Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

De novo phasing of two crystal forms of tryparedoxin II using the anomalous scattering from S atoms: a combination of small signal and medium resolution reveals this to be a general tool for solving protein crystal structures

Elena Micossi,[a] William N. Hunter[b] and Gordon A. Leonard[a]*

[a]Macromolecular Crystallography, European Synchrotron Radiation Facility, BP 220, F-38043 Grenoble CEDEX, France, and [b]School of Life Sciences, The Wellcome Trust Building, University of Dundee, Dundee DD1 5EH, Scotland

Correspondence e-mail: leonard@esrf.fr

Received 20 July 2001
Accepted 9 October 2001

The *de novo* phasing of the structures of two crystal forms of tryparedoxin II from *Crithidia fasciculata* has been carried out using single-wavelength anomalous diffraction techniques exploiting only the small anomalous signal from the S atoms intrinsic to the native protein. Data were collected at 1.77 Å wavelength, where the Bijvoet ratio is approximately 1.2%. Data collected to $d_{\min}$ = 2.5 Å from a crystal of form I, which has a diffraction limit of $d_{\min}$ = 1.5 Å and a solvent content of ∼46%, produced readily interpretable electron-density maps. When these phases were extended to the resolution limit of the crystals, almost the entire model could be traced automatically. Crystals of form II have a much higher solvent content, ∼72%, and a much lower diffraction limit than form I and at 1.77 Å wavelength yielded data only to $d_{\min}$ = 2.7 Å. Despite the medium resolution of the data for this crystal form, it was possible both to determine the heavy-atom partial structure and then use it to produce, still at $d_{\min}$ = 2.7 Å, an excellent quality interpretable electron-density map. This was then improved by phase extension to the $d_{\min}$ = 2.35 Å diffraction limits of a different crystal for which data were collected on a more intense beamline. The success of this latter structure solution markedly increases the potential use in macromolecular crystal structure determination of the anomalous signal available from S atoms that occur naturally in proteins and, as is discussed, has significant implications for structure determination in the high-throughput era.

1. Introduction

The advent of structural genomics and the consequent necessity for high-throughput structural biology means that the demand for the *de novo* phasing of macromolecular crystal structures will increase. Experimental phases have traditionally resulted from isomorphous replacement (IR) (Blundell & Johnson, 1976). More recently, with the greater accessibility of tuneable synchrotron beamlines, anomalous diffraction (AD) techniques using either multi-wavelength (MAD; Smith, 1991; Hendrickson, 1991, 1999) or single-wavelength data (SAD; Rice *et al.*, 2000) have also become commonplace. Both methods usually require that heavy atoms are added or substituted into crystals of a native protein. However, IR experiments, even when combined with AD techniques (see Blundell & Johnson, 1976; Helliwell, 1992; Fourme *et al.*, 1999 for details of these methods), can be fraught with difficulty. Problems such as non-isomorphism, deterioration of diffraction quality and non-specific binding of the heavy atoms are

Table 1
Data-collection statistics.

(*a*) Crystal form I. Values in parentheses refer to the highest resolution shell (2.54–2.50 and 1.53–1.50 Å, respectively).

Beamline	BM14, ESRF	BM14, ESRF
Wavelength (Å)	1.777	0.977
Unit-cell parameters (Å)	$a = b = 111.2$, $c = 56.3$	$a = b = 111.74$, $c = 56.55$
Resolution limits (Å)	40.0–2.5	40.0–1.5
Unique reflections	23393†	57687
Reflections measured	311699†	469429
Completeness (%)	99.8 (97.4)†	99.7 (95.5)
$\langle I\rangle/\langle\sigma(I)\rangle$	59.4 (16.8)†	37.6 (2.1)
$R_{\rm sym}$ (%)	5.2 (13.4) †	4.5(35.4)

(*b*) Crystal form II. Values in parentheses refer to the highest resolution shell (2.76–2.70 and 2.39–2.35 Å, respectively).

Beamline	BM14 CRG, ESRF	ID14-2, ESRF
Wavelength (Å)	1.777	0.933
Unit-cell parameters (Å)	$a = b = 114.2$, $c = 102.6$	$a = b = 114.29$, $c = 101.97$
Resolution limits (Å)	30.0–2.7	40.0–2.35
Unique reflections	35567†	28025
Reflections measured	531442†	136117
Completeness (%)	100 (100)†	97.5 (99.9)
$\langle I\rangle/\langle\sigma(I)\rangle$	54.0 (9.5)†	29.8 (9.6)
$R_{\rm sym}$ (%)	5.8 (29.9)†	5.3 (24.9)

† With I_{hkl} and $I_{\overline{hkl}}$ scaled and merged separately.

commonplace, particularly if the heavy atom is introduced *via* soaking or co-crystallization experiments. The use of AD methods alone is generally considered less problematic. Non-isomorphism is not an issue, as all intensity data are generally collected from the same crystal. Nevertheless, despite some spectacular successes (Deacon *et al.*, 2000; Alphey *et al.*, 2000) the preparation of crystals containing selenomethionine – the most common method of introducing heavy atoms for AD methods – is not always straightforward. In some cases, particularly with eukaryotic or mammalian proteins, the polypeptide may not express or fold properly and not every selenomethionine-containing protein will crystallize as readily as its native counterpart.

