Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structural effects of radiation damage and its potential for phasing

Sankaran Banumathi,[a] Petrus H. Zwart,[b] Udupi A. Ramagopal,[a]‡ Miroslawa Dauter[b] and Zbigniew Dauter[a]*

[a]Synchotron Radiation Research Section, MCL, National Cancer Institute, Brookhaven National Laboratory, Upton, NY 11973, USA, and [b]SAIC Frederick Inc, Basic Research Program, Brookhaven National Laboratory, Upton, NY 11973, USA

‡ Current address: Department of Biochemistry, Albert Einstein College of Medicine, Yeshiva University, 1300 Morris Park Avenue, Bronx, NY 10461, USA.

Correspondence e-mail: dauter@bnl.gov

Received 15 January 2004
Accepted 31 March 2004

A detailed analysis of radiation-damage-induced structural and intensity changes is presented on the model protein thaumatin. Changes in reflection intensities induced by irradiation display a parabolic character. The most pronounced structural changes observed were disulfide-bond breakage and associated main-chain and side-chain movements as well as decarboxylation of aspartate and glutamate residues. The structural changes induced on the sulfur atoms were successfully used to obtain high-quality phase estimates *via* an RIP procedure. Results obtained with *ACORN* suggest that the contribution originating from the partial structure may play an important role in phasing even at less than atomic resolution.

1. Introduction

Radiation damage has been a curse of macromolecular crystallography from its early days. Until the 1990s, all diffraction data were collected from macromolecular crystals mounted in glass or quartz capillaries and kept at ambient temperature (in the range 277–293 K). Initially, collecting data from protein crystals using precession cameras or multi-circle diffractometers required a very long time. The introduction of the screenless rotation method (Arndt & Wonacott, 1977) considerably sped up this process by recording more reflections simultaneously. Still, only the most robust crystals were able to deliver all necessary data from one specimen and very often a complete data set had to be combined from data measured on several crystals.

The advent of cryocooling techniques (Hope, 1988; Teng, 1990) to a large extent alleviated the problem of crystal deterioration during the progress of data collection in home laboratories or at second-generation synchrotron sources. Substantial radiation damage to crystals cooled to about 100 K (Gonzalez & Nave, 1994) remained a serious problem only in cases requiring very long exposures, such as weakly diffracting crystals of large protein complexes or atomic resolution projects.

This situation changed with the construction of the third-generation synchrotron sources at the ESRF, APS and SPring8. It was quickly realised that the X-ray beam from the new undulator beamlines induces substantial radiation damage to cryocooled crystals after a few minutes of irradiation. Since exposure times at these beamlines are very short (of the order of seconds), the collection of complete data is possible in a short time, but users of those beamlines have to be wary of the danger of overexposing their crystals.

The acute problem of deterioration of cryocooled crystals of macromolecules at undulator beamlines revived interest and prompted several investigations into the effects of

radiation damage. A number of publications devoted to this problem have appeared recently (*e.g.* Burmeister, 2000; Ravelli & McSweeney, 2000; Weik *et al.*, 2000; Garman & Nave, 2002) and dedicated radiation-damage workshops were organized in 1999 at the ESRF, in 2001 at the APS (with proceedings published in the November 2002 issue of the *Journal of Synchrotron Radiation*) and in 2003 at the ESRF.

The X-ray radiation damage of crystals can be caused by the breakage of covalent bonds as an immediate consequence of the absorption of an X-ray quantum (a primary effect) or by the destructive effect of the propagation of radicals throughout the crystal (a secondary effect). The secondary effects are diminished by cryocooling, but the primary effects do not depend on the temperature of the sample (Garman, 2003). Sliz *et al.* (2003) recently reported that the radiation damage to cryocooled crystals depends on the total dose but not on the dose rate. However, results of other investigations (Leiros *et al.*, 2001) suggest that both the total dose and the dose rate play a role in the amount of radiation damage inflicted on a protein crystal.

Apart from the obvious deterioration in the diffraction power of protein crystals, irradiation by X-rays causes a systematic increase of the crystal unit-cell parameters (Yonath *et al.*, 1998; Weik *et al.*, 2001; Müller *et al.*, 2002; Ravelli *et al.*, 2002) as well as certain changes in the crystal structure. The most characteristic specific structural consequences of irradiation reported were the rupture of disulfide bonds and decarboxylation of glutamic and aspartic acid residues (Burmeister, 2000; Ravelli & McSweeney, 2000; Weik *et al.*, 2000; O'Neill *et al.*, 2002).

Whereas previous structural radiation-damage studies have been carried out at third-generation synchotron sources (Burmeister, 2000; Weik *et al.*, 2000; Ravelli & McSweeney, 2000), the effects of radiation damage on protein crystals can be also observed after prolonged exposures at second-generation bending-magnet synchrotron beamlines. We have analyzed the effects of radiation damage on crystals of thaumatin, a medium-sized protein containing 207 amino acids including eight disulfide bridges. A number of high-resolution data sets were collected that allowed a detailed investigation of the structural changes induced by irradiation. Furthermore, we empirically modelled intensity changes by means of a polynomial correction. The order of the polynomial is determined *via* cross-validation techniques.

Some papers concerning radiation-damage effects include an estimate of the amount of radiation absorbed by the sample (O'Neill *et al.*, 2002; Sliz *et al.*, 2003). In this study, the total amount of incident radiation is assumed to be proportional to the sum of scale factors of sequentially measured data sets up to a given set. This sum of scale factors is denoted as the cumulative scale. The proportionality of the total amount of incident radiation to the cumulative scale results from the fact that the scale of an individual data set is directly proportional to the average intensity of all reflections in this data set. An estimate of the total dose was computed assuming a flux of 1.3×10^{10} photons s^{-1} (BNL Protein Crystallography Research Resource; http://www.px.nsls.bnl.gov) with the program *RADDOSE* (Murray *et al.*, 2004). The resulting estimate is 6.8×10^5 Gy for a single data set, which corresponds to about 3.4% of the Henderson limit (Henderson, 1990). The total dose after collecting all 20 data sets constitutes about 70% of the Henderson limits. These dose estimates should be seen as an upper limit, not taking into account any decay of the ring current and beam intensity during data collection in the time mode.

2. Materials and methods

2.1. Crystallization and data-collection strategy

The 22 kDa protein thaumatin (commercially available from Sigma) was crystallized as described previously (Ko *et al.*, 1994) by the hanging-drop method using 30 mg ml^{-1} protein solution, 0.8 *M* sodium/potassium tartrate and 0.1 *M* sodium *N*-(2-acetamidoimino)diacetic acid (ADA) buffer pH 6.5. The $P4_12_12$ crystals have a solvent content of 56%. A crystal of about $0.2 \times 0.2 \times 0.25$ mm was flash-cooled to 100 K using its mother liquor containing in addition 25% glycerol as a cryoprotectant. All X-ray data were collected to a resolution of 1.45 Å with a wavelength of 0.979 Å (12 658 eV) from the same crystal at the X9B beamline, NSLS, Brookhaven National Laboratory using an ADSC Quantum 4 CCD detector. A series of 20 consecutive data sets were obtained and were processed with the *HKL*2000 suite (Otwinowski & Minor, 1997). The same rotation range of 90° was used for each data set, with a constant exposure time of 30 s per image of 1.0° width. However, the effective exposure varied with changes in the synchrotron ring current. All data sets have a completeness of above 99%, a redundancy of about 7 and an $R_{\rm merge}$ in the range 4.3–6.3%, as shown in Table 1.

