Accounting for partiality in serial crystallography using ray-tracing principles

Loes M. J. Kroon-Batenburg,[a]* Antoine M. M. Schreurs,[a] Raimond B. G. Ravelli[b] and Piet Gros[a]

[a]Crystal and Structural Chemistry, Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands, and [b]M4I Division of Nanoscopy, Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands. *Correspondence e-mail: l.m.j.kroon-batenburg@uu.nl

Serial crystallography generates 'still' diffraction data sets that are composed of single diffraction images obtained from a large number of crystals arbitrarily oriented in the X-ray beam. Estimation of the reflection partialities, which accounts for the expected observed fractions of diffraction intensities, has so far been problematic. In this paper, a method is derived for modelling the partialities by making use of the ray-tracing diffraction-integration method *EVAL*. The method estimates partialities based on crystal mosaicity, beam divergence, wavelength dispersion, crystal size and the interference function, accounting for crystallite size. It is shown that modelling of each reflection by a distribution of interference-function weighted rays yields a 'still' Lorentz factor. Still data are compared with a conventional rotation data set collected from a single lysozyme crystal. Overall, the presented still integration method improves the data quality markedly. The R factor of the still data compared with the rotation data decreases from 26% using a Monte Carlo approach to 12% after applying the Lorentz correction, to 5.3% when estimating partialities by *EVAL* and finally to 4.7% after post-refinement. The merging R_{int} factor of the still data improves from 105 to 56% but remains high. This suggests that the accuracy of the model parameters could be further improved. However, with a multiplicity of around 40 and an R_{int} of ~50% the merged still data approximate the quality of the rotation data. The presented integration method suitably accounts for the partiality of the observed intensities in still diffraction data, which is a critical step to improve data quality in serial crystallography.

1. Introduction

X-ray free-electron lasers and high-brilliance undulator beamlines at synchrotrons have been used to perform serial (femtosecond) crystallography, collecting diffraction data from a large number (thousands up to millions) of micrometre-sized or nanometre-sized crystals (Chapman *et al.*, 2011; Boutet *et al.*, 2012; Redecke *et al.*, 2013; Gati *et al.*, 2014; Demirci *et al.*, 2013). Individual crystals may be hit by an X-ray pulse, thereby producing a diffraction pattern within the 10–50 fs pulse duration, before being vaporized by the transferred energy. This principle of 'diffraction before destruction' has been demonstrated by experiments on the Linac Coherent Light Source (LCLS) hard X-ray free-electron laser (Chapman *et al.*, 2011). Since the X-ray pulses are shorter than it takes for radiation-induced structural changes to occur, this approach of serial crystallography overcomes radiation damage, which has become a major problem with highly brilliant synchrotron sources (Weik *et al.*, 2000; Ravelli & McSweeney, 2000; Burmeister, 2000) using conventional rotation methods of collecting data from one or very few larger crystals. The diffraction images in serial crystallography
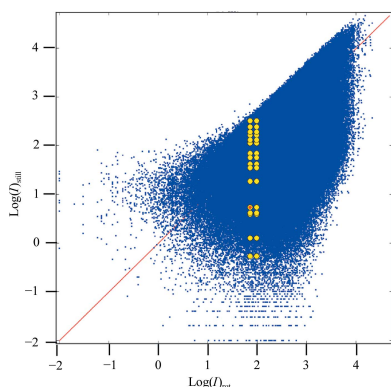
are single snapshots of nonrotating crystals: so-called still images. As opposed to the conventional rotation data, the reflections are not fully integrated but are partials, except possibly when using future pink XFEL beams (Dejoie *et al.*, 2015). The particular orientation of the crystal lattice determines the extent of this partiality, which is a great unknown in the data-reduction process.

The specific challenges in data processing are the indexing of the stills, the reconstruction of full intensities and the merging of data obtained from different crystals, in addition to the handling of huge amounts of data. Three software packages are available to process serial X-ray diffraction patterns: *CrystFEL* (White *et al.*, 2012, 2013; White, 2014), *cctbx.xfel* from the *Computational Crystallographic Toolbox* (Sauter *et al.*, 2013; Hattne *et al.*, 2014) and *nXDS* (Kabsch, 2014). For indexing, rotation-method indexing packages such as *MOSFLM* (Leslie & Powell, 2007), *DirAx* (Duisenberg, 1992) and *LABELIT* (Sauter *et al.*, 2004) are being used. In 2010, Kirian and coworkers proposed a Monte Carlo integration method that, by averaging large numbers of diffraction spots, averages out the unknown partialities as well as differences in crystal size, beam flux and the incident spectrum (Kirian *et al.*, 2010). Thousands of diffraction images are needed for this method to converge (Boutet *et al.*, 2012). It is generally believed (White, 2014) that estimation of partialities could reduce the number of images needed for the Monte Carlo integration method and could improve the data quality. Three approaches have been proposed to estimate partialities. All three use post-refinement to improve the partiality

correction factors and scale factors for each image. Kabsch (2014) derived an analytical expression for partiality from a Gaussian mosaic spread function. Comparison of ultrafine-sliced rotation images treated as stills or as normal rotation images gave satisfactory results. Kabsch includes a Lorentz factor for still data explicitly. The still data processing is not as good as one would expect, according to Kabsch. He concludes that this may be caused by two-dimensional rather than three-dimensional profile fits and the lack of other unimplemented corrections. White (2014) considers the overlap of reciprocal reflection volumes with a nest of Ewald spheres and calculates partialities from the distance of reciprocal-lattice points to the two limiting Ewald spheres. Using modelled data, White shows that the partiality estimation improve the data, with significant improvement of the statistics upon post-refinement. Most recently, Sauter (2015) and Uervirojnangkoorn *et al.* (2015) presented a partiality model that is implemented in *cctbx.xfel*. They calculated the intersection with the Ewald sphere of a spherical reciprocal-lattice point, where the radii of the lattice points are determined by mosaic spread and (asymmetric) beam divergence. Sauter (2015) also includes a parameter for the coherently scattering volume of mosaic blocks. Using this approach on XFEL data with post-refinement of crystal orientations, scale factors and beam parameters, the data are improved in quality as judged from molecular-replacement scores, structural refinement and anomalous difference maps (Uervirojnangkoorn *et al.*, 2015). Moreover, they show that reliable structures can be obtained with a lower number of images. Unfortunately, these authors do not mention merging *R* factors. Another correction that potentially improves Monte Carlo integration convergence for nanocrystals is explored by estimation of the crystal sizes and their corresponding diffraction power, as described by Qu *et al.* (2014). They show that the geometric correction factor, solely based on the maximum of the Laue interference function for each crystal with size $N_x \times N_y \times N_z$, is superior to Monte Carlo integration for simulated data. Although the above efforts were made to improve processed serial crystallography still data, many questions still need to be addressed. Why do the data not improve rigorously with the current partiality-correction models? What factors exactly determine the partiality? Which errors dominate the partiality-estimation schemes?

Here, we describe an extension of the *EVAL* profile-prediction algorithm to process still images. *EVAL* is a data-reduction method designed for integrating reflection intensities through profile fitting using ray-tracing simulations (Duisenberg *et al.*, 2003; Schreurs *et al.*, 2010). We derived a general interference function that is valid for crystals of any size and effectively includes the shape transform. The diffraction process is simulated by typically 10 000 rays, which are diffracted by an equal amount of reciprocal-lattice vectors. In the rotation method, we bring reciprocal-lattice vectors onto the Ewald sphere by rotation around the spindle axis. However, in the still diffraction method we calculate the deviation from the exact Bragg condition for each ray and estimate its contribution to the total diffracted intensity using the interference function. By summation, the partiality of a
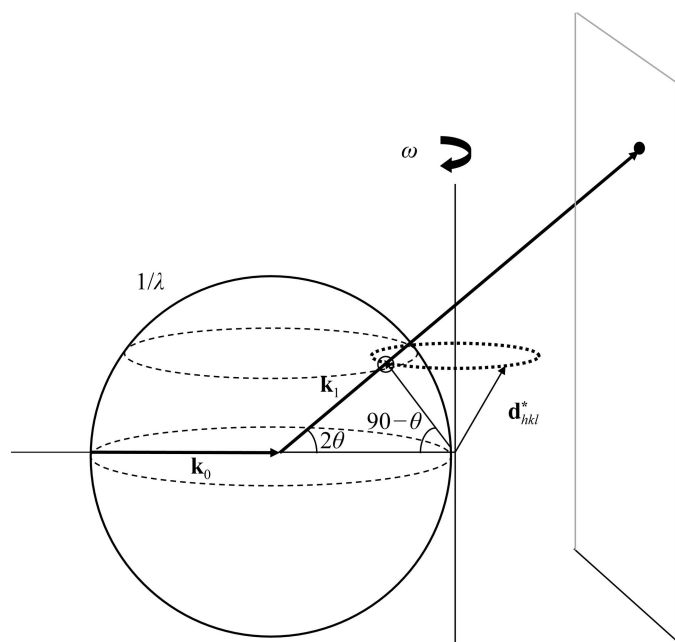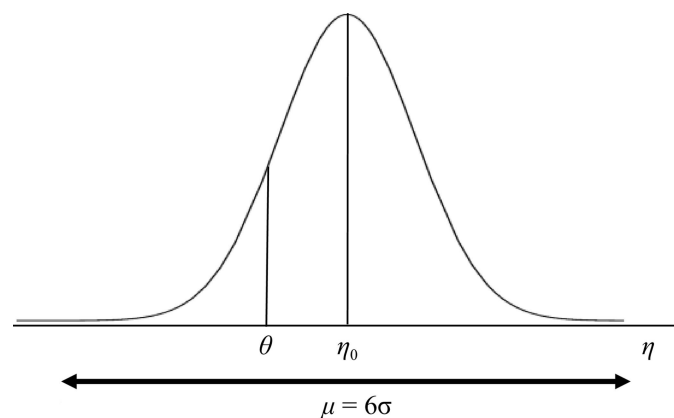

Figure 1
Ewald construction for a rotation experiment. An incident X-ray beam $\mathbf{k}_0$ with length $1/\lambda$ is reflected by a series of net planes in the direction $\mathbf{k}_1$ if the corresponding reciprocal-lattice vector $\mathbf{d}_{hkl}^*$ coincides with the Ewald sphere. In an arbitrary orientation of the crystal lattice $\mathbf{d}_{hkl}^*$ is not in a reflecting position but can be rotated around the spindle axis so that the Bragg condition is met and $\mathbf{d}_{hkl}^*$ makes an angle of $90° - \theta$ with the X-ray beam $\mathbf{k}_0$.

reflection is obtained and, as we will show, also the still Lorentz factor. To test the approach, we used two still data sets collected on our in-house diffractometer using a single lysozyme crystal: one consisting of consecutive, stepwise stills and one consisting of stills from arbitrary orientations. Both were compared with conventional rotation data collected under the same conditions. We show that for these data sets the reflection partialities can be estimated by the ray-tracing simulation method and that the presented approach significantly improves the mean intensities of the observed reflections.

