Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Phasing power at the *K* absorption edge of organic arsenic

Pascal Retailleau[a]* and Thierry Prangé[a,b]

[a]Laboratoire LURE, Centre Universitaire Paris-Sud, Bâtiment 209D, BP 34, 91898 Orsay CEDEX, France, and [b]Laboratoire de Cristallographie et RMN Biologiques (UMR 8015), Faculté de Pharmacie/Université René Descartes-Paris V, 4 Avenue de l'Observatoire, 75270 Paris CEDEX 06, France

Correspondence e-mail: retailleau@lure.u-psud.fr

Received 27 November 2002
Accepted 17 February 2003

PDB Reference: hen egg-white lysozyme, arsenic derivative, 1n4f, r1n4fsf.

Single/multiple-wavelength anomalous dispersion (SAD/MAD) experiments were performed on a crystal of an organic arsenic derivative of hen egg-white lysozyme. A *para*-arsanilate compound used as a crystallizing reagent was incorporated into the ordered solvent region of the lysozyme molecule. Diffraction data were collected to high resolution ($\leq$2.0 Å) at three wavelengths around the *K* edge (1.04 Å) of arsenic at beamline BM30A, ESRF synchrotron. Anomalous Patterson maps clearly showed the main arsanilate site to be between three symmetry-related lysozyme molecules, at a location previously occupied by a *para*-toluenesulfonate anion. MAD phases at 2 Å derived using the program *SHARP* led to an electron-density map of sufficient quality to start manual building of the protein model. Amplitudes from a second crystal measured to a resolution of 1.8 Å at the peak wavelength revealed two additional heavy-atom sites, which reinforced the anomalous subset model and therefore dramatically improved the phasing power of the arsenic derivative. The subsequent solvent-flattened map was of such high accuracy that the program *ARP*/*wARP* was able to build a nearly complete model automatically. This work emphasizes the great potential of arsenic for *de novo* structure determination using anomalous dispersion methods.

1. Introduction

Advances in diffraction-intensity recording and maximum-likelihood-based phasing techniques continue to lower the anomalous signal contribution required for use in crystallographic phasing (Brodersen *et al.*, 2000). In the case of wild-type macromolecular crystals, internal sources of anomalous scattering at wavelengths used for data collection are restricted to S atoms in proteins (Hendrickson & Teeter, 1981; Dauter *et al.*, 1999; Micossi *et al.*, 2002) or P atoms in the case of nucleic acids (Dauter & Adamiak, 2001), without taking into account the case of metalloproteins. However, this may suffice to solve the phase problem in a way that would lead to the crystallographic Grail: to readily turn measured intensities from unmodified crystals of wild-type proteins into an electron-density map.

To date, the power of anomalous scattering has been fully and readily exploited in MAD experiments when the wavelength can be tuned around the resonant peak absorption and can also provide an additional dispersive signal (Hendrickson *et al.*, 1985; Burling *et al.*, 1996). In this case, engineered substitution of methionine S atoms by selenium (Hendrickson *et al.*, 1990; Doublié, 1997) or substitution of bromouracyl/bromocytosin for nucleic acids (Peterson *et al.*, 1996; Shepard *et al.*, 1998) has exemplified the evident potential of MAD for

de novo structure determination (Hendrickson, 1999; Walsh, Dementieva *et al.*, 1999; Walsh, Evans *et al.*, 1999). Indeed, selenomethionylated proteins closely resemble the native owing to the relatively similar sizes of Se and S, although their solubility may be affected. Well ordered selenium sites together with their large number (methionines naturally occur about every 69 residues) and successful (>90%) substitution compensate for their relatively weak individual anomalous potential and they can be sufficient for the phasing of a large protein structure (Deacon & Ealick, 1999). Furthermore, their *ab initio* location is rendered possible with direct methods (Blessing & Smith, 1999; Schneider & Sheldrick, 2002). Finally, the Se *K* edge features a sharp resonance that can be enhanced under oxidized conditions (Smith & Thompson, 1998; Sharff *et al.*, 2000) at 0.98 Å and is quite suitable for high-resolution measurement. All these advantages spurred the worldwide construction of high-brilliance MAD-dedicated synchrotron beamlines at energies around the Se *K* edge. As a direct consequence, utilization of the SeMet tool accounts for more than a third of the skyrocketing number of anomalous dispersion structure deposits. Nevertheless, to address cases of lack of methionines, of biochemical skills, of synchrotron beamtime and so on, it is appreciable that other scatterers may offer alternatives. For instance, it is appropriate to study crystals of metalloproteins that carry endogenous transition metals such as zinc (Hosfield *et al.*, 1999), copper (Guss *et al.*, 1996; Walter *et al.*, 1996) or iron (Iwata *et al.*, 1996), all of which have *K* edges that can resonate even with a rotating-anode laboratory source. The binding of counterions to protein surfaces has also been demonstrated in numerous high-resolution structures or from anomalous data (Badger *et al.*, 1994). Their inclusion during an appropriate short soaking step, for example bromine (Dauter & Dauter, 1999; Dauter *et al.*, 2000) or rubidium (Korolev *et al.*, 2001), provides an easy method for the incorporation of anomalous scatterers. Another promising 'almost ready-to-go' technique is the pressurization of protein crystal in an atmosphere of a noble gas such as Xe or Kr (Schiltz *et al.*, 1994; Schiltz, Kvick *et al.*, 1997; Schiltz, Shepard *et al.*, 1997; Cohen *et al.*, 2001). Many further candidates (lanthanides) are regularly reported as successful compounds for derivatizing crystals (for reviews, see Fourme *et al.*, 1996; Girard *et al.*, 2002).

Of all these potential anomalous scattering candidates, arsenic surprisingly appears to have been underrated despite its objectively numerous adequate properties, probably owing to its notorious toxicity (Bentley & Chasteen, 2002). Historical poisoners, with Nero at their head, provided the first evidence that arsenic binds to blood or organ proteins with strong efficiency. Furthermore, with 33 electrons, arsenic just precedes selenium in the periodic classification and is thus likely to feature a white line for its *K*-edge absorption at a slightly higher wavelength than selenium (∼1.04 Å compared with ∼0.98 Å), but which should still be manageable for Se-dedicated beamlines.

Arsenic is an ubiquitous element in nature, existing in various forms, both organic or inorganic, such as minerals in the soil, arsine (AsH_3) in air and salts of arsenate or arsenite in water. Its inevitable presence in living cells is metabolized by enzymes such as arsenate reductase (Martin *et al.*, 2001; Messens *et al.*, 2002) or ArsA ATPase (Zhou *et al.*, 2001) that help in detoxification. A particularly widespread form among the biochemical community is dimethylated arsenic acid, otherwise known as cacodylic acid, which is largely exploited for its neutral pK_a in buffering protein solutions. Cacodylate anions in the crystallizing medium are thus the main opportunistic source of arsenic reported in macromolecular structures. About 70 deposited macromolecular coordinate files (including redundant structures; from the PDB, July 2002) include As atoms, the presence of which has been confirmed by anomalous Fourier map calculations in some cases. However, in contrast, only one crystal structure to date, that of HIV integrase, has been solved by MAD at the As *K* edge (Greenwald *et al.*, 1999). Choe's group took advantage of the reactivity of cacodylate anions under reducing conditions with accessible sulfhydryl groups to form a cysteine adduct as reported by Maignan *et al.* (1998) and first showed the feasibility of exploiting the anomalous property of arsenic for phasing when introduced into frozen crystals.

