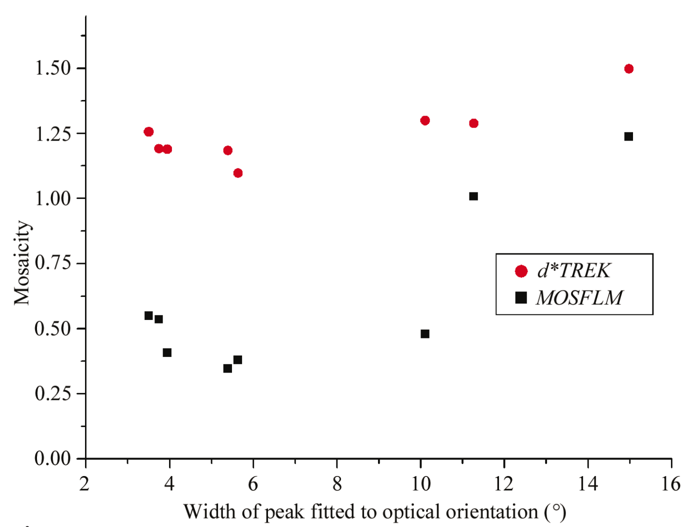

Figure 18
Plot of crystal mosaicity against optical peak width for the glucose isomerase crystals. The mosaicities as determined by both *MOSFLM* and *d*TREK* are shown.

considered as separate and the shape of the molecules becomes less important as the electric field in the solvent becomes dominated by that in the protein molecules.

Oldenbourg & Ruiz (1989) expanded the above treatment to particles with corrugated surfaces and interstitial solvent in order to predict the birefringence of DNA and tobacco mosaic virus. They showed that the effect of this more realistic treatment is to reduce the magnitude of the form birefringence.

Thus, the phase difference, δ in equation (1), is the result of superposition of both the intrinsic birefringence of the protein molecules and the form birefringence.

The two contributions to the birefringence and the possibility of variations in the form birefringence arising from solvent composition even if the crystal structure remains unchanged, illustrate further why the optical anisotropy $\sin\delta$ cannot be used to quantify the crystal quality, even if the crystal thickness were to be determined accurately, as $\sin\delta$ may vary with, for example, the duration of a soak in cryoprotectant.

The orientation of the slow optical axis is defined by the crystal symmetry. However, a protein crystal is not a perfect crystal; it is a mosaic of many submicroscopic arrays in rough alignment with each other (McPherson, 2001). The individual arrays are too small to be separately resolved, but they result in a smearing of the observed optical axis. The indicatrix can also be used to describe this phenomenon: when the crystal is subject to stress, caused by for example crystal handling or osmotic shock upon soaking, the size, shape and orientation of the indicatrix changes across the crystal. This is described in detail for a crystal belonging to space group *I*222 by Nye (1984).

Owing to the weak nature of the bonds holding protein molecules together in a crystal, poor handling and mounting techniques are enough to increase this smearing, increasing variations in the slow optical axis position and decreasing diffraction quality. This is reflected by reduced values of $I/\sigma(I)$ and increased *R* values. The lack of correlation between both the mosaicity and the slow optical axis position and the mosaicity and other indicators of diffraction quality indicate that the mosaicity is not a robust indicator of data quality.

This postulate can be examined further by looking at the effect on SOAP of cryocooling crystals. In two cases, diffraction and optical data were collected from a single crystal at both room temperature and 100 K (H1, H7 and H2, H4). Initially, crystals H1 and H2 were of similar quality. However, upon cooling the diffractive quality of crystal H1 almost halved (as judged by the decrease in $I/\sigma(I)$ and increased *R* values), while H2 remained of a similar quality. These changes in diffractive quality were reflected by a small increase in the variation in SOAP in the case of H1 but a large increase in the case of H2.

It is also possible that crystal symmetry plays a role in increasing variations in slow optical axis position. HEWL and fibronectin are uniaxial crystals, while glucose isomerase crystallizes in space group *I*222, a biaxial system; there are therefore two possible primary optical axes. The symmetry of a cubic crystal means that the indicatrix is a sphere; polarized light microscopy is therefore of limited interest for proteins crystallizing in cubic space groups. Crystallization and screening are also areas where there is great potential for the use of polarizing microscopes (Echalier *et al.*, 2004), although a supply of strain-free non-birefringent trays is required. Such trays are in development; for example, see Cherezov & Caffrey (2003).