2. Diffraction theory

Reflection profiles from a crystal in *EVAL* are simulated by generating ray traces. We consider a crystal to be built up from small crystallites by dividing the crystal on a three-dimensional grid (sampled from a distribution **K**) that can have random orientations taken from a mosaic distribution (**M**). Incident X-rays are emitted from a virtual focus (*e.g.* a square area **F**) in direction $\mathbf{k}_0$ with respect to the crystallite and with wavelength λ (sampled from a spectral distribution **L**). A crystallite with an orientation of the reciprocal-lattice vector $\mathbf{d}^*_{hkl}$ gives rise to a diffracted ray in direction $\mathbf{k}_1$ as determined by the Ewald construction. For **F**, **L**, **K** and **M** several statistical distributions are available (Schreurs *et al.*, 2010). In the simulation of rotation data the vectors $\mathbf{d}^*_{hkl}$ are rotated around the spindle axis so as to match the Bragg condition and then touch the Ewald sphere (Fig. 1).

In case of still diffraction experiments with one particular orientation of the crystal, none of the crystallite lattice vectors $\mathbf{d}^*_{hkl}$ are exactly on the Ewald sphere. However, some vectors are within a certain tolerance in angular deviation (ε) of the Bragg angle θ and may give rise to diffracted intensity that is a function of ε (see below). The integrated intensity of all vectors $\mathbf{d}^*_{hkl}$ from the various crystallites depends on the mosaic spread, the wavelength dispersion, the beam size and divergence, the crystal size and the crystallite size itself. The latter corresponds to the coherently diffracting volume of the

Figure 2
Angular spread of mosaic vectors. η_0 denotes the angle of the central mosaic vector (*i.e.* that which makes an angle of $90^\circ - \eta_0$ with the X-ray beam), which may be deviating from the ideal Bragg angle θ. The full mosaic spread is defined as six times the standard deviation of a Gaussian distribution.

mosaic blocks, and the total reflected intensity of the crystal is the incoherent sum of all diffracted rays.

2.1. Still diffraction images

The scattered intensity of a crystal at Bragg angle θ can be thought of as the coherent sum of scattering by s layers of thickness d, according to what we call the James–Buerger theory (James, 1958; Buerger, 1960). The scattered intensity of a single layer is

$$I = I_0 F^2 p \left(\frac{e^2}{mc^2}\right)^2 \lambda^2 \left(\frac{nsd}{\sin\theta}\right)^2, \qquad (1)$$

where F^2 is the squared structure factor, I_0 is the incident photon flux, p is the polarization owing to the reflection, e^2/mc^2 is the Thomson scattering length of one electron and n is the number of unit cells per unit volume. A mosaic crystal is made up of tiny crystallites with associated reciprocal-lattice vectors $\mathbf{d}^*_{hkl}$ that are spread over an angular range μ, and each of them may not be perfectly oriented to be in Bragg condition. The s layers within such a crystallite then scatter slightly out of phase and their scattered intensity is given by the interference function

$$I(\varepsilon) = I_0 F^2 p \left(\frac{e^2}{mc^2}\right)^2 n^2 \lambda^3 \frac{1}{2\sin^2\theta} V_{\text{crystallite}} \frac{1}{s} \frac{\sin^2(sB\varepsilon)}{(B\varepsilon)^2}, \qquad (2)$$

where ε is the deviation of the Bragg angle θ, $B = 2\pi d\cos\theta/\lambda$ and $V_{\text{crystallite}}$ is the reflecting volume of the crystallite.

The James–Buerger theory can be extended by writing the total diffracted intensity as an integral over all possible orientations of the crystallite vectors $\mathbf{d}^*_{hkl}$ that make angles of $90^\circ - \eta$ with the incident X-ray beam ($90^\circ - \theta$ in the Bragg condition) and replacing $V_{\text{crystallite}}$ by the volume of the crystal V and ε by $\theta - \eta$ (see Fig. 2). This results in

$$I = I_0 F^2 p \left(\frac{e^2}{mc^2}\right)^2 V_{\text{crystal}} n^2 \lambda^3 \frac{1}{2\sin^2\theta}$$
$$\times \frac{1}{s} \int_{\eta_{\min}}^{\eta_{\max}} \frac{\sin^2[sB(\theta - \eta)]}{[B(\theta - \eta)]^2} P(\eta|\eta_0) \, d\eta, \qquad (3)$$

where $P(\eta|\eta_0)$ is the probability distribution of η angles given the angle η_0 of the central reciprocal-lattice vector $\mathbf{d}^*_{hkl}(0)$ of the crystal as obtained from the unit-cell matrix.

The mosaic spread, the divergence of incident rays, the wavelength variations and the crystallite positions being slightly off-centre in a larger crystal are the cause of deviations ε for the individual rays. For the discussion here, we will concentrate on the mosaic spread, but the other parameters are accounted for as well in our ray-tracing simulations.

The distribution function can take several forms. Suppose that $P(\eta|\eta_0)$ is uniform, while $\int P(\eta|\eta_0)d\eta = 1$, then the integral over $d\eta$ in (3) reduces to $s\pi/B$ and

$$I = C \frac{1}{2\sin^2\theta} \frac{\pi}{B} = C \frac{1}{2\sin\theta\cos\theta}, \qquad (4)$$

where $C = I_0 F^2 p(e^2/mc^2)^2 \lambda^3 n^2$. The last term in (4) is familiar: it is the Lorentz factor for rotation in the equatorial plane and

research papers

is equal to the powder Lorentz factor. Thus, for still images the powder Lorentz factor applies. When $P(\eta|\eta_0)$ is a normal or an otherwise monotonous distribution, we should explicitly include it in the calculation of (3). However, it cannot be reduced to a simple trigonometric function nor to an erf (see Kabsch, 2014) because of the presence of the sinc function in the integral. Instead, it can be evaluated numerically.

2.2. Rotation diffraction images

Rotation images can be regarded as a superposition of many stills separated by an infinitesimal rotation angle ω. The integrated intensity for these is

$$I = C \frac{1}{2\sin^2\theta} \frac{1}{s} \int_{\omega_{\min}}^{\omega_{\max}} \int_{\eta_{\min}}^{\eta_{\max}} \frac{\sin^2[sB(\theta-\eta_\omega)]}{[B(\theta-\eta_\omega)]^2} P(\eta_\omega|\eta_{0,\omega}) \, d\eta_\omega \, d\omega.$$
(5)

In a sufficiently large ω scan each $\mathbf{d}_{hkl}^*$ vector makes a complete pass through the Ewald sphere, so that we can write

$$
\begin{aligned}
I &= C \frac{1}{2\sin^2\theta} \frac{1}{s} \int_{\eta_{\min}}^{\eta_{\max}} \int_{\omega_{\min}}^{\omega_{\max}} \frac{\sin^2[sB(\theta-\eta_\omega)]}{[B(\theta-\eta_\omega)]^2} \, d\omega \, P(\eta|\eta_0) \, d\eta \\
&= C \frac{1}{2\sin^2\theta} \frac{1}{s} \int_{\eta_{\min}}^{\eta_{\max}} P(\eta|\eta_0) \int_{\eta'_{\min}}^{\eta'_{\max}} \frac{\sin^2[sB(\theta-\eta')]}{[B(\theta-\eta')]^2} \frac{d\omega}{d\eta'} \, d\eta' \, d\eta \\
&= C \frac{1}{2\sin^2\theta} \frac{1}{s} \int_{\eta_{\min}}^{\eta_{\max}} P(\eta|\eta_0) \int_{\eta'_{\min}}^{\eta'_{\max}} \frac{\sin^2[sB(\theta-\eta')]}{[B(\theta-\eta')]^2} L' \, d\eta' \, d\eta \\
&= C \frac{1}{2\sin^2\theta} \frac{1}{s} \int_{\eta_{\min}}^{\eta_{\max}} P(\eta|\eta_0) \, s\pi/BL' \, d\eta = CL \int_{\eta_{\min}}^{\eta_{\max}} P(\eta|\eta_0) \, d\eta = CL,
\end{aligned}
$$
(6)

where L' is the duration component of the Lorentz factor L for the rotation experiment (equal to the reflection range expansion factor in Kabsch, 2014). The rotation Lorentz factor may alternatively be written as $1/[\mathbf{d}^*(\mathbf{k}_0 \times \boldsymbol{\omega})]$ (Milch & Minor, 1974). In case of rotation in the equator, (6) reduces to (4). In rotation data, therefore, the specific distribution function $P(\eta)$ is irrelevant to the integrated intensity and complete reflections are obtained.

2.3. Implementation of the interference function in EVAL

In EVAL a large number of vectors $\mathbf{d}_{hkl}^*(i)$ are generated from a Gaussian or Lorentzian two-dimensional mosaic distributions (M) and combined with vectors $\mathbf{k}_0$ from F, L, K distributions. The contribution of each of these to the scattered intensity is calculated with

$$I_i = \frac{1}{2\sin^2\theta} \frac{1}{s} \frac{\sin^2(sB\varepsilon_i)}{(B\varepsilon_i)^2}.$$
(7)

Summing all contributions gives the total scattered normalized intensity (i.e. $C = 1.0$; see text below equation 4), which is effectively an integral over $d\eta$, $d\mathbf{k}_0$ and, to a minor extent, $d\lambda$, because our beam is almost monochromatic. This normalized intensity is stored in the parameter 'partiality' after correction for the still Lorentz factor, i.e. the partiality is $\sum_i I_i/L_{\text{still}}$. The only new parameter introduced is the number of unit cells in

the crystallite N_{cell}, where $s = N_{\text{cell}}(|h| + |k| + |l|)$, the number of reflecting planes, while the crystallite size equals sd_{hkl}.

