



## Resolution and dose dependence of radiation damage in biomolecular systems

Hakan Atakisi,<sup>a</sup> Lauren Conger,<sup>b</sup> David W. Moreau<sup>a</sup> and Robert E. Thorne<sup>a\*</sup><sup>a</sup>Physics Department, Cornell University, Ithaca, NY 14853, USA, and <sup>b</sup>Cornell University, Ithaca, NY 14853, USA.

\*Correspondence e-mail: ret6@cornell.edu

Received 4 March 2019

Accepted 19 June 2019

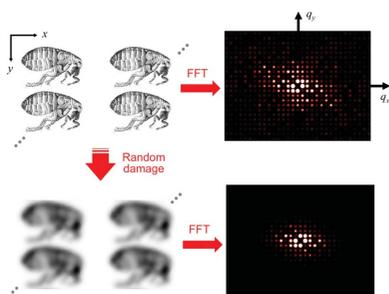
Edited by J. L. Smith, University of Michigan, USA

**Keywords:** protein crystallography; radiation damage; resolution; X-ray imaging.**Supporting information:** this article has supporting information at [www.iucrj.org](http://www.iucrj.org)

The local Fourier-space relation between diffracted intensity  $I$ , diffraction wavevector  $q$  and dose  $D$ ,  $\tilde{I}(q, D)$ , is key to probing and understanding radiation damage by X-rays and energetic particles in both diffraction and imaging experiments. The models used in protein crystallography for the last 50 years provide good fits to experimental  $I(q)$  versus nominal dose data, but have unclear physical significance. More recently, a fit to diffraction and imaging experiments suggested that the maximum tolerable dose varies as  $q^{-1}$  or linearly with resolution. Here, it is shown that crystallographic data have been strongly perturbed by the effects of spatially nonuniform crystal irradiation and diffraction during data collection. Reanalysis shows that these data are consistent with a purely exponential local dose dependence,  $\tilde{I}(q, D) = I_0(q)\exp[-D/D_c(q)]$ , where  $D_c(q) \propto q^\alpha$  with  $\alpha \simeq 1.7$ . A physics-based model for radiation damage, in which damage events occurring at random locations within a sample each cause energy deposition and blurring of the electron density within a small volume, predicts this exponential variation with dose for all  $q$  values and a decay exponent  $\alpha \simeq 2$  in two and three dimensions, roughly consistent with both diffraction and imaging experiments over more than two orders of magnitude in resolution. The  $B$ -factor model used to account for radiation damage in crystallographic scaling programs is consistent with  $\alpha = 2$ , but may not accurately capture the dose dependencies of structure factors under typical nonuniform illumination conditions. The strong  $q$  dependence of radiation-induced diffraction decays implies that the previously proposed 20–30 MGy dose limit for protein crystallography should be replaced by a resolution-dependent dose limit that, for atomic resolution data sets, will be much smaller. The results suggest that the physics underlying basic experimental trends in radiation damage at  $T \simeq 100$  K is straightforward and universal. Deviations of the local  $I(q, D)$  from strictly exponential behavior may provide mechanistic insights, especially into the radiation-damage processes responsible for the greatly increased radiation sensitivity observed at  $T \simeq 300$  K.

## 1. Introduction

Radiation damage is a key issue in all diffraction and imaging methods that illuminate biological samples with energetic particles such as X-ray photons, electrons, neutrons and positrons. Absorption and inelastic scattering processes transfer energy to the sample [quantified as dose  $D$ , the energy deposited per unit sample mass, in grays (Gy), where  $1 \text{ Gy} = 1 \text{ J kg}^{-1}$ ], with initial energy deposition from each scattering event confined to a small volume that depends on the energy transfer (Nave & Hill, 2005; Sanishvili *et al.*, 2011; Finprock *et al.*, 2013). Energetic electrons and reactive atomic and molecular species are generated, and diffuse and react, causing additional chemical, bond-scale damage (Holton, 2009; Garman, 2010). Accumulation of bond-scale damage causes degradation of sample order on larger and larger length scales (Warkentin *et al.*, 2013). In the absence of macroscopically



OPEN ACCESS

disruptive damage processes such as the eruption of hydrogen bubbles (Massover, 2007; Meents *et al.*, 2009) (for example at very large dose rates; Warkentin *et al.*, 2017) or fracturing, this accumulation of damage manifests in imaging as a loss of image contrast or ‘blurring’, evident first in short length-scale features and then moving to larger and larger scales. In diffraction, damage manifests as a decrease in the diffracted intensity at large angles  $2\theta$  or large diffraction wavevectors  $q$  that progresses to smaller and smaller  $q$ . In addition to these ‘global’ effects of radiation damage, metal sites, disulfide bonds and other structures within the sample may be particularly sensitive to damage, giving rise to ‘site-specific’ damage (Ravelli & McSweeney, 2000; Weik *et al.*, 2000; Burmeister, 2000; Banumathi *et al.*, 2004).

In biomolecular crystallography, radiation damage limits the amount of diffraction data that can be collected per unit sample volume and introduces errors in experimental structure factors. The effects of radiation damage on individual Bragg peak intensities can be complex (Warkentin *et al.*, 2017). Aside from global  $q$ -dependent intensity decay and differential decays associated with site-specific damage, measured Bragg peak intensities depend on radiation-damage-induced broadening of crystal mosaicity and strain distributions, which can cause partially illuminated reflections to initially brighten with dose (Warkentin *et al.*, 2017). They also depend on the spatial pattern of sample irradiation during data collection (Bury *et al.*, 2018), which can lead to heterogeneous sample dose states within the illuminated volume and to measured diffraction intensities that reflect a complex convolution of spatially (and, at high dose rates, temporally) nonuniform damage effects (Warkentin *et al.*, 2017). Current crystallographic processing software (Evans, 2006; Otwinowski *et al.*, 2012; Diederichs *et al.*, 2003; Diederichs, 2006) incompletely accounts for these effects. Refined structural models based on data with typical maximum resolutions of  $\sim 2$  Å are at least somewhat insensitive to these experimental errors, as models obtained from nominally zero-dose XFEL data sets show good general agreement with those from finite-dose synchrotron data sets (see, for example, Hirata *et al.*, 2014; Keedy *et al.*, 2015).

The most commonly used metric for characterizing the radiation sensitivity of biomolecular crystals is the half-dose  $D_{1/2}$ , which is equal to the dose at which the integrated intensity within all observed diffraction peaks drops to half of its zero-dose value. Typical reported half-dose values for protein crystals are  $\sim 10$ – $30$  MGy at 100 K and  $100$ – $400$  kGy at room temperature (Leal *et al.*, 2013). Half-doses depend on the initial diffraction resolution of the crystal (Howells *et al.*, 2009), which depends on the crystal quality and size, the X-ray beam size and the amount of background scatter from air and surrounding liquid. Experimental half-doses also depend on the spatial distribution of dose within the crystal, which depends on the beam intensity profile, on the size and shape of the crystal and on whether the crystal is fixed or rotated during data collection (Warkentin *et al.*, 2017; Bury *et al.*, 2018), and on how dose is defined when irradiation is spatially non-uniform (Warkentin *et al.*, 2017). Furthermore, the way in

which the half-dose weights the contribution of diffraction intensities at different  $q$  values is not obviously related to their information content or to their utility in defining and constraining the final structural model. Consequently, half-doses are at best a crude metric of radiation sensitivity and provide at best a rough rule-of-thumb limit in crystallographic data collection. Other metrics of radiation sensitivity such as the change in scaling  $B$  factor (Kmetko *et al.*, 2006; Leal *et al.*, 2013), scale (Leal *et al.*, 2013) and decay  $R$  factor (Diederichs, 2006) are used, but have related limitations.

Radiation damage has been more fully characterized using the  $q$ -dependent integrated intensity  $I(q)$ , which is obtained by integrating sample diffraction within a wavevector shell of width  $\Delta q$ , with values of diffraction resolution  $d = \lambda/2\sin(\theta) \propto 1/q$  rather than  $q$  typically quoted (Sliz *et al.*, 2003; Bourenkov & Popov, 2010; Liebschner *et al.*, 2015). Plots of experimental  $I(q)$  for a given  $q$  (or resolution) shell versus time, time-integrated flux density (in photons  $\text{cm}^{-2}$ ) or nominal dose typically show an initial linear or exponential decay and then a more gradual decay at larger times or doses, giving an overall ‘hockey-stick’ shape when plotted on semi-log axes (for example as in Blake & Phillips, 1962; Warkentin *et al.*, 2017). The rate of the initial intensity decay with dose increases with increasing  $q$  (increasing resolution, corresponding to a decreasing numerical value of  $d$ ). Kinetics-inspired models (Blake & Phillips, 1962; Hendrickson, 1976; Sygusch & Allaire, 1988) have been used for nearly 50 years to obtain good fits to measured  $I(q)$  versus exposure time or nominal dose, reproducing both the overall ‘hockey-stick’ shape at fixed  $q$  and its  $q$  or resolution dependence. However, the physical significance of the models and the fit parameters obtained have been unclear.

Recent experiments and analyses have emphasized the profound effects of nonuniform crystal irradiation during data collection on measured integrated intensity–dose curves (Warkentin *et al.*, 2017). For thaumatin and lysozyme crystals held in a fixed orientation and illuminated using an X-ray beam with a Gaussian intensity profile, the measured integrated (over all  $q$ ) intensity versus time, time-integrated flux density or nominal dose has the ‘hockey-stick’ shape. Simulations show that the Gaussian illumination profile generates this shape even if the underlying relation between diffracted intensity and dose is strictly exponential. Since previous radiation-damage experiments have seldom, if ever, provided perfectly uniform crystal irradiation (even when the X-ray beams had nominally ‘top-hat’ profiles), interpretation of their intensity–dose curves using kinetics-inspired models is now suspect.

Our goal here is to determine the underlying ‘local’ Fourier-space relationship between damage and dose,  $\tilde{I}(q, D)$ , as it would be measured from a uniformly illuminated crystal. This relationship plays an analogous role in radiation-damage studies to the local relation between conductivity and electric field,  $\sigma(E)$ , of conducting materials: it is the key to detailed understanding of damage and its mechanisms. Beginning with previously published experimental data at  $T = 100$  K, we account for spatially nonuniform illumination during data

collection and use this to estimate  $\tilde{I}(q, D)$ . We then consider a simple model for radiation damage involving random, local disordering interactions. This model predicts a purely exponential  $\tilde{I}(q, D)$  relation, and a decay constant that scales with  $q$  in a manner roughly consistent with experiment. Fits of previous models (Blake & Phillips, 1962; Hendrickson, 1976; Sygusch & Allaire, 1988) to experimental intensity versus resolution and dose data are invalidated as the primary trends have a different physical origin to that assumed. Experimental deviations from the predictions of the present model should illuminate how damage mechanisms evolve between cryogenic and room or biological temperatures.