The protein model taken from the PDB (PDB code 1thw) was refined against the first data set using *REFMAC5* (Murshudov *et al.*, 1999) to obtain model phases for statistical comparisons. This model was then used for further analysis of the occupancies of selected atoms. The occupancies of all S atoms, carboxylate groups of aspartate and glutamate residues, hydroxyl groups of tyrosines and some of the water O atoms were refined against data sets 1, 5, 10, 15 and 20 using *SHELXL* (Sheldrick & Schneider, 1997) along with the positional coordinates and isotropic temperature factors of all atoms in the protein. Most programs used were from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994) except where explicitly referenced.

2.2. Substructure solution and phasing procedure

The program *XPREP* (Sheldrick, 2001) was used to calculate the isomorphous intensity differences between the following pairs of data sets: 1 and 5, 1 and 10, and 1 and 15. *SHELXD* (Schneider & Sheldrick, 2002) was used to determine the isomorphous substructure using a resolution cutoff of 2.2 Å.

The determined substructure sites were treated as sulfurs in the heavy-atom refinement and phasing was carried out by

Table 1
R_{merge} values, correlation coefficients, B factors and $I/\sigma(I)$ values.

R_{merge} values (%) from merging pairs of data sets are within the upper triangle. On the diagonal are the R_{merge} values of the individual data sets. The lower triangle contains values of correlation coefficients between normalized amplitudes (E values) in pairs of corresponding data sets. Also given are B factors estimated from the Wilson plot and $I/\sigma(I)$ values for all data and in the highest resolution shell (HR; 1.50–1.45 Å) for all data sets.

Data set	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	**4.3**	1.2	1.9	2.7	3.4	3.9	4.4	4.8	5.2	5.4	5.8	6.2	6.6	6.9	7.3	7.5	7.9	8.4	8.5	8.8
2	0.992	**4.3**	1.2	1.9	2.6	3.0	3.5	4.0	4.4	4.6	5.0	5.4	5.8	6.1	6.5	6.7	7.1	7.6	7.8	8.1
3	0.990	0.992	**4.2**	1.2	1.8	2.3	2.7	3.2	3.6	3.8	4.2	4.6	5.0	5.4	5.8	6.0	6.3	6.9	7.1	7.3
4	0.987	0.990	0.992	**4.0**	1.0	1.5	2.0	2.4	2.9	3.1	3.5	3.9	4.3	4.7	5.1	5.3	5.6	6.2	6.4	6.7
5	0.982	0.986	0.989	0.990	**4.2**	0.9	1.3	1.7	2.1	2.4	2.8	3.2	3.5	4.0	4.3	4.6	4.9	5.5	5.7	6.0
6	0.979	0.983	0.987	0.990	0.989	**4.4**	0.9	1.3	1.8	1.9	2.3	2.7	3.1	3.5	3.9	4.1	4.5	5.1	5.3	5.6
7	0.975	0.980	0.984	0.987	0.988	0.988	**4.5**	0.8	1.2	1.4	1.8	2.2	2.6	3.0	3.4	3.6	4.0	4.6	4.7	5.0
8	0.970	0.976	0.981	0.984	0.985	0.986	0.987	**4.7**	0.9	1.1	1.5	1.8	2.2	2.6	2.9	3.2	3.5	4.2	4.3	4.6
9	0.967	0.974	0.978	0.982	0.984	0.985	0.986	0.985	**4.9**	1.0	1.1	1.5	1.8	2.1	2.5	2.8	3.1	3.7	3.8	4.1
10	0.963	0.969	0.974	0.978	0.980	0.981	0.983	0.982	0.982	**5.3**	1.0	1.3	1.7	2.0	2.4	2.6	3.0	3.6	3.7	4.0
11	0.963	0.971	0.976	0.981	0.983	0.985	0.986	0.986	0.986	0.983	**4.5**	0.9	1.1	1.5	1.9	2.1	2.4	3.1	3.2	3.5
12	0.959	0.967	0.973	0.978	0.980	0.983	0.984	0.984	0.985	0.982	0.987	**4.8**	0.9	1.2	1.6	1.8	2.1	2.8	2.9	3.2
13	0.954	0.962	0.968	0.974	0.977	0.979	0.981	0.981	0.982	0.981	0.985	0.985	**5.1**	0.9	1.2	1.4	1.7	2.4	2.5	2.8
14	0.951	0.960	0.966	0.972	0.975	0.979	0.981	0.981	0.982	0.980	0.985	0.985	0.984	**4.9**	0.9	1.1	1.3	1.9	2.1	2.4
15	0.946	0.954	0.961	0.967	0.970	0.974	0.976	0.978	0.978	0.977	0.981	0.981	0.981	0.981	**5.1**	0.9	1.0	1.6	1.7	2.0
16	0.945	0.954	0.961	0.967	0.971	0.974	0.976	0.977	0.979	0.978	0.982	0.983	0.983	0.984	0.981	**5.2**	0.9	1.5	1.6	1.9
17	0.943	0.952	0.959	0.966	0.970	0.974	0.976	0.977	0.978	0.977	0.982	0.983	0.982	0.983	0.981	0.984	**4.6**	1.2	1.2	1.5
18	0.934	0.943	0.950	0.957	0.962	0.966	0.969	0.969	0.971	0.971	0.976	0.977	0.977	0.978	0.976	0.979	0.979	**4.6**	1.1	1.2
19	0.930	0.939	0.946	0.953	0.959	0.962	0.966	0.967	0.970	0.969	0.974	0.974	0.975	0.976	0.976	0.977	0.978	0.974	**5.6**	0.9
20	0.925	0.934	0.942	0.949	0.954	0.958	0.962	0.964	0.966	0.966	0.970	0.971	0.972	0.973	0.973	0.974	0.974	0.971	0.973	**5.8**
B_{Wilson} (Å^2)	13.2	13.2	13.5	13.2	13.8	13.5	14.0	14.2	14.4	14.2	14.5	14.2	14.6	14.9	15.3	14.9	15.1	15.6	15.8	16.1
$I/\sigma(I)$	40.5	39.7	42.2	43.8	43.6	41.7	40.9	39.7	38.7	35.9	41.3	39.2	36.7	38.3	36.9	36.6	43.2	40.2	33.6	33.2
$I/\sigma(I)$ HR	7.4	6.7	6.7	7.0	5.8	5.3	4.8	4.4	4.1	3.3	4.1	4.0	3.4	3.5	3.2	3.3	3.8	3.8	2.5	2.3

SHARP 2.0 (de La Fortelle & Bricogne, 1997) using an SIR-like protocol. The data set with the lowest cumulative exposure (set 1) was treated as a derivative with respect to the other data set in each pair, which was treated as the native. The effect of the anomalous scattering was neglected in the phasing procedure at this stage, owing to the very low value of f'' for sulfur (0.25 e) at the wavelength used. The resulting experimental phases were improved by solvent flattening with *SOLOMON* and *DM* according to the standard *SHARP* protocol. Results from the automated model-building suite *ARP/wARP* (Perrakis *et al.*, 1999) served as an indicator of the interpretability of the electron-density maps.

