Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Radiation-induced site-specific damage of mercury derivatives: phasing and implications

Udupi A. Ramagopal,[a]* Zbigniew Dauter,[b] Radhakannan Thirumuruhan,[c] Elena Fedorov[a] and Steven C. Almo[a,d,e]*

[a]Department of Biochemistry, Albert Einstein College of Medicine, Yeshiva University, Bronx, NY 10461, USA, [b]Synchrotron Radiation Research Section, National Cancer Institute, Argonne National Laboratory, Building 202, Argonne, IL 60439, USA, [c]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA, [d]Department of Physiology and Biophysics, Albert Einstein College of Medicine, Yeshiva University, Bronx, NY 10461, USA, and [e]Center for Synchrotron Bioscience, Albert Einstein College of Medicine, Yeshiva University, Bronx, NY 10461, USA

Correspondence e-mail: acharya@bnl.gov, almo@aecom.yu.edu

Received 9 February 2005
Accepted 12 July 2005

The behavior of mercury-derivatized triclinic crystals of a 60 kDa protein target from the New York Structural GenomiX Research Consortium provides novel insights into the mechanism of heavy-atom-specific radiation damage and its potential exploitation for *de novo* structure solution. Despite significant anomalous signal, structure solution by classic SAD and MAD phasing approaches was not successful. A detailed analysis revealed that significant isomorphic variation of the diffracted intensities was induced by X-ray irradiation. These intensity changes allowed the crystal structure to be solved by the radiation-damage-induced phasing (RIP) technique. Inspection of the crystal structure and electron-density maps demonstrated that the covalent S—Hg bonds at all four derivatized cysteine sites were much more susceptible to radiation-induced cleavage than other bonds typically present in native proteins. A simple diagnostic is described to identify the fingerprint of such decay at the time of data collection/processing. The rapid radiation-induced decomposition of mercury adducts is consistent with the difficulties frequently associated with the experimental phasing of mercury derivatives and suggests a straightforward solution to overcome this limitation by radiation-damage-induced phasing with anomalous scattering (RIPAS). These results indicate that historically recalcitrant and newly emerging difficulties associated with Hg phasing should be revisited.

1. Introduction

The intense X-ray fluxes associated with synchrotron sources can result in significant damage to macromolecular crystals, even when cooled to cryogenic temperatures (about 100 K). Numerous efforts have focused on understanding the underlying physical mechanism of radiation damage and the development of methods to abate it or to extrapolate the data to zero dose (Gonzalez & Nave, 1994; Burmeister, 2000; Ravelli & McSweeney, 2000; Garman & Nave, 2002; Murray & Garman, 2002; O'Neill *et al.*, 2002; Diederichs *et al.*, 2003).

X-ray absorption is directly responsible for primary damage, which manifests itself by the cleavage of covalent bonds and the production of free radicals. The free radicals react further to produce a cascade of secondary radicals that can affect lattice contacts, resulting in the degradation of crystalline order, and cause further bond breakage (Henderson, 1990; Gonzalez & Nave, 1994; Murray & Garman, 2002). At cryogenic temperatures the mobility of free radicals is restricted and the secondary effects are to a large extent diminished, which leads to a considerable improvement of the crystal lifetime. However, the primary effects are independent of temperature and it has been esti-

mated that a total absorbed dose of 2×10^7 Gy will halve the diffraction power of a protein crystal, even if it is cooled to ~80 K (Henderson, 1990). This estimate is consistent with the observations of Teng & Moffat (2000).

Even at third-generation synchrotron sources this intrinsic limitation allows sufficient time, in most cases, to obtain the required data from reasonably diffracting crystals. However, the difficulties associated with radiation damage become more acute when relatively long data-collection times are required, as is the case in collecting atomic resolution data (Leiros *et al.*, 2001). It has also been observed that the presence of heavy atoms can significantly decrease crystal lifetimes, which is likely to be a consequence of the increased X-ray absorption cross-section (Garman & Murray, 2003; Ennifar *et al.*, 2002). Moreover, specific damage involving heavy atoms is highly detrimental to the relatively weak anomalous signals generated in MAD or SAD experiments and radiation damage appears to be the major cause of unsuccessful MAD experiments performed at third-generation synchrotron sources (Murray & Garman, 2002). In contrast to these apparent difficulties, it has also been demonstrated that if the damage is specific or more pronounced at the sites of heavy atoms (*e.g.* sulfur), it can be exploited for phasing macromolecular structures through a recently developed methodology known as 'radiation-damage-induced phasing' (RIP; Ravelli *et al.*, 2003).

The presence of heavy or anomalously scattering atoms is essential for the *de novo* solution of novel macromolecular structures, unless atomic resolution data are available and the size of the asymmetric unit is within the current reach of direct methods. The anomalous scatterer may be intrinsic to the structure, such as various metals in metalloproteins appropriate for MAD or SAD phasing, or sulfur (Hendrickson & Teeter, 1981; Wang, 1985; Dauter *et al.*, 1999; Liu *et al.*, 2000; Ramagopal *et al.*, 2003 and references therein; Doan & Dokland, 2003; Yang *et al.*, 2003; Debreczeni, Bunkóczi, Girmann *et al.*, 2003; Debreczeni, Bunkóczi, Ma *et al.*, 2003; Debreczeni, Girmann *et al.*, 2003; Weiss *et al.*, 2004; Olsen *et al.*, 2004; Szyk *et al.*, 2004; Djinovic Carugo *et al.*, 2005; Leonard *et al.*, 2005) and phosphorus (Dauter & Adamiak, 2001), which are present in most proteins and nucleic acids, respectively. Most notably, the ability to routinely achieve the biosynthetic incorporation of selenomethionine (SeMet) has revolutionized X-ray crystallography and is typically the method of choice for *de novo* structure determination (Hendrickson *et al.*, 1990). In situations where SeMet-derivatized material is not available and the protein is known to contain free cysteine residues, mercurials have historically been the first choice for derivatization (Blundell & Johnson, 1976; Boggon & Shapiro, 2000). Despite their widespread utility, mercury derivatives exhibit a relatively high rate of failure when employed in MAD experiments (Ji *et al.*, 2001). Considering such difficulties, Ji *et al.* (2001) suggested a 'double-edge' strategy for HgMAD data, in which they propose data acquisition at two edge wavelengths differing by 4 eV; however, no definitive conclusions were presented to explain the factors responsible for these failures. Furthermore, a comparative study of phasing the AP2 complex with Xe, Hg and Se suggested that phases obtained with Hg were much inferior to those obtained with Xe (Evans, 2003).

In this report, we describe our experiences with the structure solution of the 60 kDa *Escherichia coli* yidA (P09997), a member of the haloacid dehydrogenase superfamily and a target protein of the New York Structural GenomiX Research Consortium (NYSGXRC), which was solved by exploiting the heavy-atom-specific radiation damage of mercury-derivatized crystals. As these experiments were performed with cryo-cooled crystals on a second-generation bending-magnet beamline with a relatively short exposure time (10 s per 1° frame), there was no expectation of such radiation-induced damage. The inability of standard SAD and MAD protocols to yield a structure solution prompted a careful examination of the data. In particular, the anomalous signal-to-noise ratio showed a substantial decrease as the data collection proceeded, suggesting the possibility of radiation damage. Remarkably, this radiation-induced decay allowed the successful phasing of this previously unknown protein structure. Given the widespread use of mercury derivatives for *de novo* phasing, we performed a more detailed study to reveal the underlying mechanistic basis for the frequent occurrence of HgMAD failures. In additional experiments, a number of successive 360° data sets were collected at the wavelength corresponding to the maximum f'' contribution of the Hg $L_{\rm III}$ absorption edge. In an independent experiment, data were collected on the long-wavelength side of the mercury absorption edge to minimize the anomalous scattering contribution and to exploit only the radiation-damage-induced isomorphous differences for phasing. A retrospective analysis of data from the mercury-derivatized crystals of another NYSGXRC target (*E. coli* yjcF) revealed similar behavior. A simple diagnostic is described that identifies such damage at the time of data collection/processing.

