

X-ray absorption by macromolecular crystals: the effects of wavelength and crystal composition on absorbed dose

James W. Murray,^a Elspeth F. Garman^{a*} and Raimond B. G. Ravelli^{b*}

^aLaboratory of Molecular Biophysics, Department of Biochemistry, Oxford University, Rex Richards Building, South Parks Road, Oxford OX1 3QU, UK, and ^bEuropean Molecular Biology Laboratory (EMBL) Grenoble Outstation, 6 rue Jules Horowitz, BP 181, F38042 Grenoble Cedex 9, France. Correspondence e-mail: elspeth@biop.ox.ac.uk, ravelli@embl-grenoble.fr

Radiation damage restricts the useful lifetime for macromolecular crystals in the X-ray beam, even at cryotemperatures. With the development of structural genomics pipelines, it will be essential to incorporate projected crystal lifetime information into the automated data collection software routines. As a first step towards this goal, a computer program, *RADDOSE*, is presented which is designed for use by crystallographers in optimizing the amount of data that can be obtained from a particular cryo-cooled crystal at synchrotron beamlines. The program uses the composition of the crystal and buffer constituents, as well as the beam energy, flux and dimensions, to compute the absorption coefficients and hence the theoretical time taken to reach an absorbed dose of 2×10^7 Gy, the so-called 'Henderson limit'. At this dose, the intensity of the diffraction pattern is predicted to be halved. A 'diffraction-dose efficiency' quantity is introduced, for the convenient comparison of absorbed dose per diffracted photon for different crystals. Four example cases are considered, and the implications for anomalous data collection are discussed in the light of the results from *RADDOSE*.

© 2004 International Union of Crystallography
Printed in Great Britain – all rights reserved

1. Introduction

Three-dimensional macromolecular structures can be solved by the collection of X-ray diffraction patterns from crystals. Since the X-rays used to collect these diffraction patterns are ionizing radiation, they not only scatter elastically but they also deposit energy in the crystal. Free radicals created by this deposited energy can diffuse through the crystal, causing a chain reaction of radiation damage events. Cooling to cryotemperatures reduces this diffusion, thus slowing the damage process. However, even at cryotemperatures, synchrotron wiggler and undulator X-ray beams cause noticeable radiation damage during standard data collections (Garman & Murray, 2003). Radiation damage is the limiting experimental factor for many data collections *e.g.* Wimberly *et al.* (2000), Rudenko *et al.* (2002).

Prior to the wide use of synchrotrons for X-ray diffraction data collection, in-house X-ray generators generally provided one of two incident X-ray wavelengths, corresponding to the copper and molybdenum $K\alpha$ emission lines [8.04 keV (1.54 Å) and 17.46 keV (0.71 Å), respectively]. Experimental phases were usually determined by MIR (multiple isomorphous replacement) and used a restricted selection of different heavy atoms, *e.g.* Pt, Hg, Au and Pb. Arndt (1984) considered the optimal incident X-ray energy under these conditions and

concluded that the diffracted intensity per unit energy deposited in the crystal increased by a factor of 1.8 between 4.8 and 31 keV energy, favouring data collection at higher X-ray energies.

At the start of the 21st century the choices are much wider. Many different types of atoms are used as sources of phase information for MIR and MAD (multiple-wavelength anomalous dispersion), and the anomalous scattering information is almost always exploited [MIRAS or SIRAS (multiple or single isomorphous replacement with anomalous scattering) and SAD (single-wavelength anomalous dispersion)]. Synchrotron X-ray sources provide beamlines that can produce a wide range of X-ray energies. Therefore, it is now possible to match the X-ray energy to the experimental requirements more closely.

The structure of a macromolecular crystal is disrupted by the deposited X-ray energy, so there is an intrinsic limit on the amount of diffraction data from coherent scattering that may be collected before the crystalline diffraction is destroyed by radiation damage (Blake & Philips, 1962). The theoretical limit for a cryo-cooled crystal is commonly taken to be 2×10^7 Gy [1 Gray (Gy) = 1 J kg^{-1}] following Henderson (1990), calculated by analogy with cryo-electron microscopy as the dose which reduces the diffraction pattern to half of its original intensity.

The energy absorbed by a crystal when irradiated by an X-ray beam is a function of the incident X-ray energy and of μ_{abs} , the X-ray absorption coefficient, which is in turn a function of the atomic composition (macromolecule and solvent) of the crystal. Using these known data, as well as tabulated values for X-ray atomic cross sections, it is possible to calculate absorption coefficients for a particular protein crystal.

In order to plan an optimal data collection strategy, it is necessary to not only have an accurate estimate of μ_{abs} but also to include the relevant beam conditions for that particular experiment, *i.e.* the incident X-ray flux, beam size and beam profile, as well as the physical dimensions of the crystal. Given these pieces of information, the expected lifetime of the crystal in the beam can be predicted.

Other programs that calculate μ_{abs} exist, such as *FHKL* (Soyer, 1995) which computes scattering factors but is not optimized for use with protein crystals. *SHELXL* (Sheldrick & Schneider, 1997) also calculates μ_{abs} but does not take into account the solvent content and non-ordered part of the unit cell. In some cases this leads to poor estimates of μ_{abs} , for instance for holoferritin which contains a core of up to 4500 Fe atoms per molecule. In this extreme example, the absorption above the iron *K*-edge becomes so large that data collection at room temperature has only been possible below this edge (Smith *et al.*, 1989).

We have written a program *RADDOSE*, which is tailored to the needs of crystallographers so that they can conveniently calculate the absorbed dose and obtain an estimate of the theoretical lifetime of their crystals under particular incident beam and slit conditions. *RADDOSE* uses library values (McMaster *et al.*, 1969) of the absorption coefficients of different atoms to calculate μ_{abs} for macromolecular crystals taking into account the solvent region and the solvent constituents, as well as any heavy atoms that are incorporated in or bound to the macromolecules.

The value of μ_{abs} for the crystal calculated by *RADDOSE* and the beam conditions entered by the user are then used to compute a number of parameters relevant to the experimental design. These include the total exposure time at which the Henderson limit is reached, as well as an estimate of the maximum heating of the crystal as calculated using 'lumped model' solutions (Kuzay *et al.*, 2001). The lumped model is an analytical approach to isothermal convective heat exchange between the cold nitrogen gas stream and the crystal in the loop. Thermal gradients within the crystal are neglected; this is not valid for highly absorbing crystals or for a non-uniform beam-intensity profile. However, although the model makes a number of simplifying assumptions, it is realistic enough to give an idea of relative temperature rises, and results from it have been born out by more sophisticated treatments (Kriminski *et al.*, 2003).