2.3. Cross-validated zero-dose extrapolation

The changes in intensity arising from radiation damage can be modelled by a polynomial (Abrahams & Marsh, 1987) or linear dependence (Blake & Phillips, 1962; Diederichs *et al.*, 2003) of the intensity on the total exposure time. The disadvantage of these dose-dependent scaling models with a fixed number of parameters is that the intensity changes can be overfitted depending on the choice of the order of the polynomial used. A standard technique to detect overfitting is cross-validation, in which a small subset of observations is set aside while the remaining observations are used for parameter-estimation purposes using a number of different models. On the basis of the agreement between the observed and calculated 'free' data points, a decision can be made on which model is preferred. A well known crystallographic application of cross-validation is the refinement R_{free} (Brünger, 1992) used as an unbiased quality indicator of the fit of the model to the observed X-ray data. An application related to the method proposed here is the normalization procedure presented by Cowtan (2002). To determine the order of the polynomial best fitting the effects of radiation damage on the intensities, a cross-validation set of three out of 20 observed measurements for each reflection was generated at random for all reflections. For each reflection, the coefficients of the polynomial were obtained *via* standard least-squares methods (Weisstein, 1999) from the working set. The choice of the order of the polynomial was carried out in two different ways.

Firstly, the order was fixed for all reflections and the residual $\sum_h \sum_{free}(|I_{obs} - I_{inter}|)/\sum_h \sum_{free}(|I_{obs}|)$ based on the free observations only was used to gauge the order of the polynomial leading to the best agreement with the observations. Secondly, the optimal order of the polynomial was determined for each reflection individually. In this case, a mean residual $\langle\sum_{free}(|I_{obs} - I_{inter}|)/\sum_{free}(|I_{obs}|)\rangle$ obtained by averaging the results from 50 fits for each reflection using different randomly chosen cross-validation sets of three out of 20 measurements for each reflection was calculated. The average of 50 free residuals was used because of the relative scarcity of the test set and its sensitivity to statistical fluctuations.

The variance of the overall residual in the fixed-order interpolation was estimated using resampling techniques (Dudewicz & Mishra, 1988), by repeating the fitting procedure 50 times with different free-reflection sets. These variances can be used to assess the significance of the difference between residuals corresponding to different orders of polynomials.

3. Results and discussion

3.1. Diffraction data

20 consecutive data sets were collected from one crystal of thaumatin. The same crystal orientation, rotation range, oscillation width and exposure time were used throughout in order to minimize the effect of the systematic measurement errors. The *c* axis of the crystal was oriented approximately 10° away from the spindle axis, so that each individual data set spanned twice the minimal rotation range required for a complete data set in the 422 crystal class, leading to an approximately sevenfold redundancy of the intensity measurements.

The unit-cell parameters obtained from the post-refinement procedure for each data set are shown in Fig. 1. Each individual data set was processed with the same wavelength and crystal-to-detector distance and in the same crystal orientation, so that the differences between the estimated unit-cell parameters are significantly more accurate than their absolute values. The *a* and *b* parameters increased by 0.20% and *c* increased by 0.18% between data sets 1 and 20; the unit-cell volume therefore expanded by 0.59%. The expansion of the unit cell is proposed to originate from the build-up of electrostatic potential in the crystal (Ravelli *et al.*, 2002), although the build-up of internal pressure owing to decarboxylation might also play a role.

The 'crossword' Table 1 shows in the upper triangle the values of $R_{\rm merge}$ obtained from merging together pairs of individual data sets. In the lower triangle, correlations between pairs of data sets based on normalized structure-factor amplitudes (*E* values) are given. On the diagonal of the table the $R_{\rm merge}$ values from the scaling and merging of individual sets are given. It is evident that the radiation damage caused a gradual change in the reflection intensities. Pairs of consecutive data sets differ by about 1% or less in $R_{\rm merge}$ values and have a correlation coefficient of about 98%, whereas sets 1 and 20 differ by about 9% in terms of $R_{\rm merge}$. The correlation coefficient based on normalized structure-factor amplitudes is in the latter case 0.93. This correlation coefficient might be a more objective overall measure of the difference between data sets because of its independence of the scaling procedures and the more uniform weight of individual contributors over the whole resolution range.

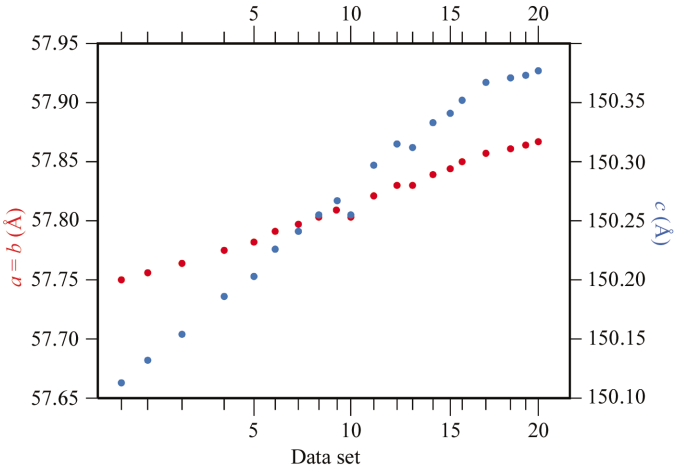

Figure 1
Change in unit-cell parameters over the course of collecting 20 data sets. The positions of the individual data-set points on the horizontal axis are proportional to the cumulative scale (see text).

Table 2
Change of intensities of selected reflections caused by radiation damage.

Reflection data set	1 0 37	1 1 25	2 2 13	3 0 62	8 1 22	8 1 31	8 2 16
1	360	1153	939	2198	1585	1202	548
2	453	1025	936	2145	1581	1113	587
3	561	923	966	2053	1635	1003	695
4	677	796	960	1975	1560	867	717
5	789	731	1077	1888	1605	797	842
6	871	694	1059	1835	1618	747	914
7	994	644	1119	1845	1653	723	980
8	1087	618	1188	1786	1708	697	1007
9	1157	583	1256	1709	1755	660	1083
10	1155	568	1242	1701	1708	635	1042
11	1313	519	1349	1643	1768	585	1156
12	1434	477	1361	1587	1789	568	1188
13	1549	461	1467	1551	1865	527	1174
14	1672	429	1560	1477	1871	501	1297
15	1815	424	1605	1467	1951	425	1284
16	1873	398	1632	1423	1931	478	1329
17	2050	377	1738	1417	2035	480	1395
18	2186	330	1833	1333	2028	441	1447
19	2285	371	1927	1327	2054	414	1476
20	2366	311	2011	1288	2119	436	1537

Table 3
Working ($R_{\rm work}$) and free ($R_{\rm free}$) residuals between the observed intensities and those estimated from the fitted polynomials of different order.

