Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

A magic triangle for experimental phasing of macromolecules

Tobias Beck,* Andrius Krasauskas, Tim Gruene and George M. Sheldrick

Department of Structural Chemistry, Georg-August-Universität Göttingen, Tammannstrasse 4, 37077 Göttingen, Germany

Correspondence e-mail: tbeck@shelx.uni-ac.gwdg.de

Received 18 August 2008
Accepted 19 September 2008

PDB References: lysozyme, 3e3d, r3e3dsf; thaumatin, 3e3s, r3e3ssf; elastase, 3e3t, r3e3tsf.

Obtaining phase information for the solution of macromolecular structures is still one of the bottlenecks in X-ray crystallography. 5-Amino-2,4,6-triiodoisophthalic acid (I3C), in which three covalently bound iodines form an equilateral triangle, was incorporated into proteins in order to obtain phases by single-wavelength anomalous dispersion (SAD). An improved binding capability compared with simple heavy-metal ions, ready availability, improved recognition of potential heavy-atom sites and low toxicity make I3C particularly suitable for experimental phasing.

1. Introduction

Protein structure solution using X-ray crystallography can only succeed if a suitable model is available or experimental phases can be obtained. Although molecular replacement has proven to be a powerful technique, it often shows model bias at medium and low resolution and can only be applied if a similar structure is already known. The traditional experimental phasing method is MIR (multiple isomorphous replacement), which involves soaking the crystals with several different heavy-atom reagents. It can suffer from difficulties in obtaining suitable heavy-atom derivatives and lack of isomorphism. The SAD (single-wavelength anomalous dispersion) and MAD (multiple-wavelength anomalous dispersion) methods for experimental phasing have evolved substantially in recent years. Although S or metal atoms, if present in the native protein, can provide phase information, particularly high-quality data are required for sulfur-SAD phasing because of the weak anomalous signal. Where it is possible to express a selenomethionine (SeMet) mutant this often enables the structure to be solved by MAD or SAD methods (Ogata, 1998; Dambe *et al.*, 2006) and, in principle, the same applies to the use of brominated nucleobases for DNA structures (Escalante *et al.*, 1998), although especially in the latter case radiation damage can be a limiting factor. Anomalously scattering atoms can also be introduced by other chemical modifications of the protein or nucleic acid (Xie *et al.*, 2004; Miyatake *et al.*, 2006).

A less demanding approach to SAD or MAD phasing is to soak crystals in a heavy-atom solution to incorporate anomalous scatterers such as mercury (Boggon & Shapiro, 2000), uranium (Wernimont *et al.*, 2000), iodine (Dauter *et al.*, 2000) or a tantalum bromide cluster (Szczepanowski *et al.*, 2005) into the crystal lattice. However, most heavy-metal ion soaks suffer from nonspecific binding of the heavy atoms, resulting in many binding sites that are not occupied in every unit cell. This leads to a weak anomalous signal and disruption of the crystal lattice, with loss of isomorphism or even destruction of the crystal. In addition, many such soaks require the use of toxic chemicals and require stringent safety precautions.

We chose the iodine-containing compound 5-amino-2,4,6-triiodoisophthalic acid (hereafter referred to as I3C; Fig. 1*a*) as a representative of a new class of derivatives for phasing that combine heavy atoms with functional groups for interaction with proteins and nucleic

acids. Three functional groups (two carboxylates and one amino group) of I3C interact through hydrogen bonds with the main chain as well as with the side chains of proteins. This results in a relatively high occupancy of the bound ligands. With three I atoms per molecule it provides a strong anomalous signal, even for in-house X-ray sources (the anomalous signal f'' of iodine is 6.85 e$^-$ at the Cu $K\alpha$ wavelength). The I atoms form an equilateral triangle (with a side of 6.0 Å) and thus are readily identified in the heavy-atom substructure. In comparison to other heavy-atom reagents, I3C has low toxicity. Its derivatives are employed as X-ray contrast reagents in medical diagnosis (Yu & Watson, 1999). We have described the synthesis and crystallographic characterization of I3C elsewhere (Beck & Sheldrick, 2008) and it is inexpensive and readily available commercially. Initial indications are that it may be successful over a wide pH range.

