Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Applications of anomalous scattering from S atoms for improved phasing of protein diffraction data collected at Cu *Kα* wavelength

Cheng Yang* and J. W. Pflugrath

Rigaku/MSC Inc., The Woodlands, TX 77381, USA

Correspondence e-mail: cyang@rigakumsc.com

Received 27 April 2001
Accepted 6 August 2001

The anomalous signal of S atoms is easily detected at the Cu $K\alpha$ wavelength of a non-synchrotron source with current data-collection methods. The position of sulfur and other anomalous scatterers can be located through an anomalous difference Fourier map ($F^+ - F^-$, $\varphi_{\rm calc} - 90^\circ$). It has been discovered experimentally that even low-quality preliminary phases are often sufficient to find anomalous scatterers. Their anomalous signal in the native crystal can contribute to significant improvement in phase refinement. This technique has been applied to solve the crystal structures of orthorhombic lysozyme and thaumatin. Furthermore, the structure of trypsin was solved using only the diffraction data set from a native crystal collected at a single wavelength (Cu $K\alpha$) from a rotating-anode X-ray generator. The anomalous scattering of sulfur was essential to solve the structure of trypsin which was initially phased from a single intrinsic Ca^{2+} atom. The positions of the S atoms of lysozyme and thaumatin were found using the initial SIRAS phases and used in phase refinement. The overall figures of merit and those in each resolution shell were consistently improved. This resulted in much improved electron-density maps even when the diffraction data were limited to 2.5 Å resolution or worse. Furthermore, peaks from S atoms and other anomalous scatterers in anomalous difference Fourier maps can confirm the tracing of the peptide chain and also provide independent unbiased confirmation of molecular-replacement results. Thus, the anomalous signal of S atoms can contribute to many aspects of solving protein structures and should be used routinely.

1. Introduction

Anomalous or Bijvoet differences are widely used to help determine the phases of Bragg reflections collected in an X-ray diffraction experiment. With the advent of synchrotron radiation and the use of selenomethionine, the crystallographic phase problem is readily solved by accurately measuring Bijvoet differences in a multiwavelength anomalous dispersion (MAD) experiment conducted with wavelengths near the absorption edges of an anomalous scatterer (Hendrickson, 1991). Even before the widespread use of synchrotron radiation in structural biology, anomalous differences were used to help determine phases with diffraction data collected away from absorption edges with Cu $K\alpha$ and other radiation from sealed-tube and rotating-anode X-ray generators (Argos & Mathews, 1973; Hendrickson & Teeter, 1981; Wang, 1985). In the isomorphous replacement technique, anomalous differences are especially important in

© 2001 International Union of Crystallography
Printed in Denmark – all rights reserved

resolving the phase ambiguity from a single heavy-atom derivative (SIRAS). Under favorable circumstances, the anomalous differences alone can provide enough initial phase information to successfully determine the structure of a macromolecule (SAS, ISAS; Hendrickson & Teeter, 1981; Liu *et al.*, 2000; Brodersen *et al.*, 2000; Wu *et al.*, 2001).

Diffraction experiments that rely on anomalous differences require a combination of a detector capable of measuring the signal, one or more elements in the crystal with imaginary structure-factor components ($\Delta f''$) and an appropriate wavelength in order to produce a measurable signal. In a well designed experiment, the wavelength can be chosen to maximize $\Delta f''$. However, in many experiments the wavelength is fixed and cannot be adjusted. This is the case with a rotating-anode or microfocus X-ray generator with a copper target. The element present in the crystal need not possess a large anomalous scattering contribution in order to be useful. For example, S, Ca, Zn and Se atoms collected with Cu $K\alpha$ radiation have $\Delta f''$ values of 0.557, 1.286, 0.678 and 1.139, respectively (International Tables for X-ray Crystallography, 1974).

Early work by Teeter and Hendrickson led to the structure determination of the 46-residue protein crambin directly from the anomalous scattering of S atoms found in six cysteine residues (Hendrickson & Teeter, 1981). More recently, Dauter *et al.* (1999) have shown that small anomalous differences can be used to solve the phase problem for diffraction data collected from tetragonal hen egg-white lysozyme crystals. They used the anomalous signal from S and Cl atoms measured at the wavelength of 1.54 Å from a synchrotron source to 1.53 Å resolution. Brodersen *et al.* (2000) used the weak anomalous signal of a bound Zn atom to determine the structure of psoriasin from the native data alone. Wang and colleagues determined the crystal structures of obelin from sulfur anomalous scattering collected at the wavelength of 1.74 Å and of ferrochelatase from iron anomalous scattering collected at the wavelength of 1.54 Å (Liu *et al.*, 2000; Wu *et al.*, 2001). Stuhrmann and colleagues (Stuhrmann *et al.*, 1995, 1997; Behrens *et al.*, 1998) have exploited the relatively heroic data collection from single crystals of ribosome, trypsin and bacteriorhodopsin near the sulfur *K* absorption edge (5.02 Å) where the $\Delta f''$ is maximized. This required a great deal of modification of the hardware setup at a synchrotron beamline.

Many of these earlier efforts to use weak anomalous signals for solving the phase problem have been limited to well diffracting crystals of relatively small proteins. In this work, we show that even poorly diffracting crystals of large proteins can benefit from the inclusion of sulfur anomalous scattering in the phase determination. In contrast to previously published work, we do not attempt to completely phase the observed amplitudes from the sulfur anomalous differences alone and thereby determine the crystal structure with no other information. Nor in most cases do we attempt to locate the positions of the weak anomalous scatterers from the observed structure-factor amplitudes alone. Instead, we use the Bijvoet differences arising from the presence of weak anomalous scatterers as an adjunct to improve phases initially determined from either the isomorphous replacement or the molecular-replacement methods. These weak anomalous signals for the most part have been ignored in phase refinement/determination, though they are almost always present. We demonstrate that inclusion of this information in the phase calculation leads to a dramatic improvement in the phases and the Fourier syntheses calculated with them. Indeed, previously uninterpretable electron-density maps become readily interpretable with the new information, while reasonably good maps become even better. The improvement is seen with phases initially determined with the MIRAS, SIRAS, SAS, SIR and MR methods.

We also show that in the favorable case of a bovine pancreatic trypsin crystal the anomalous signal of a structural Ca atom is evident in an anomalous difference Patterson synthesis. The phases derived from this signal enable us to identify the positions of the S atoms. It followed that the anomalous signals from the Ca and S atoms were measured well enough to phase the entire trypsin molecule and generate an easily interpreted electron-density map. In other words, the trypsin crystal structure was solved with one diffraction data set collected at the Cu $K\alpha$ wavelength without resorting to heavy-atom substitution. The implications of this are obvious: there was no need for selenomethionine, molecular replacement and/or a heavy-atom derivative to solve the trypsin crystal structure. Since so many proteins are metalloproteins with intrinsic anomalous scatterers, the approach used in phasing the trypsin diffraction data presented herein may potentially be used to determine the crystal structures of members of this class of proteins.

Although replacement of methionine with selenomethionine and a subsequent synchrotron data-collection experiment is a powerful tool to determine crystal structures, it is not always possible. In some instances, recombinant protein is not available. In others, the SeMet protein does not crystallize. These cases can benefit from careful data collection to measure the sulfur anomalous signal.

2. Materials and methods

2.1. Crystallization and data collection

The three test proteins used in the experiment are bovine pancreatic trypsin, hen egg-white lysozyme and *Thaumatococcus deniellii* thaumatin. The proteins were purchased from commercial sources and have been crystallized according to the conditions documented in the literature, with slight modifications (Bode & Schwager, 1975; Van der Wel *et al.*, 1975). The crystallization conditions are listed in Table 1. In particular, the lysozyme was crystallized in the presence of $KAu(CN)_2$, which yields an orthorhombic crystal form (unpublished results; CSHL Macromolecular Crystallography Course).

