research papers

On the routine use of soft X-rays in macromolecular crystallography. Part II. Data-collection wavelength and scaling models

Christoph Mueller-Dieckmann,[a] Maurizio Polentarutti,[b] Kristina Djinovic Carugo,[b] Santosh Panjikar,[a] Paul A. Tucker[a] and Manfred S. Weiss[a]*

[a]EMBL Hamburg Outstation, c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany, and [b]Structural Biology Laboratory, Sincrotrone Trieste in AREA Science Park, I-34102 Basovizza (TS), Trieste, Italy

Correspondence e-mail: msweiss@embl-hamburg.de

Complete and highly redundant data sets were collected at nine different wavelengths between 0.80 and 2.65 Å on a xenon derivative of porcine pancreatic elastase in both air and helium atmospheres. The magnitude of the anomalous signal, as assessed by the xenon-peak height in the anomalous difference Patterson synthesis, is affected by the wavelength of data collection as well as by the scaling model used. For data collected at wavelengths longer than 1.7 Å, the use of a three-dimensional scaling protocol is essential in order to obtain the highest possible anomalous signal. Based on the scaling protocols currently available, the optimal wavelength range for data collection appears to be between 2.1 and 2.4 Å. Beyond that, any further increase in signal will be compensated for or even superseded by a concomitant increase in noise, which cannot be fully corrected for. Data collection in a helium atmosphere yields higher $I/\sigma(I)$ values, but not significantly better anomalous differences, than data collection in air.

1. Introduction

In the past 2–3 y, the use of longer X-ray wavelengths (λ = 1.5–3.0 Å) has gained quite some popularity in macromolecular crystallography. This is mainly owing to the increased interest in utilizing the weak anomalous signals provided by S and P atoms present in native protein and nucleic acid molecules for phase determination. After the successful determination of the structure of crambin based on the anomalous scattering of the native S atoms (Hendrickson & Teeter, 1981), it took almost 20 years until this approach was rediscovered in the structure determination of the model protein lysozyme (Dauter *et al.*, 1999), although Wang (1985) had previously postulated that the method should be generally applicable. The structure of crambin was solved using diffraction data collected using Cu $K\alpha$ radiation (λ = 1.54 Å) on a conventional laboratory X-ray source, whereas lysozyme was solved using synchrotron radiation of the same wavelength. The photoprotein obelin was to our knowledge the first protein for which the structure was determined *de novo* by sulfur anomalous scattering at a somewhat longer wavelength (λ = 1.74 Å; Liu *et al.*, 2000). Other recent examples of successful structure determination by sulfur or other light-atom anomalous scattering include not only model proteins such as thermolysin at various wavelengths between 1.5 and 2.64 Å (Weiss, Sicker & Hilgenfeld, 2001), trypsin, insulin and thaumatin at λ = 1.54 Å (Debreczeni *et al.*, 2003) and glucose isomerase and xylanase at λ = 1.54 and 1.74 Å, respectively (Ramagopal *et al.*, 2003), but also real-life cases such as the IGF2R fragment at λ = 1.77 Å (Brown *et al.*, 2002), the C$_1$ subunit of α-crusta-

cyanin (Gordon *et al.*, 2001) the CAP-Gly domain at λ = 1.74 Å (Li *et al.*, 2002) and tryparedoxin at λ = 1.77 Å (Micossi *et al.*, 2002). Apart from increasing the anomalous scattering of the light atoms P, S, Cl, Ca *etc.*, the use of longer wavelengths may also provide advantages when dealing with xenon or iodide derivatives of proteins or with derivatives of the very heavy elements such as uranium. Cianci *et al.* (2001), for instance, used a diffraction data set collected at λ = 2.00 Å to aid in the structure determination of lobster apocrustacyanin A1.

Needless to say, there are also difficulties associated with the collection of diffraction data at longer wavelengths. Purely technical difficulties include, for instance, the physical mono-chromator limits, limitations caused by the energy spectrum of synchrotron sources or simply lower detector efficiency. Difficulties which are inherently connected to the experiment include the increased absorption of X-rays at longer wave-lengths, air scatter, the larger scattering angles and the occurrence of third harmonic reflections when Si(111) is used as the monochromator. In the absence of an experimental or an analytical absorption correction, the increased absorption certainly poses the largest problem. Fig. 1 illustrates schema-tically that one has to deal with four different absorption coefficients for the crystal (μ_1), the mother liquor (μ_2), the loop (μ_3) and air (μ_4). Only the absorption in air can be treated analytically during data processing; for the remaining three coefficients, one has to rely on the scaling and merging step. Nevertheless, in an initial study, Weiss, Sicker, Djinovic Carugo *et al.* (2001) were able to show that most of these difficulties can be dealt with. Without making time-consuming changes to a typical beamline setup, they were able to collect good diffraction data up to a wavelength of about 2.6 Å and, by employing a suitable processing protocol, managed to obtain decent anomalous differences even at these long
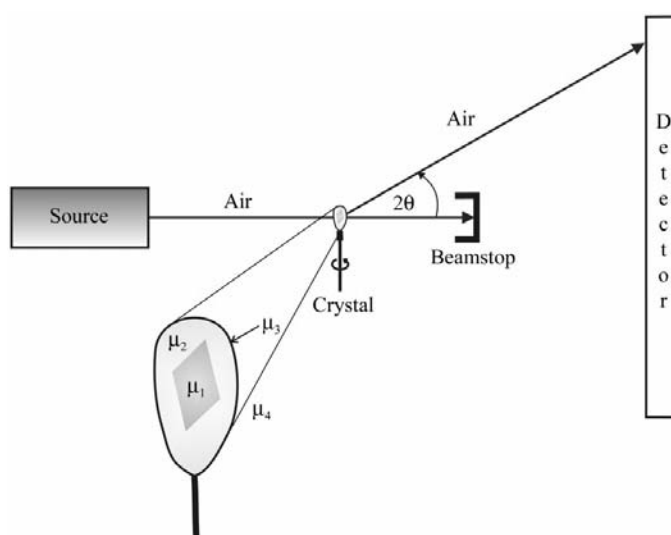
Figure 1
Schematic representation of the relevant absorption coefficients in a diffraction experiment. The four different coefficients are μ_1 for absorption in the crystal, μ_2 for absorption in the mother liquor surrounding the crystal, μ_3 for absorption by the nylon loop used and μ_4 for absorption in air (or helium).

wavelengths. Extending this first study, we would like to demonstrate here that processing the longer wavelength diffraction data requires the use of elaborated three-dimen-sional scaling protocols instead of the rather simple protocols which are often used in macromolecular crystallography. Furthermore, with the currently available scaling protocols, data-collection wavelengths of 2.1–2.4 Å yield the largest anomalous signal-to-noise ratio. At wavelengths longer than these, any further increase in signal will be compensated for or even superseded by a concomitant increase in noise, which the currently available scaling protocols are not able to fully correct for.

2. Materials and methods

2.1. Crystallization

Commercially available porcine pancreatic elastase (PPE; Serva Product No. 20929, Lot No. 12000) was used for crys-tallization without further purification. The protein was dissolved in deionized water to a concentration of 20 mg ml^{-1}. Crystals of PPE were grown at 293 K using a vapour-diffusion hanging-drop setup in the presence of 100 mM sodium acetate buffer pH 5.1 and 200 mM Na$_2$SO$_4$ as precipitant (Weiss *et al.*, 2002). These conditions yield crystals of the Na complex of PPE in a couple of days. The crystals belong to space group $P2_12_12_1$, with unit-cell parameters a = 50.2, b = 58.1, c = 74.3 Å. They typically diffract X-rays to better than 1.5 Å resolution.

2.2. Preparation of the xenon derivative

One crystal of PPE was transferred into cryosolution containing the reservoir liquid and 20%(v/v) glycerol, covered with paraffin oil and exposed to 1.4 MPa of xenon for 5 min in a home-made pressure chamber (Djinovic Carugo *et al.*, 1998). The crystal was then flash-cooled to 77 K in liquid nitrogen and subsequently transferred into a gaseous nitrogen stream at 100 K using an in-house mounting arc.

2.3. Data collection in air

Diffraction data sets were collected at the X-ray diffraction beamline at the Elettra Synchrotron Radiation Facility (Trieste, Italy) using a MAR CCD detector with a diameter of 165 mm. All relevant data-collection parameters are given in Table 1. The choice of data-collection wavelengths (Table 1 and Fig. 2) was made firstly to collect one data set at a very short wavelength (λ = 0.80 Å) and secondly to cover the whole wavelength range of the three xenon L absorption edges. The exposure time and attenuation were chosen to (i) utilize the full dynamic range of the detector without allowing too many overloaded reflections to occur on each image and (ii) keep the spindle-axis rotation speed close to approximately 6.0° min^{-1}. By doing so, a very crude normalization with respect to the energy spectrum of Elettra was achieved. In order to compensate for varying incident-beam intensity, the data collection was carried out in dose mode, which means that the speed of rotation of the spindle axis was made dependent on the incident-beam intensity. No special data-

Table 1
Data-collection and processing parameters.

