Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

On the routine use of soft X-rays in macromolecular crystallography. Part III. The optimal data-collection wavelength

Christoph Mueller-Dieckmann, Santosh Panjikar, Paul A. Tucker and Manfred S. Weiss*

EMBL Hamburg Outstation, c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany

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

Received 6 April 2005
Accepted 5 July 2005

PDB References: ConA-Xe, 2a7a, r2a7asf; adaptin-Xe, 2a7b, r2a7bsf; PPE-Xe, 2a7c, r2a7csf; HEL-Xe, 2a7d, r2a7dsf; DNA, 2a7e, r2a7esf; HEL, 2a7f, r2a7fsf; thermolysin, 2a7g, r2a7gsf; trypsin, 2a7h, r2a7hsf; thaumatin, 2a7i, r2a7isf; PPE-Ca, 2a7j, r2a7jsf.

Complete and highly redundant data sets were collected at different wavelengths between 0.80 and 2.65 Å for a total of ten different protein and DNA model systems. The magnitude of the anomalous signal-to-noise ratio as assessed by the quotient $R_{\rm anom}/R_{\rm r.i.m.}$ was found to be influenced by the data-collection wavelength and the nature of the anomalously scattering substructure. By utilizing simple empirical correlations, for instance between the estimated $\Delta F/F$ and the expected $R_{\rm anom}$ or the data-collection wavelength and the expected $R_{\rm r.i.m.}$, the wavelength at which the highest anomalous signal-to-noise ratio can be expected could be estimated even before the experiment. Almost independent of the nature of the anomalously scattering substructure and provided that no elemental X-ray absorption edge is nearby, this optimal wavelength is 2.1 Å.

1. Introduction

Whereas most novel protein structures are nowadays determined based on the multiple-wavelength anomalous diffraction (MAD) method (Hendrickson, 1991) applied to selenomethionine (SeMet) derivatives of the protein of interest, in the past half decade a renaissance of structure determination based on native proteins has been observed. This is mainly owing to the fact that the signal provided by anomalous scattering of the natively present sulfur (in proteins) or phosphorus (in nucleic acids) atoms, albeit small, can be measured fairly reliably even using Cu $K\alpha$ radiation (Hendrickson & Teeter, 1981; Dauter *et al.*, 1999; Yang & Pflugrath, 2001; Lemke *et al.*, 2002; Debreczeni *et al.*, 2003; Olsen *et al.*, 2004) provided the data redundancy is high so that the noise in the data is minimized to the largest extent possible. Longer wavelengths than Cu $K\alpha$ (λ = 1.5418 Å) would of course produce an even larger signal, but at the same time the experimental difficulties increase (Djinovic Carugo *et al.*, 2005 and references therein) as does the noise level in the data. Nevertheless, it has been reported recently that data-collection wavelengths in the range λ = 1.5–3.0 Å are fairly easy to handle in a diffraction experiment and that they can be used routinely at many synchrotron sources (Cianci *et al.*, 2001; Weiss, Sicker, Djinovic Carugo *et al.*, 2001; Weiss, Sicker & Hilgenfeld, 2001; Mueller-Dieckmann *et al.*, 2004; Djinovic Carugo *et al.*, 2005) and even at home sources using for instance Cr $K\alpha$ radiation (λ = 2.29 Å; Yang *et al.*, 2003). Consequently, a number of structures have been determined based on diffraction data collected in this longer wavelength regime (Liu *et al.*, 2000; Brown *et al.*, 2002; Li *et al.*, 2002; Micossi *et al.*, 2002; Ramagopal *et al.*, 2003; Weiss *et al.*, 2004; Chen *et al.*, 2004; Phillips *et al.*, 2004). Apart from increasing

the sulfur (or phosphorus) anomalous scattering, the use of longer wavelengths may also provide advantages when chloride or calcium ions are present in the structure or when dealing with xenon or iodide derivatives of proteins or with derivatives of the very heavy elements such as uranium (Djinovic Carugo *et al.*, 2005).

Within this longer wavelength range of $\lambda = 1.5$–3.0 Å, one of the most important questions remains: which of the wavelengths is the best for diffraction data collection? The concept of an optimal wavelength has been addressed before, but mainly in the context of optimizing the scattered intensity relative to the X-ray dose absorbed by the crystal (Arndt, 1984; Polikarpov, 1997; Polikarpov *et al.*, 1997; Teplyakov *et al.*, 1998). For the purpose of this manuscript, the best wavelength will be defined as the one yielding the highest anomalous signal-to-noise ratio. If an elemental X-ray absorption edge is present in the wavelength range of interest, the answer to this question may be trivial, but as we will discuss here, the complete anomalously scattering substructure also has to be taken into account. Initial experiments pointed to an optimal wavelength of 1.9 Å for thermolysin (Weiss, Sicker & Hilgenfeld, 2001), where the largest part of the phasing signal originates from calcium ions, and between 2.0 and 2.3 Å for a xenon complex of porcine pancreatic elastase (PPE), where one Xe atom located in the active site of PPE is the dominant anomalous scatterer (Mueller-Dieckmann *et al.*, 2004). In order to arrive at a more general statement with respect to the optimal data-collection wavelength, a total of 74 diffraction data sets collected from ten different protein and DNA model systems were examined with respect to their anomalous signal-to-noise ratio.

2. Materials and methods

2.1. Crystallization

Six of the seven proteins used for crystallization were commercially available and were used without further purification. The oligonucleotide CCCTAGGG was synthesized by MWG Biotech AG (Ebersberg, Germany). Crystallization experiments were all carried out at 293 K using the hanging-drop vapour-diffusion technique.

2.1.1. Hen egg-white lysozyme (HEL). Crystals of HEL (Calbiochem product No. 4403, lot No. B43266) were grown as described by Weiss *et al.* (2000) by mixing equal amounts of protein solution (30 mg ml^{-1} in water) and reservoir solution containing 50 m*M* sodium acetate pH 4.5 and 5%(*w*/*v*) NaCl. The tetragonal crystals (space group $P4_32_12$) appeared within a few days and exhibited the standard unit-cell parameters of $a = 78.7$, $c = 37.2$ Å. They were transferred to a solution of 25%(*v*/*v*) 2-methyl-2,4-pentanediol (MPD) prior to flash-cooling to 100 K. They typically diffract X-rays to better than 1.3 Å resolution.

2.1.2. Porcine pancreatic elastase (PPE). PPE (Serva product No. 20929, lot No. 16461) was crystallized in the presence of either sodium or calcium ions as described by Weiss *et al.* (2002). For the sodium complex of PPE, a solution of 20 mg ml^{-1} PPE in water was mixed with an equal amount of reservoir solution containing 0.1 *M* sodium acetate pH 5.1 and 0.2 *M* sodium sulfate as the precipitant. For the calcium complex of PPE, a protein solution of 12 mg ml^{-1} was mixed with an equal volume of reservoir solution composed of 0.1 *M* sodium acetate pH 5.1, 0.2 *M* sodium citrate and 0.04 *M* calcium chloride. Under both conditions, crystals belonging to space group $P2_12_12_1$ with unit-cell parameters $a = 49.3$, $b = 57.0$, $c = 73.3$ Å appeared within a few days. Crystals were cryoprotected by transferring them into a solution containing 25%(*v*/*v*) glycerol in water. They typically diffract X-rays to better than 1.5 Å resolution.