## 2. Materials and methods

### 2.1. Modeling and fitting experimental data for intensity versus dose and resolution

In an ideal experiment to measure the local  $\tilde{I}(q, D)$  relation in a bulk crystal [where photoelectron escape from the crystal (Nave & Hill, 2005; Finprock *et al.*, 2010; Sanishvili *et al.*, 2011) can be neglected], the entire crystal volume is illuminated with a fixed and uniform flux-density X-ray beam, and the crystal thickness along the incident beam direction is small compared with the X-ray absorption length. Under these illumination conditions, every crystal region receives the same dose  $D$  regardless of whether the crystal is rotated or held in a fixed orientation.

In actual experiments (Fig. 1), the X-ray beam may be smaller than the crystal, the incident flux density within the beam may be nonuniform (even when nominally flat or ‘top-hat’ beams are used) and the crystal may be rotated during data collection. These result in different regions of the crystal experiencing different (and time-dependent) incident photon flux densities and dose rates, and accumulating different total doses  $D(x, y, z, t)$  (Diederichs, 2006; Bury *et al.*, 2018). This leads to spatially nonuniform damage. The measured

diffracted intensity at any time is determined by both the incident X-ray flux-density distribution and by the distribution of damage states within the X-ray-illuminated sample volume (Diederichs, 2006; Warkentin *et al.*, 2017) [which can be characterized using an incident flux-density-weighted dose (Zeldin *et al.*, 2013) or, more meaningfully, by a diffraction-weighted dose (Warkentin *et al.*, 2017)]. Because the local relation between diffracted intensity and dose is in general nonlinear, knowing only the measured  $I(q)$  versus exposure time or nominal fluence (in photons  $\text{mm}^{-2}$ ) or nominal dose and the spatiotemporal pattern of crystal irradiation during data collection is not sufficient to uniquely determine the local  $\tilde{I}(q, D)$  relation.

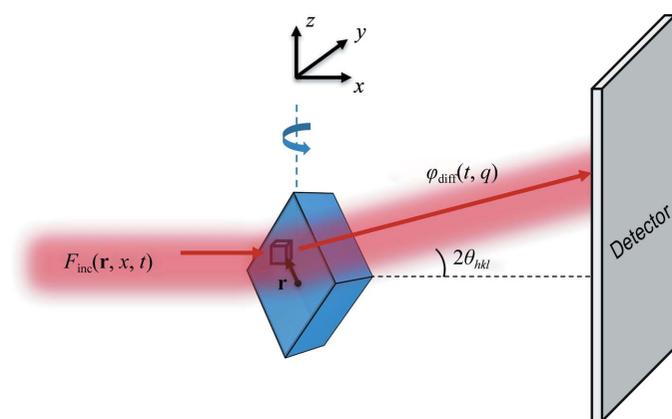
Thus, to analyze previous experimental  $I(q)$  data, we (i) calculate the spatial distribution of sample irradiation and dose from given experimental details, (ii) define a (somewhat) general expression for the local  $\tilde{I}(q, D)$  relation, with adjustable parameters, (iii) calculate the diffracted intensity  $I(q)$  versus nominal fluence or dose using this relation and the calculated dose distribution within the sample, and (iv) refine the parameters to optimize the quality of the fit to the  $I(q)$  data.

### 2.2. Calculating experimental dose distributions

To calculate the diffracted flux from a crystal after a given exposure time, we first need to know the X-ray beam size and flux-density profile, the crystal size, shape, initial orientation and location in the X-ray beam, and how the crystal is rotated during the exposure. Using this information, we can calculate the total dose (in  $\text{J kg}^{-1}$ ) delivered to each volume element (voxel) of the crystal after time  $t$ :  $D(\mathbf{r}, t)$ . This dose determines the damage state of the voxel and the diffracted flux (in photons  $\text{s}^{-1}$ ) that it will produce per unit incident flux density. The diffracted flux from the crystal at time  $t$  is then obtained by summing the product of the incident flux density at each voxel and the diffracted flux per unit flux density of that voxel.

As shown in Fig. 1, a crystal of arbitrary shape is illuminated by an X-ray beam propagating along the  $x$  direction. During irradiation/data collection, the crystal may be rotated about an axis perpendicular to the beam direction, and we define a (stationary with respect to beam and detector) coordinate system oriented as shown with  $x = 0$  located on the rotation axis and with  $y = z = 0$  at the beam center.  $F_{\text{inc}}[\mathbf{r}(t), x, t]$  is the incident X-ray photon flux density (photons  $\text{m}^{-2} \text{s}^{-1}$ ), where  $\mathbf{r}$  is a vector pointing from the coordinate system origin to the position of a voxel within the crystal at time  $t$ . The incident X-ray flux density decreases owing to scattering and absorption as it propagates through the crystal, and so in general is a function of the position  $x$  along the propagation direction.

Each voxel within the crystal is labeled with its initial ( $t = 0$ ) Cartesian coordinates  $\mathbf{r}_0$ . As the crystal is rotated about the axis during data collection, a voxel at location  $\mathbf{r}_0$  will rotate to location  $\mathbf{r}(t) = \mathbf{M}[\varphi(t)]\mathbf{r}_0$  at time  $t$ , where  $\mathbf{M}$  is a linear rotation matrix and  $\varphi(t)$  is the rotation angle about the axis. The dose that has been delivered at time  $t$  to a sample voxel initially located at  $\mathbf{r}_0$  is given by



**Figure 1**  
Experiment and simulation setup. A crystal (simulated as a rectangular prism) is located at  $x, y, z = 0$  and a volume element (voxel) of the crystal is located at position  $\mathbf{r}$ . The crystal is illuminated by an X-ray beam with a top-hat or Gaussian profile, and the incident photon flux at the voxel position is  $F_{\text{inc}}(\mathbf{r}, x, t)$ . Crystal planes with spacing  $d_{hkl}$  generate Bragg scattering at angle  $2\theta_{hkl}$ , corresponding to a scattering wavevector  $q_{hkl} = 2\pi/d_{hkl}$ . The scattered flux (in photons  $\text{s}^{-1}$ ) from each voxel is  $\varphi_{\text{diff}}(t, q)$ .

$$D(\mathbf{r}_0, t) = k \int_0^t F_{\text{inc}}\{\mathbf{M}[\varphi(t)]\mathbf{r}_0, x, t\} dt, \quad (1)$$

where the constant  $k$  depends on the X-ray energy and on the atomic composition and density of the crystal, and can be calculated using *RADDOSE-3D* (Bury *et al.*, 2018) or standard tables (Hubbell & Seltzer, 2004). We assume that the X-ray beam flux density is time-independent and that the crystal is thin compared with the X-ray attenuation length [a reasonable approximation when using 10–15 keV X-rays and 50–300  $\mu\text{m}$  crystals; at 10 keV the beam is attenuated by an average of 17% on passing through 300  $\mu\text{m}$  crystals of the proteins (lysozyme, thaumatin, apoferritin, HLA,  $\lambda 3$  and US2) of relevance here]. The incident flux density  $F_{\text{inc}}(\mathbf{r}, x, t)$  can then be written as  $F_{\text{inc}}(\rho)$ , where  $\rho = (y^2 + z^2)^{1/2}$  is the radial distance from the beam center.

### 2.3. Calculating diffracted intensities

The crystal diffraction pattern measured on the detector consists of a large number of bright ‘spots’ corresponding to Bragg diffraction at angles  $(2\theta)_{hkl}$  for which the scattering wavevector satisfies  $q_{hkl} = 2\pi/d_{hkl}$ , where  $d_{hkl}$  is the spacing of a diffracting crystal lattice plane. Integrating the detector photon counts about the incident beam direction in radial bins of fixed width  $dq$  (variable width in detector coordinates) and dividing by the detector exposure time then gives a photon flux  $\varphi_{\text{diff}}(q, t)$ . At  $t = 0$  the decay of  $\varphi_{\text{diff}}(q, 0)$  with increasing  $q$  owing to thermal atomic motions and to static crystal disorder can be approximated by a Debye–Waller factor  $DWF(B, q)$ .

At time  $t > 0$ , the diffraction at scattering wavevector  $q$  from each crystal voxel will be proportional to the incident flux density  $F_{\text{inc}}(\mathbf{r}, x, t)$  at the position of the voxel and to the diffracting power of the voxel (the number of diffracted photons per incident photon), which will depend on the dose received by the voxel from  $t = 0$  to time  $t$ . Let  $\Gamma[D(\mathbf{r}_0, t), q]$  be the factor by which the diffracting power of a voxel at initial position  $\mathbf{r}_0$  is reduced by radiation damage. The total diffracted flux from all crystal voxels at wavevector  $q$  per unit volume of reciprocal space,  $\varphi_{\text{diff}}(q, t)$ , is then

$$\varphi_{\text{diff}}(q, t) \propto \int_{\text{crystal volume}} F_{\text{inc}}\{\mathbf{M}[\varphi(t)]\mathbf{r}_0\} \times DWF(B, q) \times \Gamma[D(\mathbf{r}_0, t), q] dV_0. \quad (2)$$

We assume that the diffracting power at a given  $q$  decays exponentially with dose,

$$\Gamma[D(\mathbf{r}_0, t), q] \propto \exp[-D(\mathbf{r}_0, t)/D_c(q)], \quad (3)$$

where  $D_c(q)$  determines the decay rate. Experimentally, the scattered intensity decays more rapidly as  $q$  and  $2\theta$  increase, corresponding to a more rapid decay of short-wavelength Fourier components of the unit cell’s electron density. Based on a fit to experimental data from crystallography and X-ray imaging, Howells *et al.* (2009) suggested that  $D_c(q) \propto 1/q$  holds and that the resolution-dependent half-dose is given by  $D_{1/2}(d) \simeq d \times 10 \text{ MGy } \text{\AA}^{-1}$ . We assume a more general relation

$$D_c(q) = K/q^\alpha, \quad (4)$$

and then fit experimental data from crystallography to determine the exponent  $\alpha$ .

The total diffracted photon flux of the crystal  $\Phi_{\text{diff}}$  is obtained by integrating (3) over all  $q$  as

$$\Phi_{\text{diff}}(t) \propto \int_{q_{\text{min}}}^{q_{\text{max}}} \varphi_{\text{diff}}(t, q) dq. \quad (5)$$

### 2.4. Fitting experimental data

Fitting of reported data for diffracted intensity versus resolution  $d$  and nominal dose/fluence from prior experiments (Sliz *et al.*, 2003; Bourenkov & Popov, 2010; Liebschner *et al.*, 2015) was performed as follows, using the parameters given in Supplementary Tables S1 and S2.

