Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Single isomorphous replacement phasing of selenomethionine-containing proteins using UV-induced radiation damage

Santosh Panjikar,[a]* Hubert Mayerhofer,[a] Paul A. Tucker,[a] Jochen Mueller-Dieckmann[a] and Daniele de Sanctis[b]

[a]EMBL Hamburg Outstation, c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany, and [b]ESRF, Structural Biology Group, 6 Rue Jules Horowitz, 38043 Grenoble CEDEX, France

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

Received 15 July 2010
Accepted 21 October 2010

The most commonly used heavy-atom derivative, selenium, requires the use of a tunable beamline to access the Se *K* edge for experimental phasing using anomalous diffraction methods, whereas X-ray diffraction experiments for selenium-specific ultraviolet radiation-damage-induced phasing can be performed on fixed-wavelength beamlines or even using in-house X-ray sources. Several nonredundant X-ray diffraction data sets were collected from three different selenomethionine (Mse) derivatized protein crystals at energies far below the absorption edge before and after exposing the crystal to ultraviolet (UV) radiation using 266 nm lasers of flux density 1.7×10^{15} photons s^{-1} mm^{-2} for 10–50 min. A detailed analysis revealed that significant changes in diffracted intensities were induced by ultraviolet irradiation whilst retaining crystal isomorphism. These intensity changes allowed the crystal structures to be solved by the radiation-damage-induced phasing (RIP) technique. Inspection of the crystal structures and electron-density maps demonstrated that covalent bonds between selenium and carbon at all sites located in the core of the proteins or in a hydrophobic environment were much more susceptible to UV radiation-induced cleavage than other bonds typically present in Mse proteins. The rapid UV radiation-induced bond cleavage opens a reliable new paradigm for phasing when no tunable X-ray source is available. The behaviour of Mse derivatives of various proteins provides novel insights and an initial basis for understanding the mechanism of selenium-specific UV radiation damage.

1. Introduction

Isomorphous replacement was the first phasing technique in macromolecular crystallography. This phasing method was introduced by Max F. Perutz in 1954 (Bragg & Perutz, 1954) and is still popular for determining the phases of structure factors of Bragg reflections from a macromolecular crystal using the single or multiple isomorphous replacement (SIR or MIR) methods. An extension of this method is the use of the anomalous signal (AS) from a variety of derivatives (*e.g.* SIRAS and MIRAS) by measuring diffraction data at wavelengths close to an absorption edge of a derivative element. These methods depend not only on the nature of the derivative but also on the isomorphism between the native and derivative crystals. Whether the degree of isomorphism is sufficient for phasing or not can only be tested by X-ray diffraction experiments.

Recent studies have demonstrated that the intensity differences induced during an X-ray scattering experiment, which are generally understood to be one of the effects of radiation damage, can be successfully exploited for phasing (Ravelli *et al.*, 2003). This method (radiation-damage-induced phasing; RIP) has the advantage that data are collected from a single crystal of the native macromolecule, largely minimizing the problems of non-isomorphism (Ravelli *et al.*, 2003). Variations of this method coupled with anomalous scattering, termed radiation-damage-induced phasing with anomalous scattering (RIPAS), are also suitable for phasing macromolecular crystallographic data (Zwart *et al.*, 2004; Nanao *et al.*, 2005; Abbani *et al.*, 2007; Holton, 2009; Garman, 2010). The reported applications of radiation damage in phasing have been based on the site-specific effects on sulfurs in disulfide bridges (Ravelli *et al.*, 2003; Banumathi *et al.*, 2004; Weiss *et al.*, 2004), triiodides (Evans *et al.*, 2003), brominated uridine (Ravelli *et al.*, 2003; Schiltz *et al.*, 2004) and mercury derivatives (Ramagopal *et al.*, 2005). In all these cases, phasing macromolecular X-ray data exploits the site-specific effects of radiation damage by splitting up a high-redundancy data set collected from a single crystal into smaller parts and using the resulting isomorphous intensity differences for phasing in a traditional SIR fashion to obtain the positions of the substructure *via* direct or Patterson methods. However, a drawback of the X-ray-based RIP method is that X-rays introduce many small changes to both solvent and macromolecule which complicate the comprehensive location of the pertinent sites (Nanao & Ravelli, 2006).

More recently, ultraviolet (UV) radiation has been shown to induce specific changes in the macromolecule alone. This technique allows an elegant phasing scheme that leads to precise experimental phase information (Nanao & Ravelli, 2006). The most striking similarity is that UV radiation, like X-ray radiation, breaks disulfide bonds. X-rays, however, do not only affect disulfide bonds but also carboxyl groups, including the C-terminus, and ordered waters surrounding the protein. In contrast, suitable UV radiation only shows structural changes at the disulfide bonds. A number of novel structures containing disulfides have been determined using UV and X-ray RIP (Rudiño-Piñera *et al.*, 2007; Schönfeld *et al.*, 2008; Fütterer *et al.*, 2008). The approach has been extended to a non-disulfide-bridge-containing protein (photoactive yellow protein) which contains a chromophore, *p*-coumaric acid, covalently bound through a thioester linkage to a cysteine. Upon UV irradiation the sulfur–carbon bond is disrupted (Nanao & Ravelli, 2006).

UV light can be damaging to proteins by photolysis or photo-oxidation mechanisms (Dose, 1968; Permyakov, 1993; Vernede *et al.*, 2006; Kehoe *et al.*, 2008). However, short exposure (less than 1 s) of protein crystals to UV light does not generate detectable structural damage. Use of short-pulse UV laser-excited fluorescence has been proposed as a tool for the visualization of protein crystals mounted in loops for detection and centring. The protein crystals strongly absorb UV light. 99% of the photons at 266 nm are absorbed within a layer of 50 µm thickness in protein crystals (assuming a 10 m*M* protein concentration in crystals of a protein of 500 amino acids with 1.32% tryptophan; Vernede *et al.*, 2006). The involvement of selenomethionine (Mse) in quenching Trp fluorescence has been demonstrated for a calmodulin–target complex and it was shown that the Trp fluorescence intensity in the complex is proportional to the atomic weight of the atom in place of the sulfur in methionine. The effectiveness of quenching proceeds in the order Se > S > C (Weljie & Vogel, 2000). It was also shown that substitution of S by Se causes enhanced protein absorption in the UV region (Giordano *et al.*, 2004). This led us to believe that selenium-containing groups may be preferentially damaged with a UV laser and can be used for experimental phasing in macromolecular crystallography. Since Mse is the most popular heavy atom in X-ray crystallography for SAD/MAD phasing (Ogata, 1998), which requires accessibility and tunability around the selenium absorption edge at a synchrotron beamline (Hendrickson, 1991), we think that the applicability of Mse may be exploited even more effectively by using the UV-RIP method, which does not require tunability around the selenium absorption edge.

In this paper, we show that Mse protein crystals can be sufficiently damaged by UV radiation either in the presence or the absence of tryptophan and can be used for UV-RIP. The method exploits the apparent 'depletion' of Mse caused by radiation with a UV laser. Complete data sets were collected from Mse protein crystals before and after UV radiation. In each case the differences between the two data sets were successfully used for phase determination. We also propose a minimum dose of UV radiation for successful exploitation of the UV-RIP method.

2. Experimental procedure

2.1. Proteins used

Three different selenium-labelled proteins, the feruloyl esterase (FAE) module of xylanase 10B from *Clostridium hermocellum* (PDB code 1gkk; Prates *et al.*, 2001), H35, a recently characterized protein from our laboratory, and chorismate synthase (CHSYNT) from *Mycobacterium tuberculosis* (PDB code 2o11; M. Bruning, G. P. Bourenkov, N. I. Strizhov & H. D. Bartunik, unpublished work), were selected for this study. Mse-derivatized FAE crystallizes in space group $P2_12_12_1$. It contains two molecules in the asymmetric unit, with a solvent content of 58%. Each molecule is composed of 297 residues and contains no disulfide bridges, but contains three free Cys residues, eight Mse residues, four Cd atoms, 25 Tyr residues, 18 Phe residues and four Trp residues per molecule. Purification and crystallization protocols have been reported by Prates *et al.* (2001). The dimensions of the crystal used in this experiment were about 300 × 50 × 40 µm. Mse-containing H35 crystallizes in space group $P2_12_12$, with unit-cell parameters a = 65.9, b = 69.2, c = 108.3 Å. The asymmetric unit contains four molecules and the solvent content is 35%. Each protomer consists of 98 residues, including seven Mse residues. The protein does not contain Cys, Trp or Tyr, but does contain

two Phe residues per molecule (Mayerhofer *et al.*, in preparation). The crystal dimensions were 150 × 50 × 50 µm. Mse-derivatized CHSYNT contains 407 residues. It crystallizes in space group $P6_422$, with one molecule in the asymmetric unit and a solvent content of 73%. 11 Mse residues are present (Bruning *et al.*, in preparation) as well as three Trp, seven Tyr and six Phe residues. The crystal dimensions were 80 × 80 × 50 µm. The crystallization conditions of the three proteins did not contain chemicals that strongly absorb UV radiation at 266 nm. The FAE protein was crystallized in the presence of cadmium acetate. This chemical does not show significant UV absorption (Olmo *et al.*, 2001), but does absorb X-rays strongly.