2. Methods

2.1. Crystallization and data collection

In this study, the incorporation of I3C into three protein samples by either cocrystallization or soaking was investigated. Stock solutions of I3C were obtained by dissolving the solid material in water, adding double the equimolar amount of sodium hydroxide solution or lithium hydroxide solution to fully deprotonate the carboxyl groups. A 0.2 *M* sodium I3C solution and a 1 *M* lithium I3C solution were prepared. Protein crystals were obtained using the sitting-drop vapour-diffusion method.

Hen egg-white lysozyme (129 residues, 14.3 kDa) was crystallized at 293 K by mixing 5 µl 20 mg ml^{-1} protein solution with an equivalent volume of precipitant containing 0.1 *M* HEPES pH 7.0, 15%(*w*/*w*) PEG 3350 and 8 m*M* sodium I3C solution. Crystals containing I3C appeared within one week. Data were collected in-house on a MAR345 image-plate detector using Cu $K\alpha$ radiation from a MacScience rotating-anode X-ray generator equipped with Osmic optics.

Thaumatin (207 residues, 22.2 kDa) was crystallized at 293 K by mixing 2 µl 40 mg ml^{-1} protein solution with an equivalent volume of precipitant solution containing 0.05 *M* ADA pH 6.8 and 0.8 *M* potassium sodium tartrate. A quick soak (2 min) was performed with mother liquor containing 0.5 *M* lithium I3C. Data were collected in-house as for lysozyme.

Porcine pancreatic elastase (240 residues, 25.9 kDa) was crystallized at 277 K by mixing 2 µl 40 mg ml^{-1} protein solution with an equivalent volume of precipitant solution containing 0.1 *M* HEPES pH 8.0 and 0.6 *M* sodium sulfate. A gradient soak was performed to incorporate I3C into the crystal starting with 0.15 *M* lithium I3C solution (2 min) containing mother liquor, continuing with 0.25 *M* lithium I3C solution (5 min) and then adding the cryoprotectant [30%(*v*/*v*) glycerol]. Data were recorded in-house with a Bruker Smart 6000 CCD detector using Cu $K\alpha$ radiation from a MacScience rotating-anode generator equipped with Incoatec Helios optics.

Table 1
Data-collection details for lysozyme, thaumatin and elastase.

For lysozyme and thaumatin, a comparison of the different data sets shows that a much lower multiplicity is adequate to solve the heavy-atom substructure. With increasing map correlation coefficient and decreasing mean phase error, more residues can be traced by automated procedures. A similar comparison for elastase was not carried out because of the different detector and goniometer settings (three-circle goniometer). For refinement statistics, see Table 2. Values in parentheses are for the highest resolution shell.

(*a*) Lysozyme.

Unit-cell parameters (Å, °)	$a = b = 76.83$, $c = 38.87$, $\alpha = \beta = \gamma = 90$				
Space group	$P4_32_12$				
Resolution (Å)	19.4–1.55 (1.65–1.55)				
Degrees collected (°)	30	45	60	90	360
$R_{\rm merge}$†	0.060 (0.14)	0.063 (0.15)	0.066 (0.16)	0.068 (0.18)	0.069 (0.15)
Completeness	0.86 (0.83)	0.96 (0.93)	0.98 (0.98)	0.99 (0.98)	0.99 (0.99)
Multiplicity‡	2.14 (1.99)	3.23 (2.96)	4.32 (3.97)	6.50 (5.94)	26.1 (24.0)
$I/\sigma(I)$	10.3 (5.06)	12.0 (5.62)	13.8 (6.43)	17.1 (7.80)	38.2 (21.2)
$R_{\rm anom}$§	0.117 (0.238)	0.115 (0.236)	0.114 (0.229)	0.109 (0.205)	0.0974 (0.122)
$R_{\rm anom}/R_{\rm p.i.m.}$¶	3.68 (2.17)	3.61 (2.20)	3.71 (2.12)	4.31 (2.19)	7.12 (3.00)
No. of sites found	10 [83%]	12 [100%]	12 [100%]	12 [100%]	12 [100%]
Mean phase error (°)	38.6	32.6	32.1	30.4	31.1
Mean map CC	0.731	0.801	0.817	0.833	0.852
No. of residues built	53 [41%]	83 [64%]	105 [81%]	111 [86%]	120 [93%]

