The minimum crystal size needed for a complete diffraction data set

James M. Holton[a,b]* and Kenneth A. Frankel[b]

[a]Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94158-2330, USA, and [b]Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

Correspondence e-mail: jmholton@lbl.gov

In this work, classic intensity formulae were united with an empirical spot-fading model in order to calculate the diameter of a spherical crystal that will scatter the required number of photons per spot at a desired resolution over the radiation-damage-limited lifetime. The influences of molecular weight, solvent content, Wilson B factor, X-ray wavelength and attenuation on scattering power and dose were all included. Taking the net photon count in a spot as the only source of noise, a complete data set with a signal-to-noise ratio of 2 at 2 Å resolution was predicted to be attainable from a perfect lysozyme crystal sphere 1.2 μm in diameter and two different models of photoelectron escape reduced this to 0.5 or 0.34 μm. These represent 15-fold to 700-fold less scattering power than the smallest experimentally determined crystal size to date, but the gap was shown to be consistent with the background scattering level of the relevant experiment. These results suggest that reduction of background photons and diffraction spot size on the detector are the principal paths to improving crystallographic data quality beyond current limits.

1. Introduction

The last 15 years have seen many experimental estimates of how small a protein crystal can be and still yield a complete data set (Gonzalez & Nave, 1994; Glaeser *et al.*, 2000; Teng & Moffat, 2000, 2002; Facciotti *et al.*, 2003; Sliz *et al.*, 2003; Li *et al.*, 2004; Nelson *et al.*, 2005; Sawaya *et al.*, 2007; Coulibaly *et al.*, 2007; Standfuss *et al.*, 2007; Moukhametzianov *et al.*, 2008; reviewed by Holton, 2009) and this size has been decreasing as technology improves. But is there a theoretical limit? The work presented here establishes a firm theoretical framework for computing the absolute signal available from very small macromolecular crystals and every effort is made to explicitly and unambiguously spell out the definitions and derivations. The *International Tables for Crystallography* (Wilson & Prince, 1999) contain most of the critical pieces of the puzzle assembled here and the original references are spread out over nearly a century of literature.

Here, we endeavor to keep the theory general and independent of the limitations of current diffraction hardware. For example, the time-honored practice of recording the three-dimensional diffraction pattern on as few images as possible was not simply an effort to save money on film, but to minimize noise intrinsic to the detection process such as 'fog' on the film or the read-out circuit of a charge-coupled device (CCD). Counting detectors such as multi-wire (Cork *et al.*, 1974) and pixel arrays (Kraft *et al.*, 2009) do not have this kind of noise and the optimal data-collection strategy with these detectors is different (Xuong *et al.*, 1985; Schulze-Briese *et al.*,

2007). For simplicity, in the present work we consider the X-ray detector and indeed the entire diffractometer to be an ideal device subject only to the shot noise of the net spot photons themselves (the square root of the number of counts). All other sources of noise, including background scattering, are neglected until the discussion in §3.2.

The formula for the integrated intensity of a spot was introduced by Darwin (1914), but much subsequent work was required to fill out the original theory. For example, Darwin's variable '*f*' required the development of quantum theory to explain its observed value (Debye, 1915, 1988). The resulting orbital shapes (Slater, 1929) led directly to the cross-sections needed to compute absorption effects in the 1960s and steady improvements continue to this day (Hubbell, 2006). Only recently has it become clearly established that radiation damage at cryogenic temperatures is proportional to dose (Henderson, 1990; Gonzalez & Nave, 1994; Glaeser *et al.*, 2000; Sliz *et al.*, 2003; Leiros *et al.*, 2006; Owen *et al.*, 2006; Garman & McSweeney, 2007; Garman & Nave, 2009; Holton, 2009) and this understanding enabled the present work.

The intensity of a Bragg spot is not simply the square of the structure factor, but depends on several other factors including exposure time, crystal volume and the geometry of diffraction. Consequently, the absolute number of photons in a spot (which determines the maximum possible signal-to-noise ratio) depends on exactly where the spot falls on the detector surface. Algorithms for computing these intensity 'correction' factors are encoded into most data-processing programs, but the source codes are not always available and in many cases the implemented corrections only apply to particular camera geometries. Therefore, the reproducibility and generality of the results presented here requires a clear description of each correction factor and we begin by defining the relevant coordinate system.

2. Methods

2.1. Coordinate system

There are many possible ways to assign *xyz* coordinates to a diffractometer; unfortunately, most of them have been employed at one time or another and few data-processing programs share exactly the same convention. Here, we will adopt a 'classic' coordinate system essentially identical to that described in chapter 7 of Arndt & Wonacott (1977), which is also the coordinate system used by the data-processing program *MOSFLM* (Leslie, 2006). In this system, *x* is the direction of the X-ray beam, *z* is the (horizontal) spindle axis and *y* is 'up' (opposing gravity) or perpendicular to the page in Fig. 1.

2.2. Spot intensity

Typically, crystallographic data-processing and model-refinement programs assign an arbitrary 'scale factor' for the observed spot intensities to put them on the same scale as the structure factors calculated from the model, but the exact relationship between the intensity of a fully recorded spot and

the square of the structure factor is given by Darwin's formula (Darwin, 1914, 1922; Blundell & Johnson, 1976) and instructive re-derivations can be found in textbooks by James (1962) and Woolfson (1997),

$$I = I_{\text{beam}} r_e^2 \frac{V_{\text{xtal}}}{V_{\text{cell}}} \cdot \frac{\lambda^3 L}{\omega V_{\text{cell}}} P \cdot A \cdot |F|^2, \qquad (1)$$

where I is the integrated spot intensity (photons/spot), I_{beam} is the intensity of the incident beam (photons s^{-1} m^{-2}), r_e is the classical electron radius (2.818×10^{-15} m), V_{xtal} is the illuminated volume of the crystal (in m^3), V_{cell} is the volume of the crystal unit cell (in m^3), λ is the X-ray wavelength (in m), ω is the angular velocity of the crystal (radians s^{-1}; §2.8), L is the Lorentz factor (speed/speed; §2.3), P is the polarization factor (photons/photons; §2.4), A is the X-ray transmittance of the path through the crystal to the spot (photons/photons; §2.5) and F is the structure factor of the unit cell at the relp of interest (electron equivalents; §2.7).

The abbreviation 'relp' (reciprocal-lattice point) is used to denote a particular point in reciprocal space, distinct from its symmetry mates (Ramachandran & Wooster, 1951; Helliwell, 1999), and here we use 'spot' to refer to a single observation of a relp and '*hkl*' to indicate the sum of all symmetry-equivalent spots (merging anomalous pairs). Note that all quantities entered into (1) are in metre–kilogram–second (MKS) units, including the X-ray wavelength (λ), and that the units of 'intensity' for spots (photons/spot) are not the same as those for either the incident beam (photons s^{-1} m^{-2}) or classical electron scattering (photons sr^{-1}). Despite this, all of these quantities remain commonly referred to as 'intensity', leading
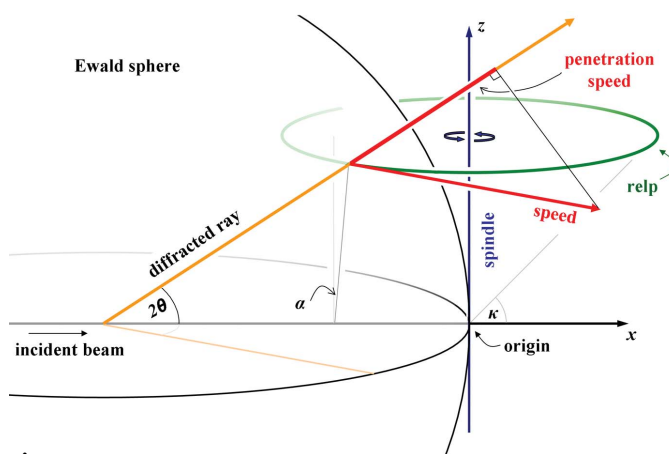

Figure 1
Coordinate system. The *x* axis is occupied by the X-ray beam and the spindle rotates the crystal (at the origin) about the *z* axis. The *y* axis is not shown as it is very nearly perpendicular to the page. The reciprocal-lattice point (relp) of interest is described here by the circle it traces out as the crystal is rotated. Note that it intersects the Ewald sphere twice and that the 'penetration speed' is the component of the relp's velocity that is perpendicular to the Ewald sphere surface. The ratio of the 'actual speed' to the 'penetration speed' is the Lorentz factor. The diffracted ray passes through the point of intersection, but evolves from the center of the Ewald sphere (not the origin!), which is an unfortunate conceptual flaw in Ewald's construction. Nevertheless, the take-off angle (2θ) obtained is the same as that observed in real space. The angles α and κ used in (3) and Appendix C are shown.

to a considerable amount of confusion if the units are not given explicitly. The change of units arises because the full spot intensity (photons/spot) is obtained by integrating over the relp as it moves through the Ewald sphere (Ewald, 1913; Arndt & Wonacott, 1977; Helliwell, 1999) and therefore several geometric factors must be taken into account.

Experimental confirmation of Darwin's formula has been presented by Moseley & Darwin (1913), Bragg *et al.* (1921*a,b*, 1922), Compton & Freeman (1922) and many others since. For an example calculation using (1), consider a 100 µm diameter spherical protein crystal with all three unit-cell edges 50 Å long. Assume that for a particular relp at 2 Å resolution we have F = 170 electron equivalents (see §2.7) and further assume some crystal orientation that assigns L = 2.2, P = 0.92 and A = 96% to this relp (see §§2.3, 2.4 and 2.5, respectively). If the crystal rotates at $1°\,s^{-1}$ in a uniform beam of 1 Å X-rays with 10^{12} photons s^{-1} passing into the 100 µm diameter circular cross-section of the crystal, then (1) predicts an integrated full spot intensity of 109 011 photons. This calculation was found to be in remarkable agreement with experimentally observed spot intensities from a lysozyme crystal (not shown) on the protein crystallography beamline 8.3.1 at the Advanced Light Source (instrument described by MacDowell *et al.*, 2004). Once I_{beam} had been calibrated (Owen *et al.*, 2009), the discrepancy between calculation and experiment was essentially the uncertainty in our visual estimate of V_{xtal} (about 15%).

The flux density I_{beam} is a constant in (1), which implies that the crystal is 'bathed' in a 'flat-top' or 'top-hat' beam. Real X-ray beams are seldom this perfect, but any crystal in any beam may be formally broken up into tiny cubes small enough for I_{beam} to be considered constant over each cube and the total spot intensity obtained by summing the results of (1) for all the cubes. However, if I_{beam} is the same for every cube there is clearly no need to break up the crystal; conversely, if the crystal has constant thickness along the beam direction then the average flux density experienced by the crystal (regardless of beam shape) may be used as I_{beam} in (1). Only if both the crystal shape and the beam profile have irregular shapes does (1) need to be integrated over the beam profile and crystal volume. However, we show in §2.11 and Appendix C (deposited as supplementary material[1]) that the damage-limited spot intensity is independent of I_{beam}, obviating the need to consider beam and crystal shapes, so for simplicity in this work we will consider a spherical crystal 'bathed' in a top-hat beam.

Note that (1) does not depend on the mosaic structure of the crystal and indeed a crystal consisting of a single mosaic domain or thousands of mosaic domains will still yield exactly the same integrated spot intensity (I) as long as the mosaic domains are small when compared with the attenuation depth (μ^{-1}) of the X-rays in the crystal. This depth is typically several millimetres for 1 Å X-rays (see the end of §2.5) and

protein crystals this large are very rare, let alone single-domain crystals (Snell *et al.*, 2003). A common misconception that protein microcrystals consisting of a single mosaic domain will produce more intense spots than expected from Darwin's formula seems to have arisen from the above-mentioned confusion over the several possible meanings of the word 'intensity' (discussed further in §2.7). In truth, however, (1) was derived for small and single-domain crystals and also applies to the 'ideally imperfect' case of a large crystal with many mosaic domains (Darwin, 1922). Large single-domain crystals that approach the length scale of the attenuation depth of the X-rays actually produce weaker spots than predicted by (1) owing to extinction effects (James, 1962; Woolfson, 1997; Sabine, 1999; Authier, 2004).