In the present work, following Choe and coworkers, we establish the potential benefits of arsenic derivatization as a phasing tool that can be readily exploited on synchrotron beamlines traditionally tuned around the Se *K* edge. Here, arsenic was introduced into the solvent-ordered part of tetragonal lysozyme crystals from a crystallizing solution containing 0.2 *M* *para*-arsanilate (pTAs) salt. This salt was chosen in order to guarantee strong substitution and well ordered site(s) in the crystal packing and also to provide an oxidized form of arsenic, which is likely to have a more intense white line. In this salt, arsenic substitutes for the SO_3 group of *para*-toluenesulfonate (pTS), which is known to be an effective crystallizing reagent for lysozyme (Vaney *et al.*, 2001) and binds tightly to three symmetry-related molecules in the crystal packing. The optimized anomalous signal at the As *K* edge appeared to be sufficient to locate other crystallizing anions at the surface of the lysozyme and then to phase the protein structure successfully with full automation.

2. Materials and methods

2.1. Crystallization

We used thrice-crystallized hen egg-white lysozyme chloride commercially available from Sigma (Cat. No. L-2879), which was characterized as containing less than 0.3%(*w*/*w*) contaminant (Riès-Kautt & Ducruix, 1997). Lysozyme was first dialyzed and then deionized according to the above-mentioned work and was stored freeze-dried. The isoionic protein powder was then dissolved to 100 mg ml^{-1} in 100 m*M* cacodylate pH 6.5. Note that diffraction of lysozyme crystals grown in 1 *M* cacodylate, used as buffer and crystallizing agent, did not reveal the presence of cacodylate ions bound to the lysozyme surface in any type of calculated difference Fourier maps. Hanging-drop vapour diffusion was used to crystallize lysozyme in a drop equilibrated against

250 m*M* sodium *para*-arsanilate (also known as 4-aminophenylarsonic acid; ACROS) solution neutralized with NaOH. Crystals were further quickly dipped into a 250 m*M* salt solution containing 30%(*w*/*v*) glycerol (in the case of crystal 1) or 15%(*w*/*v*) ethyleneglycol (in the case of crystal 2) prior to flash-cooling in a 100 K nitrogen stream. We also managed to co-crystallize lysozyme with the *ortho* form of arsanilate (2-aminophenylarsonic acid; ACROS) following a similar protocol.

2.2. X-ray data collection and data processing

X-ray data recordings were carried out at ESRF on the anomalous dispersion dedicated beamline BM30A using a MAR CCD (165 mm) detector. A fluorescence spectrum was measured directly from lysozyme crystal 1 (Fig. 1). Note the white line of the oxidized arsenic *K* absorption edge in lysozyme and the shift of the absorption at the *K* edge toward lower wavelengths compared with that of the reduced form in the HIV integrase crystal (Greenwald *et al.*, 1999): 1.0438 Å compared with 1.0458 Å. Selection of the X-ray wavelengths for data collection was made on the basis of f' and f'' anomalous scattering factors in the XANES region, which were derived from the fluorescence spectrum using the program *CHOOCH* (Evans & Pettifer, 2001). Three wavelengths (one at the adsorption-edge peak for which $f' = -8$ and $f'' = 8.1$ e$^-$, the second at the inflection point for which $f' = -11.5$ and $f'' = 4.9$ e$^-$ and the third at a remote higher energy wavelength) were selected to carry out MAD experiments on both crystals. For comparison, f'' and f' are typically <7 and ~ -11 e$^-$ for fully reduced selenomethionylated crystals and $\sim$15 and ~ -16 e$^-$, respectively, when fully oxidized (Thomazeau *et al.*, 2001).

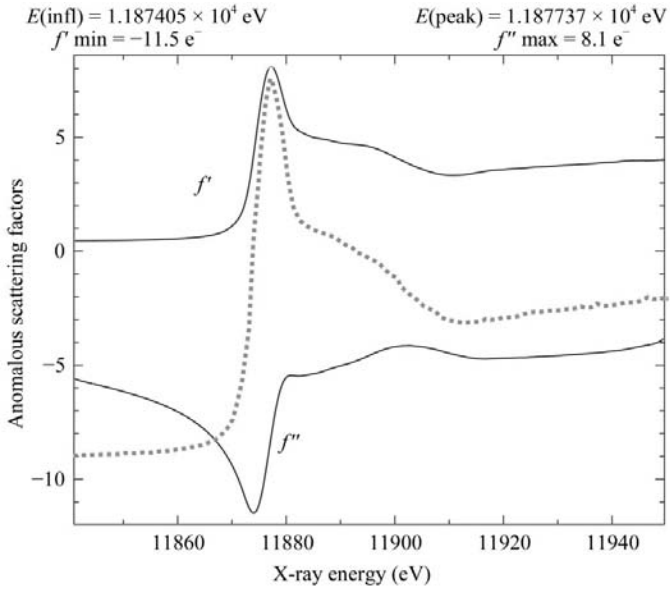

Figure 1
Dotted line: raw fluorescence spectrum from a *p*-arsanilated lysozyme crystal measured at the As *K* edge (arbitrary units). Solid line, experimental values (in electrons) of anomalous-scattering factors about the *K* absorption edge of As^V as output by *CHOOCH* (Evans & Pettifer, 2001).

Calculation of the theoretical dispersive and anomalous signals according to Hendrickson *et al.* (1985) indicates an anomalous ratio of 5.3–9.6% when the number of fully occupied As^V sites in lysozyme crystals is varied from 1 to 3 (the dispersive ratio is 2.3–4%).

All intensities were indexed and integrated using *MOSFLM* (Leslie, 1992). The scaling step was performed using a *SCALA* procedure suitable for MAD data (Evans, 1997): the data set at each wavelength was scaled in two steps against a reference set consisting of all-wavelength previously scaled and merged intensities, first *via* a batch scaling and then by local scaling across the detector. Data analysis is summarized in Table 1. Structure factors and anomalous differences were calculated using *TRUNCATE* (French & Wilson, 1978) and a final round of structure-factor scaling and analysis was performed using *SCALEIT* (Howell & Smith, 1992).

2.3. Determination of arsenic positions and phase determination

Bound arsenic was first investigated in anomalous difference Patterson maps using peak-wavelength amplitudes; the main site positions could be determined graphically and confirmed using the program *RANTAN* (Yao, 1981) as implemented in the *autoSHARP* package procedure (Vonrhein *et al.*, 2003). Coordinates were input into *SHARP* 1.4.0 (de La Fortelle & Bricogne, 1997) for heavy-atom refinement and maximum-likelihood-based phasing. Residual peak-wavelength anomalous maps output by *SHARP* were further investigated to locate other sites, whose presence was double-checked by calculating Bijvoet difference maps with phases from relevant refined models. Phase information was developed independently by *SHARP* by testing either the three-wavelength MAD protocol or the peak-wavelength SAD protocol. For the MAD procedure, it was recommended to input wavelengths in the same order as they were collected. For crystal 1, taking the remote-wavelength data set (because its scattering-factor values are quite well estimated and stable and also because of its better data statistics) rather than the peak-wavelength as the reference for estimating scale and lack of isomorphism parameters was checked and found to make little difference. For crystal 2, which was recorded at half the speed of crystal 1, the effects of radiation damage could not be underestimated. In this case, the SAD protocol was obviously the best refinement to start with. Additional wavelengths were gradually introduced to verify perturbation by non-isomorphism.