2.2. Measurement of the UV absorption spectrum of Mse and simulation of the penetration depth of protein crystals

The UV absorption spectra of Met, Mse, Cys and all aromatic amino acids (Trp, Tyr and Phe) were measured using an ND-1000 NanoDrop spectrophotometer. The spectra of Met and Mse were measured in a similar manner to that in which the spectra of the aromatic amino acids were measured. This was to ensure that all spectra, irrespective of the amino acid, remained on the same scale and could be used for cross-validation of the measurements. From the spectrum (Fig. 1) it is clear that the Mse absorbs UV. It does not contain a maximum at 230 nm as many other amino acids do; rather, the maximum absorption peak is located at 246 nm. Mse absorbs significantly more strongly than Tyr at the 266 nm wavelength used.

Fig. 2 shows a simulation of the transmission of 266 nm light across native and selenomethionine protein crystals. The concentration of the protein in the crystals was calculated to be 10, 32.5 and 5 m*M* for FAE, H35 and CHSYNTH, respectively. The values for the extinction coefficients of aromatic amino acids (Trp, Tyr and Phe), Cys and Mse at 266 nm used in the calculation were taken from the measured values shown in Fig. 1. Molar extinction coefficients of 48 036, 216 and 22 950 M^{-1} cm^{-1} for the native FAE, H35 and CHSYNTH proteins, and of 64 902, 13 334 and 43 564 M^{-1} cm^{-1} for the Mse proteins, respectively, at 266 nm were calculated. The penetration depth of UV light for the Mse crystals of FAE, H35 and CHSYNTH is expected to be about 40, 60 and 100 µm, respectively (Fig. 2).

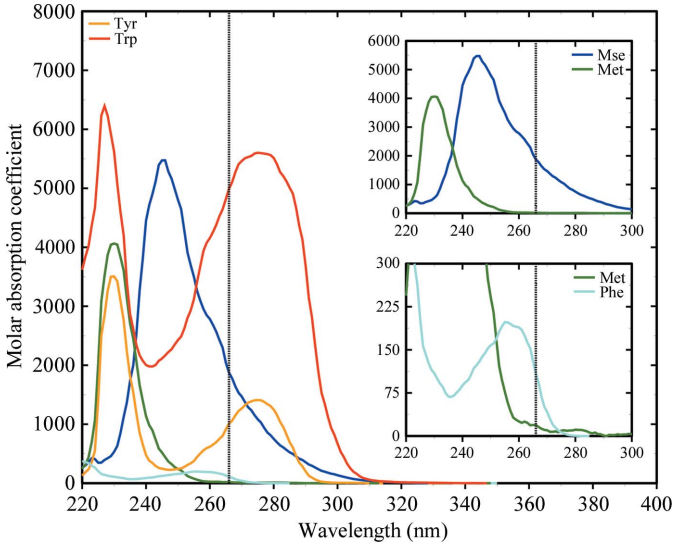

Figure 1
Plot of the UV absorption spectra of Trp, Tyr, Phe, Met and Mse residues against the wavelength; the absorption spectra were measured using an ND-1000 NanoDrop spectrophotometer. The dotted line in the graph indicates the 266 nm wavelength that was used for UV irradiation of the protein crystals.

2.3. Beamline arrangement and UV laser setup

All data were collected on European Synchrotron Radiation Facility (ESRF) beamline ID23-1 (Nurizzo *et al.*, 2006). ID23-1 is a fully tuneable beamline (5.2–20 keV) with a typical photon flux of about 5 × 10^{12} photons s^{-1} at 12 keV and a typical spot size of 40 × 30 µm at the sample position. A 266 nm pulsed microchip laser (TEEM Photonics; SNU-02p) was installed in the back cover of the mini-diffractometer. The average power of the laser source was 5 mW, the repetition rate was 7 kHz and the pulse width was 400 ps. The UV laser was reflected with a 45° mirror in the direction of the sample through the on-axis viewer (for reference, see Fig. 1*b* of Vernede *et al.*, 2006).

UV and X-ray exposures were controlled with a multi-purpose unit for synchronization, sequencing and triggering (MUSST), which was custom-built at the ESRF, with a normal user-defined oscillation angle. The user interface for UV exposure control was implemented as an additional brick in *MxCuBE* (Gabadinho *et al.*, 2010). The required fields are UV exposure time and sample-rotation range. UV exposure can be carried out for the chosen time (in seconds) both across the chosen angular oscillation range and an equivalent oscillation range 180° away in order to maximize the volume of the crystal exposed to the laser. The resulting UV spot on the crystal is larger than the X-ray beam and, since this is a laser, it is expected that the spot size at the sample is the same as that

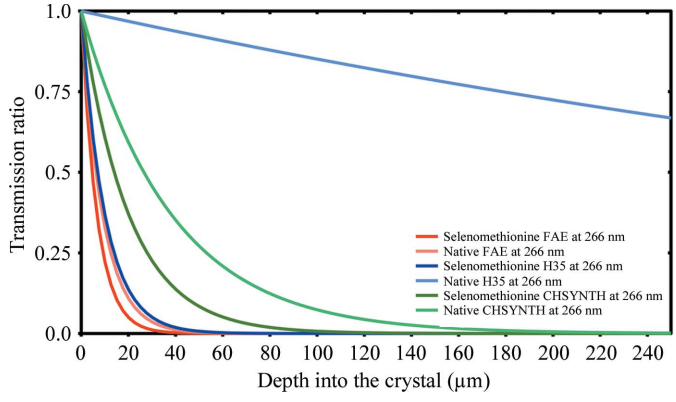

Figure 2
Simulated transmission of 266 nm UV laser light across native (light) and selenomethionine (dark) protein crystals of FAE (red), H35 (blue) and CHSYNTH (green).

before the reflecting mirror. Its intensity was measured to be about 1.4 mW at the sample position, corresponding to about 10^{15} photons s^{-1} over an area of 880 × 670 µm, so the flux density is 1.7×10^{15} photons s^{-1} mm^{-2}.

2.4. Data-collection strategy and processing

Diffraction data sets from FAE and CHSYNTH crystals were collected at a wavelength of 1.0332 Å and those from H35 crystals were collected at a wavelength of 0.9919 Å, which is below the absorption edge of selenium. These crystals were flash-cooled in a 100 K Oxford Cryosystems 700 nitrogen-gas stream prior to data collection.

The data-collection strategy was determined using the program *BEST* (Bourenkov & Popov, 2010) based on two diffraction images (90° apart) collected using 20% of the incident X-ray flux upon each protein crystal. The minimum degree of oscillation and X-ray dose suggested by *BEST* were determined prior to UV exposure and the same oscillation range and X-ray dose were maintained after UV exposure. Between adjacent data sets, the crystals were exposed to UV light while being rotated through the same rotation range as used for X-ray data collection. The X-ray and UV beams were coaxial.