2. Theoretical considerations

At the X-ray energies typically used in a crystallographic experiment (around 10 keV), the incoming photon can

interact with atoms in the crystal in three principal ways (Nave, 1995; Shmueli, 2001). Firstly it can collide elastically (Thomson or Rayleigh scattering) with the atom, scattering coherently and giving rise to the observed diffraction pattern.

Secondly, the photon may be inelastically scattered (Compton scattering) with a cross section comparable with the Thomson cross section for light elements. The incident photon transfers energy to electrons in atoms and a second photon of lower energy is emitted incoherently, giving rise to a component of the background on diffraction images. However, Compton scattering deposits negligible amounts of energy in the crystal compared with the third possibility; the photoelectric interaction, in which the incident X-ray is totally absorbed, and a photoelectron is ejected. The photoelectric effect is the dominant interaction at the energies of interest to X-ray crystallographers, and the entire energy of the photon is often deposited into the crystal. For the purposes of the calculations described here, we have assumed that the energy of the fluorescent X-ray (produced by some photoelectric interactions) is absorbed in the crystal, whereas in reality it may sometimes escape from the crystal. The energy of the photoelectron is dissipated in the crystal, contributing to mechanical disruption of the crystal lattice, thermal heating, ionizations and bond scissions. Although the relative probabilities of these various events is not known, it has been estimated that a single absorbed photon could produce ~500 ionization events (O'Neill *et al.*, 2002). The absorbed energy is not uniformly distributed but is deposited in small volumes known as 'spurs' (Ravelli & McSweeney, 2000).

An important but often neglected point is that the attenuation coefficient for the crystal, μ_{att} , is not the same as the absorption coefficient, μ_{abs} . Photons that are coherently or Compton scattered contribute little to absorption (and are neglected in *RADDOSE*), but they do attenuate the beam. *RADDOSE* uses the total cross section to obtain the attenuation coefficient, but only the photoelectric cross section is included in calculating the absorption coefficient. Neglecting the Compton scattering contribution leads to an underestimate of the absorbed dose, whereas neglecting the possible escape of fluorescent X-rays leads to an overestimate of absorbed dose.

2.1. The diffraction efficiency

To a first approximation, the total linear attenuation coefficient, μ_{att} , of the crystal can be estimated from the sum of the atomic cross sections, σ_j , of all the atoms, N , in the unit cell of volume V_c , as shown in equation (1).

$$\mu_{\text{att}} = (1/V_c) \sum_{j=1}^N \sigma_j, \quad (1)$$

where $\sigma = \sigma_{\text{Thomson}} + \sigma_{\text{Compton}} + \sigma_{\text{photoelectric}}$.

This approximation makes the assumption that the cross section of an atom is independent of its environment. This is valid for light atoms, but for heavy atoms close to absorption energies it results in a poor description. For these cases, empirical measurements of cross section may be used. For

macromolecular crystallographers, the easiest way to measure absorption coefficients is by the fluorescence spectrum produced by a scan prior to a MAD experiment. The fluorescence is directly proportional to the photoelectric absorption coefficient. Therefore, accurate values of $\sigma_{\text{photoelectric}}$ may be estimated from the fluorescence spectrum scaled using values of the absorption coefficient far from the absorption edge.

The contents of the unit cell in macromolecular crystallography are usually well defined, since the sequence and thus composition of the macromolecule is known, and the proportion of solvent may be estimated (Matthews, 1968). An uncertainty in the number of molecules in the unit cell will be reflected in a predictable uncertainty in the theoretical lifetime of the crystal: an example of this problem is given in §4.3. Theoretical and quasi-empirical tables of X-ray scattering cross sections are available (Cromer & Liberman, 1970). Away from absorption edges, these are typically accurate to within a few percent. However, some elements have an intense maximum in their absorption spectrum at the absorption peak, a so called 'white line', where the scattering cross sections deviate significantly from the tabulated values. Even so, using the tables it is possible to estimate the absorption of a protein crystal as a function of incident X-ray energy and of unit-cell contents.

Following the argument of Arndt (1984), the integrated intensity, I_{scatt} , of a diffraction spot from a crystal of volume V and path length in the beam t , of incident intensity I_0 , wavelength λ and Bragg angle θ , is

$$I_{\text{scatt}} \propto I_0 \frac{V\lambda^3 \exp(-\mu_{\text{att}} t)}{\sin 2\theta}, \quad (2)$$

where μ_{att} is the X-ray attenuation coefficient of the sample. At small Bragg angles $\sin 2\theta \simeq 2 \sin \theta = \lambda/d$, so

$$I_{\text{scatt}} \propto I_0 V \lambda^2 \exp(-\mu_{\text{att}} t). \quad (3)$$

The energy absorbed by a crystal, E_{abs} , where μ_{abs} is the X-ray absorption coefficient, is proportional to the incident-beam intensity and is

$$E_{\text{abs}} \propto \frac{I_0}{\lambda} [1 - \exp(-\mu_{\text{abs}} t)]. \quad (4)$$

For protein crystals, the important parameter is the dose: the absorbed energy per unit mass, rather than the total absorbed energy (this is larger for large crystals). For a given photon fluence, the dose is smaller the larger the crystal, as a result of the attenuation of the beam by the crystal. This effect is particularly important at lower energies where the absorption can be very large. The dose (energy per unit mass) absorbed by an isometric crystal is

$$D \propto \frac{I_0}{\lambda V} [1 - \exp(-\mu_{\text{abs}} t)]. \quad (5)$$

A naïve formulation of dose effects would maximize the diffraction as a function of energy. However, Polikarpov *et al.* (1997) proposed a diffraction efficiency measure based on the intensity divided by the energy deposited [equivalent to

equation (6)]. The diffracted intensity per energy absorbed ($I_{\text{scatt}}/E_{\text{abs}}$) is

$$I_E \propto V\lambda^3 \frac{\exp(-\mu_{\text{att}} t)}{[1 - \exp(-\mu_{\text{abs}} t)]}. \quad (6)$$

We suggest that a more appropriate measure would be the diffracted intensity per absorbed dose, as in equation (7), which we call the diffraction-dose efficiency, I_{DE} . The diffracted intensity per absorbed dose (I_{scatt}/D) is

$$I_{\text{DE}} \propto \lambda^3 \frac{\exp(-\mu_{\text{att}} t)}{[1 - \exp(-\mu_{\text{abs}} t)]}. \quad (7)$$

As can be seen from equation (5), the consequence of using the value of μ_{att} instead of μ_{abs} in dose calculations is an overestimate of the absorbed dose. The error becomes more significant at high energies, where the photoelectric cross section is reduced in proportion to the coherent scattering.