The macromolecular crystallography community is therefore increasingly interested in methods of structure solution which require neither the chemical modification of macromolecules nor the introduction of very heavy atoms into crystals. Recent reports (Dauter *et al.*, 2000, 2001; Nagem *et al.*, 2001, Korolev *et al.*, 2001) have indicated that short cryo-soaking of protein crystals with halides or monovalent cations can provide experimental phase information for macromolecular crystal structures. Indeed, it has even been suggested that this could become a general method for structure solution. This method requires that crystals are soaked in high concentrations (0.5–1.0 *M*) of halide ions and is unfortunately not foolproof. It is not always the case that the halide ions are sufficiently ordered to lead to structure solution (Dauter *et al.*, 2001) and there has been at least one report (Korolev *et al.*, 2001) that soaking crystals in 1.0 *M* NaBr results in a substantial lowering of their diffraction quality.

Much better, therefore, would be methods to derive experimental phase information from unmodified crystals of the unmodified protein. Unfortunately, the most obvious technique for this, direct methods, is of limited use in macromolecular crystallography. One of the underlying concepts of the technique is that of atomicity. This requires the collection of near-atomic resolution diffraction data, a feat which is still relatively rare in macromolecular crystallography, and the apparent current upper limit of $d_{\rm min} = 1.2$ Å for the technique to be successful means that it can be applied to only a small fraction of macromolecular structures.

However, the vast majority of proteins and all nucleic acids contain atoms (sulfur and phosphorus, respectively) which have significant anomalous scattering properties. It has been shown in an increasing number of cases (Hendrickson & Teeter, 1981; Wang, 1985; Dauter *et al.*, 1999; Lui *et al.*, 2000; Bond *et al.*, 2001; Gordon *et al.*, 2001) that the presence of sulfur can be exploited to derive *de novo* experimental phases for macromolecular crystal structures without either the need for incorporation of heavy atoms or, in most cases, the measurement of atomic resolution data. In these reports, small anomalous signals resulting from the presence of S atoms in the crystals were measured using X-rays of either 1.54 or 1.77 Å wavelength. These were used to determine the S-atom partial structure using either direct or Patterson methods and thus to initiate phasing, usually *via* SAD techniques. Although experimental phases were in some cases initially derived at relatively low resolution, in all cases the crystals diffracted to much better than $d_{\rm min} = 2$ Å and the method would thus appear to require crystals that diffract well. It is also unclear just what the limits are on the size of anomalous signal that can successfully be used in this procedure. It should now be asked just how general S-SAD (sulfur single-wavelength anomalous dispersion) can become and what the limits of the technique are.

To test the limits of S-SAD, we have attempted the *de novo* phasing of two crystal forms of tryparedoxin II (Trx II) from *C. fasciculata*. Although it has been possible to solve both

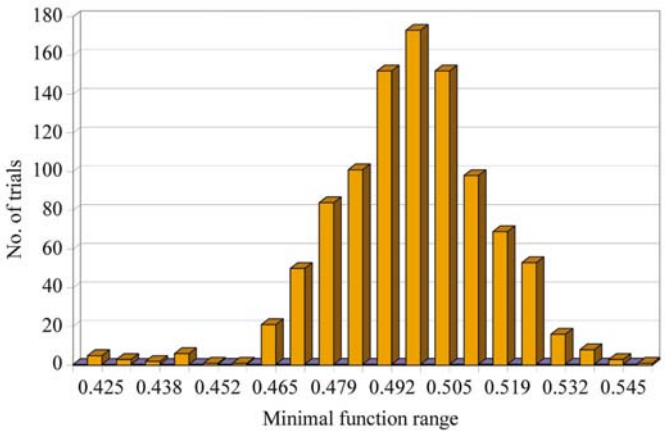

Figure 1
The histogram of the minimal function $R_{\rm min}$ resulting from the *SnB* search for the S-atom partial structure using data collected at 1.77 Å wavelength from crystal form I. The bimodal distribution is characteristic of a successful result of the search.

crystal forms using either molecular replacement (Micossi, Leonard & Hunter, unpublished work) or, in the case of crystal form I described below, isomorphous replacement (Hofmann *et al.*, 2001), the structures represent an excellent test for the generality of the S-SAD technique. The asymmetric units of both crystal forms consist notionally of 332 residues, of which 16 (*i.e.* ∼5%) are either cysteine or methionine. The maximum Bijvoet difference ratio ($\langle \Delta F/F \rangle$) available at 1.77 Å, as calculated according to (1) (Hendrickson *et al.*, 1985), where N_A and N_T are the number of anomalous scattering atoms and the total number of atoms in the asymmetric unit, respectively, and $Z_{\rm eff}$ is the effective atomic scattering factor (6.7 for proteins), is therefore approximately 1.2%,

$$\langle \Delta F/F \rangle = (N_A/2N_T)^{1/2}(2f''_A/Z_{\rm eff}). \quad (1)$$

The signal available in the experiments we describe here is thus somewhat smaller than that (≥1.5%) which allowed the

