

Novel approach to phasing proteins: derivatization by short cryo-soaking with halides

Zbigniew Dauter,* Mirosława Dauter and K. R. Rajashankar

SAIC/National Cancer Institute, Frederick and Brookhaven National Laboratory, Building 725A-X9, Upton, NY 11973, USA

Correspondence e-mail: dauter@bnl.gov

A quick (less than 1 min) soak of protein crystals in a cryo-solution containing bromide or iodide anions leads to incorporation of these anomalous scatterers into the ordered solvent region around the protein molecules. These halide anions provide a convenient way of phasing through their anomalous scattering signal: bromides using multiwavelength anomalous dispersion (MAD) and bromides and/or iodides using single-wavelength anomalous dispersion (SAD) or single isomorphous replacement with anomalous scattering (SIRAS) methods. This approach has been tested successfully on four different proteins and has been used to solve the structure of a new protein of molecular weight 30 kDa.

Received 28 October 1999

Accepted 14 December 1999

1. Introduction

Anomalous dispersion was recognized as a powerful tool for phasing macromolecular diffraction data even in the early years of protein crystallography (Blow, 1958; Blow & Rossmann, 1961; Rossmann, 1961; Phillips & Hodgson, 1980; reviewed by Fourme *et al.*, 1996). For many years, it was used mainly to supplement the isomorphous replacement method (multiple isomorphous replacement with anomalous scattering, MIRAS; North, 1965; Matthews, 1966) or to locate the anomalous scatterers in protein structures [*e.g.* manganese in pea lectin (Einspahr *et al.*, 1985), iron in myohemerythrin (Sheriff & Hendrickson, 1987), molybdenum and iron in nitrogenase (Bolin *et al.*, 1993)]. However, the structures of crambin (Hendrickson & Teeter, 1981) and neurophysin II (Chen *et al.*, 1991) demonstrated the power of anomalous scattering as a sole source of phasing. The full potential of anomalous dispersion is utilized in the multiple-wavelength anomalous dispersion (MAD) method (Hendrickson, 1991; Smith, 1991). MAD requires a tunable synchrotron source of X-radiation and, with the increased availability of appropriate synchrotron beamlines, it has become the most popular method of solving novel macromolecular crystal structures (Deacon & Ealick, 1999; Hendrickson, 1999; Cassetta *et al.*, 1999; Fourme *et al.*, 1999).

In principle, MAD experiments can be performed using any scatterer. However, technical limitations narrow down the choices to those having an X-ray absorption edge in the wavelength range accessible at a synchrotron beamline. The most common absorption edges are the *K* edges of selenium and bromine or metals such as iron and copper, *L*_{III} edges of

heavy atoms (Hg, Au, Pt) and *L*_{III} edges of lanthanides. Atoms with *K* edges in the accessible range are usually introduced into the macromolecule by chemical methods (*e.g.* bromouracil in DNA) or by genetic engineering (*e.g.* selenomethionine in proteins) or are inherent in the native sample (*e.g.* Zn, Cu or Fe in metalloproteins). The heavier metals are usually soaked into the native crystals, as in the classic MIR approach, where they bind covalently to certain functional groups. Lanthanides often replace calcium and can be more or less tightly coordinated by carboxyl or carbonyl groups. By pressurizing the crystals of macromolecules in an atmosphere of xenon it is possible to introduce Xe atoms into the crystal and to use them for phasing through their dispersive and anomalous signal (Prangé *et al.*, 1998). The presence of anomalous scatterers in the disordered solvent region is the basis for the multiwavelength anomalous solvent contrast (MASC) method (Fourme *et al.*, 1995). MASC allows the definition of the boundary between the ordered protein and solvent at low resolution (usually less than 15 Å), providing a molecular mask which can be used for subsequent phasing through density modification.

In the present work, we demonstrate an easy and rapid way of introducing anomalously scattering halide ions into protein crystals, thereby providing a fast and simple route to solve the phase problem. This work resulted from previous observations that halide ions present in the crystallization medium occupy ordered positions around the molecule of lysozyme in the crystal (Lim *et al.*, 1998; Steinrauf, 1998; Dauter *et al.*, 1999; Dauter & Dauter, 1999). In the present case, the bromide or iodide anions were present in

Table 1
Crystallization conditions for four test proteins.