Six data sets were collected from each crystal of the three Mse-labelled proteins. 100 frames were collected per data set

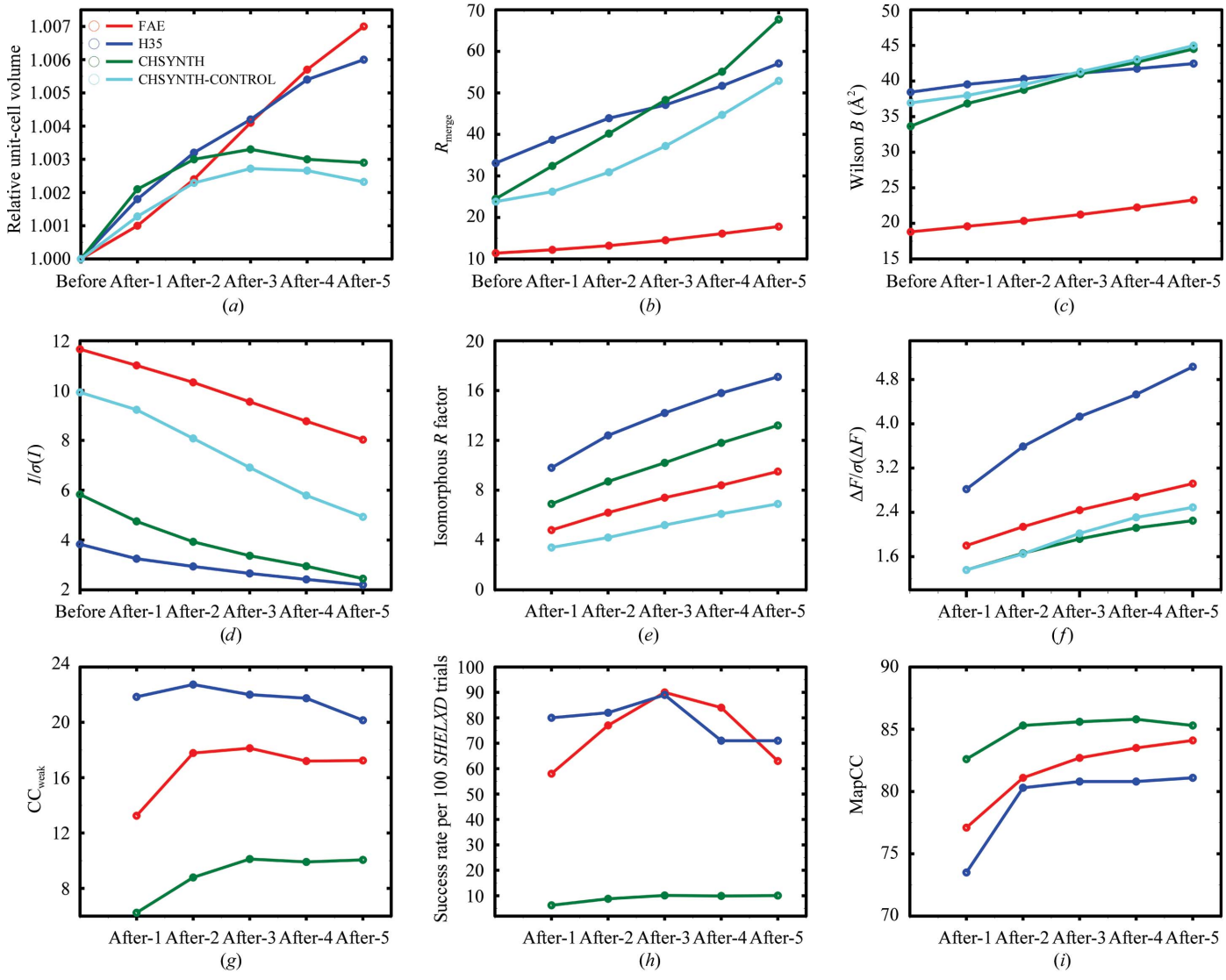

Figure 3
(*a*) Plot of relative unit-cell volume. The plot for the FAE data is shown in red, that for H35 in blue, that for CHSYNTH in green and that for CHSYNTH-CONTROL in cyan. The same colouring scheme is used in the subsequent plots. Data collected before UV radiation are marked 'before'. Data collected after successive 10 min UV exposures are marked 'after-1' to 'after-5'. The same data-set names are used in the subsequent plots. (*b*) Plot of R_{merge} for the highest resolution shell of the processed data. (*c*) Plot of overall Wilson *B* factor. (*d*) Plot of $I/\sigma(I)$ for the highest resolution shell of the processed data. (*e*) Plot of cross *R* factor between 'before' and 'after-*i*' (*i* = 1–5) data sets. (*f*) Plot of *SHELXC* output, showing overall isomorphous signal $\langle d'/\mathrm{sig}\rangle$ as $\Delta F/\sigma(\Delta F)$. (*g*) Plot of CC_{weak} from *SHELXD* for each pair ('before' and 'after-*i*', where *i* = 1–5) of RIP data sets. (*h*) Plot of success rate per 100 *SHELXD* trials for each pair of data sets. (*i*) Plot of map correlation calculated between the phases generated from each pair of RIP data with the final model phase generated from the 'before' data set.

Table 1
Data-collection, processing, phasing and refinement statistics for the FAE crystal.

Values in parentheses are for the highest resolution shell.

Data	Before	After-1	After-2	After-3	After-4	After-5
Wavelength (Å)	1.0332					
Total frames per data set	100					
X-ray dose per data set (MGy)	1.99					
Space group	$P2_12_12_1$					
Unit-cell parameters (Å)	a = 65.25, b = 108.37, c = 113.09	a = 65.27, b = 108.42, c = 113.13	a = 65.29, b = 108.47, c = 113.18	a = 65.33, b = 108.53, c = 113.25	a = 65.36, b = 108.58, c = 113.31	a = 65.39, b = 108.63, c = 113.36
Resolution (Å)	2.00					
Mosaicity (°)	0.136	0.133	0.130	0.128	0.126	0.124
Total reflections	220315 (26528)	221048 (25570)	221087 (24615)	221946 (23421)	222490 (21896)	222584 (20321)
Unique reflections	54331 (6332)	54488 (6078)	54515 (5843)	54727 (5525)	54864 (5152)	54892 (4763)
Multiplicity	4.06 (4.19)	4.06 (4.21)	4.06 (4.21)	4.06 (4.24)	4.06 (4.25)	4.05 (4.27)
Completeness (%)	99.2 (72.3)	99.2 (69.5)	99.2 (66.8)	99.1 (62.8)	99.1 (58.4)	99.0 (53.90)
$I/\sigma(I)$	19.16 (11.66)	18.52 (11.03)	17.74 (10.33)	16.66 (9.55)	15.77 (8.77)	14.84 (8.03)
$R_{\rm merge}$ (%)	5.7 (11.4)	5.9 (12.2)	6.2 (13.2)	6.6 (14.5)	7.0 (16.1)	7.5 (17.8)
$R_{\rm meas}$ (%)	6.5 (13.1)	6.8 (14.0)	7.1 (15.1)	7.6 (16.6)	8.0 (18.4)	8.7 (20.4)
Wilson B (Å^2)	18.81	19.59	20.35	21.24	22.24	23.29
Phasing						
$\Delta F/\sigma(\Delta F)$	—	1.80	2.14	2.44	2.68	2.92
Isomorphous R factor (%)	—	4.8	6.2	7.4	8.4	9.5
Scale factor	—	0.963	0.964	0.963	0.964	0.964
Resolution cutoff (Å)	—	2.3	2.3	2.3	2.3	2.3
$CC_{\rm all}$	—	22.57	29.92	31.70	31.14	30.50
$CC_{\rm weak}$	—	13.25	17.78	18.13	17.20	17.24
No. of selenium sites	—	16				
After 5 iterations using *SHELXE* (total sites)	—	100	82	82	81	79
After 5 iterations using *SHELXE* (negative sites)	—	6	14	16	21	18
No. of residues built in *SHELXE*	—	542	547	543	559	550
Phase error (°)/map CC (%)	—	42.50/77.1	39.30/81.1	38.08/82.7	37.42/84.3	36.94/84.1
Refinement						
Refinement resolution range (Å)	20–2.0					
$R_{\rm work}$ (%)	13.90					
$R_{\rm free}$ (%)	17.61					
Reflections, working	52836					
Reflections, free	988					
Non-H atoms	5437					
Water molecules	759					
Average B factor (Å^2)	14.15					
R.m.s.d. bond lengths (Å)	1.34					
R.m.s.d. bond angles (°)	0.018					

in 10 s from the FAE and H35 crystals and a total of 40 frames were collected per data set in 8 s from the CHSYNTH crystal. Between two data collections, the crystal was exposed to the UV laser for 10 min. The first data set is referred to as 'before' and the second to sixth data sets are referred to as 'after-1', 'after-2', 'after-3', 'after-4' and 'after-5', respectively. A total of 1.9, 1.36 and 0.49 MGy X-ray dose per data set was used for the FAE, H35 and CHSYNTH crystals, respectively. The collection of multiple data sets was undertaken in order to obtain an idea about the minimum accumulated UV exposure time required for successful phasing and to obtain some idea as to the extent of radiation damage. Doses have been calculated with *RADDOSE* (Paithankar & Garman, 2010).