We used either the actual reported beam profile (measured by scanning a slit across the beam) or else a profile matching the stated shape and width parameters (Supplementary Table S1). A ‘top-hat’ profile was represented as  $F_{\text{inc}}(\mathbf{r}) = F_0$  for  $\rho < \rho_{\text{max}}$  (or  $|y| < y_{\text{max}}$ ,  $|z| < z_{\text{max}}$ ) and 0 otherwise, and a Gaussian profile by

$$F_{\text{inc}}(\mathbf{r}) = F_0 \exp\left[-\left(\frac{y^2}{2\sigma_y^2} + \frac{z^2}{2\sigma_z^2}\right)\right]. \quad (6)$$

Each crystal was divided into cubic voxels, with roughly 100 voxels in each dimension. Crystals were assumed to be rectangular prisms with reported dimensions (Supplementary Table S1). Prisms were oriented as reported, and if no orientation was specified the orientation was adjusted to obtain the best fit to the intensity data (Section S1, supporting information). As a check, calculations were also performed assuming cylindrical crystals with their axes corresponding to the rotation axis.

The reported experiments either repeatedly oscillated the crystal through a small angle (*e.g.*  $2^\circ$ ), collecting one or more frames for each oscillation and returning to the starting orientation before collecting the next set of frames, or else continuously rotated the crystal during data collection (*e.g.* by  $60^\circ$  with  $1^\circ$  rotation per diffraction frame). The former method gives a more uniform dose distribution in the irradiated and diffracting crystal regions. For simulations with repeated oscillations through an angular wedge, the simulation time step was set to 1/1000 of the total exposure time for the entire set of oscillations, the crystal was rotated 1/10 of its total oscillation in the wedge (*e.g.* by  $0.5^\circ$  for a  $5^\circ$  oscillation) in each step, and after every ten steps the rotation angle was reset to the starting angle of the wedge. For simulations with continuous rotations, the crystal was rotated by 1/100 of its maximum rotation in each simulation step. The diffracted flux versus  $q$ ,  $\varphi_{\text{diff}}(t_n, q)$ , the total diffracted flux  $\varphi_{\text{diff}}(t_n)$  and a nominal dose (an average over the voxels that have received nonzero dose) corresponding to the reported dose were calculated at each step. Supplementary Fig. S2 shows example dose distributions for a cylindrical crystal held in a fixed

orientation and when rotated, when illuminated with flat-top and isotropic Gaussian beams.

Reported Bragg intensity data were integrated within resolution shells bounded by upper and lower  $d$  values. Reported incident beam profiles and diffracted intensity versus resolution shell and fluence/dose plots were digitized using the software *WebPlotDigitizer* (A. Rohatgi; <https://automeris.io/WebPlotDigitizer/>). Diffracted intensities in each shell were normalized by their zero-dose extrapolation, eliminating both the Debye–Waller factor  $DWF(B, q)$  in (2) as well as the Lorentz–polarization correction to the measured intensities, which was not consistently applied in all studies. The resulting normalized plots only show variations with resolution shell ( $q$ ) and fluence or dose owing to radiation damage.

### 2.5. A simple physics-based model for radiation damage

The primary model used to analyze global radiation damage to protein crystals for the last 50 years is due to Blake & Phillips (1962) and Hendrickson (1976), with an additional extension proposed later (Sygusch & Allaire, 1988). As shown in Supplementary Fig. S1, an undamaged crystal becomes disordered at ‘rate’  $k_1$  (proportional to the volume fraction of crystal disordered per unit dose). This damaged crystal continues to exhibit Bragg diffraction, but its intensities decrease with increasing  $q$  as  $I(q) = I_0 \exp(-B_{\text{disorder}} q^2)$ , where  $B_{\text{disorder}}$  is a (fixed) average  $B$ -factor increase in the disordered regions. Disordered crystal becomes completely amorphous and ceases to generate Bragg diffraction [ $I(q) = 0$ ] at ‘rate’  $k_2$ . Undamaged crystal can also proceed directly to the amorphous state at rate  $k_3$ . The resulting diffracted intensity versus  $q$  and dose  $D$  is given by

$$\frac{I(q, D)}{I(q, 0)} = \exp[-(k_1 + k_3)D] + \frac{k_1}{k_1 + k_3 - k_2} \times \exp(-k_2 D) \{1 - \exp[-(k_1 + k_3 - k_2)D]\} \times \exp(-B_{\text{disorder}} q^2). \quad (7)$$

With four adjustable parameters, this model has provided good fits to reported  $I(q)$  versus nominal dose data, including to the ‘hockey-stick’ dose dependence evident for data at larger  $q$  or in higher (smaller numeric) resolution shells (Blake & Phillips, 1962; Hendrickson, 1976; Sliz *et al.*, 2003; Bour-enkov & Popov, 2010; Warkentin & Thorne, 2010; Liebschner *et al.*, 2015; Owen *et al.*, 2014; Warkentin *et al.*, 2017). However, the connection of this model to underlying damage processes is opaque and the physical significance of the obtained fit parameters is unclear.

We thus considered a very simple model that captures some essential physical aspects of radiation damage. Incident X-ray photons interact with the sample at random times and locations. Each interaction ejects a photoelectron, which then generates lower energy secondary electrons within a volume (determined by the photoelectron mean free path) of a few micrometres. These secondary electrons then diffuse, break bonds and create free radicals that can diffuse and cause additional damage. Experiments using X-ray microbeams

(Sanishvili *et al.*, 2011; Finrock *et al.*, 2013) show that the resulting perturbations to electron densities from those of the original, undamaged crystal, as reflected in the degradation of the diffraction properties, are confined to within a length comparable in magnitude to the photoelectron mean free path, even at room temperature (Warkentin *et al.*, 2017) where diffusion occurs freely.

To model damage, we thus assume that X-ray photon interactions occur at random locations in the crystal. We model the effect of each interaction as a modest ‘blurring’ of the real-space electron density within a finite region around each interaction point. The number of such interactions per unit crystal volume determines the dose  $D$ . We calculate the Fourier transform of the electron density of the crystal and evaluate the integrated Bragg intensity within  $q$  (or resolution) shells and plot this versus dose  $D$ . Simulations were performed in two dimensions, using  $m \times m$  arrays of unit cells containing a grayscale image of a flea (Fig. 5), and in three dimensions, using unit cells obtained by discretizing the protein in PDB entry 3e4h (Wang *et al.*, 2009): tetragonal crystals of the 29-residue plant protein cyclotide varv F at 1.8 Å resolution (Supplementary Fig. S6).

In two dimensions, for each X-ray hit a Gaussian spatial filter was applied to a small  $n \times n$  pixel interaction region centered at a randomly chosen crystal location  $(x_i, y_i)$  as shown in Supplementary Fig. S7. FFTs of the crystal were periodically calculated as hits accumulated. The simulations were continued until the random hits caused the diffraction peaks in the highest resolution shell to fall below the background level, which corresponded to roughly 5–10 hits per pixel. A similar procedure was followed for the computationally more intensive three-dimensional simulations. Details of the simulations are given in Section S2 of the supporting information.

## 3. Results

### 3.1. Fitting experimental intensity versus dose data

Experimental data for intensity in resolution shells versus dose at  $T \simeq 100$  K from three previous studies were analyzed, modeled and fitted using the approach in Sections 2.1–2.5. Supplementary Tables S1 and S2 give the experimental details for each reference and our model parameters. We focused on cryogenic temperature data because crystals of different proteins are comparably radiation sensitive, damage is independent of dose rate and because free-radical diffusion and relaxation of protein and lattice structure following bond-scale damage are strongly constrained by the frozen solvent matrix, so that the overall behavior should be simpler and more consistent between protein crystals than at 300 K.

**3.1.1. Liebschner *et al.* (2015).** Liebschner *et al.* (2015) reported the most optimally executed experiments of those examined here. Data were collected from thaumatin crystals at 100 K by repeatedly oscillating the crystals through the same  $2^\circ$  range. Their measured beam profile (their Fig. 1, reproduced here as Supplementary Fig. S3) was nominally flat-topped but had significant tails, and the full widths at half

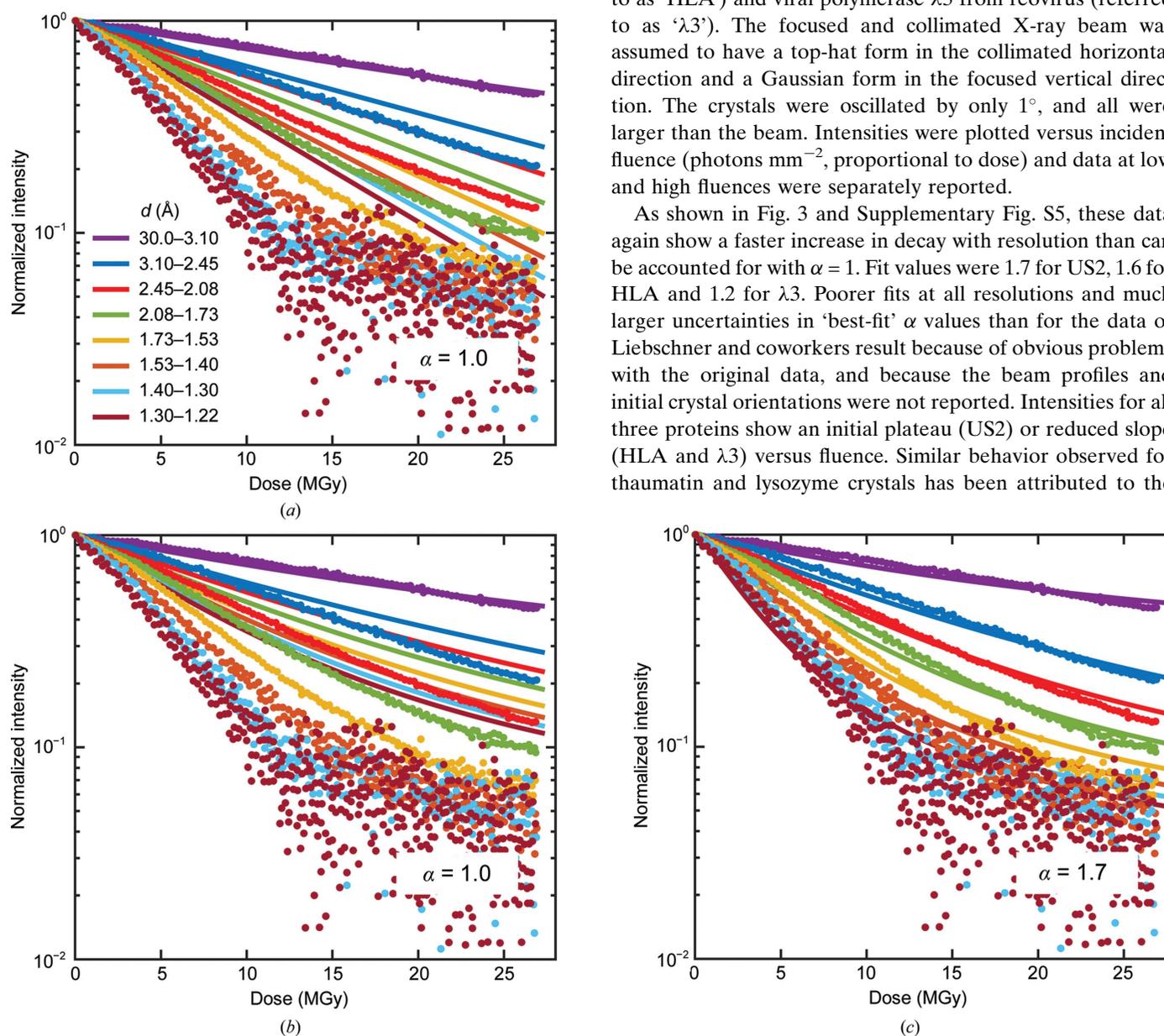
maximum (FWHMs) were much smaller than the crystal dimensions. Fig. 2 shows their data for normalized integrated intensity in resolution shells versus nominal dose. As the resolution of the shell increases, the initial decay rate with dose becomes more rapid, and deviations above exponential behavior become evident at smaller doses.