Every $\mathbf{d}_{hkl}^*(i)$ produces its own impact on detector pixel coordinates (x, y) and is weighted by contribution I_i. All impacts together build the two-dimensional reflection profile that is used as the model profile in the EVAL least-squares fit to obtain the observed integrated intensity for each diffraction spot on an image. Both the observed intensity and the summed interference function (7) contain the still Lorentz factor, and by dividing one by the other we extract F^2. We also correct for the polarization and apply possible incidence corrections.

2.4. Laue interference function

In this paper, we follow the James and Buerger approach, as explained in §2.1, for deriving diffracted intensities by crystals. The resulting interference function only depends on the deviation ε from the Bragg angle θ and the number of unit cells contained in the crystal. An alternative is to use the three Laue conditions, and the squared sinc function in (3) is replaced by

$$I(\Delta\mathbf{k}) = \left(\frac{\sin N_1\pi\Delta\mathbf{k}\cdot\mathbf{a}}{\sin\pi\Delta\mathbf{k}\cdot\mathbf{a}} \frac{\sin N_2\pi\Delta\mathbf{k}\cdot\mathbf{b}}{\sin\pi\Delta\mathbf{k}\cdot\mathbf{b}} \frac{\sin N_3\pi\Delta\mathbf{k}\cdot\mathbf{c}}{\sin\pi\Delta\mathbf{k}\cdot\mathbf{c}} \right)^2.$$
(8)

Here, $\Delta\mathbf{k} = \mathbf{k}_1 - \mathbf{k}_0$ and N_1, N_2 and N_3 are the number of unit cells in the three periodic axis directions. In Appendix A, we show that the two approaches are exactly the same.

(8) is often referred to as the shape transform of the crystal (Kirian et al., 2010; Spence et al., 2011).

2.5. Impact positions and refinement

Peak-position refinement in PEAKREF (Schreurs, 1999) minimizes the peak-position residuals and the Bragg angle deviation ε_0 of the central reciprocal-lattice vector either using peak maxima from the peak search or using optimized profile centroids from the EVAL profile fit. Inclusion of ε_0 in the unit-cell matrix refinement avoids divergence of unit-cell orientations through rotations perpendicular to the incident beam, as discussed by Sauter et al. (2014). Similarly, Kabsch (2014) uses the angular deviation τ divided by the mosaic spread σ_M in the target function for peak refinement. All three approaches use the β axis, defined for each reflection as the axis perpendicular to the incident and diffracted beams, to calculate the deviation from the Bragg angle θ (Schutt & Winkler, 1977). For each still image, the following target function was minimized to refine the unit-cell matrix,

$$\Psi = \frac{\sum_i^{N_{\text{peak}}} [(\Delta x_i)^2 + (\Delta y_i)^2]^{1/2}}{N_{\text{peak}}} + \frac{\sum_i^{N_{\text{peak}}} |\theta_i - \eta_{0,i}|}{N_{\text{peak}}},$$
(9)

where Δx_i and Δy_i are the differences of observed and calculated peak positions. PEAKREF can optimize many instrumental parameters such as detector-offset positions, primary beam direction and crystal position, which in the current analysis were fixed in the still data refinement and based on the rotation data (see below).

We found that the peak-position residuals from the post-refinement were much smaller than from the peak maxima found on a single still image, despite the much larger number of peaks. This was caused by a shift in the observed θ value for the partial reflections with large ε_0. For large mosaic crystals such as our lysozyme crystal measured with a divergent beam, these shifts in θ occur because only distinct directions of the primary rays or distinct points on the crystal are active dependent on the deviation ε_0 (Fig. 3). A negative value for ε_0 results in an apparent larger θ and a positive ε_0 in an apparent smaller θ. This θ-divergence effect has to be taken into account when the cell matrix is determined and refined from peak positions. We introduced a parameter 'flex' in *PEAKREF* that is jointly refined and takes account of this shift. The 'flex' parameter turned out to have a constant value for all still images and it appears to be a property typical for the crystal and the beam divergence of the particular experiment.

2.6. Post-refinement

We implemented a post-refinement procedure in which both the peak positions from the *EVAL* integration and the partialities could be refined. For this purpose, we calculate the mean intensity of all equivalent reflections h as

$$\langle I^h \rangle = \frac{\sum_e^{N^h} w_e s_f / p_e I_e^h}{\sum_e^{N^h} w_e s_f / p_e}. \tag{10}$$

Weights are obtained from the standard deviations from the *EVAL* profile fit (Schreurs *et al.*, 2010) and are given by $w_e = 1/\sigma_e^2$. The partialities p_e arise from the *EVAL* ray-tracing simulation, and the image scale factors s_f are determined in *ANY* (Schreurs, 2007), assuming a constant sum of Bragg intensities in each frame. The summation in (10) runs over all equivalents of reflection h (N^h) in the data set. In *PEAKREF* image scale factors s_f' and unit-cell parameters and crystal orientation angles are refined using the target function

$$\Psi = \frac{\sum_i^{N_{\text{peak}}} [(\Delta x_i)^2 + (\Delta y_i)^2]^{1/2}}{N_{\text{peak}}} + \frac{\sum_i^{N_{\text{peak}}} w_i |s_f' I_i - p_i' \langle I^h \rangle|}{\sum_i^{N_{\text{peak}}} w_i s_f' I_i}. \tag{11}$$

We specifically include peak positions in this refinement step in order to avoid unwanted divergence from peak position-derived unit-cell parameters and orientations. In this post-refinement step the *EVAL* ray-tracing procedure is not repeated to obtain partialities; instead, we use a fitted partiality *versus* ε curve with a single Gaussian. The parameters in the Gaussian were kept fixed in the refinement; ε changes with the unit-cell parameters from which we recalculate the partiality (p_i').

3. Materials and methods

3.1. Crystal preparation

Hen egg-white lysozyme (Sigma–Aldrich, Schnelldorf, Germany) was crystallized using the hanging-drop vapour-diffusion method with a protein concentration of 75 mg ml^{-1} in 0.1 M sodium acetate buffer pH 4.8. The precipitant consisted of 0.1 M sodium acetate buffer pH 4.8, 10–15%(w/v) sodium chloride, 30%(v/v) ethylene glycol (Sutton *et al.*, 2013). Drops of 4 µl were set up with a 1:1 protein:precipitant ratio.

3.2. Data collection

A crystal of dimensions $250 \times 250 \times 150$ µm was vitrified in a cold N_2-gas stream from an Oxford Instruments 700 series jet operated at 100 K. Data were collected on a Bruker–AXS X8 Proteum in-house source with Cu $K\alpha$ radiation. The rotating anode was operated at 45 kV and 60 mA. The reference rotation data set was collected by rotating over 190° in φ in 0.5° steps per frame. Data were recorded on a PLATINUM^{135} CCD detector with a sample-to-crystal distance of 52 mm. 380 still images were collected with identical angular settings as the starting angles for each of the rotation frames; thus, 380 still images were collected at 0.5° intervals. An additional 394 stills were recorded by random selections from ω scans 0–7° in ω apart at 15 different ω, κ and φ goniometer settings. The exposure time for all images was 5 s.

3.3. Data processing and analysis

VIEW was used for image display and peak search (Schreurs, 1998). Both the rotation images and stills were
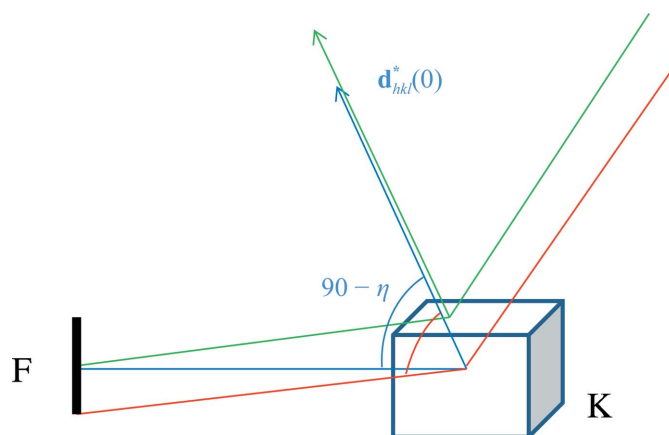

Figure 3
Schematic illustration of conditions causing an apparent shift in 2θ. A mosaic vector $\mathbf{d}_{hkl}^*(0)$ is slightly off the Bragg condition for an X-ray beam from the centre of the focus $\mathbf{F}$ (blue line) and makes an angle of $90° - \eta$ with the X-ray beam. However, the Bragg condition tends to be fulfilled with rays emerging from the lower part of the focus (red line; the arc corresponds to an angle of $90° - \theta$) or for the upper part of the crystal (green line). Therefore, the reflected beam appears at a larger 2θ angle. Owing to averaging of many rays this effect is not, or is hardly, visible for rotation images, but it is for still images or ultrafine-sliced images. As the shift in apparent 2θ is always in the direction of 2η, we can account for this by the 'flex' parameter in *PEAKREF*.

Table 1
Indexing and peak-refinement statistics.