A control experiment was performed on a CHSYNTH crystal in order to assess the nature of the X-ray damage. This experiment was designated CHSYNTH-CONTROL. The diffraction data sets were collected at a wavelength of 1.0332 Å and were collected with 1° rotation per frame. Six data sets were collected. Each data set consisted of 100 frames collected in 15 s. The total X-ray dose per data set was 2.3 MGy. The first data set is referred to as 'before' and the second to sixth data sets are referred to as 'after-1', 'after-2', 'after-3', 'after-4' and 'after-5', respectively, as described above, except that the crystal was not exposed to the UV laser between two data collections.

All data sets were processed and scaled in *XDS*/*XSCALE* (Kabsch, 2010) using the input file provided by the beamline-control graphical user interface *MxCuBE*. After data processing and scaling, the intensities were converted to structure factors using the *CCP*4 program *TRUNCATE* (French & Wilson, 1978).

2.5. Substructure determination, phasing and model building

The resolution for phasing was chosen such that $\langle\Delta F/\sigma(\Delta F)\rangle$ was greater than 1.5. The number of sites for substructure determination was set to be equal to the known number of Mse residues (which, if it were a correct assumption, would be a great advantage in substructure determination). ΔF values were calculated using the RIP option in

SHELXC (Sheldrick, 2008) and were used in *SHELXD* (Schneider & Sheldrick, 2002) to determine the most susceptible part of the specific radiation-damage substructure. The overall $\Delta F/\sigma(\Delta F)$ seems to increase monotonically over the five UV radiation exposures of 10 min each (Fig. 3*f*). The graph suggests that substantial changes were induced in the very first 'after' data set and continued to increase with UV dose.

All *SHELXD* processes were run for 100 cycles in Patterson seeding mode. The quality of the results was checked by examination of (i) the best correlation coefficient between the observed and the calculated weak E values $CC_{weak}(E_{obs}, E_{calc})$ (Fig. 3*g*) and (ii) the number of correct solutions per 100 trials (Fig. 3*h*). The solutions from *SHELXD* were subsequently submitted to *SHELXE* (Sheldrick, 2010) with 20 cycles of density modification. Difference Fourier analysis within *SHELXE* was used to update the RIP substructure, which was resubmitted to *SHELXE*, and the procedure was repeated five times. This allowed stepwise improvement of the substructure. The final RIP substructures, including both positive and negative sites, were then run through a beta version of *SHELXE* (Sheldrick, 2010) with four cycles of autotracing, in which each tracing cycle included 50 cycles of density modification. The entire process of ΔF calculation (*SHELXC*) to phase and substructure improvement (*SHELXE*) was performed using a scale factor K, which was calculated using the empirical formula rw + [(1 − rw)/1.5], where rw is the scale factor calculated using *SCALEIT* (Collaborative Computational Project, Number 4, 1994) between the 'after' and the 'before' data set. The scale factor is used to downscale the radiation-damage-induced data set for better RIP substructure determination (Nanao *et al.*, 2005). The resulting polyalanine model and the phases from *SHELXE* were then used in *ARP*/*wARP* (Perrakis *et al.*, 1999) as a starting model and restraining phase during iterative refinement for model building and side-chain docking.

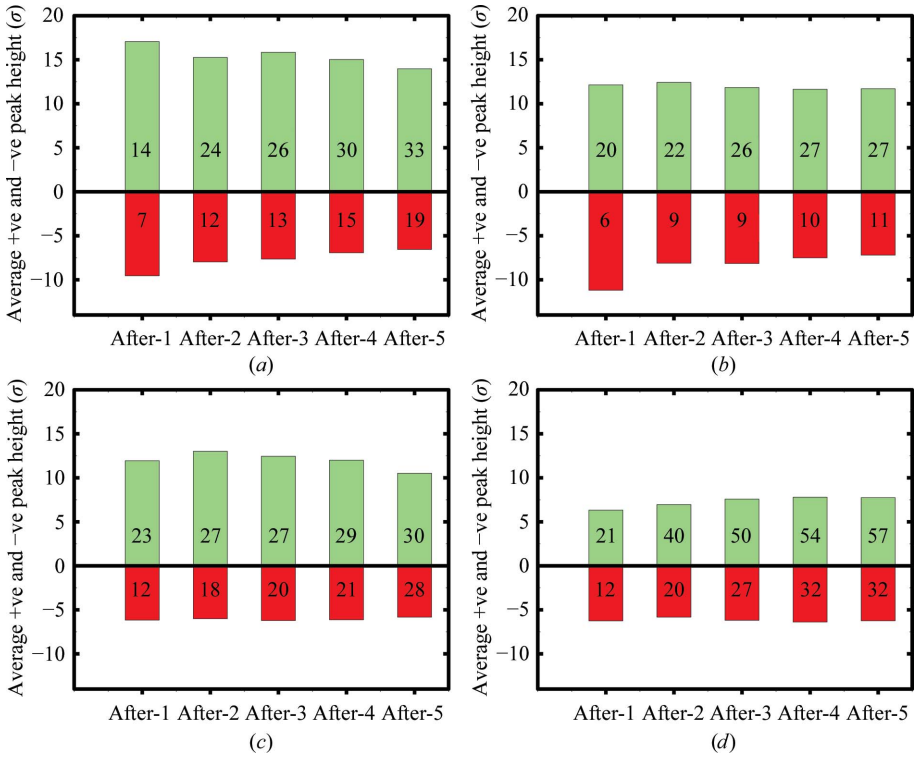

Figure 4
(*a*) Histogram of average peak height of positive and negative difference electron-density ($F_{before} - F_{after\text{-}i}, \varphi_{calc}$) map (where i = 1–5). The y axis shows the average peak height in σ. The total numbers of positive and negative peaks found are given in the green and red bars, respectively, for FAE. The same colouring scheme is used for (*b*) H35, (*c*) CHSYNTH and (*d*) CHSYNTH-CONTROL.

2.6. Refinement and difference Fourier analysis

The models of FAE, H35 and CHSYNTH resulting from RIP phasing using the corresponding 'before' and 'after-1' data sets were used for further model completion in the graphics program *Coot* (Emsley & Cowtan, 2004). These models were refined against the 'before' data set. The model was iteratively improved and alternated with refinement in *REFMAC*5 (Murshudov *et al.*, 1999). Refinement statistics for each model are listed in Tables 1, 2, 3 and 4.

For each pair of data sets ('before' and 'after-*i*', where i = 1–5), the isomorphous difference Fourier maps were computed using the final phase from the refined model to identify 'radiation-damaged' atoms. The extent of radiation damage is compared both on an absolute (e^{-1} Å^{-3}) and a relative (map r.m.s.) scale (Supplementary Table S1[1]). The final phase was also used to assess the quality of the experimental phases using the program *CPHASEMATCH* (Collaborative Computational Project, Number 4, 1994).

3. Results

3.1. Control experiment

A control experiment was performed using CHSYNTH crystals and data were collected to 2.7 Å resolution. The monotonic increase in unit-cell dimensions, Wilson B factor and the R_{merge} value in the highest resolution shell and the decrease of $I/\sigma(I)$ in the highest resolution shell (Figs. 3*a*–3*d*; Table 4) showed that the crystals had suffered from X-ray radiation damage during each step of the data collection.

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: BE5155). Services for accessing this material are described at the back of the journal.

Table 2
Data-collection, processing, phasing and refinement statistics for the H35 crystal.

Values in parentheses are for the highest resolution shell.