2.3. Lorentz factor

The Lorentz factor L in (1) is always greater than one and is the ratio of the speed of a rotating relp to the 'penetration speed' at which it transits the Ewald sphere (Fig. 1). This Lorentz factor in crystallography[2] is not to be confused with its inverse, the Lorentz correction L^{-1} which data-processing programs such as *MOSFLM* (Leslie, 2006) use to 'correct' for this effect by multiplying observed integrated intensities by L^{-1}. The description of the Lorentz factor in *International Tables for Crystallography* (Lipson & Langford, 1999) notes that some confusion has arisen over the definition of the Lorentz factor because Lorentz never published it. Instead, it seems he wrote a letter to Debye, who included it as a second note added in proof (Debye, 1914, 1988).

Essentially, the Lorentz factor accounts for how the integrated intensity (photons/spot) of a relp will be higher if it moves slowly through the Bragg condition than if it moves quickly. Indeed, the angular velocity of the crystal (ω) divided by the Lorentz factor (L) is the angular velocity of the relp as 'seen' from the origin (see Fig. 1). This geometric correction is therefore grouped with other geometric factors in (1) such as ω. The cube of the wavelength (λ^3) and one of the unit-cell volume (V_{cell}) terms are also geometric corrections since these are involved in the size of the integration volume in reciprocal space (chapter 6 of Woolfson, 1997).

It is instructive to consider the relationship between the Lorentz factor and the spot position on the detector. This will obviously depend on the camera geometry, but in the common case in which the crystal rotation axis is perpendicular to the X-ray beam the Lorentz factor (L) is given by

$$L = \frac{1}{(\sin^2 2\theta - \zeta^2)^{1/2}} \qquad (2a)$$

$$\zeta_\perp = \cos 2\theta Z_{det}/X_{stf}, \qquad (2b)$$

where θ is the Bragg angle, $\zeta\,(\lambda\mathbf{d}*\cdot\hat{\mathbf{z}})$ is a normalized projection of the relp vector onto the rotation axis (z), $\zeta_\perp$ is ζ in terms of spot coordinates on a flat detector normal to the incident beam, Z_{det} is the coordinate of the diffraction spot on the

---

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: BA5148). Services for accessing this material are described at the back of the journal.

[2] Note that there is also a 'Lorentz factor' in the Theory of Relativity, which has nothing to do with the Lorentz factor in crystallography other than sharing the same namesake.

detector along the axis parallel to the rotation axis (relative to the beam center in mm) and X_{stf} is the sample-to-detector distance along the direct-beam path (in mm).

The Bragg angle θ is defined as half of the angle between the direct-beam path and the diffracted ray (see Fig. 1). Any given relp can be represented as a vector $\mathbf{d}*$ that will always have length $d* = 1/d$, where d is the d-spacing (in Å) of the spot. No matter how the crystal is rotated, the d-spacing of a spot does not change. The polar coordinate ζ (Helliwell, 1999) is calculated by taking the z component of $\mathbf{d}*$ ($\hat{\mathbf{z}}$ is the unit vector along the z axis) and multiplying it by the X-ray wavelength λ (in Å). This is because the z component of $\mathbf{d}*$ has dimensions of Å^{-1} and ζ must be dimensionless to be meaningfully related to $\sin\theta$.

In the also common case in which the detector is a flat plane and normal to the incident X-ray beam ζ may be conveniently replaced with $\zeta_\perp$ from (2b). However, moving the detector does not change the L of a given relp and $\zeta_\perp$ serves simply as a convenient way to map the Lorentz factor onto the detector face. For arbitrary detector positions ζ must be computed from the spindle geometry and in the general case of the beam not being perfectly normal to the rotation axis L must be calculated by taking the projection of the relp velocity vector along the diffracted ray (as shown in Fig. 1).

Arbitrary rotations of the crystal will rotate the vector $\mathbf{d}*$ by exactly the same angles and if the crystal is oriented such that $\mathbf{d}*$ approaches the spindle axis (z axis) it will eventually cross into a 'blind region' (Arndt & Wonacott, 1977; Helliwell, 1999) where spindle rotation alone cannot bring the relp onto the Ewald sphere. As the relp approaches this blind region the denominator of (2a) becomes smaller and smaller and the Lorentz factor approaches infinity. Crossing into the blind region, the quantity under the square root in (2a) becomes zero or less and the Lorentz factor becomes undefined.

It is important to note, however, that an infinite Lorentz factor does not actually imply an infinite spot intensity. This is because the relps are not infinitely sharp points, but rather occupy a volume in reciprocal space that must pass completely through the Ewald sphere for (1) to be valid. In fact, the size and shape of this reciprocal-space volume is simply the Fourier transform of the size and shape of the mosaic domain producing it, but a detailed discussion of spot shapes is beyond the scope of this work. It will suffice here to say that the blind region is effectively enlarged by an angle comparable to the crystal mosaic spread, 'swallowing' the infinite Lorentz factors. The few spots that are close to the rotation axis will indeed have very large Lorentz factors, but also a very wide angular range of reflection (rocking width), so on a typical diffraction image these high-L spots are roughly the same intensity (photons/spot) as any other. A discussion of rotation range will continue in §2.8.

2.4. Polarization factor

The polarization factor P is always less than one and accounts for losses of scattering efficiency when the incident-beam and scattered-beam E-vectors do not line up. That is,

the E-vector of any electromagnetic wave must always be perpendicular to the direction of travel (Maxwell, 1865; Purcell, 1985), but the direction of travel changes upon scattering. P is simply the dot product of the E-vectors of the incident and scattered waves (averaged over all incident E-vectors) and here we use the convenient expression given by Drenth (1999) (Azároff, 1955; Kahn et al., 1982),

$$2P = 1 + \cos^2 2\theta - \mathscr{I}\cos 2\alpha \sin^2 2\theta, \qquad (3)$$

where P is the polarization factor used in (1) (photons/photons), θ is the Bragg angle, α is the angle between the projections of the z axis and the diffracted ray onto a plane normal to the incident beam and $\mathscr{I}$ is the degree of polarization.

Note that the polarization factor P varies from spot to spot whereas $\mathscr{I}$ is the 'polarization' entered into most diffraction data-processing programs. $\mathscr{I}$ ranges from 1 to 0 to -1 as the incident E-vector varies from 'horizontal' (along the z axis) to unpolarized to 'vertical', respectively. The 'plane normal to the incident beam' invoked to define α here is any plane parallel to both the y and z axes (see α in Fig. 1 as well as Arndt & Wonacott, 1977).

Many synchrotron-based diffractometers are designed with horizontal spindle axes (as defined here) because in this geometry the strong horizontal polarization of synchrotron radiation ($\mathscr{I}$ close to 1) tends to cancel the Lorentz factor and the 'hole' in scattering owing to polarization at $2\theta = 90°$ and $\alpha = 0°$ coincides with the blind region (§2.3). However, the average value of the product LP is independent of $\mathscr{I}$ (see §2.6) and therefore spindle orientation has no effect on average intensity (photons/spot) in a given resolution bin. The only practical concern is that many data-processing programs throw out spots with large L because such spots are very sensitive to small errors in crystal orientation, but even when $L > 5$ spots are rejected the 'penalty' of a vertical spindle ($\mathscr{I} = -1$) in the 2 Å bin using 1 Å radiation is only a 10% drop in photons/hkl (not shown). Indeed, for such data P ranges from 1 to 0.77 and this variation diminishes further as the pattern is compressed into lower angles at shorter wavelengths because (3) depends purely on the geometry of the camera and not on the X-ray wavelength used. The mechanical stability advantages of a vertical spindle for small crystals therefore come at only a marginal cost to photons/spot.

2.5. Sample attenuation

The attenuation factor A in (1) is an average optical transmittance and is always less than one. For full accuracy, photons from each point in the X-ray source must be ray-traced to every accessible part of the crystal volume and from there out into the spot. The transmittance along each path depends on the size, shape and atomic composition of the crystal and any other substances it traverses (including air). The profile of the beam acts as a 'weighting function' and A is the average transmittance over all possible paths. Given the potential complexity of the shapes involved, the only general expression for A is the triple integral

$$A = \frac{1}{V_{\text{xtal}} I_{\text{beam}}} \int \int \int_{\text{xtal}} I_{\text{prof}}(y, z) \exp[-\mu_{\text{air}} t_{\text{air}}(x, y, z) \quad (4)$$

$$- \mu_{\text{xtal}} t_{\text{xtal}}(x, y, z) - \mu_{\text{loop}} t_{\text{loop}}(x, y, z) - \ldots] \, dx \, dy \, dz,$$

where A is the attenuation factor (photons/photons), V_{xtal} is the volume of the crystal (m^3), I_{beam} is the total intensity of the incident beam (photons s^{-1} m^{-2}), I_{prof} is the intensity of the beam profile at the coordinate 0, y, z (photons s^{-1} m^{-2}), μ_x is the attenuation coefficient of substance x, μ_x^{-1} is the attenuation length (m) and t_x is the component of the total path taken by X-rays through substance x via crystal coordinate x, y, z (m).

The complexity arises because the scattering and attenuation processes must be co-integrated over the illuminated volume of the crystal (V_{xtal}). The path taken by the incident beam is only important up to the point location of the 'scattering event' and from there the materials between the scattering event and the location of the diffraction spot must be considered. This integral can be solved analytically for the simple case of a flat slab-shaped crystal with uniform μ and the formula for this solution is presented in *International Tables for Crystallography* (Maslen, 1999). However, for anything other than a flat slab there is no analytic solution for (4) and even a perfect sphere must be evaluated numerically. Nevertheless, a sphere is a convenient 'average shape' for a protein crystal and look-up tables are available for this integral (Dwiggins, 1975; Flack & Vincent, 1978; Maslen, 1999). For the calculation at hand, we consider a spherical crystal of radius R with uniform attenuation coefficient μ_{xtal} in a uniform 'flat-top' beam and denote the total transmission of a beam diffracting at angle 2θ simply as

$$A = T_{\text{sphere}}(2\theta, \mu_{\text{xtal}}, R), \quad (5)$$

where A is the attenuation factor (photons/photons), T_{sphere} is the numerical solution to (4) for a sphere in a vacuum, 2θ is the angle between the incident and diffracted beams, μ_{xtal} is the attenuation coefficient of the crystal (m^{-1}) and R is the radius of the spherical crystal (m).

The value of μ for each substance is obtained using its density (ρ) and the tabulated X-ray cross-sections (Storm & Israel, 1970; Berger & Hubbell, 1987; Creagh & Helliwell, 1999) of the chemical elements comprising it (reviewed by Hubbell, 2006). A convenient program for the accurate calculation of μ for a particular protein crystal is *RADDOSE* (Murray *et al.*, 2004; Paithankar *et al.*, 2009); for the calculations presented here we use an average empirical formula for protein, $H_{49.8}C_{31.8}N_{8.56}O_{9.54}S_{0.249}$, determined from a survey (not shown) of the Protein Data Bank (Berman *et al.*, 2002). Taking 1 Å X-rays, for example, the values for μ in protein, water and the 50% solvent protein crystal used in this work are 2.78, 2.85 and 2.81 cm^{-1}, respectively. This yields an attenuation depth μ_{xtal}^{-1} of 3.6 mm, so a 2.5 mm thick protein crystal is required to reduce a spot intensity (photons/spot) by half and a 100 µm crystal reduces no spot intensity by more than ∼2.7%. Therefore, A is a small correction in typical cases and only becomes significant if strongly absorbing atoms are soaked into the crystal (see Holton, 2009) or if long-wavelength X-rays are used. For example, at the S K edge (5 Å wavelength) $\mu_{\text{xtal}}^{-1} \simeq 32$ µm and attenuation can reduce the spot intensities from a 100 µm crystal by as much as 96% ($A = 0.04$).