The solid lines in Fig. 2(a) show the calculated intensities assuming a top-hat incident beam profile, a  $2^\circ$  oscillation and an exponent  $\alpha = 1$  for the  $q$  (resolution) dependence of the diffracted intensity decay with dose in (4). The calculated dose variation is nearly perfectly exponential and thus does not capture the large deviations from exponential behavior at

higher resolutions and doses. By using the measured beam profile [Fig. 2(b)], the non-exponential behavior at higher resolutions is qualitatively reproduced. However, in both Figs. 2(a) and 2(b) the choice of  $\alpha = 1$ , motivated by the results of Howells *et al.* (2009), seriously underestimates the observed increase in decay rate with increasing resolution. As shown in Fig. 2(c), relaxing this constraint yields a best-fit value of  $\alpha \simeq 1.7$  and agreement with the data that is generally excellent in all resolution shells.

**3.1.2. Sliz *et al.* (2003).** Sliz *et al.* (2003) collected data at 100 K from crystals of three different proteins: the ternary US2–HLA–A2–Tax peptide complex (referred to as ‘US2’), HLA–A2 with a bound melanoma decamer peptide (referred to as ‘HLA’) and viral polymerase  $\lambda 3$  from reovirus (referred to as ‘ $\lambda 3$ ’). The focused and collimated X-ray beam was assumed to have a top-hat form in the collimated horizontal direction and a Gaussian form in the focused vertical direction. The crystals were oscillated by only  $1^\circ$ , and all were larger than the beam. Intensities were plotted versus incident fluence (photons  $\text{mm}^{-2}$ , proportional to dose) and data at low and high fluences were separately reported.

As shown in Fig. 3 and Supplementary Fig. S5, these data again show a faster increase in decay with resolution than can be accounted for with  $\alpha = 1$ . Fit values were 1.7 for US2, 1.6 for HLA and 1.2 for  $\lambda 3$ . Poorer fits at all resolutions and much larger uncertainties in ‘best-fit’  $\alpha$  values than for the data of Liebschner and coworkers result because of obvious problems with the original data, and because the beam profiles and initial crystal orientations were not reported. Intensities for all three proteins show an initial plateau (US2) or reduced slope (HLA and  $\lambda 3$ ) versus fluence. Similar behavior observed for thaumatin and lysozyme crystals has been attributed to the



**Figure 2** Experimental data (solid circles) for integrated intensity in resolution shells versus dose for thaumatin crystals at 100 K, as measured by Liebschner *et al.* (2015). Absolute intensities in each resolution shell (Fig. 4 in the original manuscript) have been normalized by the first (approximately zero-dose) intensity point; non-normalized data are shown in Supplementary Fig. S4. The solid lines indicate results from simulations assuming (a) a perfect top-hat incident X-ray beam profile and an exponent  $\alpha = 1$  in (4), (b) the measured beam profile (Fig. 1 in the original manuscript, reproduced as Supplementary Fig. S3) and  $\alpha = 1$ , and (c) the measured beam profile and a best-fit exponent  $\alpha = 1.7$ .

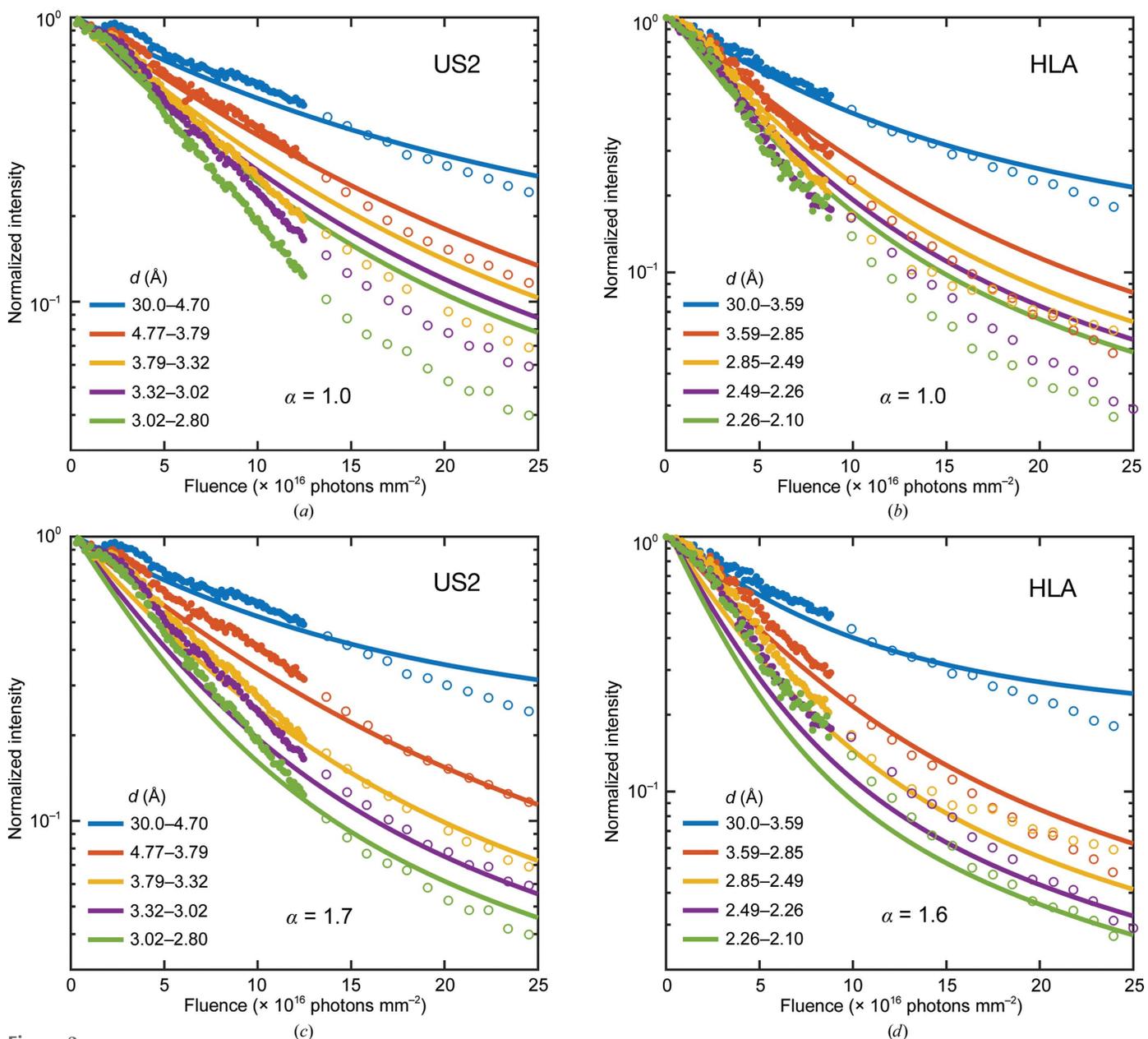
effects of dose-dependent mosaicity broadening and cell expansion (Warkentin *et al.*, 2017), which is not accounted for by scaling algorithms or by our modeling.

**3.1.3. Bourenkov & Popov (2010).** Bourenkov & Popov (2010) collected data at 100 K from crystals of insulin, P19-siRNA, FAE and FtsH. Crystals of the first three were rotated during exposure by a total angle of between 35 and 300°, with 0.5–1° rotation per frame. Large rotations are not ideal for our modeling because the dose distribution within and diffraction from the illuminated volume will have a larger dependence on the detailed crystal shape and initial orientation than when crystals are oscillated through a small angle. Crystals of FtsH were both rotated and translated (after each 30°, by an

unknown amount) perpendicular to the beam direction, and their data were not modeled. Crystals were illuminated by a nominally Gaussian beam with dimensions that were equal to or smaller than the largest crystal dimension. Beam profiles and initial crystal orientations were not reported. Despite many uncertainties the model calculations yield good fits to the data for all three crystals (Fig. 4), with best-fit exponents of  $\alpha = 1.4, 1.6$  and  $1.8$  for insulin, P19 and FAE, respectively.

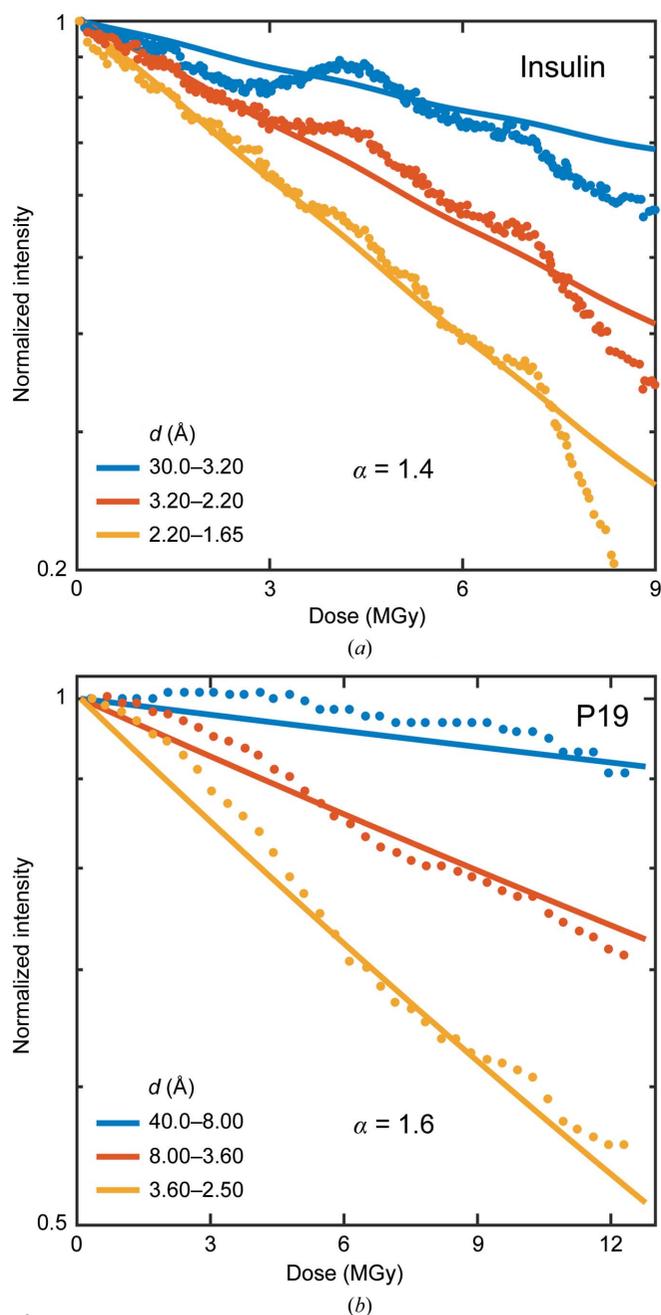
**3.2. Simulations of radiation damage**