Ideally the most useful information for the crystallographer would be how to maximize the signal-to-noise ratio for the integrated diffraction intensities in order to obtain the best possible data during the limited lifetime of the crystal. This calculation would require a more complete description of the diffraction experiment which is beyond the current scope of *RADDOS*, since additional terms should be added to take into account factors such as air scatter, sample-to-detector distance and detector response. The background is very much harder to model than the diffraction, as it consists of many independent components including fluorescence, Compton scattered photons, detector readout noise, thermal diffuse scatter and air scatter. The program *BEST* (Popov & Bour-enkov, 2003) allows such modelling of the experiment, but currently neglects absorption and radiation damage.

2.2. Beam profiles

The intensity of synchrotron X-ray beams is not uniform, having spatial and temporal structure. The intensity profile of a beam, $I(x, y)$, can be modelled as an elliptical Gaussian function. The parameters of the Gaussian function can be derived from the full-width half-maximum (FWHM), W , of the beam profile in two orthogonal directions. If X and Y are the beam width and height, and W_x and W_y are the full-width half-maxima with the beam centre at the origin, the relative intensity at any point, $I(x, y)$, can be derived using a Gaussian model.

$$\sigma_x = \frac{W_x}{2(2 \ln 2)^{1/2}} \text{ and } \sigma_y = \frac{W_y}{2(2 \ln 2)^{1/2}}, \quad (8)$$

$$I(x, y) = \frac{1}{C} \exp \left[-\left(\frac{x^2}{2\sigma_x^2} \right) - \left(\frac{y^2}{2\sigma_y^2} \right) \right], \quad (9)$$

where C is a normalizing factor and A is the area of the beam:

$$C = \frac{1}{A} \int_{-X/2}^{X/2} \int_{-Y/2}^{Y/2} \exp \left[-\left(\frac{x^2}{2\sigma_x^2} \right) - \left(\frac{y^2}{2\sigma_y^2} \right) \right] dx dy. \quad (10)$$

For a uniform beam, the relative positional intensity of the centre of the beam is the same as that at the edges. However, for a non-uniform beam, the centre may be much more intense than the edges. Given the beam profile parameters, it is possible to calculate some consequences of the beam profile.

Fig. 1 shows three cases for which the beam-defining slits have been set to $100 \times 100 \mu\text{m}$. The beam focused with an FWHM of $50 \mu\text{m}$ shows a peak intensity 20 times higher than that at the edge, whereas the beam focused with an FWHM of $250 \mu\text{m}$ is almost flat across the slits. Assuming an equal total integrated intensity, the tightly focused beam has a peak intensity almost two times higher than that of the uniform beam. Therefore, the effective crystal lifetime in the centre of the tightly focused beam would be half that of a uniform beam, but ten times more at the edges of the beam. It is well known that radiation damage affects the diffraction pattern, one cause being specific structural damage. If the crystal is differentially irradiated, this specific structural damage will occur at different rates across the beam profile. Measured reflections would thus be sampling a continuum of structures rather than a uniform state of damage at any given time, possibly leading to worse crystallographic *R* values.

3. RADDOSE

A computer program *RADDOSE* has been written to estimate the absorption coefficients and to use these to calculate the theoretical absorbed dose for a given crystal and beam. *RADDOSE* then computes the exposure time to the Henderson limit as well as the predicted temperature rise in the crystal. The program is written in FORTRAN77 using CCP4 library routines (Collaborative Computational Project, No. 4, 1994). The X-ray cross sections used are those tabulated by McMaster *et al.* (1969), incorporated into the 'mucal' Fortran routine (Badyopadhyay, 1995).

The program requires information about the crystal, which is described by three orthogonal dimensions, two being perpendicular to the beam. The volume and contents of the

unit cell are required. The unit-cell contents are estimated on the basis of the number of amino acids in the protein and the number of RNA or DNA bases present. The solvent content is computed by the program and is taken to be water as well as any additional solvent constituents specified by the user. Alternatively, the precise numbers of each type of atom per unit cell may be entered explicitly. The unit-cell contents are then used to estimate the total cross section as well as separate cross sections for the photoelectric effect, and for Compton and Rayleigh scattering. The total cross section is used to calculate the attenuation coefficient, and the photoelectric cross section alone is used to compute the absorption coefficient.

For heavy atoms at energies close to absorption edges the tabulated values for the absorption coefficients are not accurate. If desired by the experimenter, a normalized fluorescence spectrum, derived from an experimental fluorescence spectrum by the program *CHOOCH* (Evans & Pettifer, 2000; *SPLINOR* file output), may be input. *RADDOSE* then scales the high- and low-energy limits of this spectrum to the tabulated absorption cross sections to obtain more realistic values for the absorption coefficients for the heavier elements over their absorption edges.

The photon flux and energy, beam size, exposure time per image, and number of images are also input. The program calculates the dose per image and data set, and the exposure time available for data collection before a given dose limit (defaulting to 2×10^7 Gy) is reached, as well as the diffraction-dose efficiency. A range of energies can be used to study the effect of incident energy on the predicted lifetime of the crystal.

The calculations do not take into account any rotation of the crystal during the experiment. For a crystal that is much larger than the beam, the calculation is therefore only valid for the part of the crystal that stays in the beam throughout the data collection. Using the calculated absorption, the crystal and beam dimensions, the photon flux, and the total exposure time, the program will calculate the number of photons absorbed in the crystal, and this value is converted to the amount of energy deposited in the crystal. This energy divided by the mass of the exposed part of the crystal gives the absorbed dose. The default beam profile is a top-hat function, but the parameters of non-uniform Gaussian beam profiles can be entered if required, in which case the peak beam intensity is used to calculate the dose limits.