The estimates of the standard deviations (in parentheses) were obtained *via* resampling methods. The fourth column (Δ/σ) shows the difference in the free residual for the given entry and the quadratic function divided by the estimated standard deviation of the difference, indicating the significance of the observed differences between residuals.

Correction	$R_{\rm work}$ (σ)	$R_{\rm free}$ (σ)	Δ/σ
None	0.06371 (3.7×10^{-5})	0.06742 (1.8×10^{-4})	218
Linear	0.02635 (2.4×10^{-5})	0.03010 (7.8×10^{-5})	46.3
Quadratic	0.02113 (9.3×10^{-6})	0.02585 (4.8×10^{-5})	0
Cubic	0.01989 (1.2×10^{-5})	0.02669 (6.4×10^{-5})	10.6

The diffraction power of the crystal diminished considerably between data sets 1 and 20, as evidenced in Table 1 by the change in the $I/\sigma(I)$ ratio from 40.5 to 33.2 for all data and from 7.4 to 2.3 in the highest resolution shell (1.50–1.45 Å) and the increase of $R_{\rm merge}$ from 4.3 to 5.8% overall and from 24 to 72% in the last resolution shell. However, even set 20 fulfilled the standard data-quality and significance criteria (Dauter, 1999). The overall *B* value estimated from the Wilson plot changed from 13.2 to 16.1 Å^2 between the first and last data sets, equivalent to a change in the Debye–Waller factor $\exp(-B\sin^2\theta/\lambda^2)$ at 1.45 Å resolution from 0.21 to 0.15, which results in a decrease in intensities of about 30% in the highest resolution shell.

Close inspection of individual reflections revealed dramatic changes in intensity, as shown for example in Table 2. The variation of individual intensities was fitted as a polynomial function of the cumulative scale, as explained in §2.3. The cumulative scales of individual data sets obtained from the

overall merging of all sets together were used as a measure of the integrated incident flux.

The results of the fitting procedure when the same order of the polynomial was imposed for each reflection are given in Table 3 and show that a quadratic function fits the observations best, as judged from the cross-validated residual. This is in keeping with expectation based on the simplified reasoning, illustrated in Fig. 2 and outlined below. The total structure factor $F_{\rm tot}$ can be expressed as the vector sum of the constant part, $F_{\rm const}$, not influenced by radiation damage and the variable part, $F_{\rm var}$, representing atoms with occupancy factors diminishing owing to radiation damage,

$$F_{\rm tot} = F_{\rm const} + pF_{\rm var},$$

and, in terms of intensities,

$$I_{\rm tot} = I_{\rm const} + 2p(I_{\rm const}I_{\rm var})^{1/2}\cos(\varphi_{\rm const} - \varphi_{\rm var}) + p^2 I_{\rm var}.$$

If the occupancy p is linearly dependent on the cumulative exposure x, one obtains the expression

$$I_{\rm tot} = a + bx + cx^2.$$

A quadratic relation is thus obtained in a model where all changing occupancies diminish by the same amount in parallel with the cumulative exposure. In reality, the loss of occupancy of various atoms does not need to be proportional to the cumulative exposure and different atoms may respond differently to irradiation (§3.3). Furthermore, the effective exposure was not strictly uniform throughout the collection of all 20 data sets, but varied depending on the actual synchrotron ring current; therefore, in all calculations we used the cumulative scale resulting from merging all data sets as a quantity approximating the total effective dose received by the crystal. Other sources of non-isomorphism such as changes in atomic displacement factors, side-chain and/or domain movement are not taken into account in this simple model and require a full statistical treatment. It can be noted that the linear term in the parabolic expression depends on the difference between $F_{\rm var}$ and $F_{\rm const}$ phases: a source of information potentially useful in phasing.

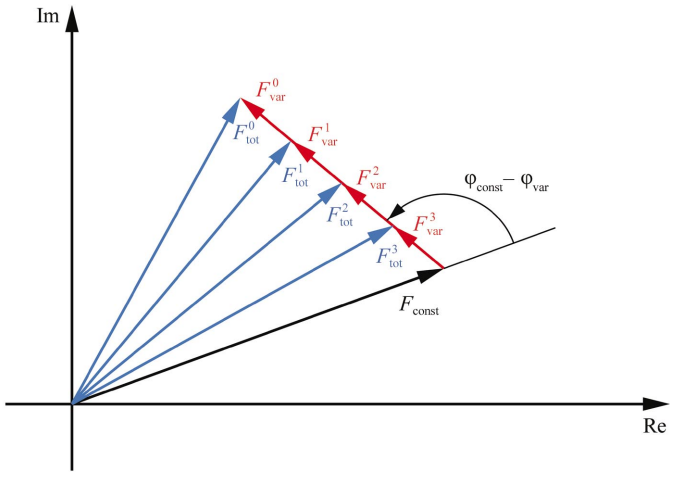

Figure 2
A structure factor in the complex plane, consisting of a constant and a variable part. See §3.1.

Blake & Phillips (1962) and Hendrickson (1976) proposed an exponential model of the intensity decay caused by radiation damage,

$$I(t)/I(0) = \exp[-(k_1 + k_3)t] + [k_1/(k_1 + k_3 - k_2)] \times \exp(-k_2 t)\{1 - \exp[-(k_1 + k_3 - k_2)t]\} \times \exp(-D\sin^2\theta/\lambda_2),$$

where D is an atomic displacement parameter and k_1, k_2 and k_3 are the rates of change between three states of the exposed sample: fully ordered, disordered and amorphous. They correspond to respective transitions from ordered to