(a) Air atmosphere.

Data set	A-Air ($\lambda = 0.80$ Å)	B-Air ($\lambda = 1.50$ Å)	C-Air ($\lambda = 1.70$ Å)	D-Air ($\lambda = 1.90$ Å)	E-Air ($\lambda = 2.10$ Å)	F-Air ($\lambda = 2.25$ Å)	G-Air ($\lambda = 2.40$ Å)	H-Air ($\lambda = 2.50$ Å)	I-Air ($\lambda = 2.65$ Å)
Crystal-to-detector distance (mm)	150	70	55	40	36	36	36	36	36
$\Delta\varphi$ (°)	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
No. of images	360	360	360	360	360	360	360	360	360
Attenuation (μm Al)	0	140	60	40	20	0	0	0	0
Total time (min)	137	104	107	115	105	94	117	105	107
Unit-cell parameters									
a (Å)	50.35	50.31	50.31	50.31	50.32	50.34	50.36	50.38	50.38
b (Å)	58.02	57.98	57.98	57.98	57.98	57.99	58.01	58.02	58.01
c (Å)	74.74	74.69	74.68	74.68	74.69	74.69	74.71	74.72	74.72
Mosaicity (°)	0.41	0.40	0.40	0.41	0.42	0.39	0.42	0.41	0.40
d_{min} (Å)	1.65	1.82	1.82	1.82	1.94	2.08	2.22	2.30	2.44

(b) Helium atmosphere.

Data set	B-He ($\lambda = 1.50$ Å)	C-He ($\lambda = 1.70$ Å)	D-He ($\lambda = 1.90$ Å)	E-He ($\lambda = 2.10$ Å)	F-He ($\lambda = 2.25$ Å)	G-He ($\lambda = 2.40$ Å)	H-He ($\lambda = 2.50$ Å)	I-He ($\lambda = 2.65$ Å)
Crystal-to-detector distance (mm)	70	55	40	36	36	36	36	36
$\Delta\varphi$ (°)	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
No. of images	360	360	360	360	360	360	360	360
Attenuation (μm Al)	140	80	60	40	20	0	0	0
Total time (min)	103	128	108	103	103	103	114	103
Unit-cell parameters								
a (Å)	50.43	50.44	50.46	50.45	50.47	50.50	50.52	50.51
b (Å)	58.07	58.07	58.07	58.06	58.07	58.08	58.08	58.08
c (Å)	74.79	74.80	74.81	74.80	74.82	74.84	74.86	74.86
Mosaicity (°)	0.37	0.39	0.39	0.38	0.39	0.41	0.40	0.40
d_{min} (Å)	1.82	1.82	1.82	1.94	2.08	2.22	2.30	2.44

collection strategy was employed. Since the plan was to collect 360° in 1° images for each data set, the starting spindle angle for the first data set was chosen randomly. For the later data sets, the same starting φ-angle was used. The time sequence of the data sets was from A-Air to I-Air according to Table 1 and
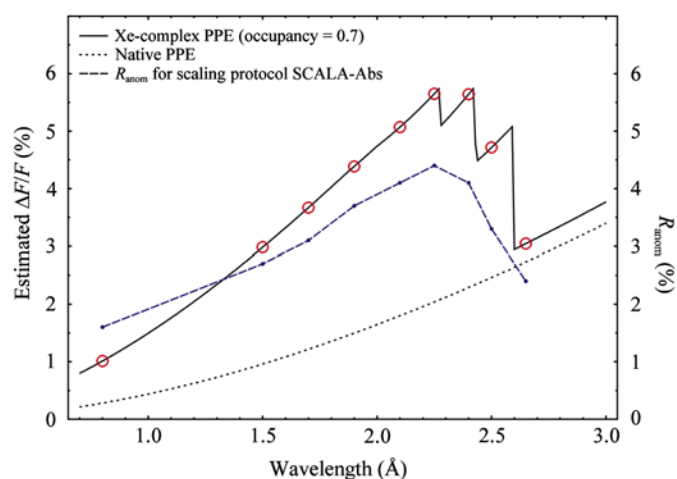
Figure 2
Estimated values for $\Delta F/F$ as a function of data-collection wavelength according to the equation given in Weiss, Sicker & Hilgenfeld (2001) for a single-site xenon derivative of PPE with an estimated xenon occupancy of 0.7 (solid line) and for native PPE (dotted line). The diffraction experiments conducted in this work are highlighted as red circles. The blue data points denote the observed values for the anomalous R factor R_{anom}.

Fig. 2, starting at the shortest wavelength and then moving towards the longer ones. Owing to the restrictions with respect to the shortest possible crystal-to-detector distance, the maximum resolution of the data sets collected at the longer wavelengths is lower. For data sets B-Air, C-Air, D-Air and E-Air the beam had to be attenuated to avoid the recording of overloaded reflections.

2.4. Data collection in a helium atmosphere

In order to assess the influence of the air scatter on the quality of the data, data sets B-He to I-He were collected from the same crystal. After dismounting the crystal from the beamline, the cold gas stream was switched to helium at 100 K, a windowless helium-purged beampath (Polentarutti & Djinovic Carugo, 2001, 2003; see also http://www.elettra.trieste.it/organization/experiments/laboratories/structural_biology/softxrays/phasing_with_soft_x.htm) was mounted on the beamline and the crystal was remounted, keeping the sample at cryotemperature during the entire operation. Owing to the cooling of the crystal with gaseous helium, a positive helium pressure was always present in the helium chamber, ensuring that no or little air would diffuse into the chamber. The time sequence of data collection was the same as for data collection in the air atmosphere, except that the data set at $\lambda = 0.80$ Å was skipped. All the geometric data-collection parameters ($\Delta\varphi$, crystal-to-detector distance) were kept the same, except that as a result of the dismounting and remounting procedure the starting φ-angle for the helium data sets was shifted by

about 16° with respect to the data sets collected in air. Furthermore, in contrast to the data sets collected in air, the data collection had to be carried out in time mode, *i.e.* a constant exposure time of 10 s was used for each 1° image irrespective of the incident-beam intensity. This was necessary because the presence of helium gas in the ionization chambers made the collection in dose mode impossible. As for the data sets collected in air, all relevant data-collection parameters are given in Table 1.

2.5. Data processing

All data sets were indexed and integrated using *DENZO* (Otwinowski & Minor, 1997). Care was taken to preserve the same axis setting for the indexing between data sets A-Air to I-Air and between data sets B-He to I-He. The post-refinement procedure in *SCALEPACK* (Otwinowski & Minor, 1997) was used to refine the unit-cell parameters and the mosaicity for each data set. For scaling and merging, two scaling programs, *SCALEPACK* (Otwinowski & Minor, 1997) and *SCALA* (Collaborative Computational Project, Number 4, 1994), were used, with a total of eight different scaling protocols. These protocols are as follows.

(i) SCALEPACK-kB. In this protocol, one constant scale factor k and one resolution-dependent scale factor B are determined and refined for each image. Friedel pairs are treated identically during scaling.

(ii) SCALEPACK-ano. Friedel pairs are scaled separately; otherwise, this protocol is identical to SCALEPACK-kB.

(iii) SCALA-Batch. A constant scale factor for each image and a resolution-dependent scale factor smoothed along the rotation axis used for data collection are determined. Except for the B-smoothing, this model is identical to SCALE-PACK-kB.

(iv) SCALA-Baniso. A constant scale factor and a resolution-dependent anisotropic scale factor is applied to each image. It is important to note that the anisotropic B factor determined here contains a centre of symmetry at the centre of the image.

(v) SCALA-Det. Here, the detector surface is divided into nine panels, each of which is given a separate scale factor. Neighbouring scale factors, both on the detector surface as well as between neighboring images, are restrained to similar values. The actual scale factor applied to each reflection is then calculated by three-dimensional interpolation. This scaling model is preceded by a SCALA-Batch run.

(vi) SCALA-Abs. In this model, the scale factor is allowed to vary as a function of the secondary beam direction in crystal coordinates expanded in spherical harmonics. Like SCALA-Det, this correction is three-dimensional and capable of correcting for absorption effects. SCALA-Abs is also preceded by a SCALA-Batch run.

(vii) SCALA-Sec. This protocol is very similar to SCALA-Abs (inclusive of the preceding SCALA-Batch run) except that the spherical harmonics expansion is performed in detector coordinates rather than crystal coordinates.

(viii) SCALA-Ref. In this case, the data set collected at $\lambda = 0.80$ Å was used as a reference data set to utilize local scaling procedures. It was combined with SCALA-Batch and SCALA-Abs.

