Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Phasing on anomalous signal of sulfurs: what is the limit?

Udupi A. Ramagopal,[a] Miroslawa Dauter[b] and Zbigniew Dauter[a]*

[a]Synchrotron Radiation Research Section, National Cancer Institute, Brookhaven National Laboratory, Building 725A-X9, Upton, NY 11973, USA, and [b]SAIC-Frederick Inc., Basic Research Program, Brookhaven National Laboratory, Building 725A-X9, Upton, NY 11973, USA

Correspondence e-mail: dauter@bnl.gov

Received 6 February 2003
Accepted 2 April 2003

Recent years have witnessed significant advancements in X-ray data-acquisition techniques and phasing algorithms, which have made possible the successful use of a very small anomalous diffraction signal for the solution of crystal structures of macromolecules. Two crystal structures, a 44 kDa glucose isomerase containing nine sulfurs and a 33 kDa xylanase containing five sulfurs, have been solved from single-wavelength anomalous data using widely available methods and programs. These two enzymes contain less sulfur than most proteins in the bacterial or eukaryotic proteomes, providing a Bijvoet ratio of about 0.6%. For glucose isomerase the automatically interpretable electron-density maps could be obtained at high as well as low resolution. The S-SAD approach relies on the anomalous signal of sulfur naturally occurring in proteins and alleviates all need for sample derivatization. It may therefore be applicable to all protein crystals able to provide accurate diffraction data.

1. Introduction

The first successful attempts to phase macromolecular structures based on the anomalous scattering signal of sulfur date from the early 1980s. The structure of crambin (Hendrickson & Teeter, 1981), containing five sulfurs in 45 amino acids, was solved by the resolved anomalous scattering approach. The expected amount of anomalous signal, in the form of the calculated Bijvoet ratio $\langle \Delta F^{\rm anom}/F \rangle$, was 1.5%. From simulations based on a 12 kDa protein containing one disulfide bridge, Wang (1985) concluded that a Bijvoet ratio of 0.6% may lead to successful structure solution by the iterative single-wavelength anomalous scattering (ISAS) technique. The anomalous effect of sulfur has often been observed and has been used to confirm the protein chain tracing (*e.g.* Greenwald *et al.*, 1999), but was not used for phasing novel structures until the late 1990s.

The necessity for very accurate diffraction data has been the main reason for such a long gap in the use of the sulfur anomalous signal in macromolecular crystallographic practice. However, in recent years there has been a revival of interest in this approach, mostly owing to advances in methodology, such as the cryocooling of crystals, the availability of synchrotron beamlines and advances in phasing algorithms. Table 1 summarizes the published macromolecular structures (novel and test cases) that have been solved with the use of the anomalous signal of sulfur. Phosphorus and chlorine, which display an anomalous diffraction behavior similar to that of sulfur, are also included.

The *K* absorption edges of P, S and Cl lie in the long-wavelength region of X-ray radiation, at 5.78, 5.02 and 4.39 Å,

Table 1
Structures solved using the anomalous signal of P, S and Cl atoms.

Protein	Atoms in ASU	Scatterers in ASU	Solvent content (%)	λ (Å)	f'' (e)	Bijvoet ratio (%)	Reference
Crambin	344	6S	33	1.54	0.56	1.45	Hendrickson & Teeter (1981)
Rhe†	833	2S	51	1.54	0.56	0.58	Wang (1985)
Lysozyme	1001	10S + 7 Cl	40	1.54	0.56 + 0.70	1.55	Dauter *et al.* (1999)
Oligonucleotide	290	10P	25	1.54	0.43	1.68	Dauter & Adamiak (2000)
CAP-Gly domain	813	3S + 1 Cl	54	1.74	0.70 + 0.86	1.11	Li *et al.* (2002)
Apocrustacyanin C1	2911	18S	45	1.77	0.72	1.00	Gordon *et al.* (2001)
IGF2R fragment	1006	11S	42	1.77	0.72	1.60	Brown *et al.* (2002)
Vancomycin aglycon	320	7Cl + 1S	30	1.91	1.00	3.40	Loll (2001)
Lysozyme	1001	12S + 7 Cl	40	1.54	0.56 + 0.70	1.55	de Graaff *et al.* (2001)
Oligonucleotide	290	10P	25	1.54	0.43	1.68	de Graaff *et al.* (2001)
Thiostrepton	105	5S	15	1.54	0.56	2.0	Bond *et al.* (2001)
Tryparedoxin form I	2376	13S	46	1.77	0.72	1.2	Micossi *et al.* (2002)
Tryparedoxin form II	2360	13S	72	1.77	0.72	1.2	Micossi *et al.* (2002)
Thaumatin	1551	17S	55	1.54	0.56	1.7	Yang & Pflugrath (2001)
EAS	3627	16S	55	1.54	0.56	0.86	Lemke *et al.* (2002)
Obelin	1553	8S + 1 Cl	67	1.74	0.70 + 0.86	1.12	Liu *et al.* (2000)
LBTI	501	14S	73	1.54	0.56	1.96	Debreczeni *et al.* (2003)
Xylanase	2300	5S	36	1.74	0.70	0.69	Current work
Xylanase	2300	5S	36	1.49	0.52	0.53	Current work
Glucose isomerase	3050	9S + 1 Cl	54	1.54	0.56	0.68	Current work

† This work used the error-free data calculated from the atomic model of Rhe protein.