(*b*) Thaumatin.

Unit-cell parameters (Å, °)	$a = b = 57.71$, $c = 149.58$, $\alpha = \beta = \gamma = 90$				
Space group	$P4_12_12$				
Resolution (Å)	19.7–1.73 (1.83–1.73)				
Degrees collected (°)	30	45	60	90	360
$R_{\rm merge}$†	0.038 (0.095)	0.041 (0.099)	0.044 (0.11)	0.046 (0.11)	0.056 (0.12)
Completeness	0.88 (0.86)	0.97 (0.95)	0.99 (0.97)	0.99 (0.97)	0.99 (0.98)
Multiplicity‡	2.19 (1.99)	3.29 (2.97)	4.40 (4.00)	6.65 (6.12)	26.8 (24.2)
$I/\sigma(I)$	16.3 (8.36)	18.7 (9.75)	20.8 (10.9)	26.0 (13.9)	47.5 (25.5)
$R_{\rm anom}$§	0.0774 (0.157)	0.0750 (0.143)	0.0712 (0.135)	0.0668 (0.115)	0.0620 (0.0886)
$R_{\rm anom}/R_{\rm p.i.m.}$¶	3.41 (2.04)	3.29 (1.85)	3.07 (1.88)	3.36 (1.90)	4.96 (2.69)
No. of sites found	15 [100%]	15 [100%]	15 [100%]	15 [100%]	15 [100%]
Mean phase error (°)	50.2	32.5	26.7	26.0	28.4
Mean map CC	0.608	0.811	0.886	0.891	0.860
No. residues built	0 [0%]	199 [96%]	198 [96%]	199 [96%]	196 [95%]

(*c*) Elastase.

Unit-cell parameters (Å, °)	$a = 50.10$, $b = 57.97$, $c = 74.40$, $\alpha = \beta = \gamma = 90$
Space group	$P2_12_12_1$
Resolution (Å)	37.2–1.60 (1.70–1.60)
Degrees collected (°)	360
$R_{\rm merge}$†	0.038 (0.10)
Completeness	0.99 (0.92)
Multiplicity‡	11.2 (2.76)
$I/\sigma(I)$	40.6 (8.38)
$R_{\rm anom}$§	0.0310 (0.109)
$R_{\rm anom}/R_{\rm p.i.m.}$¶	2.61 (1.17)
No. of sites found	12 [100%]
Mean phase error (°)	35.3
Mean map CC	0.806
No. of residues built	227 [95%]

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$. ‡ Friedel opposites merged. § $R_{\rm anom} = \sum_{hkl} |I(hkl) - I(\bar{h}\bar{k}\bar{l})| / \sum_{hkl}\langle I(hkl)\rangle$. ¶ $R_{\rm p.i.m.} = \sum_{hkl}[1/(N-1)]^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$.

3. Results and discussion

3.1. Substructure solution and data analysis

Substructure solution with *SHELXD* (Schneider & Sheldrick, 2002) using Patterson seeding and dual-space direct methods followed by density modification with *SHELXE* (Sheldrick, 2002)

resulted in high-quality starting phases (Table 1). Model building was carried out with *ARP*/*wARP* and *REFMAC* (Perrakis *et al.*, 1999; Murshudov *et al.*, 1997) as part of the *CCP*4 suite (Potterton *et al.*, 2003) and *Coot* (Emsley & Cowtan, 2004). Refinement was carried out with the *SHELX* suite (Table 2; Sheldrick, 2008).