	EVAL, rotation	CrystFEL, all stills	EVAL, consecutive stills Single unit cell	EVAL, consecutive stills Unit cell per image	EVAL, all stills	Post-refined
No. of images	380	774	380	380	774	
No. of peaks used for indexing	649	300–490†	10728	259–300†	289–300†	
No. of peaks used in refinement	101246		217272	113418	230759	396201
Unit-cell parameters						
$\quad a, b$ (Å)	78.666	78.735–79.453	78.834	78.765–79.041	78.679–79.119	78.719–79.049
$\quad c$ (Å)	36.819	36.819–37.481	36.842	36.706–36.906	36.706–36.989	36.810–36.980
Orientation r.m.s. (°)						
$\quad a*$			0.028		0.102	0.103
$\quad b*$			0.038		0.108	0.095
$\quad c*$			0.035		0.107	0.111
Average residuals						
$\quad$ Horizontal, vertical (mm)	0.0622		0.0884	0.0785	0.0766	0.099
$\quad \varepsilon_0$ (°)	0.0390		0.1850‡	0.1474‡	0.1496‡	0.173‡

† Per image. ‡ Average angular deviation of central reciprocal-lattice vector **d*** with Ewald sphere (see text for explanation).

indexed using *DirAx* (Duisenberg, 1992). Almost all stills could be indexed without manual intervention. Bravais lattice constraints were applied and the unit cells were made congruent (using the goniometer positions), ensuring a consistent choice of unit-cell axes. The unit-cell matrix and detector positions were refined from 649 peak positions in the rotation data. For the still peak positions we used different refinement options. In the first approach we made use of our knowledge of the relative positions of the goniometer axes, so that a global single unit-cell matrix could be refined against 10 728 peak positions. In the second approach, we determined and refined a unit-cell matrix for each image from 300 peak positions, as would be the normal procedure in serial crystallography. The detector-offset positions were taken from the peak-position refinement of the rotation data. The unit-cell matrix was refined against the observed peak positions, using the 'flex' parameter to account for apparent shifts in θ, simultaneously minimizing the off-Bragg angle ε_0 (9). Using the unit-cell matrix, we extracted three-dimensional and two-dimensional reflection boxes for rotation and still images, respectively, and processed these with *EVAL*. For every reflection, 10 000 rays were simulated and the impacts were collected in pixels contained in the box. In case of still data every individual ray is associated with a reciprocal-lattice vector $\mathbf{d}^*_{hkl}(i)$ with a small angular deviation from the Ewald sphere ε_i and is weighted by the interference function (7). The impact position on the detector is given by the direction of the shortest distance of the reciprocal-lattice vector to the Ewald sphere. The divergence effects are accounted for in the ray tracing and thus the profiles are generated correctly at deviating positions in θ (i.e. without the need for a 'flex' parameter as used at the peak-refinement stage).

The parameters for crystal size, mosaic spread and beam divergence were optimized automatically in the reflection profile fitted to ∼50 reflections with $I/\sigma(I) > 20$ using a simplex method (see Schreurs *et al.*, 2010). For comparison reasons, identical values of parameters in the ray-tracing simulations were used for both types of data sets, although a similar optimization can be performed for still images. In addition, for the still images we used $N_{cell} = 25$ in the interference function (the number of unit cells in a crystallite as described in §2.3). Sampling of the interference function converges much faster with low values of N_{cell}, typical for nano-sized crystals. The current data imply a larger value of N_{cell}, which in the current implementation would require many more rays (up to 10^6 instead of 10 000) to sample reflection profiles smoothly. The integrated intensities are obtained by a least-squares fit of the three-dimensional and two-dimensional model profiles to the observed pixel intensities for the rotation and still reflections, respectively. *EVAL* then delivers the profile-fitted, Lorentz- and polarization-corrected intensity values in an XML-type data-file that is further processed in *ANY* (Schreurs, 2007). In this program, we determine image scale factors, correct for the partiality factor and output the intensities and standard deviations to an hkl- or mtz-type file. Many of the graphical plots and statistical analyses are made using *ANY*.

All still images were also processed with the *CrystFEL* software suite v.0.5.1 (White *et al.*, 2012). Structural refinements were carried out with *REFMAC*5 (Murshudov *et al.*, 2011) and scaling between data sets with *SCALEIT* (Howell & Smith, 1992), both from the *CCP*4 suite (Winn *et al.*, 2011). *ANODE* (Thorn & Sheldrick, 2011) was used to calculate anomalous difference densities.

4. Results

We collected rotation and still diffraction data from one lysozyme crystal and formed three data sets for analysis: a 190° rotation data set collected in 380 images in ranges of 0.5° for reference, a consecutive still data set of 380 images collected in steps of 0.5°, and this consecutive data set combined with 15 wedges of separate arbitrary orientations totalling 774 images (Table 1). Indexing and peak refinement by *PEAKREF*/*EVAL* and *CrystFEL* yielded unit-cell dimensions that varied by ∼0.3–0.7% between the separate stills. Initially, the average residuals in peak positions for the still data were significantly larger than for the rotation data. Introduction of the 'flex' parameter, which takes into account the apparent shift in θ owing to divergence effects (see §2.5), reduced the positional residuals of peak maxima on still images significantly: from 0.13 to 0.07–0.08 mm on average. This deviation is only slightly larger than that observed for the rotation data, which was 0.06 mm. The residual in rotation angle for the rotation data was 0.039° (for 0.5° scan width). For the still data the deviation from the Bragg angle ε_0 was 0.18°, which is consistent with a mosaic spread of 0.5°. Relaxing the unit cells by a separate matrix for each image lowered the ε_0 residuals to 0.14°

Table 2
Statistics for the integration and merging of the rotation and still data.

Unit cell	EVAL, rotation	CrystFEL, all stills	EVAL, Monte Carlo		EVAL, Lorentz corrected		EVAL, partiality corrected		
			Consecutive stills	All stills	Consecutive stills	All stills	Consecutive stills		All stills
			Image	Image	Image	Image	Single	Image	Image
No. of reflections	106021	733504					320139	325753	657782
No. unique†	8291						8301	8328	8352
Multiplicity	12.8	88.4					38.6	39.1	78.8
Completeness (%)	100	100					100	99.0	99.2
$CC_{1/2}$ (%)	100	—	87.4 (80.3)	94.1 (88.6)	88.6 (79.9)	94.3 (88.5)	97.3 (90.9)	95.3 (88.1)	97.5 (90.9)
$\langle I/\sigma(I)\rangle$	13.9 (1.7)		‡	‡	‡	‡	‡	‡	‡
$\langle I/\sigma(I)\rangle$, merged	47.7 (20.3)	10.1 (5.7)	8.0 (5.1)	11.5 (7.4)	8.1 (5.4)	11.8 (7.7)	12.0 (7.4)	11.0 (6.6)	15.1 (9.0)
R_{int}§ (%)	3.8 (7.1)	—	106.2 (94.5)	104.9 (91.2)	108.0 (97.1)	106.5 (94.4)	57.5 (58.1)	62.6 (62.6)	63.8 (63.4)
Post-refinement									
$\quad\langle I/\sigma(I)\rangle$, merged							13.2 (7.9)	13.8 (8.2)	18.4 (10.8)
$\quad R_{int}$§ (%)							55.7 (56.1)	54.4 (54.5)	55.7 (55.1)

† Point group 4/*mmm*. ‡ Still data are not scaled by *SADABS* like the rotation data and no error model is determined for σ. In the merging step σ is determined from the internal standard deviation $\sum |I - \langle I\rangle|/N(N-1)^{1/2}$. § $R_{int} = \sum_h^N \sum_e^{N^h} (|I_e - \langle I^h\rangle|)/\sum_h^N \sum_e^{N^h} I_e$, where the summations runs over all N unique reflections h and equivalents.

(compare 'single unit cell' *versus* 'unit cell per image' in Table 1). In our setup the orientation of each matrix was known because the crystal orientation was set using a goniometer. The r.m.s. deviations between the set and refined orientations of the reciprocal-lattice vectors **a***, **b*** and **c*** were 0.03° for the consecutive still data and increased to ~0.10° using all still data (consecutive and random orientations). Overall, the number of observations taken into account by *EVAL* were 106×10^3 for the rotation data set, 325×10^3 for the consecutive still data and 657×10^3 for all still data, whereas *CrystFEL* took 733×10^3 into account for all still data (Table 2). All processed sets resulted in ~8300 unique reflections. The multiplicity of the consecutive still data was roughly three times that of the rotation data, indicating that reflections were, on average, sliced through three times in our still data-collection experiment.

The statistics for the integration and merging of data for the rotation and still data are shown in Table 2. Processing of the reference rotation data yielded an internal merging R_{int} of 3.8% with an $\langle I/\sigma(I)\rangle$ after merging of 47.7. Processing of the still diffraction data without correction, referred to as Monte Carlo averaging, produced R_{int} values exceeding 100% and $\langle I/\sigma(I)\rangle$ values that were about fourfold lower than that for the rotation data using the same number of images. Application of the still Lorentz correction (4) slightly increased the R_{int} (Table 2).

To estimate partialities, we determined the parameters for mosaic spread, divergence of the incident beam, crystal size and N_{cell} by optimizing two-dimensional profile fits using figures of merit (Schreurs *et al.*, 2010) on a subset of reflections in *EVAL*. Mosaic spread was set to 0.5°, beam divergence to 8.6 mrad, crystal size to $130 \times 130 \times 130$ μm (although we estimated a slightly larger size when selecting the crystal under the microscope) and N_{cell} to 25. The ray-tracing procedure yielded partialities which showed a Gaussian-like distribution with ε_0 (Fig. 4a). Notably, the computed still partialities are not normalized and exceed a value of 1, and hence are used as relative scale factors. In rotation data the partiality is defined up to 1 for a fully observed reflection (Rossmann & Beek,

1999); in contrast, the partiality in still diffraction is determined by the angular width of the intersection with the Ewald sphere, which depends on various instrumental and crystal parameters such as those given by (3). Lorentz-corrected still and (Lorentz-corrected) rotation reflections on average give the same absolute intensities. Fig. 5 shows that some still partialities are larger than 1.0 and the still intensities scatter around the rotation intensity. Further, to illustrate that the partialities depend strongly on the precise ray-tracing model parameters, Fig. 4(b) shows the partialities as a function of ε_0 in the case of a long focus for the incident beam, which results in two Gaussian-like curves superimposed. This implies that a simple Gaussian model for the partiality is not always correct. When divided into ε_0 bins, the observed average intensities correlate well with the estimated partialities (Fig. 6). Application of the partiality model resulted in average $I/\langle I\rangle$ values that varied around the ideal value of 1.0. Subsequent merging of these data, *i.e.* with both Lorentz and partiality corrections applied, reduced the R_{int} values to 57 and 63% for the data sets with consecutive and all stills, respectively.