Protein	Lysozyme	RNAase A	Subtilisin	Xylanase
No. of amino acids	129	124	269	301
Space group	$P4_32_12$	$P2_1$	$P2_12_12_1$	$P2_1$
Crystallization conditions	1 M NaCl, 100 mM NaAc pH 4.7	50% MPD, 100 mM NaAc pH 5.4	12% PEG 4K, 100 mM citrate, 1 M NaCl, 10 mM CaCl ₂ pH 6.0	10% (NH ₄) ₂ SO ₄ , 100 mM Tris-HCl pH 7.4
Cryo-soaking conditions	100 mM NaAc, 30% glycerol, 1 M NaBr or 0.5 M NaCl + 0.5 M KI	50% MPD, 100 mM NaAc, 1 M NaBr or 0.35 M KI	12% PEG 4K, 10 mM citrate, 10 mM CaCl ₂ , 25% glycerol, 1 M NaBr or 1 M KI	10% (NH ₄) ₂ SO ₄ , 100 mM Tris-HCl, 25% glycerol, 1 M NaBr or 1 M KI

the cryo-protecting solution and diffused into the protein crystals during a short soak prior to cryogenic freezing of the sample for data collection. It is shown that these introduced anomalous scatterers can be successfully utilized for phasing the diffraction data, leading to solution of crystal structures.

2. Materials and methods

2.1. Crystallization and X-ray data collection

The four test proteins have been crystallized according to methods described in the literature: lysozyme (McPherson, 1982), RNAase A (Kantha *et al.*, 1967), subtilisin (Betzl *et al.*, 1988) and xylanase (Viswamitra *et al.*, 1993). The crystallization conditions for each protein are described briefly in Table 1.

All data were collected at the National Synchrotron Light Source, Brookhaven National Laboratory, beamline X9B, using either ADSC Quantum 4 or MAR Research 165 mm CCD detectors. The diffraction images were processed with the *HKL2000* suite (Otwinowski & Minor, 1997). The results are summarized in Table 2. In all cases, crystals were in arbitrary orientations and no inverse-beam approach was used.



Figure 1
The anomalous difference Fourier synthesis (ΔF^{anom} , $\varphi_{\text{calc}} - 90^\circ$) for RNAase A showing bromide sites, contoured at 5σ .

(Collaborative Computational Project, Number 4, 1994). Atomic models and electron density maps were visualized using *QUANTA* (Molecular Simulations Inc, San Diego, USA) or *O* (Jones *et al.*, 1991).

3. Results

3.1. Crystal preparation and data collection

The crystallization conditions of the four test proteins were the same as used in their original structure determinations. As evident from Table 1, they are crystallized from various concentrations of protein, diverse precipitating agents and at different pHs. The cryoprotecting solutions appropriate for freezing crystals for data collection included various amounts of glycerol in the mother liquor (except for RNAase A, which was crystallized from an MPD solution not requiring any additional cryoprotectant). In addition, the cryoprotecting solution also contained NaBr or KI. A 1 M concentration of NaBr did not affect the quality and diffracting power of crystals, but 1 M KI negatively influenced crystals of lysozyme and xylanase. Hence, the KI concentration was reduced to 0.35 and 0.5 M, respectively, in these cases. For crystals grown in high salt concentration, the other salts may be partly or completely substituted by halides.

The crystals were transferred to cryoprotectant solution for a short time (between 15 and 45 s). Diffraction data were collected from frozen crystals (at 100 K) in arbitrary orientations, using a wavelength of 1.54 Å for iodide derivatives and an appropriate three wavelengths (inflection point, white peak and near-remote high energy) around the absorption edge of bromine for bromide-soaked crystals. Only one data set was collected at near-remote wavelength for the Br-derivatized RNAase A crystal. Near-remote energy was only 50 eV beyond the edge, as a result of hardware limitations of the monochromator on the X9B NSLS beamline.

The rotation range was selected to yield complete anomalous data, but no inverse-beam technique was used. In spite of the fact that the crystals of all four test proteins are able to diffract to atomic resolution, the exposure times were short and diffraction data were recorded to a maximum resolution ranging from 1.4 to 1.7 Å (Table 2).

3.2. Identification of anomalous sites

To obtain model phases, the four test protein models were refined against respective data sets; for bromide MAD data, the

Table 2
X-ray data and phasing details.