Table 2
Anomalous scattering of P, S and Cl.

	P		S		Cl	
Wavelength (Å)	f' (e)	f'' (e)	f' (e)	f'' (e)	f' (e)	f'' (e)
2.29 (Cr $K\alpha$)	0.377	0.899	0.375	1.141	0.333	1.423
2.00	0.354	0.703	0.374	0.897	0.372	1.123
1.74	0.318	0.544	0.349	0.697	0.368	0.876
1.54 (Cu $K\alpha$)	0.282	0.433	0.317	0.556	0.345	0.701
1.49	0.271	0.406	0.307	0.522	0.337	0.659
1.28 (Au $L\alpha$)	0.228	0.304	0.261	0.393	0.294	0.498
0.98	0.153	0.180	0.183	0.234	0.213	0.299
0.71 (Mo $K\alpha$)	0.082	0.094	0.102	0.124	0.122	0.159

respectively, according to *CROSSEC* (Cromer, 1983). The anomalous scattering contributions of these atoms at various wavelengths accessible at the laboratory or synchrotron sources are presented in Table 2. At wavelengths shorter than 2.0 Å these elements provide an anomalous signal smaller than 1.0 electron units.

A stronger signal can be obtained at longer wavelengths. At wavelengths of up to 2.0 Å, no unduly special experimental precautions are needed, *e.g.* thinner detector windows or helium beam paths (Einspahr *et al.*, 1985; Helliwell, 1993; Cianci *et al.*, 2001). At longer wavelengths, *e.g.* 3–5 Å, this is no longer true (Stuhrmann *et al.*, 1997; Liu *et al.*, 2001). In the longer wavelength region 1.5–3.0 Å, the range 1.5–2.0 Å appears to be optimal for sulfur SAD phasing (Weiss, Sicker, Djinovic Carugo *et al.*, 2001; Kwiatkowski *et al.*, 2000; Weiss, Sicker & Hilgenfeld, 2001).

The only way to utilize the anomalous signal of sulfur for phasing diffraction data is through the single-wavelength anomalous diffraction (SAD) approach, since no other 'native' data are available and the small variation of wavelength used in MAD technique does not lead to any significant difference in the scattering of S, P or Cl far from their X-ray absorption edges. The SAD phasing consists of two steps. Firstly, the anomalous substructure has to be solved; that is, the positions of the anomalous scatterers have to be located. This can be performed either by direct methods or by the anomalous difference Patterson interpretation. Neither of these methods can differentiate between the two possible enantiomeric solutions and both possibilities have to be checked for correctness.

The second step of the SAD procedure is the evaluation of protein phases. In principle, three different measurements and knowledge of the model responsible for the observed differences are necessary for unequivocal phase evaluation. A single-wavelength data set provides at best (for acentric reflections) two independent intensity measurements for each unique reflection (or a pair of Friedel mates). This leads to the ambiguity in the phase estimation, which can be expressed by the following sine relationship: $\Delta F^{\rm anom} = 2F''\sin(\varphi_T - \varphi_A)$, where $\Delta F^{\rm anom}$ is measured and F'' and φ_A can be calculated from the already known partial structure of anomalous scatterers. Two solutions are possible for $\varphi_T = \varphi_A + 90^\circ \pm \cos^{-1}(\Delta F^{\rm anom}/2F'')$. A weak discrimination between the two phase possibilities may be provided by the theory of the phase probability resulting from the known partial structure (Sim, 1964) or from a direct-methods approach (Hauptman, 1982, 1996; Fan *et al.*, 1990).

The usual procedure is based on the observation of Wang (1985) that the map calculated with the average phase values, $\langle\varphi_T\rangle = \varphi_A + 90^\circ$, should reproduce the correct image of the macromolecule superimposed on the featureless noise. This is the basis of the iterative procedure of solvent flattening (or noise filtering; Wang, 1985), which may provide a selection of the proper phase from two alternatives. The whole procedure should be carried out for both enantiomers, since only one of

Table 3
Diffraction data.

Values in parentheses are for the highest resolution range.

Crystal	GI_1.54	Xyl_1.74	Xyl_1.49
Beamline	X9B	X9A	X9B
Protein atoms	3050	2300	2300
S atoms	9†	5	5
Space group	*I*222	$P2_1$	$P2_1$
Unit-cell parameters			
a (Å)	92.90	41.19	41.07
b (Å)	97.95	67.18	67.14
c (Å)	102.71	50.88	50.81
β (°)		113.5	113.5
Wavelength (Å)	1.540	1.743	1.488
Resolution (Å)	20–2.0 (2.07–2.00)	25–1.63 (1.69–1.63)	25–1.75 (1.81–1.75)
f''(S) (e)	0.56	0.70	0.52
$\langle\Delta F^{\rm anom}\rangle/\langle F\rangle$	0.65	0.69	0.56
Total rotation range (°)	767	540 + 437‡	900
Multiplicity	17.4	12.0	15.9
Completeness (%)	100.0 (100.0)	99.8 (99.0)	99.9 (99.8)
$R_{\rm merge}$ (%)	5.0 (10.3)	4.5 (9.1)	4.1 (8.6)
$R_{\rm anom}$ (%)	0.9 (2.1)	1.1 (2.8)	1.3 (3.2)
$I/\sigma(I)$	61.2 (26.4)	48.0 (16.2)	65.3 (24.4)
$B_{\rm Wilson}$ (Å^2)	14.6	11.3	10.6

† The N-terminal methionine is totally disordered. ‡ 437° of high-resolution and 540° of low-resolution data were collected.

them will lead to a correct solution (except when the anomalous substructure itself is centrosymmetric).