2.1.3. DNA. Prior to purification of the oligonucleotide, an annealing step was performed by heating the solution to 368 K for 15 min and letting it cool to 273 K. This solution was then applied to a Sep-Pak cartridge (Waters GmbH, Eschborn, Germany), which was pre-treated first with 10 ml methanol and then with 10 ml water. After a wash step with 2%(*v*/*v*) methanol in water, the oligonucleotide was eluted with 80%(*v*/*v*) methanol. The eluted oligonucleotide was subsequently dried in a Speedvac centrifuge overnight and resuspended in water. The oligonucleotide concentration was determined at 260 nm. Crystallization was performed by mixing a solution of 50 m*M* sodium cacodylate pH 7.0, 20 m*M* $MgCl_2$, 2 m*M* oligonucleotide (single-strand concentration) and 1 m*M* spermine tetrachloride and letting the mixture equilibrate against 40%(*v*/*v*) MPD (Tippin & Sundaralingam, 1996). Tetragonal crystals of space group $P4_32_12$, with unit-cell parameters $a = 41.6$, $c = 24.7$ Å grew within four weeks. For flash-cooling to 100 K, the crystals had to be transferred to dry paraffin oil in order to remove surrounding mother liquor. They diffract X-rays to better than 1.0 Å resolution.

2.1.4. Concanavalin A (ConA). Crystals of the lectin ConA (Fluka product No. 61760, lot No. 420479/1) were grown within a period of two months from drops made from equal amounts of protein solution (10 mg ml^{-1} in water) and reservoir solution [34%(*v*/*v*) PEG 1500]. They were flash-cooled to 100 K directly from the crystallization drop and belong to space group $I222$, with unit-cell parameters $a = 62.0$, $b = 85.1$, $c = 88.4$ Å. They diffract X-rays to better than 1.5 Å resolution.

2.1.5. Thaumatin. Thaumatin (Sigma product No. T-7638, lot No. 108 F0299) crystals grew within a few days after mixing equal amounts of protein solution [15 mg ml^{-1} in 0.1 *M* *N*-(2-acetamido)iminodiacetic acid (ADA) pH 6.5] and reservoir solution (0.1 *M* ADA pH 6.5, 1 *M* sodium/potassium tartrate) (Charron *et al.*, 2002). The tetragonal crystals were transferred to a cryoprotection solution consisting of 0.75 *M* sodium/potassium tartrate and 25%(*v*/*v*) glycerol prior to flash-cooling in liquid nitrogen at 77 K. The tetragonal crystals (space group $P4_12_12$) exhibit unit-cell parameters $a = 57.3$, $c = 148.8$ Å and diffract X-rays to beyond 1.5 Å resolution.

2.1.6. Trypsin. Crystals of trypsin (Fluka product No. 93610, lot No. 433274–023) were obtained from a solution of 7.5 mg ml^{-1} protein in 0.1 *M* Tris pH 7.4, 0.08 *M* $(NH_4)_2SO_4$, 10%(*w*/*v*) PEG 6000 and 15%(*v*/*v*) ethylene glycol equilibrated against 0.16 *M* $(NH_4)_2SO_4$, 20%(*w*/*v*) PEG 6000 and

Table 1
Relevant data-collection and processing statistics for the 74 data sets.

	λ (Å)	d_{min} (Å)	N†	$R^*_{r.i.m}$ (%)	R_{anom} (%)
ConA-Xe	1.00	2.40	13.3	4.3	1.5
	1.50	2.40	13.2	3.6	2.4
	1.70	2.40	13.0	3.6	2.9
	1.90	2.40	12.3	3.8	3.1
	2.10	2.40	9.1	4.1	3.6
	2.30	2.40	10.4	4.6	3.5
	2.50	2.40	6.1	6.0	3.8
Adaptin-Xe	0.80	2.44	12.0	4.6	1.4
	1.50	2.44	11.4	2.7	2.2
	1.70	2.44	11.0	2.8	2.8
	1.90	2.44	10.8	3.0	3.2
	2.10	2.44	10.6	4.2	4.3
	2.25	2.44	10.2	4.7	4.3
	2.40	2.44	9.9	8.6	4.7
	2.50	2.44	9.3	8.2	5.9
	2.65	2.44	9.1	11.6	6.1
PPE-Xe	0.80	2.44	13.1	4.8	1.5
	1.50	2.44	12.5	4.4	2.5
	1.70	2.44	12.5	4.1	2.9
	1.90	2.44	12.1	4.4	3.4
	2.10	2.44	11.3	4.7	3.8
	2.25	2.44	10.4	4.8	4.3
	2.40	2.44	9.9	5.5	4.0
	2.50	2.44	9.7	5.2	3.2
	2.65	2.44	9.2	5.7	2.4
HEL-Xe	0.80	2.44	24.6	4.1	0.8
	1.50	2.44	22.5	3.6	1.5
	1.70	2.44	24.7	3.6	1.7
	1.90	2.44	24.5	3.9	2.1
	2.10	2.44	24.0	4.3	2.4
	2.25	2.44	22.6	4.7	2.7
	2.40	2.44	20.8	5.7	2.9
	2.50	2.44	29.7	6.3	2.9
	2.65	2.60	17.2	11.7	3.7
DNA	0.80	2.30	19.3	2.7	0.6
	1.50	2.30	18.2	3.4	1.3
	1.70	2.30	18.1	4.8	1.4
	1.90	2.30	15.9	5.2	1.7
	2.10	2.30	16.5	11.1	2.6
	2.30	2.30	15.2	17.4	4.1
	2.50	2.30	13.6	28.8	4.8
HEL	1.50	2.20	23.3	3.0	1.2
	1.70	2.20	23.8	3.2	1.5
	1.90	2.20	23.2	4.1	1.8
	2.10	2.20	22.2	4.7	2.1
	2.30	2.20	20.7	6.9	2.4
Thermolysin	1.28	2.24	35.0	5.3	2.1
	1.50	2.24	38.6	5.1	1.2
	1.70	2.24	40.1	3.5	1.3
	1.80	2.24	39.3	3.4	1.4
	1.90	2.24	39.1	3.4	1.6
	2.00	2.24	38.1	4.0	1.7
	2.10	2.24	35.7	6.8	2.0
	2.64	2.56	35.5	6.2	2.8
Trypsin	1.00	2.32	18.9	3.4	0.7
	1.50	2.32	19.2	3.5	1.1
	1.90	2.32	18.0	3.1	1.5
	2.10	2.32	16.2	3.8	1.7
	2.30	2.32	14.8	3.9	1.9
	2.50	2.32	13.4	4.4	2.2
Thaumatin	1.00	2.32	25.8	2.4	0.6
	1.50	2.32	26.3	3.1	1.0
	1.70	2.32	25.7	2.8	1.2
	1.90	2.32	23.5	4.2	1.5
	2.10	2.32	24.1	4.1	1.6
	2.30	2.32	21.6	5.5	2.0
	2.50	2.32	8.4	5.8	2.7
PPE-Ca	0.80	2.32	12.8	5.4	1.1
	1.50	2.32	12.0	4.9	1.4
	1.70	2.32	11.7	4.8	1.5
	1.90	2.32	11.7	4.6	1.6
	2.10	2.32	11.1	4.8	1.8
	2.30	2.32	10.0	5.5	2.2
	2.50	2.32	7.9	7.3	3.3

† Redundancy of the data set

30%(*v*/*v*) ethylene glycol as described by Debreczeni *et al.* (2003). The trigonal crystals (space group $P3_121$) with unit-cell parameters a = 54.2, c = 105.9 Å were flash-cooled in a nitrogen-gas stream at 100 K directly from the crystallization drop. They diffract X-rays to better than 1.0 Å resolution.