Protein	Lysozyme		RNAse A		Subtilisin		Xylanase	
	Br†	I	Br	I	Br†	I	Br†	I
Resolution (Å)	1.70	1.62	1.50	1.62	1.40	1.60	1.50	1.60
R_{merge} (%)	3.8	5.0	3.5	5.9	5.1	4.4	2.4	4.4
Completeness (%)	100.0	99.5	99.9	95.1	99.9	93.6	88.5	94.4
$I/\sigma(I)$	30.4	20.0	28.7	15.7	21.2	26.5	30.0	25.6
Multiplicity‡	3.8 (7.0)	3.2 (5.9)	3.7 (7.1)	3.2 (6.2)	3.1 (6.0)	3.3 (6.1)	1.9 (3.4)	3.0 (5.9)

Anomalous scatterers and phasing (2.0 Å data used for all calculations).

Sites in (ΔF^{anom} , $\varphi_{\text{calc}} - 90^\circ$) map									
No. of sites > 10σ	8	8	9	5	17	16	18	11	
No. of sites > 5σ	14	16	18	21	40	39	34	31	
No. of sites > 3σ	67	73	67	125	172	105	164	128	
Phasing method									
No. of sites used for phasing	MAD	SIRAS	SAD	SIRAS	MAD	SIRAS	MAD	SIRAS	
FOM after <i>DM</i>	7	11	12	9	7	8	16	14	
Correlation coefficient with F_o map	0.87	0.77	0.73	0.74	0.85	0.76	0.78	0.75	
	0.89	0.73	0.68	0.63	0.80	0.68	0.65	0.69	

† Statistics for the remote wavelength are given. Data for other wavelengths are very similar. ‡ Multiplicity calculated with Friedel-related reflections separated (or, in parentheses, treated as equivalent).

near-remote wavelength sets were used. The models were refined by *REFMAC*; *ARP* was used to select solvent waters in an automatic manner. All solvent atoms were treated as waters, even when some of them acquired

very low B factors suggesting their true identity as heavier atoms. The resulting R -factor values were in the range 16–20% and R_{free} values were in the range 20–25%.

The phases calculated from the refined atomic models were used to compute the anomalous difference Fourier maps, of which one example is shown in Fig. 1. In all cases, the maps showed several anomalous peaks surrounding the protein surface (Table 2). The anomalous solvent peaks had variable heights and all of them corresponded to the water positions automatically found by *ARP* during model refinement. The anomalous peak heights generally did not correlate with refined water B factors, suggesting that the halide ion sites are partially occupied, shared between Br^- or I^- ions and water molecules.

In an independent attempt to solve the anomalous scatterer partial structures, the direct-methods program *SHELXS* was used. In all cases, clearly indicated correct solutions (within the equivalent sets of possible origin shifts or enantiomers) were obtained. In contrast to the usual situation in small-molecule crystallography, the peak heights in the corresponding E maps varied from a maximum decreasing gradually to lower values, in agreement with the presence of many partially occupied halide sites.

3.3. Environment of anomalous sites

Bromide and iodide anions generally occupy the same sites around the protein surface within the ordered solvent shell. Apparently, the halide sites are shared with water molecules, as judged from the refined B factors of water O atoms at these positions and from the height of the corresponding peaks in the anomalous difference Fourier syntheses.

The occupancies of halide sites were refined using *SHELXL*. Only halide sites corresponding to peaks higher than 5σ in the anomalous difference map were included. Each site was filled by a halide and a water with constrained coordinates, but with their occupancies and B factors refined independently. The results are presented in Fig. 2. In each case, occupancy of several sites refined to above 50%, generally with good correlation with the height of their peaks in the anomalous difference map. The peaks at the 5σ map level acquired occupancies in the range 10–20%. This agrees with the height of sulfur peaks visible in ΔF^{anom} maps calculated with data collected at 1.54 Å, *i.e.* from iodide-soaked crystals. They appeared at about the 6σ level, roughly 10% of the highest occupied iodide sites, in keeping with the corresponding values of $\delta f''$ for sulfur (0.56 e) and for iodine (6.8 e).

The halide sites are not specific. Most of them have hydrogen-bonding contacts with hydrogen-donor groups of protein or water on one side and hydrophobic interactions in non-polar niches at the protein surface. Those halide ions which form ion pairs with positively charged arginine or lysine side chains have the highest occupancies. A few typical examples of halide sites are illustrated in Fig. 3.