Investigations of the data show that only a small multiplicity of data is required to find the heavy-atom positions (Table 1), although of course the ratio of $R_{\rm anom}$ to $R_{\rm p.i.m.}$, a measure of the quality of the anomalous signal (Weiss, 2001), increases with higher multiplicity. Since the heavy atoms form an equilateral triangle, a successful solution is readily identified when inspecting the heavy-atom positions (Fig. 1*b*), which facilitates structure solution. It is planned to incorporate a search for triangles and similar patterns into *SHELXD*, similar to the resolution of 'super-sulfurs' (Debreczeni *et al.*, 2003); this should be particularly beneficial for the solution of larger structures.

Table 2
Refinement statistics for lysozyme, thaumatin and elastase.

	Lysozyme	Thaumatin	Elastase
PDB code	3e3d	3e3s	3e3t
Resolution (Å)	19.4–1.55	19.7–1.73	37.2–1.60
No. of reflections	17445	27271	28897
$R_{\rm cryst}/R_{\rm free}$	17.1/22.0	15.8/21.0	15.7/21.3
No. of protein atoms	993	1531	1860
No. of ligand/ion atoms	79	94	75
No. of waters	147	262	300
B factors (Å^2)			
Protein	9.58	12.80	14.22
Ligands	11.00	21.90	44.68
Waters	21.23	25.50	30.10
R.m.s. deviations			
Bond length (Å)	0.007	0.008	0.008
Angle distance (Å)	0.023	0.024	0.024
Ramachandran plot			
Favoured (%)	99.2	98.5	97.9
Disallowed (%)	0	0.5	0

3.2. I3C-binding sites

The molecular structures are similar to those previously reported (*e.g.* lysozyme, Weiss *et al.*, 2000; thaumatin, Ko *et al.*, 1994; elastase, Würtele *et al.*, 2000), since I3C mostly replaces solvent water molecules in the crystal lattice. Inspection of the I3C sites reveals several modes of interaction. The amino group of I3C forms hydrogen bonds to a main-chain carbonyl O atom (*e.g.* Fig. 1*c*) or to a hydrogen-bond acceptor in a side chain (*e.g.* the O atom of asparagine or glutamine). The carboxylate groups interact with side chains of the protein *via* hydrogen bonds to donor groups in serine, lysine, tyrosine or threonine residues. The prominent hydrogen bonding of the carboxylate group is its interaction with arginine (Fig. 1*d*). Additional hydrogen bonds are also formed to solvent water molecules. These strong interactions lead to relatively high occupancies of the sites (Fig. 2). The occupancies of I3C were refined with *SHELXL* to values of 0.60, 0.32, 0.32 and 0.25 for lysozyme, 0.49, 0.44, 0.27, 0.26 and 0.19 for thaumatin and 0.45, 0.23, 0.21 and 0.16 for elastase. From the limited evidence available, it appears that for incorporation of I3C it will be worth trying both soaking and cocrystallization in practice.