In the case of *SCALEPACK* scaling, the adjustment of the error model was performed using the program *SCERROR* (V. Lamzin, personal communication). In this procedure, both the overall ERROR SCALE FACTOR as well as the resolution-dependent ESTIMATED ERROR parameters are adjusted simultaneously in order to obtain χ^2 values close to 1. When *SCALA* was used for scaling, the adjustment of the error model was performed automatically based on a normal probability analysis starting from default values for the parameters SDADD and SDFAC. Thus, it was ensured that for similar scaling models (*e.g.* SCALEPACK-kB and SCALA-Batch) the standard deviations in both cases are on the same scale and that they are similar. The redundancy-independent merging R factor $R_{r.i.m.}$ as well as the precision-indicating merging R factor $R_{p.i.m.}$ (Weiss, 2001) were calculated using the program *RMERGE* (available from http://www.embl-hamburg.de/~msweiss/projects/msw_qual.html or from MSW upon request).

2.6. Magnitude and usefulness of the anomalous signal

Fig. 2 shows the expected anomalous signal as expressed as the estimated $\Delta F/F$ for a PPE xenon derivative with one xenon site and an occupancy of the xenon of 0.7 (upper line) as well as for native PPE (lower line). As the first and most obvious measure of the magnitude of the anomalous signal the anomalous R factor R_{anom} was used, which describes the intensity differences between the Friedel mates of acentric reflections. An indicator for the usefulness of the anomalous signal for later application, *e.g.* phase determination, is the quotient $R_{anom}/R_{p.i.m.}$ (Weiss, Sicker & Hilgenfeld, 2001; Panjikar & Tucker, 2002), where R_{anom} describes the signal and $R_{p.i.m.}$ is a measure of the noise in the data. The peak height of the single Xe atom in the anomalous difference Patterson synthesis (see below) was used as a gauge of the magnitude of the obtained anomalous signal, which is independent of the statistics provided by the scaling and merging step of the data.

2.7. Anomalous difference Patterson synthesis

After scaling and merging, structure-factor amplitudes and anomalous amplitude differences were calculated using the program *TRUNCATE* (French & Wilson, 1978; Collaborative Computational Project, Number 4, 1994). The obtained anomalous amplitude differences were then used to calculate an anomalous difference Patterson synthesis in the program *FFTBIG* (Collaborative Computational Project, Number 4, 1994). In order to make the different data sets comparable, the maximum resolution for this calculation was set to 2.44 Å in all cases, which was the maximum resolution for data sets I-Air and I-He.

Table 2
Selected scaling statistics.

Full information concerning the total number of observations, the total number of reflections, the number of unique reflections, the number of rejected reflections, the redundancy, the completeness, the $I/\sigma(I)$ values, the merging R factor R_{merge}, the redundancy-independent merging R factor $R_{r.i.m.}$, the precision-indicating merging R factor $R_{p.i.m.}$ and the anomalous R factor R_{anom} can be found in the supplementary information to this paper. The values given for redundancy, $R_{p.i.m.}$ and R_{anom} describe each data set to its maximum resolution (see Table 1). The two numbers refer to the full 360° of data and to the first 180° only (in italics).

(*a*) Air atmosphere.

Scaling protocol		A-Air ($\lambda = 0.80$ Å)	B-Air ($\lambda = 1.50$ Å)	C-Air ($\lambda = 1.70$ Å)	D-Air ($\lambda = 1.90$ Å)	E-Air ($\lambda = 2.10$ Å)	F-Air ($\lambda = 2.25$ Å)	G-Air ($\lambda = 2.40$ Å)	H-Air ($\lambda = 2.50$ Å)	I-Air ($\lambda = 2.65$ Å)
SCALEPACK-kB	Redundancy	14.3 *7.2*	13.8 *6.9*	13.1 *6.6*	12.0 *6.0*	11.2 *5.6*	11.0 *5.5*	10.8 *5.4*	10.8 *5.4*	11.0 *5.5*
	$R_{p.i.m.}$ (%)	1.5 *2.0*	1.6 *2.3*	1.8 *2.6*	2.2 *3.2*	2.5 *3.7*	2.9 *4.2*	3.1 *4.5*	3.1 *4.7*	3.4 *5.1*
	R_{anom} (%)	1.6 *2.2*	2.7 *3.2*	3.1 *3.7*	3.7 *4.4*	4.1 *4.9*	4.4 *5.3*	4.2 *5.3*	3.5 *5.0*	2.7 *4.4*
SCALEPACK-ano	Redundancy	14.3 *7.2*	13.8 *6.9*	13.1 *6.6*	11.9 *6.0*	11.2 *5.6*	10.9 *5.5*	10.8 *5.4*	10.8 *5.4*	11.0 *5.5*
	$R_{p.i.m.}$ (%)	1.5 *2.0*	1.6 *2.3*	1.7 *2.5*	2.2 *3.2*	2.5 *3.7*	2.9 *4.2*	3.1 *4.5*	3.1 *4.7*	3.4 *5.1*
	R_{anom} (%)	1.6 *2.2*	2.7 *3.2*	3.1 *3.7*	3.7 *4.5*	4.1 *4.9*	4.4 *5.4*	4.2 *5.4*	3.6 *5.1*	2.8 *4.7*
SCALA-Batch	Redundancy	13.4 *6.7*	12.6 *6.3*	12.1 *6.1*	11.1 *5.6*	10.1 *5.1*	9.6 *4.9*	9.4 *4.8*	9.3 *4.7*	9.2 *4.6*
	$R_{p.i.m.}$ (%)	1.5 *2.0*	1.6 *2.3*	1.7 *2.5*	2.2 *3.2*	2.5 *3.7*	3.0 *4.3*	3.2 *4.6*	3.3 *4.9*	3.5 *5.3*
	R_{anom} (%)	1.6 *2.2*	2.7 *3.2*	3.1 *3.7*	3.8 *4.4*	4.2 *5.1*	4.7 *5.8*	4.5 *5.9*	3.8 *5.6*	3.4 *5.6*
SCALA-Baniso	Redundancy	13.4 *6.7*	12.6 *6.3*	12.1 *6.1*	11.1 *5.6*	10.1 *5.1*	9.6 *4.9*	9.4 *4.8*	9.3 *4.7*	9.2 *4.6*
	$R_{p.i.m.}$ (%)	1.5 *2.0*	1.5 *2.2*	1.6 *2.2*	1.9 *2.6*	12.1 *3.0*	2.4 *3.5*	2.6 *3.6*	2.5 *3.7*	2.6 *3.9*
	R_{anom} (%)	1.6 *2.2*	2.7 *3.1*	3.1 *3.5*	3.7 *4.1*	4.1 *4.6*	4.5 *5.1*	4.2 *5.0*	3.5 *4.4*	2.7 *4.1*
SCALA-Det	Redundancy	13.4 *6.7*	12.6 *6.3*	12.1 *6.1*	11.1 *5.6*	10.1 *5.1*	9.6 *4.8*	9.4 *4.8*	9.3 *4.7*	9.2 *4.6*
	$R_{p.i.m.}$ (%)	1.4 *2.0*	1.5 *2.0*	1.5 *2.0*	1.7 *2.4*	1.9 *2.7*	2.0 *2.9*	2.2 *3.0*	1.9 *2.8*	1.9 *2.8*
	R_{anom} (%)	1.6 *2.2*	2.7 *3.1*	3.1 *3.4*	3.7 *4.1*	4.1 *4.5*	4.4 *4.9*	4.1 *4.7*	3.3 *4.0*	2.5 *3.4*
SCALA-Abs	Redundancy	13.4 *6.7*	12.6 *6.3*	12.1 *6.1*	11.1 *5.6*	10.1 *5.1*	9.6 *4.8*	9.4 *4.8*	9.3 *4.7*	9.2 *4.6*
	$R_{p.i.m.}$ (%)	1.4 *2.0*	1.5 *2.0*	1.5 *2.1*	1.8 *2.4*	1.9 *2.7*	2.0 *2.9*	2.1 *3.0*	1.8 *2.7*	1.8 *2.7*
	R_{anom} (%)	1.6 *2.2*	2.7 *3.1*	3.1 *3.5*	3.7 *4.1*	4.1 *4.5*	4.4 *4.9*	4.1 *4.7*	3.3 *3.9*	2.4 *3.3*
SCALA-Sec	Redundancy	13.4 *6.7*	12.6 *6.3*	12.1 *6.1*	11.1 *5.6*	10.1 *5.1*	9.6 *4.8*	9.4 *4.8*	9.3 *4.7*	9.2 *4.6*
	$R_{p.i.m.}$ (%)	1.4 *2.0*	1.5 *2.0*	1.5 *2.1*	1.8 *2.4*	1.9 *2.7*	2.0 *2.9*	2.1 *3.0*	1.8 *2.7*	1.8 *2.7*
	R_{anom} (%)	1.6 *2.2*	2.7 *3.1*	3.1 *3.5*	3.7 *4.1*	4.1 *4.5*	4.4 *4.9*	4.1 *4.7*	3.3 *3.9*	2.4 *3.3*
SCALA-Ref	Redundancy	—	12.6 *6.3*	12.1 *6.1*	11.1 *5.6*	10.1 *5.1*	9.6 *4.8*	9.4 *4.8*	9.3 *4.7*	9.2 *4.6*
	$R_{p.i.m.}$ (%)	—	1.5 *2.1*	1.6 *2.2*	1.8 *2.4*	1.9 *2.7*	2.1 *3.0*	2.2 *3.2*	2.0 *3.0*	2.2 *3.2*
	R_{anom} (%)	—	2.7 *3.1*	3.1 *3.5*	3.7 *4.1*	4.1 *4.6*	4.5 *5.0*	4.2 *4.9*	3.4 *4.2*	2.7 *4.0*

(*b*) He atmosphere.