2. Experimental

Crystals of *E. coli* yidA were grown by hanging-drop vapor diffusion by mixing equal volumes of 20 mg ml^{-1} protein solution (in 10 m*M* HEPES pH 7.5, 150 m*M* NaCl, 10 m*M* methionine, 10% glycerol, 5 m*M* DTT) and well solution (20% PEG 3350 and 0.2 *M* $MgCl_2$). Native crystals were flash-cooled in mother liquor supplemented with 10% ethylene glycol, while mercury-derivatized crystals were directly mounted at 100 K in their original liquor without back-soaking or addition of cryoprotectant. All diffraction data were collected on the X9A beamline at the National Synchrotron Light Source (Brookhaven National Laboratory) using a MAR165 CCD detector. The data were processed and merged with *HKL*2000 (Otwinowski & Minor, 1997). The crystals exhibited diffraction consistent with space group *P*1 and had rather high mosaicity, generally exceeding 1°. The statistics of all the diffraction data are presented in Table 1.

The heavy-atom constellation was solved with *SHELXD* (Schneider & Sheldrick, 2002) and protein phases were calculated with *MLPHARE* (Otwinowski, 1991; Collaborative

Table 1
Diffraction data.

Crystal	NatS	Hg1a	Hg1b	Hg1c	Hg2aL	Hg2aH	Hg2bL	Hg2bH	Hg2cL	Hg2cH	Hg2dL	Hg2dH	Hg3a180	Hg3b
Space group	*P*1	*P*1	*P*1	*P*1	*P*1	*P*1	*P*1	*P*1	*P*1	*P*1	*P*1	*P*1	*P*1	*P*1
Unit-cell parameters														
a (Å)	45.04	45.10	45.05	45.07	44.95	44.96	44.97	44.96	44.97	44.98	44.99	45.00	45.06	45.04
b(Å)	54.76	54.58	54.51	54.55	53.92	53.95	53.93	53.93	53.93	53.97	53.96	53.99	54.81	54.78
c (Å)	67.58	67.42	67.45	67.47	67.26	67.27	67.27	67.27	67.27	67.29	67.29	67.32	67.53	67.47
α (°)	112.6	112.4	112.4	112.4	112.1	112.2	112.1	112.1	112.1	112.3	112.1	112.2	112.6	112.6
β (°)	96.4	96.3	96.2	96.2	96.6	96.6	96.7	96.7	96.7	96.7	96.7	96.7	96.4	96.4
γ (°)	106.8	107.2	107.2	107.2	106.4	106.4	106.4	106.4	106.4	106.4	106.4	106.4	106.7	106.7
Exposure (s per degree)	12	20	8	8	10	20	10	20	10	20	10	20	20	20
Mosaicity range (°)	0.43–0.60	0.93–1.56	0.88–1.57	0.87–1.56	0.55–0.84	0.63–0.82	0.57–0.92	0.61–0.90	0.57–0.94	0.61–0.83	0.57–0.89	0.61–0.85	0.78–1.04	0.81–1.0
Wavelength (Å)	1.740	1.009	1.009	1.001	1.001	1.001	1.001	1.001	1.001	1.001	1.001	1.001	1.254	1.254
Resolution (Å)	1.86	2.3	2.85	2.75	2.1	1.6	2.1	1.6	2.1	1.6	2.1	1.6	2.0	2.0
Total rotation (°)	1666	360	360	360	360	360	360	360	360	360	360	360	180	180
Multiplicity	7.6	1.9	1.94	1.94	1.98	2.0	1.98	2.0	1.98	2.0	1.98	2.0	1.91†	1.91†
Completeness (%)	94.8	94.8	91.0	96.8	98.0	96.3	97.7	95.9	98.0	96.5	98.0	96.3	94.4	95.5
Highest resolution shell	87.0	80.0	83.2	97.6	96.9	94.9	96.9	94.8	96.9	94.8	96.4	94.8	89.4	92.5
$R_{\rm merge}$ (%)	5.5	4.9	4.2	4.4	4.7	5.2	3.4	4.1	3.2	3.3	3.3	4.2	4.8	3.3
Highest resolution shell	30.3	12.3	8.1	8.7	9.0	21.2	6.6	21.15	7.8	35.1	10.0	44.1	13.7	14.8
$I/\sigma(I)$	29.0	12.7	15.1	14.2	11.64	14.3	12.14	16.9	18.16	21.9	18.18	19.9	10.8	17.2
Highest resolution shell	3.5	5.0	7.9	8.3	8.2	3.7	7.6	3.9	9.8	2.4	7.71	2.1	6.6	5.1
Wilson *B* factor (Å^2)	25.7	35.3	50.8	49.8	23.9	20.7	26.9	21.8	26.9	23.4	28.6	25.1	21.3	22.4

† Multiplicity after merging Friedel mates.

Computational Project, Number 4, 1994), *SHELXE* (Sheldrick, 2002), *SOLVE* (Terwilliger & Berendzen, 1999) or *SHARP* (de La Fortelle & Bricogne, 1997). The final structure was refined at 1.4 Å resolution against data collected from a native crystal and deposited with PDB code 1rkq.

3. Results and discussion

3.1. Initial characterization: native data and sulfur-SAD attempt

The yidA protein contains eight methionines and two cysteines among 281 residues. Assuming two molecules per unit cell, the $V_{\rm M}$ value is 2.3 Å^3 Da^{-1} and the calculated solvent content is 46%. Considering the strong diffraction and the presence of several S atoms in the protein, data were collected at a wavelength of 1.74 Å radiation (Table 1, denoted NatS), with the goal of solving the structure from the sulfur anomalous signal. As the acquisition of highly redundant data is difficult from crystals belonging to space group *P*1, a total of 1666° of data were collected, changing the crystal orientation twice by adjusting the goniometer arcs. The exposure time (10 s per frame of 1° width) was selected to avoid overloads. These data were processed, merged and input to *SHELXD* to find the sulfur sites. All 20 expected sites were located in multiple phasing trials and a clear dimer could be identified from the sulfur substructure (Fig. 1). Unfortunately, attempts to obtain protein phases from this substructure with programs *MLPHARE*, *SHELXE* or *SHARP* did not result in interpretable electron-density maps even after utilization of non-crystallographic symmetry averaging procedures. The calculated correlation coefficient of this map with the refined model was 0.39 after initial phasing and 0.50 with density modification (0.54 with incorporation of non-crystallographic symmetry). Notably, all sulfur sites were at distances greater than 2.0 Å from each other, which indicated the presence of free cysteine residues, suggesting the possibility of derivatization with mercury.