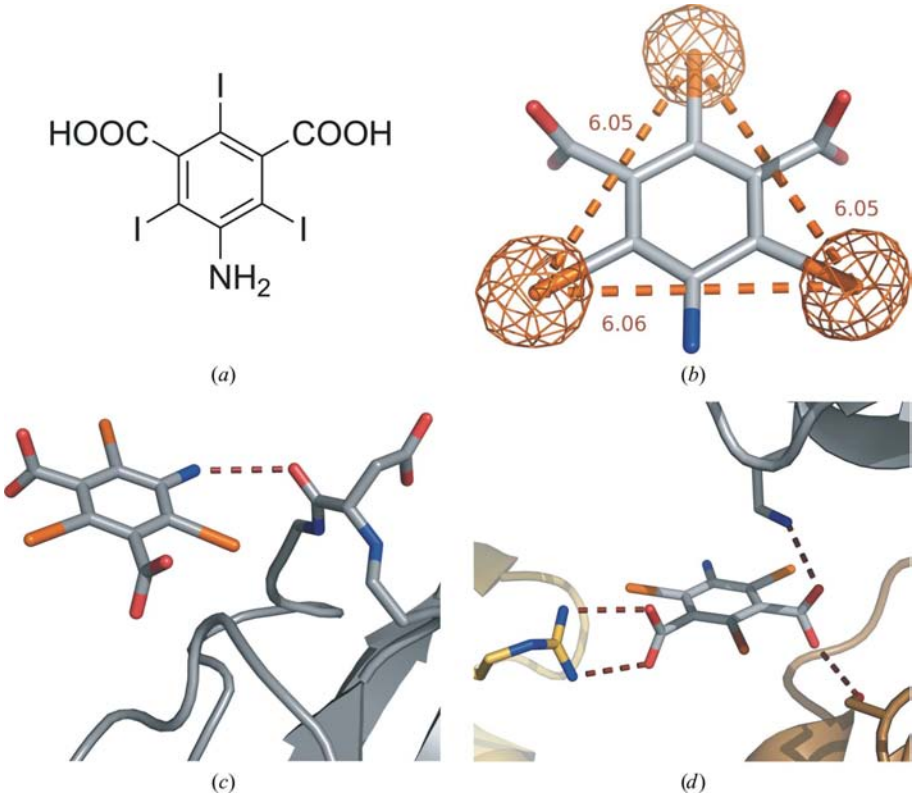

Figure 1
I3C and its interaction with proteins. (*a*) 5-Amino-2,4,6-triidoisophthalic acid (I3C) with three I atoms for anomalous scattering and three functional groups for hydrogen bonding. (*b*) One molecule of I3C in lysozyme; anomalous electron density at 4σ (orange), distances in Å. The equilateral triangle formed by the I atoms is clearly visible. (*c*) Hydrogen bonding of the amino group of I3C with the main-chain carbonyl O atom of Asp21 in thaumatin (other interactions are not shown). (*d*) Interaction of I3C with three lysozyme molecules. Hydrogen bonds to Arg73 (left), Lys33 (top right) and Ser24 (bottom right) are shown by dashed lines.

Restraints for refinement with *REFMAC* (CIF format) or *SHELXL*, together with coordinates in PDB format for the I3C molecule, may be downloaded from http://shelx.uni-ac.gwdg.de/tbeck. These are based on the small-molecule crystal structure of I3C (Beck & Sheldrick, 2008).

4. Conclusion

I3C represents a new class of compound that may be used for heavy-atom derivatization for SAD or SIRAS (single isomorphous replacement plus anomalous scattering) phasing, combining an easily recognizable arrangement of three anomalous scatterers with functional groups for hydrogen bonding to a protein molecule. Low toxicity and ready commercial availability are further advantages. Recently, I3C was used to solve an unknown structure of a 38 kDa protein that had resisted other phasing attempts (Sippel *et al.*, 2008, in this issue). It is also possible that the hydrogen donor and acceptor groups might promote crystal growth, since we observed several sites where I3C molecules acted as bridges between different protein molecules (*e.g.* Fig. 1*d*). It has been shown that similar small molecules promote crystal growth (McPherson & Cudney, 2006) and a new crystallization screen was recently introduced that includes such molecules (Silver