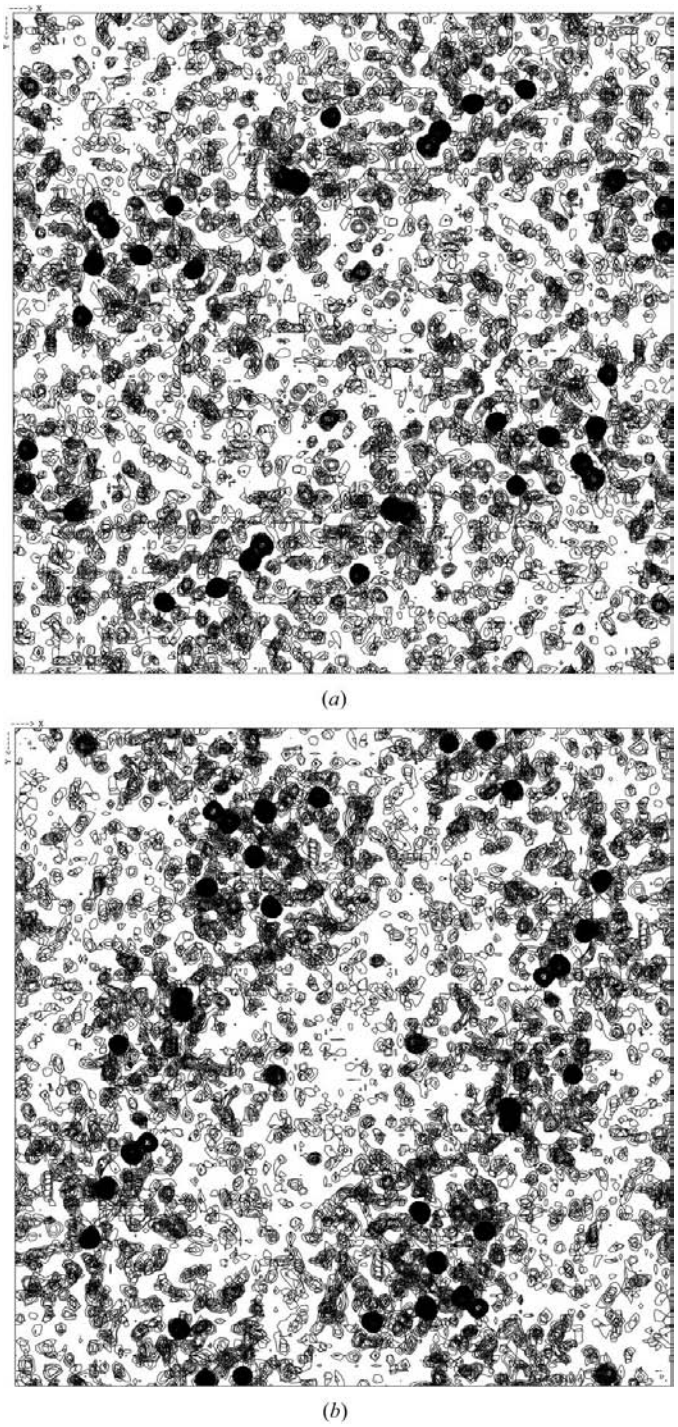

Figure 2
The electron-density maps for crystal form I using the raw experimental phases ($\alpha_{\rm S\text{-}SAD}$, $\rm FOM_{S\text{-}SAD}$) at 2.5 Å resolution for (*a*) the wrong hand and (*b*) the correct hand of the sulfur partial substructure model. Although both maps are dominated by density at the positions of the S atoms used to derive the phase-probability distributions, the protein–solvent boundary is clearly more evident when the hand of the substructure model is correct.

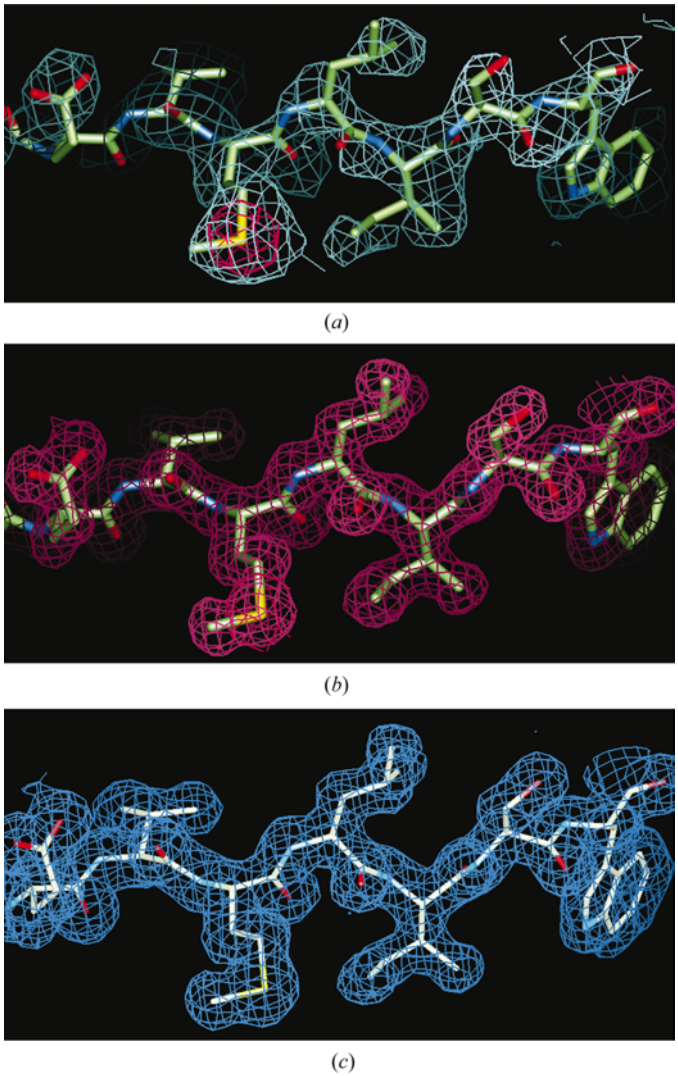

Figure 3
(*a*) A part of the electron-density map (cyan chicken wire) for crystal form I calculated using solvent-flattened phases at $d_{\rm min}$ = 2.5 Å. No NCS averaging was used in the phase-improvement procedure that produced this electron-density map. (*b*) The same region of the electron-density map (purple chicken wire) calculated at $d_{\rm min}$ = 1.5 Å after the phase-extension procedure described in the text. (*c*) The same region of the electron-density map (blue chicken wire) calculated at $d_{\rm min}$ = 1.5 Å after the refinement procedure. In all figures, the model in its current state of refinement has been superimposed on the electron density.

de novo solutions of the crystal structures of crambin and lysozyme using data collected at 1.54 Å wavelength (Hendrickson & Teeter, 1981; Dauter *et al.*, 1999), but similar in magnitude to that used by Wang and coworkers to solve the structure of obelin using data collected at λ = 1.74 Å (Lui *et al.*, 2000). The two crystal forms of Trx II diffract to $d_{\min}$ = 1.5 Å and $d_{\min}$ = 2.7 Å, respectively, and, as will be described, both experiments produced extremely high-quality electron-density maps. The success of the latter experiment extends the resolution limits of the technique markedly.