In this experiment, one native trypsin crystal was used. Trypsin contains an intrinsic weak anomalous scatterer in the form of a structural Ca atom. At 1.54 Å wavelength (*i.e.* Cu $K\alpha$) calcium and selenium have very similar values of $\Delta f''$

of 1.286 and 1.139 electrons, respectively (*International Tables for X-ray Crystallography*, 1974).

Two crystals were used for each of the lysozyme and thaumatin determinations. The lysozyme orthorhombic crystal form grown for our studies incorporates an Au atom into the crystal lattice. It appears that this Au atom is lost as the crystals age. A relatively old 'gold-depleted' crystal was used for a native diffraction data set, while a second similar crystal was soaked in the crystallization solution saturated with potassium dicyanoaurate, $KAu(CN)_2$, for 4 h in order to reincorporate the gold moiety. A thaumatin crystal was derivatized by soaking in 10 m*M* methylmercury chloride dissolved in crystallization solution overnight.

All diffraction data from native and derivative crystals of the test proteins were collected on either a 5.0 kW Rigaku RUH2R or RUH3R rotating-anode generator equipped with MSC Blue-3 or MSC Green-3 multilayer optics at the wavelength of Cu $K\alpha$ radiation (1.54 Å). The detector used was either a Rigaku R-AXIS IV^{++} imaging plate or MSC Jupiter-140 CCD detector. All the crystals were mounted in arbitrary orientations and collected at cryogenic temperatures in a single continuous scan; that is, no inverse-beam approach was employed. A highly redundant native data set (>tenfold) was collected for each test crystal. All the diffraction data were processed with *d*TREK* (Pflugrath, 1999). The data-processing statistics are summarized in Table 1.

Table 1
Data-collection statistics.

	Trypsin	Orthorhombic lysozyme		Thaumatin	
	Native	Native	Au derivative	Native	Hg derivative
Radiation (Å)	1.54	1.54		1.54	
XG power (kW)	5.0	5.0		5.0	
Detector and optics	MSC Jupiter CCD-140, MSC Green-3	R-AXIS IV++, MSC Blue-3		R-AXIS IV++, MSC Blue-3	
Unit-cell parameters (Å, °)	54.1, 58.3, 69.7, 90.0, 90.0, 90.0	29.8, 55.9, 72.7, 90.0, 90.0, 90.0		57.8, 57.8, 150.6, 90.0, 90.0, 90.0	
Space group	$P2_12_12_1$	$P2_12_12_1$		$P4_12_12$	
Diffraction resolution (Å)	2.0	1.50	1.50	2.0	2.0
R_{merge} (I) (%)	5.6 (23.4)†	4.0 (12.0)	2.5 (6.3)	6.8 (18.4)	6.0 (16.4)
Completeness (%)	98.7 (96.4)	96.0 (68.2)	83.7 (33.4)	99.9 (98.9)	99.6 (97.1)
Redundancy	13.9 (8.9)	11.4 (5.4)	5.3 (4.2)	17.9 (9.9)	6.1 (3.7)
$I/\sigma(I)$	22.1 (12.5)	42.3 (11.0)	56.5 (27.1)	36.1 (7.9)	25.9 (9.9)
No. of atoms (per ASU)	1765	1011		1551	
No. of anomalous scatterers (per ASU)	1 Ca^{2+}, 14 S	10 S, 8 Cl^-		17 S	
Resolution of phasing (Å)	15.0–2.5	15.0–2.5		15.0–2.5	
Experimental $\langle\Delta F\rangle/\langle F\rangle$ (%)	1.9	1.7		1.7	
Calculated $\langle\Delta F\rangle/\langle F\rangle$ (%)	1.6	1.2		1.2	
Experimental $\langle\lvert\Delta F\rvert\rangle/\langle\sigma(\Delta F)\rangle$ (%)	1.3	1.3		1.0	
Crystallization conditions	4% PEG 4000, 0.2 *M* Li_2SO_4, 15% ethylene glycol, 0.1 *M* MES pH 6.5	6.5% NaCl, 4 m*M* $KAu(CN)_2$, 100 m*M* Na acetate pH 4.8		1 *M* KNa tartrate, 100 m*M* ADA pH 6.5	

† Values in parentheses are for the highest resolution shell.

2.2. Phase determination

All the programs used in finding the locations of heavy atoms and anomalous scatterers and phase refinement were from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994). Every data set was treated equally through the *DTREK2MTZ* and *TRUNCATE* (French & Wilson, 1978) programs to calculate the amplitude of structure factors ($|F|$) and anomalous difference (ΔF) from the measured intensities (I). For all calculations described below, the diffraction data within the resolution range 15.0–2.5 Å was used.

The initial position of the Ca^{2+} ion in the trypsin crystal was located from inspection of Harker sections of an anomalous difference Patterson map. The position and anomalous occupancy of the Ca^{2+} were refined using the anomalous difference of the native crystal with *MLPHARE* (Otwinowski, 1991) before the initial single-wavelength anomalous scattering (SAS) phases were calculated using the position of Ca^{2+} only.

Anomalous difference Fourier maps ($F^+ - F^-, \varphi_{SAS} - 90°$) were generated using the *FFT* program (Ten Eyck, 1973). *PEAKMAX* (from the *CCP*4 suite) was employed to search for peaks in the anomalous difference Fourier map in order to locate the positions of anomalous scatterers such as S atoms. The potential positions of the anomalous scatterers were refined along with the parameters for Ca^{2+} and new SAS phases were calculated. This cycle was iterated several times. The details will be described in §3.

The positions of heavy atoms in the lysozyme and thaumatin crystals were located from their isomorphous difference Patterson maps. A similar procedure used in phasing the trypsin crystal structure was applied here. The positions of anomalous scatterers were also found by peak searching on the anomalous difference Fourier maps ($F^+ - F^-$, $\varphi_{SIRAS} - 90°$). The positions and anomalous occupancies of anomalous scatterers were refined using the anomalous difference of the native data as second derivative data. New phases were calculated with information from both the heavy-atom and the anomalous scatterers.

The improvement of electron-density maps was visualized with the program *O* (Jones *et al.*, 1991). The map correlation coefficient of the F_o map was calculated using *OVERLAPMAP* (Jones & Stuart, 1991) with the $2F_o - F_c$ map (F_c from the final atomic model) used as the reference map.

3. Results and discussion

As stated in the introduction, nearly all proteins have methionine and/or cysteine residues. S atoms are inherently present in almost all proteins. However, the sulfur K absorp-

tion edge is at 5.02 Å, which is far from the energy typically available at crystallographic beamlines or with laboratory X-ray generators. It is very difficult to solve a structure using the *K* edge of S atoms (Stuhrmann *et al.*, 1995, 1997; Behrens *et al.*, 1998). This paper will use several examples to show the anomalous scattering signal of S atoms can be easily measured on a laboratory source at the wavelength of Cu $K\alpha$ radiation (1.54 Å). The anomalous scattering of S atoms can be essential to solve a protein structure.

3.1. Phasing orthorhombic trypsin structure using a single data set

More than a third of the current protein population is known as metalloproteins that contain one or more bound metal ions such as Fe, Co, Ca, Zn and Mn (The Metalloprotein Structure and Design Group, 1999). Some nonmetalloproteins can bind one or more metal ions during their crystallization process. The bovine pancreatic trypsin molecule has one noncovalently bound Ca^{2+} ion. In this experiment, a single data set of a native trypsin crystal was collected on a 5 kW laboratory source at Cu $K\alpha$ wavelength (1.54 Å) and its structure was phased from the intrinsic anomalous scattering of Ca^{2+} and 14 S atoms.

The trypsin crystal was about 0.1 × 0.1 × 0.2 mm in size. It was crystallized in the expected space group $P2_12_12_1$. During X-ray diffraction data collection, the crystal was maintained at cryogenic temperatures and no noticeable decay was observed. The data-collection statistics are listed in Table 1.