Data	Before	After-1	After-2	After-3	After-4	After-5
Wavelength (Å)	0.9919					
Total frames per data set	100					
X-ray dose per data set (MGy)	1.36					
Space group	$P2_12_12$					
Unit-cell parameters (Å)	a = 65.92, b = 69.20, c = 108.27	a = 65.93, b = 69.26, c = 108.35	a = 65.95, b = 69.30, c = 108.41	a = 65.97, b = 69.32, c = 108.45	a = 65.99, b = 69.35, c = 108.50	a = 66.00, b = 69.37, c = 108.53
Resolution (Å)	2.15					
Mosaicity (°)	0.222	0.216	0.214	0.214	0.214	0.213
Total reflections	109136 (16951)	109304 (16974)	109807 (17081)	110040 (17173)	110166 (17115)	110350 (17038)
Unique reflections	27137 (4328)	27161 (4345)	27282 (4384)	27344 (4397)	27356 (4391)	27403 (4374)
Multiplicity	4.02 (3.92)	4.02 (3.91)	4.02 (3.90)	4.02 (3.91)	4.03 (3.90)	4.03 (3.90)
Completeness (%)	98.1 (98.6)	98.1 (99.0)	98.0 (98.6)	98.0 (98.8)	97.9 (98.6)	97.9 (98.3)
$I/\sigma(I)$	19.29 (3.83)	18.45 (3.25)	17.60 (2.94)	17.19 (2.66)	16.57 (2.42)	16.20 (2.20)
R_{merge} (%)	5.0 (33.1)	5.3 (38.7)	5.6 (43.9)	5.8 (47.1)	6.0 (51.7)	6.1 (57.1)
R_{meas} (%)	5.7 (38.0)	6.0 (44.4)	6.4 (50.4)	6.6 (54.1)	6.8 (59.3)	7.0 (65.5)
Wilson B (Å^2)	38.44	39.53	40.30	41.12	41.73	42.45
Phasing						
$\Delta F/\sigma(\Delta F)$	—	2.82	3.59	4.13	4.53	5.03
Isomorphous R factor (%)	—	9.8	12.4	14.2	15.8	17.1
Scale factor	—	0.957	0.960	0.963	0.965	0.962
Resolution cutoff (Å)	—	2.4	2.4	2.4	2.4	2.4
CC_{all}	—	40.80	42.24	41.30	39.91	38.58
CC_{weak}	—	21.84	22.73	22.00	21.74	20.15
No. of selenium sites	—	28				
After 5 iterations using *SHELXE* (total sites)	—	54	43	46	43	49
After 5 iterations using *SHELXE* (negative sites)	—	12	14	15	11	17
No. of residues built using *SHELXE*	—	300	330	331	333	326
Phase error (°)/map CC (%)	—	55.45/73.5	48.07/80.3	47.74/80.8	47.65/80.8	47.58/81.1
Refinement						
Refinement resolution range (Å)	20–2.15					
R_{work} (%)	24.1					
R_{free} (%)	27.4					
Reflections, working	26049					
Reflections, free	1039					
Non-H atoms	3119					
Water molecules	86					
Average B factor (Å^2)	27.9					
R.m.s.d. bond lengths (Å)	0.018					
R.m.s.d. bond angles (°)	1.725					

Although the isomorphous R factor and the overall $\Delta F/\sigma(\Delta F)$ both increased over the five successive data sets (Figs. 3*e* and 3*f*), a correct substructure could not be determined even after 10 000 trials of *SHELXD* using the 'before' and any 'after' data sets.

The 'before' data set was used to refine a CHSYNTH model. The model phase was used for isomorphous difference Fourier analysis for all 'after' data sets. The analysis shows that the subsequent data were indeed affected by X-ray radiation damage. However, the average positive peak height in the isomorphous difference Fourier map for each 'after' data set is about half of the average positive peak height shown for the CHSYNTH data sets from the UV experiment (Fig. 4). It should be noted that the X-ray dose used for each data set of CHSYNTH in the UV experiment was 4.7 times lower than that used in the control experiment (although it would have been better if it were the same). It can be inferred that the expected X-ray damage in the UV experiment is minimal and does not contribute significantly to phasing.

3.2. X-ray data quality and analysis of the UV experiment

All three of the 'before' diffraction data sets are of rather good quality (Tables 1, 2 and 3). The maximum resolution ranges from 2.70 Å for CHSYNTH to 2.00 Å for FAE. All 15 'after' data sets also show reasonable data quality. The monotonic increase in unit-cell dimensions, Wilson B factor and the R_{merge} value in the highest resolution shell and the decrease in $I/\sigma(I)$ in the highest resolution shell (Figs. 3*a*–3*d*, Tables 1, 2 and 3) indicated that the crystals had suffered from radiation damage during each step of UV exposure. This also includes some X-ray radiation damage arising from the increase in the accumulative X-ray dose to the protein crystals. The 'before' and 'after-*i*' (where i = 1–5) data sets were scaled together on a common scale using *SCALEIT* and the merging statistics on structure-factor amplitudes shows that substantial differences have evolved between the respective data sets (Fig. 3*e*, Tables 1, 2 and 3). In comparison to FAE and H35, the CHSYNTH crystal shows inferior data-collection statistics

Table 3
Data-collection, processing, phasing and refinement statistics for the CHSYNTH crystal.

Values in parentheses are for the highest resolution shell.

Data	Before	After-1	After-2	After-3	After-4	After-5
Wavelength (Å)	1.0332					
Total frames per data set	40					
X-ray dose per data set (MGy)	0.49					
Space group	$P6_422$					
Unit-cell parameters (Å)	$a = b = 132.19$, $c = 160.56$	$a = b = 132.29$, $c = 160.68$	$a = b = 132.33$, $c = 160.72$	$a = b = 132.35$, $c = 160.73$	$a = b = 132.34$, $c = 160.70$	$a = b = 132.34$, $c = 160.68$
Resolution (Å)	2.70					
Mosaicity (°)	0.069	0.068	0.074	0.081	0.089	0.095
Total reflections	109987 (17622)	109956 (17661)	109584 (17575)	108881 (17466)	108252 (17265)	107933 (17219)
Unique reflections	23163 (3644)	23126 (3656)	23060 (3649)	22940 (3631)	22830 (3597)	22769 (3586)
Multiplicity	4.75 (4.83)	4.75 (4.83)	4.75 (4.82)	4.75 (4.81)	4.74 (4.80)	4.74 (4.80)
Completeness (%)	98.9 (99.3)	98.9 (99.3)	98.9 (99.3)	98.9 (99.1)	98.9 (98.9)	98.9 (98.9)
$I/\sigma(I)$	14.81 (5.83)	14.18 (4.75)	13.15 (3.93)	13.05 (3.37)	12.59 (2.95)	11.61 (2.45)
R_{merge} (%)	8.8 (24.5)	9.0 (32.4)	9.9 (40.2)	10.5 (48.3)	11.2 (55.1)	12.6 (67.7)
R_{meas} (%)	9.8 (27.4)	10.1 (36.2)	11.1 (45.0)	11.7 (54.0)	12.6 (61.7)	14.1 (75.7)
Wilson B (Å^2)	33.65	36.84	38.76	40.99	42.65	44.51
Phasing						
$\Delta F/\sigma(\Delta F)$	—	1.36	1.66	1.92	2.12	5.03
Isomorphous R factor (%)	—	6.9	8.7	10.2	11.8	13.8
Scale factor	—	0.979	0.984	0.983	0.986	0.993
Resolution cutoff	—	3.0	3.0	3.0	3.0	3.0
CC_{all}	—	16.34	19.58	21.13	21.58	19.87
CC_{weak}	—	6.24	8.79	10.12	9.91	10.06
No. of selenium sites	—	11				
After 5 iterations using *SHELXE* (total sites)	—	57	47	67	60	54
After 5 iterations using *SHELXE* (negative sites)	—	19	15	27	22	21
No. of residues built using *SHELXE*	—	287	309	293	320	293
Phase error (°)/map CC (%)	—	38.46/82.6	35.51/85.3	35.35/85.6	34.73/85.8	35.20/85.3
Refinement						
Refinement resolution range (Å)	20–2.7					
R_{work} (%)	17.4					
R_{free} (%)	20.1					
Reflections, working	21564					
Reflections, free	1046					
Non-H atoms	2959					
Water molecules	115					
Average B factor (Å^2)	43.15					
R.m.s.d. bond lengths (Å)	0.018					
R.m.s.d. bond angles (°)	1.733					

and higher increases in the R_{merge} and Wilson B factor (Figs. 3*b* and 3*c*) for each data set collected after UV exposure.

3.3. Substructure determination and RIP phasing

In order to locate the specific sites of radiation damage, we treated our data as a case of SIR and used the program *SHELXD* (Schneider & Sheldrick, 2002) to find the sites of the largest differences between the 'before' and 'after-*i*' (i = 1–5) data sets.