The ability of halide ions to (at least partially) substitute water molecules within the ordered solvent region around the protein in the crystal explains why so many sites can be identified on the basis of their anomalous scattering signal. Heavy transition metals as well as lanthanides require rather specific coordination around their sites. This can only rarely be provided at the protein surface. Similarly, complex anions like SeO_4^{2-} or AsO_4^{3-} require being surrounded by a polar environment and in this respect do not show the flexibility of the monoatomic polarizable halide anions. However, some ordered metal or complex anion sites at the protein surface were identified from their anomalous signal; for example, selenates or ytterbium cations (Ramin *et al.*, 1999) or thallium ions in cubic insulin (Badger *et al.*, 1994). Selenates have

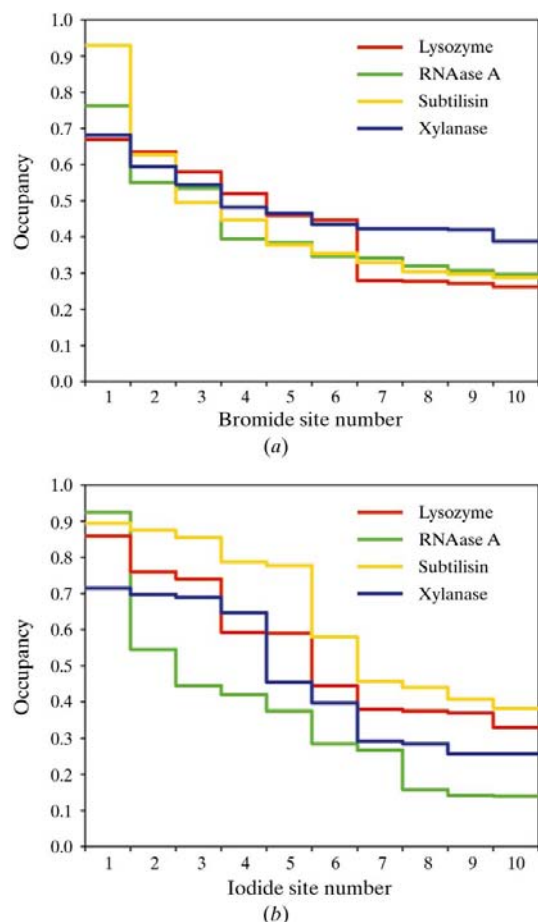


Figure 2
The refined occupancies of ten strongest (a) bromide and (b) iodide sites in the four test structures.