The trypsin crystal diffracted beyond 1.6 Å resolution, but the data were collected and processed up to the edge of the detector (2.0 Å). The overall redundancy was about 14-fold

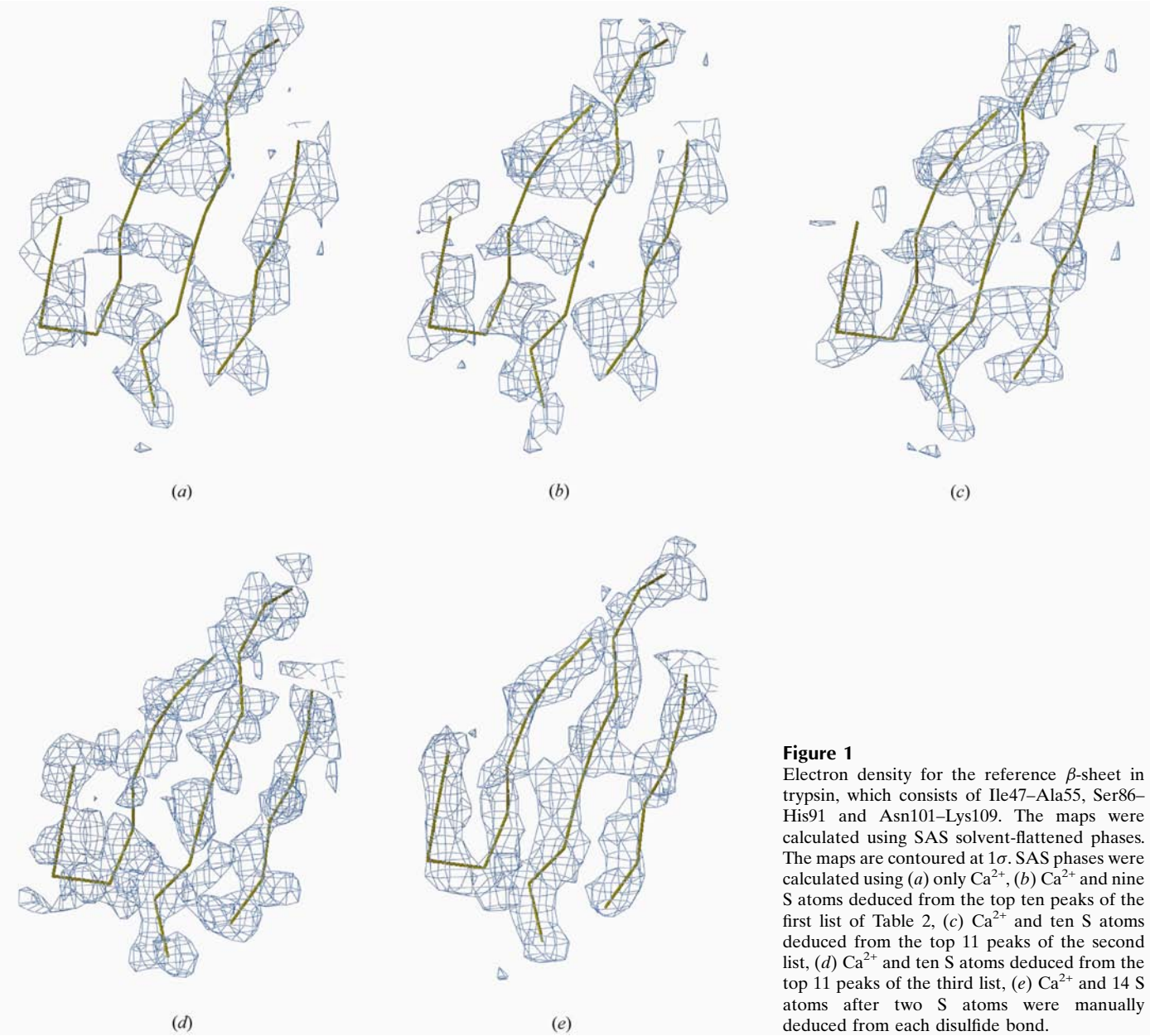

Figure 1
Electron density for the reference β-sheet in trypsin, which consists of Ile47–Ala55, Ser86–His91 and Asn101–Lys109. The maps were calculated using SAS solvent-flattened phases. The maps are contoured at 1σ. SAS phases were calculated using (*a*) only Ca^{2+}, (*b*) Ca^{2+} and nine S atoms deduced from the top ten peaks of the first list of Table 2, (*c*) Ca^{2+} and ten S atoms deduced from the top 11 peaks of the second list, (*d*) Ca^{2+} and ten S atoms deduced from the top 11 peaks of the third list, (*e*) Ca^{2+} and 14 S atoms after two S atoms were manually deduced from each disulfide bond.

Table 2
Result of peak searching on an anomalous difference Fourier map of trypsin.

Peak list number	1		2		3		4	
SAS phases of	Ca^{2+}		Ca^{2+} and peaks 2–10 of list 1 as S atoms		Ca^{2+} and peaks 2–11 of list 2 as S atoms		Ca^{2+} and peaks 2–11 of list 3 as S atoms	
FOM†	0.26		0.30		0.32		0.43	
Peak	Height	Anomalous scatterers	Height	Anomalous scatterers	Height	Anomalous scatterers	Height	Anomalous scatterers
1	51.9	Ca^{2+}	45.1	Ca^{2+}	40.7	Ca^{2+}	27.6	Ca^{2+}
2	6.2		15.0	Met104	16.1	Cys42–Cys58	16.5	Cys42–Cys58
3	5.3	Met104	13.9	Met106	15.7	Cys22–Cys157	16.3	Cys22–Cys157
4	5.1	Met106	12.9	Cys168–Cys182	14.8	Cys136–Cys202	16.0	Cys168–Cys182
5	4.9		12.7	Cys136–Cys201	14.0	Met104	14.6	Cys136–Cys202
6	4.9	Cys168–Cys182	12.5	Cys128–Cys232	13.8	Cys168–Cys182	13.8	Cys191–Cys220
7	4.6	Cys128–Cys232	6.7		13.6	Met180	13.8	Met180
8	4.5		6.3	Cys42–Cys58	13.2	Cys191–Cys220	13.3	Cys232–Cys128
9	4.4	Cys136–Cys201	5.9	Cys22–Cys157	11.2	Cys232–Cys128	12.3	Met104
10	4.2		5.5		10.4		3.8	
11	4.1		5.4	Cys191–Cys220	7.6		3.7	
12	4.0		4.2		3.9		3.6	
13	4.0		4.1		3.8		3.6	
14	4.0		4.0		3.8		3.6	
15	4.0		3.7		3.7		3.4	
16	3.9		3.7		3.7		3.3	

† FOM is the figure of merit of the SAS phases used to generate the anomalous difference Fourier map.

with 98% overall completeness. The overall $R_{\rm merge}$ and that in the highest resolution (2.0–2.14 Å) shell are 5.6 and 23.6%, respectively. The overall $\langle I/\sigma(I)\rangle$ for the unaveraged observations is 22.1.

Two ratios of $\langle|\Delta F|\rangle/\langle F\rangle$ and $\langle|\Delta F|\rangle/\sigma\langle|\Delta F|\rangle$ were used to evaluate the strength and significance of anomalous scattering. According to Hendrickson & Teeter (1981), the expected average ratio of $\langle|\Delta F|\rangle/\langle F\rangle$ can be calculated as $2^{1/2}(N_A^{1/2}\Delta f_A'')/(N_P^{1/2}Z_{\rm eff})$. In the case of trypsin, the number of anomalous scatterers (N_A) in trypsin is 15, the number of total non-H protein atoms (N_P) is 1765 and the effective atomic number ($Z_{\rm eff}$) is ~6.7 for non-H protein atoms. One Ca^{2+} ion with $\Delta f_A'' = 1.283$ electrons and 14 S atoms with $\Delta f_A'' = 0.563$ electron results in a value of only 1.6% for the ratio of $\langle|\Delta F|\rangle/\langle F\rangle$ at the wavelength of Cu $K\alpha$ radiation. The overall experimental ratio of $\langle|\Delta F|\rangle/\langle F\rangle$ is 1.9% up to the resolution of 2.0 Å because of the experimental errors in the measured intensities. The experimental ratios of $\langle|\Delta F|\rangle/\langle F\rangle$ and $\langle|\Delta F|\rangle/\sigma\langle|\Delta F|\rangle$ are equal to 1.93 and 1.34%, respectively, in the resolution range 15.0–2.5 Å used to phase the trypsin data. At higher resolution, the ratio of $\langle|\Delta F|\rangle/\langle F\rangle$ increases, but the ratio of $\langle|\Delta F|\rangle/\sigma\langle|\Delta F|\rangle$ decreases. It remains to be tested what resolution is optimal for phasing protein data with the anomalous scattering of S atoms.