Since the anomalous signal within the sulfur SAD data is usually very small, around or below 1%, the diffraction intensities have to be measured as accurately as possible. It has been pointed out that the best way to enhance the data accuracy (apart from following proper experimental procedures) is through averaging multiple measurements within a highly redundant data set (Dauter & Adamiak, 2001; Weiss, Sicker & Hilgenfeld, 2001; Ramagopal *et al.*, 2003). However, this approach is also not unconditionally beneficial owing to the effects of crystal radiation damage. Long irradiation of protein crystals leads to the disruption of disulfide bridges (Weik *et al.*, 2000), as well as other manifestations of non-isomorphism, such as decarboxylation of acidic groups and changes in unit-cell parameters (Burmeister, 2000; Ravelli *et al.*, 2002), and a compromise between the enhanced measurement multiplicity and the increased radiation damage should be achieved. Particular care should be exercised during the scaling and merging procedure; the use of spherical harmonics (Blessing, 1995; Evans, 1997; Otwinowski *et al.*, 2003) usually gives significantly more accurate results.

In the present work, attempts to phase medium-size proteins by S-SAD are presented in which the amount of anomalous signal is significantly lower than 1.0% and is close to 0.6%, a limit formulated by Wang (1985) as a theoretical minimum for successful phasing of protein crystals.

2. Diffraction data

All diffraction data were collected on beamlines X9A and X9B at NSLS (Brookhaven National Laboratory) using Quantum4 ADSC CCD (X9B) or MAR CCD (X9A) detectors and were processed with *HKL*2000 (Otwinowski & Minor, 1997). Glucose isomerase from *Streptomyces rubiginosus* (Carrell *et al.*, 1994) was purchased from Hampton Research, treated with 0.2 *M* EDTA and dialyzed to remove the Mn^{2+} ions naturally present in the enzyme. It was crystallized in the common *I*222 form from a drop containing 15 mg ml^{-1} protein, 0.1 *M* $MgCl_2$, 11% MPD and 0.05 *M* Tris buffer pH 7.0. Crystals of xylanase from *Thermoascus aurantiacus* (Viswamitra *et al.*, 1993; Natesh *et al.*, 2003) were a kind gift from Professor S. Ramakumar (Indian Institute of Science, Bangalore, India). All data were collected from crystals frozen at 100 K in a stream of cold nitrogen gas.

The diffraction data statistics are presented in Table 3. Data for glucose isomerase (hereafter named GI_1.54) and for xylanase with λ = 1.49 Å (Xyl_1.49) were collected in several passes with constant exposure, with the goniostat *κ* angle changed each time by 10°. Each data-collection pass for the xylanase with λ = 1.74 Å (Xyl_1.74) with a particular *κ* orientation was repeated with two different exposure times in order to ensure the proper measurement of the strongest reflections.

To be able to judge the quality of SAD phases, the protein structures were refined against the measured data using *REFMAC* (Murshudov *et al.*, 1999) and *ARP* (Perrakis *et al.*, 1999), starting from the known models of glucose isomerase (PDB code 1xib; Carrell *et al.*, 1994) and xylanase (PDB code 1i1w; Natesh *et al.*, 2003). The final values of $R/R_{\rm free}$ were as follows: 15.0/20.1% for GI_1.5, 13.2/17.4% for Xyl_1.74 and 12.7/17.8% for Xyl_1.49.

3. Anomalous signal

Glucose isomerase, MW 43 kDa, consists of 388 residues, of which eight are methionines and one is cysteine. The

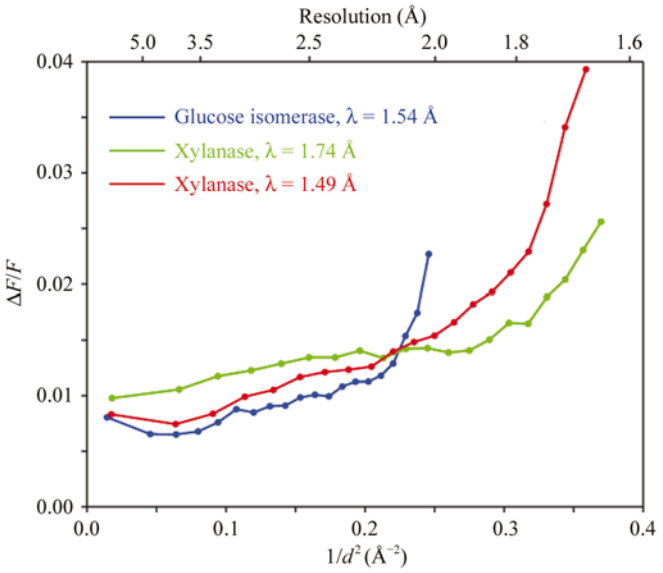

Figure 1
The amount of the anomalous signal in the diffraction data. The Bijvoet ratio $\langle\Delta F^{\rm anom}\rangle/\langle F\rangle$ is plotted as a function of resolution for the three data sets discussed in the text.