Figs. 5 and 6 show the results of simulations of our model for radiation damage as a sequence of random, local Gaussian



**Figure 3** Experimental data in the low-dose (solid circles) and high-dose (open circles) regions for integrated intensity in resolution shells versus incident fluence (in photons mm<sup>-2</sup>, proportional to dose) reported by Sliz *et al.* (2003) (Fig. 1 in the original manuscript) for crystals of US2 and HLA at 100 K. The solid lines indicate results from simulations assuming a top-hat incident beam profile in the horizontal direction and a Gaussian profile in the vertical direction (based on descriptions of the experimental setup), with (a), (b)  $\alpha = 1$  and (c), (d) ‘best-fit’ values chosen based on visual comparison.

blurs and corresponding to a condition of spatially uniform irradiation. Fig. 5(a) shows four unit cells of an initial, undamaged  $16 \times 16$  cell two-dimensional crystal. Its diffraction (proportional to the square of the FFT amplitudes) has strong peaks extending out to the maximum  $q$  or resolution of the initial image. After some large number of hits, the real-space density is blurred throughout the crystal and its diffraction decays much more rapidly with  $q$ . Movies of the evolution of the electron density of the crystal and its diffraction are provided as Supplementary Movies S1 and S2.



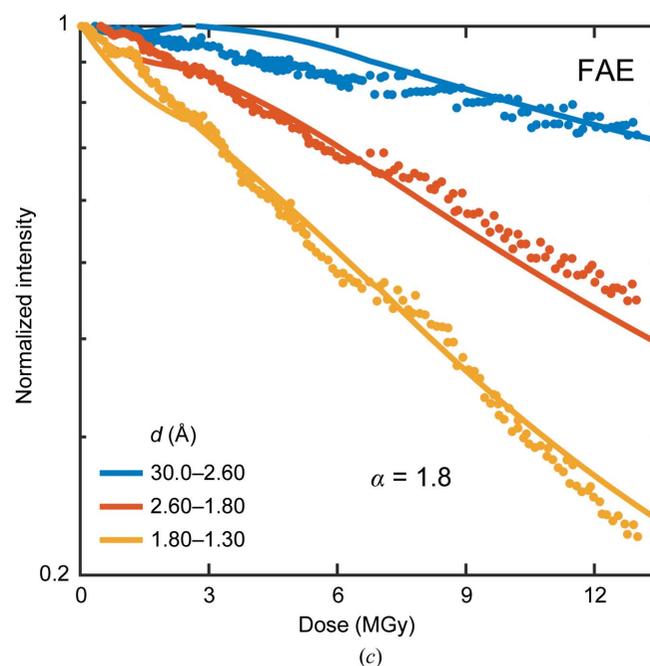
**Figure 4**  
Experimental data (solid circles) for integrated intensity in resolution shells versus dose for crystals of insulin, P19 and FAE at 100 K, as measured by Bourenkov & Popov (2010) (Figs. 3, 4 and 6 in the original paper). Crystals were rotated continuously during irradiation as for crystallographic data collection. Oscillations in the data may be owing to irregular crystal shapes and large rotations that produce complex dose distributions within the X-ray-illuminated volume. The solid lines indicate the results from simulations with best-fit exponents  $\alpha$  of (a) 1.4, (b) 1.6 and (c) 1.8.

Figs. 6(a) and 6(c) give results in two and three dimensions, respectively, for the predicted diffracted intensity in a resolution shell versus dose. In both two and three dimensions, the diffracted intensity within a resolution shell has a strictly exponential decay with dose, consistent with our assumption for the behavior of the local  $\tilde{I}(q, D)$  in fitting the data in Figs. 2, 3 and 4. Figs. 6(b) and 6(d) give results in two and three dimensions, respectively, for the half-dose in a given resolution shell versus resolution, determined from plots as in Figs. 6(a) and 6(c) as the dose at which the intensity in a given resolution shell drops to half of its initial value. Except at the highest (lowest numerical value) resolutions, in both two and three dimensions the half-dose varies with resolution  $d$  approximately as  $D_{1/2}(d) \propto d^\alpha$  with an apparent low-resolution asymptote of  $\alpha \simeq 2$ ; best-fit values to the near-linear regions in the two- and three-dimensional results are  $\alpha \simeq 1.96$  and 1.86, respectively. These results do not change when the Gaussian blur is replaced by a uniform blur (Supplementary Fig. S8). Deviations from simple power-law behavior at the highest (lowest numerical value) resolutions depend on the Gaussian width, with larger widths causing deviations at lower resolutions. Similar results were obtained using other two-dimensional images and using other PDB entries as the basis for the three-dimensional unit cell.

## 4. Discussion

### 4.1. The local $\tilde{I}(q, D)$ relation: dose dependence

By assuming a purely exponential local dependence of diffracted intensity on dose of the form  $\tilde{I}(q, D) = I_0(q) \exp[-D/D_c(q)]$  with  $D_c(q) = K/q^\alpha$ , and accounting for



the nonuniform pattern of crystal irradiation during data collection, we obtain good fits to experimental  $I(q)$  versus nominal dose/fluence relations measured for several protein crystals at  $T = 100$  K under diverse data-collection conditions. Deviations of the calculated dose/fluence dependence from the data may arise because the actual crystal shapes and initial crystal orientations (which were not given or adequately described) deviate from those assumed, and because of issues in data collection and processing that cause measured intensities to deviate from the actual dose-dependent structure factors (Warkentin *et al.*, 2017). Consequently, based on the available data at 100 K, there is no reason to believe that the local  $\tilde{I}(q, D)$  relation at  $T \simeq 100$  K is anything but purely exponential in all resolution shells over the resolution and dose range relevant in biomolecular crystallography.

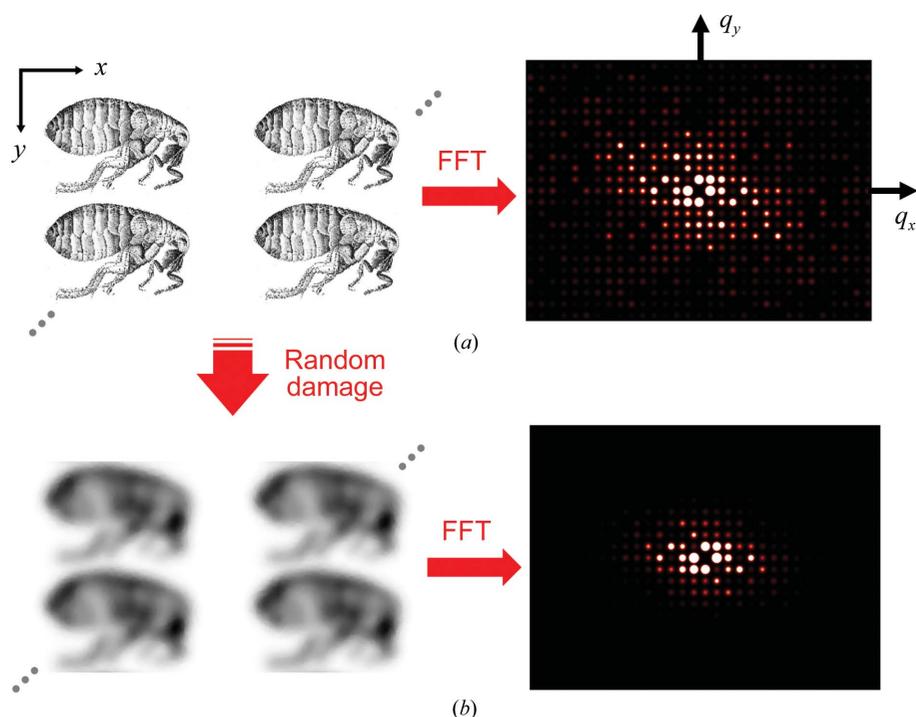
The present analysis also shows that even relatively small deviations of the profile of an X-ray beam from an ‘ideal’ top-hat form can have a substantial effect on the dose dependence of the intensity at larger doses and higher resolutions. This is particularly evident in the fits to the data of Liebschner and coworkers in Fig. 2; using the actual profile of the nominally top-hat beam dramatically improves the fit quality.

#### 4.2. The local $\tilde{I}(q, D)$ relation: $q$ dependence

Howells *et al.* (2009) presented a summary of available data for resolution-dependent maximum tolerable doses, obtained from published half-dose values in biomolecular crystallography and from X-ray and electron imaging studies. These results, spanning resolutions from  $\sim 2$  to 700 Å, show large scatter but are roughly consistent with a linear resolution dependence corresponding to  $D_e(q) = K/q^\alpha$  with  $\alpha = 1$ .

The present analysis shows that data for protein crystallography with resolutions between  $\sim 1$  and 10 Å are unambiguously inconsistent with  $\alpha = 1$ , and yield best-fit values of between  $\sim 1.4$  and 1.8, with the most ‘ideal’ data of Liebschner and coworkers yielding a value of  $\sim 1.7$ .

Fig. 7 summarizes the results for half-dose versus resolution at 100 K, deduced from previous  $I(q)$  measurements using the methods of Sections 2.1–2.5 (solid symbols) or as originally reported (open symbols). These half-dose values are for diffraction within resolution shells, rather than overall half-dose values obtained by integrating the entire diffraction pattern over all resolutions. The data between 1 and 10 Å are well described by  $\alpha \simeq 2$ , with a best-fit value of 1.86; the only



**Figure 5** Simulation of radiation damage to a two-dimensional crystal formed of images of a flea (from Robert Hooke’s *Micrographia*). Each ‘hit’ corresponds to the application of a local Gaussian blur at a randomly selected location (Supplementary Fig. S7). (a) Four unit cells (each 1024 × 512 pixels) of the undamaged two-dimensional crystal (left) and the square of the FFT amplitude, proportional to the diffracted intensity, of a 16 × 16 cell crystal (right). (b) After receiving a large number of hits ( $\gg 1$  per pixel) the electron density has been blurred and the high-resolution (large  $q$ ) diffraction peaks have faded out. A full video of the evolution of the crystal and its diffraction pattern with dose is given in Supplementary Movies S1 and S2.

data that appreciably deviate from this fit are those first reported by Howells and coworkers. Extrapolating the  $\alpha = 1.86$  fit from 1 to 600 Å yields a half-dose of  $\sim 3 \times 10^{11}$  Gy and using  $\alpha = 2$  gives  $\sim 7 \times 10^{11}$  Gy. This compares with a reported overall half-dose value (obtained by integrating over all resolutions) in X-ray imaging of cells to this resolution of  $5 \times 10^{11}$  Gy (Maser *et al.*, 2000) and with a value of only  $\sim 1 \times 10^{10}$  Gy based on the best fit in Howells and coworkers (their Fig. 3) with  $\alpha = 1$ .