2.6. Average Lorentz–polarization factor and completeness

Since we are concerned here with the average value of a spot intensity (photons/spot) at a given resolution, we must know the average value of the product of the Lorentz and polarization factors (LP). It is also important to account for relps that fall into the 'blind region' (§2.3) as these will not contribute to the merged signal of an hkl index at one wavelength but may contribute at another. The fraction of all relps in a given resolution bin that can be observed by rotating about a single axis (f_{obs}) is simply $\cos\theta$ (see Appendix A) and if we average the product of (2a) and (3) for these accessible relps (Appendix B) we obtain the exact expressions

$$\langle LP \rangle = \frac{\pi}{2} \left(\frac{1}{\sin 2\theta} - \frac{\sin 2\theta}{2} \right) \quad (6a)$$

$$\langle LP \rangle f_{\text{obs}} = \frac{\pi(3 + \cos 4\theta)}{16 \sin \theta}, \quad (6b)$$

where f_{obs} is the fraction of relps at this resolution that will cross the Ewald sphere using a single axis ($\cos\theta$) and θ is the Bragg angle. Note the use of angle brackets $\langle \rangle$ to denote average values and that $\langle LP \rangle$ and f_{obs} depend only on the Bragg angle (θ) and thus are independent of wavelength (λ) and the degree of polarization $\mathscr{I}$ from (3). However, as Bragg's law relates λ to θ, $\langle LP \rangle f_{\text{obs}}$ tends to cancel one of the λ terms in (1), but not exactly.

2.7. Average structure factor

The 'structure factor' has been defined (Debye & Scherrer, 1918; Hartree, 1925; Coppens, 1999) as the ratio of the amplitude of an electromagnetic wave scattered by an object of interest to that scattered by a single classical electron (Thomson, 1906; chapter 2 of Woolfson, 1997; Maslen *et al.*, 1999a), hence Thomson's classical electron cross-section (r_e^2) is included in (1). The F in (1) is the structure factor of one unit cell, which must be isolated in space for the intensity (photons sr^{-1}) to be computed directly from F. The other terms in (1) represent the ratio of the intensity scattered from a single unit cell to that of the entire crystal.

The apparent amplification from one V_{cell} term in (1) is effectively cancelled by the average square structure factor $\langle F^2 \rangle$, which is proportional to V_{cell} when the number of atoms per unit volume is fixed. This cancellation arises because the average scattering from a macromolecule at d-spacings better than ∼4 Å is essentially the same as that of a random distribution of atoms (Wilson, 1942, 1949; Shmueli & Wilson, 1999) and the total structure factor of a random arrangement of atoms rapidly approaches the structure factor of one atom (f_a) multiplied by the square root of the number of atoms. That is, when the scattered waves from a group of atoms are in no way 'correlated' with each other, the total scattered intensity (photons s^{-1} sr^{-1}) is the sum of the intensities that would be

seen from individual atoms and the square root of this total intensity is (by definition) proportional to the structure factor of the group. Conversely, if the atomic positions are perfectly correlated (such as in a regular lattice) then the amplitudes add in a nonrandom way and the intensity scattered in some directions (diffraction spots) becomes proportional to the square of the number of atoms. It is important to remember that this intensity has units of photons $s^{-1} sr^{-1}$, where steradians (sr) are the units of solid angle. For example, 10^6 photons s^{-1} emitted in completely random directions are described by an 'intensity' of $10^6/4\pi = 79\,577$ photons $s^{-1} sr^{-1}$ and a square detector pixel 100 μm in size and 100 mm from the sample (10^{-6} sr) will intercept about one photon every 12.6 s. Although the intensity (photons $s^{-1} sr^{-1}$) scattered by a crystal of N atoms can be very large, this is only true over a very small solid angle and as the size of the crystal (or mosaic domain) increases this solid angle becomes proportionally smaller. In general, this patch of high intensity is much smaller than a pixel, but the observed intensity (in photons) is given by the integral of photons $s^{-1} sr^{-1}$ over the entire pixel and rocking width of the relp (chapters 2 and 6 of Woolfson, 1997). The change in units whilst using the same word 'intensity' has historically led to some confusion, no doubt arising in part from Darwin's formula appearing more than half a century before the first use of the word 'pixel' in the scientific literature.

It is instructive here to examine how the terms in (1) interrelate as the properties of the crystal change. For example, as atoms are added to random locations in the unit cell (keeping V_{cell} fixed for the moment) the structure factor of the unit cell (F) increases as the square root of the number of atoms in the unit cell (N_{cell}) and hence the intensity of a fully recorded spot (I, in photons) is proportional to N_{cell}. Conversely, if V_{cell} is increased while keeping V_{xtal} and the total number of atoms in the crystal constant, then the number of unit cells (V_{xtal}/V_{cell}) decreases while N_{cell} increases. This causes F to increase as the square root of V_{cell}, so F^2 cancels one V_{cell} term and the net effect of reorganizing a fixed number of atoms into larger cells is that individual spot intensities decrease proportionally to V_{cell}. Since the number of relps in a given volume of reciprocal space is also proportional to V_{cell}, the total summed intensity of all spots does not change and remains proportional to the number of atoms in the X-ray beam regardless of how these atoms are divided into unit cells. Another way to reach the same conclusion is by the simple fact of conservation of scattered photons: a given number of atoms will scatter a fixed number of photons and this number is dictated by the elastic scattering cross-section of these atoms. The arrangement of the atoms affects the direction in which these photons are scattered but cannot change their number and in the limiting case of very small unit cells that have no relps intersecting the Ewald sphere all of these photons are scattered in the forward direction (the relp with index $hkl = 000$).

The number of scattering atoms per unit volume in protein crystals varies with solvent content because the atoms of disordered solvent contribute only very weakly to high-angle Bragg peaks (Tronrud, 1997; Afonine et al., 2005). Therefore,

the number of atoms contributing to spots at a given resolution beyond ∼4 Å can be taken as the number of ordered (protein) atoms in the unit cell,

$$N_{cell} = n_{symop} n_{ASU} \frac{M_r}{\langle M_a \rangle} = \frac{V_{cell}}{V_M \langle M_a \rangle}, \qquad (7)$$

where N_{cell} is the total number of ordered atoms in the unit cell (including hydrogen), n_{symop} is the number of symmetry operators in the space group, n_{ASU} is the number of protein molecules in the asymmetric unit, M_r is the molecular weight of the protein (Da or g mol^{-1}), $\langle M_a \rangle$ is the number-averaged protein-atom mass ($M_r/N_{protein} \simeq 7.13$ g mol^{-1}), $N_{protein}$ is the total number of ordered atoms in the protein (including hydrogen), V_{cell} is the volume of the unit cell (in $Å^3$) and V_M is the Matthews coefficient ($Å^3 Da^{-1}$; Matthews, 1968). Since protein consists of more than one kind of atom, the effective per-atom structure factor f_a is given by the number-weighted average of the square structure factors of each atom type,

$$N_{cell} \langle f_a^2 \rangle \cong N_C f_C^2 + N_N f_N^2 + N_O f_O^2 + N_H f_H^2 \dots, \qquad (8)$$

where $\langle f_a^2 \rangle$ is the number-averaged squared atomic structure factor of protein (electron^2), N_{Ee} is the number of ordered atoms of element Ee and f_{Ee} is the atomic structure factor of element Ee (electron equivalents). In this work, atomic form factors were calculated using the five-Gaussian fit approximation used by the CCP4 suite (Collaborative Computational Project, Number 4, 1994; Winn, 2003) and tabulated in International Tables for Crystallography Vol. C (Maslen et al., 1999b). Given the atomic composition of protein provided in §2.5, this average atomic structure factor of protein is roughly equivalent to that of boron ($f_a \simeq 5$ electrons for forward scattering). This is because half of the atoms in protein are hydrogen and this brings down the number-averaged quantities $\langle f_a^2 \rangle$ and $\langle M_a \rangle$. However, the quotient $f_N^2/14$ is at worst 14% greater than $\langle f_a^2 \rangle/\langle M_a \rangle$ between 1.5 and 4 Å resolution, so if 14% error in calculated intensity is tolerable then protein can be considered to be made of an equal mass of nitrogen.

Note that (8) only applies for ∼4 Å resolution and better, where the approximations of Wilson (1942, 1949) hold, and recall that the structure factors F and f_a depend on the d-spacing of the spot (d). The contribution of each atom is also modified by an atomic B factor (Maslen et al., 1999a) identical to those listed in the Protein Data Bank (PDB; Berman et al., 2002). It is important to note that the B factor is the only model of intrinsic crystal disorder used in this work. Although there is reason to believe that disorder in crystals is more complicated than this (Welberry, 2004), B factors remain the formalism for describing disorder in crystallographic refinement (Tronrud, 2007; Brunger, 2007; Murshudov et al., 1997, 1999; Winn et al., 2003; Zwart et al., 2008). Fundamentally, Debye's argument (Debye, 1915) was that the effect of atomic displacements from their ideal lattice points is dominated by the mean square atomic displacement $\langle u_x^2 \rangle$, a result that Waller (1923, 1925) related to temperature and Ott (1935) derived rigorously (James, 1962). B factors form a resolution-dependent 'weight' for the contribution of each atom and atoms with low B factors will contribute a larger fraction of the total

scattering at high angles than atoms with high B factors. However, as long as the contribution of each protein atom is similar at a given resolution of interest we may substitute the Wilson B factor (Wilson, 1949; Shmueli & Wilson, 1999) for all the atomic B factors and arrive at a general expression for the average square structure factor of a unit cell,

$$\langle F^2 \rangle \cong \frac{V_{\text{cell}}}{V_M \langle M_a \rangle} \langle f_a^2 \rangle \exp\left[-2B\left(\frac{\sin\theta}{\lambda}\right)^2\right], \qquad (9)$$

where $\langle F^2 \rangle$ is the average value of the squared structure factor of the unit cell (electrons^2), V_{cell} is the volume of the unit cell (Å^3), V_M is the Matthews coefficient ($\text{Å}^3 \, \text{Da}^{-1}$ or $\text{Å}^3 \, \text{mol g}^{-1}$; Matthews, 1968), $\langle M_a \rangle$ is the number-averaged protein-atom mass ($M_r/N_{\text{protein}} \simeq 7.1 \, \text{g mol}^{-1}$), $\langle f_a^2 \rangle$ is the number-averaged squared atomic structure factor of protein (electrons^2), B is the average (Wilson) B factor (Å^2), θ is the the Bragg angle and λ is the X-ray wavelength (Å).

Since $\langle f_a \rangle$ and $\langle M_a \rangle$ are essentially constants for protein and V_M also has a restricted range (Matthews, 1968; Kantardjieff & Rupp, 2003), it is readily apparent that substituting $\langle F^2 \rangle$ from (9) for $|F|^2$ in (1) does indeed cancel one of the $1/V_{\text{cell}}$ terms. For example, if $V_M = 2.5 \, \text{Å}^3 \, \text{Da}^{-1}$, $d = 2.5 \, \text{Å}$ and $B = 0$, (9) reduces to $\langle F^2 \rangle \simeq 0.2 V_{\text{cell}}$. That is, given two protein crystals with the same V_{xtal} (and Wilson B factor) but one with V_{cell} twice that of the other, the average spot intensity from the large unit-cell crystal will be half of that from the smaller unit-cell crystal.

2.8. Exposure time and multiplicity

The exposure time (t) does not appear explicitly in (1) because it is hidden in the rotation speed $\omega = \Delta\Phi/t$, where $\Delta\Phi$ is the rotation covered during an exposure (in radians). What happens if the crystal is not rotated during the exposure? Does the spot intensity become infinite? Of course not, but in reality it does approach the intensity of the incident beam as the mosaic spread approaches zero, the mosaic domain volume becomes large and the X-ray beam becomes perfectly monochromatic and parallel. This limiting case is routinely achieved with the perfect silicon crystals used in monochromators, where nearly 100% of X-rays at a desired wavelength are reflected, a treatment which requires the dynamical theory of diffraction (Authier, 2004). (1) is based on what is known as the kinematical approximation to the dynamical theory and assumes that the mosaic domains are small compared with the attenuation length of the X-rays in the crystal and that the drop in the main-beam intensity owing to diffraction is negligible, which is generally a very good assumption for protein crystals (see μ^{-1} values at the end of §2.5).