2.1.7. Adaptin. The γ-adaptin appendage domain was kindly provided by Dr Phil Evans (Structural Studies Division, MRC Laboratory of Molecular Biology, Hills Road, Cambridge) and used without further purification. Crystals were obtained by employing the method of Kent *et al.* (2002). Equal amounts of protein solution (5 mg ml^{-1}) and reservoir solution containing 50 m*M* 2-[4-(2-hydroxyethyl)-1-piperazinyl]ethanesulfonic acid (HEPES) pH 7.4, 20%(*v*/*v*) glycerol and 1.0 *M* sodium/potassium tartrate were mixed and equilibrated against each other. The obtained crystals belong to the orthorhombic space group $P2_12_12_1$, with unit-cell parameters a = 33.4, b = 53.8, c = 66.8 Å. They were flash-cooled in a gaseous nitrogen stream at 100 K after they had been transferred to a solution consisting of 22%(*v*/*v*) glycerol in water.

2.1.8. Thermolysin. Thermolysin was crystallized according to the protocol of Schiefner (2000). 30 mg thermolysin powder (Calbiochem, Lot No. B28249) was dissolved in 120 µl 45%(*v*/*v*) dimethylsulfoxide (DMSO) in water. 4 µl of this protein solution was mixed with 2 µl reservoir solution containing 1.4 *M* calcium acetate, 10 m*M* zinc acetate, 1 m*M* NaN_3 and 50 m*M* Tris pH 7.3 and equilibrated over 1 ml reservoir solution. Crystals belonging to space group $P6_122$ with unit-cell parameters a = 92.7, c = 129.6 Å usually appeared within a few days. Prior to flash-cooling in a nitrogen-gas stream at 100 K, they were transferred to a solution consisting of 10 m*M* Tris pH 7.3, 10 m*M* calcium acetate, 7%(*v*/*v*) DMSO and 20%(*v*/*v*) glycerol. They typically diffract X-rays to better than 1.5 Å resolution.

2.2. Preparation of the xenon derivatives

Crystals of ConA, adaptin, PPE (sodium complex), HEL, DNA and thaumatin were fished out from their cryoprotecting solutions, transferred to a xenon chamber (Hampton Research, California, USA) and exposed to 1.2–1.4 MPa xenon pressure for 1–2 min. After gently releasing the pressure, the crystals were flash-cooled to 77 K in liquid nitrogen.

2.3. Diffraction data collection

Diffraction data sets were collected at the X-ray diffraction beamline at the ELETTRA Synchrotron Radiation Facility (Trieste, Italy) using a 165 mm MAR CCD detector. The condensed relevant data-collection and processing parameters

Table 2
Refinement statistics for the ten model systems against the data sets collected at the shortest wavelength for each of the systems.

Data set	ConA-Xe	Adaptin-Xe	PPE-Xe	HEL-Xe	DNA	HEL	Thermolysin	Trypsin	Thaumatin	PPE-Ca
Wavelength (Å)	1.00	0.80	0.80	0.80	0.80	1.50	1.28	1.00	1.00	0.80
Space group	*I*222	$P2_12_12_1$	$P2_12_12_1$	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P6_122$	$P3_121$	$P4_12_12$	$P2_12_12_1$
Unit-cell parameters										
a (Å)	61.97	33.43	50.35	77.05	41.63	78.71	92.99	54.19	57.31	49.32
b (Å)	85.06	53.76	58.02	77.05	41.63	78.71	92.99	54.19	57.31	56.97
c (Å)	88.37	66.77	74.74	37.21	24.72	37.21	130.21	105.86	148.75	73.32
Resolution (Å)	40–1.75	40–1.65	40–1.65	40–1.65	40–1.67	40–1.85	40–1.85	40–2.1	40–1.75	40–1.65
Total No. of reflections	23439	14704	26471	13401	2742	10173	28417	10801	25377	24957
Reflections in working set	22939	14411	25959	13113	2694	9946	27815	10585	24869	24449
Reflections in test set	500	293	512	288	48	227	602	216	508	508
R_{cryst} (%)	18.0	18.2	15.5	17.2	18.3	16.6	16.1	16.6	19.6	14.7
R_{free} (%)	20.2	22.8	18.2	20.2	19.7	23.2	18.9	23.0	21.9	18.9
No. of protein atoms	1812	947	1860	1013	161	1013	2444	1626	1562	1849
No. of water molecules	119	87	222	88	32	173	178	118	154	296
No. of ions and other atoms	7	2	3	11	0	8	7	3	0	2
Average *B* factor (Å^2)										
All atoms	8.8	21.8	12.4	23.1	23.6	21.7	21.1	10.3	15.8	13.5
Protein atoms	8.7	21.0	11.3	22.4	21.5	19.9	20.7	9.9	13.9	11.7
Ions/other atoms	6.3	17.4	10.4	19.9	—	17.7	16.8	35.5	—	5.3
R.m.s.d. bond lengths (Å)	0.016	0.025	0.012	0.022	0.011	0.021	0.017	0.025	0.014	0.016
R.m.s.d. bond angles (°)	1.73	2.03	1.44	1.80	1.83	1.74	1.43	2.19	1.60	1.48
PDB code	2a7a	2a7b	2a7c	2a7d	2a7e	2a7f	2a7g	2a7h	2a7i	2a7j

are given in Table 1. A complete listing of the parameters is provided as supplementary material.[1] The choice of data-collection wavelengths was made to cover the interesting wavelength range between 0.80 and 2.65 Å. For the Xe data sets we also wanted to avoid being close to the three *L* absorption edges. Exposure time and attenuation were chosen in order to utilize the full dynamic range of the detector without allowing too many overloaded reflections to occur on each image and to keep the spindle-axis rotation speed below approximately 10.0° min^{-1}. Compensation for varying incident beam intensity was performed by carrying out the data collection 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-collection strategy was employed. Since the plan was to collect 360° in either 1 or 0.5° 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 as for the first data set was used. Since the shortest possible crystal-to-detector distance at the beamline was 35 mm, the longer wavelength data sets could not be collected to as high a resolution as the shorter wavelength data sets. All data set series were collected from one crystal each, except for the thermolysin series, for which a total of five crystals were used at different visits to ELETTRA.

2.4. Data processing

All data sets were indexed and integrated using *DENZO* (Otwinowski & Minor, 1997). The post-refinement procedure in *SCALEPACK* (Otwinowski & Minor, 1997) was used to refine the unit-cell parameters and the mosaicity for each data set. Scaling and merging of the data were carried out using the program *SCALA* (Collaborative Computational Project, Number 4, 1994) and the scaling protocol *SCALA-Sec* as described in Mueller-Dieckmann *et al.* (2004). 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). $R^*_{r.i.m.}$, which is $R_{r.i.m.}$ based on treating I^+ and I^- as separate observations, was obtained from *SCALA*.