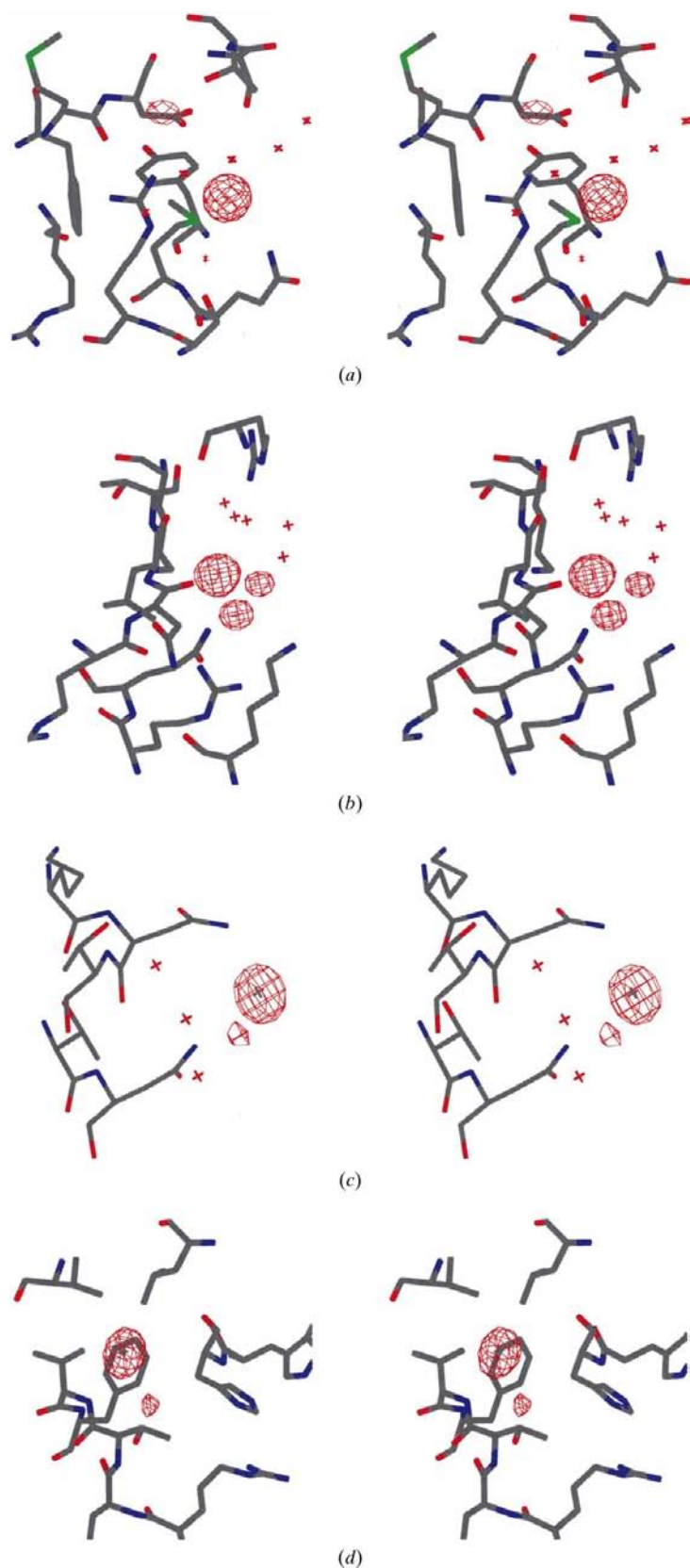


Figure 3
The environment of halides in protein structures. The anomalous difference density is also shown at 5σ . (a) and (b) Bromide ions in RNAase A near the arginine side chains, (c) iodide ion coordinated by two asparagine NH_2 groups at the surface of xylanase and (d) iodide ion located in a hydrophobic niche and hydrogen bonded to a main-chain amide of xylanase.

been identified in phosphate- or sulfate-binding sites (Ericksson *et al.*, 1993; Qiu *et al.*, 1996). Xenon usually occupies sites located in protein hydrophobic cavities and is not able to substitute water around the protein surface (Prangé *et al.*, 1998)

In this respect, bromide and iodide anions (as well as chlorides; Dauter *et al.*, 1999) seem to be unique in their ability to substitute water molecules within the ordered solvent shell in the protein crystals. The number of halide sites in the four tested structures is proportional to the surface area of the protein studied.

3.4. Phasing

All phasing calculations were performed with the program *MLPHARE*. For Br-MAD data of lysozyme, subtilisin and xylanase, the data collected at the inflection-point wavelength were accepted as native and the two other wavelengths as derivatives. Single-wavelength anomalous dispersion (SAD) phasing was applied to RNAase A soaked in bromide. For iodine derivatives, the available native data were utilized as in the classic SIRAS approach. The selection of peaks from *SHELXS* for use for phasing in *MLPHARE* was somewhat arbitrary, since their heights diminish gradually from the maximum value without clear contrast between correct and false sites. About ten highest sites were accepted in each phasing trial.

The phase sets obtained from *MLPHARE* were subjected to density modification by *DM* in the *COMBINE PERTURB* mode. The values of the figure of merit and correlation coefficient with the F_{obs} map for all protein residues are given in Table 2. An example of the resulting *DM*-phased electron density is shown in Fig. 4(a).

3.5. Application to unknown structure

The quick cryo-soaking approach was also applied to the crystals of a novel protein consisting of about 300 amino acids (von Buelow *et al.*, in preparation). The protein was crystallized from 30% PEG 4K, 0.2 M ammonium sulfate with 0.1 M cacodylate buffer pH 5.8 and the crystals were soaked for 1 min in a cryo-solution containing 1 M KI. Diffraction data were collected to a resolution of 2.2 Å and the subsequent process followed the protocol outlined above for the test structures. The resulting phase set, characterized by a FOM of 0.77, yielded a clearly interpretable map, of which a portion is shown in Fig. 4(b).

4. Conclusions

Bromide or iodide ions, when present in the cryoprotectant solution, easily diffuse into protein crystals during short (~30 s) soaking. This has been observed for four different proteins crystallized under diverse conditions.

Bromine and iodine display significant anomalous dispersion effects. The K absorption edge of bromine (at 0.92 Å or 13 474 eV) is easily accessible at MAD-capable synchrotron beamlines. The quick and easy incorporation of bromide ions into a number of ordered sites of the protein crystal by short soaking in appropriate cryo-solution may be proposed as an addition to the traditional selenomethionine-based MAD approach, which requires an elaborate and time-consuming preparation of the sample.