N-terminal methionine is disordered. Of 17 structures of this enzyme deposited in the PDB, 11 lack this residue; in the remaining six, all atoms have B factors higher than 60 Å^2. The expected Bijvoet ratio provided by the eight S atoms contributing to the anomalous signal of this protein at 1.54 Å wavelength is 0.60%, according to the formula $\langle \Delta F^{\rm anom} \rangle / \langle F \rangle = (2N_A/N_T)^{1/2}(f''_A/Z_{\rm eff})$ (Hendrickson & Teeter, 1981), where N_A and N_T are the number of anomalous atoms and the total number of atoms in the molecule, f''_A is the imaginary scattering contribution of sulfur (0.556 e) and $Z_{\rm eff}$ is the effective number of electrons of the 'average' protein atom (6.7). Xylanase, MW 30 kDa, has 302 residues with three methionines and two cysteines which form a disulfide bridge. The expected Bijvoet ratio is therefore 0.69% with 1.74 Å wavelength and 0.53% at a wavelength of 1.49 Å. The actual amount of anomalous signal in the diffraction data is shown in Fig. 1 as a function of resolution. At low resolution, where the intensities are strong and measured very accurately, the calculated $\langle \Delta F^{\rm anom} \rangle / \langle F \rangle$ values are close to the expected level and at high resolution they are significantly higher, as a result of the less accurate estimation of the reflection intensities, which is a typical behavior (Dauter *et al.*, 2002).

The probable reason for the lower value of $R_{\rm anom}$ for Xyl_1.74 than for Xyl_1.49 data is similar. In general, the overall quality of the Xyl_1.74 data is higher, with smaller errors in measured intensities; therefore, the value of $R_{\rm anom}$ is lower; that is, closer to the expected level of the anomalous signal. For the Xyl_1.49 data the expected amount of the anomalous signal is smaller, but the data display a higher value of $R_{\rm anom}$ as a result of the relatively less accurate intensity estimations.

The average frequency of sulfur-containing amino acids in 24 bacterial proteomes (Jáuregui *et al.*, 2000) is 1.1% for cysteine and 2.2% for methionine, which gives about 3.3 sulfurs per 100 amino acids and a Bijvoet ratio of about 0.75% for 1.54 Å radiation. The eukaryotic proteomes contain a slightly greater number of Met and Cys residues, *e.g.* 4.42% for *Homo sapiens* (according to the EMBL-EBI web site http://www.ebi.ac.uk/proteome/). Glucose isomerase, with nine (eight useful) sulfur-containing residues among 388 amino acids (2.3%), and xylanase, with five such residues among 302 amino acids (1.7%), are therefore below average in terms of the anomalous scattering signal provided by their S atoms.

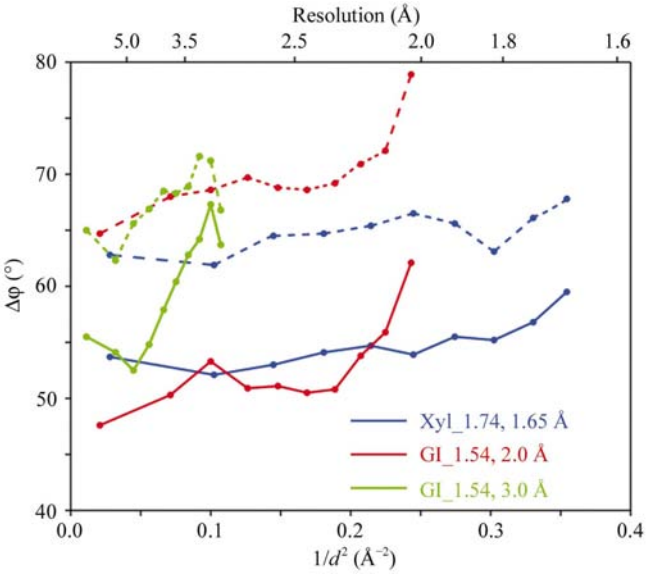

Figure 2
Comparison of phases obtained after initial estimation (dashed lines) and density modification (solid lines) with the refined values for GI_1.54 and Xyl_1.74 data, as a function of resolution.

Table 4
Identification of anomalous scatterers by *SHELXD*.

(*a*) The percentage of successful solutions of sulfur sites in 200 phase trials is given for each data set and each resolution limit. About 1500 reflections were used in each *SHEXLD* run.

	Data set		
Resolution (Å)	GI_1.54	Xyl_1.74	Xyl_1.48
1.63	—	72	—
1.8	—	100	92
2.0	69	100	100
2.5	27	99	99
2.7	65	66	83
3.0	61	96	99
3.5	38	0	0

(*b*) Difference between the positions of S atoms in the refined models and those located by *SHELXD* with different data resolution cutoffs. SG255 and SG261 in xylanase form a disulfide bridge.