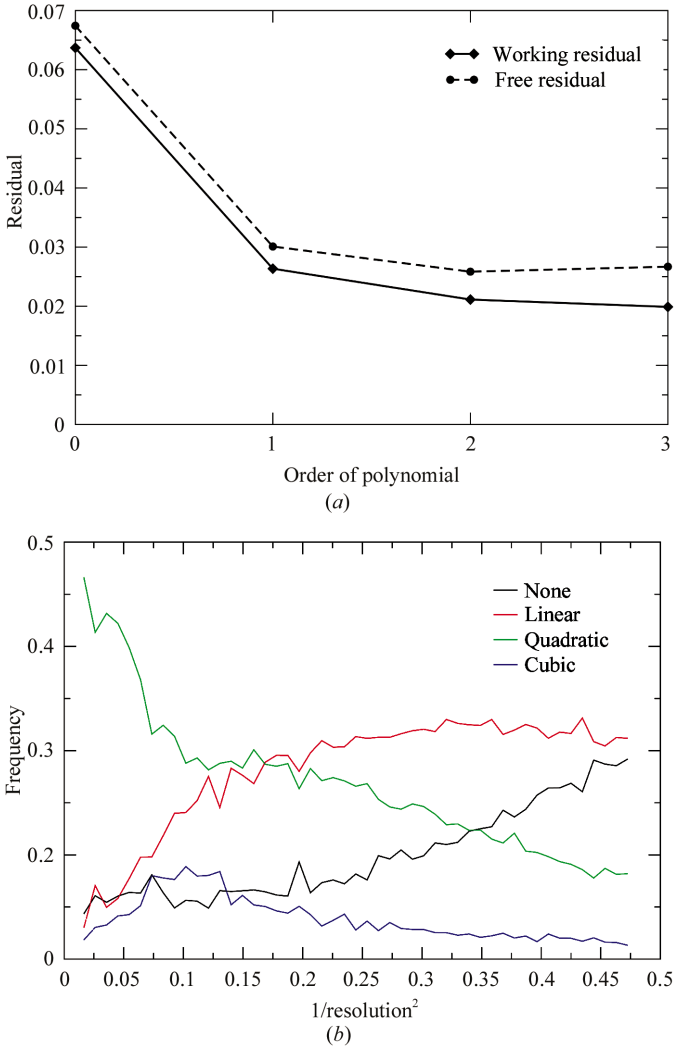

Figure 3
(*a*) Residuals (continuous lines) and free residuals (dashed lines) obtained from the polynomial fit of reflection intensities imposing the same order of the function on all reflections. (*b*) Results of selection of the best polynomial fit for individual reflections, presented in resolution shells. The black curve shows the percentage of reflections for which no radiation-damage correction leads to the best fit as judged by the cross-validated residual. The red line corresponds to a linear correction, the green line to a quadratic function and the blue line to a cubic correction being preferred on the same basis.

disordered (k_1), from disordered to amorphous (k_2) and directly from ordered to amorphous (k_3) states.

The Blake–Phillips–Hendrickson approach treats the diffraction decay globally and their exponential model of intensity changes applies to the average diffracted intensity of all reflections (in a particular resolution range). It correctly predicts that the diffracted intensity cannot be negative and does not allow it to increase as a consequence of radiation damage. In contrast, our model applies to individual reflection intensities and permits their increase as well as decrease, a behavior clearly seen in Table 2. In fact the quadratic model also excludes negative intensities, since the branches of the parabola point upwards ($c = I_{\rm var} > 0$) and its minimum $\{a - b^2/(4c) = I_{\rm const}[1 - \cos^2(\varphi_{\rm const} - \varphi_{\rm var})]\}$ is always positive.

The polynomials describing the behavior of reflection intensities are expressed as a function of the cumulative scale. Taking into account the decrease of the average reflection intensities, as described by the Blake–Phillips–Hendrickson model, the cumulative scale is not strictly proportional to the integrated incident flux, since the amount of the non-diffracting amorphous phase gradually increases during the course of data collection. However, the cumulative scales are appropriate for the zero-dose extrapolation procedure, since they take into account the total integrated flux weighted by the amount of diffracting matter.

Although the cross-validated residuals for the four polynomial orders are apparently very close, as shown in Fig. 3(*a*), the estimated standard deviations in Table 3 indicate that their differences are significant.

A better way of modelling radiation-damage effects might be by determining the order of the polynomial fit for each reflection rather than imposing it on the basis of a global residual. Deciding on a single-reflection basis has the advantage that modelling radiation-damage effects for poorly defined intensities will not be overfitted by a high-order polynomial. The results of this procedure are shown in Fig. 3(*b*), where the frequency of reflections for which a given order of the polynomial leads to the best fit is plotted as a function of resolution for all reflections. To about 2.2 Å resolution, a quadratic polynomial occurs most frequently. At higher resolution, lower order polynomials are more prevalent, most likely owing to inaccuracies in the estimation of intrinsically weak intensities and to resolution-dependent non-isomorphism effects, for example unit-cell changes or disorder, the effects of which are more prominent at high resolution.

3.2. Structural changes caused by irradiation

The atomic model of thaumatin was refined against data sets 1, 5, 10, 15 and 20 with *REFMAC*5 (Murshudov *et al.*, 1999). The resulting values of R and $R_{\rm free}$ were comparable (Table 4), although the average values of the B factors showed an increasing tendency from data set 1 to 20, in keeping with the estimated $B_{\rm Wilson}$ values (Table 1). The average difference of calculated phases between pairs of refined models is included in Table 4.

Table 4
Statistics of refined models.

Refinement R (%) and $R_{\rm free}$ (%) values are shown on the diagonal. The upper triangle contains the r.m.s. phase difference between pairs of phase sets.

Data set	1	5	10	15	20
1	15.6/17.5	4.7	7.4	9.6	11.3
5		15.8/17.8	3.7	5.8	7.6
10			16.1/18.2	3.6	5.3
15				16.2/18.5	3.5
20					16.4/18.6

Table 5
Occupancies and atomic displacement parameters (B values, Å^2) for the cysteine sulfurs in data sets 1, 5, 10, 15 and 20 refined by *SHELXL*.

Cysteines in thaumatin are paired in the following way: 9–204, 56–66, 71–77, 121–193, 126–177, 134–145, 149–158 and 159–164.

Cysteine	Data set 1	Data set 5	Data set 10	Data set 15	Data set 20
9	1.00 (11.1)	1.00 (12.0)	0.97 (12.4)	0.95 (13.3)	0.93 (14.0)
56	0.94 (13.2)	0.89 (14.0)	0.87 (14.5)	0.84 (15.3)	0.85 (16.5)
66	0.94 (13.3)	0.82 (15.0)	0.80 (16.3)	0.77 (18.0)	0.75 (19.0)
71	0.91 (16.0)	0.83 (17.0)	0.81 (19.1)	0.80 (20.6)	0.79 (22.5)
77	0.90 (14.1)	0.86 (15.0)	0.84 (16.0)	0.82 (17.0)	0.80 (17.5)
121	0.97 (16.0)	0.95 (18.0)	0.91 (20.0)	0.87 (22.0)	0.85 (24.0)
126	0.85 (9.5)	0.72 (13.0)	0.63 (15.0)	0.59 (17.0)	0.56 (19.0)
134	0.92 (9.5)	0.82 (11.0)	0.76 (12.0)	0.68 (13.1)	0.64 (14.5)
145	0.90 (10.4)	0.84 (12.0)	0.82 (13.3)	0.79 (14.1)	0.78 (15.0)
149	0.90 (10.5)	0.84 (12.0)	0.76 (14.0)	0.69 (16.0)	0.64 (18.0)
158	0.99 (11.0)	0.87 (12.0)	0.83 (14.0)	0.76 (16.0)	0.72 (18.0)
A159	0.65 (19.0)	0.56 (24.0)	0.57 (28.4)	0.60 (33.0)	0.59 (35.0)
B159	0.57 (22.0)	0.47 (26.0)	0.42 (28.0)	0.43 (32.0)	0.40 (33.0)
164	0.96 (16.1)	0.88 (20.0)	0.80 (22.0)	0.73 (24.0)	0.68 (26.0)
177	0.84 (9.4)	0.69 (12.0)	0.62 (14.0)	0.56 (17.0)	0.54 (19.0)
193	0.98 (14.5)	0.89 (16.0)	0.89 (18.0)	0.84 (19.0)	0.84 (21.0)
204	0.92 (11.0)	0.85 (12.0)	0.79 (13.0)	0.73 (14.0)	0.69 (15.0)