In order to make the different data sets comparable with respect to the maximum resolution, all data sets of each wavelength series were processed to the maximum resolution of the data set in each series that was collected at the longest wavelength and thus exhibited the lowest maximum resolution. The only exception to this was thermolysin, where the maximum resolution of the λ = 2.10 Å data set (d_{min} = 2.24 Å) was used as the threshold, whereas the maximum resolution of the λ = 2.64 Å data set of thermolysin was only 2.56 Å. The maximum resolution of the λ = 2.65 Å data set of HEL had to be reduced to 2.60 Å owing to a signal-to-noise ratio that was too small at the edge of the detector. Consequently, these two data sets have to be treated separately. The maximum resolution of the remaining 72 data sets considered here was between 2.20 and 2.44 Å.

2.5. The anomalous signal-to-noise ratio

The estimated anomalous diffraction ratio $\Delta F/F$ as a function of the data-collection wavelength was calculated using $\Delta F/F = (2\sum_i N_i \Delta f_i''^2 / \sum_i N_i f_i^2)^{1/2}$, where both sums run over all atoms in the structure, as described in Weiss, Sicker & Hilgenfeld (2001). The estimate is for zero scattering angle and assumes resolved and independent anomalous scatterers. As an indicator for the usefulness of the anomalous signal for later application, *e.g.* phase determination, the quotient

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: DZ5049). Services for accessing these data are described at the back of the journal.

$R_{\rm anom}/R_{\rm p.i.m.}$ was calculated, where $R_{\rm anom}$ [$R_{\rm anom} = 100\sum_{hkl}|I(hkl)-I(\overline{hkl})|/\sum_{hkl}\langle I(hkl)\rangle$] expresses the magnitude of the anomalous signal and $R_{\rm p.i.m.}$ {$R_{\rm p.i.m.} = 100\sum_{hkl}[1/(N-1)]^{1/2}\sum_i|I_i(hkl)-\langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, with N being the number of times each reflection hkl has been observed} is a measure of the noise in the data. This quotient has been shown previously (Weiss, Sicker & Hilgenfeld, 2001; Panjikar & Tucker, 2002; Mueller-Dieckmann *et al.*, 2004) to be a very good predictor for the success rate in substructure determination, for the quality of the phases which can be obtained and for the efficiency with which the model can be built automatically. In order to make $R_{\rm anom}/R_{\rm p.i.m.}$ independent of the redundancy of the data, $R_{\rm p.i.m.}$ in the denominator was replaced by $R_{\rm r.i.m.}$ {$R_{\rm r.i.m.} = 100\sum_{hkl}[N/(N-1)]^{1/2}\sum_i|I_i(hkl)-\langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$} or $R^*_{\rm r.i.m.}$, with $R^*_{\rm r.i.m.}$ being equivalent to $R_{\rm r.i.m.}$ but calculated with the Friedel pairs kept separate.

2.6. Refinement

All ten protein structures were refined to the highest resolution possible against the data set collected at the shortest wavelength using the program *REFMAC*5 (Collaborative Computational Project, Number 4, 1994). Relevant refinement statistics are given in Table 2. To estimate the occupancy of the anomalously scattering substructure atoms or ions (Xe, Mn^{2+}, Zn^{2+}, Ca^{2+}, Cl^- and SO_4^{2-}), the refined structures were refined again against the respective data sets collected at a wavelength of 1.50 Å, except for HEL, where 1.50 Å was the shortest wavelength collected. Occupancies were estimated as described by Weiss *et al.* (2002) based on the relative peak heights in the anomalous difference Fourier syntheses, with the sulfur peaks serving as a reference. The respective relevant substructure occupancies are given in the legend to Fig. 1. All ten refined coordinate files as well as the underlying structure-factor amplitudes have been deposited with the PDB (for PDB codes see Table 2).

3. Results

3.1. Diffraction data collection and processing

Most of the collected data sets are of good quality: they have close to 100% completeness and exhibit excellent merging statistics. A short summary of the most relevant statistics is presented in Table 1, while the detailed statistics are available as supplementary information to this manuscript. Five of the thermolysin data sets as well as the nine data sets of PPE-Xe have already been described in previous publications (Weiss, Sicker & Hilgenfeld, 2001; Mueller-Dieckmann *et al.*, 2004), although they had to be reprocessed and rescaled in order to make them comparable to the other data sets described here. The λ = 1.7 Å data set of trypsin is missing from the series owing to a mistake in backing up the raw data. A potential problem with the data sets is the somewhat reduced redundancy of the longer wavelength data sets. This resulted from the necessity of using the shortest possible crystal-to-detector distance in order to collect the data to the highest resolution possible. Consequently, large parts of the detector surface (up to 25%) were covered by the shadow of the cryo-nozzle, reducing the active area available for data collection. A 2θ stage was not available at the beamline. In some cases, problems during data collection prevented the utilization of the full 360° of data, which also led to some reduced values for the redundancy.

3.2. Refinement and anomalously scattering substructures

All ten model structures have been well refined and exhibit very good geometrical parameters (Table 2). Thus, the model phases derived from these refinements are reliable and the corresponding anomalously scattering substructures which have been derived based on the anomalous difference electron-density map using data collected at λ = 1.5 Å are also reliable in all cases. The detailed composition of the anomalously scattering substructures is given in the legend to Fig. 1.

For the two model systems thaumatin and DNA, only protein or nucleic acid atoms were found to constitute the