In *RADDOSE* we have assumed that the rate of energy deposition does not affect the radiation dose limit. There is disagreement between experiments reported in the literature as to whether the rate of deposition of dose in the crystal is important rather than just the total dose: the so-called dose/dose-rate effect (Leiros *et al.*, 2001; Sliz *et al.*, 2003). One less disputed possible cause of dose/dose-rate effects is crystal heating induced by the beam. This heating is proportional to the dose rate, and can become large for very intense beams combined with strongly absorbing crystals (Ravelli *et al.*, 2002), resulting in increased mobility of free-radical species within the crystal and thus increased rates of radiation

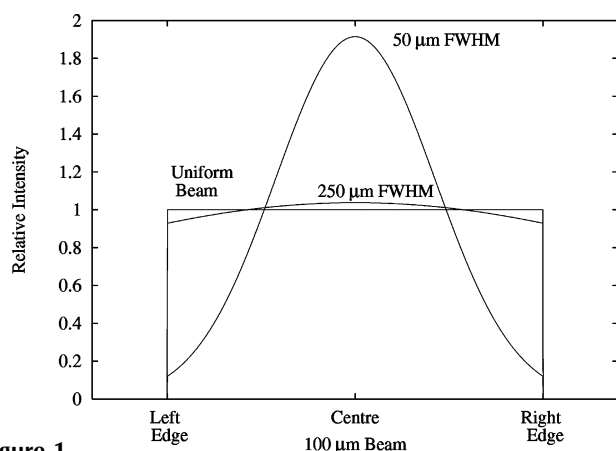


Figure 1
Graphs to show three beam profiles for hypothetical beams across a slit width of $100 \mu\text{m}$. The beam profile shapes are: a Gaussian with a full-width half-maximum (FWHM) of $50 \mu\text{m}$, a Gaussian with an FWHM of $250 \mu\text{m}$ and a top hat (uniform beam).

damage. Additionally, if the temperature rise takes the sample above the phase transition of the solvent glass in the crystal at around 155 K (Weik *et al.*, 2001), crystalline ice will form and the diffraction pattern will be degraded. For the calculations of the temperature rise induced by the X-ray beam in the sample using the isothermal 'lumped model' of Kuzay *et al.* (2001) in *RADDOSE*, two parameters are required to specify the thermal properties of the crystal: the heat capacity, c_p of the protein, estimated to be $5 \times 10^2 \text{ J K}^{-1} \text{ kg}^{-1}$, and the heat-transfer coefficient, h , taken to be $320 \text{ W m}^{-2} \text{ K}^{-1}$ after Kriminski *et al.* (2003). In the examples below, the initial temperature of the crystal has been taken as 100 K.

Using the exposure time per frame and all the parameters given or calculated above, *RADDOSE* will calculate the steady-state temperature increase and the system response time, t_{sys} , which is the characteristic time constant for the convectively cooled and spatially uniform lumped body. The body reaches 95% of the final temperature in three system-time constants. Note that a single exposure may be less than

Table 1

The times taken for protein crystals derivatized with different heavy atoms to reach an absorbed dose of $2 \times 10^7 \text{ Gy}$ at different beam energies.

Protein	No. of amino-acid residues	Heavy-atom compound	Buffer	Energy (keV)	Time (s)
DegP	448 ($\times 2$)	Native	0.1 M Tris-HCl	9.0	249
		Ta ₆ Br ₁₄		9.0	142
		PtCl ₄		9.0	191
		SeMet		9.0	220
		SeMet		12.5 (LE remote)	434
		SeMet		Inflection	280
		SeMet		Peak	127
		SeMet		12.8 (HE remote)	187
LDL receptor	699	Native	0.5 M CaCl ₂	9.0	222
		O ₄₀ PW ₁₂		9.0	161
		O ₄₀ PW ₁₂		10.2 (W L_{III})	78
		SeMet		9.0	220

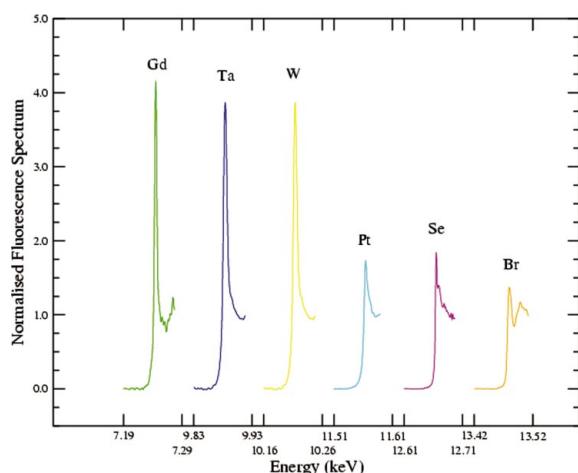


Figure 2
Graph showing the normalized experimental fluorescence spectra from the Gd L_{III} , Ta L_{III} , W L_{III} , Pt L_{III} , Se K and Br K edges.

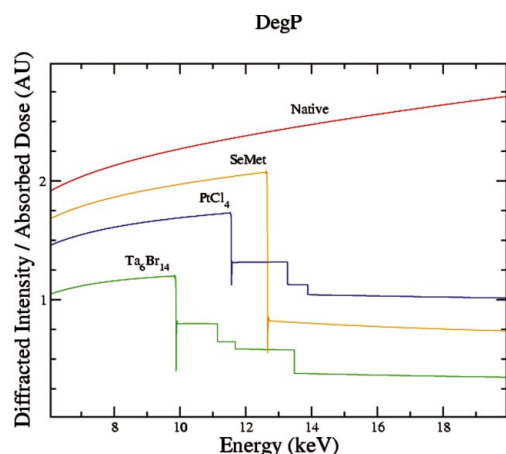


Figure 3
Graph showing the diffraction-dose efficiency for native and derivative crystals of DegP.

this time, so the maximum estimated temperature rise may not be reached.

The beam flux may be estimated from a beamline counter reading if this is calibrated against a reference photodiode on the beamline. The calibration may drift over time, so estimates of the total incident flux should be made in consultation with the local beamline scientist.

4. Examples

To investigate the effects of a variety of strongly absorbing atoms, we selected a number of structures from the literature that were solved using a range of different heavy-atom compounds or anomalous scatterers, and investigated the theoretical crystal lifetimes as a function of X-ray energy [6.0–20 keV (2.07 to 0.62 Å)] using *RADDOSE*. Since the beam flux was not calibrated for these data collections, it is unfor-

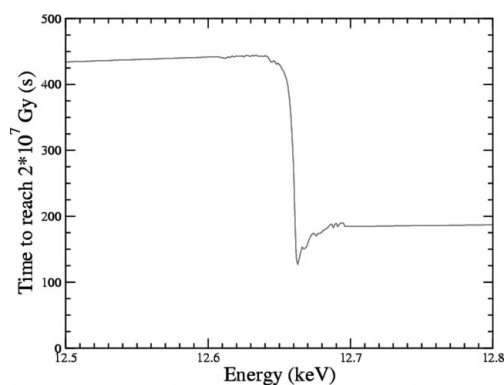


Figure 4
Graph showing the predicted time taken for a $100 \times 100 \times 100 \mu\text{m}$ crystal of SeMet-DegP to reach an absorbed dose of $2 \times 10^7 \text{ Gy}$ over a small range of energies on a high-flux beamline ($10^{12} \text{ photons s}^{-1}$, $100 \times 100 \mu\text{m}$ beam).

tunately not possible to compare the computed results with experimental reality, but the calculations do illustrate the variation in lifetimes and beam-induced heating to be expected for protein crystals.