The structural features induced by irradiation were visualized in the isomorphous difference maps using an $(|F_{1,\rm obs} - F_{15,\rm obs}|, \varphi_{1,\rm calc})$ synthesis. The most prominent features were present at the sites of disulfide bonds (Fig. 4). As observed previously (Burmeister, 2000; Weik *et al.*, 2000; Ravelli & McSweeney, 2000), the arrangement of positive and negative peaks indicated disulfide-bond breakage and conformational changes of the side chains of freed cysteines. Other prominent difference density features were found at the sites of carboxyl groups of aspartatic and glutamic acids as well as at the C-terminal carboxylate group (Fig. 4*b*). Additional features are present at the sites of ordered solvent water molecules. Less prominent features corresponded to the side-chain oxygen atoms of tyrosines, threonines and serines.

Interestingly, main-chain and side-chain conformational rearrangements could be identified where new positions of freed cysteine side chains would cause steric clashes with surrounding residues. Fig. 4(*a*) illustrates the movement of peptide planes and a phenylalanine side chain that make more space available for the released side chains of Cys134 and Cys145.

The models obtained from *REFMAC*5 were subsequently refined with *SHELXL* against data sets 1, 5, 10, 15 and 20 with the full resolution of 1.45 Å using a block-matrix least-squares option. The occupancies of sulfur atoms were refined along

with the atomic displacement and positional parameters of all non-H atoms. The estimated standard deviations of the S-atom occupancies obtained were about 0.03. Table 5 shows that the sulfur occupancies diminished by 10–20% between the first and last data set. Such a change is significant in the light of the estimated standard deviations of the occupancies. Even the changes of the occupancies between the first and fifth data set seem to be significant. A similar procedure was applied to carboxylate groups of aspartate and glutamate residues, where the COO-group occupancies were refined, as well as to a number of solvent oxygen atoms with low atomic displacement parameters. However, for those atoms the occupancy changes were less prominent in comparison to the behavior of the sulfur atoms.

Whereas the difference map clearly shows the breakage of S—S bridges and is suggestive of the new locations of the released cysteine side chains (Fig. 4*a*), attempts to model and refine the cysteines as a mixture of bonded and broken species were unsuccessful, probably owing to low occupancies and possible disorder of their new positions. The disulfide bridges 126–177 and 149–158 were most susceptible to radiation damage over the whole course of exposure. No correlation can be observed between the solvent-accessible surface of the cysteine residues and their susceptibility to radiation damage, similarly to the observations of Ravelli & McSweeney (2000), but in contrast to the results of Weik *et al.* (2000) and Leiros *et al.* (2001), who reported a low-level correlation between these two parameters.

The difference maps made it possible identify a number of less prominent features. Some carboxylate groups of aspartates were clearly affected by the radiation damage (Fig. 4*b*), whereas other were not. Asp70 and Asp147, which are in close proximity to cysteine residues, seemed to suffer more from radiation damage than the other aspartates. A number of glutamate residues were also affected (noticeably Glu89), but fewer in comparison with the aspartates. Some threonine, tyrosine and serine residues seem to lose part of their hydroxyl moieties. For two lysines (67 and 106) the terminal amino group was affected. A significant amount of negative and positive density appears for methionine 112, showing that this residue is affected significantly, as also shown from the occupancy refinement. Whereas the electron density around the C-terminus was poorly defined and no conclusion about the possible decarboxylation of the terminal carboxylate can be drawn, the N-terminus showed a gradual build-up of negative difference density in the difference maps over the course of exposure, indicating an increase of disorder of this residue.