2. Experimental

2.1. Protein preparation, purification and crystallization

The protein was prepared in a similar manner to that previously described for tryparedoxin I (Alphey *et al.*, 1999). Crystals were grown using the vapour-diffusion method in hanging drops at 277 K. The drops consisted of a 2:1 protein:precipitant ratio and the reservoirs contained 500 µl of the precipitant solution. Rod-shaped crystals (form I) grew in conditions where the precipitant was 30% PEG 8000, 100 m*M* sodium cacodylate pH 6.5, 200 m*M* ammonium sulfate. Tetragonal bipyramids (form II) resulted from a precipitant solution consisting of 1.4 *M* sodium citrate, 100 m*M* Na HEPES pH 7.5. In both cases, diffraction-quality crystals appeared after approximately one week.

2.2. Data collection and processing

2.2.1. Crystal form I. These crystals are tetragonal with space group $P4_22_12$. A single crystal was cryoprotected in a solution consisting of the crystallization mother liquor and 15%(*v*/*v*) glycerol and then flash-frozen by plunging it into liquid nitrogen. It was then mounted on beamline BM14 of the ESRF and held at a temperature of 100 K in a nitrogen-gas stream. Two data sets were then collected from the same crystal. The wavelength was set to λ = 1.77 Å and 360° of data were collected using a MAR CCD-130 as the detector. Hardware constraints meant that at this wavelength the maximum resolution of the data that could be collected was 2.5 Å. No attempt was made to align the crystal and measure Bijvoet related reflections on the same or even neighbouring images. Data were integrated with *DENZO* and scaled with *SCALEPACK* (Otwinowski & Minor, 1997), where the *hkl* and $\overline{hkl}$ reflections in an anomalous pair were treated as separate reflections both during scaling and merging. In order to collect data for structure refinement, the wavelength was set to 0.977 Å and a data set with $d_{\min}$ = 1.5 Å was measured from the same crystal. To ensure proper completeness, these high-resolution data were collected in two passes. As for the data set collected at λ = 1.77 Å, the data were integrated with *DENZO* and scaled with *SCALEPACK*, this time in the standard fashion. Further details are shown in Table 1(*a*).

2.2.2. Crystal Form II. This crystal form adopts the space group $P4_12_12$. A single crystal was flash-frozen in a similar fashion to crystal form I, except that the cryoprotectant consisted of the crystallization mother liquor and 10%(*v*/*v*) glycerol. As for crystal form I, it was then mounted on BM14. Data for S-SAD phasing were then collected and processed as for crystal form I, with the only difference being that data were collected to $d_{\min}$ = 2.7 Å, the diffraction limit of the crystals. In order to obtain higher resolution data for refinement, a second crystal was then mounted on the ESRF undulator beamline ID14-EH2, a much more intense beamline than BM14. Here, diffraction data to $d_{\min}$ = 2.35 Å were collected and processed using *DENZO* and *SCALEPACK* in the usual manner. Full details are given in Table 1(*b*).

For all data sets, mean structure-factor amplitudes and anomalous differences were derived using the program *TRUNCATE* (Collaborative Computational Project, 1994; French & Wilson, 1978).

2.3. Determination of S-atom partial structure, phase determination, model building and refinement

2.3.1. Crystal form I. For this crystal form, the Matthews coefficient for two molecules in the asymmetric unit is 2.34 Å^3 Da^{-1} and the resulting solvent content is 46.5%. Hence, a S-atom partial structure determination searching for a total of 14 sites (it was assumed that the two N-terminal methionines would be disordered) was carried out using the program *SnB* (Howell *et al.*, 2000) in a similar fashion to that described previously for the C1 subunit of α-crustacyanin (Gordon *et al.*, 2001). The minimum acceptable diffE/σ(diff$E$) was reduced to 2.0 to allow the phasing of 420 reflections from which triplet relationships could be derived. At the end of the *SnB* procedure, the distribution of the minimum function $R_{\min}$ is bimodal (Fig. 1), with a small subset of trials having values of $R_{\min}$ clearly lower than those for the large majority of trials, for which the distribution is centred around $R_{\min}$ = 0.5. This bimodal nature of the distribution of $R_{\min}$ values indicates that the partial structure determination had been successful (Miller *et al.*, 1994; Howell *et al.*, 2000). As a result, the top 12 putative S-atom positions from the trial giving the lowest value of the minimum function ($R_{\min}$ = 0.425) were then input into the program *SHARP* (La Fortelle & Bricogne, 1997) for refinement and phase-probability distribution calculation using the anomalous differences in the 1.77 Å wavelength data. This initial model of the partial structure was then modified based on the appearance of log-likelihood-based residual maps (La Fortelle & Bricogne, 1997), with the final model consisting of 13 atoms, of which four were cysteines involved in two disulfide bridges and the rest were methionines. A final round of *SHARP* was then carried out testing both hands of the partial structure with, in the correct case, a much clearer protein–solvent boundary visible in weighted electron-density maps calculated with the raw experimental phases ($\alpha_{\text{S-SAD}}$, $\text{FOM}_{\text{S-SAD}}$) (Fig. 2) at 2.5 Å resolution. The overall figure of merit at the end of the *SHARP* procedure was 0.33 for all reflections in the range 40.0–2.5 Å.