For all the example calculations, a default crystal size of $100 \times 100 \times 100 \mu\text{m}$ has been used along with a uniform (top-hat) beam profile and a flux of 10^{12} photons per second through 100 by 100 μm slits. This flux can be obtained on the brightest undulator beamlines at third-generation synchrotron sources. The flux depends critically on the resonant energies of the undulator cavity, which produces a series of relatively sharp maxima as a function of energy (Spencer & Winick, 1980), and the undulator gap can usually be tuned to obtain the maximum of a fixed harmonic over a limited energy range. In order to illustrate the application and utility of *RADDOS*, in the calculations presented below we have used the same incident flux for each X-ray energy. However in practice experimenters would have to know the real variation of the beam flux over their desired energy range.

The tabulated photoelectric cross sections of heavy atoms at energies close to their commonly used absorption edges have been improved by fitting them to a number of normalized experimental fluorescence spectra obtained on the EMBL ESRF Joint Structural Biology Group beamlines using the auto-edge-scan routines (Beteva *et al.*, 2003). The normalized spectra are shown in Fig. 2.

4.1. Heat shock protein DegP

The structure of the heat shock protein DegP (HtrA) was solved recently (Krojer *et al.*, 2002). Crystals diffracted extremely anisotropically to 4.5 Å. However, controlled dehydration resulted in isotropic diffraction to 2.8 Å resolution on a high-flux undulator beamline (ID14-EH4 ESRF). Initial phases were determined from crystals soaked in a saturated solution of $\text{Ta}_6\text{Br}_{14}$ that yielded two cluster sites, and those phases allowed the location of four platinum sites in PtCl_4 -soaked crystals. Improved phases were obtained from a MAD experiment performed on an SeMet crystal using the 14

methionine residues per 448-residue monomer. Both the crystallization buffer and the dehydration liquor contained chloride ions but included no other atoms heavier than oxygen.

Fig. 3 shows the diffraction-dose efficiency, I_{DE} , as a function of energy for the four different cases: native DegP crystal, $\text{Ta}_6\text{Br}_{14}$ soak, PtCl_4 soak and the SeMet derivative. Experimental fluorescence scans have been included to improve the tabulated values of μ_{abs} for the L_{III} -edges of tantalum and platinum, and the K -edge of selenium.

It can be seen that the crystals soaked with the large $\text{Ta}_6\text{Br}_{14}$ clusters are calculated to absorb more than either the platinum- or the SeMet-derivatized crystals in almost the whole energy range tested.

The times taken to reach the Henderson dose limit ($2 \times 10^7 \text{ Gy}$) at an energy of 9 keV (1.37 Å), which is below the absorption edges for all these elements, are shown in Table 1.

Fig. 4 shows the expected lifetime (in s) of the SeMet crystal in a small region around the energy of the selenium K absorption edge. Just below the edge, at 12.5 keV, the expected lifetime is 434 s. At the inflection point it is 280 s, whereas at the peak it is further reduced to 127 s. Above the absorption edge, at 12.8 keV, the expected lifetime would be 187 s. Compared with a data set collected at the absorption peak, we would thus expect to collect 3.4, 2.1 or 1.5 times more data just below the absorption edge, at the inflection point or just above the absorption edge, respectively.

The maximum temperature rise as calculated using the lumped-body model is about 7 K at 13.0 keV for the native protein. This value is predicted to increase extremely sharply from 23 K just below the tantalum L_{III} -edge (9.9 keV) to almost 60 K at the absorption maximum, taking it near the solvent-glass transition mentioned earlier. These calculated temperature rises are strongly dependent on the incident flux, beam size and crystal shape. However, the temperature of a $100 \times 100 \mu\text{m}$ crystal with a thickness of 20 μm would still rise by 28 K at the Ta L_{III} -edge with the high flux that was used in these examples.

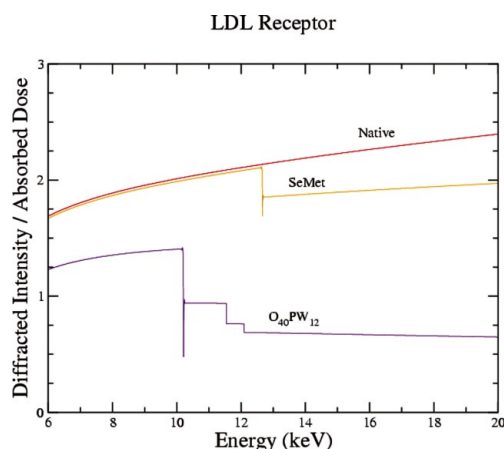


Figure 5
Graph showing the diffraction-dose efficiency for native and derivative crystals of the LDL receptor extracellular domain.

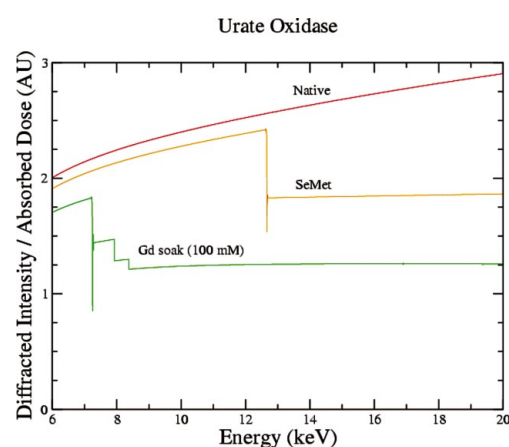


Figure 6
Graph showing the diffraction-dose efficiency for native and derivative crystals of urate oxidase.

Table 2

For different proteins and unit-cell volumes, the number of photons per unit cell required to reach a total absorbed dose of 2×10^7 Gy is shown at two different incident photon energies.