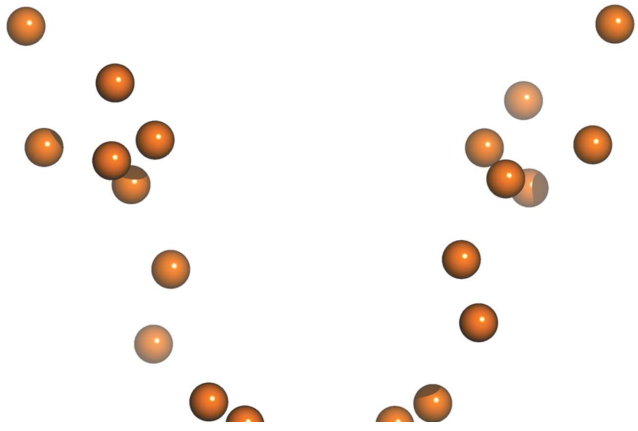

Figure 1
Sulfur substructure solution from NatS data clearly indicating a dimer. All inter-sulfur distances are greater than 2.2 Å, suggesting the availability of free cysteines.

3.2. Radiation-damage-induced phasing with anomalous scattering (RIPAS) on an Hg derivative

In an attempt to solve the structure by HgSAD, a data set (denoted Hg1a) was collected from a mercury-derivatized crystal in a 'high-throughput' quick-soak fashion without back-soaking (Sun & Radaev, 2002). These crystals were prepared by the addition of a few grains of solid mercury acetate [Hg(OAc)$_2$] to the original drop to achieve saturation,

Table 2
Quality of substructure solutions using SAD/RIP/RIPAS approaches represented by the E_{calc}/E_{obs} correlation coefficients (CC) for all/weak reflections as given by *SHELXD* for data from the first derivative crystal.

The first data set (Hg1a) was used as derivative in RIP/RIPAS. In all cases 100 phase trials were attempted and the solutions with the best CC are listed.

Data set	SAD	RIP	RIPAS
Hg1a	37.6/22.3	—	—
Hg1a + Hg1b	38.0/23.6	—	—
Hg1b	31.5/18.9	32.7/20.3	46.0/37.5
Hg1c	27.3/17.4	35.5/22.9	48.1/39.6

followed by a 10–30 min incubation. The X-ray energy was set to 12 290 eV, 6 eV above the L_{III} edge of Hg previously calibrated on a sample of mercury salt, without recording the fluorescence spectra from the crystal (the Hg fluorescence spectrum does not have a white line at any of the three L absorption edges owing to its closed d shell of electrons). The χ^2 test after merging these data clearly showed the presence of a significant anomalous signal and the substructure solution with *SHELXD* identified four strong sites. SAD attempts with these heavy-atom positions using *MLPHARE*, *SHELXE*, *SOLVE* and *SHARP* did not produce an interpretable electron-density map. Data collection was continued at the same energy with shorter exposures (set Hg1b), with the expectation that the overall quality of the data would improve with the increased redundancy. However, merging these two data sets did not enhance the substructure solution or the map. Additional data were collected from the same crystal at the high-energy remote wavelength of 12 385 eV, approximately 100 eV above the Hg L_{III} absorption edge (set Hg1c). Two-wavelength MAD phasing attempts also failed to produce an interpretable map.

Inspection of these data sets revealed that the anomalous signal continuously diminished during the prolonged data collection, suggesting a significant degree of radiation damage. This decay is likely to be the result of cleavage of the sulfur–mercury bonds, resulting in the loss of mercury. Phasing attempts analogous to the radiation-damage-induced phasing technique (RIP; Ravelli *et al.*, 2003) were therefore undertaken. The Hg1b and Hg1a data sets were treated as native and derivative data, respectively. Identification of the Hg sites was now significantly more successful (Table 2) using F_A values estimated from both anomalous and isomorphous (*i.e.* radiation-damage-induced) differences with *XPREP* (Sheldrick, 2001) as well as using only isomorphous differences input to *SHELXD* (Table 2). Phasing was continued with *SHELXE* and the resulting map was clearly interpretable (Fig. 2).

3.3. Tracking the time course of Hg-specific decay

To confirm and further evaluate the effects of radiation damage on the Hg-derivatized crystals, diffraction data were collected from another specimen. This crystal was prepared by adding 1.0 µl mercury acetate-saturated mother liquor to a ∼4.0 µl drop containing crystals, followed by a 30 min incubation. These data were measured more deliberately after

Table 3
Quality of substructure solution using SAD/RIP/RIPAS approaches represented by the E_{calc}/E_{obs} correlation coefficients (CC) as given by *SHELXD* for data from the second derivative crystal of yidA.

In all cases 100 phase trials were attempted and the solutions with the best CC are listed. Values on the diagonal refer to the SAD approach, values above the diagonal correspond to the RIP approach and those below the diagonal correspond to the RIPAS approach.

Data set	Hg2aL	Hg2bL	Hg2cL	Hg2dL
Hg2aL	42.9	32.7	32.7	28.5
Hg2bL	53.3	25.1	26.8	9.1
Hg2cL	49.9	19.8	25.4	9.2
Hg2dL	48.4	16.9	17.2	26.9

encountering the effects of radiation damage with the previous crystal. Four consecutive pairs of 360° data sets were measured, with each pair consisting of a low-exposure set (10 s per degree) followed by a high-exposure set (20 s per degree) (sets denoted Hg2aL, Hg2aH, Hg2bL, ..., Hg2dL, Hg2dH), with the idea of using the radiation-induced isomorphous differences between the low-exposure sets for phasing and the high-resolution sets to refine the occupancies of the mercury sites as data collection progressed. Based on a fluorescence scan of the crystal, the data were collected at 12 384 eV, corresponding to a 'high-energy remote' value convenient for SAD and RIP approaches. Substructure solution was performed for all four low-resolution data sets using anomalous differences (SAD) with *SHELXD* followed by RIP and

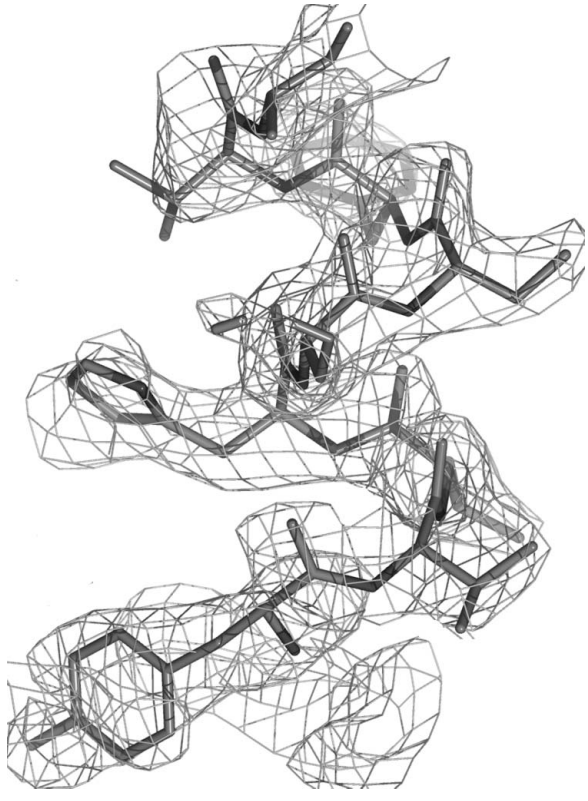

Figure 2
The density-modified map for yidA after RIPAS phasing against Hg1a and Hg1b data, contoured at the 1σ level.