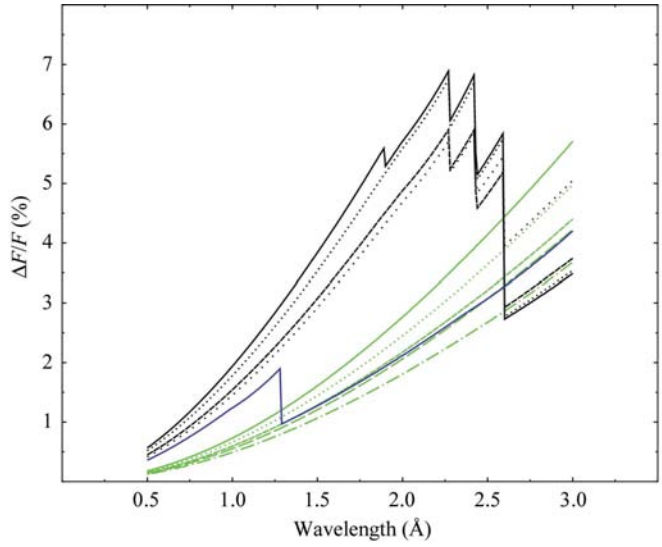

Figure 1
Estimated anomalous diffraction ratio $\Delta F/F$ in the wavelength range 0.5–3.0 Å for the ten model systems described in this paper. The estimate is for zero scattering angle and assumes resolved and independent anomalous scatterers. The ten systems are separated into two groups (Xe group and P/S/Ca group) and sorted in decreasing strength of the anomalous signal. Black lines, ConA-Xe, adaptin-Xe, PPE-Xe and HEL-Xe; blue line, thermolysin; green lines, DNA, HEL, trypsin, thaumatin and PPE-Ca. The experimentally determined anomalously scattering substructures for the ten systems were (1) ConA-Xe: two protein S atoms, one Mn^{2+} and one Ca^{2+} ion, both fully occupied, as well as six Xe atoms with occupancies (q) of 0.40, 0.30, 0.20, 0.15, 0.15 and 0.10; (2) adaptin-Xe: four protein S atoms and two Xe atoms (q = 0.32 and 0.10); (3) PPE-Xe: ten protein S atoms, one Xe atom (q = 0.72) and two SO_4^{2-} ions (q = 0.70 and 0.50); (4) HEL-Xe: ten S atoms, two Xe atoms (q = 0.24 and 0.08), the first one being situated on a crystallographic twofold axis, and eight Cl^- ions (q = 0.68, 0.58, 0.52, 0.37, 0.37, 0.34, 0.31 and 0.28); (5) DNA: eight P atoms; (6) HEL: ten S atoms and seven Cl^- ions (q = 0.80, 0.77, 0.70, 0.60, 0.57, 0.37 and 0.25); (7) thermolysin: two protein S atoms, six Ca^{2+} ions (q = 1.00, 1.00, 1.00, 1.00, 0.50 and 0.25), one fully occupied Zn^{2+} ion and two DMSO molecules (q = 0.50 and 0.40); (8) trypsin: 14 protein S atoms, one fully occupied Ca^{2+} ion and two partially occupied Cl^- ions (q = 0.40 and 0.25); (9) thaumatin: 17 protein S atoms and (10) PPE-Ca: ten protein S atoms, one Ca^{2+} ion (q = 0.81) and one Cl^- ion (q = 0.30).

anomalously scattering substructures. The substructure for HEL with ten protein S atoms and eight partially occupied Cl^- ions has been described before (PDB code 1lz8; Dauter *et al.*, 1999). The top six of the seven Cl^- ions found in our structure are identical to the previously described top six chlorides, whereas the seventh ion position has not been observed before. In contrast, two of the previously reported positions were not observed in our structure, probably owing to too low occupancy. Also, the substructure of PPE-Ca with partially occupied Ca^{2+} and Cl^- ions has been described previously (PDB code 1lka; Weiss *et al.*, 2002). In the case of trypsin, a fully occupied Ca^{2+} ion as well as two partially occupied Cl^- ions complement the protein part. To the best of our knowledge the chloride positions have not been described before. In thermolysin, one Zn^{2+} ion and six Ca^{2+} ions were found in addition to the two protein S atoms, as well as two partially occupied DMSO molecules. For the xenon derivatives ConA-Xe, adaptin-Xe, PPE-Xe and HEL-Xe, six, two, one and two Xe atoms, respectively, could be identified in the anomalous difference map. ConA-Xe contains the known Mn^{2+} and Ca^{2+} ions in addition to the Xe atoms, whereas HEL-Xe also contains eight Cl^- ions, of which the top five correspond to those in both HEL and PDB entry 1lz8 (Dauter *et al.*, 1999). Some discrepancies occur at the low-occupancy positions, but this observation is probably of no significance. For PPE-Xe, two SO_4^{2-} ions were also found in accordance with previous observations (PDB code 1lkb; Weiss *et al.*, 2002) for the Na_2SO_4 crystal form of PPE. The Cl^- position near Ser14 was found to be unoccupied, which can be explained by the careful omission of chloride during buffer equilibration. Whereas in Weiss *et al.* (2002) the acetate buffer was equilibrated with hydrochloric acid, acetic acid was used here instead. Adaptin-Xe does not contain anything besides the protein S atoms and the Xe atoms. For thaumatin and DNA, Xe derivatization has also been attempted, but no Xe atoms could be located in the respective Xe-pressurized crystals (data not shown).

From the data presented it is clear that longer wavelength data, carefully collected to high redundancy and properly processed and scaled, can be very useful in establishing which ions from the buffer or even from the purification protocol might be bound to the protein molecule of interest. This has already been demonstrated several times in the literature (Einspahr *et al.*, 1985; Weiss, Sicker, Djinovic Carugo *et al.*, 2001; Weiss *et al.*, 2002; Kuettner *et al.*, 2002; Ferreira *et al.*, 2004; Sekar *et al.*, 2004). For this reason, it would be desirable that any macromolecular structure determination be complemented with a long-wavelength data set.

3.3. Correlation between estimated $\Delta F/F$ and observed $R_{\rm anom}$

The anomalous diffraction ratio $\Delta F/F$ as a function of the data-collection wavelength estimated based on the atomic composition of the asymmetric unit, including the complete anomalously scattering substructure described and discussed in the previous paragraph for all ten protein and DNA model systems, is shown in Fig. 1. The ten curves cluster into two groups: the Xe group with a relatively larger signal and the P/S/Ca group with a smaller signal. For the Xe data sets, the three *L* edges are clearly discernible, as is the Zn *K* edge for thermolysin. The Mn *K* edge in the case of ConA-Xe almost vanishes behind the stronger contribution of the six Xe atoms.

In Fig. 2, all observed $R_{\rm anom}$ values are compared with the estimated $\Delta F/F$ values for the ten model systems. As is evident

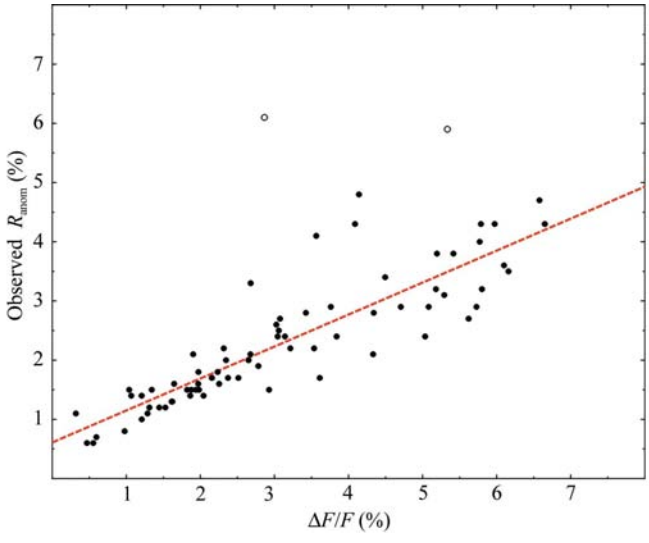

Figure 2
Observed $R_{\rm anom}$ values as a function of estimated $\Delta F/F$ values. The distribution can be fitted with a linear dependence of $R_{\rm anom} = 0.61 + 0.54\Delta F/F$ and a correlation of 0.87. The data points for adaptin-Xe at wavelengths of 2.50 and 2.65 Å have been excluded as obvious outliers (open circles). Owing to counting statistics and other errors in data collection as well as in data processing, the intercept of the line is different from its theoretical value of zero.