These experimental phases were then improved by carrying out one round of solvent flattening using the program *SOLOMON* (Abrahams & Leslie, 1996) as incorporated in the *SHARP* suite. The resulting electron-density maps were

readily interpretable (Fig. 3*a*). However, to take advantage of the high-resolution ($d_{\rm min}$ = 1.5 Å) data available, phase extension was carried out as follows: the S-atom partial structure was used to calculate an approximate non-crystallographic symmetry (NCS) relationship for the two molecules in the asymmetric unit. An approximate molecular mask for a monomer was then constructed by assuming the radius of each S atom to be 12 Å and then removing any consequent overlap between symmetry- and NCS-related masks. This mask and the NCS relationships derived above were then used in conjunction with solvent flattening and histogram matching to extend the 2.5 Å experimental phases to the limit of the high-resolution data using the program *DM* (Cowtan, 1994). The resulting figure of merit was 0.693 and the electron-density map at this stage was of excellent quality (Fig. 3*b*). The major part of the atomic model for the protein dimer was built automatically using the program *wARP* (Perrakis *et al.*, 1999). Further refinement and manual rebuilding of the model was carried out using *CNS* (Brunger *et al.*, 1998) and *QUANTA* (Accelrys Inc.; http://www.accelrys.com). The model in its current state of refinement results in R = 20.4% and $R_{\rm free}$ = 22.5% (5% of data) for all data in the resolution range 40–1.5 Å and contains 292 amino-acid residues (residues 14–164 for molecule *A*, 18–160 for molecule *B*), one sulfate ion and 337 water molecules. An example of the electron density at the end of the refinement procedure is shown in Fig. 3(*c*).

2.3.2. Crystal form II. For this crystal form, the Matthews coefficient for two molecules in the asymmetric unit is 4.4 Å^3 Da^{-1} and the resulting solvent content is 72.2%. These values are at the higher extremes usually found for protein crystals but are consistent with the markedly reduced diffracting power compared with crystal form I. Hence, it was thought that the number of S atoms to be found to initiate phasing would be similar to that for form I. Several attempts at S-atom partial structure determination were then carried out using *SnB* (Howell *et al.*, 2000), varying the resolution of the data, the number of reflections to be phased and the number of triplet relationships to be used. None were successful. However, examination of anomalous difference Patterson syntheses calculated at $d_{\rm min}$ = 2.7, 3.0 and 4.0 Å showed a number of constant features and an attempt was made to solve the heavy-atom partial structure at 4 Å resolution using the program *SOLVE* (Terwilliger & Berendzen, 1999). This was successful and provided nine sites as an initial model for refinement in the program *SHARP* at $d_{\rm min}$ = 2.7 Å. As for crystal form I, this initial model of the partial structure was then modified based on log-likelihood residual maps, with the final model consisting of 12 atoms, of which four were cysteines involved in two disulfide bridges and the rest were methionines. The final figure of merit from the *SHARP* procedure was 0.264 for all reflections in the range 30.0–2.7 Å and, as for form I, a clear protein–solvent boundary was seen in weighted electron-density maps calculated with the raw experimental phases ($\alpha_{\rm S\text{-}SAD}$, FOM$_{\rm S\text{-}SAD}$) (Fig. 4). As for crystal form I, these experimental phases were then improved by carrying out one round of solvent flattening using the program *SOLOMON* (Abrahams & Leslie, 1996) as incorporated in *SHARP* to produce fully interpretable electron-density maps (Fig. 5*a*). Phase extension to $d_{\rm min}$ = 2.35 Å was then carried out as described for crystal form I, giving a figure of merit of 0.680. From the resulting maps (Fig. 5*b*) an initial model was built manually using the program *QUANTA*. Refinement was then carried out using *CNS* and, when manual rebuilding of the model was required, *QUANTA*. Our current model has R = 20.1% and $R_{\rm free}$ = 21.5% (5% of data) for all data in the range 30–2.35 Å and contains 290 amino-acid residues (residues 16–163 for molecule *A*, 17–160 for molecule *B*) and 133 water molecules. An example of the electron density at the end of the refinement procedure is shown in Fig. 5(*c*).

3. Conclusions

A full report of the structures will appear elsewhere and we limit ourselves to a discussion of the implications of the phasing procedures described above.

We have confirmed that (as initially shown by Lui *et al.*, 2000) initial medium-resolution phases derived from the anomalous signal from sulfur at λ = 1.77 Å can be extended to the true diffraction limit of the crystals such that building of the protein structure was straightforward. However, the most important result is that we have produced, with no prior phase information, interpretable electron-density maps from intensity data collected from a native protein crystal for which the diffraction limit ($d_{\rm min}$ = 2.7 Å) can at best be described as medium resolution. Although our refinement of the structure of crystal form II has been carried out at somewhat higher