What value then should we choose for $\Delta\Phi$? It cannot be smaller than the mosaic spread if we are to fully record a spot, but since we are interested in collecting a complete data set we must set $\Delta\Phi$ to the full rotation range of the data set and set t to the total accumulated exposure time of the data set (t_{DS}). The average angular velocity for recording each spot is then simply $\omega = \Delta\Phi/t_{\text{DS}}$. Now, several spots belonging to the same unique hkl index may be observed in a given data set, so

account must be taken of the extra signal available from merging equivalent observations. Any relp that is not in the blind region (see §2.3) will cross the Ewald sphere twice during a 360° rotation, as will the Friedel mate. Therefore, a crystal belonging to a space group with n_{symop} symmetry operators will produce a total of $4n_{\text{symop}}$ observations of each accessible unique hkl index (merging Friedel mates). For simplicity, we will use 360° for $\Delta\Phi$ and multiply the single-spot intensity by $4n_{\text{symop}}$,

$$\omega_{\text{eff}} = \frac{2\pi}{4n_{\text{symop}} t_{\text{DS}}}, \qquad (10)$$

where ω_{eff} is the effective angular velocity for the data set (radians s^{-1}), $2\pi = 360°$, n_{symop} is the number of symmetry operators in the space group and t_{DS} is the total accumulated exposure time of a complete data set (s). That is, ω_{eff} is the angular velocity of a 360° data set. In practice, a data-collection strategy (Dauter, 1999) is often devised to take advantage of reciprocal-space symmetry and collect a complete data set with $\Delta\Phi < 360°$, but such strategies are generally planned to finish at the end of the crystal's useful life (discussed in Appendix C) so t_{DS} is the same. The per-image exposure time is increased and this decreases ω, but it also decreases the number of observations, so ω_{eff} formally does not change. That is, a strategized data set will contain fewer but proportionally brighter spots and the radiation-damage-limited photon count is independent of the collection strategy.

This does not mean a data-collection strategy is useless! A well designed strategy minimizes the noise accumulation and resource consumption inherent in using a given set of equipment, such as the read-out noise of a CCD chip or the time required to collect the data, but a discussion of these concerns is beyond the scope of this work. Here we are interested in the absolute minimum crystal size, even given an ideal diffractometer, so we assume that the only source of noise in a spot is the photon-counting noise (shot noise) of the Bragg-scattered photons themselves and all other sources of noise, including the contribution of background scattering, are assumed to be negligible.

2.9. Absorption and dose

The attenuation factor A described in §2.5 is often incorrectly referred to as an 'absorption factor', but attenuation refers to every process for removing photons from a beam of light, including scattering. Absorption is the process of transferring energy from the beam into the substance of the crystal and the amount of energy 'deposited' into a sample per unit mass is the dose (Gy or J kg^{-1}). The mass of our spherical crystal is simply its density (ρ) multiplied by its volume $V_{\text{xtal}} = 4\pi R^3/3$ and the available energy is the photon energy (E_{ph}) multiplied by the number of photons that were not transmitted. The latter is the number of incident photons ($I_{\text{beam}} \times \pi R^2$) multiplied by the fraction $1 - T_{\text{sphere}}(0, \mu, R)$ (see equation 5). In this way, the calculation of dose is related to that of the attenuation factor (A) because the process of dose deposition begins with a photon–atom interaction, but

not every interaction deposits the full photon energy as dose. Some photons are merely scattered, depositing little or no energy, and in some cases absorbed energy is fluoresced away (Paithankar *et al.*, 2009). Seltzer (1993) accounted for such energy-loss mechanisms by assuming that only low-energy charged particles represent a 'deposit' of dose and tabulated the result as the mass energy-absorption coefficient μ_{en}. Operationally, calculating absorption instead of attenuation amounts to substituting μ_{en} for μ_{xtal} in (5), which leads to

$$D_{en} = \frac{3}{4} \frac{q_e E_{ph} I_{beam} t}{R \rho} [1 - T_{sphere}(0, \mu_{en}, R)], \qquad (11)$$

where D_{en} is the dose in Gy (J kg^{-1}), q_e is the electron charge (1.6022×10^{-19} J eV^{-1}), E_{ph} is the photon energy (eV/photon), I_{beam} is the incident-beam intensity (photons s^{-1} m^{-2}), t is the exposure time (s), ρ is the density of the sphere material (kg m^{-3} or g l^{-1}), R is the radius of the sphere (m) and μ_{en} is the mass energy-absorption coefficient of the sphere material (m^{-1}). The subscript 'en' denotes the use of the Seltzer (1993) coefficient. Note that the $1/R$ term in (11) is effectively cancelled by the T_{sphere} term for typical wavelengths and crystal sizes. Take, for example, a cube-shaped crystal of the same width as our sphere, which will transmit $T_{cube} = \exp(-\mu \cdot 2R)$, and since the limit of $1 - \exp(-x)$ as $x \to 0$ is x, one can see that the $(1 - T)$ term approaches $\mu \cdot 2R$ when most of the beam is transmitted. This is generally the case for protein crystals, but we will keep (11) in its exact form and continue to use the spherical crystal model for dose and attenuation to avoid complicating our analysis of the attenuation factor (A) against resolution with the corners of a rotating cube-shaped crystal.

If the beam profile is not flat (the constant I_{beam} case assumed here and in §2.2) then some parts of the crystal will absorb more dose than others and these high-dose regions will 'count' more in the diffraction pattern than the low-dose regions because they experience a brighter part of the beam (see equation 1). Formally, we may deal with non-uniform beams as discussed in §2.2 by breaking up the crystal into tiny cubes that do experience a constant I_{beam} and then summing the resulting diffraction patterns [using equation (4) to account for the attenuation of each incident and diffracted beam]. However, we shall see in §2.11 and Appendix C that such a treatment is unnecessary because the damage-limited photon yield per spot is independent of I_{beam}, obviating the need to integrate over the beam profile. That is, given a long enough exposure time every part of the crystal will eventually 'burn out' and contribute whatever it will contribute to the diffraction pattern. Therefore, for simplicity, we keep the 'average dose' given by (11) and assume that the entire crystal is 'evenly cooked' with no significant microscopic variation in the dose across the crystal.

2.10. Photoelectron escape and the meaning of 'dose'

Cowan, Nave and Hill (Nave & Hill, 2005; Cowan & Nave, 2008) have pointed out that as the size of a protein crystal (R) is reduced it eventually approaches the size of a primary photoelectron track (R_{PE}) and the electrons themselves will start to escape. When this happens, the energy 'deposited' within the crystal (dose) will be less than that predicted by (11).

In general, dose calculations are not simple and although a sphere is the simplest possible shape, (11) comes with certain caveats. For example, if R becomes large compared with μ_{en}^{-1} of the crystal material then some fraction of the photons scattered from the core will be absorbed before escaping the sphere and some of the energy discounted to scattering by Seltzer must be added back to the dose. A similar correction must also be made for energy assumed to be lost to fluorescence if R becomes large compared with μ_{en}^{-1} for the energy of the fluorescent photons (Paithankar *et al.*, 2009). Conversely, as R becomes comparable to R_{PE} the dose given by using μ_{en} will be too high.

Fundamentally, the flow of energy between attenuation and radiation damage is a shower of particles which quickly divides the energy of the initial photon among a large number of atoms distributed in space. For example, a photoelectric absorption event results in an excited atom and a photoelectron (Einstein, 1905; Hubbell, 2006) and the excited atom then relaxes by emitting a fluorescent photon (Moseley, 1913) or more electrons *via* Auger (Meitner, 1922; Auger, 1925) or Coster–Kronig (Coster & Kronig, 1935) processes (ICRU, 1983). These particles travel some distance before colliding with another atom and this cascade continues, with the number of excited atoms increasing and the magnitude of transferred energy decreasing with each subsequent collision. However, the distribution of events is not entirely random, as energy transfer requires an allowed electronic transition in the material. Initially, at high energies, the number of allowed transitions is small (photoelectric absorption by deep shells and scattering), but the list of possible transitions increases dramatically at lower energy. Chemical transformations take place once the magnitude of energy transfer approaches that of the strongest chemical bonds in the sample ($\sim$1 eV or 100 kJ mol^{-1}) and there are a very large number of such states excited by a single X-ray photon.

Unfortunately, such a complete treatment of energy flow is not only beyond the scope of this work but is beyond the current understanding of radiation physics in complex substances. For example, the available transitions or 'oscillator strength' in pure water between 30 and 100 eV are still poorly understood (Garrett *et al.*, 2004). Dose calculations with particle-tracking simulation codes such as *EGS* (Nelson *et al.*, 1985; Kawrakow & Rogers, 2001; Edimo *et al.*, 2008) or *MCNP* (Hendricks *et al.*, 2000; Chiavassa *et al.*, 2005; Chibani & Li, 2002) take into account carefully tabulated single- and double-differential cross-sections of all known interactions between atoms, photons and electrons, but once a particle energy drops below 1 keV it is added to the 'dose' because this is where most of the cross-section tabulations end. This means that even these highly sophisticated dose calculations will systematically underestimate track lengths by the range of 1 keV electrons. Cole (1969) measured this to be $\sim$0.06 µm in collodion plastic, so *MCNP* will overestimate the dose to crystals of the order of 60 nm and smaller.

Perhaps the most important caveat is that photoelectron escape formally violates the fundamental dosimetric principle of charged-particle equilibrium (CPE; Attix, 1986; Moussa et al., 2006), making simulation results difficult to interpret. The concern over violating CPE arises because more than half of the energy 'deposited' by a photoelectron is not in the form of ionizations but rather charge-neutral electronic excitations. Significantly more energy is deposited in this non-ionizing form at the beginning of an electron track than at the end (ICRU, 1983). No doubt this energy destabilizes the molecules that receive it, but probably not in the same way as energy deposited by ionizing interactions. Since it is not clear which kind of energy transfer is relevant to the fading of diffraction spots, the impact of 'dose' may vary along the track.

To date, all dose-calibrated radiation-damage measurements have been conducted with samples larger than the relevant photoelectron tracks and the dose has been calculated using coefficients such as μ_{en}, so we shall continue to use μ_{en} for dose in this work. However, in anticipation of future developments we shall introduce a Nave–Hill 'capture fraction' f_{NH} to represent the fraction of the conventionally calculated dose D_{en} from (11) remaining in the crystal and contributing to the 'true' dose (D_{reso}) that is relevant to resolution-degrading chemical transformations. For large crystals in ~1 Å X-ray beams we assert that $f_{NH} = 1$ and in our highly symmetric case of a uniform beam and a spherical crystal in a vacuum this correction can only depend on the radius of the crystal R and the X-ray photon energy (E_{ph}). Although an exact expression cannot be derived at this time, a rough estimate of f_{NH} is useful for detecting when a crystal has reached a size where the Nave–Hill effect may have a significant impact. Since photoelectrons are preferentially emitted in a direction normal to the incident beam and deposit energy more-or-less evenly along their track, it is assumed here that the rough effect of photoelectron escape will be to enlarge the volume over which the dose is deposited in a single direction and thereby reduce the dose to the crystal by a fraction

$$f_{NH}(R, E_{ph}) = \frac{D_{reso}}{D_{en}} \simeq \frac{R}{R + R_{PE}(E_{ph})}, \quad (12)$$

where E_{ph} is the photon energy (eV/photon), R is the radius of the spherical crystal (m) and $R_{PE}(E)$ is the range of a photoelectron of energy E derived by Cole (1969) (m). Note that for simplicity the K-shell energy of the atom that emits the photoelectron has not been deducted from the photon energy before applying it to Cole's formula, nor have Compton electrons been considered, but these are not likely to be the largest source of error in (12). It must be stressed that this equation is a very rough estimate only and could easily be off by a factor of two or more when $R \ll R_{PE}$. However, it is instructive to show that f_{NH} is expected to reduce the dose roughly as the first power of R once R becomes less than R_{PE}.

To demonstrate the potential variability of f_{NH} calculations, we conducted MCNP (Hendricks et al., 2000) simulations of a sphere with radius R and the density and atomic composition of a protein crystal given in §2.5 illuminated in a vacuum by X-rays of various energies. The resulting minimum crystal

sizes are plotted against those obtained using (12) in Fig. 2. Note that certain conclusions such as the optimum photon energy to use clearly depend on how f_{NH} is calculated. The MCNP calculation is probably more reliable than the simplistic model in (12), but the caveats mentioned above have yet to be addressed.