2.4. Phase improvements

The experimental phase probability distribution encoded by *SHARP* in the Hendrickson–Lattman (HL) coefficients (Hendrickson & Lattman, 1970) was improved by a density-modification procedure implemented into the *SHARP* interface which begins with a simple run of *DM* (solvent flattening and histogram matching; Cowtan, 1994) before switching to an iterative solvent-flipping procedure using *SOLOMON* (Collaborative Computational Project, Number 4, 1994;

Table 1
Statistics of data collection.

Values in parentheses refer to the last shell.

	Crystal 1 + glycerol			Crystal 2 + ethyleneglycol		
	Peak	Inflection	Remote	Peak	Inflection	Remote
Data collection, 100 K, BM30A						
Wavelength (Å)	1.0438	1.0442	1.0411	1.0438	1.0442	1.0411
Angular increment per frame (°)	1			1		
Total rotation range (°)	180			180		
Crystal-to-detector distance (mm)	120			120		
Exposure time per frame (s)	10			20		
Unit-cell parameter refinement						
Space group	$P4_32_12$					
Unit-cell parameters (Å)	a = 76.75, b = 76.75, c = 37.47			a = 76.76, b = 76.76, c = 38.25		
Data reduction						
No. of measured reflections	102554	105374	103830	155910	155304	155042
Resolution limits (Å)	15.6–2.0 (2.11–2.0)	15.6–2.0 (2.11–2.0)	15.6–2.0 (2.11–2.0)	15.7–1.78 (1.87–1.78)	15.4–1.78 (1.87–1.78)	15.4–1.77 (1.87–1.77)
No. of unique reflections	7979	7980	7983	11323	11322	11434
Data completeness (%)	99.8 (100)	99.8 (100)	99.8 (100)	99.4 (97.6)	99.4 (98.1)	99.2 (98.3)
$I/\sigma(I)$	7.6 (2.4)	8.3 (2.8)	8.0 (2.6)	8.5 (3.3)	9.1 (3.9)	8.6 (4.0)
Redundancy	12.8	13.2	13.0	13.7	13.7	13.5
$R_{\rm sym}$†	0.075 (0.310)	0.068 (0.273)	0.069 (0.278)	0.054 (0.165)	0.049 (0.146)	0.051 (0.140)
$R_{\rm anom}$‡	0.051 (0.109)	0.031 (0.083)	0.027 (0.080)	0.057 (0.120)	0.025 (0.075)	0.023 (0.071)
$R_{\rm r.i.m.}$§	0.093 (0.393)	0.076 (0.353)	0.076 (0.355)	0.079 (0.298)	0.056 (0.236)	0.057 (0.216)
$R_{\rm p.i.m.}$¶	0.026 (0.107)	0.021 (0.096)	0.021 (0.094)	0.021 (0.136)	0.015 (0.107)	0.015 (0.092)

† $R_{\rm sym}$ is defined as $\sum_{hkl}\sum_i |I_i(hkl) - \overline{I(hkl)}|/\sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith observation of reflection hkl and $\langle I(hkl)\rangle$ is the weighted mean of all observations (after rejection of outliers). ‡ $R_{\rm anom} = \sum_{hkl} |I(hkl) - \overline{I(hkl)}|/\sum_{hkl} \overline{I(hkl)}$. § $R_{\rm r.i.m.}$ is the redundancy-independent merging R factor (Weiss & Hilgenfeld, 1997), which is identical to the $R_{\rm meas}$ of Diederichs & Karplus (1997) defined as $\sum_{hkl}[N/(N-1)]^{1/2}\sum_i |I(hkl) - \overline{I(hkl)}|/\sum_{hkl}\sum_i I_i(hkl)$, with N being the number of times a given reflection has been observed. ¶ $R_{\rm p.i.m.} = \sum_{hkl}[1/(N-1)]^{1/2}\sum_i |I(hkl) - \overline{I(hkl)}|/\sum_{hkl}\sum_i I_i(hkl)$ is the precision-indicating merging R factor (Weiss & Hilgenfeld, 1997; Weiss, 2001).

Abrahams & Leslie, 1996). The solvent fraction, which is a sensible input parameter in addition to the solvent-flipping factor to determine the solvent envelope, does not have to be the exact solvent content (28% from *TRUNCATE*) of the tetragonal lysozyme crystal. In fact, it can be optimized according to a simple procedure proposed in *SHARP* from three density-modification runs with different input values. Electron-density maps were calculated with *FFT* (Collaborative Computational Project, Number 4, 1994) and displayed using the program *O* (Jones *et al.*, 1991) *via* the *SHARP* interface.

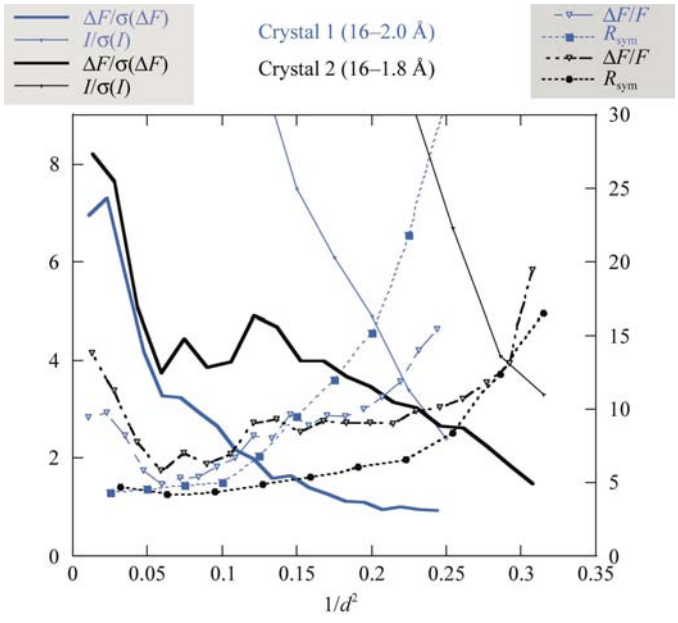

Figure 2
Resolution-dependent statistics of the anomalous signal and of quality indicators for two different data sets (blue for crystal 1 and black for crystal 2): $|\Delta F_{\rm anom}|/\sigma(\Delta F_{\rm anom})$ and $I/\sigma(I)$ on the left y axis are absolute numbers, whereas $R_{\rm sym}$ and $|\Delta F_{\rm anom}|/F$ are measured in percentages on the right y axis.

2.5. Refinement

The best solvent-modified maps for each case were subjected to *warpNtrace* (*ARP*/*wARP*; Perrakis *et al.*, 1999) called by the *SHARP* interface in order to automatically build the structure to the maximum resolution using the known primary sequence. With the current automatic implementation, the program task was restricted to 100 cycles and interpretation of the density/model coupled with *REFMAC* (Murshudov *et al.*, 1997) was permitted every ten cycles. Maximum-likelihood refinement of the structure model was performed using the program *BUSTER* 1.0 (Bricogne, 1993, 1997; Roversi *et al.*, 2000) interfaced with *TNT* (Tronrud *et al.*, 1987). Water and solvent molecules were subsequently included from inspection of Fourier differences maps and their atomic temperature factors and occupancy with respect to arsanilate anions were then refined. SAD or MAD phase information output by *SHARP* as the Hendrickson–Lattman coefficients constrained the partial structure refinement, whose stereochemical and geometrical restraints were handled in *TNT*. Final statistics of the model parameter refinements are given in Table 4.