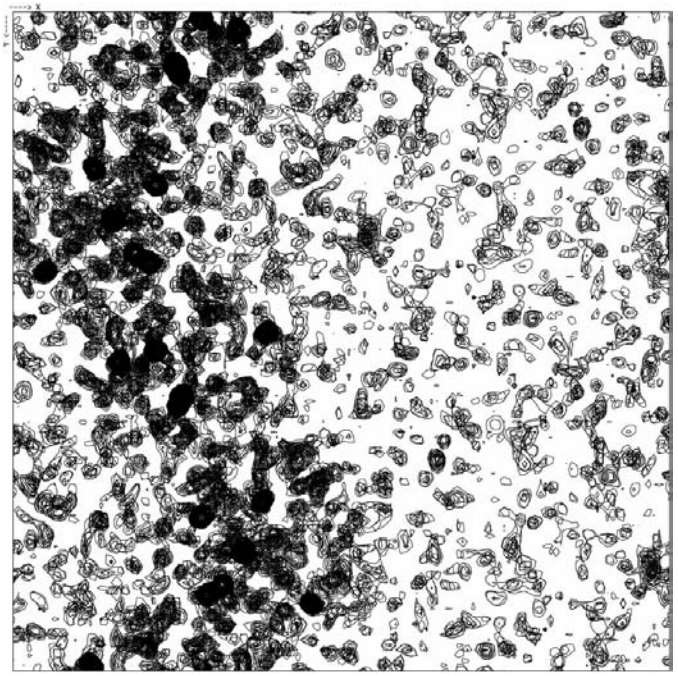

Figure 4
The electron-density map for crystal form II using the raw experimental phases ($\alpha_{\rm S\text{-}SAD}$, FOM$_{\rm S\text{-}SAD}$) at 2.7 Å resolution for the correct hand of the sulfur partial substructure model.

resolution (d_{min} = 2.35 Å), it is clear from Fig. 5(*a*) that even the initial 2.7 Å solvent-flattened but unaveraged electron-density map could easily be used as a basis for the construction of an initial model prior to refinement. Table 2, which details phase differences between those calculated from our final models and those found at various stages of the phasing procedure as well as giving values of the map correlation coefficients at the various intermediate stages, confirms this visual impression.

It may be thought from an examination of Table 1(*b*) that we are being somewhat pessimistic in describing the diffraction limit of the form II on BM14 as only 2.7 Å; after all, the average $I/\sigma(I)$ in the highest resolution data bin is 9.5. It should be noted, however, that this high value is entirely a result of the very high redundancy of our data set. Processing statistics from a subset of this data chosen using the program *STRATEGY* (Ravelli *et al.*, 1997) to give 98% completeness results in a redundancy of around 3 and $I/\sigma(I)$ in the highest resolution data bin of 3.7. Thus, this crystal would under normal circumstances be judged to have a diffraction limit close to 2.7 Å. Our solution of the structure using S-SAD with data collected from this crystal therefore represents a considerable increase over the previously mooted limit of 2.2 Å (Dauter *et al.*, 1999). In a similar manner to Dauter and colleagues, we have carried out a survey of the structures/data sets in the PDB (Berman *et al.*, 2000). At the time of writing

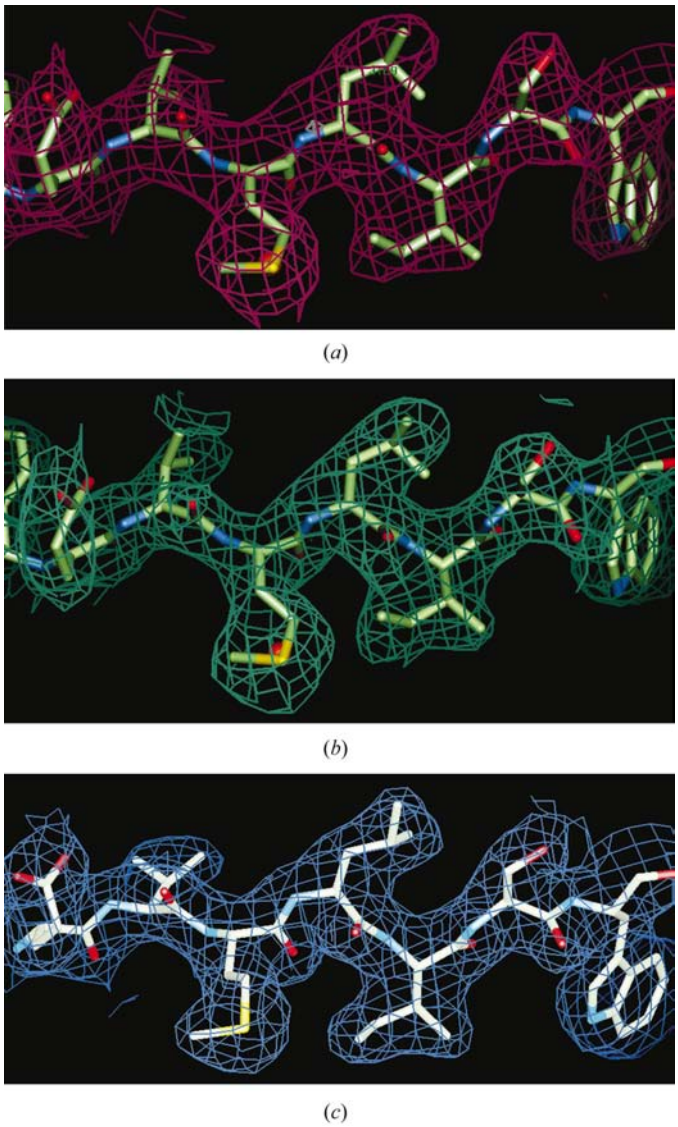

Figure 5
(*a*) Part of the electron-density map (red chicken wire) for crystal form II calculated using solvent-flattened phases at d_{min} = 2.7 Å. Note that no NCS averaging was used in the phase-improvement procedure that produced this electron-density map. (*b*) The same region of the electron-density map (green chicken wire) calculated at d_{min} = 2.35 Å after the phase-extension procedure described in the text. (*c*) The same region of the electron-density map (blue chicken wire) calculated at d_{min} = 2.35 Å after the refinement procedure. In all figures the model in its current state of refinement has been superimposed on the electron density.