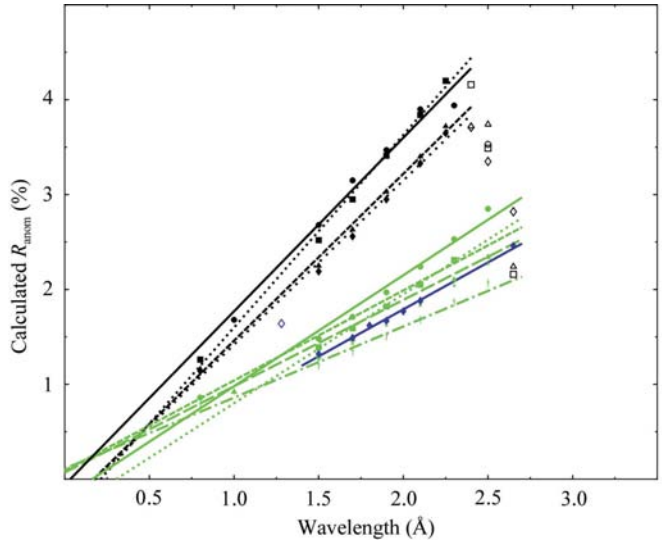

Figure 3
Calculated $R_{\rm anom}$ values as a function of data-collection wavelength for each of the ten model systems. In the cases of the Xe derivatives the Xe *L* edges have been neglected in the linear fit, as has the Zn *K* edge in the case of thermolysin (open symbols). The Mn *K* edge in the case of ConA-Xe is included since it is more or less overshadowed by the presence of the six Xe atoms which dominate the anomalous signal in this case. The parameters of the ten linear fits are given in Table 3 as is the correlation coefficient for the fit in each case. The same colouring scheme applies as in Fig. 1.

from the figure, there is a linear correlation between the observed R_{anom} values and the estimated $\Delta F/F$ values. The correlation coefficient of 0.87 for the fit of the equation $R_{anom} = 0.61 + 0.54\Delta F/F$ is very good, which proves that the estimated $\Delta F/F$ values are a very good predictor for the observed R_{anom} values. This has to be taken with a pinch of salt, however, since the observed R_{anom} values also depend to some extent on the redundancy of the data. However, in our cases the redundancy of all data sets is so high (mostly between 10 and 40, with a few exceptions for the data sets collected at the longest wavelengths) that any further increase will have no further lowering effect on the observed R_{anom}.

3.4. Estimation of R_{anom} as a function of data-collection wavelength

With the correlation discussed in the previous paragraph and the wavelength-dependence of the estimated $\Delta F/F$ values, it is now possible to calculate R_{anom} for each model system of known anomalously scattering substructure for any data-collection wavelength. The calculated R_{anom} values for all ten model systems are shown in Fig. 3 for all wavelengths at which a data set was collected. For each model system a linear fit with correlation coefficients of close to 1.0 can be formulated (Table 3) when the absorption edges present (L edges of Xe derivatives and K edge of Zn in thermolysin) are excluded.

3.5. Dependence of $R^*_{r.i.m.}$ on data-collection wavelength

A plot of the observed values for $R^*_{r.i.m.}$ as a function of data-collection wavelength is shown in Fig. 4. This distribution can be fitted conveniently with an exponential function and a good R^2 value of 0.57. The reason why the three DNA data sets (collected at $\lambda = 2.10$, 2.30 and 2.50 Å, respectively) do not obey the fit very well is not clear at present, especially in the light of the fact that they were collected under identical conditions and from the same crystal as the first four DNA data sets. Nevertheless, the exponential fit described explains the distribution well. The fitted exponential curve displays a value of 3.8% at a hypothetical wavelength of 0 Å. This number describes the expected merging statistics of a crystal which is diffracting X-rays rather well. This is certainly the case for all of our ten model systems, but for less strongly diffracting crystals this value might turn out to be significantly higher. The steep increase beyond a wavelength of 2.5 Å reflects the difficulties of the scaling programs in coping with the increased absorption at the longer wavelengths. This phenomenon has been discussed before a number of times (Weiss, Sicker, Djinovic Carugo *et al.*, 2001; Weiss, Sicker & Hilgenfeld, 2001; Mueller-Dieckmann *et al.*, 2004). It should also be noted that scaling protocols inferior to that used here result in the shift of the curve to the left (data not shown).

Table 3
Linear fit parameters for calculating R_{anom} values as a function of data-collection wavelength for the ten model systems [R_{anom} (%) = $a + b\lambda$ (Å)] including the correlation coefficients for the respective fits (see also Fig. 3).

The three Xe L edges in the Xe-derivative data sets have been neglected in the linear fit, as has the Zn K edge in the case of thermolysin.

No.	Model System	a	b	CC
1	ConA-Xe	−0.06	1.83	0.99
2	Adaptin-Xe	−0.43	2.03	1.00
3	PPE-Xe	−0.30	1.76	1.00
4	HEL-Xe	−0.31	1.73	1.00
5	DNA	−0.19	1.17	0.99
6	HEL	−0.35	1.15	1.00
7	Thermolysin	−0.09	0.95	0.99
8	Trypsin	−0.19	0.99	1.00
9	Thaumatin	−0.07	0.91	0.99
10	PPE-Ca	+0.11	0.75	0.99

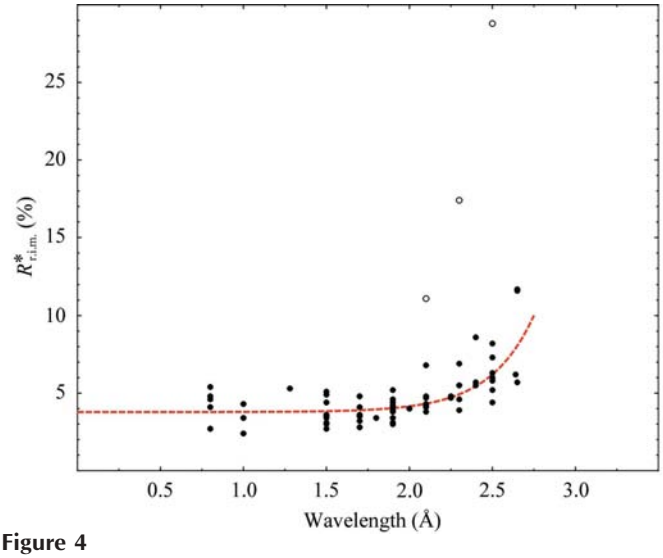

Figure 4
Dependence of the observed $R^*_{r.i.m}$ values on the data-collection wavelength. The data distribution can be fitted with the exponential function $R^*_{r.i.m} = 3.78 + 0.0002\exp(3.76\lambda)$ and an R^2 value of 0.57. The three outliers (DNA data sets collected at $\lambda = 2.10$, 2.30 and 2.50 Å, respectively) not obeying the exponential fit are shown as open circles.

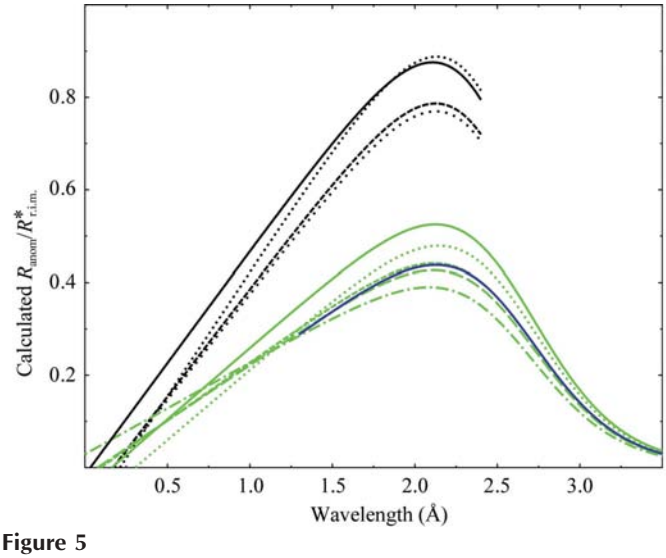

Figure 5
Ratio of calculated $R_{anom}/R^*_{r.i.m}$ as a function of data-collection wavelength. The same colouring scheme applies as in Fig. 1. The curves for the Xe derivatives have been truncated before the first Xe L absorption edge; the thermolysin curve starts past the K absorption edge of Zn.