The absorption edges of iodine are beyond the easily accessible wavelength range. However, at 1.54 Å X-radiation

iodine has an anomalous $\delta f''$ value of 6.8 e (Cromer, 1983). Although it is not susceptible to MAD, it provides a significant anomalous effect which can be utilized in the SAD or SIRAS approaches.

Diffusion of halide ions into protein crystals seems to occur rapidly. They tend to occupy several ordered solvent sites around the protein surface with varying occupancy, presumably sharing their sites with water molecules. The sites are not particularly specific. Those near positively charged arginine or lysine side chains form ion pairs and usually display higher occupancy, while others substitute general solvent-water sites and may form hydrogen bonds with various polar functions of the protein and/or attach to the hydrophobic patches on the protein surface. The latter effect can be attributed to a significant polarizability of bromine and iodine. Since arginines and lysines have high pK_a values (12.0 and 10.0, respectively), they remain positively charged up to a relatively high pH and are able to bind halide anions

over a wide pH range of crystallization conditions.

Halides, being small monoatomic ions, are able to substitute water in the solvent sites around the protein surface. They do not have any preference for specific coordination geometry. In contrast, most metal ions display a preference for a particular coordination and do not easily find appropriate ordered sites at the surface of a protein in the crystal. Similarly, complex anions such as (potentially anomalous) selenates or arsenates may also require specific coordination for binding in a more ordered fashion. In this respect, bromides and iodides are more suited as ordered anomalous scatterers for soaking into protein crystals.

This approach requires very little preparative effort and may be particularly applicable for high-throughput crystallographic projects, such as structural genomics. The quick cryo-soaking with halides described here may be an alternative method for phasing protein crystal structures.

We thank Boris Strokopytov for providing crystals of RNAase A, Professor Viswamitra and Professor Ramakumar for crystals of xylanase, and Mike Sullivan for help with the beamline and detector setup.

References

- Badger, J., Li, Y. & Caspar, D. L. D. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 1224–1228.
- Betzel, C., Dauter, Z., Dauter, M., Ingelman, M., Papendorf, G., Wilson, K. S. & Branner, S. (1988). *J. Mol. Biol.* **204**, 803–804.
- Blow, D. M. (1958). *Proc. R. Soc. London Ser. A*, **247**, 302.
- Blow, D. M. & Rossmann, M. G. (1961). *Acta Cryst.* **14**, 1195–1202.
- Bolin, J. T., Ronco, A. E., Morgan, T. V., Mortenson, L. E. & Xuong, N.-H. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 1078–1082.
- Cassetta, A., Deacon, A. M., Ealick, S. E., Helliwell, J. R. & Thompson, A. W. (1999). *J. Synchrotron Rad.* **6**, 822–833.
- Chen, L., Rose, J. P., Breslow, E., Yang, D., Chang, W. R., Furey, W. F., Sax, M. & Wang, B.-C. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 4240–4244.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- 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. & Dauter, M. (1999). *J. Mol. Biol.* **289**, 93–101.
- Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.
- Deacon, A. M. & Ealick, S. E. (1999). *Structure*, **7**, R161–R166.