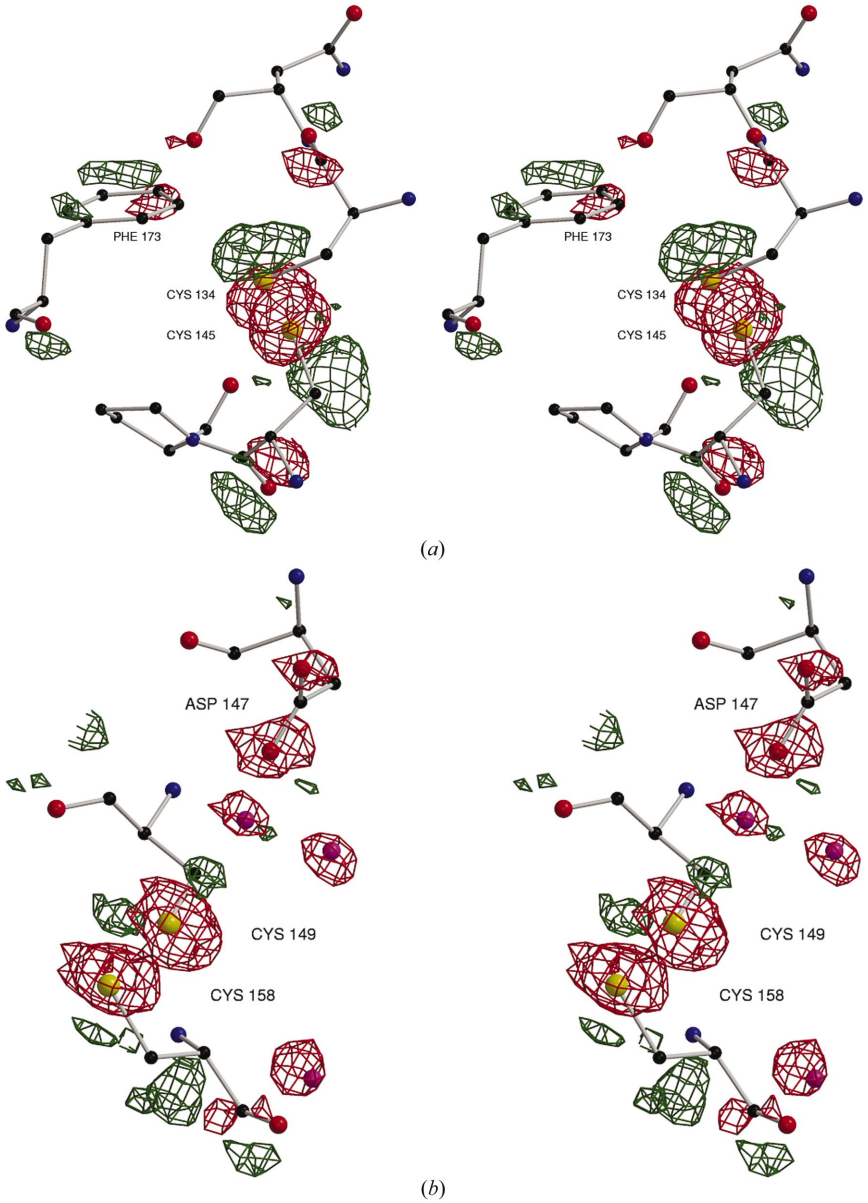

Figure 4
(*a*) ($|F_{1,\mathrm{obs}}| - |F_{15,\mathrm{obs}}|$, φ_1) maps at -4.5σ (red) and $+4.5\sigma$ (green) levels around the disulfide bridge Cys134–Cys145. The breakage of the disulfide bond results in conformational changes of the side chains of the cysteine residues, which induce further rearrangements of the surrounding main-chain and side-chain atoms. (*b*) Difference density as in (*a*) illustrating the dissociation of the carboxyl group of Asp147. The density present at the three water sites shown here is typical for most well ordered waters surrounding the protein molecule.

3.3. SIR-like phasing (RIP)

The intensity differences at 2.2 Å between data sets 1 and 5, data sets 1 and 10, and data sets 1 and 15 were used to determine positions of the substructure responsible for the isomorphous difference signal. The eight strongest sites generated by *SHELXD* for the data-set pairs 1 and 10, and 1 and 15 corresponded to all disulfide bridges present in the known structure of

thaumatin. The results of *SHELXD* on the data set pair 1 and 5 revealed six clear sites corresponding to all disulfides except cysteine pairs 9–204 and 121–193.

The eight (superatom) peaks from the data-set pairs 1 and 10, and 1 and 15, as well as six peaks from pair 1 and 5 were fed into *SHARP* for phasing at 1.45 Å resolution with an SIR protocol similar to that used by Ravelli *et al.* (2003). Subsequent solvent flattening with *SOLOMON*/*DM* resulted in a phase set with an average phase error of 31.7 and a map correlation to the final map of 91% for data pair 1/10. *ARP*/*wARP* was able to automatically build 195 out of 207 residues in this map. The phasing and automated model building for the pairs 1/5 and 1/15 were also successful, although a lower map correlation was obtained (70 and 80%, respectively). The phasing statistics are summarized in Table 6. The poorer performance for the 1/5 pair can be attributed to the lower amount of isomorphous signal present (only six out of eight superatoms were used in the phasing), whereas the lower performance for the 1 and 15 data set pair could originate from the relatively lower quality of data set 15 (Table 1). Nevertheless, Fig. 5 shows the remarkably high quality of the electron-density maps originating from the data set pair 1 and 15.

Table 6
RIP phasing results for various pairs of data sets.

The phasing power as well as the mean phase difference ($\Delta\varphi$) and the map correlation coefficients (m.c.c.) to the final phases are given.

Data sets	1 and 5	1 and 10	1 and 15	1 (anomalous)
SHARP phasing power	0.50	0.68	0.74	0.25
SHARP $\Delta\varphi$ (°)/m.c.c. (%)	70.2/29.3	77.0/24.5	70.6/31.3	73.6/25.0
SOLOMON $\Delta\varphi$ (°)/m.c.c. (%)	51.5/69.7	31.7/90.9	41.7/82.0	27.2/91.9
Residues built out of 207/ No. chains	192/5	195/5	188/5	195/5

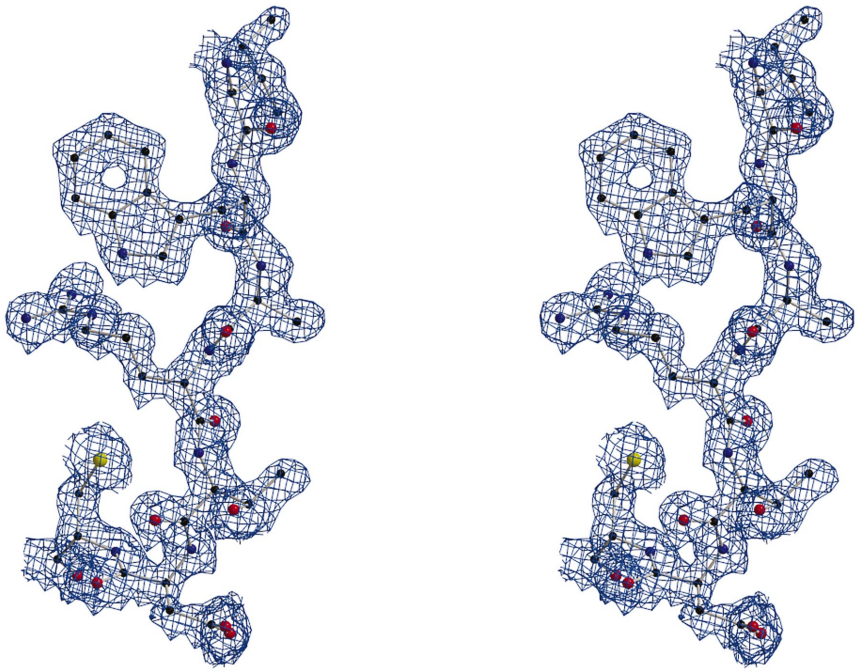

Figure 5
$2F_o - F_c$ map contoured at the 1.5σ level obtained from radiation-damage-induced phasing (RIP) and solvent flattening around residues 50–56.

3.4. Anomalous phasing attempts

It was recently reported by Ramagopal *et al.* (2003) that the limit of anomalous signal sufficient for successful SAD phasing is of the order of 0.6%, in agreement with the earlier estimations of Wang (1985). Although the data were collected with $\lambda = 0.98$ Å, corresponding to f''(S) of about 0.25 e and an expected Bijvoet ratio $\langle\Delta F^{\pm}\rangle/\langle F\rangle$ of 0.51%, the anomalous difference maps computed with data set 1 and model phases showed substantial peaks (at about the 7σ level) around the S atoms. In addition, the correlation coefficient between the anomalous differences in pairs of data sets as computed by *XPREP* reaches about 60% at low resolution and still exceeds 30% to a resolution of about 2.5 Å. This indicates that in spite of the short wavelength used, the data sets contain a significant amount of anomalous signal. This signal was, however, not sufficient to locate S-atom positions by direct methods from the Bijvoet differences.