#### 4.3. The local $\tilde{I}(q, D)$ relation: connection to ‘kinetic’ models

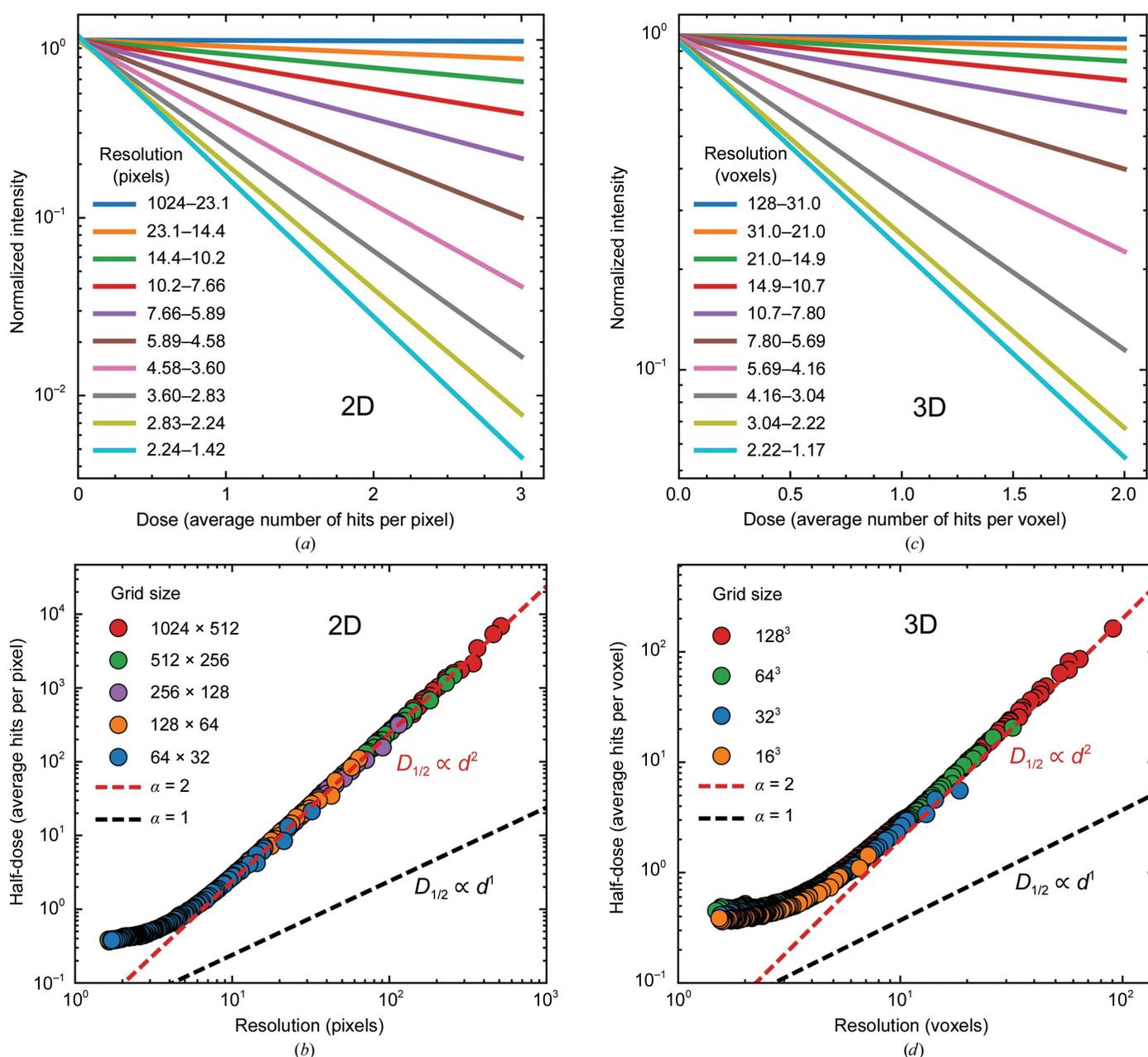
The ‘kinetic’ models of Blake & Phillips (1962), Hendrickson (1976) and Sygusch & Allaire (1988) all implicitly assume uniform sample irradiation, and so should not have been used to fit experimental data that were collected under conditions of substantially nonuniform illumination. The present analysis shows that the local  $\tilde{I}(q, D)$  is consistent with a purely exponential dose dependence for all  $q$  at 100 K and that the dose scale for intensity at a given  $q$  varies as a power of  $q$ . Equation (7) cannot replicate these features with any sensible parameter choices. The deviations from exponential behavior that these models have proved so successful at fitting are owing to nonuniform sample irradiation (and possibly also to data-processing errors), which these models

do not include. These models are thus inconsistent with experiment at  $T = 100$  K and should no longer be used.

#### 4.4. The ‘dose limit’ in biomolecular crystallography

Based upon experience in cryoelectron microscopy, Henderson (1990) suggested that the maximum tolerable dose in X-ray cryocrystallography, beyond which diffraction would be seriously degraded, would be roughly 20 MGy (Henderson, 1990). Teng & Moffat (2000, 2002), using perhaps the most nearly ideal irradiation conditions to date – a beam with a  $2\sigma$

width much larger than their crystal size (250  $\mu\text{m}$  versus 100  $\mu\text{m}$ ), giving nearly uniform illumination of the entire crystal volume – obtained a  $T = 100$  K half-dose of  $\sim 17$  MGy for lysozyme crystals diffracting to 1.6  $\text{\AA}$  resolution. Based on diffraction statistics they suggested a dose limit of  $\sim 10$  MGy be used in macromolecular crystallography. Burmeister (2000) obtained a  $T = 100$  K half-dose of  $\sim 21$  MGy for myrosinase crystals diffracting to 2.0  $\text{\AA}$  resolution. Owen *et al.* (2006) obtained half doses of 40 and 48 MGy for holoferritin and apoferritin crystals diffracting to  $\sim 2.3$   $\text{\AA}$ , and based on examination of diffraction statistics and electron-density maps



**Figure 6** Simulation results for the random Gaussian blur model of radiation damage applied to (a, b) a two-dimensional crystal formed of pixelated images of a flea and (c, d) a three-dimensional crystal formed of a pixelated electron-density map for PDB entry 3e4h. In both two and three dimensions, the intensity within a given resolution shell decays exponentially with dose, and the half-dose varies with resolution  $d$  approximately as  $D_{1/2}(d) \propto d^\alpha$ , with  $\alpha$  approaching 2 at low resolutions (large numeric values of resolution). The best-fit values in the near-linear region are  $\alpha = 1.95$  and  $\alpha = 1.86$  in (c) and (d), respectively. Simulation details are given in the supporting information.

suggested that a maximum dose of  $\sim 30$  MGy be used. Liebschner *et al.* (2015) using a nearly flat-top beam, obtained a half-dose of 18.5 MGy for thaumatin crystals diffracting to 2.1 Å. Warkentin *et al.* (2017), using a Gaussian microbeam with no crystal oscillation and correcting for the effects of the Gaussian beam profile, obtained half-doses of 10 MGy for lysozyme crystals diffracting to 1.4 Å and 13 MGy for thaumatin crystals diffracting to 1.6 Å. Many other half-dose measurements have been reported but have generally involved substantially nonuniform crystal irradiation, which can make measured half-doses substantially larger than the true, local half-dose that would be measured under conditions of purely uniform irradiation (Warkentin *et al.*, 2017).

Our conclusion based these previous studies is that diffraction half-doses at  $T = 100$  K for crystals diffracting to  $\sim 1.5$ – $2$  Å are  $\sim 15$ – $20$  MGy, and that the  $\sim 10$  MGy dose limit suggested by Teng and Moffat is appropriate. However, the 30 MGy ‘Garman limit’ reported by Owen *et al.*, rather than the 10 MGy limit of Teng and Moffat, has been by far the most widely cited, and has become the accepted standard dose limit. It far exceeds the dose that would be required to severely degrade diffraction in the highest resolution shells of, for example, the overwhelming majority of PDB entries, which have a median refined resolution of  $\sim 2.0$  Å.

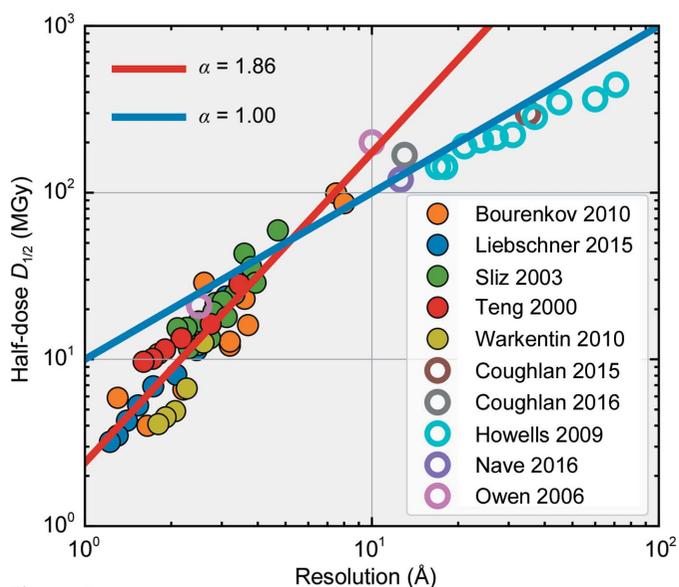
As shown in Fig. 7, using the *local* half-dose in a given resolution shell as a more meaningful and robust metric, we find that the dose limit at  $T = 100$  K increases from  $\sim 2$ – $3$  MGy at 1 Å to 8 MGy at 2 Å, 16 MGy at 3 Å and 30 MGy at 4 Å. Half-doses obtained by integrating over all resolutions up to the maximum available resolution are somewhat larger than but track these values; for lysozyme crystals diffracting to a

maximum resolution of 1.4 Å the half-dose is  $\sim 10$  MGy (Warkentin *et al.*, 2017). These resolution-dependent dose limits should be used as rules of thumb in place of the previous 20 or 30 MGy limit when determining exposure strategies in crystallographic measurements.