3. Results and discussion

3.1. Phase-determination results

Fig. 2 plots some resolution-dependent statistics taken from Table 1 in order to assess the diffraction quality of both crystals subsequently used for distinct experiments. Although both crystals met the initial objective of collecting routine (2 Å) resolution data in a redundant fashion but without pre-orientating the crystal, they differed in quality. Crystal 2 clearly benefitted from better cryoprotection in 15%(*v*/*v*) ethyleneglycol than crystal 1, which was protected by glycerol, and yet displayed an obviously lower mosaicity. In contrast, strong reflections [$I/\sigma(I) > 2$] for crystal 1 did not exceed the 2 Å resolution sphere and the post-refined mosaicity in *MOSFLM* was estimated to be of the same order as the oscillation (0.81° compared with 0.39° for the ethyleneglycol-protected crystal).

The main concern, however, resided in obtaining an anomalous signal that could be used for phasing. The signal-to-noise level of the anomalous differences from crystal 2 appeared to be quite strong [overall $\Delta F/\sigma(\Delta F) \simeq 4$] and remained significant (>1.0) over the entire resolution range (20–1.78 Å). Furthermore, only for the highest resolution bin did the increased measurement noise assessed by the $R_{\rm sym}$ curve start to bury the anomalous contribution. Starting at a comparable level [$\Delta F/\sigma(\Delta F) > 7$], the anomalous signal for crystal 1 fades away quickly at medium resolution and levels off around the noise level in the 3–2 Å range. The reliability of this source of information therefore remains questionable above 2.6 Å. However, its overall level remains appreciable [$\langle \Delta F/\sigma(\Delta F) \rangle \simeq 2.6$] compared with the non-optimized differences obtained from four disulfide bonds at 1.5418 Å (∼1.8), which was still sufficient to phase the lysozyme structure (Dauter *et al.*, 1999).

Calculation of anomalous difference Patterson maps (Fig. 3) is also a pertinent indicator to judge the reliability of the anomalous signal and eventually to estimate the number of scatterers present in the structure. Based on the pTS molecule as observed in the lysozyme structure (Vaney *et al.*, 2001), one peak related to a high-occupancy anion site in the structure was expected in the Harker sections defined for the *P*4/*mmm* Patterson space group. Indeed, in both cases one common peak appeared at the 14σ (using data to 1.78 Å) and 10σ (using data to 2.0 Å) levels, respectively. The Patterson map derived from crystal 2 difference amplitudes unexpectedly revealed a second peak at 8σ. A third minute peak was confirmed in a residual anomalous map at the end of a *SHARP* refinement. Neither in a Bijvoet difference calculation nor in a final difference Fourier map could these secondary sites be retrieved in crystal 1. In contrast, a similar experiment conducted on a complex glycerol-cryo-preserved crystal with *ortho*-arsanilate at 1.004 Å on DW32 beamline at LURE showed the second binding site to be occupied.

Note that over the 5–2 Å range of both data sets, $\langle \Delta F/F \rangle$ takes a value of 9% which is close – and in reasonably good agreement with crystal 2 – to the theoretical value estimated according to Hendrickson & Teeter (1981).

Table 2
Refined heavy-atom parameters.

Parameters were calculated using *SHARP*. Parameters given in parentheses were calculated using *BUSTER*/*TNT*.

	MAD crystal 1, 16–2 Å	SAD crystal 2, 16–1.78 Å		
As sites	1	1	2	3
Occupancy	1.18 (0.84)	0.70	—	—
		0.59	0.37	—
		0.82 (0.92)	0.37 (0.62)	0.31 (0.49)
B factor	28.6 (44.6)	20.1	—	—
		19.7	21.8	—
		23.9 (30.0)	18.0 (36.3)	14.8 (38.9)
Scattering factors (e^-)				
f'	−8.8	−8 (not refined)		
f''	9.2	7.6	8.9	7.5

Table 3
Phasing statistics output by *SHARP*.

	Remote		Edge		Peak				
Wavelength, MAD 2.0 Å	Iso	Ano	Iso	Ano	Iso	Ano	Ano SAD 1.8 Å site No.		
							1	2	3
$R_{\rm Cullis}$†									
Centric	—	—	0.64	—	0.58	—			
Acentric	—	0.80	0.61	0.76	0.61	0.65	0.72	0.67	0.56
$R_{\rm Kraut}$‡									
Centric	0.071	0.141	0.113	0.143	0.083	0.186			
Acentric	0.018	0.035	0.029	0.036	0.021	0.047	0.050	0.052	0.043
Phasing power§									
Centric	—	—	0.91	—	0.88	—			
Acentric	—	2.12	1.59	2.41	1.39	2.79	1.64	2.00	2.82

† $R_{\rm Cullis} = \langle$phase-integrated lack of closure$\rangle / \langle |F_{PH} - F_P| \rangle$. ‡ $R_{\rm Kraut} = \langle$phase-integrated lack of closure$\rangle / \langle |F_{PH}| \rangle$. § Phasing power = $\langle [|F_H(\text{calc})|/$phase-integrated lack of closure$] \rangle$.

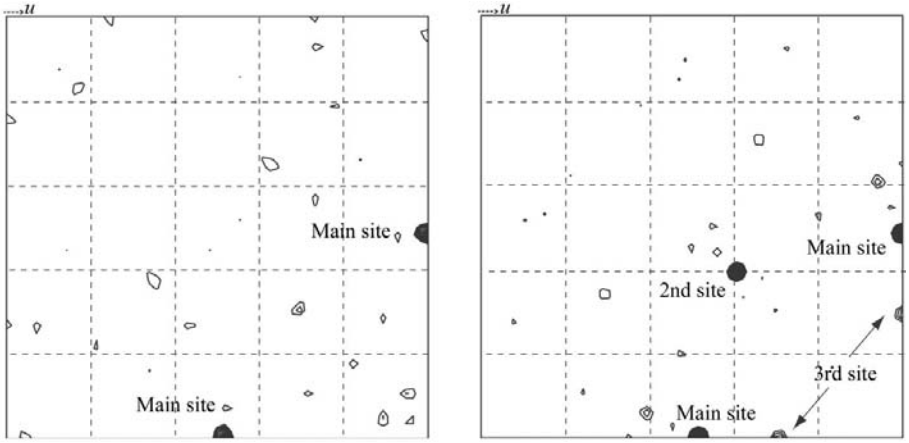

Figure 3
$\{w = 1/2, 0 \le u \le 1/2, 0 \le v \le 1/2\}$ Harker sections of anomalous difference Patterson maps for (*a*) crystal 1 and (*b*) crystal 2. Levels are contoured in 1σ steps starting at 2σ.