Next, the effects of Lorentz and partiality correction were evaluated by comparing the data with the reference rotation data set. The uncorrected and the Lorentz-corrected intensities have high internal R_{int} values of 104.9 and 106.5%, respectively, consistent with the scattering in Fig. 5. The Lorentz- and partiality-corrected intensities have an R_{int} of 63.8%. Upon merging the data to unique reflections the agreement with the rotation data improved dramatically; the scatter diagrams in Fig. 7(a) and 7(b) reflect the improvement corresponding to the uncorrected (Monte Carlo) and corrected (Lorentz and partiality) data. The effects from the still data corrections are more clearly demonstrated by the R factors with respect to the reference rotation data, which we refer to as R_{comp} (Table 3). R_{comp} (on intensities) was 26% using Monte Carlo averaging. Application of the Lorentz correction alone decreased the R_{comp} to 12%. Application of both Lorentz and partiality corrections yielded an R_{comp} of 5.3%.

Table 3
Comparison of stills with *EVAL* rotation data.

|	CrystFEL, Monte Carlo	EVAL, Monte Carlo		EVAL, Lorentz correction only		EVAL, partiality corrected		EVAL, partiality corrected, post-refined
	All stills	Consecutive stills	All stills	Consecutive stills	All stills	Consecutive stills	All stills	All stills
R_{comp}† (%)	32.0	26.7	26.4	12.9	12.0	6.5	5.3	4.7
R on F‡ (%)	18.9	14.5	15.2	10.0	9.0	4.9	4.1	3.1

† Data merged in point group 4/*mmm*. ‡ From *SCALEIT*: overall scale.

Although the Lorentz and partiality corrections significantly improved the quality of the merged data, the merging R_{int} value remained high (*i.e.* 63.8% for all still data). To
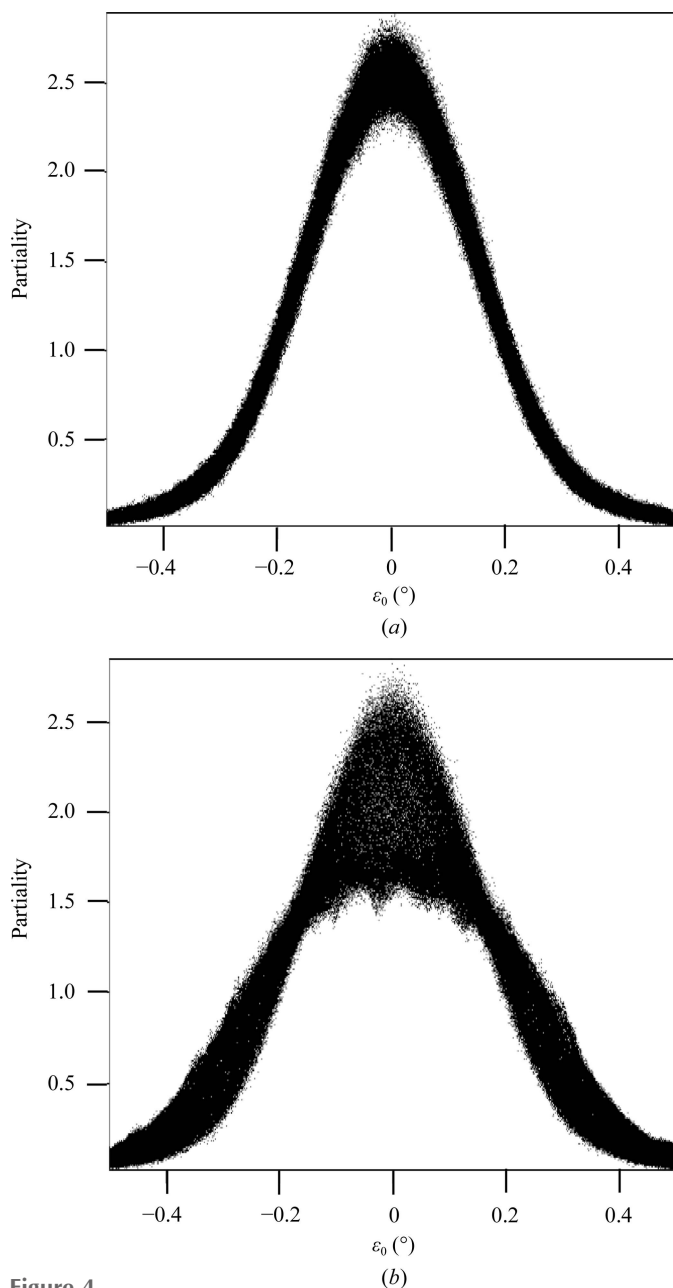
Figure 4
Partialities of reflections plotted against their ε_0. (*a*) From the ray-tracing model used in this paper, (*b*) from ray tracing using a long focus.

improve the partialities, we performed post-refinement of the image scale factor, unit-cell parameters and orientations, minimizing the target function of (11). Post-refinement of the 'all stills' data gave scale factors of 0.84–1.35 (additional to the scale factor s_f used in equation 10) and sharpened the distribution of unit-cell dimensions, with virtually no effect on the variation of crystal orientations (Table 1). These adjustments resulted in a significant, but modest, reduction of R_{int} from 63.8 to 55.7% (Table 2). The progress in the precision of processing the data is reflected by the distributions $I(hkl)/\langle I(hkl)\rangle$ shown in Fig. 8. Ideally, $I(hkl)/\langle I(hkl)\rangle$ values form a sharp distribution around 1 (as a reference, we depict the distribution resulting from the rotation data in Fig. 8*e*). Figs. 8(*b*) and 8(*c*) reflect the striking improvement obtained by modelling the partiality in *EVAL* and subsequent post-refinement. Fig. 8(*d*) shows that mainly the weak data do not profit from the post-refinement. Comparison of the merged data sets shows that the improvement in precision is matched
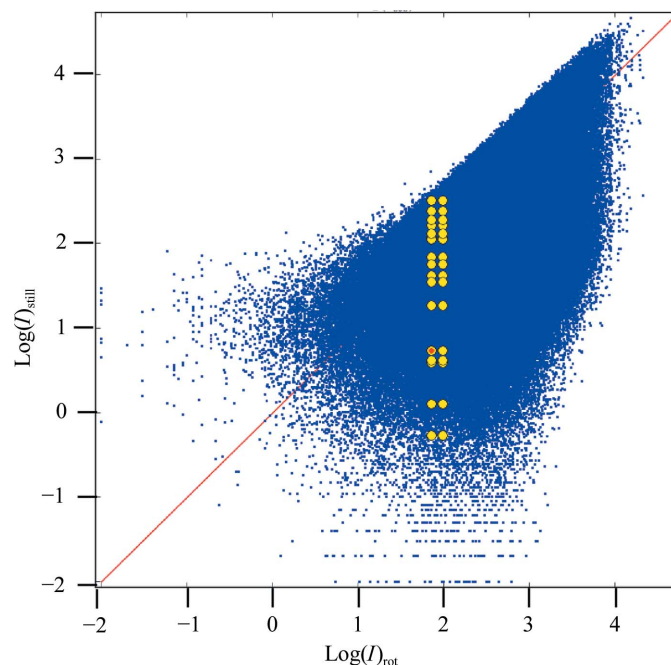

Figure 5
Lorentz-corrected still (not partiality-corrected) *versus* Lorentz-corrected rotation intensities. All occurrences of reflection 1, −6, 5 are marked by yellow dots. No scaling was applied so that intensities are on an absolute scale. Dots above the red line with slope = 1.0 indicate that the observed still intensity is larger than its corresponding rotation intensity, resulting in a partiality larger than 1.0. On average, the equivalent observations of a unique still reflection are equal to the rotation intensity (see text).

Table 4
Comparison of data quality.

	EVAL	*CrystFEL*, Monte Carlo	*EVAL*, Monte Carlo		*EVAL*, Lorentz correction only		*EVAL*, partiality corrected		*EVAL*, partiality corrected, post-refined
	Rotation	All stills	Consecutive stills	All stills	Consecutive stills	All stills	Consecutive stills	All stills	All stils
Refinement†									
R_{work} (%)	15.9	17.4	16.8	16.7	16.7	16.6	15.9	16.0	16.0
R_{free} (%)	19.8	21.4	20.3	20.2	20.2	20.2	20.0	20.0	20.2
$\langle B \rangle$ (Å^2)	13.8	25.0	20.7	20.6	11.8	11.8	13.0	13.2	13.4
Average anomalous densities‡ (σ)									
No. of reflections§	6420	6614	6293	6595	6266	6571	6091	6403	6406
Met SD	13.3	3.2	3.2	4.2	3.6	4.6	5.2	6.3	7.2
Cys SG	10.9	2.1	2.9	3.3	2.8	3.3	3.8	5.2	6.0
Cl$^-$	5.0	1.0	1.0	1.0	1.1	1.2	1.7	2.1	2.3
Na$^+$	1.8	—	1.0	0.8	2.0	0.8	2.0	1.5	1.5