Scaling protocol		B-He ($\lambda = 1.50$ Å)	C-He ($\lambda = 1.70$ Å)	D-He ($\lambda = 1.90$ Å)	E-He ($\lambda = 2.10$ Å)	F-He ($\lambda = 2.25$ Å)	G-He ($\lambda = 2.40$ Å)	H-He ($\lambda = 2.50$ Å)	I-He ($\lambda = 2.65$ Å)
SCALEPACK-kB	Redundancy	13.7 *6.9*	12.5 *6.3*	11.9 *6.0*	11.8 *6.0*	11.7 *5.9*	10.8 *5.4*	11.6 *5.9*	11.7 *5.9*
	$R_{p.i.m.}$ (%)	1.0 *1.6*	1.4 *2.1*	1.7 *2.5*	2.0 *2.9*	2.3 *3.3*	3.1 *4.5*	2.6 *3.7*	2.8 *4.0*
	R_{anom} (%)	2.1 *2.3*	2.7 *3.0*	3.1 *3.5*	3.4 *3.9*	3.7 *4.3*	4.2 *5.4*	2.9 *3.8*	2.0 *3.3*
SCALEPACK-ano	Redundancy	13.7 *6.9*	12.4 *6.3*	11.9 *6.0*	11.8 *6.0*	11.6 *5.9*	10.8 *5.4*	11.6 *5.8*	11.7 *5.9*
	$R_{p.i.m.}$ (%)	1.0 *1.6*	1.4 *2.1*	1.7 *2.5*	1.9 *2.9*	2.3 *3.3*	3.1 *4.5*	2.6 *3.7*	2.8 *4.0*
	R_{anom} (%)	2.1 *2.2*	2.7 *3.0*	3.2 *3.5*	3.4 *3.9*	3.7 *4.4*	4.2 *5.4*	2.9 *3.9*	2.1 *3.3*
SCALA-Batch	Redundancy	12.9 *6.5*	11.6 *5.9*	11.2 *5.6*	10.9 *5.5*	10.6 *5.4*	10.3 *5.2*	10.1 *5.1*	10.1 *5.1*
	$R_{p.i.m.}$ (%)	1.0 *1.5*	1.4 *2.0*	1.7 *2.4*	2.0 *2.9*	2.3 *3.4*	2.7 *3.9*	2.8 *4.0*	2.9 *4.3*
	R_{anom} (%)	2.1 *2.3*	2.6 *3.0*	3.2 *3.5*	3.5 *4.0*	3.9 *4.6*	3.9 *4.9*	3.3 *4.5*	2.7 *4.1*
SCALA-Baniso	Redundancy	12.9 *6.5*	11.6 *5.9*	11.2 *5.6*	10.9 *5.6*	10.6 *5.4*	10.3 *5.2*	10.0 *5.1*	10.1 *5.1*
	$R_{p.i.m.}$ (%)	0.9 *1.4*	1.2 *1.7*	1.4 *2.0*	1.6 *2.3*	1.9 *2.7*	2.1 *3.0*	2.1 *3.0*	2.1 *3.0*
	R_{anom} (%)	2.1 *2.2*	2.6 *2.8*	3.1 *3.3*	3.4 *3.6*	3.7 *4.0*	3.7 *4.1*	3.0 *3.5*	2.2 *2.9*
SCALA-Det	Redundancy	12.9 *6.4*	11.6 *5.9*	11.2 *5.6*	10.9 *5.5*	10.5 *5.3*	10.2 *5.2*	10.0 *5.1*	10.0 *5.1*
	$R_{p.i.m.}$ (%)	0.8 *1.2*	1.1 *1.5*	1.2 *1.7*	1.3 *1.9*	1.5 *2.1*	1.6 *2.3*	1.4 *2.1*	1.3 *1.9*
	R_{anom} (%)	2.1 *2.2*	2.6 *2.8*	3.1 *3.2*	3.3 *3.5*	3.6 *3.8*	3.6 *3.8*	2.8 *3.1*	1.9 *2.3*
SCALA-Abs	Redundancy	12.9 *6.4*	11.6 *5.9*	11.2 *5.6*	10.9 *5.5*	10.5 *5.3*	10.2 *5.2*	10.0 *5.1*	10.1 *5.1*
	$R_{p.i.m.}$ (%)	0.9 *1.2*	11.1 *1.5*	1.2 *1.8*	1.3 *1.9*	1.5 *2.1*	1.5 *2.2*	1.4 *2.0*	1.2 *1.7*
	R_{anom} (%)	2.1 *2.2*	2.6 *2.8*	3.1 *3.3*	3.3 *3.6*	3.6 *3.9*	3.5 *3.8*	2.8 *3.1*	1.8 *2.2*
SCALA-Sec	Redundancy	12.9 *6.4*	11.6 *5.9*	11.2 *5.6*	10.9 *5.5*	10.5 *5.3*	10.2 *5.2*	10.0 *5.1*	10.1 *5.1*
	$R_{p.i.m.}$ (%)	0.9 *1.2*	1.1 *1.5*	1.2 *1.8*	1.3 *1.9*	1.5 *2.1*	1.5 *2.2*	11.4 *2.0*	1.2 *1.7*
	R_{anom} (%)	2.1 *2.2*	2.6 *2.8*	3.1 *3.3*	3.3 *3.6*	3.6 *3.9*	3.5 *3.8*	2.8 *3.1*	1.8 *2.2*
SCALA-Ref	Redundancy	12.9 *6.4*	11.7 *5.9*	11.2 *5.6*	10.9 *5.5*	10.6 *5.4*	10.3 *5.3*	10.0 *5.1*	10.1 *5.1*
	$R_{p.i.m.}$ (%)	0.9 *1.3*	1.1 *1.6*	1.3 *1.8*	1.4 *2.0*	1.6 *2.2*	1.7 *2.4*	1.6 *2.3*	1.5 *2.1*
	R_{anom} (%)	12.1 *2.2*	2.6 *2.8*	3.1 *3.3*	3.4 *3.6*	3.7 *4.0*	3.5 *3.9*	2.9 *3.3*	2.0 *2.6*

2.8. Refinement

In order to assess the influence of radiation damage on the xenon occupancy and therefore the magnitude of the anomalous signal, the structure of the PPE–Xe complex was first refined against the data set A-Air starting from the coordinate set 1lkb (Weiss *et al.*, 2002) and using the program *REFMAC* (Collaborative Computational Project, Number 4, 1994). The refined structure ($R = 17.6\%$, $R_{free} = 20.6\%$) contained all 240 amino acids of PPE, one Na^+, one Cl^-, two SO_4^{2-}, one Xe, one glycerol and 250 water molecules. This structure, along with

the processed data A-Air, has been deposited in the PDB (PDB code 1uo6). The Xe atom was then omitted and the remaining coordinates refined against all other data sets (scaled using the protocol SCALA-Ref) in order to use the absolute xenon-peak height in the $(2F_{obs} - F_{calc}, \alpha_{calc})$ map, the $(F_{obs} - F_{calc}, \alpha_{calc})$ map as well as the anomalous difference Fourier $(F_{obs}^+ - F_{obs}^-, \alpha_{calc} - 90°)$ map as an indicator of a possible loss of the anomalous signal owing to radiation damage. The absolute peak heights in units of e Å^{-3} were then normalized by division by $f(Xe) + \Delta f'(Xe)$ in the case of the

Table 3
Theoretical scattering coefficients $\Delta f'$ and $\Delta f''$ of xenon and sulfur in units of electrons at the nine data-collection wavelengths used in this work.

Wavelength (Å)	$\Delta f'$(Xe) (e)	$\Delta f''$(Xe) (e)	$\Delta f'$(S) (e)	$\Delta f''$(S) (e)
0.80	−0.2	2.4	0.2	0.2
1.50	−0.4	7.0	0.3	0.5
1.70	−1.1	8.6	0.4	0.7
1.90	−2.4	10.3	0.4	0.8
2.10	−4.5	11.8	0.4	1.0
2.25	−7.3	13.1	0.4	1.1
2.40	−10.2	12.9	0.4	1.2
2.50	−11.4	10.1	0.4	1.3
2.65	−12.1	3.5	0.3	1.5

$(2F_{obs} - F_{calc}, \alpha_{calc})$ and the $(F_{obs} - F_{calc}, \alpha_{calc})$ maps and by $\Delta f''$(Xe) in the case of the $(F_{obs}^+ - F_{obs}^-, \alpha_{calc} - 90°)$ map. In order to make these refinements and the corresponding Fourier summations comparable, they were all carried out in the resolution range 40.0–2.44 Å.