3.3.1. FAE. Using the 'before' and first 'after' data set (after-1), 58 solutions of 16 sites were found from 100 *SHELXD* trials with a clear distinction between the correct and incorrect solutions. There was no sharp drop in the occupancy of the resulting sites as would be expected for Se-SAD or Se-MAD substructure solutions. The success rate clearly improved with increased exposure to UV, but started to degrade (Fig. 3*h*) after the third step of UV irradiation, which is also consistent with the highest CC_{all}/CC_{weak} for the third 'after' data set (after-3), although the differences in CC_{all}/CC_{weak} for the different 'after' data sets are minimal (Table 1). This implies that prolonged exposure of the protein crystals to UV introduces nonspecific structural changes that are indicated by an increase in non-isomorphism (Fig. 3*e*), which has implications for the success rate. The resulting substructure from *SHELXD* from each 'after' data set yielded an interpretable electron-density map after 20 cycles of *SHELXE*. An additional five rounds of iterative improvement of substructures in *SHELXE* gave sufficiently improved phase quality for autotracing. The chain tracing in *SHELXE* resulted in an almost complete polyalanine model for each 'after' data set. In each case, the map correlation between the final and the calculated phases before chain tracing was 77% or better (Fig. 3*i*).

3.3.2. H35. In this case, the overall $\Delta F/\sigma(\Delta F)$ also increases monotonically with the five successive exposures to UV light (Fig. 3*f*). However, it shows larger non-isomorphism after each irradiation (Fig. 3*e*). Nevertheless, the substructure was easily solved from each data set and the best CC_{weak} in each case is better than 20% (Fig. 3*g*), with minor differences

Table 4
Data-collection, processing and refinement statistics for the CHSYNTH crystal (control experiment, no UV radiation).

Values in parentheses are for the highest resolution shell.

Data	Before	After-1	After-2	After-3	After-4	After-5
Wavelength (Å)	1.0332					
Total frames per data set	100					
X-ray dose per data set (MGy)	2.3					
Space group	$P6_422$					
Unit-cell parameters (Å)	$a = b = 132.05$, $c = 160.21$	$a = b = 132.11$, $c = 160.27$	$a = b = 132.16$, $c = 160.31$	$a = b = 132.18$, $c = 160.33$	$a = b = 132.18$, $c = 160.32$	$a = b = 132.17$, $c = 160.29$
Resolution (Å)	2.70					
Mosaicity (°)	0.068	0.069	0.075	0.084	0.095	0.108
Total reflections	271142 (43455)	271958 (43823)	272176 (43752)	272237 (43604)	272237 (43299)	272181 (42978)
Unique reflections	23068 (3616)	23099 (3638)	23121 (3634)	23121 (3629)	23101 (3620)	23054 (3596)
Multiplicity	11.75 (12.01)	11.77 (12.05)	11.77 (12.04)	11.77 (12.02)	11.77 (11.96)	11.76 (11.95)
Completeness (%)	99.8 (99.1)	99.9 (99.7)	99.9 (99.8)	99.9 (99.6)	99.9 (98.4)	99.7 (98.7)
$I/\sigma(I)$	23.18 (9.93)	23.78 (9.23)	23.77 (8.08)	24.57 (6.91)	23.19 (5.79)	22.49 (4.93)
R_{merge} (%)	8.5 (23.8)	8.3 (26.2)	8.5 (30.9)	8.6 (37.2)	9.3 (44.7)	9.9 (52.9)
R_{meas} (%)	8.8 (24.8)	8.7 (27.3)	8.9 (32.3)	9.0 (38.9)	9.7 (46.6)	10.4 (55.3)
Wilson B (Å^2)	36.93	37.98	39.5	41.31	43.05	44.98
$\Delta F/\sigma(\Delta F)$	—	1.36	1.65	2.02	2.31	2.49
Isomorphous R factor (%)	—	3.4	4.2	5.2	6.1	6.9
Refinement						
Refinement resolution range (Å)	20–2.7					
R_{work} (%)	17.3					
R_{free} (%)	18.9					
Reflections, working	21916					
Reflections, free	1038					
Non-H atoms	2959					
Water molecules	115					
Average B factor (Å^2)	28.16					
R.m.s.d. bond lengths (Å)	0.023					
R.m.s.d. bond angles (°)	1.85					

among the data sets. The effect of the longer exposure is only reflected in the success rate of the substructure determination, with the third data set being the best (Fig. 3*h*). Phase calculation proceeded in a similar manner as described previously. Of the five 'after' data sets, 20 min of accumulative UV dose seems to be the best choice. Although the phase quality continues to improve marginally upon further UV exposure, this gain is offset by continuous nonspecific radiation-induced damage (Fig. 3*i*). Autotracing using the final phases after five rounds of iterative improvement of the substructure resulted in *SHELXE* building more than 75% of the complete model (Table 2) and *ARP*/*wARP* rebuilding the model with side-chain docking for more than 90% of the complete model.

3.3.3. CHSYNTH. The maximum resolution of the data is 2.7 Å and the resolution cutoff was set to 3.0 Å for substructure determination. There were only two successful trials out of 100 *SHELXD* trials for the first data set, with the best CC_{weak} being 6.24%. Despite the low CC_{weak}, a clear distinction between correct and incorrect solutions was found. The resulting substructure was improved after five rounds of iteration in *SHELXE*, resulting in a total of 57 sites, which included 19 negative sites. The peaks where electron density has been lost are positive and those where it has not been modelled are negative. In the final round, *SHELXE* traced 287 polyalanine residues of the possible 407. This partial model was used as an initial model in *ARP*/*wARP* and the resulting phases from *SHELXE* were used to restrain the refinement during model building. 335 of 407 residues were rebuilt and all of the residues were docked with the protein sequence. The success rate for substructure determination improved for the second to fifth 'after' data sets, with the highest rate being 11% for the second data set (Fig. 3*h*). However, in each case more than 74% of the complete model was traced using *SHELXE* and subsequent rebuilding with *ARP*/*wARP* generated more than 86% of the complete model. The resulting phase quality after five rounds of substructure iteration with *SHELXE* showed map correlations with the final model phase (Fig. 3*i*) of better than 82%.

3.4. Analysis of structural damage after 10 min of UV exposure

3.4.1. FAE. The FAE crystals contain two molecules in the asymmetric unit which are related by twofold noncrystallographic symmetry. The root-mean-square deviation (r.m.s.d.) between the two molecules is 0.12 Å over the 282 superimposed C^{α} atoms. The extent of damage to Mse is similar in each molecule (Supplementary Table S1*a*). Therefore, the environment of the Mse residues from a single molecule will be discussed here. Most of the Mse residues are located (as expected) in a hydrophobic environment. The maximum radiation damage is for Mse863, which shows the highest positive difference peak height (38σ) and a nearby negative peak (19σ) in an isomorphous difference Fourier ($F_{before} - F_{after\text{-}1}$, φ_{calc}) map. The residue is surrounded by His864, Phe903, Phe913, Phe950 and Thr959 (Supplementary

Fig. S1*a*). The second highest positive peak is located at the selenium position of residue Mse964. The minor difference in peak height of this residue between the two independent molecules is likely to arise from slight differences in the side-chain orientations. In each molecule the residues are also located in a hydrophobic environment, which includes a neighbouring Mse residue (Mse1031). The Mse residue Mse1024 shows relatively lower damage and in this case the side chains of aromatic residues are in close proximity, but the side chain of the Mse residue points towards residue Thr1014 (Supplementary Fig. S1*c*). The spatial neighbours of the Mse residue (Mse889) are Glu825 and Cys823. The side chain of residue Cys823 points towards Mse889 and the positive peak at this residue is slightly above 5σ (Supplementary Table S1*a*). Decarboxylation of Asp980 and Glu868 is observed and these residues point towards the N atoms of the indole rings of Trp982 and Trp1060, respectively (Supplementary Figs. S1*d* and S1*e*). Other Glu and Asp residues distant from the N atom of the indole ring of Trp were not affected by irradiation.