Prior to locating the positions of S atoms of trypsin, the position of Ca^{2+} was located on the Harker sections of an anomalous difference Patterson map. The position and anomalous occupancy of the Ca^{2+} ion were initially refined using the anomalous difference of its native data for ten cycles while its B factor was set to 40 Å^2. The error-prone weak data with $|F_h|$ or $|F_{-h}| < 2\sigma(I)$ were excluded. A total of 7698 anomalous differences up to the resolution of 2.5 Å were used. Since the $\Delta f''$ of Ca^{2+} is 1.286 electrons and there is one Ca^{2+} ion among 1765 non-H atoms, the overall figure of merit of SAS phasing of the Ca^{2+} ion was only 0.26. The solvent-flattening procedure *DM* (Cowtan, 1994) was employed to improve the quality of the electron-density map, but the electron density was still very noisy and the boundary of the molecule was still ambiguous because of low solvent content in the trypsin crystal (27%). One β-sheet of the core of trypsin structure was chosen as the reference to display the change in quality of the electron-density maps (Fig. 1), which consists of Ile47–Ala55, Ser86–His91 and Asn101–Lys109. The correlation coefficient between the F_o map and the $2F_o - F_c$ map of the final refined structure was as low as 0.15. This first electron-density map was uninterpretable (Fig. 1*a*).

When SAS phases calculated from the position of the Ca^{2+} ion were used to calculate the anomalous Fourier map ($F^+ - F^-$, $\varphi_{\rm SAS} - 90$), some 'anomalous' peaks with a considerable height were found by the program *PEAKMAX*. The highest peak has a value more than six standard deviations above the average value in the map. A trypsin molecule has S atoms from two methionine and 12 cysteine residues. The first peak list in Table 2 shows the 16 highest found 'anomalous' peaks and their corresponding residues. No obvious decrease in peak height was seen between peaks 15 and 16. The positions of the first ten peaks after the peak of Ca^{2+} (>4σ) were input into the program *MLPHARE* as S-atom positions. A second round of phase refinement and calculation were performed with a Ca^{2+} and ten S atoms deduced from first ten non-Ca^{2+} 'anomalous' peaks of the first peak list. The figure of merit increased by 15% from 0.26 to 0.30 and the map correlation coefficient also improved. These second SAS phases were also modified by the solvent-flattening procedure (*DM*). The map of the reference β-sheet shown in Fig. 1(*b*) still appears to be noisy and uninterpretable.

The same procedure to generate the first list of anomalous peaks was used to produce the second list. Five noise peaks included in the second phase calculation were no longer

present, while the peaks corresponding to the positions of S atoms remained with an increased height. An obvious decrease in height between peaks 11 and 12 was also seen in the second list of anomalous peaks in Table 2. The positions of top ten peaks below the Ca^{2+} peak in the second list were input (along with the Ca^{2+} position) to the program *MLPHARE* as the sulfur positions for a third round of SAS phase refinement and calculation. The overall figure of merit of the SAS phases was increased to 0.32. The boundary of the protein molecule became clear and some secondary structure became distinguishable, such as the electron density of the reference β-sheet shown in Fig. 1(*c*).

The same procedure was iterated twice more to produce a third and fourth list of anomalous peaks (Table 2). Ten anomalous peaks besides the Ca^{2+} peak remained with increased peak heights. The overall figure of merit of the fourth SAS phases was increased to 0.43. The map correlation coefficient reached 0.41 and the map became interpretable, but many discontinuous regions were still seen in the electron-density map. The electron density of the reference β-sheet is shown in Fig. 1(*d*).

The top eight out of ten anomalous peaks of the third list were found again in the fourth anomalous difference map (fourth list of Table 2). For most of the disulfide bonds, a single anomalous peak instead of two peaks for two S atoms can be found in an anomalous difference Fourier map at a low resolution. The peak position is usually referred to as the position of a 'super S atom' (Chen *et al.*, 2000) and located at the center of a disulfide bond (—SS—). Including higher resolution data can resolve the one peak into two peaks from which two sulfur positions can be deduced. The positions of six S atoms of crambin were recognized only on the anomalous difference Patterson maps at the resolution of 1.5 Å. It still remains to be investigated what resolution is high enough to deduce the positions of two S atoms of a disulfide bond on an anomalous difference Fourier map. In this experiment, an anomalous difference Fourier map up to the resolution of 2.5 Å was calculated using the fourth SAS phases. This map was visually inspected along with the position of anomalous peaks in the fourth list. In addition to the density of the Ca^{2+} ion, two individual spherical densities and six elongated ellipsoidal densities were seen (Fig. 2). One anomalous peak rests in each density. The two peaks in two individual spherical densities were recognized as the S atoms of Met180 and Met104. The peaks in the center of an ellipsoidal density were recognized as the —SS— atoms of a disulfide bond and replaced with two S atoms whose bond distance was 2 Å. Therefore, the positions of 14 S atoms and one Ca^{2+} were input into the fifth round of SAS phase refinement and calculation. The overall figure of merit increased to 0.46. The map correlation coefficient increased to 0.48. The map improved considerably, as evident in Fig. 1(*e*) which shows the electron density of the reference β-sheet. Fig. 3 displays the electron density of one β-strand (Gln51–Ala55) of the reference β-sheet with the side chains in place. The main chain of 245 amino acids was manually traced within a few hours. Fig. 4 shows the direct comparison between SAS phases from the Ca^{2+} ion only and from the Ca^{2+} ion and 14 S atoms. Fig. 4(*a*) displays the map of the C-terminal helix (Tyr234–Asn245) calculated using only SAS solvent-flattened phases derived from the anomalous scattering of the Ca^{2+} ion. It was the most recognizable region of the entire electron-density map. The density of the main chain was discontinuous and side chains were not covered by electron density at the 1σ level. It was very difficult to recognize the region as an α-helix. Fig. 4(*b*) shows the electron density of the same helix calculated with the anomalous scattering of the Ca^{2+} ion and 14 S atoms and

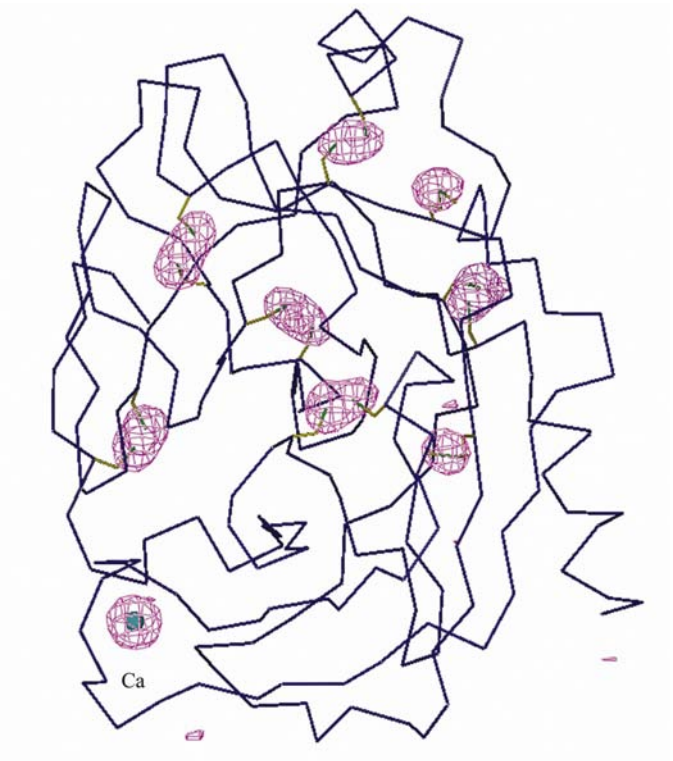

Figure 2
Anomalous difference Fourier map generated using SAS solvent-flattened phases for trypsin (C_α backbone with Ca^{2+} and side chains of Cys and Met). The anomalous peaks of the fourth list were found on this map. The map was contoured at 5σ.