Table 2
Phase differences (°) and map correlation coefficients between those calculated at the current stage of refinement and those derived at various stages of the phasing procedure for crystal forms I and II.

The numbers in parentheses refer to phase differences and map correlation coefficients calculated at d_{min} = 2.5 Å and d_{min} = 2.7 Å for crystal forms I and II, respectively.

	Phasing stage	d_{min} (Å)	Phase difference	Map correlation coefficient
Crystal form I	Raw S-SAD phases	2.5	62.0	0.45
	S-SAD + solvent flattening	2.5	52.7	0.68
	S-SAD + solvent flattening + NCS averaging	1.5	53.9 (40.6)	0.81 (0.84)
Crystal form II	Raw S-SAD phases	2.7	69.1	0.40
	S-SAD + solvent flattening	2.7	49.7	0.82
	S-SAD + solvent flattening + NCS averaging	2.35	50.4 (43.6)	0.83 (0.84)

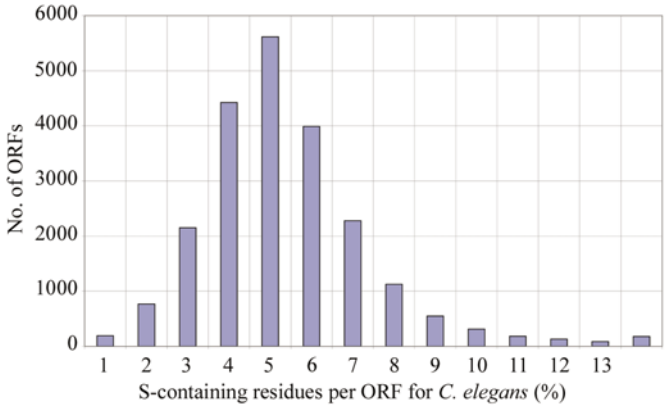

Figure 6
A histogram showing the distribution of the number of open reading frames (ORFs) as a function of the percentage content of sulfur-containing residues for the genome of *C. elegans*. '1' signifies between 0 and 1% *etc*. As can be seen, more than 50% of all ORFs in this genome are predicted to produce proteins containing at least as high a ratio of sulfur-containing residues as Trx II.

(May 2001) there are 12 501 entries for structures determined by X-ray diffraction. Of these, 10 776 are determined to a resolution of $d_{\min} = 2.7$ Å or better. This represents 86% of all entries and it is clear that, even allowing for the skewing effect of amino-acid mutant structures *etc.*, the vast majority of crystal structures would, in terms of resolution of data available, be amenable to solution by S-SAD.

It is clear from Figs. 3(*a*) and 5(*a*) as well as from Table 2 that the electron-density map obtained after solvent flattening but without averaging is considerably better for crystal form II than that obtained at the same stage for crystal form I. This is despite the fact that the S-SAD phasing data for the latter is of higher quality (see Table 1). The reason for this is probably that the high solvent content of crystal form II (∼72%) allows for a much better improvement of the phases than does the rather lower solvent content (∼46%) of crystal form I. This perhaps suggests that low solvent content may become a limiting factor in the use of S-SAD as a means of *de novo* structure determination. However, as can be seen in Table 2, the increase in diffraction limit ($d_{\min}$) and better quality data normally associated with crystals of a low solvent content should, in principle, produce more accurate raw experimental S-SAD phases. These should need less improvement before interpretable maps can be obtained and it may be that low solvent content will not become an issue for S-SAD phasing, although this clearly needs to be investigated further.

In addition to the resolution of the data available, one must also take into account amino-acid composition when considering whether a protein crystal structure might be suitable for solution by S-SAD. The asymmetric unit of crystal form II contains 300 ordered amino-acid residues, of which 14 are either cysteine or methionine. The sulfur-containing amino-acid composition of the protein is thus 4.7%. This yielded, at $\lambda = 1.77$ Å, a strong enough signal to solve the crystal structure. Of the five eukaryotic genomes for which there is full or partial sequence information, the frequency of occurrence of sulfur-containing amino-acid residues is as follows (see http://www.ebi.ac.uk/proteome/ for details): *Homo sapiens*, 4.4%; *Arabidopsis thaliana*, 4.3%; *Caenorhabditis elegans,* 4.7%; *Drosophila melanogaster*, 4.2%; *Saccharomyces cerevisiae*, 3.4%. Thus, large numbers of proteins from all these genomes should, in principle, be amenable to structure solution by S-SAD (see Fig. 6 for a breakdown of percentage of sulfur-containing residues for *C. elegans*). Most bacteria and archea have a frequency of occurrence of sulfur-containing amino acid residues of between 3 and 3.5% and, at first glance, may not contain as many proteins as eukaryotes that have the requisite ratio of sulfur-containing residues. However, it is unlikely that the frequency of occurrence of sulfur-containing residues of around 4.5% is the lower limit to give enough of an anomalous signal for S-SAD to be successful. On the high-energy side of an absorption edge the imaginary part of the scattering factor $\Delta f''$ increases as λ^2 and thus the use of longer wavelength X-rays should yield a greater anomalous signal from a smaller sulfur composition. Indeed, it is possible that the decreased resolution necessary for structure solution described above may be a consequence of the increased value of $\Delta f''$ (0.72 e^-) at $\lambda = 1.77$ Å compared with that (0.56 e^-) at $\lambda = 1.54$ Å at which Dauter and colleagues performed their measurements on lysozyme (Dauter *et al.*, 1999). Recently, the routine use of soft ($\lambda = 2.6$ Å) X-rays in protein crystallography has been described (Weiss *et al.*, 2001). It could therefore be that this, coupled with the S-SAD method, could make a considerable contribution to high-throughput structural biology and to structural genomics itself.