Table 2 lists the refined parameters of the heavy-atom subset on an absolute scale (one molecule of lysozyme per asymmetric unit) and scattering factors performed by *SHARP* with respect to the phasing protocol that led to the best result for each case. The ultimate criterion for judging the phasing quality was the advancement of the model autobuilding in the finest density-modified electron-density map as described below. The main site (*i.e.* highest peak in the specific Harker sections) is almost fully occupied, while the minor sites in crystal 2 have occupancy of about a half. Subsequent refinements of the entire structures confirmed the level of occupancy and increased the *B* values of the two other sites to twice that for the overall atomic model, testifying to their looser binding. Final phasing statistics depending on the number of sites subjected to refinement in *SHARP* are reported in Table 3. Fig. 4 illustrates the phasing power by showing the evolution of the map tracing for the best phasing protocols applied to both crystal data sets; Fig. 5 compares the experimental phases at successive stages during the phase-improvement process with those pertinent to the refined relevant structures.

In fact, using the SAD protocol for the refinement of a subset that included As 1 and 2 against data to 1.78 Å (crystal 2) provided the case which led, after density modification with *SOLOMON* (starting solvent ∼40%), to a map which was interpretable under full automation by *ARP*/*wARP*/*REFMAC*

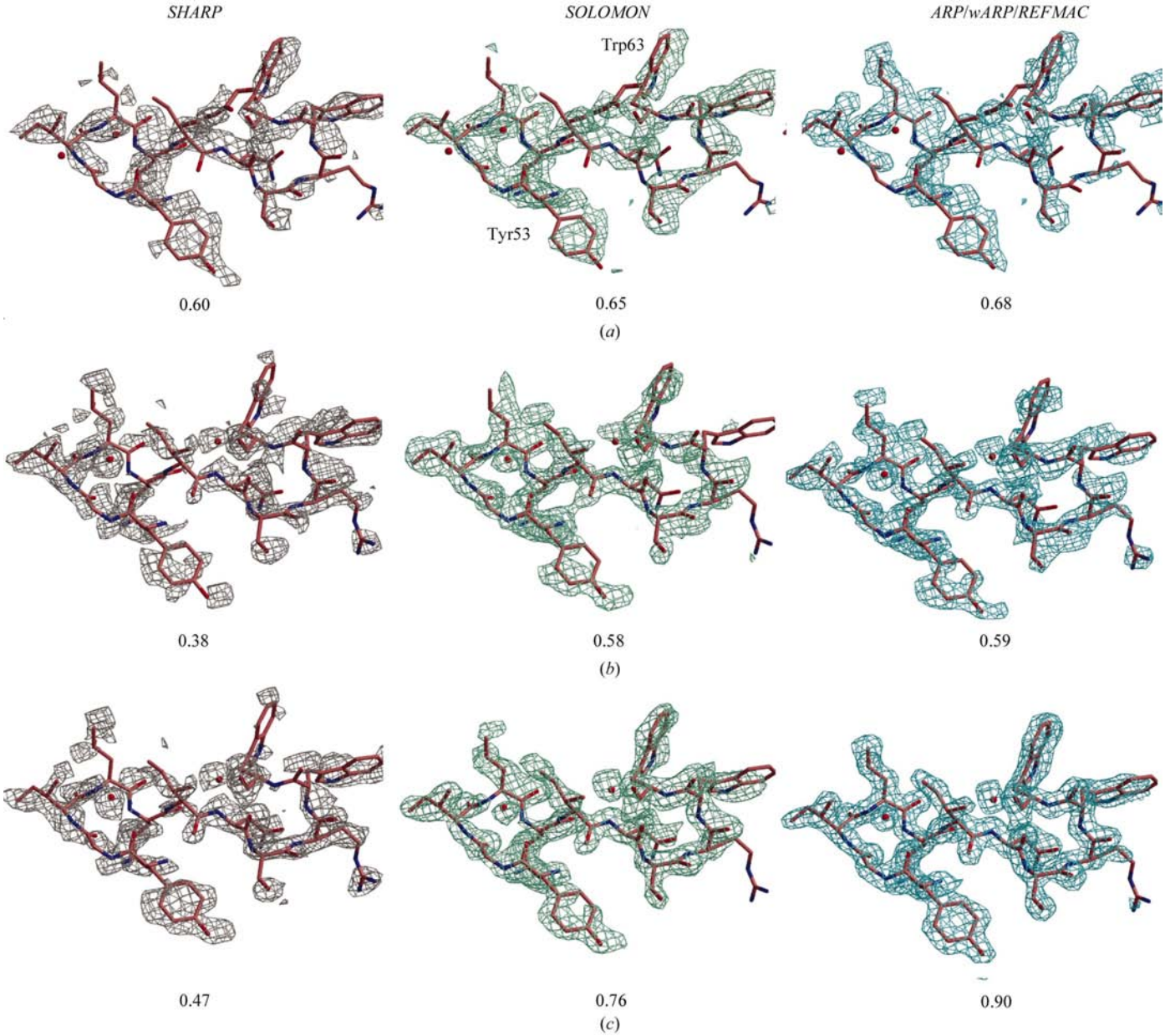

Figure 4
Electron-density maps of the region 52–63 superimposed with the relevant lysozyme model calculated subsequently from left to right with experimental phases after *SHARP*, then after solvent flipping by *SHARP*/*SOLOMON* and finally the automatic *SHARP*/*ARP*/*wARP*/*REFMAC* procedure. (*a*) MAD protocol using 2.0 Å amplitudes from crystal 1 that contains one arsenic site. (*b*) SAD protocol using 1.8 Å amplitudes from crystal 2 with one arsenic site and (*c*) with two arsenic sites. The contouring level is 1.2σ above the mean background. Values below the figures are ⟨RSCC⟩ values for residues 1–129.

(Fig. 4*c*). Indeed, 115/119 residues were built out of 129 and the phase error with the final refined model was only 19° (Fig. 5). Addition of the third minor site to the treatment offered only marginal improvement. As highlighted by the dramatic improvement of the map quality (Figs. 4*b* and 4*c*), the phases estimated by the SAD method benefitted much more from the improvement of the anomalous subset, even with a partially occupied additional site, than from attempts to reduce phase ambiguity using the dispersive signal, which were likely to be corrupted by insidious non-isomorphism arising from radiation damage. This reinforces the current tendency to suggest that, in many respects, a SAD approach is actually safer and less time-consuming than a MAD approach in order to obtain readily a rapid structure solution.

In the 2 Å experiment (crystal 1) recorded at twice the speed of crystal 2 and comprising only one As site, combination of anomalous and dispersive signals appeared to be more successful in deriving a Fourier difference electron-density map which benefitted from accurate low-resolution centroid phases (acentric reflections to 3.0 Å have figures of merit >0.73): it distinctly features a protein tracing as assessed by the 0.6 value for the overall correlation fit (real-space correlation coefficient) to the corresponding refined model. This value increased slightly to 0.7 after application of the density-modification procedure, followed by an attempt at automated model building that nevertheless aborted after <20 residues were placed. With 2 Å resolution data, we may be on the boundary for successfully exploiting the *ARP* implementation into *SHARP*. Fig. 5 possibly suggests in the present case that an overall phase error above 50° after density modification is the frontier for model auto-building to proceed to completion.