RIPAS phasing trials. In RIP(AS) attempts, the earliest collected data were always treated as derivative and the later data as native (*e.g.* when Hg2aL was treated as derivative, Hg2bL, Hg2cL or Hg2dL were all individually treated as native).

The results shown in Table 3 indicate that the substructure solutions obtained using RIPAS are superior to those obtained by SAD using only Hg2aL data. Also, individual RIP/RIPAS attempts treating Hg2bL, Hg2cL or Hg2dL as native data with Hg2aL as derivative did not show considerable difference in the E_{calc}/E_{obs} correlation coefficients (CC) produced by *SHELXD*. Furthermore, it can be seen from Table 4 that the isomorphous differences between the data sets Hg2bL, Hg2cL and Hg2dL are negligible. It was not possible to obtain good protein phases using SAD with these three data sets and only the initial Hg2aL set led to a satisfactory map using *SHARP* or *SOLVE*, although it failed to give an interpretable map with *MPLHARE* and *SHELXE*. In addition, whereas the *SHELXD* substructure obtained from the Hg2aL data set contains four strong sites, the substructure solutions with Hg2bL, Hg2cL and Hg2dL sets gave only two strong positions, suggesting that two of the four mercury sites decayed faster than the other two.

Fig. 3 illustrates the Hg-omit maps calculated with the Hg2aL and Hg2bL data in the vicinity of Cys67 and Cys141 in one yidA molecule. The Hg atom bound to solvent-exposed Cys141 is evidently 'damaged' much faster than that bound to the predominately buried Cys67. The almost complete loss of the mercury bound to Cys141 is likely to be a consequence of covalent-bond rupture and diffusion of the released mercury. In contrast, the density around Cys67 suggests that only a fraction of the corresponding S—Hg bond is cleaved under these conditions and that the released Hg atom is maintained at significant occupancy at a distance of more than 2.2 Å from its original position (Fig. 3). Evidently, the loop segment composed of residues 77–81, which in the static structure appears to shield Cys67 from solvent, is sufficiently dynamic at room temperature to allow the Hg ion to penetrate and react with the thiol of Cys67. At cryogenic temperatures this loop is significantly less mobile and the heavy atom is 'caught in the cold' and cannot diffuse away after release from the cysteine.

Table 4
R_{merge} values from merging the pairs of data sets from the second derivative crystal of yidA.

Data set	Hg2aL	Hg2bL	Hg2cL	Hg2dL
Hg2aL	0	11.6	11.7	13.8
Hg2bL		0	2.9	4.1
Hg2cL			0	4.1
Hg2dL				0

Table 5 contains the occupancies of Hg atoms refined against all Hg2 low- and high-resolution data sets. Occupancies of Hg atoms were refined by fixing the temperature factor to 1.2 times the average temperature factors of the cysteine SG atoms using the program *SHELXL* (Sheldrick & Schneider, 1997). In the case of Cys67, partial occupancies were refined for the original position of the Hg atom and that arising from decay. The total occupancy for the two positions was not fixed to unity, considering the possibility of reduced total occupancy with exposure.

These results and the cross R_{merge} values (Table 4) indicate that most of the decay takes place between the first and the second low-resolution data sets, *i.e.* during collection of the first high-resolution data set. The statistics in Table 1 also show that the overall quality of the Hg2bL, Hg2cL and Hg2dL data is better than for the Hg2aL set. The influence of radiation damage on reflection intensities appears to level out, suggesting that the structural changes responsible for this effect are rapid and complete prior to the second low-resolution data set. These observations further suggest that the rupture of susceptible Hg—S bonds is completed much earlier than the detrimental global effects (*i.e.* loss of diffraction power). This kinetic behavior differs from that observed previously on systems in which the functionalities most susceptible to radiation damage were disulfide bridges or carboxylates (Ravelli *et al.*, 2003;

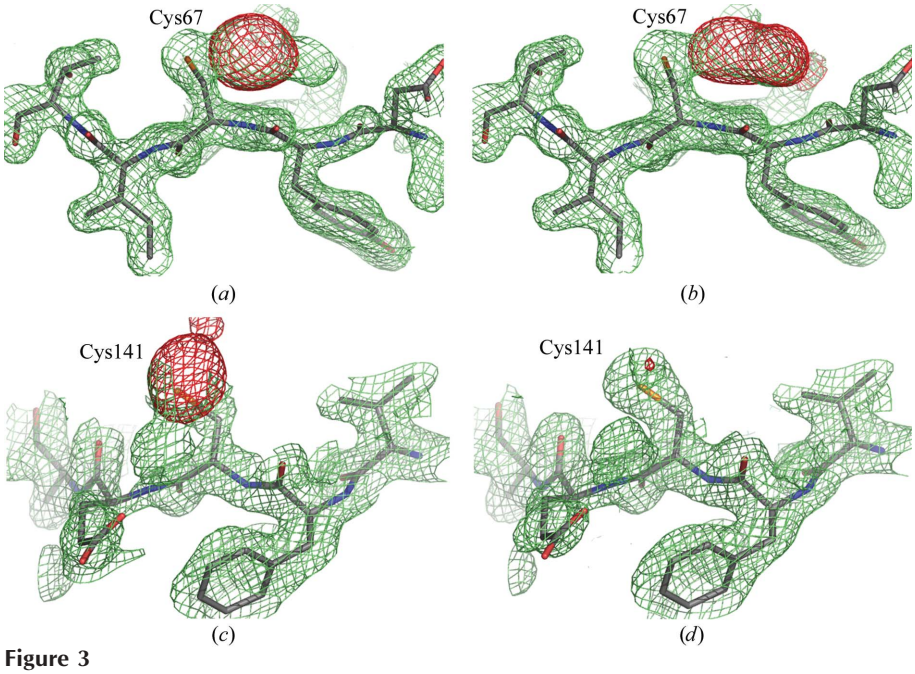

Figure 3
The vanishing mercury (Cheshire cat). The Hg omit map contoured at the 5σ level (red) (*a*, *b*) in the vicinity of Cys67 of yidA, calculated with the Hg2aL and Hg2bL data sets, respectively, and (*c*, *d*) in the vicinity of Cys141, calculated with the Hg2aL and Hg2bL data sets, respectively. The Hg atom that is bound to solvent-exposed Cys141 is evidently 'damaged' much faster than that bound to the predominately buried Cys67. A loop composed of residues 77–81 appears to shield Cys67 from solvent and hence the Hg atom cannot readily diffuse, resulting in a significant occupancy at a distance around 2.2 Å from its original position.

Table 5
Variation of occupancies of Hg-atom sites in the second derivative crystal of yidA refined with *SHELXL*.

Partial occupancies of the original mercury site and the resulting new site arising from decay near Cys67 were refined against Hg2 data sets.