In addition to *de novo* structure solution, one of the major benefits of the S-SAD technique is that it could lead to a situation in which working with electron-density maps derived from experimental phases could become the norm, even for cases where molecular replacement would normally be the preferred method for structure solution. Even poor molecular-replacement solutions will yield reliable positions for the majority of S atoms in a structure, whether they be taken from the resulting model directly or derived from anomalous difference Fourier syntheses ($\Delta F_{\rm obs}$, $\alpha_{\rm MR} - 90^\circ$) calculated using the resulting molecular-replacement phases ($\alpha_{\rm MR}$). These, coupled with the observed anomalous differences ($\Delta F_{\rm obs}$) could then be used to provide electron-density maps with no model-based bias. These could help prevent errors in map interpretation (see Brodersen *et al.*, 2000 for some examples of these) and would provide a good guide to the correctness of molecular-replacement derived models.

The authors would like to thank Dr Magnus Alphey for discussions, Professor Peter Lindley for critical reading of the manuscript and Dr Sean McSweeney for his encouragement and providing much of the data relating to the sulfur-containing amino-acid composition of various genomes. WNH is funded by the Wellcome Trust.

References

Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* D**52**, 30–42.
Alphey, M. S., Bond, C. S., Tetaud, E., Fairlamb, A. H. & Hunter, W. N. (2000). *J. Mol. Biol.* **300**, 903–916.
Alphey, M. S., Leonard, G. A., Gourley, D. G., Tetaud, E., Fairlamb, A. H. & Hunter, W. N. (1999). *J. Biol. Chem.* **274**, 25613–25622.
Berman, H., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., Shindyalov, I. & Bourne, P. (2000). *Nucleic Acids Res.* **28**, 235–242.
Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography*, ch. 8. London: Academic Press.
Bond, C. S., Shaw, M. P., Alphey, M. S. & Hunter, W. N. (2001). *Acta Cryst.* D**57**, 755–758.
Brodersen, D., de La Fortelle, E., Vonrhein, C., Bricogne, G., Nyborg, J. & Kjeldgaard, M. (2000). *Acta Cryst.* D**56**, 431–441.
Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cowtan, K. (1994) *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.
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.
Dauter, Z., Li, M. & Wlodawer, A. (2001). *Acta Cryst.* D**57**, 239–249.
Deacon, A. M., Ni, Y. S., Coleman W. G. Jr & Ealick, S. E. (2000). *Structure*, **8**, 453–462.
Fourme, R., Shepard, W., Schiltz, M., Prangé, T., Ramin, M., Kahn, R., de La Fortelle, E. & Bricogne, G. (1999). *J. Synchrotron Rad.* **6**, 834–844.
French, S. & Wilson, K. (1978). *Acta Cryst.* A**34**, 517–525.
Gordon, E. J., Leonard, G. A., McSweeney, S. M. & Zagalsky, P. F. (2001). *Acta Cryst.* D**57**, 1230–1237.
Helliwell, J. R. (1992). *Macromolecular Crystallography with Synchrotron Radiation.* Cambridge University Press.
Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
Hendrickson, W. A. (1999). *J. Synchrotron Rad.* **6**, 845–851.
Hendrickson, W. A., Smith, J. L. & Sheriff, S. (1985). *Direct Phase Determination Based on Anomalous Scattering*, edited by H. W. Wycoff, C. H. Hirs & S. N. Timasheff, Vol. 115, pp. 41–55. New York: Academic Press.
Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
Hofmann, B., Budde, H., Bruns, K., Guerrero S. A., Kalisz, H. M., Menge, U., Montemartini, M., Nogoceke, F., Steinert, P., Wissing, J. B., Flohe, L. & Hecht, H. J. (2001). *Biol. Chem.* **382**, 459–471.
Howell, P. L., Blessing, R. H., Smith, G. D. & Weeks, C. M. (2000). *Acta Cryst.* D**56**, 604–617.
Korolev, S., Dementieva, I., Sanishvili, R., Minor, W., Otwinowski, Z. & Joachimiak, A. (2001). *Acta Cryst.* D**57**, 1008–1012.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Lui, Z.-J., Vysotski, E. S., Chen, C.-J., Rose, J. P., Lee, J. & Wang, B. C. (2000). *Protein Sci.* **9**, 2085–2093.
Miller, R., Gallo, S. M., Khalak, H. G. & Weeks, C. M. (1994). *J. Appl. Cryst.* **27**, 613–621.
Nagem, R. A. P., Dauter, Z. & Polikarpov, I. (2001). *Acta Cryst.* D**57**, 996–1002.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature. Struct. Biol.* **6**, 458–463.
Ravelli, R. B. G., Sweet, R. M., Skinner, J. M., Duisenberg, A. J. M. & Kroon, J. (1997). *J. Appl. Cryst.* **30**, 551–554.
Rice, L. M., Earnest, T. N. & Brunger, A. T. (2000). *Acta Cryst.* D**56**, 1413–1420.
Smith, J. L. (1991). *Curr. Opin. Struct. Biol.* **1**, 1002–1011.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Wang, B.-C. (1985). *Methods Enzymol.* **115,** 90–112.
Weiss, M. S., Siecker, T., Djinovic Carugo, K. & Hilgenfeld, R. (2001). *Acta Cryst.* D**57**, 689–695.