Inclusion of the anomalous scattering effect in SIRAS-like phasing increased the quality of the phases only marginally and therefore this effect has not been used in the previously discussed phasing strategy. However, using the known positions of all individual S atoms and the anomalous signal from data set 1, *SHARP* produced a phase set with an average phase error close to 70°. Subsequent solvent flattening resulted in electron-density maps with a correlation to the final electron density of about 90% (Table 6). These results suggest that the amount of signal required to locate the anomalous scatterers on the basis of the Bijvoet differences is higher than that sufficient to estimate phases accurate enough to initiate successful density modification. This may be explained by the fact that the Bijvoet differences only approximate the F_A values, properly representing the scattering contribution of the anomalous atoms. Moreover, in the *SHARP* phasing process the anomalous scatterers additionally contribute to the protein phase estimation *via* their known partial structure contribution. In this case, the sulfur substructure was large enough to phase the data set using *ACORN* (Foadi *et al.*, 2000) and data set 1 without anomalous signal, in spite of the data extending only to 1.45 Å resolution. This procedure resulted in an electron density with a correlation of 60% to the final map, which was easily interpretable in an automatic way by *ARP*/*wARP*.

4. Conclusion

The effect of radiation damage on the protein structure is considerable. The occupancy changes of the affected parts of the

protein as well as side-chain and main-chain movement are reflected in sometimes substantial intensity changes over time. Although the changes in occupancy are not equal for each atom, the intensity changes are best described by a parabolic function, as judged from the results of the cross-validation procedures. Clearly, more sophisticated models taking into account changes in atomic displacement and positional parameters caused by radiation damage should be developed.

Diederichs *et al.* (2003) proposed a method of extrapolation intensities to a zero-dose exposure based on overall scaling of a single redundant data set. Estimation of the parameters describing the linear dependence of the intensity changes owing to radiation damage for each reflection is performed as a separate step, independent of scaling. Our procedure is somewhat simpler, with averaging of the variation of reflection intensities within a single data set. It may be more advantageous to conduct both the scaling and the zero-dose extrapolation steps in a single procedure.

Even the simple SIR-like phasing approach, also known as RIP, results in maps of extremely high quality. This can be ascribed to the relatively large number of missing electrons responsible for the 'isomorphous' differences present in measured intensities. For instance, the difference between data set 1 and 10 is caused by about two electrons per S atom, which is comparable to the anomalous signal (f'') of sulfur at a wavelength of 3.2 Å, which is too long for practical use owing to very severe absorption effects. In this particular case, the total occupancy change of S atoms amounts to about 32 electrons generating the isomorphous difference. This is approximately equivalent to the addition of a fully occupied Zn atom and is clearly large enough to solve this particular protein. The RIP phasing is thus able to provide a signal which is substantially larger than that obtainable from sulfur anomalous scattering at practically available wavelengths on most synchrotron beamlines.

Interestingly, the sulfur substructure was sufficient to phase the full structure against a single native data set at 1.45 Å resolution without the use of the anomalous or isomorphous signal. However, it is impossible to locate the substructure using only the native data at this resolution. The use of isomorphous and anomalous signal permits the location of the substructure and provides additional phasing power. The amount of information originating from the partial structure in the phasing of the native data will be investigated further. In favourable cases, the anomalous or isomorphous signal is only necessary for the location of the substructure and can be treated as an auxiliary source of phase information.

This work was supported in part with Federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. NO1-CO-12400.

References

Abrahams, S. C. & Marsh, P. (1987). *Acta Cryst.* A**43**, 265–269.
Arndt, U. W. & Wonacott, A. J. (1977). *The Rotation Method in Crystallography.* Amsterdam: North Holland.
Blake, C. C. F. & Phillips, D. C. (1962). *Biological Effects of Ionizing Radiation at the Molecular Level*, pp. 183–191. Vienna: IAEA.
Brünger, A. (1992). *Nature (London)*, **355**, 472–474.
Burmeister, W. P. (2000). *Acta Cryst.* D**56**, 328–341.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cowtan, K. (2002). *J. Appl. Cryst.* **35**, 655–663.
Dauter, Z. (1999). *Acta Cryst.* D**55**, 1703–1717.
Diederichs, K., McSweeney, S. & Ravelli, R. G. B. (2003). *Acta Cryst.* D**59**, 903–909.
Dudewicz, E. J. & Mishra, S. N. (1988). *Modern Mathematical Statistics.* New York: Wiley.
Foadi, J., Woolfson, M. M., Dodson, E. J., Wilson, K. S., Jia-Xing, Y. & Chao-De, Z. (2000). *Acta Cryst.* D**56**, 1137–1147.
Garman, E. (2003). *Curr. Opin. Struct. Biol.* **13**, 545–551.
Garman, E. & Nave, C. (2002). *J. Synchrotron. Rad.* **9**, 327–328.
Gonzalez, A. & Nave, C. (1994). *Acta Cryst.* D**50**, 874–877.
Henderson, R. (1990). *Proc. R. Soc. London Ser. B*, **241**, 6–8.
Hendrickson, W. A. (1976). *J. Mol. Biol.* **106**, 889–893.
Hope, H. (1988). *Acta Cryst.* B**44**, 22–26.
Ko, T.-P., Day, J., Greenwood, A. & McPherson, A. (1994). *Acta Cryst.* D**50**, 813–825.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Leiros, H. K., McSweeney, S. M. & Smalås, A. O. (2001). *Acta Cryst.* D**57**, 488–497.
Müller, R., Weckert, E., Zellner, J. & Drakopoulos, M. (2002). *J. Synchrotron Rad.* **9**, 368–374.
Murray, J., Garman, E. & Ravelli, R. B. G. (2004). Submitted.
Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.
O'Neill, P., Stevens, D. L. & Garman, E. F. (2002). *J. Synchrotron Rad.* **9**, 329–332.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003). *Acta Cryst.* D**59**, 1020–1027.
Ravelli, R. B. G. & McSweeney, S. M. (2000). *Structure*, **8**, 315–328.
Ravelli, R. B. G., Schrøder-Leiros, H. K., Pan, B., Caffrey, M. & McSweeney, S. (2003). *Structure*, **11**, 217–224.
Ravelli, R. B. G., Theveneau, P., McSweeney, S. & Caffrey, M. (2002). *J. Synchrotron Rad.* **9**, 355–360.
Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.
Sheldrick, G. M. (2001). *XPREP* program, v. 6.13. Bruker–Nonius Inc., Madison, Wisconsin, USA.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Sliz, P., Harrison, S. C. & Rosenbaum, G. (2003). *Structure*, **11**, 13–19.
Teng, T.-Y. (1990). *J. Appl. Cryst.* **23**, 387–391.
Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
Weik, M., Kryger, G., Schreurs, A. M. M., Bouma, B., Silman, I., Sussman, J. L., Gros, P. & Kroon. J. (2001). *Acta Cryst.* D**57**, 566–573.
Weik, M., Ravelli, R. B. G., Kryger, G., McSweeney, S., Raves, M. L., Harel, M., Gros, P., Silman, I., Kroon, J. & Sussman, J. L. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 623–628.
Weisstein, E. W. (1999). *Least-Squares Fitting – Polynomial.* http://mathworld.wolfram.com/LeastSquaresFittingPolynomial.html.
Yonath, A. *et al.* (1998). *Acta Cryst.* A**54**, 945–955.