2.11. Radiation damage

The radiochemical mechanism behind the fading of diffraction spots is not presently clear (Garman & Nave, 2009), but the connection to dose has been calibrated experimentally. Specifically, it was pointed out by Holton (2009) and Howells et al. (2009) that the general trend reported by Howells et al. (2009), namely $D_{1/2} \simeq 10d$ MGy, where d is the feature size in Å, is remarkably consistent with the independent observations of both Owen et al. (2006) and Kmetko et al. (2006) (see Fig. 3) if the average spot intensity at a given resolution fades exponentially,

$$\langle I \rangle = \langle I \rangle_{ND} \exp\left[-\ln(2) \frac{D_{reso}}{Hd} \right], \quad (13)$$

where $\langle I \rangle$ is the average spot intensity (photons/spot) after absorbing a dose D_{reso}, $\langle I \rangle_{ND}$ is the average spot intensity (photons/spot) expected in the absence of radiation damage, $\ln(2)$ is the natural log of two (~0.7), D_{reso} is the deposited dose that is relevant to spot fading (MGy), H is the criterion of Howells et al. (2009) (10 MGy Å^{-1}) and d is the d-spacing in Å.

Note that here we use D_{reso} because it was defined in the last section as the resolution-degrading dose, but for currently
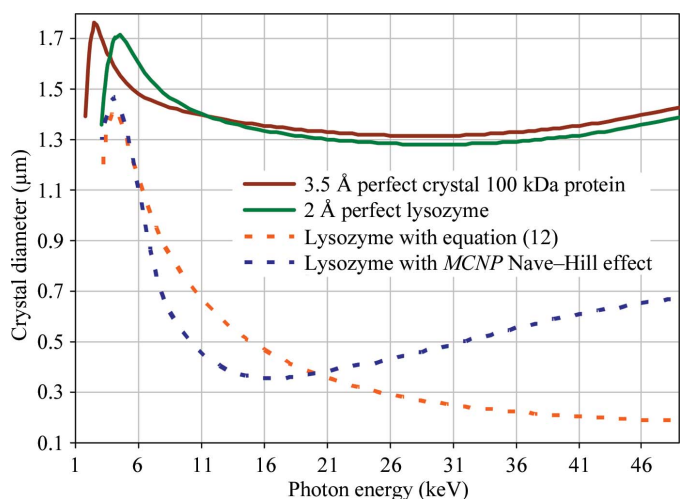

Figure 2
Wavelength-dependence of the minimum required crystal size. All plotted calculations used $V_M = 2.4$ Å^3 Da^{-1}, Wilson $B = 0$ and four photons/hkl in the indicated resolution bin. The crystal size required for 2 Å data from lysozyme and 3.5 Å data from a 100 kDa protein are essentially identical as these cases balance scattering power with data-quality requirements. Solid lines were calculated neglecting photoelectron escape ($f_{NH} = 1$) and dotted lines represent two different models for photoelectron loss: that given by (12) (orange) and a full particle-tracking dose calculation with the program MCNP (blue). The sharp reversal of the curves at low energy is a consequence of the onset of backscattering, where the Lorentz factor spikes.

available spot-fading data this is the same as D_{en} from (11) ($f_{NH} = 1$). We use angle brackets $\langle\rangle$ to emphasize that (13)

Data taken from Owen *et al.* (2006)

(a)

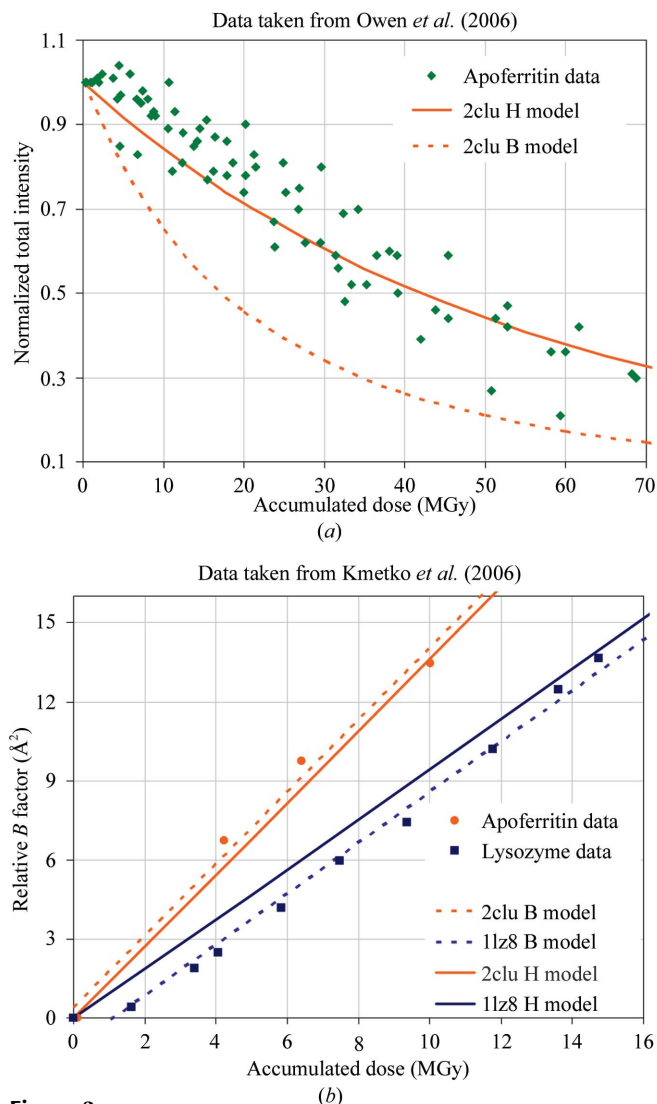Data taken from Kmetko *et al.* (2006)

(b)

Figure 3

Radiation-damage model. The observations made by Owen *et al.* (2006) and Kmetko *et al.* (2006) are reproduced with permission from the original publishers and plotted against predicted curves derived from two alternative radiation-damage models. The 'H model' is an exponential decay of spot intensity with dose and the 'B model' is the dose-dependent *B*-factor model suggested by Kmetko *et al.* (2006). The 'H model' predictions were made by applying (13) to intensities derived from the observed structure-factor file deposited with the indicated PDB entry and then computing the sum of all intensities (*a*) followed by scaling the 'simulated damage' intensities to the 'zero-dose' intensities (*b*) using the procedure described by Kmetko *et al.* (2006). The 'B model' prediction curves (dotted lines) were prepared similarly except that the 'simulated damage' intensities were generated by applying the relevant dose-dependent *B* factor reported by Kmetko *et al.* (2006). All 'H model' curves (solid lines) used the same value of *H* (10 MGy Å^{-1}) and therefore may explain the dissimilar 'sensitivity parameter' observed by Kmetko *et al.* (2006) for apoferritin and lysozyme (orange circles *versus* blue squares, respectively). It is clear from (*a*) that the 'B model' is at odds with the observations of Owen *et al.* (2006) (green diamonds), although the same predicted intensities are in very good agreement with the data points from Kmetko *et al.* (2006) (orange circles). Agreement between these two studies is restored, however, if we accept the 'H model' where the resolution-dependence of radiation damage is exponential as opposed to a Gaussian (B model).

describes the decay of average spot intensity at a given *d*-spacing, as opposed to the decay of any particular spot. Realistically, individual spots may follow different paths of decay that are not necessarily exponential (Blake & Phillips, 1962; Banumathi *et al.*, 2004), but in this work we are only interested in the average spot intensity in a given resolution bin and the argument for (13) is based largely upon spot-fading measurements.

The meta-analysis of Howells *et al.* (2009) did not include the observations made by Owen *et al.* (2006) or Kmetko *et al.* (2006), but we reproduce in Fig. 3 the observations presented in these works superimposed on predictions made by our radiation-damage model (H model) and the dose-dependent *B*-factor model (B model) suggested by Kmetko *et al.* (2006). We selected PDB entries 2clu and 1lz8 as representative of apoferritin and lysozyme, respectively, because 2clu claims a similar resolution limit to that observed in Owen *et al.* (2006) and 1lz8 is the entry for lysozyme reported by Kmetko *et al.* (2006). It should be noted that the same value of *H* (10 MGy Å^{-1}) was used for all 'H model' curves in Fig. 3 and this was not 'fitted' to the plotted data points in any way, so the agreement between all observations and the 'H model' predictions (solid lines) is quite remarkable. In fact, the 'H model' predictions in Fig. 3(*b*) were intentionally offset to pass through the origin so that the 'H model' lines would not obscure the least-squares fitted lines of the 'B model'. In this work we use the 'H model' because it is in best agreement with both these studies as well as 20 other radiation-damage experiments surveyed by Howells *et al.* (2009).

However, spot-fading experiments measure the same spots over and over again and we are interested in the total accumulated intensity $\langle I\rangle_{DL}$ at the 'damage limit' (T_{DL}), so we must integrate (13) over time. This integral is performed in Appendix C, where we show that integrating over an exponential decay is equivalent to accumulating a nondecaying intensity for less time, and applying the proportionality constant gives

$$\langle I\rangle_{DL} = \frac{\langle I\rangle_{ND}}{t_{DS}}\frac{0.1f_{decayed}4Hd\lambda R\rho}{3\ln(2)f_{NH}hcI_{beam}[1 - T_{sphere}(0, \mu_{en}, R)]}, \quad (14)$$

where $\langle I\rangle_{DL}$ is the average damage-limited intensity (photons/spot) at a given resolution, $\langle I\rangle_{ND}$ is the average spot intensity (photons/spot) expected in the absence of radiation damage, t_{DS} is the exposure time for the data set (s), 0.1 is a factor for converting three units λ from Å to m, ρ from g cm^{-3} to kg m^{-3} and MGy to Gy, $f_{decayed}$ is the fractional progress toward completely faded spots at end of the data set, *H* is Howells's criterion (10 MGy Å^{-1}), *d* is the resolution of interest (Å), λ is the X-ray wavelength (Å), *R* is the radius of the spherical crystal (m), ρ is the density of the crystal ($\sim$1.2 g cm^{-3}), f_{NH} is the Nave–Hill dose-capture fraction, *h* is Planck's constant (6.626 × 10^{-34} J s), *c* is the speed of light (299 792 458 m s^{-1}), I_{beam} is the incident-beam intensity (photons s^{-1} m^{-2}) and μ_{en} is the mass energy-absorption coefficient of the sphere material (m^{-1}). Note that the 'damage limit' was defined in Appendix C as the point when spot intensity has decayed by

some fraction (f$_{decayed}$) of the initial 'undamaged' value. For example, Owen *et al.* (2006) recommended ending the data collection when the average spot intensity fades to ~0.7 of the undamaged value (f$_{decayed}$ = 0.3), but the level of concern over radiation damage for a particular project may inspire some investigators to exceed this limit or set a more conservative limit (Holton, 2009).