† Restraint refinement, individual isotropic B factors, input structure PDB entry 193l. ‡ *ANODE* with data merged in point group 422. Averaged densities over similar atom types (two for Met SD, eight for Cys SG, 14 for Cl$^-$ and one for Na$^+$). § Selected by *SHELXC*.
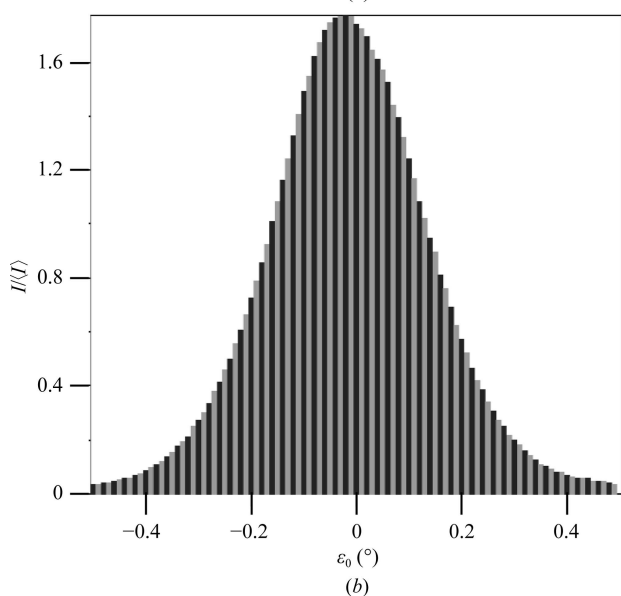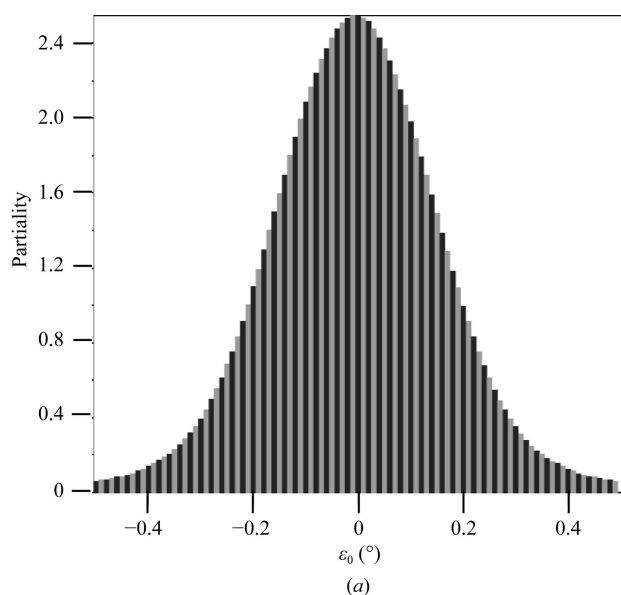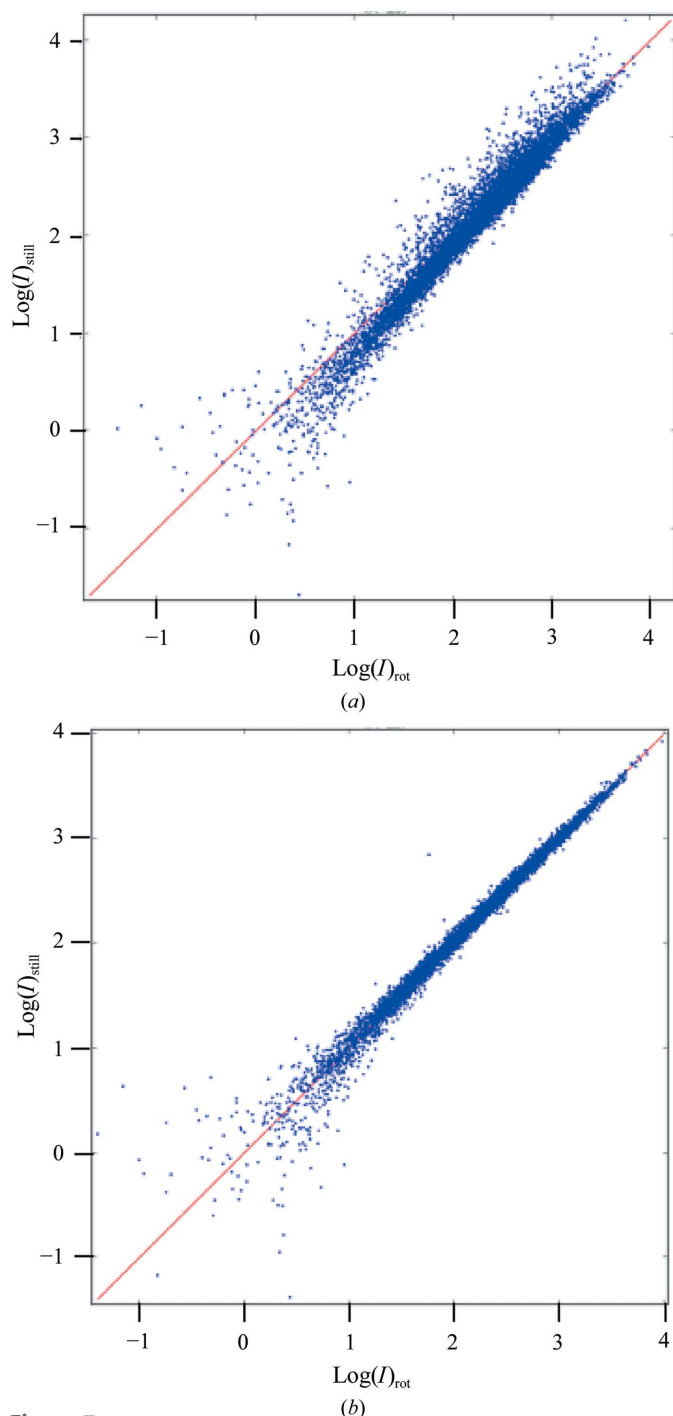

by an improvement in accuracy. Post-refinement reduced the R_{comp} from 5.3 to 4.7% (Table 3).

To illustrate the data quality, we refined the lysozyme crystal structure and computed anomalous difference densities. The structure was refined starting from PDB entry 193l (Vaney *et al.*, 1996) against the reflection data using *REFMAC*, and we observed similar R_{work} and R_{free} values for the differently processed data (Table 4). Significant differences between the methods were observed for the resulting average isotropic B factors. Monte Carlo averaging of the data in *CrystFEL* and *EVAL* yielded increased B factors (21–25 Å^2) compared with the reference defined by the rotation data set ($\langle B \rangle = 13.8$ Å^2). The Lorentz correction had a large effect on the B factors and produced an average B factor of 11.8 Å^2; this large effect on the B factors is explained by a comparable fall-off in θ of the Lorentz factor and the temperature factor. When the Lorentz
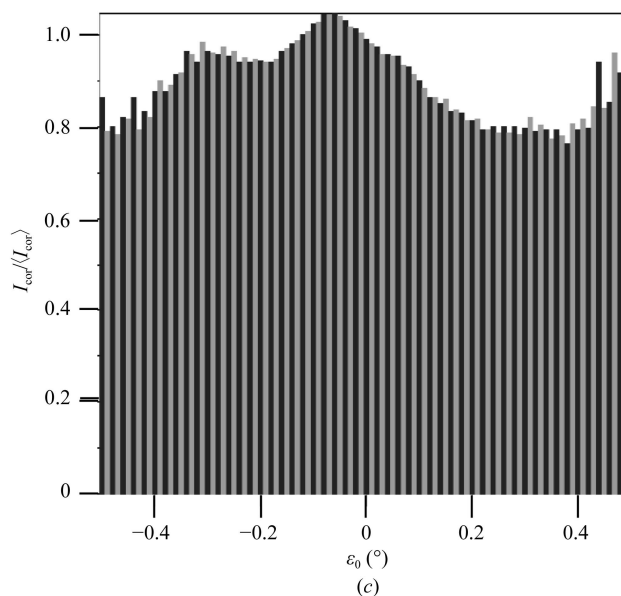
Figure 6
Histogram of reflection data *versus* ε_0. (*a*) Partialities, (*b*) $I/\langle I \rangle$ for equivalent reflections and (*c*) $I/\langle I \rangle$ after partiality correction.

and partiality corrections were both applied, the B factors became more similar to those obtained when using the rotation data (13.2 *versus* 13.8 Å^2). Anomalous differences are much more sensitive to the accuracy of the data than structure refinement. We generated anomalous difference densities based on the processed data sets using phases from the refined structure by *ANODE*. For the methionine sulfur positions the anomalous density from the rotation data gave a peak height

Table 5
Comparison of data with reduced statistics (60 frames).

	EVAL, Monte Carlo		*EVAL*, partiality corrected	
	Consecutive stills	Random stills	Consecutive stills	Random stills
Refinement†				
R_{work} (%)	20.7	25.6	17.3	20.7
R_{free} (%)	25.1	32.3	21.8	26.3
$\langle B \rangle$ (Å^2)	20.9	21.6	14.3	13.8
Average anomalous densities‡ (σ)				
No. of reflections§	862	1265	2212	2166
Met SD	0.6	0.8	—	1.6
Cys SG	—	0.8	1.1	—
Cl$^-$	—	—	—	—
Na$^+$	—	—	—	—

† Restraint refinement, individual isotropic B factors, input structure PDB entry 193l.   ‡ *ANODE* with data merged in point group 422. Averaged densities over similar atom types (two for Met SD, eight for Cys SG, 14 for Cl$^-$ and one for Na$^+$).   § Selected by *SHELXC*.

of 13.3σ. The uncorrected, Monte Carlo averaged still data yielded a weak anomalous signal: a 4.2σ peak for methionine S, corresponding to 32% of the peak height using the rotation data. Lorentz correction improved the methionine S signal to 35%, whereas including partiality corrections resulted in 47% of the signal. Finally, this signal improved to 54% after post-refinement. This shows that both Lorentz and partiality correction improved the intensities deduced from the still data.

We tested the effect of data-set size by limiting the still data to 60 images (Table 5). For the reduced 'consecutive still' data we used images 250–310. For the 'random still' data 60 images from three different wedges were used. For these limited data sets (91.7 and 97.3% completeness, respectively), the R_{free} factors show that the structure quality deteriorated. Furthermore, the anomalous signal is largely lost. For both structure refinement and anomalous density analyses the Lorentz and partiality-corrected data outperform the noncorrected Monte Carlo processed data.