Resolution (Å)	1.63	1.8	2.0	2.2	2.5	2.7	3.0	3.5
GI_1_54								
SD84			0.109	0.078	0.100	0.067	0.108	0.437
SD88			0.141	0.104	0.097	0.121	0.125	0.190
SD158			0.076	0.085	0.212	0.183	0.051	0.139
SD223			0.073	0.161	0.112	0.108	0.176	0.080
SG306			0.040	0.031	0.103	0.137	0.136	0.440
SD307			0.114	0.172	0.135	0.091	0.059	0.163
SD370			0.081	0.096	0.176	0.226	0.279	0.054
SD380			0.125	0.154	0.155	0.078	0.163	0.242
Average			0.095	0.110	0.136	0.126	0.137	0.218
Xyl_1.74								
SD49	0.122	0.154	0.206	0.206	0.185	0.126	0.193	
SD108	0.140	0.101	0.127	0.127	0.093	0.182	0.211	
SD116	0.088	0.160	0.126	0.126	0.184	0.092	0.138	
SG255	0.173	0.146	0.185	0.185	0.241	0.092	0.250	
SG261	0.159	0.153	0.277	0.277	0.276	0.459	0.843	
Average	0.136	0.143	0.184	0.184	0.196	0.190	0.327	
Xyl_1_49								
SD49		0.254	0.278	0.284	0.438	0.150	0.250	
SD108		0.147	0.304	0.229	0.156	0.286	0.217	
SD116		0.246	0.112	0.239	0.238	0.291	0.415	
SG255		0.146	0.594	0.458	0.217	0.172	0.415	
SG261		0.237	0.097	0.332	0.528	0.647	Wrong	
Average		0.206	0.277	0.308	0.315	0.309	0.324	

4. Identification of sulfurs

For all three cases, the sulfur sites were located with *SHELXD* (Schneider & Sheldrick, 2002) from about 1500 largest Bijvoet differences prepared from the raw data by *XPREP* (Bruker Analytical X-ray Systems). The *SHELXD* jobs were run with data truncated to different resolution limits and the success

rates of these trials are presented in Table 4. The rate of successful solutions is generally high, although it varies depending on the data resolution cutoff and the number of used reflections; the reasons for this effect remain unclear. This follows the typical behavior of direct methods, which are often sensitive to small changes in the input parameters.

This table also shows the average distance between the refined sulfur coordinates and those given by *SHELXD*. It is somewhat surprising that both sulfurs within the single disulfide bridge in xylanase could be identified at a resolution as low as 3.0 Å, although with low accuracy. Nevertheless, the ability to locate two sulfur sites within a disulfide bond of 2.03 Å length at a modest resolution provides a valuable check of the correctness of the obtained substructure solution. In general, the success rate of finding the anomalous scattering sites was very high, especially for both xylanase data sets at the medium resolution of about 2 Å, where it reaches 100%. Furthermore, it should be noted that the procedure used to remove the Mn^{2+} ions naturally present in glucose isomerase was highly successful and the manganese position did not appear among the first ten sites in any of the solutions obtained with different resolution cutoffs.