3. Results and discussion

3.1. Data collection, expected anomalous signal and radiation damage

The total data-collection time for the nine data sets A-Air to I-Air was approximately 16.5 h and the total exposure time was 10 h (Table 1). For data sets B-He to I-He, the total data-collection time was 14.5 h and the total exposure time was 8 h. It must therefore be expected that some radiation damage will have occurred during the experiment and that the expected anomalous signal (Fig. 2) will be affected by it. Except for data sets I-Air and I-He, the dominant source of the anomalous signal is the xenon bound in the active-site cleft of PPE. Since the absorption coefficient of xenon is much larger than that of all the other atoms in the structure, it can be expected that radiation damage will predominantly occur *via* the xenon site.

The first indication of radiation damage is the increase in the unit-cell parameters and the unit-cell volume (Ravelli *et al.*, 2002) as shown in Table 1 and in Fig. 3(*a*). The deviation of the first data point (data set A-Air) may be explained by an error in the exact determination of the wavelength. Overall, the unit-cell volume increased by about 0.8% from the second data set B-Air to the last data set I-He collected from this crystal. This is considerably less than has been observed in other examples which have been significantly damaged by radiation (Ravelli *et al.*, 2002). In contrast to the increase in the unit-cell parameters, the crystal mosaicity remained constant.

Nevertheless, from the reduction of the xenon-peak heights in the $(2F_{obs} - F_{calc}, \alpha_{calc})$, the $(F_{obs} - F_{calc}, \alpha_{calc})$ and the $(F_{obs}^+ - F_{obs}^-, \alpha_{calc} - 90°)$ maps during the course of the experiment, the radiation damage which occurred can be estimated to be about 35% over the course of all 17 data sets (Fig. 3*b*). This corresponds to a radiation damage of approximately 2% per data set. The reduction of the xenon-peak height should be accompanied by a concomitant loss of

Figure 3
Indications of radiation damage during the course of the experiment. (*a*) Change of the unit-cell volume as a function of the data set collected. The dotted straight line was fitted to the data points by linear regression. The highest correlation coefficient (0.98) was obtained when the first two data points were excluded from the fitting procedure. (*b*) Reduction of the xenon-peak height in the $(2F_{obs} - F_{calc}, \alpha_{calc})$ map (blue line), the $(F_{obs} - F_{calc}, \alpha_{calc})$ map (green line) as well as the anomalous difference Fourier $(F_{obs}^+ - F_{obs}^-, \alpha_{calc} - 90°)$ map (brown line). The peak heights have been normalized with respect to the estimated peak height and with respect to the actual peak height observed in data set A-Air. The straight line was fitted to the data points from all three maps by linear regression. The correlation coefficient was −0.78.

Table 4
R_{anom}/R_{pim} values and xenon-peak heights.

The numbers given are to a maximum resolution of 2.44 Å only. The two numbers refer to the full 360° of data and to the first 180° (in italics).

(*a*) Air atmosphere.

Scaling protocol		A-Air (λ = 0.80 Å)	B-Air (λ = 1.50 Å)	C-Air (λ = 1.70 Å)	D-Air (λ = 1.90 Å)	E-Air (λ = 2.10 Å)	F-Air (λ = 2.25 Å)	G-Air (λ = 2.40 Å)	H-Air (λ = 2.50 Å)	I-Air (λ = 2.65 Å)
SCALEPACK-kB	$R_{anom}/R_{p.i.m.}$	1.2 *1.1*	1.7 *1.4*	1.8 *1.5*	1.6 *1.4*	1.7 *1.3*	1.6 *1.3*	1.3 *1.2*	1.1 *1.0*	0.8 *0.9*
	Xe-peak height (σ)	6.8 *4.1*	17.6 *11.4*	19.3 *12.3*	118.3 *12.6*	19.4 *14.3*	18.8 *13.9*	16.7 *11.3*	13.0 *7.4*	3.7 *3.2*
SCALEPACK-ano	$R_{anom}/R_{p.i.m.}$	1.2 *1.1*	1.7 *1.4*	1.8 *1.5*	1.7 *1.4*	1.7 *1.3*	1.6 *1.3*	1.4 *1.2*	1.1 *1.1*	0.8 *0.9*
	Xe-peak height (σ)	6.9 *4.0*	17.6 *11.4*	18.9 *12.7*	17.8 *12.1*	18.5 *13.6*	18.3 *13.4*	16.5 *11.1*	12.9 *7.6*	3.5 *3.4*
SCALA-Batch	$R_{anom}/R_{p.i.m.}$	1.2 *1.1*	1.7 *1.4*	1.8 *1.4*	1.8 *1.4*	1.7 *1.4*	1.6 *1.3*	1.4 *1.3*	1.2 *1.1*	1.0 *1.1*
	Xe-peak height (σ)	6.8 *4.2*	17.5 *11.3*	19.0 *14.0*	18.3 *13.9*	18.8 *14.3*	14.6 *11.8*	14.6 *8.9*	9.6 *5.4*	3.5 *3.4*
SCALA-Baniso	$R_{anom}/R_{p.i.m.}$	1.2 *1.1*	1.8 *1.5*	1.9 *1.5*	1.9 *1.6*	2.0 *1.5*	1.9 *1.5*	1.6 *1.4*	1.4 *1.2*	1.0 *1.1*
	Xe-peak height (σ)	6.9 *4.1*	17.6 *11.7*	18.9 *14.5*	18.2 *14.3*	19.3 *15.7*	17.9 *14.0*	16.4 *11.6*	12.0 *7.7*	3.5 *3.6*
SCALA-Det	$R_{anom}/R_{p.i.m.}$	1.2 *1.1*	1.8 *1.5*	2.1 *1.7*	2.1 *1.7*	2.3 *1.7*	2.2 *1.7*	1.9 *1.6*	1.7 *1.4*	1.3 *1.2*
	Xe-peak height (σ)	6.8 *4.2*	17.8 *11.9*	19.2 *14.7*	19.1 *13.8*	19.6 *16.3*	19.2 *15.0*	17.8 *13.0*	13.7 *10.4*	4.1 *<3.0*
SCALA-Abs	$R_{anom}/R_{p.i.m.}$	1.2 *1.1*	1.8 *1.5*	2.1 *1.7*	2.1 *1.7*	2.3 *1.7*	2.3 *1.7*	2.0 *1.6*	1.8 *1.5*	1.3 *1.2*
	Xe-peak height (σ)	6.8 *4.3*	17.8 *11.8*	19.2 *14.5*	19.0 *13.6*	19.9 *16.2*	19.7 *15.3*	18.2 *13.7*	13.9 *11.0*	4.2 *3.9*
SCALA-Sec	$R_{anom}/R_{p.i.m.}$	1.2 *1.1*	1.8 *1.5*	2.1 *1.7*	2.1 *1.7*	2.3 *1.7*	2.3 *1.7*	2.0 *1.6*	1.8 *1.5*	1.3 *1.2*
	Xe-peak height (σ)	6.8 *4.3*	17.8 *11.8*	19.2 *14.5*	19.0 *13.6*	19.9 *16.3*	19.7 *15.4*	18.2 *13.8*	13.9 *11.1*	4.3 *4.0*
SCALA-Ref	$R_{anom}/R_{p.i.m.}$	—	1.9 *2.2*	2.2 *1.8*	2.3 *1.9*	2.3 *1.8*	2.3 *1.9*	2.0 *1.6*	1.8 *1.5*	1.2 *1.3*
	Xe-peak height (σ)	—	17.9 *11.9*	19.4 *14.6*	19.3 *14.3*	20.3 *16.6*	20.6 *16.2*	19.0 *14.5*	13.0 *10.0*	3.9 *3.8*

(*b*) Helium atmosphere.