3.4.2. H35. The H35 protein crystals contain four molecules in the asymmetric unit. The r.m.s.d. between the various molecules over 98 superimposed C^{α} atoms varies between 1.5 and 2.5 Å. This range of r.m.s.d. values is also expected for the side-chain orientations and is reflected in the radiation-damage analysis. Two of the four protomers are more affected by UV damage (Supplementary Table S1*b*). The protein crystals contain two dimers. The dimers are related to each other by improper NCS and the crystal contact environment of the two dimers differs. The packing arrangement of the molecules in the unit cell and/or most likely the fluorescence anisotropy of the UV irradiation are responsible for the variation in the damage. Not all of the equivalent residues among the four protomers are equally affected by UV damage (Supplementary Figs. S2*a* and S2*b*). The protein contains neither Trp nor Tyr residues, but does contain two Phe residues (Phe23 and Phe42). These are in close proximity to Mse42, Mse12 and Mse76. Despite their close proximity to UV-absorbing aromatic residues, these Mse residues do not show a high level of radiation damage. Interestingly, the Mse most distant from the Phe residues shows a higher degree of radiation damage. Clearly, the UV radiation damage to Mse is not dependent upon the aromatic residues only, although it may be dependent on their relative orientations. Most of the Mse residues in the protein are located such that side chains of two Mse residues point towards each other. Such Mse residues are the most damaged (Supplementary Figs. S2*a* and S2*b*). The highest difference peak is located at Mse54*A* (Supplementary Table S1*b*) and this residue is 4.0 Å away from Mse54*B*, whereas Mse50*A* is 6.1 Å away from Mse50*B* and this pair of residues show lower damage than for the Mse54*A* and Mse54*B* pair (Supplementary Table S1*b*). The extent of damage to Mse residues varies according to the local environment in different molecules in the asymmetric unit. For example, the immediate spatial neighbours of residues Mse76 of molecules *A* and molecule *D* are Asn85 from molecule *C* and Phe23 from the symmetry-related molecule *A*, respectively, whereas in molecules *B* and *C* its spatial neighbours are its symmetry mates (Supplementary Figs. S2*a* and S2*b*). These residues show different degrees of radiation damage (Supplementary Table S1*b*). The residue Mse76*D* is not at all affected by UV radiation, perhaps because the residue lacks a spatial Mse neighbour. Similarly, some of the other Mse residues (*i.e.* Mse12*D*, Mse12*C*, Mse34*B*, Mse34*D* and Mse46*D*) lacking neighbouring Mse residues show no or very little radiation damage (Supplementary Table S1*b*). There are a number of Glu and Asp residues in the protein, but they were not affected by UV, perhaps because the protein lacks Trp residues.

3.4.3. CHSYNTH. The CHSYNTH crystals contain one molecule in the asymmetric unit, with a solvent content of 73%. The radiation damage to the protein crystal is mainly centred on Mse residues. Six Mse residues out of 11, three Asp residues out of a total of 24 and five Glu residues out of a total of 22 are damaged after 10 min exposure to UV radiation (Supplementary Table S1*c*). The radiation-affected Glu and Asp residues are located near the UV-absorbing residues. The maximum 'damage' is to the Se atom of Mse89, as can be inferred from the highest positive (36σ) and negative (20σ) peak at a distance of 2.71 Å from the Se atom in the isomorphous difference Fourier ($F_{\rm before} - F_{\rm after\text{-}1}, \varphi_{\rm calc}$) map. The Mse residue is surrounded by charged residues (Glu134, Glu9, Arg130 and His11) as well as Trp85. At this location, Glu134 and Glu9 are both affected by UV radiation damage (Supplementary Fig. S3*a*). The second highest difference peak in the structure is located at Mse281, with its spatial neighbours being Pro111, Mse121, Tyr118 and the Leu329 residue from a symmetry-related molecule (Supplementary Fig. S3*b*). Mse253, which is next to the main chain of residues 258–261, is also damaged (Supplementary Fig. S3*c*). The residue Mse357 is located in a similar environment to Mse89. Interestingly, an acetate ion bound to the structure located near Mse357 is damaged, as are two other negatively charged residues, Glu140 and Asp241 (Supplementary Fig. S3*d*). Asp185, which is located in the vicinity of Tyr197, is affected, whereas Glu184, which is distant from this Tyr, remains unaffected (Supplementary Fig. S3*e*).

4. Discussion

4.1. Scale factor, substructure and RIP phasing

The scale factor has been shown to be important in solving substructures from disulfide-containing protein crystal diffraction data (Nanao *et al.*, 2005). However, the scale factor is not crucial (de Sanctis & Panjikar, submitted work) for substructure determination of Mse-containing proteins. The substructure can be solved by a standard SIR method as implemented in *SHELXC* by introducing a scale factor as calculated from the empirical formula described here, but this does not substantially affect the substructure determination. This simplifies the methodology as it does not require the scanning of large number of scale factors in substructure determination as was shown to be necessary for disulfide-

bridge-containing protein crystals (Nanao & Ravelli, 2006). The selenium RIP substructure can be solved at a resolution of as low as 4.5 Å and an interpretable map can be achieved (data not shown).

These three examples demonstrate that the RIP phasing of Mse-containing proteins is as easy as the popular Se-MAD or Se-SAD phasing methods. This method of phasing using Mse proteins and UV-RIP provides an alternative way of obtaining phase information, which can be performed on a fixed-wavelength synchrotron beamline or even on a home source. Attempts to solve these structures using the differences between 'after-1' data sets and 'after-5' data sets did not succeed. We believe that the greatest effect occurs during the first UV burn and that after that the changes are much smaller and hence the signal is much smaller.

In the three examples described here, even with no refinement of heavy-atom positions and no noncrystallographic symmetric averaging, the phase quality was still sufficient for auto-tracing. With poorer quality data, or when phasing at lower resolutions, these two steps might be essential to improve the phase quality.

4.2. Effects of longer UV exposure to selenium-labelled protein crystals for substructure determination

Prolonged exposure to the UV laser and X-ray dose of Mse-substituted crystals results in greater non-isomorphism (Fig. 3*e*). This is reflected in an increase in the unit-cell volume and a matching increase in the isomorphous *R* factor. The effects are (i) breakage of the Se—C^{γ} bond, resulting in structured or specific damage, and (*b*) an increase in the global change in structure upon prolonged UV exposure (nonstructural or nonspecific damage). The extent of specific damage depends upon the local environment of the selenium (see below). The difference between the two effects is evident during substructure determination in *SHELXD*. In the case of FAE the success rate of substructure determination clearly initially improves with increased UV dose, but starts to degrade again after the third UV irradiation (Fig. 3*h*). In the 'after-1' data set, specific damage has been observed for most of the Se atoms and there are a smaller number of changes in the structure (Supplementary Table S1*a*) which result in global changes in the packing. When the global changes start to dominate, the success rate of substructure determination decreases. Based on the success rate, 30 min of accumulative UV dose is optimal for the protein crystal, which is consistent with the highest CC_{all}/CC_{weak} for the third 'after' data set (after-3). Clearly, prolonged UV exposure of the protein crystal increases the non-isomorphism, which in turn has implications for the success of substructure determination. Similar observations were made for H35 and CHSYNTH (Fig. 3*h*), leading us to believe that the 30 min time period is broadly applicable.

4.3. Environment of Mse and extent of UV radiation damage

It is known that methionine residues in proteins are usually in a hydrophobic environment mostly surrounded by Trp, Phe, Tyr, Leu, Ile, Val and Met, although in some cases they can be partially in the neighbourhood of charged residues. The side chain of Mse in proteins generally adopts a similar orientation to that observed for methionine (Zhang & Vogel, 1994). Not only do the aromatic residues Trp, Phe and Tyr absorb UV at 266 nm, but Mse also absorbs significantly at this wavelength (Fig. 1). Met does not show absorption at a wavelength of 266 nm (Fig. 1). Mse seems to be damaged by UV more strongly than any aromatic residues, possibly because the Se atom is a large electron-rich atom that is more easily polarizable than those of the first-row elements. Analysis of UV data shows that the Mse residues are the most affected residues in the protein structures (Supplementary Table S1). The various Mse residues in the structure show different degrees of UV radiation damage. Mse residues that are inaccessible to the solvent region are more damaged than those that are accessible to solvent. For example, the solvent-accessible areas of Mse863, Mse964, Mse975, Mse1023 and Mse1031 in FAE are between 0 and 1.5 $Å^2$ and these are the most damaged residues (peak heights in the difference map of between 11σ and 38σ) after 10 min of UV radiation, whereas Mse964, Mse955 and Mse889 show lesser UV damage (peak heights of between 5.5σ and 6.8σ) and the solvent-accessible areas of these residues are between 20 and 90 $Å^2$. However, some Mse residues that are accessible to the solvent region but are packed against symmetry-related residues show relatively high degrees of UV radiation damage. For example, the solvent-accessible areas of Mse253 and Mse281 are 90 and 110 $Å^2$, respectively, and the peak heights of these residues are 13σ and 21σ, respectively. These two residues are packed against residues from symmetry-related molecules. In such cases, even though the Mse is located on the surface of the protein it is not solvent-accessible because of the crystal-packing environment. This suggests that the broad conclusions of this work may be better explained using an analysis of packing density. Such a full analysis of the packing density of Mse residues *versus* corresponding UV radiation damage requires further work.