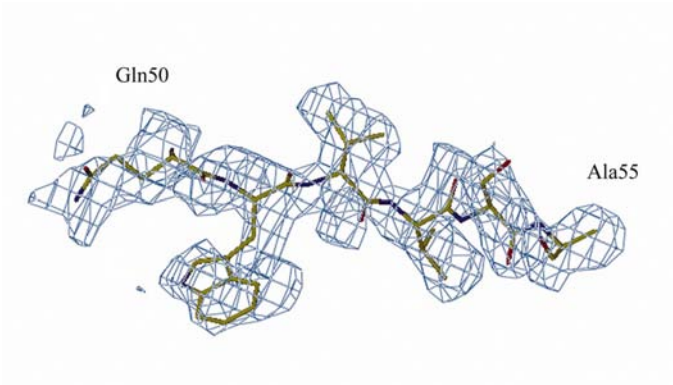

Figure 3
2.5 Å resolution electron density for one β-strand in trypsin (Gln51–Ala55) of the reference β-sheet with the side chains in the place. The map was contoured at 1σ.

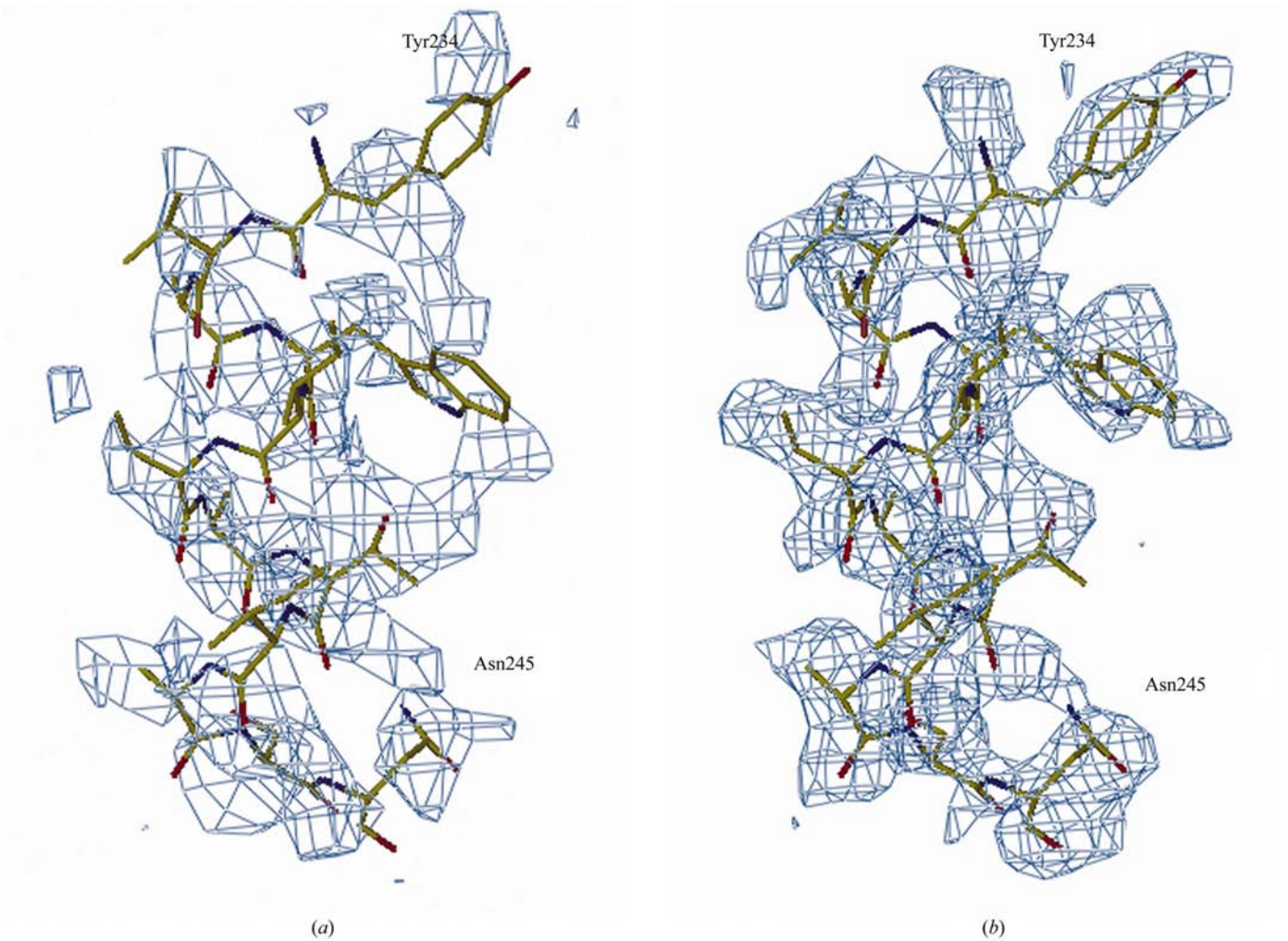

Figure 4
2.5 Å resolution electron density for the C-terminal helix (model coordinates) generated using different SAS phases. The SAS phases were improved through solvent flattening. The maps were contoured at 1σ. (*a*) SAS phases were calculated using anomalous scattering of only Ca^{2+}. (*b*). SAS phases were calculated using anomalous scattering of Ca^{2+} and 14 S atoms.

solvent flattening. The density of the main chain can be easily recognized and almost every side chain was covered by well defined density, especially Tyr234 and Trp235.

Trypsin (MW $\simeq$ 26 kDa) is a good example for testing a new approach to the use of sulfur anomalous scattering in phasing protein structures. According to the values of the imaginary dispersion correction of atomic structure factors ($\Delta f''$) for metal ions, the strength of the anomalous scattering of the Ca^{2+} ion is lower than the commonly used heavy atoms and many common metal ions of metalloproteins, such as Fe and Co atoms. A Ca atom has almost the same strength of anomalous scattering as an Se atom at the wavelength of 1.54 Å. The preparation of selenium protein has become a routine step for structural determination. The results of this experiment suggest the positions of Se atoms can be located using the data collected on a non-synchrotron source and the combination of anomalous scattering of S and Se atoms should be able to solve a protein structure without diffraction data from a synchrotron beamline.

3.2. Phasing the protein structures with one native and one derivative data set

SIR and SIRAS phasing are commonly used in solving protein structures. In this experiment, we demonstrate that including the sulfur anomalous scattering signal in the cases of SIR and SIRAS phasing can greatly improve the results. Known SIR or SIRAS phases were used to generate an anomalous difference Fourier map through which the positions of S atoms were located. The positions and anomalous occupancies of found S atoms were refined along with the positions of heavy atoms using the anomalous differences of native data. The addition of anomalous differences of native data in the phase refinement and calculation can enhance

Table 3
Phasing statistics.