The value of $\langle I \rangle_{ND}$ is simply the average value of spot intensity as given by (1) and computation of this average was accomplished by replacing the terms in (1) that vary from spot to spot with their average values and also by substituting ω_{eff} from (10) to convert spot intensities into merged *hkl* intensities,

$$\frac{\langle I \rangle_{ND}}{t_{DS}} = I_{beam} r_e^2 \frac{V_{xtal}}{V_{cell}} \cdot \frac{4 n_{symop}}{2\pi} \frac{\lambda^3}{V_{cell}} \langle LP \rangle f_{obs} \cdot A \cdot \langle F^2 \rangle. \quad (15)$$

We may now substitute $\langle I \rangle_{ND}/t_{DS}$ from (15) into (14) and then replace $\langle LP \rangle f_{obs}$, $\langle F^2 \rangle$, V_{cell} and V_{xtal} with their expanded forms from (6), (9), (7) and $4\pi R^3/3$, respectively, to yield the fully qualified expression for damage-limited spot intensity,

$$\langle I \rangle_{DL} = \frac{2\pi}{9} \frac{10^5 r_e^2}{hc} \frac{f_{decayed} \rho R^4 \lambda^4}{f_{NH} n_{ASU} M_r V_M^2} \frac{0.5 \lambda H}{\ln(2) \sin\theta} \frac{T_{sphere}(2\theta, \mu, R)}{[1 - T_{sphere}(0, \mu_{en}, R)]}$$

$$\times \frac{(3 + \cos 4\theta)}{\sin\theta} \frac{\langle f_a^2 \rangle}{\langle M_a \rangle} \exp\left[-2B \left(\frac{\sin\theta}{\lambda} \right)^2 \right], \quad (16)$$

where $\langle I \rangle_{DL}$ is the average damage-limited intensity (photons/*hkl*) at a given resolution, 10^5 is a factor for converting four units: R from μm to m, r_e from m to Å, ρ from g cm^{-3} to kg m^{-3} and MGy to Gy, r_e is the classical electron radius (2.818×10^{-15} m), h is Planck's constant (6.626×10^{-34} J s), c is the speed of light ($299\,792\,458$ m s^{-1}), f$_{decayed}$ is the fractional progress toward completely faded spots at the end of the data set, ρ is the density of the crystal (~1.2 g cm^{-3}), R is the radius of the spherical crystal (μm), λ is the X-ray wavelength (Å), f$_{NH}$ is the Nave–Hill dose-capture fraction (1 for large crystals; Nave & Hill, 2005), n_{ASU} is the number of proteins in the asymmetric unit, M_r is the molecular weight of the protein (Da or g mol^{-1}), V_M is the Matthews coefficient (~2.4 Å^3 Da^{-1}), H is Howells's criterion (10 MGy Å^{-1}), θ is the Bragg angle, $\langle f_a^2 \rangle$ is the number-averaged squared structure factor per protein atom (electron^2), $\langle M_a \rangle$ is the number-averaged atomic weight of a protein atom (~7.1 Da), B is the average (Wilson) temperature factor (Å^2), μ is the attenuation coefficient of the sphere material (m^{-1}) and μ_{en} is the mass energy-absorption coefficient of the sphere material (m^{-1}). Note that the incident-beam intensity (I_{beam}) is missing from this equation because spot intensity was integrated out to the 'damage limit' where the average spot has decayed by a given fraction (f$_{decayed}$). Note that the crystal symmetry is also missing, as the n_{symop} term from (10) was cancelled by another n_{symop} term in the expression for the average structure factor (7), implying that the damage limit is more closely related to the number of molecules in the crystal than it is to the number of unit cells. One R in the R^4 term is effectively cancelled by the $(1 - T)$ term for all but the very largest protein crystals

and one λ term is roughly cancelled (within ~30% between 7 and 17 keV) by the $\langle LP \rangle f_{obs}$ factor.

Although (16) may appear somewhat intimidating, it is both instructive and useful to examine it in this expanded form as this eases the incorporation of different macromolecule types, radiation-damage models and crystal shapes. For example, $\langle f_a^2 \rangle$, $\langle M_a \rangle$, μ and μ_{en} may be replaced with appropriate values for nucleic acids. The ln(2) term arises from the definition of H as the dose required to reduce spot intensities at a given *d*-spacing ($d = 0.5\lambda/\sin\theta$) by half, so Hd and ln(2) are grouped together. Crystals that are more sensitive than normal to radiation damage per unit of dose, as was reported for dodecin by Murray *et al.* (2005), may be represented by using a smaller value of H and a more sophisticated resolution-dependent damage model might replace Hd/ln(2) with an arbitrary function $H(d)$. Also, considering the crystal to be a cube with edge $2R$ instead of a sphere of radius R simply changes the leading $2\pi/9$ term to unity and replaces T_{sphere} with $\exp(-\mu_{en} 2R)$. The increased scattering power of the cube arises because $(2R)^3$ is roughly twice $4\pi R^3/3$ and the damage-limited intensity (photons/*hkl*) scales linearly with crystal volume.

3. Results and discussion

We are now prepared to calculate the diameter of the smallest protein crystal that can be expected to produce a complete data set on an ideal diffractometer: a very large perfect detector, a perfect shutter and a perfect spindle with a uniform and flicker-free X-ray beam bathing a spherical protein crystal in a vacuum. The noise from such a machine is dominated by photon counting, so if we require a signal-to-noise ratio (SNR) of 2.0 in the outer resolution bin of say 2 Å then the average *hkl* in this bin must accumulate at least four photons ($I/\sigma = I/I^{1/2}$). If there are other sources of noise, such as background scattering, then more than four photons will be required, but since it is theoretically possible to reduce background to a negligible level (see §3.2), we will begin with this limiting case.

3.1. Zero-background case

We begin by neglecting the Nave–Hill effect because it has yet to be measured and represents the greatest unknown in the dose calculation. With f$_{NH}$ = 1, (16) predicts that a 1.2 μm diameter sphere of perfect lysozyme crystal (B = 0; M_r = 14 300 Da; V_M = 2.0 Å^3 Da^{-1}) in a beam of 1 Å X-rays will scatter an average of 4 photons/*hkl* ($\langle I \rangle_{DL}$) at 2 Å resolution before the radiation-damage limit is reached (f$_{decayed}$ = 0.3). This limit is independent of exposure time or beam flux since the total accumulated fluence (photons/area) is dictated by the damage limit.

If we now involve f$_{NH}$ from (12) or from *MCNP* simulations then the four-photon lysozyme crystal size shrinks to 0.5 or 0.34 μm, respectively. In addition to this, if we allow the spots to fade away completely (f$_{decayed}$ = 1) then 0.81 μm (f$_{NH}$ = 1), 0.28 μm (equation 12) or 0.19 μm (MCNP) crystals will yield

4 photons/*hkl* at 2 Å. There are a number of reasons why complete decay is not a realistic damage limit, not the least of which is the biological relevance of the results (Owen *et al.*, 2006), but it is instructive to consider an infinite exposure time here because photon counting is the only kind of noise that is theoretically impossible to eliminate.

Immediately, the next questions to ask are how this limit is influenced by the choice of photon energy, desired resolution, degree of disorder in the crystal and molecular weight of the protein or combinations thereof. (16) is the exact formula for relating all these quantities, but as the questions to be asked occupy a large multidimensional parameter space it is instructive to graph the influence of each parameter separately. Since many of the variables in (16) change with the X-ray wavelength, we begin by plotting the minimum crystal size against photon energy in Fig. 2. This graph is similar to the 'I_E' quantity obtained by Arndt (1984), except that here the *y* axis is on an absolute scale. The energy-dependence is remarkably flat and this result is consistent with experimental observation (Gonzalez *et al.*, 1994). The 'spike' in crystal size at very low photon energy arises from a sharp upswing in $\langle LP \rangle$ when the relp grazes the back of the Ewald sphere just before f_{obs} drops to zero and the 2 Å curves stop at 3.1 keV because it is not possible to collect 2 Å data with wavelengths longer than 4 Å. The minimum-size curve for 4 photons/*hkl* at 3.5 Å from a perfect crystal of a 100 kDa protein is provided to fill this low-energy gap as well as demonstrate how simultaneously decreasing the scattering power and lowering the desired data quality can 'coincidentally' result in the same crystal size requirement.

Graphs of minimum crystal size against molecular weight (Fig. 4), n_{ASU}, $f_{decayed}$, H and absorption coefficients are all very similar because 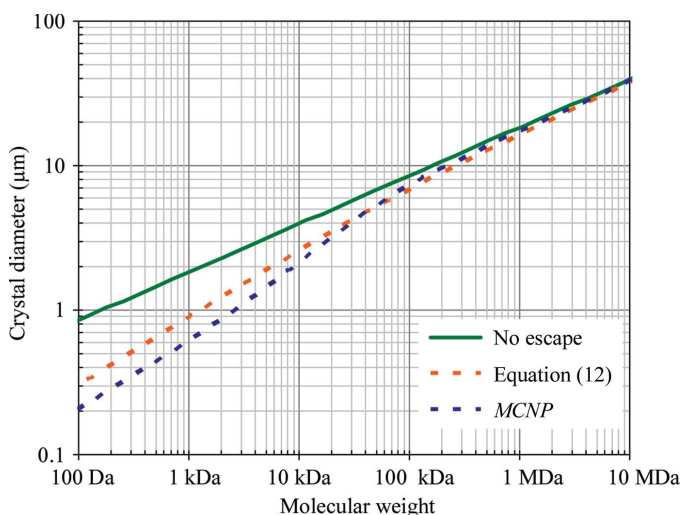each of these terms scales linearly with crystal volume. An examination of (16) reveals that these variables are not strongly coupled to any others if $R << \mu^{-1}$, as absorption is proportional to R and attenuation is negligible in this case. The solvent content V_M dependence is also not graphed because this is just a plot of a square-root function passing through 1.2 μm for $V_M = 2.0$ Å^3 Da^{-1}, $\lambda = 1$ Å, $d = 2$ Å and $B = 0$.

The graph of minimum crystal size against desired resolution may curve upward or downward depending on the value chosen for the Wilson B factor (dashed lines in Fig. 5) and indeed it is not surprising that the degree of disorder in a protein crystal has a strong influence on the diffraction limit. What is surprising is that if the B factor is always selected to follow the empirically derived formula ($B = 4d^2 + 12$) presented by Holton (2009), one obtains the straight solid lines in Fig. 5. This remarkable result appears to be a consequence of this B-factor formula effectively cancelling the resolution-dependence of the average atomic form factor (8), implying that the number of photons required to detect the weakest spots is relatively fixed from crystal to crystal. Regardless of the origin, Fig. 5 immediately suggests an empirical formula for the required crystal size given an observed resolution limit,

$$2R = 0.011(\langle I \rangle_{DL} M_r)^{1/3} \exp\left(\frac{4.74}{d}\right), \qquad (17)$$

where $2R$ is the required diameter of the crystal (μm), 0.011 is a scale factor assuming $V_M = 2.4$ Å^3 Da^{-1}, $\langle I \rangle_{DL}$ is the desired damage-limited intensity (photons/*hkl*) at a given resolution, M_r is the molecular weight of the protein (Da or g mol^{-1}) and $4.74 = 4\pi^2 r_a^2$, where r_a is the radius of gyration of a protein atom (Å) and d is the resolution of interest (Å). This is not to say that a crystal of diameter $2R$ will diffract to resolution d, but rather that a crystal of a protein with mass M_r found to diffract to resolution d probably has a Wilson B factor that will
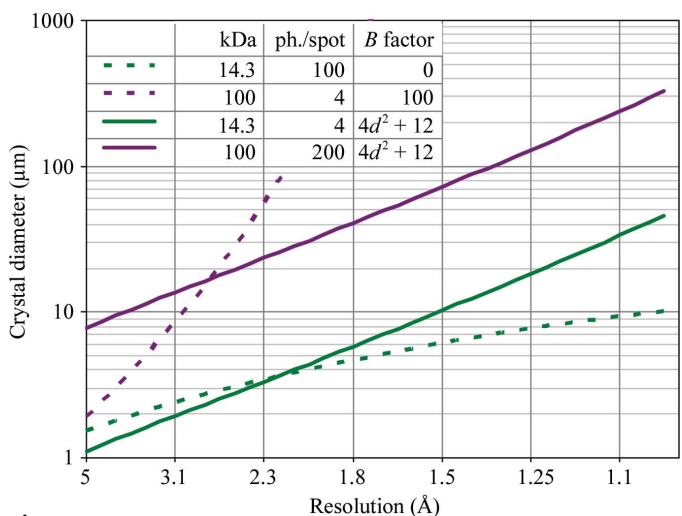
Figure 4
Molecular-weight dependence of the minimum required crystal size. All plotted calculations used $V_M = 2.4$ Å^3 Da^{-1}, 1 Å radiation, 2 Å spots and $B = 24$ Å^2. Without photoelectron escape, the required crystal volume is simply proportional to molecular weight and the two different models of photoelectron escape considered here are shown to have significant yet different effects for crystals smaller than a few micrometres wide, as this is the linear dimension of a photoelectron track (R_{PE}).

Figure 5
Resolution-dependence of the minimum required crystal size. All plotted calculations used $V_M = 2.4$ Å^3 Da^{-1} and 1 Å radiation. The Wilson B factor strongly affects the curvature of the plot of the required crystal size for a given number of photons, but applying the empirical formula shown serendipitously simplifies this analysis, as described in the text.

require the crystal to be of diameter 2R to yield a complete data set. Until now, we have assumed that an outer resolution bin ($\langle I \rangle_{DL}$) need only gather 4 photons/hkl, but it appears that the 'detection limit' of current technology is much higher than this (described in the next section) and a value of $\langle I \rangle_{DL} = 100–200$ photons/hkl is suggested for the practical use of (17) depending on the background level.