	Hg_67A	Hg_141A	Hg_67B	Hg_141B
Hg2aL	0.81/0.15	0.70	0.82/0.11	0.64
Hg2aH	0.62/0.25	0.37	0.63/0.21	0.48
Hg2bL	0.40/0.27	0.26	0.45/0.24	0.24
Hg2bH	0.41/0.31	0.20	0.41/0.28	0.28
Hg2cL	0.29/0.30	0.16	0.33/0.28	0.15
Hg2cH	0.33/0.32	0.15	0.30/0.29	0.16
Hg2dL	0.23/0.31	0.13	0.24/0.28	0.12
Hg2dH	0.23/0.32	0.10	0.21/0.32	0.11

Table 6
Phasing attempts and results with Hg3aN data as derivative and Hg3b as native data.

Values in parentheses correspond to the highest resolution bin.

Data set	Completeness (%)	Sites found	FOM (*SOLVE*)	MapCC before/after *DM*	*ARP*/*wARP*: residues built in first cycle/ninth cycle
Hg3a180	94.4 (89.4)	6	0.37	0.48/0.75	55/99
Hg3a160	93.6 (86.3)	6	0.37	0.49/0.79	89/165
Hg3a140	89.5 (82.6)	6	0.36	0.54/0.85	263/502
Hg3a120	80.4 (74.5)	6	0.33	0.51/0.83	122/475
Hg3a100	69.2 (63.9)	7	0.31	0.49/0.81	110/460
Hg3a80	58.2 (54.6)	7	0.26	0.41/0.63	25/66

Banumathi *et al.*, 2004). In these cases, even after prolonged exposures the specific radiation damage of these sites was not complete. Banumathi *et al.* (2004) observed continuous and steady deterioration of the overall data quality as data collection progressed, where the most affected disulfide bridges lost only about 20–30% of their occupancy. The Hg—S bonds appear to be significantly more susceptible to radiation damage than S—S bonds or any other bonds present in native proteins. The high vulnerability of the S—Hg bonds to X-ray radiation damage is analogous to the effect displayed by the Br atoms covalently bound to aromatic bases in nucleotides (Ennifar *et al.*, 2002).

It is possible that the observed heavy-atom-specific decay is in part a consequence of the relatively high concentration of strongly absorbing mercury ions within the disordered solvent regions that resulted from the omission of a back-soaking step. However, it should be noted that the presence of mercury in solvent channels would also be expected to increase the global decay. Further systematic studies will be required to understand whether back-soaking will increase or decrease the ratio of specific/global decay.

3.4. RIP and data redundancy

From a third derivatized crystal prepared similarly as described for Hg2, 360° of data were collected to a resolution of 2.0 Å at λ = 1.254 Å, about 2000 eV less than the mercury L_{III} edge, so as to minimize the anomalous signal and allow phasing attempts based solely on the radiation-induced-damage (RIP) effects. These data were split into wedges spanning 0–N° (N = 80, 100, 120, 140, 160 and 180; data sets denoted as Hg3a80, Hg3a100, ..., Hg3a180) and 181–360° (Hg3b). Since scaling in space group $P1$ is inefficient owing to the scarcity of equivalent reflections within a small wedge of data, the scale factors derived from merging the Hg3a180 data were used to merge the Hg3a80 and Hg3a100 sets. Six individual RIP phasing attempts were performed, treating the first six 0–N° sets as derivative data, with the final Hg3b data serving as native. The program *SOLVE* was used with the standard SIR protocol, followed by density modification with *DM*. In this particular case, the phases resulting from density modification with *DM* were superior to those from *RESOLVE*. Good-quality phases were obtained with the Hg3a140 as well as with Hg3a120 and Hg3a100 sets (see Table 6 and Fig. 4). This is corroborated by the results of the automatic model building by *ARP*/*wARP* (Perrakis *et al.*, 1999) with the use of the Hg3b amplitudes and phases obtained from all Hg3aN data sets. About 80% of the protein chain was built in ten cycles on the basis of the Hg3a140, Hg3a120 and Hg3a100 sets, but in the first cycle the Hg3a140 set appeared to give the best result, with about 50% of the

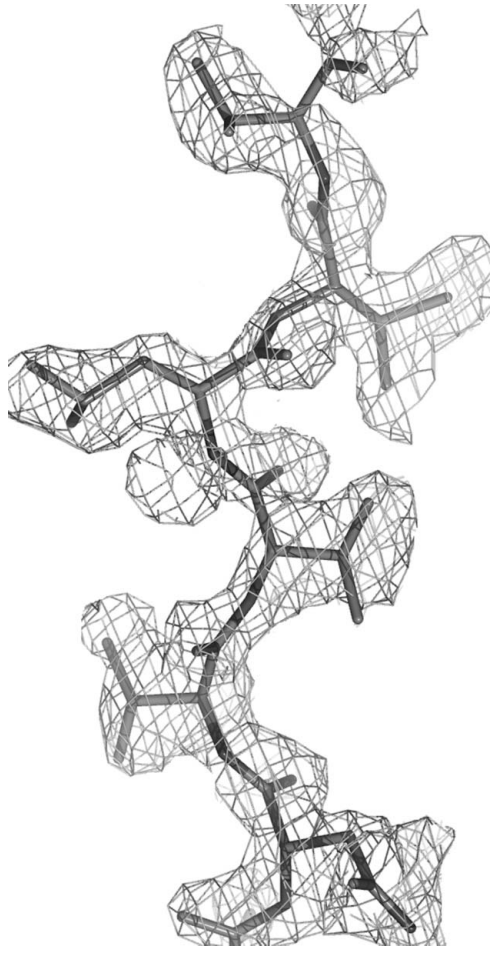

Figure 4
The density-modified map for yidA (2 Å resolution) contoured at the 1σ level after RIP phasing using data Hg3a140 as derivative and Hg3b as native.

main chain being successfully built. The more complete phase sets Hg3a180 and Hg3a160 led to poorer results, as did the less complete Hg3a80 set.

These observed results of phasing and model building can be explained as follows. The increased redundancy (and completeness), which normally enhances data quality, in this case lead to an increase in data non-uniformity resulting from progressive radiation damage. The Hg3a100–140 sets are not fully complete, but the isomorphic variation caused by radiation damage did not significantly degrade the overall data quality. The Hg3a180 and Hg3b data sets are more complete but less uniform owing to the high level of damage associated with longer accumulated exposure. Most importantly, the radiation-induced differences between the data sets Hg3a100 and Hg3b (or between Hg3a120 and Hg3b) are more pronounced than those calculated with Hg3a180 and Hg3b (when there is no gap between the data sets). These observations suggest that careful adjustment of the exposure time (*i.e.* the absorbed X-ray dose) separating data sets can be an important consideration for the execution of successful phasing experiments by RIP, as observed by Ravelli *et al.* (2003).