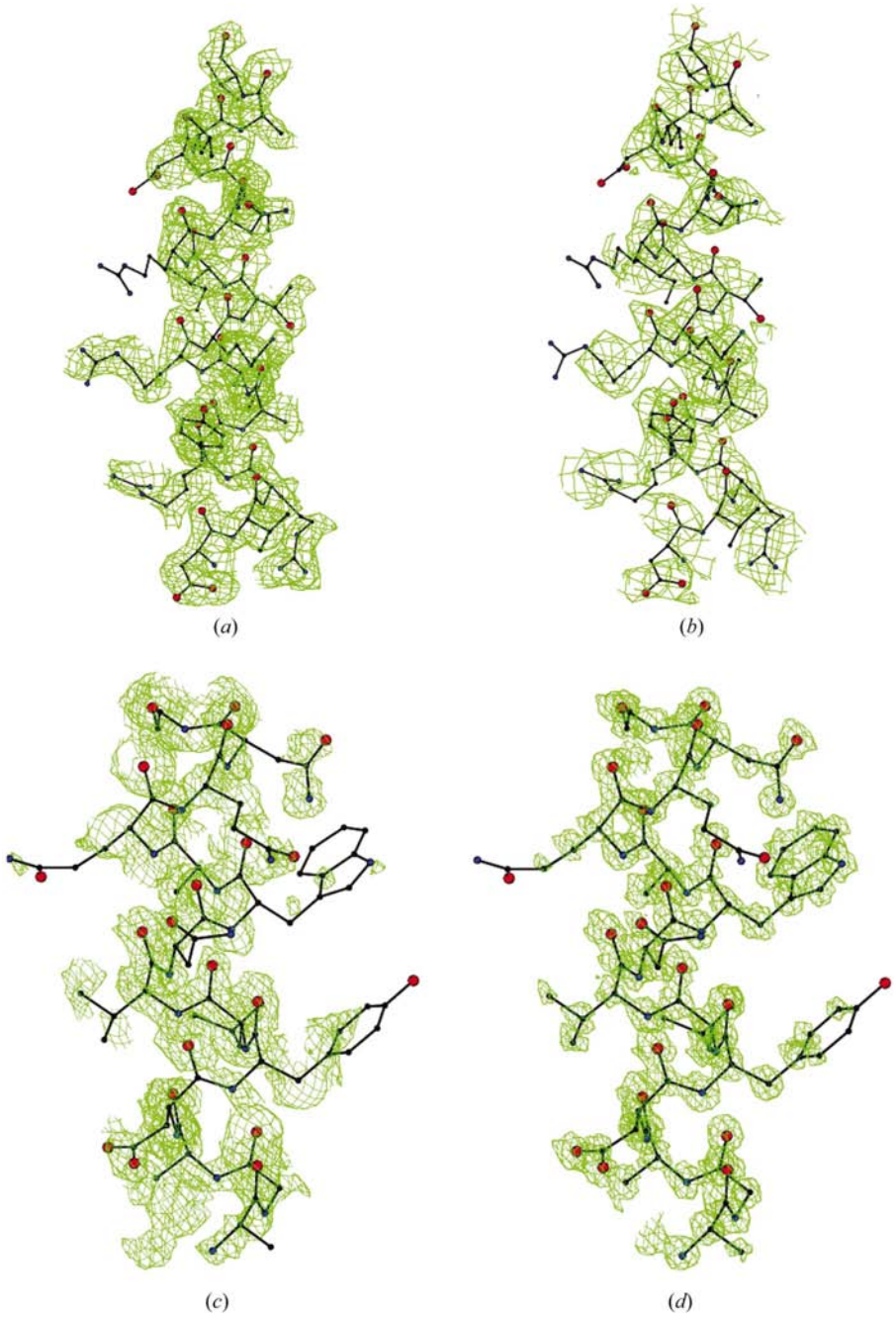

Figure 3
Representative regions of the electron-density maps after density modification. (*a*) GI_1.54 at 2.0 Å, (*b*) GI_1.54 at 3.0 Å, (*c*) Xyl_1.74 at 1.65 Å and (*d*) Xyl_1.49 extended to atomic resolution of 0.89 Å. The maps in (*a*), (*b*) and (*c*) are contoured at the 1σ level; map (*d*) is contoured at 2σ to better show its atomicity. Several side chains in these regions extend into the solvent region and their density is suppressed by the solvent flattening.

5. SAD phasing and automated model building

The protocols and programs used to obtain the protein phases were different in all three cases and are described separately below. The statistics of the experimental phases are shown in Fig. 2.

5.1. GI_1.54

All eight S atoms expected as anomalous scatterers in glucose isomerase were found by *SHELXD*; in addition, two weaker peaks appeared consistently in all substructure solutions. These anomalous scatterers were assumed to be chloride ions from the crystallization solvent and were treated as S atoms for the purpose of phasing. All data to 2.0 Å resolution and ten anomalous scatterer positions were input into *SHELXE* (Sheldrick, 2002), as well as *MLPHARE* (Otwinowski, 1991), followed by *DM* (Cowtan & Zhang, 1999). The electron-density maps obtained from *SHELXE* were not interpretable. However, *MLPHARE* phases after density modification with *DM* resulted in an interpretable map. This map was submitted to the automatic model-building program *ARP/wARP* (Perrakis *et al.*, 1999) and 378 residues of the total of 388 protein amino acids were built by the iterative free-atom density-modification and model-building procedure. Considering the success of substructure location at as low as 3.5 Å resolution, phasing with *MLPHARE* was also attempted with

data truncated to 2.5 Å as well as 3.0 Å. In both cases *RESOLVE* (Terwilliger, 2000) was able to improve the phases and build automatically more than 200 residues, mainly in the helical regions, with many other parts of the structure clearly manually interpretable. The electron density obtained at 2.0 and 3.0 Å is shown in Figs. 3(*a*) and 3(*b*).

5.2. Xyl_1.74

Xylanase has only five sulfurs in its 303 residues and a rather low solvent content of about 37%. However, the crystals diffracted very well, the data were merged from the highly redundant set of measurements and the anomalous substructure determination showed a very high success rate (Table 4). Several attempts to obtain the protein phases using five sulfur positions with the programs *MLPHARE*, *SHARP* (La Fortelle & Bricogne, 1997), *OASIS* (Hao *et al.*, 2000) and *SOLVE* (Terwilliger & Berendzen, 1999) run with default parameters and subsequent density modification were unsuccessful. The phasing and density modification with *SHELXE* produced a reasonably good map, in which considerable parts of the density could be interpreted. Owing to the low solvent content the overall phase improvement with density modification was modest, about 10°, whereas the phase improvement for GI_1.54 was about 20° (Fig. 2) where the solvent content was 55%. Phasing attempts at lower resolution followed by the extension of phases to 1.65 Å did not improve the resulting maps. Attempts to build the model in the *SHELXE* map using *ARP*/*wARP* were unsuccessful. However, density modification and automated interpretation with *RESOLVE* was able to build more than 150 residues in the map. The average difference between the *SHELXE* and the refined phases was about 55°, implying that a slight improvement of the density-modified phases might be sufficient to build the complete model in an automatic manner. Tuning some of the *SHELXE* input parameters, such as reduction of the solvent fraction to 33% and changing the perturbation γ value to 2.0 from the default of 1.0, produced a phase set on average closer by about 3–4° to the refined phases and a clearer electron-density map. With this map, *ARP*/*wARP* built 302 out of the total of 303 residues. A representative region of the electron-density maps at several stages of phasing is shown in Fig. 3(*c*).

5.3. Xyl_1.49

As can be seen from Table 1, this data set has the lowest amount of anomalous signal ever used for SAD phasing, including the other two cases described above. The substructure solution was again surprisingly successful (Table 4). However, the very small anomalous differences lacked the sufficient phasing power to produce interpretable electron-density maps with this data alone after a number of attempts with various programs and protocols. Since crystals of xylanase diffract to atomic resolution (Natesh *et al.*, 2003), the 0.89 Å data (kindly supplied by Professor S. Ramakumar, Bangalore) were used as a 'native' set, together with the Xyl_1.49 anomalous data and five sulfur positions found by *SHELXD* for phase estimation and extension to atomic resolution with *SHELXE*. This procedure led to the electron-density map illustrated in Fig. 3(*d*). The quality of this map does not require further statistical support and could be easily interpreted automatically with any available model-building program.