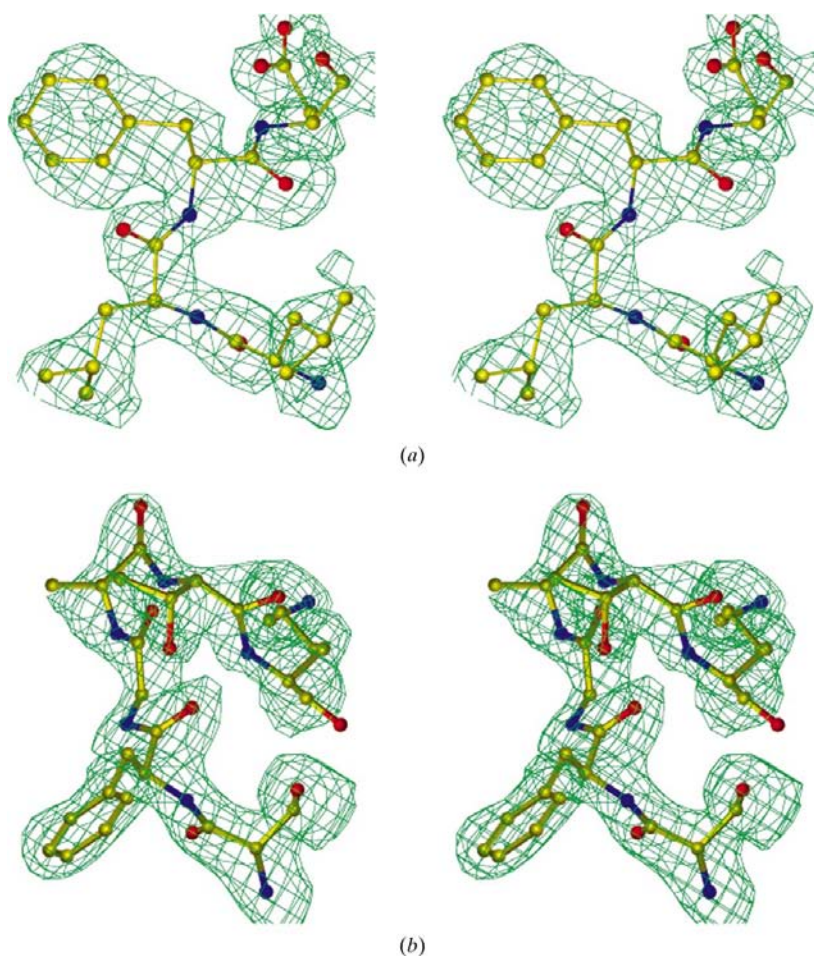


Figure 4

Representative portions of the experimental electron-density map after density modification by DM at the 1σ level for (a) xylanase and (b) a novel protein, both phased using iodides.

- Einspahr, H., Suguna, K., Suddath, F. L., Ellis, G., Helliwell, J. R. & Papiz, M. Z. (1985). *Acta Cryst.* **B41**, 336–341.
- Ericksson, A. E., Cousens, L. S. & Matthews, B. W. (1993). *Protein Sci.* **2**, 1274–1284.
- Fourme, R., Shepard, W. & Kahn, R. (1996). *Prog. Biophys. Mol. Biol.* **64**, 167–199.
- Fourme, R., Shepard, W., Kahn, R., L'Hermite, G. & de La Sierra, G. (1995). *J. Synchrotron Rad.* **2**, 36–48.
- 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.
- Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
- Hendrickson, W. A. (1999). *J. Synchrotron Rad.* **6**, 845–851.
- Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* **A47**, 110–119.
- Kartha, G., Bello, J. & Harker, D. (1967). *Nature (London)*, **213**, 862–865.
- Lamzin, V. S. & Wilson, K. S. (1997). *Methods Enzymol.* **277**, 269–305.
- Lim, K., Nadarajah, A., Forsythe, E. L. & Pusey, M. L. (1998). *Acta Cryst.* **D54**, 899–904.
- McPherson, A. Jr (1982). *The Preparation and Analysis of Protein Crystals*. New York: John Wiley.
- Matthews, B. W. (1966). *Acta Cryst.* **20**, 82–86.
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* **D53**, 240–255.
- North, A. C. T. (1965). *Acta Cryst.* **18**, 212–216.
- Otwinowski, Z. (1991). *Proceedings of the CCP4 Study Weekend*, edited by W. Wolf, P. R. Evans & A. G. W. Leslie, pp. 80–86. Warrington: Daresbury Laboratory.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Phillips, J. C. & Hodgson, K. O. (1980). *Acta Cryst.* **A36**, 856–864.
- Prangé, T., Schiltz, M., Pernot, L., Colloc'h, N., Longhi, S., Bourguet, W. & Fourme, R. (1998). *Proteins*, **30**, 61–73.
- Qiu, X., Pohl, E., Holmes, R. K. & Hol, W. G. (1996). *Biochemistry*, **35**, 12292–12302.
- Ramin, M., Shepard, W., Fourme, R. & Kahn, R. (1999). *Acta Cryst.* **D55**, 157–167.
- Rossmann, M. G. (1961). *Acta Cryst.* **14**, 383–388.
- Sheldrick, G. M. (1990). *Acta Cryst.* **A46**, 467–473.
- Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
- Sheriff, S. & Hendrickson, W. A. (1987). *Acta Cryst.* **B43**, 209–212.
- Smith, J. L. (1991). *Curr. Opin. Struct. Biol.* **1**, 1002–1011.
- Steinrauf, L. K. (1998). *Acta Cryst.* **D54**, 767–779.
- 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.