Protein	Space group	Unit-cell parameters	Volume (\AA^3)	No. of photons per unit cell	
		Parameters (\AA , $^\circ$)		6 keV	20 keV
DegP (twofold NCS)	$P6_322$	$a = b = 121.37$, $c = 233.67$	2.98×10^6	70	20
LDL receptor	$P3_121$	$a = b = 185.29$, $c = 85.19$	2.53×10^6	55	16
Urate oxidase	$I222$	$a = 78.9$, $b = 95.2$, $c = 104.1$	0.782×10^6	23	7
Xylanase	$P2_1$	$a = 41.1$, $b = 67.7$, $c = 50.8$, $\beta = 113.5$	0.128×10^6	3	1

4.2. Extracellular domain of the LDL receptor

The determination of the structure of the LDL receptor extracellular domain (LDLR) presented a challenging crystallographic problem, because of the radiation sensitivity, poor diffraction and non-isomorphism of the crystals (Rudenko *et al.*, 2002). Phases were obtained by the MAD method on crystals that were soaked with micromolar amounts of the tungsten cluster compound $\text{Na}_3\text{O}_{40}\text{PW}_{12}$. 2.5 tungsten clusters were bound per 699-residue monomer. Data were collected to 3.7 \AA on a high-intensity beamline (ID19, APS), and the final four-wavelength MAD data used to solve the structure were collected from 15 different crystals, although hundreds had to be screened.

Fig. 5 shows the diffraction-dose efficiency as a function of energy for the three different cases: native crystal, sodium 12-tungstophosphate soak and a hypothetical SeMet derivative (eight Se atoms per monomer). The plot indicates that, had it been possible to produce equally well diffracting crystals of SeMet protein, the diffraction patterns collected from such a crystal could have been almost three times stronger than those from a tungsten-cluster-soaked crystal for a given absorbed dose at an energy above the selenium *K*-edge. The crystals that were soaked with the large $\text{Na}_3\text{O}_{40}\text{PW}_{12}$ clusters are calculated to absorb much more than SeMet-derivatized crystals; the minimum expected lifetime of the $\text{Na}_3\text{O}_{40}\text{PW}_{12}$ -soaked crystal at the tungsten L_{III} -absorption peak would be

78 s. The times needed to reach the Henderson limit at an energy of 9 keV (1.37 \AA) are shown in Table 1.

The maximum temperature rise as calculated with the lumped-body model at 13.0 keV is not significantly different from that calculated for the native DegP (7 K), whereas a large temperature increase is predicted for data collected at the tungsten L_{III} -edge. Just above this absorption edge, the predicted temperature rise is 25 K, whereas on the edge, this increases to more than 45 K.

4.3. Urate oxidase

Urate oxidase has been used by Girard *et al.* (2003) to test a new class of gadolinium complexes employed to obtain high-phasing-power heavy-atom derivatives, and the results were found to be promising. However, soaks are generally carried out at rather high concentrations (100 mM), since binding to the protein is relatively weak as a result of the neutral character of most of the gadolinium complexes.

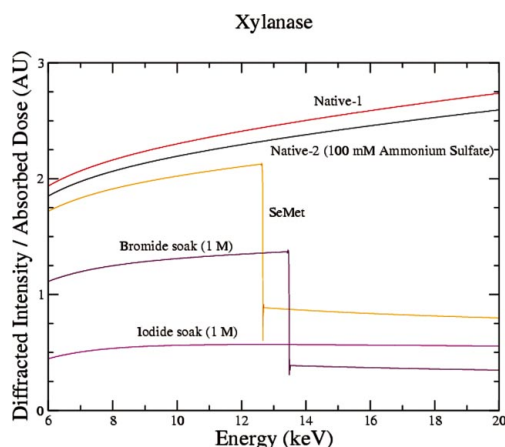
Fig. 6 shows the diffraction-dose efficiency as a function of energy for the native crystal, a 100 mM Gd soak and a hypothetical SeMet derivative. The gadolinium has an intense white line (Fig. 2), which could result in a very strong signal for phasing if required ($f'' = 28 e^-$).

The absorption of the beam by a $100 \times 100 \times 100 \mu\text{m}$ crystal can become as high as 25% of the incident power at the gadolinium L_{III} -absorption peak, whereas it would only be just over 3% at the selenium *K*-absorption maximum.

For this example, there were eight molecules of urate oxidase in the unit cell (space group $I222$). If the calculation is repeated with 16 molecules in the unit cell (implying a decrease in solvent content from 60 to 20%), the time available for data collection increases by 20% for the native crystal. However, for the SeMet crystal, the time available decreases by 10% at energies above the selenium *K*-edge (12.7 keV). Conversely, for the gadolinium-soaked crystal, the time available increases by 20%, as a result of the high concentration of gadolinium (100 mM) in the solvent channels.

4.4. Xylanase

Crystals of the protein xylanase were used by Dauter *et al.* (2000) to demonstrate that halide soaks can provide a convenient method for phasing macromolecular crystals. Bromine is an element almost as 'light' as selenium and might

**Figure 7**

Graph showing the diffraction-dose efficiency for native and derivative crystals of xylanase.

therefore affect the total X-ray absorption of a crystal less than heavier elements such as gadolinium, tantalum and tungsten. However, since in general the bromine sites are of low occupancy, a high concentration of bromide must be used.

Fig. 7 shows the diffraction–dose efficiency as a function of energy for two native crystals, a 1 M NaBr soak and a 1 M NaI soak, and for a hypothetical SeMet derivative (14 selenium sites). The difference between the two native crystals is the crystallization conditions; native-1 contains PEG, whereas native-2 contains 400 mM ammonium sulfate. The calculations for the halide soaks were carried out with ammonium sulfate present in the mother liquor, whereas for the SeMet crystal, the PEG mother liquor was used.

Fig. 7 clearly shows the impact of high concentrations of halide ions on the diffracted intensity per absorbed dose unit. There is almost a factor of four difference between the I soak and native-1 over the entire energy range. The absorption of a 100 μm -thick crystal soaked with 1 M NaI is significant, *e.g.* 8.2% at 13.0 keV, whereas the absorption is only 2.1% for native-1. The maximum temperature increase reflects the absorption: 29 K at 13.0 keV for the I soak and 7 K for native-1.

Note that the effect of ammonium sulfate in the mother liquor is significant and results, on average, in a 5% reduction of diffracted intensity per absorbed dose unit over the whole energy range.

For all the above protein crystals, the number of absorbed photons per unit cell for a total absorbed dose of 2×10^7 Gy has also been calculated. The results, which are insensitive to the exact contents of the unit cell, are shown in Table 2.

5. Discussion

In this paper we have described a computer program, *RADDose*, for the convenient prediction of the theoretical available time for data collection before radiation damage severely affects the sample.