3.5. The absorbed radiation dose and specific decay

The flux measured in 2003 for beamline X9A through a 0.2 × 0.2 mm aperture (NSLS X-ray ring running at 2.8 GeV, 200 mA) at a wavelength of 1.0 Å was $\sim 2.4 \times 10^{10}$ photons s^{-1} (http://www.px.nsls.bnl.gov/nsls_px_main.html). As these experiments were performed at almost the same time (PDB submission date for 1rkq is 23 November 2003), we have used these values in the dose calculation. The measured solubility of mercury acetate in the mother liquor was found to be approximately 10 mg ml^{-1} (∼35–40 m*M*). Since the Hg2 and Hg3 crystals were prepared by adding 1.0 µl mercury acetate-saturated solution to ∼4.0 µl drops containing crystals, we estimate a maximum Hg concentration of 10 m*M* for the dose calculations. The crystals used in these experiments were approximately 0.1 × 0.1 × 0.15 mm and a slit size of 0.2 × 0.2 mm was used. The *RADDOSE* (Murray *et al.*, 2005) program was used to calculate the total absorbed dose in each experiment. The total absorbed dose for the Hg2 crystal (with ten S atoms and two Hg atoms per monomer) calculated over all eight data sets is approximately 0.17×10^8 Gy (1 Gy = 1 J kg^{-1}), which is very near the dose limit of 0.2×10^8 Gy described by Henderson, but is only 0.41×10^7 Gy for the first two data sets (Hg2aL and Hg2aH) when the majority of Hg—S bonds were disrupted. Similarly, the total absorbed dose for the Hg3 data set (all 360°) is about 0.33×10^7 Gy. These values confirm that the rupture of susceptible Hg—S bonds has its effect on the data much earlier than the detrimental global effects become visible. Also, the observed difference between the decay of Hg atoms bound to Cys67 and Cys141 within the same molecule suggest that it is difficult to correlate the specific decay of Hg atoms based solely on the absorbed dose.

Table 7
Map correlation and phase error from different phasing approaches.

In the case of yidA the Hg2aL and Hg2bL data sets and in the case of yjcF the Hg4a and Hg4b data sets were used in the phasing attempts.

	yjcF (Hg4a and Hg4b data sets)		yidA (Hg2aL and Hg2bL data sets)	
	MapCC before/ after *DM*	Phase error (°) before/ after *DM*	MapCC before/ after *DM*	Phase error (°) before/ after *DM*
RIP	0.47/0.55	73.0/67.5	0.58/72	68.9/55.5
SAD	0.49/0.56	70.5/66.2	0.67/82	63.4/47.3
RIPAS	0.55/0.73	65.5/56.6	0.74/85	56.9/44.4
SAD + RIP†	0.62/0.79	61.9/50.6	0.79/88	52.8/39.9
Pseudo-MAD‡	0.62/0.81	61.2/50.5	0.74/85	57.6/44.0

† Phases from SAD and RIP attempts were combined using *SIGMAA* (Collaborative Computational Project, Number 4, 1994). ‡ Pseudo-MAD: treating first data as peak data and second as inflection data with an f' value of −50 e.

3.6. Second example

Another NYSGRC target protein, *B. subtilis* yjcF, a member of the GCN5-related acetyl transferase superfamily (PDB code 1q2y) was originally solved by MIR utilizing mercury as one of the derivatives. This protein has a single cysteine in 140 residues. A retrospective analysis revealed that this Hg derivative exhibited similar behavior to that described above for yidA and led to successful phasing based on the Hg data alone by the RIPAS approach. The crystallization conditions (32% PEG 550, 0.05 *M* diammonium tartrate, 0.1 *M* Bis-Tris pH 6.5) were entirely different from that of yidA and notably the reagent used for derivatization was the organomercurial ethylmercury phosphate, in contrast to the inorganic Hg^{2+} salt used for yidA. The crystal used in this study belongs to space group $P2_12_12_1$ and the derivatized crystals diffracted to 2.9 Å; peak data sets collected in two wedges covering 0–150° and 180–330° were available. These data sets were merged separately using the program *HKL*2000 and the first and second wedge were treated as derivative (Hg4a) and native (Hg4b), respectively. The best E_{calc}/E_{obs} correlation coefficients (CC) for substructure solutions produced by *SHELXD* using SAD, RIP and RIPAS approaches are 28.2/18.8, 23.8/10.8 and 32.9/25.7%, respectively. These results reiterate the influence of radiation damage on the data obtained from Hg-derivatized crystals. Remarkably, the success rate of the substructure solutions was 100% in all three cases. However, both SAD and RIP maps obtained after phasing by the programs *SHELXE* and *SOLVE* were of poor quality and only the RIPAS map was clearly interpretable (Table 7). Based on these observations, it would be valuable to re-evaluate past and existing problematic cases associated with Hg phasing in the light of the potential impact of site-specific radiation damage.

3.7. Comparison of phases from different approaches

To be able to judge the quality of phases obtained with different approaches, the structures (PDB code 1rkq for target yidA and 1q2y for target yjcF) were refined against selected derivative data sets. The substructure solution from the SAD

data was input to *SOLVE* in 'analyze_solve' mode to re-refine the occupancies in SAD, SIR and SIRAS protocols and to keep the same origin. The map correlations listed in Table 7 clearly demonstrate that inclusion of radiation-induced decay is very important in both mercury-derivatized cases, although good-quality phases can be obtained by SAD alone in the case of Hg2aL data, which was of superior quality owing to the short exposure times (*i.e.* 10 s per degree). As mentioned earlier, the structure can be solved by the SAD method using the program *SHARP* as well as *SOLVE*, although there is considerable amount of heavy-atom decay within the data set.

We also tested the possibility of combining the phases from RIP and the SAD approach prior to density modification. In general, both SAD and classical SIR phases suffer from phase ambiguity and hence SIRAS phases are expected to be of better quality, as the density-modification process is heavily reliant on the quality of the initial phases. However, it was interesting to note that the SAD + RIP phases after phase combination and density modification were superior to RIPAS density-modified phases in both cases (Table 7).

If all the heavy-atom decay is assumed to be uniform, it can be approximated that the decay of the heavy atom mimics the dispersive contribution. The case of yjcF, with only one mercury site, almost mimics the situation expected for MAD, except that the effect arising from reduced occupancy is very large compared with the dispersive difference between the peak and inflection data sets. Pseudo-MAD attempts were performed considering the set collected earlier as peak data and the later set as inflection data for both cases, yidA (Hg2aL and Hg2bL) and yjcF (Hg4a and Hg4b), using the program *SOLVE*. It was realised that the phasing results are sensitive to the value of f' and larger (as high as -40 to -80 electrons) values assumed in calculations produced better phases. It can be seen that the pseudo-MAD phases were as good as RIPAS phases both in the case of yjcF and yidA (Table 7). This might be a possible reason for many successful MAD experiments with mercury in which the heavy-atom decay was utilized together with dispersive differences without the knowledge of the experimenter. These results suggest that there are many ways to derive quality phases from data that suffer from specific damage if the variation in the first data set is within the limit of tolerance. Therefore, for mercury-derivatized crystals it is essential that the initial data be collected at a wavelength that provides a significant anomalous signal and with relatively low exposure.

3.8. Real-time identification of heavy-atom-specific decay

Fig. 5 shows a plot of χ^2 *versus* image number, as calculated by the merging program *SCALEPACK* (from the *HKL*2000 suite), for the native (NatS) and for the two derivative data sets. In the presence of radiation damage, reflections measured halfway through the experiment would be predicted to be most similar to the overall merged values, while those measured at the end or beginning of the experiment would show the largest deviation form the merged values (depending on the sensitivity of the crystal to X-ray damage and the length of exposure). The variation of χ^2 within the first 360° of data from the native crystal of yidA (Fig. 5*a*) indicates that the decay over this interval is minimal. In contrast, when all 1666° of data are merged, the χ^2 plot takes a concave form (Fig. 5*b*), suggesting significant decay owing to radiation damage over the prolonged exposure (Otwinowski, 2003; Zwart *et al.*, 2004). This is most probably the main cause for the failure of the S-SAD phasing attempts with this data set.