	Trypsin		Lysozyme		Thaumatin	
Phase of	Ca^{2+}	Ca^{2+} + 14 S	Au	Au + (2 —SS—, 6 S, 7 Cl)†	Hg	Hg + (2 —SS—, 12 S)‡
Resolution (Å)	FOM		FOM		FOM	
15.00–9.23	0.22	0.30	0.65	0.66	0.58	0.59
9.23–6.67	0.25	0.45	0.65	0.74	0.54	0.65
6.67–5.22	0.26	0.49	0.70	0.80	0.56	0.69
5.22–4.29	0.29	0.47	0.66	0.75	0.52	0.65
4.29–3.64	0.30	0.49	0.53	0.70	0.47	0.62
3.64–3.16	0.29	0.49	0.49	0.69	0.40	0.58
3.16–2.79	0.26	0.46	0.41	0.61	0.34	0.54
2.79–2.50	0.22	0.40	0.34	0.54	0.28	0.50
15.0–2.50	0.26	0.46	0.48	0.67	0.39	0.57
Map correlation coefficient	0.15	0.48	0.48	0.67	0.52	0.56

† Two —SS—, six S and seven Cl atoms were deduced from the first 15 anomalous peaks of lysozyme in Table 4. They were used as individual S atoms in phase refinement. ‡ Two —SS— and 12 S atoms were deduced from the first 14 anomalous peaks of thaumatin in Table 4. They were used as individual S atoms in phase refinement.

Table 4
Results of peak searching in anomalous difference Fourier maps of lysozyme and thaumatin.

Protein	Lysozyme		Thaumatin	
SIRAS phase of	One Au site		One Hg site	
FOM	0.48		0.39	
Peak	Height	Anomalous scatterers	Height	Anomalous scatterers
1	7.1	Cys127–Cys6	9.6	Cys193–Cys121
2	6.8	Cys76–Cys94	9.2	Met112
3	6.7	Cl^-	8.4	Cys9
4	5.6	Met105	7.8	Cys56
5	5.5	Cl^-	7.5	Cys126–Cys117
6	5.4	Cys64	7.0	Cys149
7	5.1	Cl^-	7.0	Cys71
8	4.8	Cl^-	6.9	Cys77
9	4.8	Cys30	6.5	Cys145
10	4.5	Cl^-	6.4	Cys134
11	4.3	Cys80	6.4	Cys158
12	4.3	Cl^-	6.2	Cys66
13	4.2	Cys115	5.4	Cys164
14	3.9		5.1	Cys159
15	3.9		4.5	
16	3.7		4.3	
17	3.7		4.1	
18	3.7		4.0	

resolving the phase ambiguity of SIR and SIRAS phase distributions without introducing any lack-of-isomorphism error.

The two proteins, lysozyme and thaumatin, with molecular weights of 14 and 22 kDa, respectively, were crystallized in different crystal lattices and space groups (Table 1). The crystals diffracted to 2.0 Å resolution or better. In our study, all the refinements were performed using only 15.0–2.5 Å data to match that of many crystals currently collected with laboratory X-ray generators. An iterative procedure similar to that described for trypsin above was used with the lysozyme and thaumatin diffraction data. However, instead of a single diffraction data set as in the case of trypsin, heavy-atom derivative data were collected and used along with the native data.

The expected average ratios of $\langle|\Delta F|\rangle/\langle F\rangle$ of both lysozyme and thaumatin are ~1.2% with Cu $K\alpha$ radiation, lower than those calculated for trypsin and crambin. Experimentally derived $\langle|\Delta F|\rangle/\langle F\rangle$ ratios of native data of lysozyme and thaumatin are close to their expected value (1.2%) in the resolution range to 2.5 Å. These ratios increase at higher resolution owing to the increase in error of ΔF and F measurement. The datum of orthorhombic lysozyme has an overall $R_{\rm merge}$ of 4% and $\langle I/\sigma(I)\rangle$ for unaveraged observations of 42. Thaumatin datum has an overall $R_{\rm merge}$ of 6.8% and $\langle I/\sigma(I)\rangle$ of 36. The experimental ratio $\langle|\Delta F|\rangle/\langle\sigma\Delta F\rangle$ of lysozyme and thaumatin native data are above 1.0 to 2.5 Å. The significance of the anomalous signal $\langle|\Delta F|\rangle/\langle\sigma\Delta F\rangle$ appears to be closely correlated to the quality of data. Our experience shows that the redundancy of data can play a critical role in increasing the accuracy of the $\langle|\Delta F|\rangle$ and $\langle\sigma\Delta F\rangle$ measurement, especially when the data are collected at a non-synchrotron source. This is consistent with the observation by Weiss (2001) that data redundancy improves the accuracy of averaged intensities.

The attempt to find sulfur positions by inspection of an anomalous difference Patterson map calculated from the data of lysozyme and thaumatin in the resolution range 15.0–2.5 Å failed owing to the low ratios of $\langle|\Delta F|\rangle/\langle F\rangle$ and $\langle|\Delta F|\rangle/\langle\sigma\Delta F\rangle$, although others have used more powerful procedures to do so (Dauter *et al.*, 1999; Weiss, 2001). A gold derivative of lysozyme and a mercury derivative of thaumatin were prepared using the procedure described in §2. The heavy-atom positions were located by isomorphous difference Patterson maps. The SIRAS phases were calculated with the program *MLPHARE*. The figures of merit of the SIRAS phases of each test crystal are listed in Table 3.

An anomalous difference Fourier map of each native data set from the lysozyme and thaumatin crystals was generated using SIR or SIRAS phases and the anomalous difference of the native data in the resolution range 15–2.5 Å. The peak-search program (*PEAKMAX*) was also used to find the peaks in the anomalous difference Fourier maps. Table 4 lists the found anomalous peaks and corresponding anomalous scatterers.

After the structure was aligned to the electron-density map, the height of the largest anomalous peak of the lysozyme data is 7.1σ above the average map value; this peak was found to be the —SS— of Cys6 and Cys127. In fact, the anomalous peaks of all S atoms except Cys204 and Met12 were greater than 4σ in height. Six anomalous peaks corresponding to bound chloride ions ($\Delta f''$ = 0.702) were also found along with those corresponding to S atoms (Table 4) and were originally included in the phase refinement as S atoms.

Lysozyme has eight cysteines and two methionines. Some noise peaks along with the real anomalous peaks were presumably found as top peaks because of the lack of accuracy of the initial phases and there may be also some unexpected anomalous scatterers in proteins. The peaks (>4σ) were

usually used as those of anomalous scatterers. For lysozyme, the positions of the first 13 peaks (>4σ) were initially input to *MLPHARE* and treated as the positions of S atoms. The anomalous differences of the native data for lysozyme were used as the data for the second derivative to refine the position and anomalous occupancy of S atoms. The total number of reflections used was 4511. After ten cycles of refinement, the overall figure of merit increased from 0.48 to 0.67. The figure of merit showed more improvement at high resolution than at low resolution (Table 3). This resulted in significant improvement of the electron-density map. The map correlation coefficient of the electron-density map increased by 20% to 0.53 with the $2F_o - F_c$ map of the final refined structure used as a reference map. Fig. 5 illustrates the difference of electron densities in the region Gln41–Asn44 in the presence and absence of S and Cl atoms as anomalous scatterers in the phase-calculation procedure.