Table 4
Optimal data-collection wavelengths for the ten protein and DNA model systems.

No.	Model system	λ (Å) at which the calculated $R_{anom}/R^*_{r.i.m}$ is maximal
1	ConA-Xe	2.10
2	Adaptin-Xe	2.13
3	PPE-Xe	2.12
4	HEL-Xe	2.12
5	DNA	2.12
6	HEL	2.14
7	Thermolysin	2.12
8	Trypsin	2.11
9	Thaumatin	2.11
10	PPE-Ca	2.08

3.6. Estimated $R_{anom}/R^*_{r.i.m.}$ as a function of data-collection wavelength

With the equations for R_{anom} and $R^*_{r.i.m.}$ as a function of data-collection wavelength in place (see §§3.4 and 3.5), it is now possible to estimate the ratio $R_{anom}/R^*_{r.i.m.}$ as a function of data-collection wavelength. This is shown in Fig. 5 for each of the ten model systems. The most striking feature of this figure is that all ten curves exhibit a maximum in the vicinity of 2.1 Å. From the exponential nature of the correlation between $R^*_{r.i.m.}$ and the data-collection wavelength, it might have been expected that at longer data-collection wavelengths all curves would look very similar, but it is still surprising that even the weaker signals originating from P or S atoms, which increase less with data-collection wavelength than for instance the signals originating from Xe, are still able to compensate for the exponential increase of the error. The exact location of the maximum for each of the ten systems is given in Table 4. There might be a slight correlation between the strength of the anomalous signal and the optimum data-collection wavelength. The stronger the signal, the longer the optimum wavelength. However, this effect is so small that it is probably not significant.

4. Discussion

4.1. Empirical correlations

It is clear that the presented correlations are purely empirical and that they represent no mathematical relationship between the considered quantities. The reason why this line of analysis was pursued is that there are too many sources of error to consider in a typical diffraction data-collection experiment, which makes strict mathematical treatment extremely cumbersome if not impossible. Nevertheless, the high correlation we managed to extract from the data demonstrate that our empirical approach is working and that we can use these correlations for prediction.

4.2. Estimation of the anomalous diffraction ratio $\Delta F/F$

The $\Delta F/F$ values discussed here have been estimated based on the equation given in Weiss, Sicker & Hilgenfeld (2001) for zero scattering angles and based on the assumption of resolved and independent anomalous scatterers. Furthermore, knowledge of the complete anomalous substructure was required. In the study of an unknown structure one does not necessarily know the complete anomalous substructure. Not knowing about anomalously scattering ions bound to the surface of the protein or misestimating the number and the occupancy of Xe atoms bound to the protein can easily lead to underestimation or overestimation of $\Delta F/F$. However, as is evident from Fig. 5, this will not alter the main conclusion of this study about the optimal wavelength for data collection. It does have an influence, however, if the required redundancy for structure determination is to be estimated (see §4.5).

4.3. Effect of redundancy

In the analysis presented here, we have neglected the influence of the data redundancy on the observed R_{anom}, despite the fact that it has been shown before (Dauter & Adamiak, 2001; Weiss, 2001; Debreczeni *et al.*, 2003) that the redundancy has a significant influence on the observed R_{anom}. For a rotating-anode data set of HEL, for instance (see Table 4 in Weiss, 2001), the observed R_{anom} was decreased from 2.7 to 1.4% when the data redundancy was increased from about 3 to 26. However, all data sets considered exhibit redundancies of 10 or higher, even in the lowest symmetry space groups discussed here (orthorhombic), with a few exceptions in the data sets collected at longer wavelengths (Table 1). At this level of redundancy any further increase in the redundancy will not lead to a further significant lowering of R_{anom} and thus in our experiments R_{anom} can be considered to be independent of the redundancy.

4.4. Radiation damage

One phenomenon we have neglected in our analysis here is the advent of radiation damage, which is in principle capable of altering the anomalous signal significantly owing to destruction of the anomalous substructure. However, as we have shown previously on the data-set series PPE-Xe (Mueller-Dieckmann *et al.*, 2004), radiation damage reduced the anomalous signal by roughly only 2% per data set. Since all data sets considered here have been collected under similar conditions to PPE-Xe, it can be expected that the main conclusions will not be affected significantly by radiation damage, although it is conceivable that individual numbers may change slightly. This notion is actually corroborated by the R_d plots (Diederichs, 2005), which have been calculated for a representative data set of each system (data not shown). In an R_d plot, where the subscript d stands for decay, the values for R_d, which is an R factor based on a comparison of reflection pairs separated by a given number of diffraction images, is plotted against the difference in batch number between the two respective reflections. The occurrence of radiation damage should manifest itself in an increase of R_d with increasing image distance. In all systems except thermolysin the R_d plots are practically flat, indicating that radiation damage is negligible. For thermolysin, the data sets collected are based on five different crystals; thus, it is also improbable

that the damage observed for thermolysin results in a bias of the optimum signal towards certain wavelengths.

4.5. The $R_{\rm anom}/R^{*}_{\rm r.i.m.}$ ratio

As has been discussed before (Weiss, Sicker & Hilgenfeld, 2001; Panjikar & Tucker, 2002; Mueller-Dieckmann *et al.*, 2004), the actual value of the ratio $R_{\rm anom}/R_{\rm p.i.m.}$ can give some indication of whether the information content in the collected data is sufficient for structure determination or not. As a rule of thumb, a value for $R_{\rm anom}/R_{\rm p.i.m.}$ of about 1.5 should yield phases of sufficient quality to allow structure determination to be completed. As a consequence, we should be able to estimate the redundancy necessary for this task. In order to calculate $R_{\rm anom}/R_{\rm p.i.m.}$, the square root of the redundancy N has to be taken into account for each reflection. As a quick estimate, $R_{\rm anom}/R^{*}_{\rm r.i.m.}$ can be multiplied by $N^{1/2}$ to yield $R_{\rm anom}/R_{\rm p.i.m.}$, provided that N is the same or very similar for all reflections. For the four Xe data sets the values of $R_{\rm anom}/R^{*}_{\rm r.i.m.}$ at a wavelength of 2.1 Å is about 0.90. In order to achieve an $R_{\rm anom}/R_{\rm p.i.m.}$ value of 1.5 or larger, a minimum data redundancy of 3–4 would be required. Similarly, for the weaker signals in the P/S/Ca data sets, the minimum redundancy required should be about 10–16. In the future, we will investigate this further in order to determine whether the ratio $R_{\rm anom}/R_{\rm p.i.m.}$ is not only a good predictor for the strength of the anomalous signal, but also for the likelihood at which a structure can be determined.