The χ^2 plot for the mercury-derivatized crystal of both yidA and yjcF (Figs. 5*c* and 5*d*) shows entirely different behavior, with large variations in intensities at the beginning of the experiment which plateau after a relatively short time. This analysis is consistent with our assertion that mercury-specific decay occurs well before any global decay is observed. From the above observations it appears that the specific decay of mercury-derivatized crystals can be readily identified at the time of data collection/processing, which might provide valuable criteria for selecting a particular phasing approach. However, it must be kept in mind that when only relatively small wedges of data (less than 180°) are available, the effect owing to intrinsic quality of the crystal and hence the diffraction in different crystal orientations should be taken into account, as these factors also contribute to variations in χ^2. Furthermore, at third-generation synchrotron sources it may be necessary to highly attenuate the direct beam in order to observe such variations in χ^2, otherwise one might possibly lose any information associated with mercury derivatization within the initial (10–20) frames.

4. Conclusions

In general unsuccessful experiments are infrequently reported; however, both anecdotal and some published reports suggest that the use of mercury derivatives is often troublesome in practice, despite their widespread popularity (Boggon & Shapiro, 2000). According to Ji *et al.* (2001) about 50% of HgMAD phasing attempts fail and require additional derivative data for successful structure solution. The underlying basis of these difficulties is rarely examined and in cases where they are analyzed, the interpretations are often very general, such as the lack of high-quality dispersive or anomalous signal (*e.g.* Gulbis *et al.*, 2004). Of particular relevance is the comparative study of phasing the AP2 complex with Xe, Hg and Se, which show that the Hg-derived phases were significantly inferior to those obtained with other derivatives and in this case inclusion of data from a second wavelength in a HgMAD phasing was detrimental, probably because of radiation damage (Evans, 2003).

In the current study, phasing statistics as well as direct examination of difference Fourier maps clearly demonstrate that Hg atoms covalently bound to cysteine sulfhydryl groups are highly susceptible to radiation damage, manifested by the disruption of the Hg—S bonds and rapid decline of the occupancy of the original Hg sites. The data used in this report were measured at a second-generation bending-magnet beamline employing exposure times that were two to five times shorter than routinely utilized at the same beamline. The

X-ray flux delivered by undulator sources at the third-generation synchrotrons can be two to three orders of magnitude higher; thus, owing to the high vulnerability of Hg atoms covalently bound to cysteines, our observations may have widespread implications. These mercury-specific effects are particularly detrimental for SAD and MAD phasing experiments, where decay within and between data sets degrades the quality of the experimental phases. However, since the Hg atom possesses 80 electrons, the isomorphous intensity changes induced by radiation damage of Hg-derivatized crystals may be very effectively used for phasing. The results obtained for yidA (and the additional support provided by yjcF) suggests the robustness of this approach, as it presents a rather unfavorable case with the lowest possible crystal symmetry and consequently very low redundancy of measurements. Nevertheless, the structure of yidA could be solved based exclusively on the radiation-damage-induced signal present in the diffraction data measured with very short exposures. These observations suggest specific data-collection strategies for Hg-derivatized crystals. An optimized strategy would initially focus on obtaining the highest quality anomalous data by collecting the smallest sector(s) of data at the peak energy that is consistent with the required accuracy; these data should be collected as rapidly as possible (with low exposure times/low dose) and if needed, with attenuation, especially at third-generation sources. Subsequent data, perhaps after considerable X-ray exposure to fully reduce the Hg occupancies, would represent native data and should be collected to the diffraction limit of the crystal for eventual refinement.

It is interesting to note that the vulnerability of mercury derivatives to radiation damage was observed by the team of

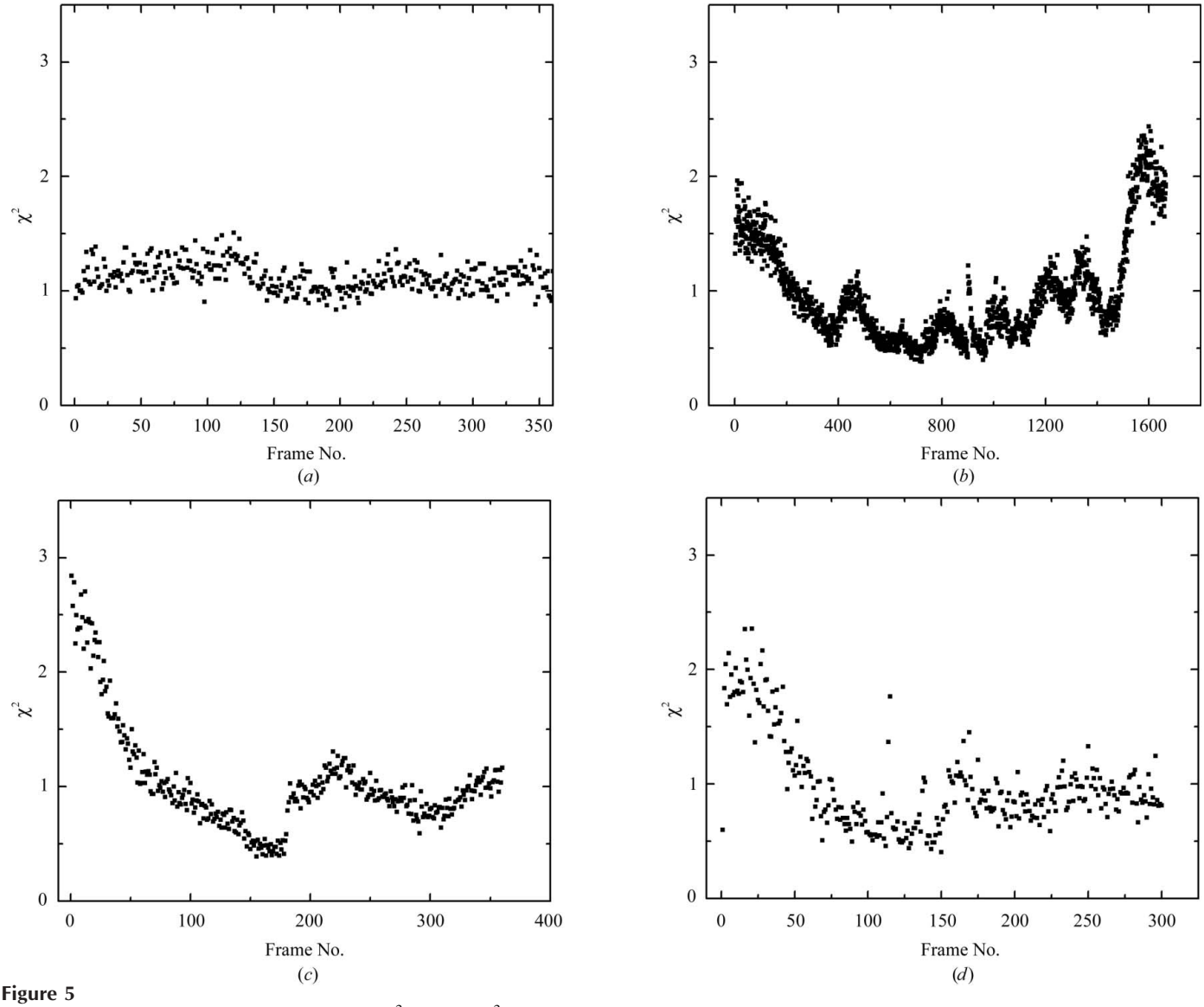

Figure 5
Real-time identification of mercury decay using χ^2 plots. (*a*) χ^2 as a function of frame number for the first 360° data of native crystals of yidA collected for sulfur phasing (NatS) suggests that decay over this interval is minimal. (*b*) The χ^2 plot takes a concave form when all 1666° of data are merged, suggesting a significant decay owing to radiation damage over the prolonged accumulated exposure. (*c*, *d*) χ^2 plots for the mercury-derivatized crystals of both yidA (Hg3 derivative) and yjcF show large variation in intensities at the beginning of the experiment, which plateaus after a relatively short time.