5.4. Phasing on weak anomalous signal

The S-SAD results described here are based on data from proteins with low sulfur content. It is clear from the above cases that the solvent fraction is one of the important factors, particularly in S-SAD phasing, where not only the refinement of phases but also the removal of phase ambiguity is required without any possibility of additional measurements of the data. This may explain why obtaining the interpretable map from the relatively high resolution Xyl_1.74 data was more difficult than from the GI-1.54 data, where reasonable protein phases could be obtained even when the resolution was truncated to 3.0 Å. However, generally more accurate and higher resolution diffraction data can be collected from tightly packed crystals with low solvent content and such data should theoretically produce good-quality raw experimental phases and may not require a very significant improvement by solvent flattening. On the other hand, the high solvent content makes the density-modification process more effective and may significantly improve even relatively poor phases. Non-crystallographic symmetry, if present, may also significantly enhance the process of phase refinement. A longer wavelength, in the vicinity of 2 Å, should provide a greater anomalous signal originating from sulfur. A significant proportion of protein crystals, which generally have a higher sulfur content than in the examples discussed here, may be expected to be solvable by the S-SAD phasing.

6. Conclusions

Simulations based on the error-free structure factors calculated for a 12 kDa protein containing one disulfide bridge led Wang (1985) to conclude that an anomalous signal of 0.6% of the total diffraction was sufficient for a successful structure solution with the SAD approach. The data from crystals of glucose isomerase and xylanase contain the lowest level of the anomalous scattering signal ever used for the successful SAD phasing based on experimentally collected diffraction data and confirm that Wang's conjecture was indeed realistic.

If an anomalous signal weaker than 1% is to be used for the estimation of phases, the errors in the estimation of intensities, or at least of the significant part of them, cannot on average exceed 2% (since $\sigma I/I \simeq 2\sigma F/F$) and in practice should be considerably lower. Such a level of accuracy can be achieved only at well performing data-collection facilities with crystals diffracting strongly enough and by averaging multiple measurements of the same or symmetry-equivalent intensities. The requirement of a well conducted experiment is obvious, since any X-ray beam instabilities, crystal misalignment, bad beam–shutter synchronization, deficient detector calibration

etc. will lead to systematic errors, precluding the appearance of such a sub-percent amount of anomalous signal.

The necessity for strong diffraction results from the principles of counting statistics, since the standard uncertainty is expected to be equal to the square root of the number of counts, $\sigma(I) = I^{1/2}$. To achieve a 1% accuracy in a single measurement, the intensity must therefore correspond to at least 10 000 X-ray counts. This is somewhat complicated by the fact that two-dimensional detectors based on CCDs or image plates do not count individual X-ray quanta; nevertheless, the detector 'gain factor' can be taken into account in the estimation of the measured diffracted intensities and their uncertainties. It can be stated qualitatively that the stronger the measured intensity (but remaining below the detector saturation limit) is, the more accurately it can be estimated.

The improvement of the data accuracy resulting from multiple measurements is again a consequence of the rule of statistics stating that the variance of the sum equals the sum of variances. Therefore, for the individual intensity measured N times and averaged, the uncertainty diminishes $N^{1/2}$ times. However, the gain in the accuracy of the intensity estimation obtained by increased redundancy of measurements may be neutralized by the effect of radiation damage and the resulting non-isomorphism after exceedingly long total crystal exposure, particularly at the brightest synchrotron X-ray sources. This effect requires some compromises to be made during diffraction experiments between the data multiplicity, the resolution limit and the total exposure.

The examples of successful SAD phasing based on the signal of weak anomalous scatterers, such as S, P or Cl (Table 1), including the present results, prove that even the very weak (below 1%) anomalous signal provided by these elements, naturally occurring in macromolecules, can be successfully used for solving crystal structures. Recent advances in macromolecular crystallography, both in hardware development and in crystallographic computing, have made the application of this approach quite easy in practice and often less complicated than the classic MIR or MAD experiments. It may be expected that the sulfur SAD method will soon gain wider popularity. This method is also applicable to high-throughput projects.

Diffraction data are available upon request from ZD.

We thank R. Thirumuruhan for help with collecting data at beamline X9A.

References

Blessing, R. H. (1995). *Acta Cryst.* A**51**, 33–38.

Bond, C. S., Shaw, M. P., Alphey, M. S. & Hunter, W. N. (2001). *Acta Cryst.* D**57**, 755–758.

Brown, J., Esnouf, R. M., Jones, M. A., Linnell, J., Harlos, K., Hassan, A. B. & Jones, E. Y. (2002). *EMBO J.* **21**, 1054–1062.

Burmeister, W. P. (2000). *Acta Cryst.* D**56**, 328–341.

Carrell, H. L., Hoier, H. & Glusker, J. P. (1994). *Acta Cryst.* D**50**, 113–123.

Cianci, M., Rizkallah, P. J., Olczak, A., Raftery, J., Chayen, N. E., Zagalsky, P. F. & Helliwell, J. R. (2001). *Acta Cryst.* D**57**, 1219–1229.

Cowtan, K. D. & Zhang, K. Y. J. (1999). *Prog. Biophys. Mol. Biol.* **72**, 245–270.