3.2. Background scattering

X-ray background consists of scattering from air, aperture walls, fluorescence, disorder in the crystal and potentially many other sources. A full theoretical treatment of background and all other possible sources of noise in a diffraction experiment is well beyond the scope of this work, but we shall briefly describe here how the large gap between our calculated absolute minimum crystal size and those that have been determined experimentally is completely explained by background scattering alone.

A summary of experimental minimum crystal-size determinations was provided by Holton (2009), who related scattering power to data quality with an empirical 'difficulty parameter' (n_0) that increases with the quality of data needed for 'success' and decreases as instrument capabilities improve. The 'record' for obtaining a complete data set was $n_0 = 3.1$, but entering the parameters obtained in the last section into equation (3) of Holton (2009), $n_{xtals} = 1$ (number of crystals used), $d = 2.0$ Å (resolution limit), $B = 0$, $V_M = 2.0$ Å^3 Da^{-1} and $\ell_{xyz} = 1.2$ μm (crystal 'size'), we obtain $n_0 = 0.2$. This is a factor of 15 improvement over the 'record' and using $\ell_{xyz} = 0.34$ μm, as expected from the more optimistic photoelectron escape model, we arrive at $n_0 = 0.0044$, which is 700-fold less scattering power than has ever been used to collect a complete data set.

There are many possible reasons why extant beamlines may not have reached the theoretical limit, but what is clear is that more than four photons are presently required to detect the faintest spots. Indeed, the $n_0 = 3.1$ case corresponds to 64 photons/hkl [if the cubic crystal volume in Holton (2009) is taken to be V_{xtal} here]. Formally, this must arise from additional noise inflating $\sigma(I)$ beyond simply $I^{1/2}$, requiring increased I (photons/hkl) to bring $I/\sigma(I)$ back up to 2.0. An obvious source of additional noise is background scattering, so we now generalize our formula for the average signal-to-noise ratio (SNR) in the outer resolution bin from simply $\langle I \rangle_{DL}^{1/2}$ to

$$ \mathrm{SNR} = \frac{\langle I \rangle_{DL}}{\left(\langle I \rangle_{DL} + m n_{pix} I_{BG} \dfrac{T_{DL}}{n_{images}} + \sigma_{other}^2 \right)^{1/2}}, \quad (18) $$

where $\langle I \rangle_{DL}$ is the average damage-limited intensity (photons/hkl), m is the mean multiplicity (spots/hkl, counting partials as distinct spots), n_{pix} is the number of pixels involved in the average spot, I_{BG} is the average background scattering rate (photons pixel^{-1} s^{-1}) at the resolution of interest, T_{DL} is the damage-limited exposure time of the data set (s), n_{images} is the number of diffraction images in the data set and σ_{other} is the

root-mean-square of all other sources of noise (placed on a one-photon scale).

For a given camera and sample, the observed background photons/pixel on a single diffraction image will be proportional to the per-image exposure time ($t_{image} = T_{DL}/n_{images}$), indicating how I_{BG} is fixed for a given experiment. Since we are considering a damage-limited experiment, the total number of background photons that fall on the detector ($I_{BG}T_{DL}$) is also fixed, regardless of how these photons are divided into images. The practice of 'fine-slicing' (Pflugrath, 1999) reduces $I_{BG}t_{image}$, at the expense of increasing m, but in the limit of 'infinite' fine-slicing the quantity $m I_{BG} t_{image}$ approaches a constant because the background that actually falls into the three-dimensional integration region of a given spot cannot be avoided by finer slicing. Very fine slicing will start to make other sources of noise important, such as detector read-out noise, so this and all other sources of noise are lumped into σ_{other} for completeness. Nevertheless, with our hypothetical ideal diffractometer σ_{other} will be negligible.

Choosing some reasonable parameters ($m = 4$, $n_{pix} = 5 \times 5$) (18) is solved for SNR = 2.0 and $\langle I \rangle_{DL} = 64$ photons/hkl by $I_{BG}t_{image} = 10$ photons pixel^{-1}. It must be stressed that this is a very rough approximation, particularly since n_0 was not claimed to be accurate to better than a factor of two and such an error propagated through (18) becomes a factor of four in background level. Nevertheless, this $I_{BG}t_{image}$ is exactly that observed near the faintest spots shown in Fig. 4 of Moukhametzianov et al. (2008), the source of our $n_0 = 3.1$ 'record' (the detector registers 1.0 pixel levels per photon and has a 'zero' offset of 20 pixel levels).

The experience of the authors of this work is that 10 photons pixel^{-1} is on the low side of the range of background levels seen on typical diffraction images. It is more common to see hundreds of photons per pixel from crystals that only diffract to modest resolutions because the same disorder that leads to faint spots also produces diffuse scattering (James, 1962; Welberry, 2004). If we keep $n_{pix} = 5 \times 5$ and $m = 4$ as above and $I_{BG}t_{image} = 25$, 100 or even 400 photons pixel^{-1}, then satisfying SNR = 2 in (18) requires $\langle I \rangle_{DL}$ to be 102, 202 or 402 photons/hkl, respectively.

Note that reducing the multiplicity (m) by collecting the bare minimum number of images will result in no net 'gain' so long as the damage limit is reached at the end of data collection because the increased exposure time per image will increase $I_{BG}t_{image}$ to exactly compensate for any reduced multiplicity (m). On the other hand, considerable gains can be had by making the absolute background counts (photons pixel^{-1} s^{-1}; I_{BG}) lower, reducing the number of pixels occupied by spots on the detector (n_{pix}) or both.

Background scattering can never be completely eliminated, but the noise it adds to the spots can be minimized by making the spot size very small. A detailed discussion of spot size is beyond the scope of this work, but theoretically very small spots can be achieved with a perfect protein crystal (no mosaic spread), a near-zero emittance beam of very short wavelength X-rays focused on an enormous and noiseless detector with no point-spread function, very small pixels and very fine rotation

steps. Therefore, I_{BG} can be reduced to near zero, or at least to the point where the noise from background is insignificant ($\langle I \rangle_{DL} \gg mn_{pix}I_{BG}t_{image}$ in equation 18), implying that (16) with $\langle I \rangle_{DL}$ set to 4 photons/hkl represents an absolute and fundamental limit. That is, unless some way is found to change one of the parameters in (16), such as increasing H by mitigating the chemistry of global damage or decreasing f_{NH} with photoelectron escape, a lysozyme crystal smaller than 1.2 µm will never yield a complete data set to 2 Å.

3.3. Implications for micro-focus beams

The 1.2 µm size limit for perfect lysozyme crystals determined here does not imply that crystals and X-ray beams smaller than ~1 µm are useless. If a complete data set cannot be obtained from one crystal then a multi-crystal strategy (Kendrew et al., 1960; Dickerson et al., 1961), a 'needle-scanning' strategy (Moukhametzianov et al., 2008) or perhaps the 'serial crystallography' approach proposed by Starodub et al. (2008) may be employed, but the total scattering volume will have to add up to the volume of a sphere given by R in (16) using f_{NH} for the individual crystal size. For example, the volume needed for one crystal of a 100-crystal data set with final merged $\langle I \rangle_{DL}$ = 4 photon/hkl is given by using $\langle I \rangle_{DL}$ = 0.04 photon/hkl in (16).

Crystals with larger unit cells or more disorder (or both) will have to be larger than their 'perfect lysozyme equivalent' volume. For example, a lysozyme crystal with a more realistic Wilson B factor of 20 Å^2 must be 2.8 µm wide to produce 4 photons/hkl in the 2 Å bin using the $f_{decayed}$ = 0.3 damage limit and a 10 MDa asymmetric unit with V_M = 2.4 Å^3 Da^{-1} and B = 61 Å^2 must form a crystal 15 µm wide to produce 4 photons/hkl at 3.5 Å. However, as the present 'detection limit' appears to be of the order of 100 photons/hkl ($I_{BG}t_{image} \simeq$ 100 photons pixel^{-1}), these realistic lysozyme crystals will have to be 8.3 µm in diameter for 2 Å data, and 3.5 Å data from the 10 MDa case will require 43 µm crystals, limiting the usefulness of X-ray beams smaller than this.

4. Conclusions

The minimum useful protein crystal size is limited by the background photons that accumulate in the detector pixels occupied by a spot and current technologies seem to require of the order of 100 photons/hkl (after merging) to attain a signal-to-noise ratio of 2. The choice of X-ray wavelength appears to have only a minor impact on the damage-limited scattering power of a crystal, which remains proportional to the crystal volume and inversely proportional to both the molecular weight of the asymmetric unit and the square of the Matthews coefficient (Matthews, 1968) for all practical purposes. The resolution-dependence is complicated by the Wilson B factor, but relating B to d-spacing empirically revealed that damage-limited scattering power is proportional to $\exp(-14.2/d)$, where d is the d-spacing of interest. Dose reduction owing to photoelectron escape appears to be theoretically promising but difficult to predict and the current detection limit for spots

will have to be overcome for this effect to be of practical use for typical single-crystal data sets at accessible photon energies.

References

Afonine, P. V., Grosse-Kunstleve, R. W. & Adams, P. D. (2005). *Acta Cryst.* D**61**, 850–855.

Arndt, U. W. (1984). *J. Appl. Cryst.* **17**, 118–119.

Arndt, U. W. & Wonacott, A. J. (1977). *The Rotation Method in Crystallography.* Amsterdam: North-Holland.

Attix, F. H. (1986). *Introduction to Radiological Physics and Radiation Dosimetry.* New York: Wiley.

Auger, P. (1925). *J. Phys. Radium*, **6**, 205–208.

Authier, A. (2004). *Dynamical Theory of X-ray Diffraction*, revised ed. Oxford University Press.

Azároff, L. V. (1955). *Acta Cryst.* **8**, 701–704.

Banumathi, S., Zwart, P. H., Ramagopal, U. A., Dauter, M. & Dauter, Z. (2004). *Acta Cryst.* D**60**, 1085–1093.

Berger, M. J. & Hubbell, J. H. (1987). *XCOM: Photon Cross Sections on a Personal Computer.* National Bureau of Standards Internal Report NBSIR-87-3597. Gaithersburg: National Bureau of Standards.

Berman, H. M. et al. (2002). *Acta Cryst.* D**58**, 899–907.

Blake, C. C. F. & Phillips, D. C. (1962). *Biological Effects of Ionizing Radiation at the Molecular Level*, pp. 183–191. Vienna: IAEA.

Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography.* New York: Academic Press.

Bragg, W. L., James, R. W. & Bosanquet, C. H. (1921a). *Philos. Mag. Ser. 6*, **41**, 309–337.

Bragg, W. L., James, R. W. & Bosanquet, C. H. (1921b). *Philos. Mag. Ser. 6*, **42**, 1–17.

Bragg, W. L., James, R. W. & Bosanquet, C. H. (1922). *Philos. Mag. Ser. 6*, **44**, 433–449.

Brunger, A. T. (2007). *Nature Protoc.* **2**, 2728–2733.

Chiavassa, S., Lemosquet, A., Aubineau-Laniece, I., de Carlan, L., Clairand, I., Ferrer, L., Bardies, M., Franck, D. & Zankl, M. (2005). *Radiat. Prot. Dosimetry*, **116**, 631–635.

Chibani, O. & Li, X. A. (2002). *Med. Phys.* **29**, 835–847.

Cole, A. (1969). *Radiat. Res.* **38**, 7–33.

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

Compton, A. H. & Freeman, N. L. (1922). *Nature (London)*, **110**, 38.

Coppens, P. (1999). *International Tables for Crystallography*, Vol. B., 2nd ed., ch. 1.2. Dordrecht: Kluwer Academic Publishers.

Cork, C., Fehr, D., Hamlin, R., Vernon, W., Xuong, N. H. & Perez-Mendez, V. (1974). *J. Appl. Cryst.* **7**, 319–323.

Coster, D. & Kronig, R. de L. (1935). *Physica*, **2**, 13–24.

Coulibaly, F., Chiu, E., Ikeda, K., Gutmann, S., Haebel, P. W., Schulze-Briese, C., Mori, H. & Metcalf, P. (2007). *Nature (London)*, **446**, 97–101.

Cowan, J. A. & Nave, C. (2008). *J. Synchrotron Rad.* **15**, 458–462.