Besides Mse, other residues that are UV damaged in the proteins are Glu or Asp residues. This radiation damage results in decarboxylation; occasionally, the carboxyl group and C^{β} in Asp and C^{γ} in Glu are no longer seen in the electron density. Decarboxylation of Glu residues is well known in X-ray radiation damage (Ravelli & McSweeney, 2000). However, upon UV radiation the Glu or Asp residue located near the indole ring of a Trp residue is preferentially decarboxylated (Supplementary Figs. S1*e* and S1*f*). In the CHSYNTH structure, Glu or Asp residues located near Mse or Tyr residues lose their carboxyl groups (Supplementary Figs. S3*a*, S3*d* and S3*f*), whereas Glu or Asp residues away from the UV-absorbing residues are not affected (Supplementary Fig. S3*f*). These changes in protein structure are located around the UV-absorbing residues. Decarboxylation of Glu or Asp was observed in the FAE and CHSYNTH structures but not in that of H35, hypothetically because it lacks Trp and Tyr.

4.4. Minimum/optimal exposure for selenomethionine protein crystals using a 266 nm UV laser

The isomorphous signal improves with longer UV laser exposure, although phase improvement was only marginal after a certain UV dose because the radiation damage has a structural component and a component involving intermolecular changes. The first effect results in increasing difference peak heights with increasing dose and the second results in an increased r.m.s. of the maps. Presumably, in the extreme the signal from the structured part of the damage is 'lost' in the noise arising from nonspecific damage (Supplementary Table S1).

With the data presented here, structure determination was successful even with five successive 10 min exposures to UV. Based on the above analysis, 20–30 min exposure to this UV laser seems to be the best choice (Figs. 3*h* and 3*i*). Additional exposure to UV results in only marginal improvement of the phase, as the radiation damage continues to increase. Consequently, the success rate of substructure solution does not improve.

5. Conclusions and future perspectives

In this paper, we have shown that UV-based radiation damage of Mse facilitates structure determination. UV brings more specific changes to the protein than X-rays. Here, we show the requirements, limits and advantages of using UV-RIP on Mse-containing crystals. In the case where the X-ray-induced damage of selenomethionine proteins did not result in successful phasing despite a high X-ray dose, brief exposure to UV did. The method discussed in this paper is a further extension of the UV-RIP method (Nanao & Ravelli, 2006) applied to disulfide-bridge-containing proteins.

The method has considerable potential and combination of Se-UV-RIP with the Se-SAD or Se-MAD methods is also likely to be attractive. Selenium-specific UV damage could serve as an extra source of phase information, being an additional or even an alternative way of experimental phasing in macromolecular crystallography. Another advantage of the method is that the experiment can be carried out on a home source using selenium-labelled protein crystals, avoiding a wait for access to tunable synchrotron beamlines for SAD or MAD phasing. Further developments of the UV-RIP method are expected to benefit traditional phasing techniques using derivatives with covalently bound heavy atoms.

A number of important questions remain open at the moment. The underlying chemistry and physics of selenium-specific UV damage is poorly understood. In the current study a 266 nm microchip pulse laser was used, but questions as to whether a continuous laser source at this wavelength or at different wavelengths away from 280 nm may be more effective as the laser source used in this study remain to be answered. The UV spectrum of Mse suggests that this method might be best with UV radiation of a wavelength at which Mse absorption is a maximum, *i.e.* around 246 nm.

DdS is grateful to Matias Guijarro and Didier Nurizzo for software and hardware support on ID23-1. Hans Bartunik is kindly acknowledged for providing crystals of the Se derivative of chorismate synthase from *M. tuberculosis* within the framework of BIOXHIT.

References

Abbani, M. A., Papagiannis, C. V., Sam, M. D., Cascio, D., Johnson, R. C. & Clubb, R. T. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 2109–2114.
Banumathi, S., Zwart, P. H., Ramagopal, U. A., Dauter, M. & Dauter, Z. (2004). *Acta Cryst.* D**60**, 1085–1093.
Bourenkov, G. P. & Popov, A. N. (2010). *Acta Cryst.* D**66**, 409–419.
Bragg, W. L. & Perutz, M. F. (1954). *Proc. R. Soc. Lond. A*, **225**, 315–329.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Dose, K. (1968). *Photochem. Photobiol.* **8**, 331–335.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Evans, G., Polentarutti, M., Djinovic Carugo, K. & Bricogne, G. (2003). *Acta Cryst.* D**59**, 1429–1434.
French, S. & Wilson, K. (1978). *Acta Cryst.* A**34**, 517–525.
Fütterer, K., Ravelli, R. B. G., White, S. A., Nicoll, A. J. & Allemann, R. K. (2008). *Acta Cryst.* D**64**, 264–272.
Gabadinho, J. *et al.* (2010). *J. Synchrotron Rad.* **17**, 700–707.
Garman, E. F. (2010). *Acta Cryst.* D**66**, 339–351.
Giordano, A., Russo, C., Raja, C. A., Kuznetsova, I. M., Stepanenko, O. V. & Turoverov, K. K. (2004). *J. Proteome Res.* **3**, 613–620.
Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
Holton, J. M. (2009). *J. Synchrotron Rad.* **16**, 133–142.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Kehoe, J. J., Remondetto, G. E., Subirade, M., Morris, E. R. & Brodkorb, A. (2008). *J. Agric. Food Chem.* **56**, 4720–4725.
Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.
Nanao, M. H. & Ravelli, R. B. G. (2006). *Structure*, **14**, 791–800.
Nanao, M. H., Sheldrick, G. M. & Ravelli, R. B. G. (2005). *Acta Cryst.* D**61**, 1227–1237.
Nurizzo, D., Mairs, T., Guijarro, M., Rey, V., Meyer, J., Fajardo, P., Chavanne, J., Biasci, J.-C., McSweeney, S. & Mitchell, E. (2006). *J. Synchrotron Rad.* **13**, 227–238.
Ogata, C. M. (1998). *Nature Struct. Biol.* **5**, *Suppl.*, 638–640.
Olmo, R., Dolores, B., Socorro, J. M., Martin, J. A. & Teijon, J. M. (2001). *J. Enzyme Inhib.* **16**, 65–80.
Paithankar, K. S. & Garman, E. F. (2010). *Acta Cryst.* D**66**, 381–388.
Permyakov, E. A. (1993). *Luminescent Spectroscopy of Proteins*. Boca Raton: CRC Press.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Prates, J. A. M., Tarbouriech, N., Charnock, S. J., Fontes, C. M. G. A., Ferreira, L. M. A. & Davies, G. J. (2001). *Structure*, **9**, 1183–1190.
Ramagopal, U. A., Dauter, Z., Thirumuruhan, R., Fedorov, E. & Almo, S. C. (2005). *Acta Cryst.* D**61**, 1289–1298.
Ravelli, R. B. G., Leiros, H.-K. S., Pan, B., Caffrey, M. & McSweeney, S. (2003). *Structure*, **11**, 217–224.
Ravelli, R. B. G. & McSweeney, S. (2000). *Structure*, **8**, 315–328.
Rudiño-Piñera, E., Ravelli, R. B. G., Sheldrick, G. M., Nanao, M. H., Korostelev, V. V., Werner, J. M., Schwarz-Linek, U., Potts, J. R. & Garman, E. F. (2007). *J. Mol. Biol.* **368**, 833–844.
Schiltz, M., Dumas, P., Ennifar, E., Flensburg, C., Paciorek, W., Vonrhein, C. & Bricogne, G. (2004). *Acta Cryst.* D**60**, 1024-1031.
Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.
Schönfeld, D. L., Ravelli, R. B. G., Mueller, U. & Skerra, A. (2008). *J. Mol. Biol.* **384**, 393–405.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.

Sheldrick, G. M. (2010). *Acta Cryst.* D**66**, 479–485.
Vernede, X., Lavault, B., Ohana, J., Nurizzo, D., Joly, J., Jacquamet, L., Felisaz, F., Cipriani, F. & Bourgeois, D. (2006). *Acta Cryst.* D**62**, 253–261.
Weiss, M. S., Mander, G., Hedderich, R., Diederichs, K., Ermler, U. & Warkentin, E. (2004). *Acta Cryst.* D**60**, 686–695.
Weljie, A. M. & Vogel, H. J. (2000). *Protein Eng.* **13**, 59–60.
Zhang, M. & Vogel, H. J. (1994). *Biochemistry*, **33**, 1163–1171.
Zwart, P. H., Banumathi, S., Dauter, M. & Dauter, Z. (2004). *Acta Cryst.* D**60**, 1958–1963.