Scaling protocol		B-He (λ = 1.50 Å)	C-He (λ = 1.70 Å)	D-He (λ = 1.90 Å)	E-He (λ = 2.10 Å)	F-He (λ = 2.25 Å)	G-He (λ = 2.40 Å)	H-He (λ = 2.50 Å)	I-He (λ = 2.65 Å)
SCALEPACK-kB	$R_{anom}/R_{p.i.m.}$	2.3 *1.6*	2.0 *1.5*	1.9 *1.5*	1.8 *1.4*	1.6 *1.3*	1.4 *1.2*	1.1 *1.0*	0.7 *0.8*
	Xe-peak height (σ)	20.0 *18.3*	17.4 *14.9*	10.8 *15.2*	17.3 *15.4*	18.2 *15.3*	14.0 *13.1*	13.1 *8.1*	3.6 *<3.0*
SCALEPACK-ano	$R_{anom}/R_{p.i.m.}$	2.2 *1.7*	2.0 *1.5*	2.0 *1.5*	1.8 *1.4*	1.6 *1.3*	1.4 *1.2*	1.1 *1.0*	0.8 *0.9*
	Xe-peak height (σ)	19.8 *18.0*	17.2 *14.2*	10.0 *14.3*	16.6 *14.5*	17.5 *14.1*	13.6 *12.5*	13.0 *7.6*	3.6 *3.5*
SCALA-Batch	$R_{anom}/R_{p.i.m.}$	2.7 *1.7*	2.1 *1.7*	2.0 *1.6*	1.9 *1.5*	1.7 *1.4*	1.4 *1.3*	1.2 *1.1*	0.9 *1.0*
	Xe-peak height (σ)	20.6 *19.2*	19.8 *18.2*	18.3 *17.2*	18.8 *16.3*	18.3 *14.9*	10.0 *9.7*	10.4 *5.2*	<3.0 *<3.0*
SCALA-Baniso	$R_{anom}/R_{p.i.m.}$	2.5 *1.8*	2.3 *1.7*	2.3 *1.7*	2.2 *1.7*	2.0 *1.6*	1.7 *1.4*	1.4 *1.2*	1.1 *1.3*
	Xe-peak height (σ)	20.7 *19.6*	19.8 *18.7*	18.6 *17.7*	19.4 *17.7*	18.7 *17.0*	12.6 *13.5*	11.3 *8.1*	<3.0 *3.4*
SCALA-Det	$R_{anom}/R_{p.i.m.}$	2.5 *1.9*	2.4 *1.9*	2.6 *1.9*	2.7 *2.0*	2.5 *1.9*	2.2 *1.7*	2.0 *1.6*	1.5 *1.2*
	Xe-peak height (σ)	20.9 *19.1*	20.1 *18.3*	18.8 *17.3*	19.5 *17.4*	18.8 *17.2*	16.0 *15.0*	13.5 *10.3*	3.6 *<3.0*
SCALA-Abs	$R_{anom}/R_{p.i.m.}$	2.5 *1.9*	2.4 *1.9*	2.6 *1.9*	2.7 *1.9*	2.5 *1.9*	2.3 *1.8*	2.1 *1.6*	1.5 *1.3*
	Xe-peak height (σ)	20.9 *19.1*	20.3 *18.2*	19.0 *17.2*	19.7 *17.1*	18.9 *16.8*	17.4 *15.2*	14.3 *11.5*	4.3 *<3.0*
SCALA-Sec	$R_{anom}/R_{p.i.m.}$	2.5 *1.9*	2.4 *1.9*	2.7 *1.9*	2.7 *1.9*	2.5 *1.9*	2.3 *1.8*	2.1 *1.6*	1.5 *1.3*
	Xe-peak height (σ)	20.9 *19.1*	20.3 *18.2*	19.0 *17.2*	19.7 *17.1*	18.9 *16.9*	17.4 *15.3*	14.2 *11.5*	4.4 *3.6*
SCALA-Ref	$R_{anom}/R_{p.i.m.}$	2.5 *1.9*	2.5 *1.9*	2.7 *1.9*	2.7 *2.0*	2.6 *1.9*	2.3 *1.7*	1.9 *1.5*	1.3 *1.2*
	Xe-peak height (σ)	20.7 *19.8*	21.1 *18.8*	20.0 *18.3*	21.3 *18.9*	20.8 *19.2*	19.5 *17.0*	16.1 *13.0*	<3.0 *3.7*

the anomalous signal. Interestingly, the reduction is almost linear and relatively smooth over the whole course of the experiment. This suggests that in this respect the signal obtained from the longer wavelength data sets is affected in a similar manner by data processing and scaling than the signal obtained from the shorter wavelength data sets. Otherwise, one would have expected a larger relative signal from data set B-He than from I-Air. As a consequence of the radiation damage, the results obtained in the later data sets ought to be corrected by the respective extent of signal loss. However, since it is not obvious how to perform such a correction properly we refrained from doing so.

3.2. Data processing

Since each set of integrated raw intensities was scaled and merged according to the eight protocols described in §2 and as a function of data redundancy, a total of 270 reduced sets of structure-factor amplitudes were produced. Table 2 contains the most important scaling and merging statistics for data sets A-Air to I-Air and B-He to I-He. A complete listing of the statistical descriptors for all data sets is provided in the supplementary material.[1]

The redundancy of the data sets (Table 2) should ideally be all the same for the 17 data collections. However, owing to the obstruction of part of the detector surface by the shadow of the cryosystem nozzle at the short crystal-to-detector distances which were used at the longer wavelengths, the data sets at the longer wavelengths exhibit a lower data redundancy. At the shortest crystal-to-detector distances of 36 mm for data sets E-Air to I-Air and E-He to I-He, the total detector surface obstructed by the nozzle approaches 20%. In addition, scaling the data with *SCALEPACK* resulted in a much larger initial number of observations than scaling them using *SCALA*. The reason for this is that *SCALEPACK* performs post-refinement and therefore includes observations

---

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: DZ0018). Services for accessing these data are described at the back of the journal.

located at the edge of an image with a partiality of 0.000. This is to allow the orientation matrix to change during the post-refinement procedure, which may result in a concomitant increase of the partiality of the observation. Since *SCALA* does not perform any post-refinement operations, these observations are excluded even before the start of the scaling procedure (Eleanor Dodson, personal communication).

As expected from counting statistics, the values for the precision-indicating merging R factor $R_{\text{p.i.m.}}$ are smaller for the 360° data sets compared with those for the 180° data sets (Table 2). This reflects the reduced noise in the averaged measurements and the increased accuracy obtained by including the extra 180° of data. $R_{\text{p.i.m.}}$ also increases with data-collection wavelength (Table 2). This means that the precision

of the averaged intensities is lower if the data are collected at longer wavelengths. Owing to the increased absorption at these longer wavelengths, this effect is expected and can be explained by the notion that absorption can only partially be taken into account during scaling. Nevertheless, the utilization of an elaborate three-dimensional scaling protocol leads to a significant decrease of the values for $R_{\text{p.i.m.}}$, especially in the longer wavelength data sets. The simpler and more conventional scaling protocols do not lead to a comparable decrease of $R_{\text{p.i.m.}}$, which reflects the need to use such elaborate protocols for the scaling of data which are affected significantly by absorption.