It may be surprising that the absorption due to just one or two heavy atoms per protein molecule may be a significant fraction of the total absorption (see Figs. 3, 5, 6 and 7). As noted elsewhere (Garman & Murray, 2003), this makes back-soaking of heavy-atom derivative compounds highly desirable in order to remove excess heavy atoms from the solvent, although some heavy atoms cannot be back-soaked, *e.g.* xenon and weak binders such as halide ions. If radiation damage is a severe problem, perhaps sufficient phasing may be obtained with a lighter atom, for example as used in the structure determination of a tubulin-colchicine:RB3 complex (Ravelli *et al.*, 2004). Even sulfur in the solvent contributes significantly to absorption; a protein crystal with 60% solvent containing 2.4 M ammonium sulfate will have an absorption coefficient 1.4 times greater than that of a protein crystallized in PEG. However, heavy atoms may also have beneficial effects. In the case of LDLR, the crystals diffracted better when cocrystallized with the tungsten clusters. Thus radiation damage is not the only important factor to consider.

When considering the choice of compounds as potential free-radical scavengers for mitigating radiation damage in cryo-cooled crystals, Murray & Garman (2002) rejected some (*e.g.* those containing transition metal ions) on the basis of their increased contribution to absorption, which would partially or wholly negate any positive scavenging effect.

Henderson derived a dose limit that is independent of X-ray energy, and although comprehensive studies have not yet been performed, there is some evidence (González *et al.*, 1994) that the radiation dose limit is indeed independent of energy. Lower X-ray energies are, however, reputed to cause worse radiation damage per dose than higher energies. We would contend that much stronger evidence based on systematic studies is required before such a conclusion could be drawn. Figs. 3, 5, 6 and 7 all indicate that for native crystals the choice of a higher X-ray energy is favoured. However if a heavy atom is present, data collection above its absorption edge has serious consequences for the diffraction–dose efficiency. The true situation may be less pessimistic than has been presented here, as we have neglected the escape of some fluorescent X-ray photons, leading to an overestimate of absorbed dose.

Our dose limit model implies that the number of photons required to destroy the diffraction of a unit cell is proportional to its volume. It follows that roughly the same crystal lifetime is predicted for xylanase and LDLR crystals, even though the xylanase crystal unit cell has one twentieth of the volume of that of the LDLR. However, it is possible that the absorption of a small number of photons may disrupt the diffraction of a unit cell irrespective of its volume. *RADDose* provides information that would be useful for further studies on the importance of unit-cell volume for radiation damage.

The calculations presented here have implications for the design of automated pipelines for structure determination. A holistic approach to automated experimental design would incorporate a complete model of the crystallographic experiment and optimize all of the experimental parameters so that the information obtained from the experiment was maximized. This approach requires information about the crystal and mother liquor composition to be passed along the automated structure solution pipeline. After the initial indexing of the diffraction pattern to give the lattice parameters, a calculation of the expected crystal lifetime for these should be incorporated and passed on to the automated procedure to calculate the best data collection strategy (Popov & Bournekov, 2003) in the presence of radiation damage.

Strategies for anomalous data collection are numerous. The data may be collected with the crystal in a random orientation or using inverse beam geometry, and at a varying number of different energies (1 for SAD, 2, 3 or 4 for MAD). It has been claimed that it may be better to collect at the inflection point of the absorption spectrum first and then at a remote high energy (González, 2003). However this is a matter for debate, and Lehmann *et al.* (2003) suggest that it is better to collect the inflection point last. In this case, the loss of occupancy of the anomalous scatterer caused by radiation damage suffered at the first energy or energies will increase the size of the dispersive difference measured at the inflection point energy,

thus using radiation damage in a beneficial way to provide better phasing power (Ravelli *et al.*, 2003). In our examples, the elements which have a white line can dramatically increase the absorption at the peak energy for MAD experiments. For these cases, maximizing the dispersive rather than anomalous differences might provide better data with a lower radiation dose. In favourable cases the precise strategy might not be important, but in marginal cases, knowledge of the relative dose and signal at each energy would be beneficial.

Today, almost all X-ray diffraction data collection is carried out between 17.7 and 5 keV (0.7 and 2.5 Å). Certain very strong anomalous edges at low energy, *e.g.* the uranium M_{IV} -edge at 3.72 keV (3.326 Å), might allow easier phasing of large complexes (Liu *et al.*, 2001), but use of these low energies presents technical challenges, as special diffraction geometries and vacuum or helium paths must be present. In addition, at low energies, the crystal absorption is very large and only the surface of the crystal contributes to the diffraction. At higher energies, such as 30 keV, existing detectors are not sufficiently sensitive, but phosphor layers on detectors could be adapted to make them suitable for higher energies (Amemiya, 1995; Gruner *et al.*, 2002). At these energies our calculations for SeMet proteins show that there would not be a significant advantage for the diffraction–dose efficiency, but it is clear that there would be for native protein crystals.

The *RADDOSE* program makes use of the beam flux, which can be a difficult value to obtain. Such information is desirable, as ideally all parameters in the crystallographic experiment should be known or amenable to estimation. It is hoped that the use of *RADDOSE* will provide an incentive for users and beamline scientists to regularly calibrate the flux of the beamline (at different energies) and to evaluate the generality of ‘Henderson’s limit’.

If the crystal is smaller than the dimensions of the beam, *RADDOSE* requires the size of the crystal to be known. This requirement imposes a need for beamlines to provide a graticule with an absolute scale so that the size can be trivially measured. However, if the crystal is smaller than the beam, the dose is almost independent of crystal size.

In the examples presented here, the calculated maximum temperature rise is highest at absorption edges *e.g.* 54 K for DegP-Ta₆Br₁₄ at the Ta L_{III} -edge. If this heating results in even a transient crystal temperature above the solvent-glass transition around 155 K, the diffraction may be compromised (Weik *et al.*, 2001). Therefore if a high temperature increase is predicted, the beam should be attenuated to a level at which the predicted temperature rise is lower than 55 K.

Note that neither the theoretical predictions of crystal lifetime given by *RADDOSE* nor the validity of the ‘Henderson limit’ have yet been verified by systematic experiments. This is an obvious future step, of great interest to the community, but which is beyond the scope of this paper. Meanwhile however, we hope that by making *RADDOSE* available to crystallographers, they will more easily be able to identify the factors in their experiments that are likely to affect their crystal lifetimes. In addition, the feedback from

our users will help us to evaluate *RADDOSE* and our model of absorbed dose.