4.6. Maximum resolution and data-collection strategy

It will of course not always be possible to collect diffraction data at the optimum wavelength of 2.1 Å. This may have device-dependent reasons originating from the synchrotron beamline or it may have strategic reasons, for instance when only a small detector is available and the aim of the data collection is to collect sufficiently high resolution, for instance in order to resolve the S atoms of disulfide bridges. In such a case it is advisable to collect the data at a slightly shorter wavelength, which also makes the data processing more straightforward. The strategy should then be to set the detector at the minimum distance to the crystal and choose the wavelength according to the maximum resolution one wants to collect. The required data redundancy for structure determination can be estimated as discussed in §4.5. With the presence of an increasing number of automated structure-determination pipelines at various synchrotron beamlines, such as *Auto-Rickshaw* (Panjikar *et al.*, 2005) at the EMBL Hamburg beamlines, it will be possible to attempt structure determination while the crystal is still at or near the beamline. The estimated required data redundancy can then be experimentally verified before the experiment is concluded and the crystal removed.

In addition to the long-wavelength data set, it is of course always advisable to collect a short-wavelength data set to the maximum resolution possible, normally on another crystal. This data set can then be of assistance in scaling the long-wavelength data (Mueller-Dieckmann *et al.*, 2004) and can also provide the basis for refinement to the highest resolution, because it suffers less from systematic errors in the reflection intensities.

5. Summary and conclusions

From the data presented here, it is clear that the wavelength at which the highest anomalous signal-to-noise ratio can be obtained using standard means of data collection and processing is about 2.1 Å. This is nearly independent of the composition of the anomalously scattering substructure, provided that no dominating absorption edge is present in the wavelength range considered.

Longer wavelengths do not only provide an increased anomalous signal for phase determination, they also allow a much clearer definition of substructures. In particular, the positions and occupancies of phosphate, sulfate, chloride, calcium *etc*. can be established reliably. In many cases this information may turn out to be very important for elucidating the function of a molecule. For this reason, we would like to recommend complementing every crystal structure determination with a properly collected and processed long-wavelength data set.

We would like to thank Maurizio Polentarutti (ELETTRA Trieste, Italy) and Dr Kristina Djinovic Carugo (University of Vienna, Austria) for their help in collecting many of the longer wavelength data sets and for many stimulating discussions. We would also like to acknowledge the support of this work by the EC 5th Framework Programmes 'Transnational Access to Major Research Infrastructures' (Contract No. HPRI-CT-1999-00033) and 'Quality of Life and Management of Living Resources' (Integrated Research Project SPINE, Project No. QLG2-CT-2002-00988), the RTD-Project EXMAD (Contract No. HPRI-CT-1999-50015) and the EC 6th Framework Programme 'Life Sciences, Genomics and Biotechnology for Health' (Integrated Research Project BIOXHIT, Contract No. LHSG-CT-2003-503420) as well as the Deutsche Forschungsgemeinschaft (DFG grant WE2520/2 to MSW). All data described in this work including the diffraction images will be available to the scientific community upon request.

References

Arndt, U. W. (1984). *J. Appl. Cryst.* **17**, 118–119.
Brown, J., Esnouf, R. M., Jones, M. A., Linnell, J., Harlos, K., Hassan, A. B. & Jones, E. Y. (2002). *EMBO J.* **21**, 1054–1062.
Charron, C., Kadri, A., Robert, M.-C., Giegé, R. & Lorber, B. (2002). *Acta Cryst.* D**58**, 2060–2065.
Chen, L., Chen, L. R., Zhou, X. E., Wang, Y., Kahsai, M. A., Clark, A. T., Edmondson, S. P., Liu, Z. J., Rose, J. P., Wang, B.-C., Meehan, E. J. & Shriver, J. W. (2004). *J. Mol. Biol.* **341**, 73–91.
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.
Dauter, Z. & Adamiak, D. A. (2001). *Acta Cryst.* D**57**, 990–995.
Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.

Debreczeni, J. É., Bunkóczi, G., Ma, O., Blaser, H. & Sheldrick, G. M. (2003). *Acta Cryst.* D**59**, 688–696.
Diederichs, K. (2005). *Acta Cryst.* D. Submitted.
Djinovic Carugo, K., Helliwell, J. R., Stuhrmann, H. & Weiss, M. S. (2005). *J. Synchrotron Rad.* **12**, 410–419.
Einspahr, H., Suguna, K., Suddath, F. L., Ellis, G., Helliwell, J. R. & Papiz, M. Z. (1985). *Acta Cryst.* B**41**, 336–341.
Ferreira, K. N., Iverson, T. M., Maghlaoui, K., Barber, J. & Iwata, S. (2004). *Science*, **303**, 1831–1838.
Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
Kent, H. M., McMahon, H. T., Evans, P. R., Benmerah, A. & Owen, D. J. (2002). *Structure*, **10**, 1139–1148.
Kuettner, E. B., Hilgenfeld, R. & Weiss, M. S. (2002). *J. Biol. Chem.* **277**, 46402–46407.
Lemke, C. T., Smith, G. D. & Howell, P. L. (2002). *Acta Cryst.* D**58**, 2096–2101.
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.
Mueller-Dieckmann, C., Polentarutti, M., Djinovic Carugo, K., Panjikar, S., Tucker, P. A. & Weiss, M. S. (2004). *Acta Cryst.* D**60**, 28–38.
Olsen, J. G., Flensburg, C., Olsen, O., Bricogne, G. & Henriksen, A. (2004). *Acta Cryst.* D**60**, 250–255.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. (2005). *Acta Cryst.* D**61**, 449–457.
Panjikar, S. & Tucker, P. A. (2002). *J. Appl. Cryst.* **35**, 261–266.
Phillips, J. D., Whitby, F. G., Warby, C. A., Labbe, P., Yang, C., Pflugrath, J. W., Ferrara, J. D., Robinson, H., Kushner, J. P. & Hill, C. P. (2004). *J. Biol. Chem.* **279**, 38960–38968.
Polikarpov, I. (1997). *J. Synchrotron Rad.* **4**, 17–20.
Polikarpov, I., Teplyakov, A. & Oliva, G. (1997). *Acta Cryst.* D**53**, 734–737.
Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003). *Acta Cryst.* D**59**, 1020–1027.
Schiefner, A. (2000). Diploma thesis. University of Jena, Jena, Germany.
Sekar, K., Rajakannan, V., Velmurugan, D., Yamane, T., Thirumurugan, R., Dauter, M. & Dauter, Z. (2004). *Acta Cryst.* D**60**, 1586–1590.
Teplyakov, A., Oliva, G. & Polikarpov, I. (1998). *Acta Cryst.* D**54**, 610–614.
Tippin, D. B. & Sundaralingam, M. (1996). *Acta Cryst.* D**52**, 997–1003.
Weiss, M. S. (2001). *J. Appl. Cryst.* **34**, 130–135.
Weiss, M. S., Mander, G., Hedderich, R., Diederichs, K., Ermler, U. & Warkentin, E. (2004). *Acta Cryst.* D**60**, 686–695.
Weiss, M. S., Palm, G. J. & Hilgenfeld, R. (2000). *Acta Cryst.* D**56**, 952–958.
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.
Yang, C. & Pflugrath, J. W. (2001). *Acta Cryst.* D**57**, 1480–1490.
Yang, C., Pflugrath, J. W., Courville, D. A., Stence, C. N. & Ferrara, J. D. (2003). *Acta Cryst.* D**59**, 1943–1957.