The improved phases produced an anomalous difference Fourier map with eight spherical and four strongly ellipsoidal densities. Since the quality of the electron density of lysozyme was significantly improved, resolving disulfide bonds into two individual S atoms was not necessary for the phasing calculation. When the electron-density map and anomalous difference Fourier maps were superimposed, two spherical and four ellipsoidal density peaks seemed to correspond to the S atoms of methionine and cysteine, as they overlaid the electron densities of side chains. These anomalous densities can be used as markers for chain tracing (Lehmann & Pebay-Peyroula, 1992; Hendrickson & Sheriff, 1987). Six spherical peaks in the anomalous difference Fourier map, corre-

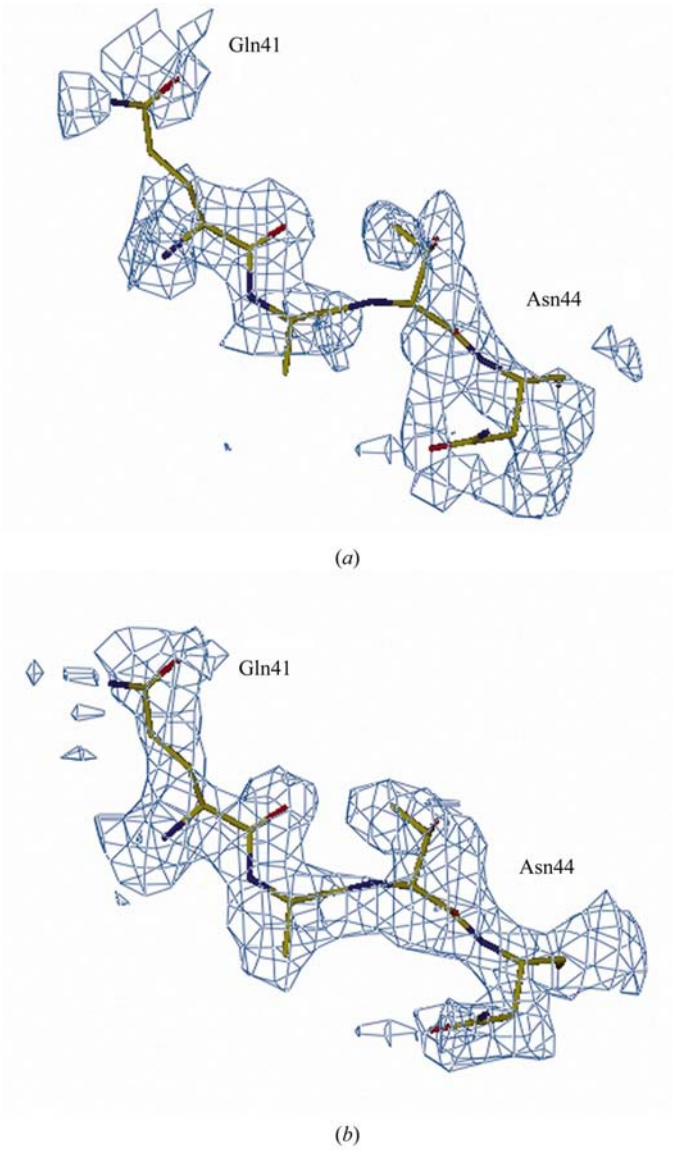

Figure 5
Electron density for the residues of Gln41–Asn44 of lysozyme. The maps were computed using different phases. (*a*) SIRAS phases of one Au site. (*b*) MIR phases of one gold site and sulfur sites of protein. The maps are contoured at 1σ.

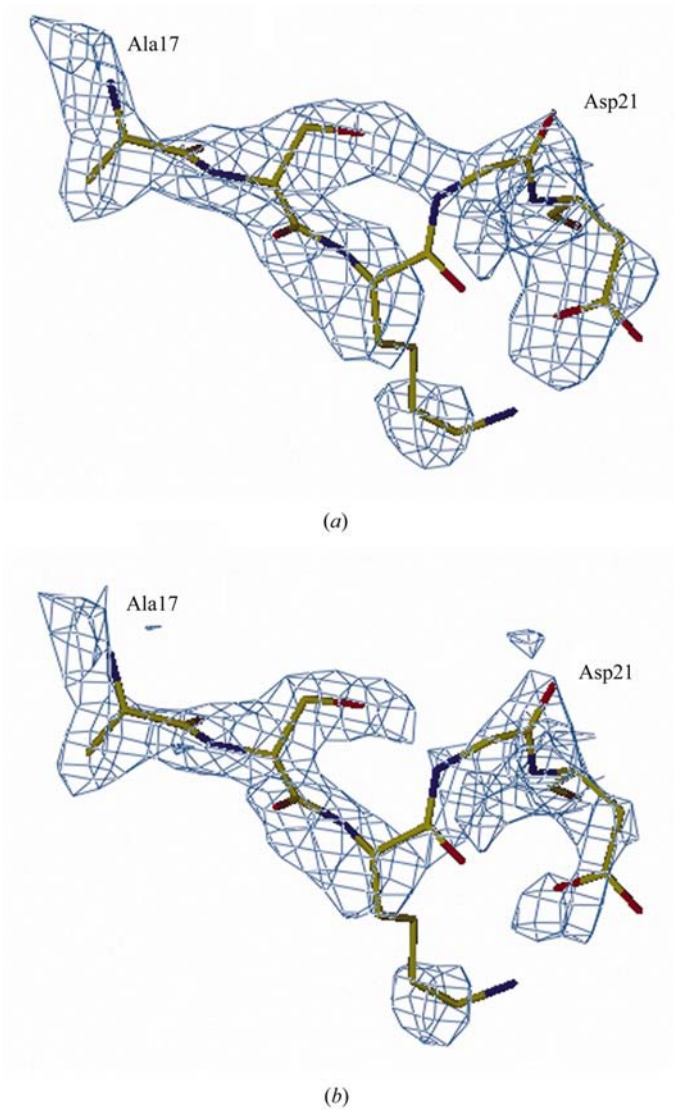

Figure 6
Electron density for the residues of Ala17–Asp21 of thaumatin. The maps were computed using different phases. (*a*) SIRAS phases of one mercury site. (*b*) MIR phases of one mercury site and sulfur sites of protein. The maps are contoured at 1σ.

sponding to peaks 3, 5, 7, 8, 10 and 12 in Table 4, might arise from some other anomalous scatterers. They were located about 2 and 4 Å away from the side-chain and main-chain protein atoms. These peaks were nearly as strong as those attributed to the protein S atoms, which serve as an internal standard. Chloride ions were inferred to be the only species possible from the crystallization medium. The chloride ions were found in the same locations in the lysozyme structure (1lz8) in the PDB (Berman *et al.*, 2000).

Diffraction data from a single mercury derivative and the native crystal of thaumatin were collected on a rotating-anode X-ray generator with Cu $K\alpha$ radiation. The procedure used here was exactly the same as that used in phasing the lysozyme structure. The figure of merit of the SIRAS phases of the thaumatin structure was 0.39. The height of the first anomalous peak was 9.6σ (Table 4). The thaumatin sequence has 16 cysteine residues and one methionine residue. No obvious separation in peak height can be seen between peaks 17 and 18. The positions of the top 14 anomalous peaks ($>5\sigma$) were input to *MLPHARE* as the initial positions of S atoms when the anomalous difference of native thaumatin data was used as the data of the second derivative. The total number of unique reflections used was 9426. The overall figure of merit was improved by ~46% to 0.57 by including these 14 anomalous signal peaks as S atoms in phase calculation. The overall map correlation coefficient increased by ~8%. The better improvement of the figure of merit was also seen in the highest resolution shell (Table 3). Fig. 6(*a*) shows the electron density for the structure near Ala17 and Asp21 without any S atoms used in the phase calculation. A section of main chain could be mistakenly traced through the density between the side chains of Ser18 and Gly20. Fig. 6(*b*) displays a clear trend in the density of the main chain and well defined density for the Ser18 side chain.

A new anomalous difference Fourier map displayed clearly eight strong ellipsoidal densities for eight disulfide bonds and one isolated spherical density for the one methionine sulfur after the phases were calculated including sulfur anomalous scattering. The first and fifth anomalous peaks in Table 3 corresponded to —SS— moieties. After replacing these —SS— positions with individual S atoms, the overall figure of merit increased to 0.63. The overall map correlation coefficient rose by ~15%.

Thus, it has been demonstrated in the examples of trypsin and thaumatin that resolving the —SS— of a disulfide bond into two individual S atoms is required to take full advantage of the strength of anomalous scattering signal of S atoms. It even may be essential in solving protein structures.