5. Discussion and conclusions

We used our ray-tracing profile-prediction methods to model partialities of the observed reflections in still diffraction data and adapted the programs *PEAKREF* and *EVAL* to process still diffraction images. By taking experimental conditions into account, we compute 10 000 rays generated from focus, crystal grid points, wavelength spectrum and mosaic distributions, and calculate the interference-function weighted contribution to an observed reflection and hence derive its partiality. Our formalism implicitly models for the Lorentz factor, mimicking the contribution of the Lorentz factor to the observed intensities. Our approach differs fundamentally from other still data-processing methods. Kabsch (2014) defined an analytical erf function for the partiality, which is the integral over a Gaussian mosaic function. It is equivalent to our integral in (3) for an infinitely sharp sinc function (implying that integration

Figure 7
Scatter plots of still *versus* rotation reflection intensities. (*a*) Uncorrected Monte Carlo and (*b*) corrected reflection intensities after merging in point group 4/*mmm*.

over this function is complete within a solid angle smaller than the pixel size of the detector), while ignoring broadening effects other than the mosaic spread. Kabsch explicitly corrects for the still Lorentz factor. White (2014) and Sauter (2015) use reciprocal-lattice point volumes for calculating the partiality. White (2014) accounts for spectral width and beam divergence by calculating the overlap of a reciprocal-lattice volume with a nest of Ewald spheres. Sauter (2015) and Uervirojnangkoorn *et al.* (2015) use a single Ewald sphere and calculate the intersection with a spherical reciprocal-lattice volume, the size of which is determined by beam divergence,

mosaic spread and spectral dispersion. Both approaches account for increase of reciprocal diffracting volume with resolution, and in this way for the wider range of acceptable off-Bragg angles ($d\varepsilon$; see Appendix *A*). However, both approaches lack the reflectivity part of the Lorentz factor ($d\Omega$; see Appendix *A*). If the spectral width of the beam becomes large, an additional Lorentz factor needs to be accounted for, as used in the Laue method (Zachariasen, 1945). Uervirojnangkoorn *et al.* (2015) very recently presented their results on XFEL data. They showed that the R_{work} and R_{free} of refined structures improved and part of the anomalous signal was retrieved. Unfortunately, they do not provide merging R_{int} or a comparison to a rotation data set, *i.e.* R_{comp}, to evaluate the resulting quality of the data more directly. In our approach, the integration of (3) is achieved by simulation of the rays that contribute to an observed reflection spot. Because of the simulation, the derivation of analytical functions for the various effects is not needed and the Lorentz effect is implicitly taken into account. Moreover, the interference function can be taken into account in our approach.

For the initial development of the method, we used an experimental setup that allowed a direct comparison to the conventional rotation method. Our analysis showed a dramatic improvement in data quality after partiality and
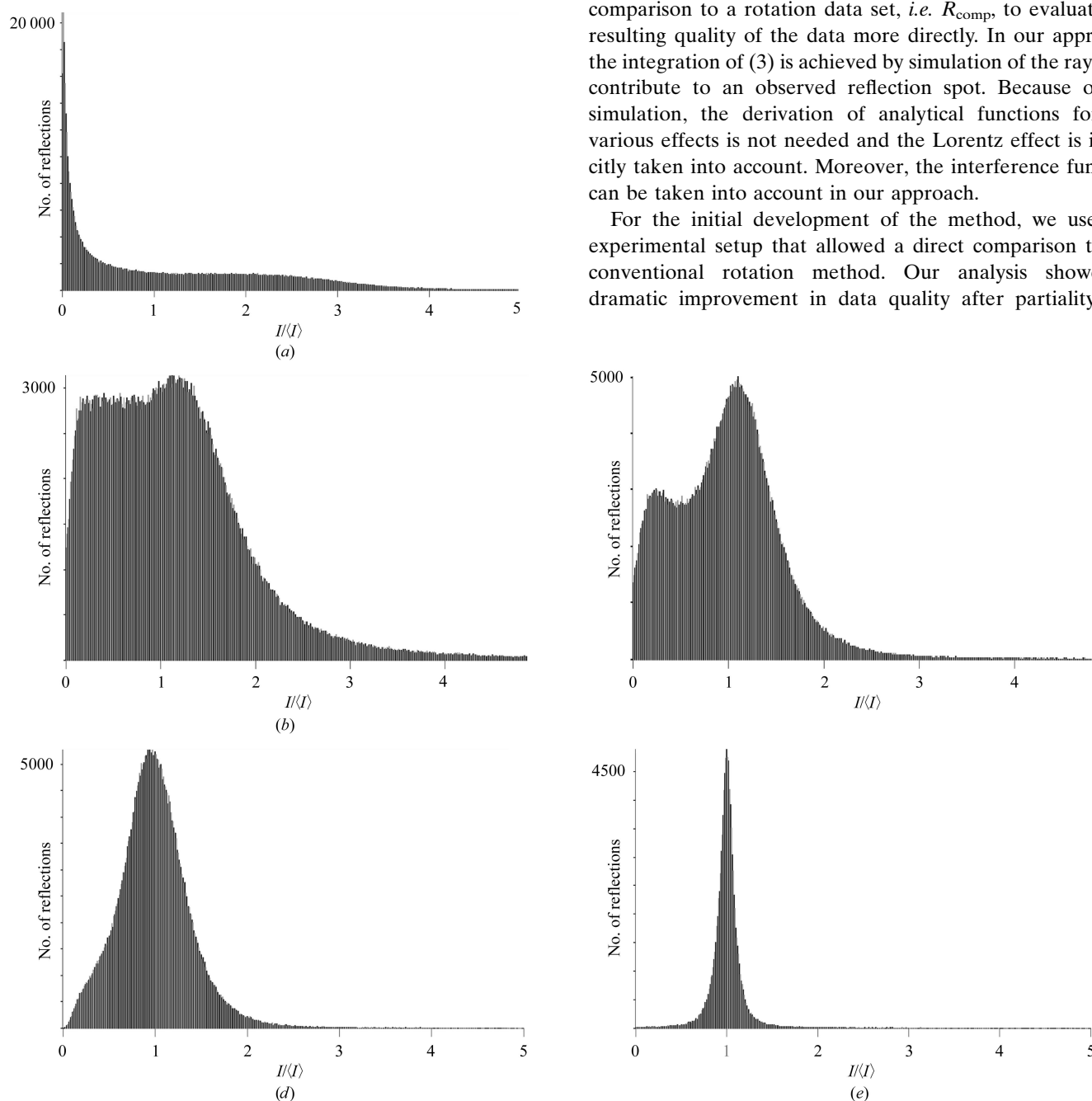
Figure 8
Histogram of $I(hkl)/\langle I(hkl)\rangle$. (*a*) Uncorrected, (*b*) partiality-corrected, (*c*) partiality and post-refined still data, (*d*) the same as (*c*) but omitting weak reflections and (*e*) rotation data

Lorentz correction. Both data processing and structure refinement showed that Lorentz correction is important and that omission of the Lorentz correction strongly affects the temperature factor. The anomalous sulfur densities increased 1.7-fold upon Lorentz and partiality correction of the still data. Overall, our approach markedly improved the R_{comp} factor between the intensities derived from rotation and still data from 26% to a final value of 4.7% after Lorentz and partiality correction and post-refinement.

Concurrent with the improvement in the final data quality upon Lorentz and partiality correction in *EVAL*, the internal merging R_{int} decreased from 105 to 64%. When we were developing the method, we hoped that post-refinement of the parameters would improve the final unique intensity data as well as further reduce the internal merging R_{int} factor. Post-refinement improved the precision of the modelled unit-cell dimensions and scale factor per image, although the error in modelled crystal orientations remained ∼0.1°. These more precise parameters indeed improved the resulting intensities (R_{comp} decreased from 5.3 to 4.7%). The internal statistics improved as well (R_{int} decreased from 64 to 56%); however, the final R_{int} factor remained high. This high R_{int} could be owing to features that were not included in our ray-tracing model, such as possible asymmetry in the focus, (anisotropic) mosaic spread or crystal form, or absorption by the crystal. Notably, crystal absorption may have a significant effect on the presented data because a relatively large crystal was used in this experiment. Crystal absorption is likely to be negligible when data are collected from microcrystals or nanocrystals, as is the case in serial crystallography. Obviously, further development of our approach is needed to account for the experimental conditions of serial (femtosecond) crystallography using XFEL or synchrotron sources. Automated schemes will be needed to model, for example, the large number of single-crystal diffraction images and fluctuations in beam spectra. In general, comprehensive modelling of the relevant experimental conditions should improve both the internal merging statistics and the resulting intensities. Not modelling significant effects that are present in the data can only be overcome by collecting more data to allow the averaging out of these effects by the Monte Carlo approach. In a real-case scenario the rotation data will not be available to evaluate the data quality, and an R_{int} of ∼50% may possibly be a practical metric to judge the resulting data quality.

Overall, we have shown that ray tracing can produce reliable partialities that improve the resulting data quality originating from still diffraction images. Moreover, our method is versatile and allows the modelling of a wide variety of effects, including those that yield non-Gaussian, asymmetric effects on the diffraction spot. In particular, the approach can take the interference function into account, which will be critical for processing data obtained from nanocrystals. Thus, in this paper we have presented the theoretical framework and demonstrated the potential of the ray-tracing methodology for processing still diffraction data.

The rotation and still diffraction images are available at http://rawdata.chem.uu.nl/c003.