Creagh, D. C. & Helliwell, J. R. (1999). *International Tables for Crystallography*, Vol. C, 2nd ed., ch. 4.2.4. Dordrecht: Kluwer Academic Publishers.

Darwin, C. G. (1914). *Philos. Mag.* **27**, 315–333.

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

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

Debye, P. J. W. (1914). *Ann. Phys.* **348**, 49–92.

Debye, P. J. W. (1915). *Ann. Phys.* **351**, 809–823.

Debye, P. J. W. (1988). *The Collected Papers of Peter J. W. Debye.* Woodbridge: Ox Bow Press.

Debye, P. J. W. & Scherrer, P. (1918). *Phys. Z.* **19**, 474–483.

Dickerson, R. E., Kendrew, J. C. & Strandberg, B. E. (1961). *Acta Cryst.* **14**, 1188–1195.

Drenth, J. (1999). *Principles of Protein X-ray Crystallography.* Berlin: Springer-Verlag.

Dwiggins, C. W. (1975). *Acta Cryst.* A**31**, 395–396.

Edimo, P., Clermont, C., Kwato, M. G. & Vynckier, S. (2008). *Phys. Med.* **25**, 111–121.

Einstein, A. (1905). *Ann. Phys.* **322**, 549–560.

Ewald, P. P. (1913). *Phys. Z.* **14**, 465–472.

Facciotti, M. T., Cheung, V. S., Nguyen, D., Rouhani, S. & Glaeser, R. M. (2003). *Biophys. J.* **85**, 451–458.

Flack, H. D. & Vincent, M. G. (1978). *Acta Cryst.* A**34**, 489–491.

Garman, E. F. & McSweeney, S. M. (2007). *J. Synchrotron Rad.* **14**, 1–3.

Garman, E. F. & Nave, C. (2009). *J. Synchrotron Rad.* **16**, 129–132.

Garrett, B. C. et al. (2004). *Chem. Rev.* **105**, 355–390.

Glaeser, R., Facciotti, M., Walian, P., Rouhani, S., Holton, J., MacDowell, A., Celestre, R., Cambie, D. & Padmore, H. (2000). *Biophys. J.* **78**, 3178–3185.

Gonzalez, A., Denny, R. & Nave, C. (1994). *Acta Cryst.* D**50**, 276–282.

Gonzalez, A. & Nave, C. (1994). *Acta Cryst.* D**50**, 874–877.

Hartree, D. R. (1925). *Philos. Mag. Ser. 6*, **50**, 289–306.

Helliwell, J. R. (1999). *International Tables for Crystallography*, Vol. C, 2nd ed., ch. 2.2. Dordrecht: Kluwer Academic Publishers.

Henderson, R. (1990). *Proc. R. Soc. Lond. B Biol. Sci.* **241**, 6–8.

Hendricks, J. S., Adam, K. J., Booth, T. E., Briesmeister, J. F., Carter, L. L., Cox, L. J., Favorite, J. A., Forster, R. A., McKinney, G. W. & Prael, R. E. (2000). *Appl. Radiat. Isot.* **53**, 857–861.

Holton, J. M. (2009). *J. Synchrotron Rad.* **16**, 133–142.

Howells, M. R., Beetz, T., Chapman, H. N., Cui, C., Holton, J. M., Jacobsen, C. J., Kirz, J., Lima, E., Marchesini, S., Miao, H., Sayre, D., Shapiro, D. A., Spence, J. H. C. & Starodub, D. (2009). *J. Electron Spectrosc. Relat. Phenom.* **170**, 4–12.

Hubbell, J. H. (2006). *Phys. Med. Biol.* **51**, R245–R262.

ICRU (1983). *Microdosimetry.* Report No. 36. Washington, DC: International Commission on Radiological Units and Measurements.

James, R. W. (1962). *The Optical Principles of the Diffraction of X-rays.* London: Bell.

Kahn, R., Fourme, R., Gadet, A., Janin, J., Dumas, C. & André, D. (1982). *J. Appl. Cryst.* **15**, 330–337.

Kantardjieff, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865.

Kawrakow, I. & Rogers, D. W. O. (2001). *The EGSnrc Code System: Monte Carlo Simulation of Electron and Photon Transport.* NRCC Report PIRS-701. Ottowa: National Research Council of Canada.

Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C. & Shore, V. C. (1960). *Nature (London)*, **185**, 422–427.

Kmetko, J., Husseini, N. S., Naides, M., Kalinin, Y. & Thorne, R. E. (2006). *Acta Cryst.* D**62**, 1030–1038.

Kraft, P., Bergamaschi, A., Broennimann, Ch., Dinapoli, R., Eikenberry, E. F., Henrich, B., Johnson, I., Mozzanica, A., Schlepütz, C. M., Willmott, P. R. & Schmitt, B. (2009). *J. Synchrotron Rad.* **16**, 368–375.

Leiros, H.-K. S., Timmins, J., Ravelli, R. B. G. & McSweeney, S. M. (2006). *Acta Cryst.* D**62**, 125–132.

Leslie, A. G. W. (2006). *Acta Cryst.* D**62**, 48–57.

Li, J., Edwards, P. C., Burghammer, M., Villa, C. & Schertler, G. F. (2004). *J. Mol. Biol.* **343**, 1409–1438.

Lipson, H. & Langford, J. I. (1999). *International Tables for Crystallography*, Vol. C, 2nd ed., ch. 6.2. Dordrecht: Kluwer Academic Publishers.

MacDowell, A. A. et al. (2004). *J. Synchrotron Rad.* **11**, 447–455.

Maslen, E. N. (1999). *International Tables for Crystallography*, Vol. C, 2nd ed., ch. 6.3. Dordrecht: Kluwer Academic Publishers.

Maslen, E. N., Fox, A. G. & O'Keefe, M. A. (1999a). *International Tables for Crystallography*, Vol. C, 2nd ed., ch. 6.1. Dordrecht: Kluwer Academic Publishers.

Maslen, E. N., Fox, A. G. & O'Keefe, M. A. (1999b). *International Tables for Crystallography*, Vol. C, 2nd ed., Table 6.1.1.4. Dordrecht: Kluwer Academic Publishers.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Maxwell, J. C. (1865). *Philos. Trans. R. Soc. Lond.* **155**, 459–512.

Meitner, L. (1922). *Z. Phys. A*, **9**, 131–144.

Moseley, H. G. J. (1913). *Philos. Mag.* **26**, 1024–1034.

Moseley, H. G. J. & Darwin, C. G. (1913). *Philos. Mag.* **26**, 210–232.

Moukhametzianov, R., Burghammer, M., Edwards, P. C., Petit-demange, S., Popov, D., Fransen, M., McMullan, G., Schertler, G. F. X. & Riekel, C. (2008). *Acta Cryst.* D**64**, 158–166.

Moussa, H. M., Eckerman, K. F. & Townsend, L. W. (2006). *Radiat. Prot. Dosimetry*, **121**, 252–256.

Murray, J. W., Garman, E. F. & Ravelli, R. B. G. (2004). *J. Appl. Cryst.* **37**, 513–522.

Murray, J. W., Rudiño-Piñera, E., Owen, R. L., Grininger, M., Ravelli, R. B. G. & Garman, E. F. (2005). *J. Synchrotron Rad.* **12**, 268–275.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.

Nave, C. & Hill, M. A. (2005). *J. Synchrotron Rad.* **12**, 299–303.

Nelson, R., Sawaya, M. R., Balbirnie, M., Madsen, A. O., Riekel, C., Grothe, R. & Eisenberg, D. (2005). *Nature (London)*, **435**, 773–778.

Nelson, W. R., Hirayama, H. & Rogers, D. W. O. (1985). *The EGS4 Code System.* Stanford Linear Accelerator Center Report SLAC-265.

Ott, H. (1935). *Ann. Phys.* **23**, 169–196.

Owen, R. L., Holton, J. M., Schulze-Briese, C. & Garman, E. F. (2009). *J. Synchrotron Rad.* **16**, 143–151.

Owen, R. L., Rudino-Pinera, E. & Garman, E. F. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 4912–4917.

Paithankar, K. S., Owen, R. L. & Garman, E. F. (2009). *J. Synchrotron Rad.* **16**, 152–162.

Pflugrath, J. W. (1999). *Acta Cryst.* D**55**, 1718–1725.

Purcell, E. M. (1985). *Electricity and Magnetism*, 2nd ed. New York: McGraw-Hill.

Ramachandran, G. N. & Wooster, W. A. (1951). *Acta Cryst.* **4**, 335–344.

Sabine, T. M. (1999). *International Tables for Crystallography*, Vol. C, 2nd ed., ch. 6.4. Dordrecht: Kluwer Academic Publishers.

Sawaya, M. R., Sambashivan, S., Nelson, R., Ivanova, M. I., Sievers, S. A., Apostol, M. I., Thompson, M. J., Balbirnie, M., Wiltzius, J. J. W., McFarlane, H. T., Madsen, A. O., Riekel, C. & Eisenberg, D. (2007). *Nature (London)*, **447**, 453–457.

Schulze-Briese, C., Brönnimann, Ch., Eikenberry, E. F., Billich, H., Diez, J., Henrich, B., Kobas, M., Näf, M., Panepucci, E. & Tomizaki, T. (2007). *Acta Cryst.* A**63**, s87.

Seltzer, S. M. (1993). *Radiat. Res.* **136**, 147–170.

Shmueli, U. & Wilson, A. J. C. (1999). *International Tables for Crystallography*, Vol. B, 2nd ed., ch. 2.1. Dordrecht: Kluwer Academic Publishers.

Slater, J. C. (1929). *Phys. Rev.* **34**, 1293.

Sliz, P., Harrison, S. C. & Rosenbaum, G. (2003). *Structure*, **11**, 13–19.

Snell, E. H., Bellamy, H. D. & Borgstahl, G. E. (2003). *Methods Enzymol.* **368**, 268–288.

Standfuss, J., Xie, G., Edwards, P. C., Burghammer, M., Oprian, D. D. & Schertler, G. F. (2007). *J. Mol. Biol.* **372**, 1179–1188.

Starodub, D., Rez, P., Hembree, G., Howells, M., Shapiro, D., Chapman, H. N., Fromme, P., Schmidt, K., Weierstall, U., Doak, R. B. & Spence, J. C. H. (2008). *J. Synchrotron Rad.* **15**, 62–73.

Storm, E. & Israel, H. I. (1970). *Nuclear Data Tables*, **7**, 565–581.

Teng, T. & Moffat, K. (2000). *J. Synchrotron Rad.* **7**, 313–317.

Teng, T.-Y. & Moffat, K. (2002). *J. Synchrotron Rad.* **9**, 198–201.

Thomson, J. J. (1906). *Conduction of Electricity Through Gases.* Cambridge University Press.

Tronrud, D. E. (1997). *Methods Enzymol.* **277**, 306–319.

Tronrud, D. E. (2007). *Methods Mol. Biol.* **364**, 231–254.

Waller, I. (1923). *Z. Phys.* **17**, 398–408.

Waller, I. (1925). *Theoretische Studien zur Interferenz- und Dispersionstheorie der Röntgenstrahlen.* Dissertation. Uppsala University, Sweden.

Welberry, T. R. (2004). *Diffuse X-ray Scattering and Models of Disorder.* Oxford University Press.

Wilson, A. J. C. (1942). *Nature (London)*, **150**, 152.

Wilson, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.

Wilson, A. J. C. & Prince, E. (1999). Editors. *International Tables for Crystallography*, Vol. *C*, 2nd ed. Dordrecht: Kluwer Academic Publishers.

Winn, M. D. (2003). *J. Synchrotron Rad.* **10**, 23–25.

Winn, M. D., Murshudov, G. N. & Papiz, M. Z. (2003). *Methods Enzymol.* **374**, 300–321.

Woolfson, M. M. (1997). *An Introduction to X-ray Crystallography.* Cambridge University Press.

Xuong, N. H., Nielsen, C., Hamlin, R. & Anderson, D. (1985). *J. Appl. Cryst.* **18**, 342–350.

Zwart, P. H., Afonine, P. V., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., McKee, E., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., Storoni, L. C., Terwilliger, T. C. & Adams, P. D. (2008). *Methods Mol. Biol.* **426**, 419–435.