In any case, the values for the data redundancy of 9–14 for the data sets comprising the full 360° of data and 5–7 for the first 180°, for a completeness of 98–100% and merging R factors of 5.0 and 11.4%, demonstrate that all data sets are of good quality. Nevertheless, there are also significant differences. For instance, data sets A-Air to I-Air, all scaled and merged using the protocol SCALA-Abs, exhibit R factors on structure-factor amplitudes calculated to a maximum resolution of 2.44 Å ranging from 1.3% (A-Air *versus* B-Air) to 10.7% (A-Air *versus* I-Air). These differences should not only be attributed to the dispersive differences caused by the Xe atom as exemplified by the different values of $\Delta f'$ (Table 3), but also to radiation damage (Fig. 3). The nine different scaling protocols lead to R factors between the scaled and merged F-Air data sets of 0.2–6.1%. The magnitude of these differences is significantly larger than the $R_{\text{p.i.m.}}$ values (Table 2), thus showing that the different scaling protocols can really cause significant differences. Finally, the data sets F-Air and F-He exhibit R factors of about 8.8% to each other
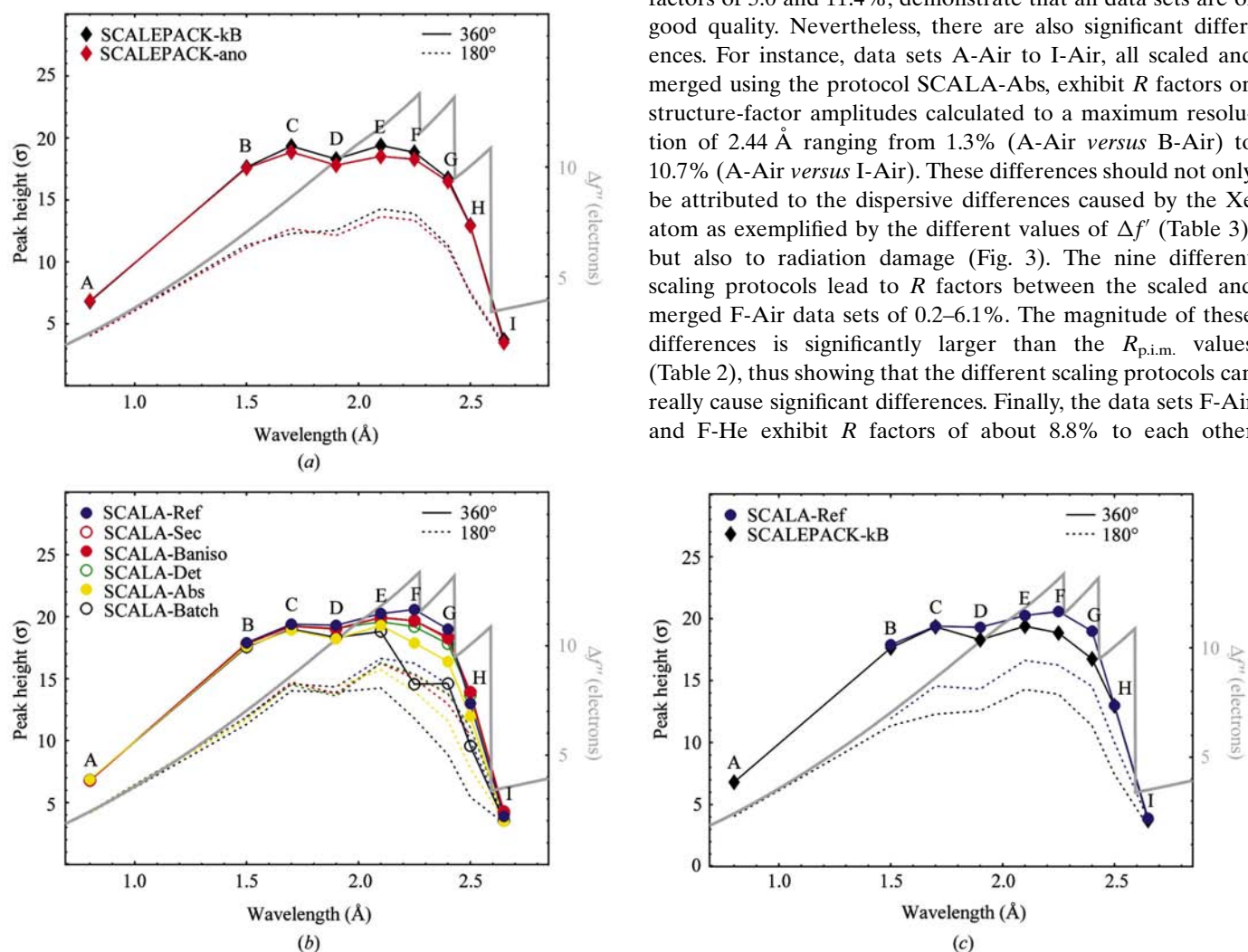
Figure 4
Peak height of the xenon site in an anomalous difference Patterson synthesis using data collected in air atmosphere given in units of σ above the mean value of the map as a function of the data-collection wavelength, the scaling protocol used and whether the full data set (360°) was used (solid lines) or just the first 180° of each data set (broken lines). In each of the three panels, the respective data sets have been labelled A–I and the theoretical $\Delta f''$ values of xenon are added as a grey line as an indicator for the expected signal. The values for $\Delta f''$ are based upon the theoretical approximation developed by Cromer & Liberman (1970). They were retrieved from the internet site of the Biomolecular Structure Center at the University of Washington, Seattle, USA (http://www.bmsc.washington.edu/scatter/AS_periodic.html). (*a*) The two *SCALEPACK* scaling models: SCALEPACK-kB in black and SCALEPACK-ano in red. (*b*) Various SCALA scaling models: SCALA-Ref (blue), SCALA-Sec (red), SCALA-Baniso (orange), SCALA-Det (green), SCALA-Abs (yellow) and SCALA-Batch (black). (*c*) The best *SCALEPACK* scaling model (SCALEPACK-kB, black symbols) *versus* the best *SCALA* scaling mode (SCALA-Ref, blue symbols).

depending on the protocol used for scaling and merging. This difference is a consequence of the reduced air scatter in helium atmosphere, but also of the difference in the two data sets caused by radiation damage (Fig. 3).

3.3. Magnitude of the anomalous signal

The obvious primary indicator of the magnitude of the anomalous signal is the anomalous R factor R_{anom}. The source of the observed anomalous signal is the Xe atom bound to PPE, the ten S atoms of PPE as well as two sulfate and one chloride ion bound to PPE. The estimated $\Delta F/F$ curves shown in Fig. 2 and the values of $\Delta f'$ and $\Delta f''$ for Xe and S presented in Table 3 demonstrate the relative contribution of the xenon to the total signal. As expected and as can be seen from the data presented in Table 2 and in Fig. 2, R_{anom} does approximately follow the estimated $\Delta F/F$ curve of a single-site xenon complex of PPE. The R_{anom} values for the 360° data sets are about 10–20% smaller compared with those for the 180° data sets. Together with the decreased $R_{p.i.m.}$ values, this reflects the reduced noise in the anomalous signal rather than the gradual loss of anomalous signal. With respect to the different scaling protocols the obtained R_{anom} values are not much different. The comparison of the data sets B-He to I-He with data sets B-Air to I-Air shows that the R_{anom} values for the data collected in a helium atmosphere are slightly smaller that those for data collected in an air atmosphere. This may be explained by some loss of some anomalous signal owing to radiation damage (Fig. 3), but also by the fact that data collection in a helium atmosphere apparently leads to some reduction of noise in the data mainly by reducing the air scatter.

As has been discussed in a qualitative manner before, a better indicator of the usefulness of the anomalous signal for later applications, *e.g.* phase determination, is the quotient $R_{anom}/R_{p.i.m.}$ (Weiss, Sicker & Hilgenfeld, 2001; Panjikar & Tucker, 2002). In order to make the different data sets comparable, the values for $R_{anom}/R_{p.i.m.}$ presented in Table 4 have been computed to the same maximum resolution (2.44 Å).

As a third measure, the magnitude of the obtained anomalous signal was gauged by the peak height of the single Xe atom in the structure in the anomalous difference Patterson synthesis. This may not be an optimal measure, since it mainly depends on the relative scaling of the $I(hkl)$ *versus* the $I(\overline{hkl})$ measurements and not so much on the relative scaling of the different $I(hkl)$ with respect to each other. Nevertheless, this is the first value that one can obtain after the scaling and merging stage and it is for this reason that we chose to use it. From the data presented in Table 4, it becomes apparent that the behaviour of the quotient $R_{anom}/R_{p.i.m.}$ matches the xenon-peak height fairly closely. The correlation coefficient between the $R_{anom}/R_{p.i.m.}$ values for all 360° data sets in this work and the corresponding xenon-peak heights is 0.81. In contrast, the correlation coefficient between R_{anom} and the xenon-peak height is only 0.40. Thus, $R_{anom}/R_{p.i.m.}$ is a significantly better predictor of the magnitude of the anomalous signal than

R_{anom}, which corroborates the qualitative statements made earlier (Weiss, Sicker & Hilgenfeld, 2001; Panjikar & Tucker, 2002).

Fig. 4 shows the xenon-peak height in the anomalous difference Patterson syntheses as a function of wavelength, scaling protocol and redundancy. In the top panel (Fig. 4a), the two *SCALEPACK*-based scaling protocols are illustrated. The xenon peak height increases as the data-collection wavelength increases, but not as much as one would expect from the wavelength-dependence of $\Delta f''(Xe)$. The signal increases up to data set C-Air and then levels off. An explanation for this is that any additional signal provided by wavelengths longer than 1.70 Å is compensated for by additional noise which the scaling protocols cannot correct for. Also, it is clear that the protocol SCALEPACK-kB is superior to the protocol SCALEPACK-ano, except for data sets A-Air and B-Air, which are little affected by absorption, and data sets H-Air and I-Air, in which the absorption problem is so bad that both protocols fail equally.

Fig. 4(b) shows the effect of the six *SCALA*-based scaling protocols. As with the *SCALEPACK*-based protocols, there is hardly any difference between them for data sets A-Air and B-Air, which are those least affected by absorption. At wavelengths longer than 1.70 Å, the different protocols yield different results. The best of all in terms of the magnitude of the anomalous signal is SCALA-Ref, in which the long-wavelength data set is scaled locally to the data set A-Air. SCALA-Ref is followed by SCALA-Sec and SCALA-Baniso; these are followed by SCALA-Det, SCALA-Abs and
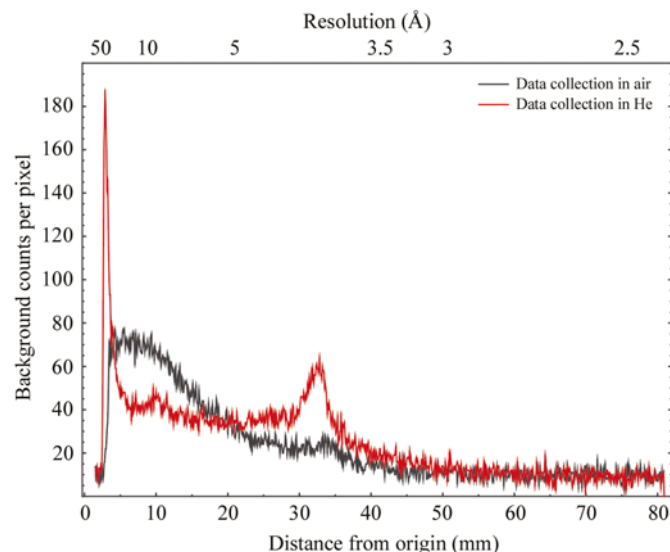
Figure 5
Background scattering as a function of the distance from the detector centre for a selected diffraction image of data sets I-Air (black line) and I-He (red line). The background is not radially averaged, but just plotted on a line from the detector centre to the right rim of the detector. The increased air scatter in data set I-Air leads to an increase in background radiation. The effect is most pronounced in the central part of the images. The high number of counts close to the centre for data set I-He is owing to a slight misalignment of the beamstop. The increased background at 3.5 Å resolution in the same data set results from the appearance of an ice ring (see also Fig. 6).