Why were the half-doses reported by Owen *et al.* so much larger than were obtained in the other measurements? As noted previously (Warkentin *et al.*, 2014), the data sets analysed had a resolution limit of 2.3 Å, somewhat lower than those used in other studies; this can account for roughly half the difference with half-dose values measured by Teng and Moffat (Teng & Moffat, 2000) and Warkentin *et al.* (Warkentin *et al.*, 2017). Owen *et al.* used a  $100 \times 100 \mu\text{m}$  X-ray beam, and stated that the beam profile on the beamline used for their studies ‘has been determined to be a top-hat shape’, citing Arzt *et al.* (Arzt *et al.*, 2005) for the profile. However, the beamline profiles reported in Fig. 4 of Arzt *et al.*, are not top hat (normally produced through collimation or slitting of a defocused beam). They are standard focused profiles. With tight focusing to  $31 (V) \times 47 (H) \mu\text{m}$  [Arzt *et al.*, 2005; Fig. 4(a)], the beam profile was roughly Gaussian in the vertical and somewhat flattened in the horizontal; with a  $100 \times 100 \mu\text{m}$  FWHM spot size as used by Owen *et al.*, the profile may have been more nearly Gaussian in both horizontal and vertical [Arzt *et al.*, 2005; Fig. 4(b)]. Since the beam size was much smaller than the  $\sim 200 \mu\text{m}$  of the holo- and apoferritin crystals examined, the crystals may thus have been nonuniformly irradiated, and this may have increased the apparent half-dose relative to the true, local half-dose by a factor close to two.

Which experimental dose should be compared with these limits, for a crystal diffracting to a given maximum resolution? In nearly all crystallographic data collection the crystal is nonuniformly irradiated owing to nonuniform flux density in the beam and owing to crystal rotation. Using *RADDOSE-3D* or simple code (written in, for example, *MATLAB*) based on (1), and knowing the incident X-ray flux-density profile, the approximate crystal dimensions and the crystal rotation or oscillation pattern, the dose distribution within a crystal during a given data collection can easily be calculated. *RADDOSE-3D* currently calculates the maximum dose (at any position) received within the X-ray-illuminated crystal volume, the average dose within the illuminated volume, the average dose within the crystal and an incident flux-density-weighted dose (Zeldin *et al.*, 2013). The most conservative choice is to use the maximum dose. The average dose within the illuminated volume and within the crystal can both yield problematically small dose estimates when the crystal is larger than the beam and when the beam has a non-top-hat (*e.g.* Gaussian) profile.

A more robust measure of average dose is the diffraction-weighted dose (Warkentin *et al.*, 2017), which weights the dose received at each location after an exposure time  $t$  by its contribution to the measured diffraction at time  $t$ , and thus appropriately downweights contributions from regions that, either owing to weak incident illumination or owing to radiation damage, contribute little to the measured diffraction. The diffraction-weighted dose is given by



**Figure 7** Experimental half-doses versus resolution obtained from several previous crystallographic studies (Bourenkov & Popov, 2010; Liebschner *et al.*, 2015; Sliz *et al.*, 2003; Teng & Moffat, 2000; Warkentin & Thorne, 2010; Coughlan *et al.*, 2015, 2016; Howells *et al.*, 2009; Nave *et al.*, 2016; Owen *et al.*, 2006), analyzed accounting for their dose distribution (solid circles) or used as reported (open circles). Aside from the data of Howells *et al.* (2009) at resolution beyond 10 Å, the overall trend is in good agreement with  $\alpha = 2$ , with a best-fit exponent of 1.86.

$$\text{DWD}(t) = \frac{\int_0^t \int_{\text{crystal}} D(\mathbf{r}, t') S(\mathbf{r}, t') F_{\text{inc}}(\mathbf{r}, t') d^3r \cdot dt'}{\int_0^t \int_{\text{crystal}} S(\mathbf{r}, t') F_{\text{inc}}(\mathbf{r}, t') d^3r \cdot dt'}. \quad (8)$$

Here,  $S(\mathbf{r}, t) = S[D(\mathbf{r}, t)]$  is the diffracted flux (in all reflections) per unit illuminated crystal volume per unit incident flux density at position  $\mathbf{r}$  and time  $t$ , and the denominator gives the total number of scattered photons up to time  $t$  (not the total diffracted intensity as stated by Warkentin and coworkers). As assumed by Warkentin and coworkers and supported by the present analysis,  $S(\mathbf{r}, t)$  decays exponentially with dose  $D$ . However, since the highest resolution diffraction peaks fade out at the smallest doses, they are down-weighted in the definition of (8). If the effects of dose within a given resolution shell centered at some  $q$  are of primary interest, then the weighting diffraction can be restricted to that  $q$ , *e.g.*

$$\text{DWD}(q, t) = \frac{\int_0^t \int_{\text{crystal}} D(\mathbf{r}, t') s(\mathbf{r}, q, t') F_{\text{inc}}(\mathbf{r}, t') d^3r \cdot dt'}{\int_0^t \int_{\text{crystal}} s(\mathbf{r}, q, t') F_{\text{inc}}(\mathbf{r}, t') d^3r \cdot dt'}, \quad (9)$$

where

$$S(\mathbf{r}, t) = \int_q s(\mathbf{r}, q, t) \times q dq$$

and, averaging over reflections at a given  $q$  as in our model in Section 2.3,  $s(\mathbf{r}, q, t) \equiv \Gamma[D(\mathbf{r}, t), q] \times \text{DWF}(B, q)$ .

#### 4.5. Mechanisms underlying the local $I(q, D)$ relation

Global radiation damage to biomolecular crystals at 100 K is robust: it does not have an appreciable dependence on any properties of the biomolecule (*e.g.* primary sequence and fold) or of the crystal (*e.g.* packing density, solvent content and composition). Each X-ray absorption or inelastic scattering event generates secondary electrons and damage in a volume of many cubic micrometres containing a large number of biomolecules and unit cells. The frozen solvent network prevents relaxation of the structure following each damage event on any but the shortest length scales. As a result, one might expect that a fairly simple physical damage model would be required to reproduce the essential features of the decay of diffraction with dose at 100 K.

This expectation is borne out by the present results. A model of sequential random damage interactions that cause local blurring of the electron density reproduces all salient trends of the available data for global radiation damage, including both its dose and  $q$  dependence (to within uncertainties arising from how available data were collected).

The apparent asymptote of the simulation results to  $\alpha = 2$  at large  $q$  in Fig. 6 can be readily understood. In the limit of each pixel or voxel having received a large number of hits  $N$ , the fractional fluctuations in the number of hits (or blurs) per voxel  $N^{1/2}/N$  will become small, and the blurring of the electron density will become nearly uniform throughout the sample. The uniformity of the blurring will be greater for long-

wavelength (large numeric resolution, small  $q$ ) Fourier components of the electron density, since these average over fluctuations in a larger volume. Instead of our elementary damage event corresponding to a local blurring in a small volume of the crystal, in this large-dose, low-resolution limit we can assume a simpler model in which our elementary damage event delivers a uniform dose  $D_0$  that produces a uniform blurring throughout the crystal. After  $n$  of these events, the total dose received by the sample is  $D_n = nD_0$ , and the electron density satisfies  $\rho(\mathbf{r}, D_n) = \rho(\mathbf{r}, D_{n-1}) \otimes G(\mathbf{r})$ , where the blur kernel  $G(\mathbf{r})$  is convolved with the real-space density. The Fourier transform of the electron density and thus the diffracted intensity will have the form  $I(q, D_n) = I(q, 0) \times [G'(q)]^{2n}$ . Taking  $G(\mathbf{r})$  and thus  $G'(q)$  as isotropic Gaussians, the diffracted intensity can be written as

$$\begin{aligned} I(q, D_n) &\propto I(q, 0) \times \exp(-4\pi^2 \sigma^2 q^2 n) \\ &\propto I(q, 0) \times \exp[-D_n/D_e(q)], \end{aligned} \quad (10)$$

where

$$D_e(q) \propto 1/q^2. \quad (11)$$

Thus, this simplified model predicts an exponential dependence of intensity on dose and an exponent  $\alpha = 2$  in (4).

Above the protein–solvent glass transition near 200 K, damage processes involving relaxations on large length scales and longer, temperature-dependent timescales may qualitatively change the evolution of disorder with dose, especially at large doses. Deviations of the local  $\tilde{I}(q, D)$ , determined by deconvolving the effects of nonuniform irradiation, from the predictions of the simple model used here should provide a useful starting point for the study of these damage processes.

#### 4.6. Implications for crystallographic data analysis

Diffraction scaling programs used in crystallography attempt to correct for changes in Bragg intensities owing to radiation damage. An early approach assumed a linear variation of peak intensity with frame number or dose and used measurements of equivalent peaks in different frames to extrapolate back to the zero-dose intensity. A second approach assumes that the decay of intensities with dose can be described by a linearly increasing  $B$  factor,  $I(\theta, n) = I_0[B(n)\sin^2(\theta)/\lambda^2]$ , where  $B(n) = (1 - n)B_0 + nB_1$  and  $n$ , the frame number, is proportional to dose (Otwinowski & Minor, 1997; Evans, 2006). This can be rewritten as

$$\begin{aligned} I(q, n) &= I_0 \exp\left(-\frac{B_0 q^2}{16\pi^2}\right) \exp\left[-\frac{(B_1 - B_0)nq^2}{16\pi^2}\right] \\ &= I(q, 0) \exp[-n/n_e(q)] \end{aligned} \quad (12)$$

with  $n_e(q) \propto 1/q^2$ . This matches the result of our model and is consistent (within experimental uncertainties) with the local  $\tilde{I}(q, D)$  relation we deduced from previous experiments.

This correction can be calculated separately for each set of symmetry-related peaks and their Friedel mates, assuming that all peaks in this set have the same decay rate (Diederichs, 2006). This requires that the multiplicity of the data set be

sufficiently high and that observations of equivalent reflections are well spaced over the data set.

The difficulty in these approaches in accounting for radiation damage is that the measured peak intensities (even when fully recorded) may not have an exponential dependence on frame number (Owen *et al.*, 2014; Warkentin *et al.*, 2017), the dose state of the diffracting crystal region does not in general vary linearly with frame number (Warkentin *et al.*, 2017), and both of these effects are by far the largest for the highest resolution data. Because the crystal is nonuniformly irradiated and nonuniformly damaged, diffraction in any frame reflects a nonlinear weighting of structure factors from crystal regions in various states of decay. Since the local  $D_e(q) \propto 1/q^2$ , the effects of nonuniform irradiation are much more pronounced, and become evident at much smaller exposures, for the highest resolution structure factors. The size, shape and position in reciprocal space of a structure-factor peak evolve with dose, because mosaicity, the spread in lattice constants within the illuminated volume, and the average lattice constant generally increase with dose. Depending on the incident beam divergence and energy spread, the initial crystal mosaicity and its rate of increase with dose, the incident beam fluence or dose per frame and the sample rotation per frame, the resulting intensity variations with exposure time can introduce large errors in nominally fully recorded peak intensities. Recording high-multiplicity data can average over these effects, but there is no reason to expect that the ‘average’ for a set of equivalent reflections, or its extrapolation to frame  $n = 0$  based on damage models implemented in current scaling programs, will correspond to the  $n = 0$  structure factor, especially in the highest resolution shells. This may contribute to the rapid degradation of  $R$  factors within each shell as the resolution limit of a data set is approached (Holton *et al.*, 2014).