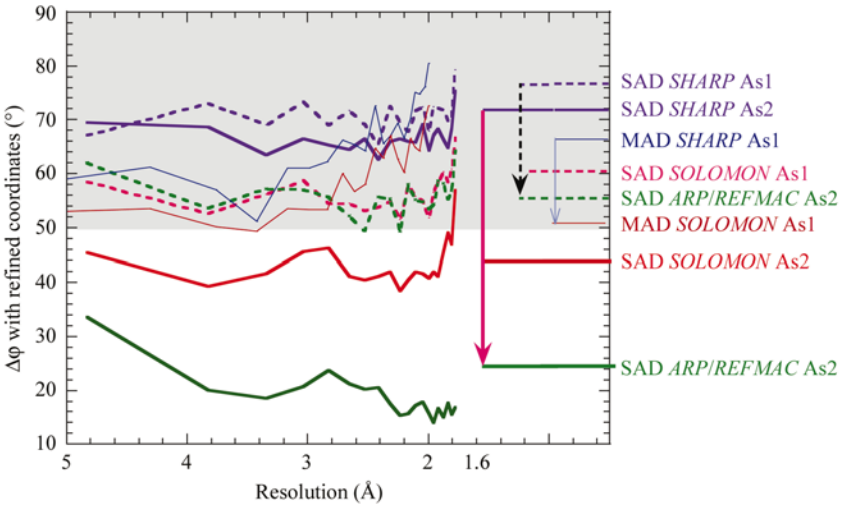

Figure 5
Average non-weighted difference $\Delta\varphi$ between the phases calculated from the corresponding refined model and the different sets of phases output first by *SHARP* (dotted lines) then by *SOLOMON* (thin solid lines) and finally by *ARP*/*wARP*/*REFMAC* (thick solid lines) *versus* resolution for each of the different phasing methods. An averaged phase difference value after solvent density modification that lies below the shaded region may lead to automatic model building.

Table 4
Refinement statistics from *BUSTER*/*TNT*.

HEWL	Crystal 1, MAD	Crystal 2, SAD
No. reflections (total)	7959	11331
No. reflections ($\|F\| > 3\sigma$)	6698	10741
No. reflections used for refinement	7181	10217
No. reflections used for R_{free}	775	1113
No. protein atoms	1004	1011
No. solvent non-H atoms (water)	126 (109)	175 (121)
Average *B* value for protein atoms (Å^2)	25.2	20.2
Average *B* value for solvent molecules (Å^2)	38.6	34.1
Range of spacings (Å)	16.0–2.0	16.0–1.78
R value† (%)	18.5	16.4
R_{free}‡ (%)	24.1	21.3
Weighted r.m.s.d. from ideality		
Bond length (Å)	0.014	0.009
Bond angle/dihedral angle (°)	1.295/15.2	1.212/23.2
Ramachandran plot: residues in		
Most favoured/additional allowed regions (%)	87.6/12.4	86.7 /13.3
Generously allowed/disallowed regions (%)	0/0	0/0

† The crystallographic *R* factor is defined as $\sum|F_o| - |F_c|/\sum|F_o|$ and indicates the accuracy of the model. ‡ The free *R* factor is a cross-validation residual calculated using 10% of the native data which were randomly chosen and excluded from the refinement.

3.2. Crystallographic model refinement and bound crystallizing anions to protein molecules

Starting from the *ARP*/*wARP*/*REFMAC* map obtained from the crystal 2 data set, a total of 1011 non-H protein atoms and three arsanilate anions were finally introduced to comply with the peak-wavelength amplitudes. The average correlation value between the model and the final *SIGMAA* map is better than 0.93. The standard deviation on coordinates estimated from the Luzzati plot (Luzzati, 1952) within *SFCHECK* (Vaguine *et al.*, 1999) is 0.16 Å and all non-glycine residues are in the favoured and allowed regions of the Ramachandran diagram (Ramachandran & Sasisekharan, 1968). In addition to these counter-anions, the final model comprises 121 water molecules with three on special positions and five ethylene glycol molecules and has an R_{free} of 21.3% and an *R* factor of 16.4% (see Table 4). Loop 70 has been modelled with an alternate flipped position at the level of the peptide bond between Arg73 and Asn74, making possible the binding of a sodium cation as previously described in Vaney *et al.* (1996). The side chains of Asn59 and Asn103 display alternate conformations.

The 2.0 Å (crystal 1) MAD model essentially differs from the crystal 2 model near Pro70, which is the cornerstone of the loop 66–73, the r.m.s. backbone coordinate difference otherwise being 0.26 Å. In this flexible region, Pro70 has moved away by more than 1.2 Å from the position it takes when the phenyl ring of the second arsanilate (Asr141) is stacked on it (Fig. 6*b*), suggesting that absence in crystal 1 of Asr141 is at least real and not blurred by inaccurate data.

3.3. Binding of arsanilate anions

All three anions bind to protein interfaces. Vaney *et al.* (2001) tentatively attempted to classify monovalent anion-binding sites into three main categories. The pTs anion was found to occupy a specific site (site 14 according to Vaney *et al.*, 2001). This site was also expected to bind pTAs. Nevertheless, in addition to its 17 additional electrons, pTAs also differs from pTS by substitution of the hydrophobic methyl group by a hydrophilic amine group. Fig. 6(*a*) shows that these changes do not hamper the binding. In fact, binding benefits from two new hydrogen bonds between the amine group and two carbonyl groups from Lys33 and from the symmetry-related Gly22. In contrast, two other anions managed to bind to 'lysozyme-specific' sites not specifically designed for pTAs and therefore explain the random and looser binding, with fewer interactions (Figs. 6*b* and 6*c*). The second site is near the sodium location inside the loop 70–75, which also accepts a chloride ion as in the 1gwd structure. The third site corresponds to Vaney's site number 5 which is known to accommodate nitrate and/or bromine and whose main contact is made to the hydroxyl group of Ser24.

We also checked the binding of the *ortho* form of arsanilate and concluded that the amine position does not really matter (*i.e.* both forms are basically interchangeable; see Figs. 6*a* and 6*b*) with respect to the overall tightness of interactions with the proteins.

4. General applicability

Finding a crystallizing agent with pertinent anomalous properties is of great interest since it would permit the reduction of two steps fundamental for crystallography to one. Co-crystallization is certainly a preferable alternative to the post-soaking method to produce strong substitution. Vaney *et al.* (2001) showed that a diffused pTS anion in a lysozyme crystal bound more weakly to the main site than when it was incorporated during crystal packing. The main drawback of this approach is that search for crystal-growth conditions may be a non-trivial preamble. Here, the availability of solubility curves for lysozyme in the presence of pTS at two pH values (Retailleau *et al.*, 2002) alleviated this task for the chemically similar pTAs. However, the dual hydrophilic/hydrophobic properties of arsanilate can be related to the non-detergent sulfobetaines, which have been acknowledged to form a class of compounds favouring molecular association (Vuillard *et al.*, 1998).

The general applicability of this method to other systems cannot yet be asserted. A preliminary test with porcine pancreatic elastase (PPE) showed that arsanilate could replace a sulfate ion usually found on the surface of PPE (unpublished results). Substitution of sulfate anions largely utilized in crystal growth by arsenate $HAsO_4^{2-}$ may be more appropriate, although preliminary tests on PPE crystals were not conclusive, probably because the protein was not previously deionized and thus carried residual sulfates competing for binding.