6. Conclusions

In the past 20 years the technology for X-ray data collection and processing has undergone a revolution, allowing the choices of phasing technique and phasing atoms to become much wider. These technologies are now being pipelined and automated. The automation process has led to a re-examination of data collection strategies and methods in general. In particular, we would advocate that a careful choice of X-ray energy allows the experimenter to maximize the information that can be obtained from each crystal. Thus we have written a computer program, *RADDOSE*, for convenient calculation of absorption cross sections for macromolecular crystals of arbitrary composition at differing X-ray energies. This program has already been useful to experimenters wanting to optimize their experiments, both in terms of beam-energy choices and in highlighting the merits of back-soaking to reduce the concentration of non-specifically bound anomalous scatterers. Programs such as *RADDOSE* may prove essential in the optimization of forthcoming beamline automation systems.

RADDOSE can be obtained from the authors.

JWM was supported by an MRC studentship and visited the EMBL Grenoble Outstation under the EU Improving Human Potential Access to Research Infrastructures Programme HPRI-CT-1999-00022. We thank Gabby Rudenko for information on LDLR, we thank all the staff of the EMBL Grenoble Outstation and the ESRF for support, maintenance and development of the JSBG beamlines, and gratefully acknowledge the ESRF for beam time under long-term project LS2047.

References

- Amemiya, Y. (1995). *J. Synchrotron Rad.* **2**, 13–21.
- Arndt, U. W. (1984). *J. Appl. Cryst.* **17**, 118–119.
- Badyopadhyay, P. (1995). ‘mucal’ subroutine. <http://ixs.csrii.iit.edu/database/programs/mcmaster.html>.
- Beteva, A., McCarthy, A., Shepard, W. & Spruce, D. (2003). *Automated Edge Scan Routines*. Unpublished results. ESRF.
- Blake, C. C. F. & Philips, D. C. (1962). *Biological Effects of Ionizing Radiation at the Molecular Level*, pp. 183–191. Vienna: International Atomic Energy Agency.
- Collaborative Computational Project, No. 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Cromer, D. T. & Liberman, D. (1970). *Phys. Rev.* **53**, 1891–1898.
- Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). *Acta Cryst.* **D56**, 232–237.
- Evans, G. & Pettifer, R. F. (2000). *J. Appl. Cryst.* **34**, 82–86.
- Garman, E. & Murray, J. (2003). *Acta Cryst.* **D56**, 1903–1913.
- Girard, E., Stelter, M., Anelli, P. L., Vicat, J. & Kahn, R. (2003). *Acta Cryst.* **D59**, 118–126.
- González, A. (2003). *Acta Cryst.* **D59**, 315–322.
- González, A., Denny, R. & Nave, C. (1994). *Acta Cryst.* **D50**, 276–282.
- Gruner, S. M., Tate, M. W. & Eikenberry, E. F. (2002). *Rev. Sci. Instrum.* **73**, 2815–2841.
- Henderson, R. (1990). *Proc. R. Soc. London Ser. B*, **241**, 6–8.

- Kriminski, S., Kazmierczak, M. & Thorne, R. E. (2003). *Acta Cryst.* **D59**, 697–708.
- Krojer, T., Garrido-Franco, M., Huber, R., Ehrmann, M. & Clausen, T. (2002). *Nature (London)*, **416**, 455–459.
- Kuzay, T. M., Kazmierczak, M. & Hsieh, B. J. (2001). *Acta Cryst.* **D57**, 69–81.
- Lehmann, C., Debreczeni, J. E., Bunkóczy, G., Dauter, M., Dauter, Z., Vértessy, L. & Sheldrick, G. M. (2003). *Helv. Chim. Acta*, **86**, 1478–1487.
- Leiros, H.-K. S., McSweeney, S. M. & Smalås, A. O. (2001). *Acta Cryst.* **D57**, 488–497.
- Liu, Y., Ogate, C. M. & Hendrickson, W. A. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 10648–10653.
- McMaster, W. H., Grande, N. K. D., Mallett, J. H. & Hubbell, J. H. (1969). *Compilation of X-Ray Cross Sections*, Vol. UCRL-50174, Section II, Revision I. Lawrence Livermore National Laboratory Report, National Technical Information Services L-3, US Department of Commerce.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Murray, J. & Garman, E. (2002). *J. Synchrotron Rad.* **9**, 347–354.
- Nave, C. (1995). *Radiat. Phys. Chem.* **45**, 483–490.
- O'Neill, P., Stevens, D. L. & Garman, E. F. (2002). *J. Synchrotron Rad.* **9**, 329–332.
- Polikarpov, I., Teplyakov, A. & Oliva, G. (1997). *Acta Cryst.* **D53**, 734–737.
- Popov, A. N. & Bourenkov, G. P. (2003). *Acta Cryst.* **D59**, 1145–1153.
- Ravelli, R. B. G., der Leiros, H.-K. S., Pan, B., Caffrey, M. & McSweeney, S. (2003). *Structure*, **11**, 217–224.
- Ravelli, R. B. G., Gigant, B., Curmi, P. A., Jourdain, I., Lachkar, S., Sobel, A. & Knossow, M. (2004). *Nature (London)*, **428**, 198–202.
- Ravelli, R. B. G. & McSweeney, S. M. (2000). *Structure*, **8**, 315–328.
- Ravelli, R. B. G., Theveneau, P., McSweeney, S. M. & Caffrey, M. (2002). *J. Synchrotron Rad.* **9**, 355–360.
- Rudenko, G., Henry, L., Henderson, K., Ichtchenko, K., Brown, M. S., Goldstein, J. L. & Deisenhofer, J. (2002). *Science*, **298**, 2353–2358.
- Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
- Shmueli, U. (2001). *International Tables for Crystallography*, Vol. B, *Reciprocal Space*, 2nd ed. Dordrecht: Kluwer Academic Publishers.
- Sliz, P., Harrison, S. C. & Rosenbaum, G. (2003). *Structure*, **11**, 13–19.
- Smith, J. M. A., Rice, D. W., White, J. L., Ford, G. C. & Harrison, P. M. (1989). *J. Appl. Cryst.* **22**, 284–286.
- Soyer, A. (1995). *J. Appl. Cryst.* **28**, 244.
- Spencer, J. E. & Winick, H. (1980). *Synchrotron Radiation Research*, edited by H. Winick & S. Doniach, ch. 21, *Wiggler Systems as Sources of Electromagnetic Radiation*. New York: Plenum Press.
- Weik, M., Kryger, G., Schreurs, A. M. M., Bauma, B., Silman, I., Sussman, J. L., Gos, P. & Kroon, J. (2001). *Acta Cryst.* **D57**, 566–573.
- Wimberly, B. T., Brodersen, D. E., Clemons, W. M. Jr, Morgan-Warren, R. J., Carter, A. P., Vornrhein, C., Hartsch, T. & Ramakrishnan, V. (2000). *Nature (London)*, **407**, 327–339.