The good news, evident from the present work, is that our understanding of radiation damage is improving. It should soon be possible to implement much more sophisticated models, and perhaps also improved crystallographic data-collection protocols that include the measurement of key damage-related parameters, to allow more accurate correction of measured intensities. This could help close the  $R$ -factor gap between protein and small-molecule structures. Whether the corrections are large enough to significantly impact structural models and mechanistic understanding remains to be determined.

## 5. Conclusions

We have shown that the experimentally observed diffracted intensity decays and their resolution or  $q$  dependence, arising from radiation damage to biomolecular crystals, can be explained by assuming a locally exponential relation  $\tilde{I}(q, D)$  between diffracted intensity and dose with a half-dose  $D_{1/2}(q) \propto 1/q^\alpha$  where  $\alpha \simeq 1.7$ , and by accounting for the effects of non-uniform irradiation, damage and diffraction during data collection. The very strong dependence of  $D_{1/2}(q)$  on  $q$  increases the effects of both radiation damage and of non-uniform irradiation on measured intensities (and the structure

factors derived from them) in the highest resolution shells. Consequently, the 20–30 MGy Henderson or Garman dose limit, which has long been used as a rule of thumb in crystallography, should be replaced with a metric that depends on the initial maximum resolution of a data set, and the application of this metric should account for nonuniform irradiation and diffraction during data collection. Radiation-damage models that have been long used to fit data for  $I(q)$  versus dose implicitly assume uniform sample irradiation and do not apply. Both an exponential dose dependence for  $\tilde{I}(q, D)$  and  $D_{1/2}(q) \propto 1/q^\alpha$  with  $\alpha \simeq 2$  follow from perhaps the simplest physically plausible model, in which damage events cause random, local blurring of the electron density. Experimental deviations from these model predictions, especially at temperatures above the protein–solvent glass transition where radiation sensitivity rapidly increases, should guide the development of a more complete model.

## Acknowledgements

We wish to thank Professor Chris Myers for advice on the simulations. This research was conducted with support from the Cornell University Center for Advanced Computing.

## Funding information

The following funding is acknowledged: National Science Foundation, Directorate for Biological Sciences (award No. MCB-1330685); National Institutes of Health, National Institute of General Medical Sciences (award No. R01-GM127528).

## References

- Arzt, S., Beteva, A., Cipriani, F., Delageniere, S., Felisaz, F., Förstner, G., Gordon, E., Launer, L., Lavault, B., Leonard, G., Mairs, T., McCarthy, A., McCarthy, J., McSweeney, S., Meyer, J., Mitchell, E., Monaco, S., Nurizzo, D., Ravelli, R., Rey, V., Shepard, W., Spruce, D., Svensson, O. & Theveneau, P. (2005). *Prog. Biophys. Mol. Biol.* **89**, 124–152.
- Banumathi, S., Zwart, P. H., Ramagopal, U. A., Dauter, M. & Dauter, Z. (2004). *Acta Cryst.* **D60**, 1085–1093.
- Blake, C. C. F. & Phillips, D. C. (1962). *Biological Effects of Ionizing Radiation at the Molecular Level*, pp. 183–191. Vienna: International Atomic Energy Agency.
- Bourenkov, G. P. & Popov, A. N. (2010). *Acta Cryst.* **D66**, 409–419.
- Burmeister, W. P. (2000). *Acta Cryst.* **D56**, 328–341.
- Bury, C. S., Brooks-Bartlett, J. C., Walsh, S. P. & Garman, E. F. (2018). *Protein Sci.* **27**, 217–228.
- Coughlan, H. D., Darmanin, C., Kirkwood, H. J., Phillips, N. W., Hoxley, D., Clark, J. N., Vine, D. J., Hofmann, F., Harder, R. J., Maxey, E. & Abbey, B. (2017). *J. Synchrotron Rad.* **24**, 83–94.
- Coughlan, H. D., Darmanin, C., Phillips, N. W., Hofmann, F., Clark, J. N., Harder, R. J., Vine, D. J. & Abbey, B. (2015). *Struct. Dyn.* **2**, 041704.
- Diederichs, K. (2006). *Acta Cryst.* **D62**, 96–101.
- Diederichs, K., McSweeney, S. & Ravelli, R. B. G. (2003). *Acta Cryst.* **D59**, 903–909.
- Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.
- Finfrock, Y. Z., Stern, E. A., Alkire, R. W., Kas, J. J., Evans-Lutterodt, K., Stein, A., Duke, N., Lazarski, K. & Joachimiak, A. (2013). *Acta Cryst.* **D69**, 1463–1469.

- Finrock, Y. Z., Stern, E. A., Yacoby, Y., Alkire, R. W., Evans-Lutterodt, K., Stein, A., Isakovic, A. F., Kas, J. J. & Joachimiak, A. (2010). *Acta Cryst.* **D66**, 1287–1294.
- Garman, E. F. (2010). *Acta Cryst.* **D66**, 339–351.
- Henderson, R. (1990). *Proc. R. Soc. Lond. B Biol. Sci.* **241**, 6–8.
- Hendrickson, W. A. (1976). *J. Mol. Biol.* **106**, 889–893.
- Hirata, K., Shinzawa-Itoh, K., Yano, N., Takemura, S., Kato, K., Hatanaka, M., Muramoto, K., Kawahara, T., Tsukihara, T., Yamashita, E., Tono, K., Ueno, G., Hikima, T., Murakami, H., Inubushi, Y., Yabashi, M., Ishikawa, T., Yamamoto, M., Ogura, T., Sugimoto, H., Shen, J.-R., Yoshikawa, S. & Ago, H. (2014). *Nature Methods*, **11**, 734–736.
- Holton, J. M. (2009). *J. Synchrotron Rad.* **16**, 133–142.
- Holton, J. M., Classen, S., Frankel, K. A. & Tainer, J. A. (2014). *FEBS J.* **281**, 4046–4060.
- 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. C. H. & Starodub, D. (2009). *J. Electron Spectrosc. Relat. Phenom.* **170**, 4–12.
- Hubbell, J. & Seltzer, S. (2004). *Tables of X-ray Mass Attenuation Coefficients and Mass Energy-Absorption Coefficients*. Gaithersburg: National Institute of Standards and Technology.
- Keedy, D. A., Kenner, L. R., Warkentin, M., Woldeyes, R. A., Hopkins, J. B., Thompson, M. C., Brewster, A. S., Van Benschoten, A. H., Baxter, E. L., Uervirojnangkoorn, M., McPhillips, S. E., Song, J., Alonso-Mori, R., Holton, J. M., Weis, W. I., Brunger, A. T., Soltis, S. M., Lemke, H., Gonzalez, A., Sauter, N. K., Cohen, A. E., van den Bedem, H., Thorne, R. E. & Fraser, J. S. (2015). *Elife*, **4**, 07574.
- Kmetko, J., Husseini, N. S., Naides, M., Kalinin, Y. & Thorne, R. E. (2006). *Acta Cryst.* **D62**, 1030–1038.
- Leal, R. M. F., Bourenkov, G., Russi, S. & Popov, A. N. (2013). *J. Synchrotron Rad.* **20**, 14–22.
- Liebschner, D., Rosenbaum, G., Dauter, M. & Dauter, Z. (2015). *Acta Cryst.* **D71**, 772–778.
- Maser, J., Osanna, A., Wang, Y., Jacobsen, C., Kirz, J., Spector, S., Winn, B. & Tennant, D. (2000). *J. Microsc.* **197**, 68–79.
- Massover, W. H. (2007). *J. Synchrotron Rad.* **14**, 116–127.
- Meents, A., Dittrich, B. & Gutmann, S. (2009). *J. Synchrotron Rad.* **16**, 183–190.
- Nave, C. & Hill, M. A. (2005). *J. Synchrotron Rad.* **12**, 299–303.
- Nave, C., Sutton, G., Evans, G., Owen, R., Rau, C., Robinson, I. & Stuart, D. I. (2016). *J. Synchrotron Rad.* **23**, 228–237.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Otwinowski, Z., Minor, W., Borek, D. & Cymborowski, M. (2012). *International Tables for Crystallography*, Vol. F, edited by E. Arnold, D. M. Himmel & M. G. Rossmann, pp. 282–295. Dordrecht: Kluwer Academic Publishers.
- Owen, R. L., Paterson, N., Axford, D., Aishima, J., Schulze-Briese, C., Ren, J., Fry, E. E., Stuart, D. I. & Evans, G. (2014). *Acta Cryst.* **D70**, 1248–1256.
- Owen, R. L., Rudiño-Piñera, E. & Garman, E. F. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 4912–4917.
- Ravelli, R. B. G. & McSweeney, S. M. (2000). *Structure*, **8**, 315–328.
- Sanishvili, R., Yoder, D. W., Pothineni, S. B., Rosenbaum, G., Xu, S., Vogt, S., Stepanov, S., Makarov, O. A., Corcoran, S., Benn, R., Nagarajan, V., Smith, J. L. & Fischetti, R. F. (2011). *Proc. Natl Acad. Sci. USA*, **108**, 6127–6132.
- Sliz, P., Harrison, S. C. & Rosenbaum, G. (2003). *Structure*, **11**, 13–19.
- Sygyusch, J. & Allaire, M. (1988). *Acta Cryst.* **A44**, 443–448.
- Teng, T. & Moffat, K. (2000). *J. Synchrotron Rad.* **7**, 313–317.
- Teng, T.-Y. & Moffat, K. (2002). *J. Synchrotron Rad.* **9**, 198–201.
- Wang, C. K., Hu, S.-H., Martin, J. L., Sjögren, T., Hajdu, J., Bohlin, L., Claeson, P., Göransson, U., Rosengren, K. J., Tang, J., Tan, N.-H. & Craik, D. J. (2009). *J. Biol. Chem.* **284**, 10672–10683.
- Warkentin, M. A., Atakisi, H., Hopkins, J. B., Walko, D. & Thorne, R. E. (2017). *IUCrJ*, **4**, 785–794.
- Warkentin, M., Hopkins, J. B., Badeau, R., Mulichak, A. M., Keefe, L. J. & Thorne, R. E. (2013). *J. Synchrotron Rad.* **20**, 7–13.
- Warkentin, M., Hopkins, J. B., Haber, J. B., Blaha, G. & Thorne, R. E. (2014). *Acta Cryst.* **D70**, 2890–2896.
- Warkentin, M. & Thorne, R. E. (2010). *Acta Cryst.* **D66**, 1092–1100.
- Weik, M., Ravelli, R. B. G., Kryger, G., McSweeney, S., Raves, M. L., Harel, M., Gros, P., Silman, I., Kroon, J. & Sussman, J. L. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 623–628.
- Zeldin, O. B., Brockhauser, S., Bremridge, J., Holton, J. M. & Garman, E. F. (2013). *Proc. Natl Acad. Sci. USA*, **110**, 20551–20556.