David Phillips during their early work on the structure of lysozyme at 6 Å resolution (Blake *et al.*, 2001, p. 754), but this effect has not been investigated further. Mercury seems to behave like the famous Cheshire cat (Carroll, 1865); it disappears but leaves a grin wide enough for successful usage (quite appropriate, since Lewis Carroll was born in Daresbury, at the future UK synchrotron site). The observations presented here provide an explanation for the widespread difficulties experienced with Hg derivatives and provides a straightforward experimental remedy.

This work is a contribution from the New York Structural GenomiX Research Consortium (NYSGXRC) and was supported by NIH Structural Genomics Pilot Center Grant 1P50 GM62529. We appreciate the access to beamline X9A at the National Synchrotron source. We would like to thank Dr Raimond Ravelli and Dr Elspeth Garman for the *RADDOSE* program.

References

Banumathi, S., Zwart, P. H., Ramagopal, U. A., Dauter, M. & Dauter, Z. (2004). *Acta Cryst.* D**60**, 1085–1093.
Blake, C. C. F., Fenn, R. H., Johnson, L. N., Koenig, D. F., Mair, G. A., North, A. C. T., Oldham, J. W. H., Phillips, D. C., Poljak, R. J., Sarma, V. R. & Vernon, C. A. (2001). *International Tables for Crystallography*, Vol. *F*, edited by M. G. Rossmann & E. Arnold, pp. 745–772. Dordrecht: Kluwer Academic Publishers.
Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography*. New York: Academic Press.
Boggon, T. J. & Shapiro, L. (2000). *Structure*, **8**, R143–R149.
Burmeister, W. P. (2000). *Acta Cryst.* D**56**, 328–341.
Carroll, L. (1865). *Alice's Adventures in Wonderland*, ch. 6. London: Macmillan & Co.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Dauter, Z. & Adamiak, D. A. (2001). *Acta Cryst.* D**57**, 990–995.
Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.
Dauter, Z., Dauter M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (2001). *J. Mol. Biol.* **289**, 83–92.
Debreczeni, J. É., Bunkóczi, G., Girmann, B. & Sheldrick, G. M. (2003). *Acta Cryst.* D**59**, 393–396.
Debreczeni, J. É., Bunkóczi, G., Ma, Q., Blaser, H. & Sheldrick, G. M. (2003). *Acta Cryst.* D**59**, 688–696.
Debreczeni, J. É., Girmann, B., Zeeck, A., Kratzner, R. & Sheldrick, G. M. (2003). *Acta Cryst.* D**59**, 2125–2132.
Diederichs, K., McSweeny, S. & Ravelli, R. B. G. (2003). *Acta Cryst.* D**59**, 903–909.
Djinovic Carugo, K., Helliwell, J. R., Stuhrmann, H. & Weiss, M. S. (2005). *J. Synchrotron Rad.* **12**, 410–419.
Doan, D. N. & Dokland, T. (2003). *Structure*, **11**, 1445–1451.
Ennifar, E., Carpentier, P., Ferrer, J.-L., Walter, P. & Dumas, P. (2002). *Acta Cryst.* D**58**, 1262–1268.
Evans, P. (2003). *Acta Cryst.* D**59**, 2039–2043.
Garman, E. & Murray, J. W. (2003). *Acta Cryst.* D**59**, 1903–1913.
Garman, E. & Nave, C. (2002). *J. Synchrotron Rad.* **9**, 327–328.
Gonzalez, A. & Nave, C. (1994). *Acta Cryst.* D**50**, 874–877.
Gulbis, J. M., Kazmirski, S. L., Finkelstein, J., Kelman, Z., O'Donnell, M. & Kuriyan, J. (2004). *Eur. J. Biochem.* **271**, 439–449.
Henderson, R. (1990). *Proc. R. Soc. London Ser. B*, **241**, 6–8.
Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.
Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
Ji, X., Blaszczyk, J. & Chen, X. (2001). *Acta Cryst.* D**57**, 1003–1007.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Leiros, H.-K. S., McSweeney, S. M. & Smalås, A. O. (2001). *Acta Cryst.* D**57**, 488–497.
Leonard, G. A., Sainz, G., de Backer, M. M. E. & McSweeney, S. (2005). *Acta Cryst.* D**61**, 388–396.
Liu, Z.-J., Vysotski, E. S., Chen, C. J., Rose, J. P., Lee, J. & Wang, B.-C. (2000). *Protein Sci.* **9**, 2085–2093.
Murray, J. & Garman, E. (2002). *J. Synchrotron Rad.* **9**, 347–354.
Murray, J. W., Rudiño-Piñera, E., Owen, R. L., Grininger, M., Ravelli, R. B. G. & Garman, E. (2005). *J. Synchrotron Rad.* **12**, 268–275.
Olsen, J. G., Flensburg, C., Olsen, O., Bricogne, G. & Henriksen, A. (2004). *Acta Cryst.* D**60**, 250–255.
O'Neill, P., Stevens, D. L. & Garman, E. F. (2002). *J. Synchrotron Rad.* **9**, 329–332.
Otwinowski, Z. (1991). *Proceedings of the CCP4 Study Weekend. Isomorphous Replacement and Anomalous Scattering*, edited by W. Wolf, P. R. Evans & A. G. W. Leslie, pp. 80–86. Warrington: Daresbury Laboratory.
Otwinowski, Z. (2003). Private communication.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Perrakis, A., Morris, R. & Lamzin, V. S. (1997). *Nature Struct. Biol.* **6**, 458–463.
Ramagopal, U. A, Dauter, M. & Dauter, Z. (2003). *Acta Cryst.* D**59**, 1020–1027.
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*, **11**, 217–224.
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. (2002). *Z. Kristallogr.* **217**, 644–650.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Sun, P. D. & Radaev, S. (2002). *Acta Cryst.* D**58**, 1099–1103.
Szyk, A., Lu, W., Xu, C. & Lubkowski, J. (2004). *J. Struct. Biol.* **145**, 289–294.
Teng, T & Moffat, K. (2000). *J. Synchrotron Rad.* **7**, 313–317.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
Weiss, M. S., Mander, G., Hedderich, R., Diederichs, K., Ermler, U. & Warkentin, E. (2004). *Acta Cryst.* D**60**, 686–695.
Yang, C., Plfugrath, P. W., Courville, D. A., Stence, C. N. & Ferrara, J. D. (2003). *Acta Cryst.* D**59**, 1943–1957.
Zwart, P. H., Banumathi, S., Dauter, M. & Dauter Z. (2004). *Acta Cryst.* D**60**, 1958–1963.