Cromer, D. T. (1983). *J. Appl. Cryst.* **16**, 437–438.

Dauter, Z. & Adamiak, D. A. (2001). *Acta Cryst.* D**57**, 990–995.

Dauter, Z., Dauter, M. & Dodson, E. J. (2002). *Acta Cryst.* D**58**, 494–506.

Dauter, Z., Dauter, M., La Fortelle, E. de, Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.

Debreczeni, J. E., Bunkoczi, G., Girmann, B. & Sheldrick, G. M. (2003). *Acta Cryst.* D**59**, 393–395.

Einspahr, H., Suguna, K., Suddath, F. L., Ellis, G., Helliwell, J. R. & Papiz, M. Z. (1985). *Acta Cryst.* B**41**, 336–341.

Evans, P. R. (1997). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **33**, 22–24.

Fan, H. F., Hao, Q., Gu, Y. X., Qian, J. Z., Zheng, C. D. & Ke, H. (1990). *Acta Cryst.* A**46**, 935–939.

Gordon, E. J., Leonard, G. A., McSweeney, S. & Zagalsky, P. F. (2001). *Acta Cryst.* D**57**, 1230–1237.

Graaff, R. A. G. de, Hilge, M., van der Plaas, J. L. & Abrahams, J. P. (2001). *Acta Cryst.* D**57**, 1857–1862.

Greenwald, J., Fischer, W. H., Vale, W. W. & Choe, S. (1999). *Nature Struct. Biol.* **6**, 18–22.

Hao, Q., Gu, Y. X., Zheng, C. D. & Fan, H. F. (2000). *J. Appl. Cryst.* **33**, 980–981.

Hauptman, H. A. (1982). *Acta Cryst.* A**38**, 289–294.

Hauptman, H. A. (1996). *Acta Cryst.* A**52**, 490–496.

Helliwell, J. R. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 80–88. Warrington: Daresbury Laboratory.

Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.

Jáuregui, R., Bolivar, F. & Merino, E. (2000). *Microb. Comput. Genomics*, **5**, 7–15.

Kwiatkowski, W., Noel, J. P. & Choe, S. (2000). *J. Appl. Cryst.* **33**, 876–881.

La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.

Lemke, C. T., Smith, G. D. & Howell, P. L. (2002). *Acta Cryst.* D**58**, 2096–2101.

Li, S., Finley, J., Liu, Z.-J., Qiu, S.-H., Chen, H., Luan, C.-H., Carson, M., Tsao, J., Johnson, D., Lin, G., Zhao, J., Thomas, W., Nagy, L. A., Sha, B., DeLucas, L. J., Wang, B.-C. & Luo, M. (2002). *J. Biol. Chem.* **277**, 48596–48601.

Liu, Y., Ogata, C. M. & Hendrickson, W. A. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 10648–10653.

Liu, Z.-J., Vysotski, E. S., Chen, C.-J., Rose, J. P., Lee, J. & Wang, B.-C. (2000). *Protein Sci.* **9**, 2085–2093.

Loll, P. J. (2001). *Acta Cryst.* D**57**, 977–980.

Micossi, E., Hunter, W. N. & Leonard, G. A. (2002). *Acta Cryst.* D**58**, 21–28.

Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.

Natesh, R., Manikandan, K., Bhanumoorthy, P., Viswamitra, M. & Ramakumar, S. (2003). *Acta Cryst.* D**59**, 105–117.

Otwinowski, Z. (1991). *Proceedings of the CCP4 Study Weekend. Isomorphous Replacement and Anomalous Scattering*, edited by W. Wolf, P. R. Evans & A. G. W. Leslie, pp. 80–86. Warrington: Daresbury Laboratory.

Otwinowski, Z., Borek, D., Majewski, W. & Minor, W. (2003). *Acta Cryst.* A**59**, 228–234.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.

Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003). *Acta Cryst.* D**59**, 868–875.

Ravelli, R. B., Theveneau, P., McSweeney, S. & Caffrey, M. (2002). *J. Synchrotron Rad.* **9**, 355–360.

Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.
Sheldrick, G. M. (2002). *Z. Kristallogr.* **217**, 644–650.
Sim, G. A. (1964). *Acta Cryst.* **17**, 1072–1073.
Stuhrmann, S., Bartels, K. S., Braunwarth, W., Doose, R., Davergne, F., Gabriel, A., Knöchel, A., Marmotti, M., Stuhrmann, H. B., Trame, C. & Lehmann, M. S. (1997). *J. Synchrotron Rad.* **4**, 298–310.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Viswamitra, M. A., Bhanumoorthy, P., Ramakumar, S., Manjula, M. V., Vithayathil, P. J., Murthy, S. K. & Naren, A. P. (1993). *J. Mol. Biol.* **232**, 987–988.
Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
Weik, M., Ravelli, R. B., Kryger, G., McSweeney, S., Raves, M. L., Harel, M., Gros, P., Silman, I., Kroon, J. & Sussman, J. L. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 623–628.
Weiss, M. S., Sicker, T., Djinovic Carugo, K. & Hilgenfeld, R. (2001). *Acta Cryst.* D**57**, 689–695.
Weiss, M. S., Sicker, T. & Hilgenfeld, R. (2001). *Structure*, **9**, 771–777.
Yang, C. & Pflugrath, J. W. (2001). *Acta Cryst.* D**57**, 1480–1490.