A quantitative advantage that arises with arsenates is their high degree of oxidation (V). Based on observations with fully oxidized selenomethionines in crystals (Smith & Thompson, 1998; Sharff *et al.*, 2000; Thomazeau *et al.*, 2001), an important gain in anomalous contribution is expected with As^{V} compared with As^{III}. On the other hand, another important asset is the spontaneous reduction by organic thiols that can lead to reactive intermediates with solvent-accessible free cysteines. Such a mechanism was highlighted in biocrystallo-

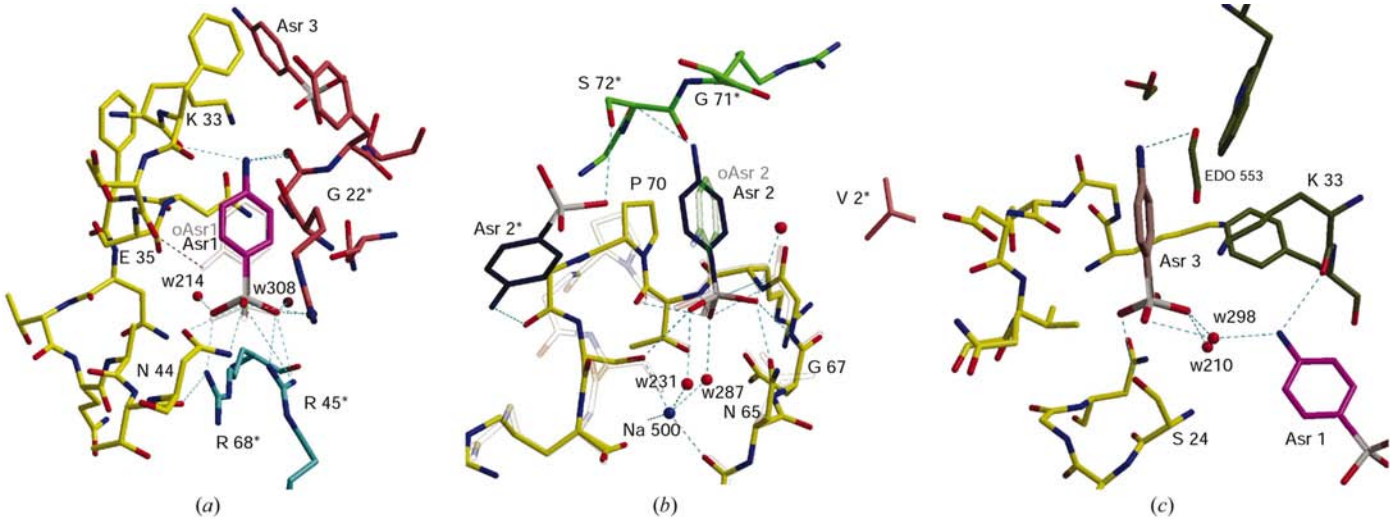

Figure 6
Details of the interactions around the three *p*-arsanilate anions found in the crystal 2 SAD refined structure (PDB code 1n4f). (*a*) A close-up of the main site Asr 1. The *ortho*-arsanilate anion is displayed as a transparency (coordinates not deposited). (*b*) The secondary Asr 2 site with the second *ortho*-arsanilate and the backbone and side chains of the crystal 1 MAD-refined structure superimposed as a transparency (coordinates not deposited), showing the flip of residues 69–72, including Pro70. (*c*) The third minor site in crystal 2, Asr 3, close to the main site located at the lower right. Graphics were created using *MOLSCRIPT* (Kraulis, 1991) and rendered by *RASTER3D* (Merritt & Murphy, 1994).

graphy by Maignan *et al.* (1998) and was further exploited by Choe and collaborators for the first MAD experiment around the *K* edge of arsenic (Greenwald *et al.*, 1999). This covalent substitution would likely compensate for the loss in f'' magnitude associated with As^{III} and thus cacodylation would be worth attempting in place of mercury derivatization, which is sometimes blamed for non-isomorphism problems. However, naturally occurring solvent-accessible cysteines are often implicated in catalytic mechanism and cacodylation may distort the active-site configuration and impede structural description of the enzyme function (Tete-Favier *et al.*, 2000), dissuading one to resort to this buffer (Maignan *et al.*, 1998).

5. Conclusions

We have confirmed that the existence of a sharp white line of organic arsenic at a wavelength very near to the Se *K* transition can be routinely exploited for accurate and nearly automatic phase determination. We have used small harmless (∼70 µg per protein droplet) amounts of an arsenic compound to derive lysozyme crystals in a co-crystallized fashion which seems to be the method of choice to strongly order particular counter-ions within the crystal packing, provided there are protein binding sites to host them. Arsenic-substituted sulfonate salts associated with a hydrophobic moiety are assumed to constitute a class of reagent that can bind to protein surfaces and may favour crystal growth. Indeed, many of them are dye molecules which are known to find application in protein-purification methods by affinity (*e.g.* recognition of the helical motif by bis-arsenical fluorescein; Thorn *et al.*, 2000) and thus may form a reservoir of potent protein markers also useful for crystallographic purposes. Another possible method of labelling the protein that is under investigation would be to pressurize crystals with the reducing gas arsine which should target free cysteines.

The present result brings further evidence that arsenic, which has thus far been underestimated and is sometimes absent from libraries called by crystallographic software, can contribute to the development of fast and full automation for *de novo* structure determination, complementing other current tools. The quality of data sets, however, remains a main criterion to ascertain how many arsenic sites are required per residue of the asymmetric unit for efficient anomalous phasing. With particularly weak substitution, efforts towards reducing systematic errors using appropriate strategies such as having Friedel pairs on the same image or over-redundancies in data collection may resolve difficulties. In this regard, the far *L* edge of As could also be exploited for in-house phase determination using soft X-rays at the Cr $K\alpha$ wavelength [$f''_{As} \simeq 2e^-$], as shown by Kwiatkowski *et al.* (2000).

This work was supported by European Commission grant No. HPRI-CT-1999-50015 within the EXMAD project. We would like to thank Madeleine Riès-Kautt for providing deionized lysozyme. We are grateful to Richard Kahn and BM30 staff for their help in using the FIP beamline at the European Synchrotron Radiation Facility (ESRF) in Grenoble. We thank Global Phasing Ltd for licenses for *SHARP* and *BUSTER*/*TNT* and particularly Clemens Vonrhein for help with *SHARP*, and Jeffrey Roach and Arnaud Ducruix for helpful comments on the manuscript.