APPENDIX A
Comparison with Laue interference function

The diffracted intensity reflected by a small crystal bathed in an incident monochromatic beam is proportional to the shape transform of the crystal. The reflected intensity received by the detector in a small cone of solid angle $d\Omega$, while the reciprocal-lattice vector $\mathbf{d}^*_{hkl}$ has a small deviation ε of the Bragg angle θ, is given by the Laue interference function (Laue, 1936; James, 1958) and is used in papers by Kirian *et al.* (2010) and White *et al.* (2012),

$$I_0|F|^2 p\left(\frac{e^2}{mc^2}\right)^2 \frac{\sin^2 \pi\xi N_1}{\sin^2 \pi\xi}\frac{\sin^2 \pi\eta N_2}{\sin^2 \pi\eta}\frac{\sin^2 \pi\zeta N_3}{\sin^2 \pi\zeta}\, d\Omega\, d\varepsilon. \quad (12)$$

N_1, N_2 and N_3 are the number of unit cells in the three dimensions of the parallepiped crystal. ξ is the scalar product $\Delta\mathbf{k} \cdot \mathbf{a}$ and in near-Bragg condition it is equal to $h + \Delta h$, and likewise for the other directions. As we are only interested in the diffracted intensity close to the Bragg condition, we introduce a local reciprocal axis system and replace ξ by the nonperiodic Δh. The terms in the denominator of (12) can then be written as $(\pi\Delta h)^2$ because they concern only small numbers,

$$I_0|F|^2 p\left(\frac{e^2}{mc^2}\right)^2 \frac{\sin^2 \pi\Delta h N_1}{(\pi\Delta h)^2}\frac{\sin^2 \pi\Delta k N_2}{(\pi\Delta k)^2}\frac{\sin^2 \pi\Delta l N_3}{(\pi\Delta l)^2}\, d\Omega\, d\varepsilon. \quad (13)$$

It is more convenient to choose the reciprocal axes system such that Δl is along the reciprocal-lattice vector $\mathbf{d}^*_{hkl}$ and Δh and Δk are parallel to the diffracting Bragg plane hkl (Authier, 2001). Such a transformation can be carried out because the normal to the Bragg plane is always a reciprocal-lattice vector. (Note that Δh, Δk and Δl are dimensionless.) The Jacobians of the transformation of the integration variables are $d\Omega = V^*_{cell}(d_{hkl}/\sin\theta)\lambda^2 d(\Delta h)d(\Delta k)$ and $d\varepsilon = \lambda/(2d_{hkl}\cos\theta)d(\Delta l)$ (Authier, 2001), leading to

$$I_0|F|^2 p\left(\frac{e^2}{mc^2}\right)^2 \frac{\lambda^3}{V_{cell}\sin 2\theta}\frac{\sin^2 \pi\Delta h N_1}{(\pi\Delta h)^2}$$
$$\times \frac{\sin^2 \pi\Delta k N_2}{(\pi\Delta k)^2}\frac{\sin^2 \pi\Delta l N_3}{(\pi\Delta l)^2}\, d\Delta h\, d\Delta k\, d\Delta l. \quad (14)$$

By integration over Δh and Δk the diffracted intensity for a given value of ε is obtained,

$$I_0|F|^2 p\left(\frac{e^2}{mc^2}\right)^2 \lambda^3 \frac{N_1 N_2}{V_{cell}\sin 2\theta}\frac{\sin^2 \pi\Delta l N_3}{(\pi\Delta l)^2}\, d\Delta l. \quad (15)$$

The equivalence of this expression to that of James (1958) and Buerger (1960), as we use in *EVAL*, is shown by the following. How does Δl depend on a small deviation $\Delta\theta = \varepsilon$? We write $\Delta l = d(l)/d\theta = d(c/d_{hkl})/d\theta = d(cd^*_{hkl})/d\theta = c2\cos\theta/\lambda$. cN_3 can be written as $ld_{hkl}N_3$, and further we use the property that the volume of the crystal $V_{crystal} = N_1 N_2 N_3 V_{cell}$ and $B = 2\pi d_{hkl}\cos\theta/\lambda$. Writing (14) in terms of $d\varepsilon$ gives

$$I_0|F|^2 p\left(\frac{e^2}{mc^2}\right)^2 \lambda^2 \frac{N_1 N_2}{V_{\text{cell}}\sin 2\theta} \frac{2ld_{hkl}\cos\theta}{\lambda}$$

$$\times \frac{\sin^2 2\pi(N_3 ld_{hkl}\cos\theta/\lambda\varepsilon)}{(2\pi\, ld_{hkl}\cos\theta/\lambda\ \varepsilon)^2}\, d\varepsilon. \qquad (16)$$

Using $N_1 N_2/V_{\text{cell}} = V_{\text{crystal}}/V_{\text{cell}}^2(1/N_3)$ (James, 1958, p. 43) and $d_{hkl} = \lambda/2\sin\theta$, we can write

$$I_0|F|^2 p\left(\frac{e^2}{mc^2}\right)^2 \lambda^3 \frac{V_{\text{crystal}}}{V_{\text{cell}}^2} \frac{1}{2\sin^2\theta} \frac{1}{N_3 l} \frac{\sin^2 \pi N_3 lB\varepsilon}{(\pi B\varepsilon)^2}\, d\varepsilon. \qquad (17)$$

(17) is exactly the equation used in *EVAL* (2), as the number of layers $s = N_3 l$, and using $n = N_1 N_2 N_3/V_{\text{crystal}}$ we can write $V_{\text{crystal}}/V_{\text{cell}}^2 = n^2 V_{\text{crystal}}$.

Acknowledgements

References

Authier, A. (2001). *Dynamical Theory of X-ray Diffraction*. Oxford University Press.

Boutet, S. *et al.* (2012). *Science*, **337**, 362–364.

Buerger, M. J. (1960). *Crystal Structure Analysis*. New York: John Wiley & Sons.

Burmeister, W. P. (2000). *Acta Cryst.* D**56**, 328–341.

Chapman, H. N. *et al.* (2011). *Nature (London)*, **470**, 73–77.

Dejoie, C., Smeets, S., Baerlocher, C., Tamura, N., Pattison, P., Abela, R. & McCusker, L. B. (2015). *IUCrJ*, **2**, 361–370.

Demirci, H. *et al.* (2013). *Acta Cryst.* F**69**, 1066–1009.

Duisenberg, A. J. M. (1992). *J. Appl. Cryst.* **25**, 92–96.

Duisenberg, A. J. M., Kroon-Batenburg, L. M. J. & Schreurs, A. M. M. (2003). *J. Appl. Cryst.* **36**, 220–229.

Gati, C., Bourenkov, G., Klinge, M., Rehders, D., Stellato, F., Oberthür, D., Yefanov, O., Sommer, B. P., Mogk, S., Duszenko, M., Betzel, C., Schneider, T. R., Chapman, H. N. & Redecke, L. (2014). *IUCrJ*, **1**, 87–94.

Hattne, J. *et al.* (2014). *Nature Methods*, **11**, 545–548.

Howell, P. L. & Smith, G. D. (1992). *J. Appl. Cryst.* **25**, 81–86.

James, R. W. (1958). *The Optical Principles of the Diffraction of X-rays*. London: G. Bell & Sons.

Kabsch, W. (2014). *Acta Cryst.* D**70**, 2204–2216.

Kirian, R. A., Wang, X., Weierstall, U., Schmidt, K. E., Spence, J. C. H., Hunter, M., Fromme, P., White, T., Chapman, H. N. & Holton, J. (2010). *Opt. Express*, **18**, 5713–5723.

Laue, M. von (1936). *Ann. Phys.* **41**, 971–988.

Leslie, A. G. W. & Powell, H. R. (2007). *Evolving Methods for Macromolecular Crystallography*, edited by R. J. Read & J. L. Sussman, pp. 41–51. Dordrecht: Springer.

Milch, J. R. & Minor, T. C. (1974). *J. Appl. Cryst.* **7**, 502–505.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Qu, K., Zhou, L. & Dong, Y.-H. (2014). *Acta Cryst.* D**70**, 1202–1211.

Ravelli, R. B. G. & McSweeney, S. M. (2000). *Structure*, **8**, 315–328.

Redecke, L. *et al.* (2013). *Science*, **339**, 227–230.

Rossmann, M. G. & van Beek, C. G. (1999). *Acta Cryst.* D**55**, 1631–1640.

Sauter, N. K. (2015). *J. Synchrotron Rad.* **22**, 239–248.

Sauter, N. K., Grosse-Kunstleve, R. W. & Adams, P. D. (2004). *J. Appl. Cryst.* **37**, 399–409.

Sauter, N. K., Hattne, J., Brewster, A. S., Echols, N., Zwart, P. H. & Adams, P. D. (2014). *Acta Cryst.* D**70**, 3299–3309.

Sauter, N. K., Hattne, J., Grosse-Kunstleve, R. W. & Echols, N. (2013). *Acta Cryst.* D**69**, 1274–1282.

Schreurs, A. M. M. (1998). *VIEW*. Utrecht University, The Netherlands.

Schreurs, A. M. M. (1999). *PEAKREF*. Utrecht University, The Netherlands.

Schreurs, A. M. M. (2007). *ANY*. Utrecht University, The Netherlands.

Schreurs, A. M. M., Xian, X. & Kroon-Batenburg, L. M. J. (2010). *J. Appl. Cryst.* **43**, 70–82.

Schutt, N. K. & Winkler, F. K. (1977). *The Rotation Method in Crystallography*, edited by U. W. Arndt & A. J. Wonacott, pp. 173–186. Amsterdam: North Holland.

Spence, J. C. H., Kirian, R. A., Wang, X., Weierstall, U., Schmidt, K. E., White, T., Barty, A., Chapman, H. N., Marchesini, S. & Holton, J. (2011). *Opt. Express*, **19**, 2866–2873.

Sutton, K. A., Black, P. J., Mercer, K. R., Garman, E. F., Owen, R. L., Snell, E. H. & Bernhard, W. A. (2013). *Acta Cryst.* D**69**, 2381–2394.

Thorn, A. & Sheldrick, G. M. (2011). *J. Appl. Cryst.* **44**, 1285–1287.

Uervirojnangkoorn, M., Zeldin, O. B., Lyubimov, A. Y., Hattne, J., Brewster, A. S., Sauter, N. K., Brunger, A. T. & Weis, W. I. (2015). *eLife*, **4**, e05421.

Vaney, M. C., Maignan, S., Riès-Kautt, M. & Ducruix, A. (1996). *Acta Cryst.* D**52**, 505–517.

Weik, M., Ravelli, R. B. G., Kryger, G., McSweeney, S., Raves, M. L., Harel, M., Gros, P., Silman, I., Kroon, J. & Sussman, J. L. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 623–628.

White, T. A. (2014). *Philos. Trans. R. Soc. B Biol. Sci.* **369**, 20130330.

White, T. A., Barty, A., Stellato, F., Holton, J. M., Kirian, R. A., Zatsepin, N. A. & Chapman, H. N. (2013). *Acta Cryst.* D**69**, 1231–1240.

White, T. A., Kirian, R. A., Martin, A. V., Aquila, A., Nass, K., Barty, A. & Chapman, H. N. (2012). *J. Appl. Cryst.* **45**, 335–341.

Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.

Zachariasen, W. H. (1945). *Theory of X-ray Diffraction in Crystals*. New York: Dover.