The dependence of the linear birefringence on the solvent content provides scope for further work in determining the dominant cause of birefringence in the protein crystals studied and such a study has previously been carried out for haemoglobin (Perutz, 1953). The relative ease with which the position of a (non-cubic) crystal and loop can be isolated on an image means that this method is well suited for development of crystal positioning and alignment software.

The results reported above indicate that variations in SOAP across (or the 'SOAPiness' of) a protein crystal can be used as an indicator of the diffractive quality of a non-cubic protein crystal. As beamtime at synchrotron sites is a limited resource, this provides an opportunity for evaluating crystals before synchrotron trips and adds a potential selection step into high-throughput pipelines prior to crystal irradiation.

In this work, we have used very different wavelengths of electromagnetic radiation to sample the optical and diffractive properties of protein crystals (580 nm and 1.54 Å, respectively). It is thus perhaps not surprising that correlations that might have been expected (*e.g.* SOAP and mosaicity) were not observed. This clearly raises further questions as to the physical interpretation of the quantities we have measured and how these relate to the lattice order.

The authors would like to thank Richard Glazer for encouragement and the extended loan of a Metripol microscope, Morten Geday and staff at Oxford Cryosystems for extensive discussions and Enrique Rudiño-Piñera for supplying the fibronectin crystals. RLO is funded by a BBSRC CASE studentship.

References

Bragg, W. L. & Pippard, A. B. (1953). *Acta Cryst.* **6**, 865–867.

Cervelle, P. B., Cesbron, F., Berthou, J. & Jolles, P. (1974). *Acta Cryst.* A**30**, 645–648.

Cherezov, V. & Caffrey, M. (2003). *J. Appl. Cryst.* **36**, 1372–1377.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Darwin, C. G. (1922). *Philos. Mag.* **43**, 800–829.

Dauter, Z. (1999). *Acta Cryst.* D**55**, 1703–1717.

Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.

Echalier, A., Glazer, R. L., Fülöp, V. & Geday, M. A. (2004). *Acta Cryst.* D**60**, 696–702.

Garman, E. F. (1999). *Protein Crystallization: Techniques, Strategies and Tips. A Laboratory Manual*, edited by T. M. Bergfors, ch. 17. La Jolla, California, USA: International University Line.

Garman, E. & Schneider, T. (1997). *J. Appl. Cryst.* **30**, 211–237.

Geday, M. A., Kaminsky, W., Lewis, J. G. & Glazer, A. M. (2000). *J. Microsc.* **198**, 1–9.

Glazer, A. M., Lewis, J. G. & Kaminsky, W. (1996). *Proc. R. Soc. London Ser. A*, **452**, 2751–2765.

Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Kobayashi, J., Asahi, T., Sakurai, M., Kagomiya, I., Asai, H. & Asami, H. (1998). *Acta Cryst.* A**54**, 581–590.
Leslie, A. G. W. (1999). *Acta Cryst.* D**55**, 1696–1702.
McPherson, A. (2001). *Methods in Macromolecular Crystallography*, edited by D. Turk & L. Johnson, ch. 1. Erice, Italy: IOS Press.
Nye, J. F. (1984). *Physical Properties of Crystals: Their Representation by Tensors and Matrices.* Oxford: Clarendon Press.
Oldenbourg, R. & Ruiz, T. (1989). *J. Biophys.* **56**, 195–205.
Perutz, M. F. (1953). *Acta Cryst.* **6**, 859–864.
Pflugrath, J. W. (1999). *Acta Cryst.* D**55**, 1718–1725.
Rudiño-Piñera, E., Schwarz-Linek, U., Potts, J. R. & Garman, E. F. (2004). *Acta Cryst.* D**60**, 1341–1345.
Ruiz, T. & Oldenbourg, R. (1988). *J. Biophys.* **54**, 17–24.
Skrzypczak-Jankun, E., Bianchet, M. A., Amzel, L. M. & Funk, M. O. Jr (1996). *Acta Cryst.* D**52**, 959–965.
Wiener, O. (1912). *Abh. Math. Phys. Klas. Köngl. Sächs. Gesellsch. Wissensch.* **23**, 509–604.