References

Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* D**52**, 30–42.
Badger, J., Li, Y. & Caspar, D. L. D. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 1224–1228.
Bentley, R. & Chasteen, T. G. (2002). *Chem. Educator*, **7**, 51–60.
Blessing, R. H. & Smith, G. D. (1999). *J. Appl. Cryst.* **32**, 664–670.
Bricogne, G. (1993). *Acta Cryst.* D**49**, 37–60.
Bricogne, G. (1997). *Methods Enzymol.* **276**, 361–423.
Brodersen, D. E., de La Fortelle, E., Vonrhein, C., Bricogne, G., Nyborg, J. & Kjeldgaard, M. (2000). *Acta Cryst.* D**56**, 431–441.
Burling, F. T., Weis, W. I., Flaherty, K. M. & Brunger, A. T. (1996). *Science*, **271**, 72–77.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cohen, A., Ellis, P., Kresge, N. & Soltis, S. M. (2001). *Acta Cryst.* D**57**, 233–238.
Cowtan, K. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.
Dauter, Z. & Adamiak, D. A. (2001). *Acta Cryst.* D**57**, 990–995.
Dauter, Z. & Dauter, M. (1999). *J. Mol. Biol.* **289**, 93–101.
Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.
Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). *Acta Cryst.* D**56**, 232–237.
Deacon, A. M. & Ealick, S. E. (1999). *Structure*, **7**, R161–R166.
Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.
Doublié, S. (1997). *Methods Enzymol.* **276**, 523–530.
Evans, G. & Pettifer, R. (2001). *J. Appl. Cryst.* **34**, 82–86.
Evans, P. (1997). *Proceedings of the CCP4 Study Weekend. Recent Advances in Phasing*, edited by K. S. Wilson, G. Davies, A. W. Ashton & S. Bailey, pp. 114–122. Warrington: Daresbury Laboratory.
Fourme, R., Shepard, W. & Kahn, R. (1996). *Prog. Biophys. Mol. Biol.* **64**, 167–199.
French, G. S. & Wilson, K. S. (1978). *Acta Cryst.* A**34**, 517–525.
Girard, E., Chantalat, L., Vicat, J. & Kahn, R. (2002). *Acta Cryst.* D**58**, 1–9.
Greenwald, J., Le, V., Butler, S. L., Bushman, F. D. & Choe, S. (1999). *Biochemistry*, **38**, 8892–8898.
Guss, J. M., Merritt, E. A., Phizackerley, R. P. & Freeman, H. C. (1996). *J. Mol. Biol.* **262**, 686–705.
Hendrickson, W. A. (1999). *J. Synchrotron Rad.* **6**, 845–851.
Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **5**, 1665–1672.
Hendrickson, W. A. & Lattman, E. E. (1970). *Acta Cryst.* B**26**, 136–143.
Hendrickson, W. A., Smith, J. L. & Sheriff, S. (1985). *Methods Enzymol.* **115**, 41–55.
Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
Hosfield, D. J., Guan, Y., Haas, B. J., Cunningham, R. P. & Tainer, J. A. (1999). *Cell*, **98**, 397–408.
Howell, L. & Smith, D. (1992). *J. Appl. Cryst.* **25**, 81–86.
Iwata, S., Saynovits, M., Link, T. A. & Michel, H. (1996). *Structure*, **4**, 567–579.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Korolev, S., Dementieva, I., Sanishvili, R., Minor, W., Otwinowski, Z. & Joachimiak, A. (2001). *Acta Cryst.* D**57**, 1008–1012.

Kraulis, P. E. (1991). *J. Appl. Cryst.* **24**, 946–950.
Kwiatkowski, W., Noel, J. P. & Cheo, S. (2000). *J. Appl. Cryst.* **33**, 876–881.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EAMCB Newsl. Protein Crystallogr.* **26**.
Luzzati, V. (1952). *Acta Cryst.* **5**, 802–810.
Maignan, S., Guilloteau, J. P., Zhou-Liu, Q., Clement-Mella, C. & Mikol, V. (1998). *J. Mol. Biol.* **282**, 359–368.
Martin, P., DeMel, S., Shi, J., Gladysheva, T., Gatti, D. L., Rosen, B. P. & Edwards, B. F. (2001). *Structure*, **9**, 1071–1081.
Merritt, E. A. & Murphy, M. E. P. (1994). *Acta Cryst.* D**50**, 869–873.
Messens, J., Martins, J. C., Van Belle, K., Brosens, E., Desmyter, A., De Gieter, M., Wieruszeski, J.-M., Willem, R., Wyns, L. & Zegers, I. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 8506–8511.
Micossi, E., Hunter, W. N. & Leonard, G. A. (2002). *Acta Cryst.* D**58**, 21–28.
Murshudov, G., Vagin, A. & Dodson, E. (1997). *Acta Cryst.* D**53**, 240–255.
Perrakis, A., Morris, R. J. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Peterson, M. R., Harrop, S. J., McSweeney, S. M., Leonard, G. A., Thompson, A. W., Hunter, W. N. & Helliwell, J. R. (1996). *J. Synchrotron Rad.* **3**, 24–34.
Ramachandran, G. N. & Sasisekharan, V. (1968). *Adv. Protein Chem.* **23**, 283–438.
Retailleau, P., Ducruix, A. & Riès-Kautt, M. (2002). *Acta Cryst.* D**58**, 1576–1581.
Riès-Kautt, M. & Ducruix, A. (1997). *Methods Enzymol.* **276**, 23–59.
Roversi, P., Blanc, E., Vonrhein, C., Evans, G. & Bricogne, G. (2000). *Acta Cryst.* D**56**, 1316–1323.
Schiltz, M., Fourme, R. & Prangé, T. (1994). *J. Appl. Cryst.* **27**, 950–960.
Schiltz, M., Kvick, A., Svensson, O., Shepard, W., de La Fortelle, E., Prangé, T., Kahn, R., Bricogne, G. & Fourme, R. (1997). *J. Synchrotron Rad.* **4**, 287–297.
Schiltz, M., Shepard, W., Fourme, R., Prangé, T., de La Fortelle, E. & Bricogne, G. (1997). *Acta Cryst.* D**53**, 78–92.
Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.
Sharff, A. J., Koronakis, E., Luisi, B. & Koronakis, V. (2000). *Acta Cryst.* D**56**, 785–788.
Shepard, W., Cruse, W. B., Fourme, R., de La Fortelle, E. & Prangé, T. (1998). *Structure*, **6**, 849–861.
Smith, J. L. & Thompson, A. (1998). *Structure*, **6**, 815–819.
Tete-Favier, F., Cobessi, D., Boschi-Muller, S., Azza, S., Branlant, G. & Aubry, A. (2000). *Structure*, **8**, 1167–1178.
Thomazeau, K., Curien, G., Thompson, A., Dumas, R. & Biou, V. (2001). *Acta Cryst.* D**57**, 1337–1340.
Thorn, K. S., Naber, N., Matuska, M., Vale, R. D. & Cooke, R. (2000). *Protein Sci.* **9**, 213–217.
Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. (1987). *Acta Cryst.* A**43**, 489–501.
Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst.* D**55**, 191–205.
Vaney, M. C., Broutin, I., Retailleau, P., Douangamath, A., Lafont, S., Hamiaux, C., Prangé, T., Ducruix, A. & Riès-Kautt, M. (2001). *Acta Cryst.* D**57**, 929–940.
Vaney, M. C., Maignan, S., Riès-Kautt, M. & Ducruix, A. (1996). *Acta Cryst.* D**52**, 505–517.
Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2003). In the press.
Vuillard, L., Rabilloud, T. & Goldberg, M. E. (1998). *Eur. J. Biochem.* **256**, 128–135.
Walsh, M. A., Dementieva, I., Evans, G., Sanishvili, R. & Joachimiak, A. (1999). *Acta Cryst.* D**55**, 1168–1173.
Walsh, M. A., Evans, G., Sanishvili, R., Dementieva, I. & Joachimiak, A. (1999). *Acta Cryst.* D**55**, 1726–1732.
Walter, R. L., Ealick, S. E., Friedman, A. M., Blake, R. C. II, Proctor, P. & Shoham, M. (1996). *J. Mol. Biol.* **263**, 730–751.
Weiss, M. S. (2001). *J. Appl. Cryst.* **34**, 130–135.
Weiss, M. S. & Hilgenfeld, R. (1997). *J. Appl. Cryst.* **30**, 203–205.
Yao, J.-X. (1981). *Acta Cryst.* A**37**, 642–644.
Zhou, T., Radaev, S., Rosen, B. P. & Gatti, D. L. (2001). *J. Biol. Chem.* **276**, 30414–30422.