SCALA-Batch, which is the least effective protocol of those tried. At data set C-Air (λ = 1.70 Å), the different protocols start to make a difference. The difference increases up to data set F-Air (λ = 2.25 Å) and then starts to decrease again. Interestingly, the signal keeps increasing for the protocol SCALA-Ref up to data set F-Air (λ = 2.25 Å), whereas for the other *SCALA* protocols except SCALA-Batch the maximum is observed at data set E-Air (λ = 2.10 Å). The signal in case of the SCALA-Batch protocol starts to decrease again after data set C-Air (λ = 1.90 Å).

The bottom panel (Fig. 4*c*) shows the behaviour of the best *SCALA* protocol, SCALA-Ref, against the best *SCALE-PACK* protocol, SCALEPACK-kB. In the wavelength range 1.70–2.25 Å, SCALA-Ref yields a clearly higher anomalous signal.

In a previous study on the Zn-metalloprotease thermolysin (Weiss, Sicker & Hilgenfeld, 2001), the wavelength at which the anomalous signal was maximized was 1.9 Å. The observation that here the 'best' wavelength appears to be shifted to slightly higher values of between 2.1 and 2.4 Å may be attributed to the fact that in the case of the PPE–Xe complex the absolute increase in signal is larger when going from a wavelength of 1.9 to that of 2.1 Å than in the case of thermolysin. This, taken together with the assumption that the increase in noise is approximately comparable in both cases, will lead to the observed shift of the optimal wavelength.

3.4. Reduction of air scatter in a helium atmosphere

In order to assess the influence of the air scatter on the quality of the data, the data sets B-He to I-He were collected from the same crystal. Fig. 5 shows the background of two selected images from data sets I-Air and I-He from the centre of the detector to the right-hand rim. The two images were chosen so that they contain approximately the same slice of
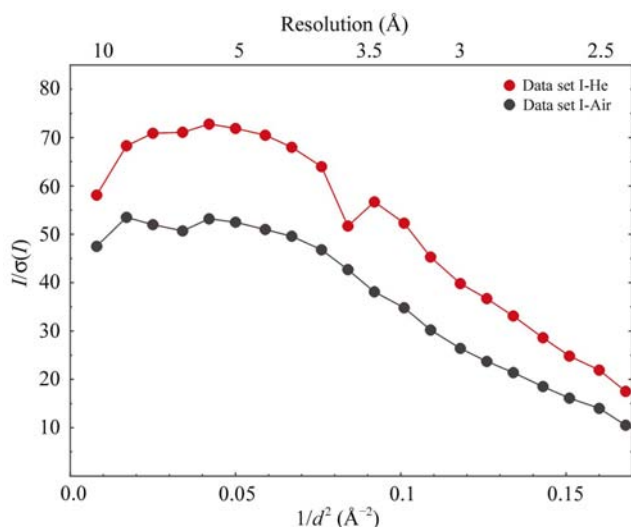
Figure 6
$I/\sigma(I)$ as a function of resolution ($1/d^2$) for the PPE–Xe data set I-Air (black lines) and I-He (red lines). The dip at 3.5 Å resolution in the upper curve is a result of the appearance of an ice ring on the images of the later data sets.

reciprocal space. Even though an ice ring had developed during the experiment and is visible in the helium curve, it is clearly evident that the background in the centre of the image is almost twice as high for data sets I-Air as for data set I-He. The explanation for this phenomenon is air scatter.

The reduction of air scatter by using the helium chamber for data collection leads to increased $I/\sigma(I)$ values (Fig. 6). For data set I-He $I/\sigma(I)$ is about 40% higher at low resolution and about 75% higher at high resolution than for data set I-Air. This increase in $I/\sigma(I)$ is reflected in a weak but consistent increase of the magnitude of the anomalous signal as estimated by the $R_{\text{anom}}/R_{\text{p.i.m.}}$ ratio and the xenon-peak height in the anomalous difference Patterson synthesis (Table 4), even when one does not take into account that the signal is reduced by about 17% as a result of radiation damage (Fig. 3). It is conceivable that this observed signal increase would have been larger if the data sets B-He to I-He had been collected before data sets A-Air to I-Air. The observation suggests that the use of the helium-purged beampath may help to obtain a better signal or a signal which is usable to higher resolution when data are collected at longer wavelengths. This effect is expected to be more pronounced for crystals that diffract less strongly than PPE crystals.

4. Summary and conclusions

The previously demonstrated fact that diffraction data collection at longer wavelengths is feasible on a routine basis without having to make time-consuming changes to the beamline setup is also confirmed in this study.

The magnitude of the anomalous signal as assessed by the quotient $R_{\text{anom}}/R_{\text{p.i.m.}}$ or the xenon-peak height in the anomalous difference Patterson synthesis is affected by the wavelength of data collection as well as by the scaling model used. For data collected at wavelengths longer than 1.7 Å the use of a three-dimensional scaling protocol is essential in order to obtain the highest possible anomalous signal.

Based on the scaling protocols currently available, the optimal wavelength range for data collection of protein–xenon complexes is between 2.10 and 2.40 Å. Beyond that, any further increase in signal will be compensated for or even superseded by a concomitant increase in noise, which cannot be fully corrected for.

The benefits of data collection in a helium atmosphere are mainly the reduction of air scatter, resulting in higher $I/\sigma(I)$ values and a weak but consistent increase in the anomalous signal.

research papers

work, including the diffraction images, are available to the scientific community on request to the authors.

References

Brown, J., Esnouf, R. M., Jones, M. A., Linnell, J., Harlos, K., Hassan, A. B. & Jones, E. Y. (2002). *EMBO J.* **21**, 1054–1062.

Cianci, M., Rizkallah, P. J., Olczak, A., Raftery, J., Chayen, N. E., Zagalsky, P. F. & Helliwell, J. R. (2001). *Acta Cryst.* D**57**, 1219–1229.

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

Cromer, D. T. & Liberman, D. A. (1970). *J. Chem. Phys.* **53**, 1891–1898.

Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.

Debreczeni, J. E., Bunkoczi, G., Ma, Q., Blaser, H. & Sheldrick, G. M. (2003). *Acta Cryst.* D**59**, 688–696.

Djinovic Carugo, K., Everitt, P. & Tucker, P. A. (1998). *J. Appl. Cryst.* **31**, 812–814.

French, S. & Wilson, K. S. (1978). *Acta Cryst.* A**34**, 517–519.

Gordon, E. J., Leonard, G. A., McSweeney, S. & Zagalsky, P. F. (2001). *Acta Cryst.* D**57**, 1230–1237.

Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.

Li, S., Finley, J., Liu, Z.-J., Qiu, S. H., Chen, H., Luan, C. H., Carson, M., Tsao, J., Johnson, D., Lin, G., Zhao, J., Thomas, W., Nagy, L. A., Sha, B., DeLucas, L. J., Wang, B.-C. & Luo, M. (2002). *J. Biol. Chem.* **277**, 48596–48601.

Liu, Z.-J., Vysotski, E. S., Chen, C.-J., Rose, J. P., Lee, J. & Wang, B.-C. (2000). *Protein Sci.* **9**, 2085–2093.

Micossi, E., Hunter, W. N. & Leonard, G. A. (2002). *Acta Cryst.* D**58**, 21–28.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Panjikar, S. & Tucker, P. (2002). *J. Appl. Cryst.* **35**, 261–266.

Polentarutti, M. & Djinovic Carugo, K. (2001). *ELETTRA Highlights 2000–2001*, pp. 26–27. Trieste, Italy: Sincrotrone Trieste.

Polentarutti, M. & Djinovic Carugo, K. (2003). In preparation.

Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003). *Acta Cryst.* D**59**, 1020–1027.

Ravelli, R. B. G., Theveneau, P., McSweeney, S. & Caffrey, M. (2002). *J. Synchrotron Rad.* **9**, 355–360.

Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.

Weiss, M. S. (2001). *J. Appl. Cryst.* **34**, 130–135.

Weiss, M. S., Panjikar, S., Nowak, E. & Tucker, P. A. (2002). *Acta Cryst.* D**58**, 1407–1412.

Weiss, M. S., Sicker, T., Djinovic Carugo, K. & Hilgenfeld, R. (2001). *Acta Cryst.* D**57**, 689–695.

Weiss, M. S., Sicker, T. & Hilgenfeld, R. (2001). *Structure*, **9**, 771–777.