4. Conclusions

The combination of the MAD method and tunable synchrotron radiation has become extremely powerful and popular in the field of structural biology, but difficulties still arise under certain circumstances. For instance, sometimes selenomethionine protein cannot be expressed in a system, crystals of selenomethionine protein cannot be grown or the crystals are too fragile to be handled. In these cases, a major benefit would be to use other anomalous scatterers found in protein crystals for phase determination.

S atoms have always possessed an anomalous signal thought to be generally useful in phasing protein diffraction intensities. However, the $\langle|\Delta F|\rangle/\langle F\rangle$ ratios calculated from S atoms in proteins are usually very low. For instance, the test proteins and crambin have values less than 2% for $\langle|\Delta F|\rangle/\langle F\rangle$. Until now, it has been very difficult to detect these signals directly in an anomalous difference Patterson synthesis for large proteins at a low resolution. In fact, the sulfur anomalous signal can be relatively easily detected in the anomalous difference Fourier map when any kind of initial phases are available such as the phases from the SIR, SIRAS, SAS, MIR and MR techniques. We have demonstrated how the sulfur anomalous scattering can be easily collected using Cu $K\alpha$ radiation from a non-synchrotron source and used in solving protein structures. Indeed, the methods described herein were essential in the determination in our laboratory of a proprietary crystal structure with more than 800 residues in the asymmetric unit.

There are some advantages of using a laboratory source over a synchrotron to collect the sulfur anomalous scattering. The 1.54 Å (Cu $K\alpha$) radiation is longer than the typical wavelength used at synchrotrons for protein single-crystal diffraction (0.8–1.1 Å). At wavelengths longer than 1.54 Å (Cu $K\alpha$) radiation, such as 2 Å, the expected $\langle\Delta|F|\rangle/\langle|F|\rangle$ can increase, but additional problems will be introduced such as the radiation absorption of crystals and air scattering. Even at the wavelength 1.54 Å, the anomalous scattering signal of S atoms can be masked by errors such as radiation damage without the protection of cryogenic temperatures (Weik & Sussman, 2000). This problem is exacerbated by the high-brilliance radiation of synchrotron beamlines. It still remains unclear how to optimize the data collection for sulfur anomalous scattering at a synchrotron beamline, although Wang and colleagues have used 1.74 Å wavelength (Liu *et al.*, 2000). Owing to advances in current X-ray generators, optics, detectors and processing software, even the weak Bijvoet differences created by S and other atoms can be measured accurately with a laboratory source. The radiation absorption of a crystal can be corrected by commonly used software. The radiation damage to cryocooled crystals is negligible, which in turn allows highly redundant diffraction data be collected. Highly redundant data are known to be essential for detecting the sulfur anomalous scattering on a conventional laboratory source and even at a synchrotron beamline. Nevertheless, it is very convenient to collect data without traveling to a synchrotron beamline.

We wish to reiterate what seems obvious but often is not practiced by structural biologists. That is, anomalous scatterers can be located readily in anomalous difference Fourier maps calculated with phases derived from other methods (Rossmann, 1961; Argos & Mathews, 1973; Hendrickson & Sheriff, 1987; Lehmann & Pebay-Peyroula, 1992). In fact, instead of looking for electron-density features that match large aromatic side chains as an aid in initially locating the primary sequence into the map, we recommend that one first looks for

sulfur peaks found in an anomalous difference Fourier map. This is analogous to finding the methionine positions in a MAD-phased map. Sulfur peaks also provide an important and easy confirmation of the map fitting – one knows that the methionine and cysteine residues are correctly placed. Furthermore, one can also distinguish metal atoms such as Mn and Ca from water molecules and ions such as Na^+. It even may be possible to distinguish between Mn and Mg ions, since the $\Delta f''$ of Mn is larger than that of both Mg and S (an internal standard!) with Cu $K\alpha$ radiation.

Besides their use as chain-tracing markers, the anomalous scattering of S atoms can also contribute to other aspects of structure determination. For instance, when the method of molecular replacement is employed to solve a structure, the position and anomalous occupancy of S atoms may provide an independent confirmation of the initial MR solution.

Finally, careful, but not heroic, data-collection experiments are required for all the techniques we discuss. With current X-ray generators, detectors and processing software, it is relatively easy to observe the Bijvoet differences arising from the weak anomalous scattering of S and other atoms. We stress that careful experiments should always be performed. The intrinsic anomalous signal from S and other atoms should be routinely measured and used in phasing protein crystal diffraction data, since it can contribute to many other aspects of solving protein crystal structures.

The authors wish to thank Dr Joseph D. Ferrara for helpful discussions and critical reading of the manuscript, Adam Courville for technical support of the X-ray equipment and Jodi DeSchepper for growing the test crystals. Since the submission of this manuscript, a diffraction data set from a single trypsin crystal that we collected as described herein has been phased by C. Vonrhein and G. Bricogne in an automated procedure that used the *SHARP* program (personal communication).

References

Argos, P. & Mathews, F. S. (1973). *Acta Cryst.* B**29**, 1604–1611.
Behrens, W., Otto, H., Stuhrmann, H. B. & Heyn, M. P. (1998). *Biophys. J.* **75**, 255–263.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Bode, W. & Schwager, P. (1975). *J. Mol. Biol.* **98**, 693–717.
Brodersen, D. E., de La Fortelle, E., Vonrhein, C., Bricogne, G., Nyborg, J. & Kjeldgaard, M. (2000). *Acta Cryst.* D**56**, 431–441.
Chen, C. J., Rose, J. P., Rosenbaum, G. & Wang, B.-C. (2000). Abstr. Am. Crystallogr. Assoc. Annu. Meet., Abstr. E0046.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cowtan, K. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–48.
Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.
French, G. S. & Wilson, K. S. (1978). *Acta Cryst.* A**34**, 517.
Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
Hendrickson, W. A. & Sheriff, S. (1987). *Acta Cryst.* A**43**, 121–125.
Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
International Tables for X-ray Crystallography (1974). Vol. IV. Birmingham: Kynoch Press. (Present distributor Kluwer Academic Publishers, Dordrecht.)
Jones, T. A., Zhou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Jones, Y. & Stuart, D. (1991). *Proceedings of the CCP4 Study Weekend. Isomorphous Replacement and Anomalous Scattering*, edited by W. Wolf, P. R. Evans & A. G. W. Leslie, pp. 39–48. Warrington: Daresbury Laboratory.
Lehmann, M. S. & Pebay-Peyroula, E. (1992). *Acta Cryst.* B**48**, 115–116.
Liu, Z.-J., Vysotski, E. S., Cheng, C. J., Rose, J. P., Lee, J. & Wang, B.-C. (2000). *Protein Sci.* **9**, 2085–2093.
The Metalloprotein Structure & Design Group (1999). *MDB (Main Page) – Metalloprotein Site Database and Browser, TSRI.* http://metallo.scripps.edu/.
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.
Pflugrath, J. W. (1999). *Acta Cryst.* D**55**, 1718–1725.
Rossmann, M. G. (1961). *Acta Cryst.* **14**, 383–388.
Stuhrmann, S., Bartels, K. S., Braunwarth, W., Doose, R., Dauvergne, F., Gabriel, A., Knöchel, A., Marmotti, M., Stuhrmann, H. B., Trame, C. & Lehmann, M. S. (1997). *J. Synchrotron Rad.* **4**, 298–310.
Stuhrmann, S., Hütsch, M., Trame, C., Thomas, J. & Stuhrmann, H. B. (1995). *J. Synchrotron Rad.* **2**, 83–86.
Ten Eyck, L. F. (1973). *Acta Cryst.* A**29**, 183–191.
Van der Wel, H., van Soest, T. C. & Royers, R C. (1975). *FEBS Lett.* **56**, 316–317.
Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
Weik, M. & Sussman, J. L. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 623–628.
Weiss, M. (2001). *J. Appl. Cryst.* **34**, 130–135.
Wu, C.-K., Dailey, H. A., Rose, J. P., Burden, A., Sellers, V. M. & Wang, B.-C. (2001). *Nature Struct. Biol.* **8**, 156–160.