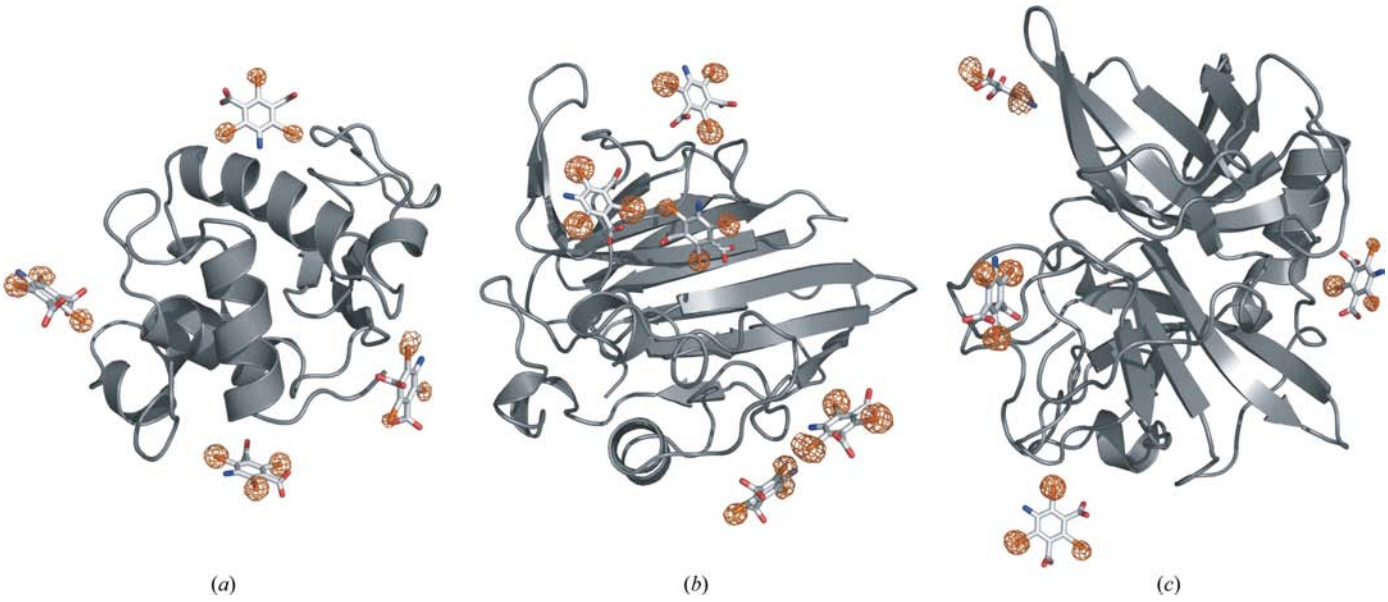

Figure 2
I3C in protein crystals: anomalous electron density shown around I3C at 4σ (orange). I3C molecules are found at the surface of the protein molecules. (*a*) Lysozyme with four molecules of I3C in the asymmetric unit. (*b*) Thaumatin with five molecules of I3C in the asymmetric unit. (*c*) Elastase with four molecules of I3C in the asymmetric unit.

Bullets Screen; HR2-096, Hampton Research, Aliso Viejo, California, USA). An introduction of I3C into standard crystallization screens is therefore desirable: it could promote crystal growth and introduce heavy atoms for phasing at the same time. Similar compounds could be used to introduce elements such as bromine or selenium that would be more suitable than iodine for MAD experiments or to exploit hydrophobic interactions as well as hydrogen bonds for binding to protein molecules.

This work was supported by the International Centre for Diffraction Data (Ludo Frevel Scholarship Award 2008 to TB), the German Research Foundation (DFG Graduate School IRTG 1422) and the Fonds der Chemischen Industrie.

References

Beck, T. & Sheldrick, G. M. (2008). *Acta Cryst.* E**64**, o1286.
Boggon, T. J. & Shapiro, L. (2000). *Structure*, **8**, R143–R149.
Dambe, T. R., Kühn, A. M., Brossette, T., Giffhorn, F. & Scheidig, A. J. (2006). *Biochemistry*, **45**, 10030–10042.
Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). *Acta Cryst.* D**56**, 232–237.
Debreczeni, J. É., Girmann, B., Zeeck, A., Krätzner, R. & Sheldrick, G. M. (2003). *Acta Cryst.* D**59**, 2125–2132.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Escalante, C. R., Yie, J., Thanos, D. & Aggarwal, A. K. (1998). *Nature (London)*, **391**, 103–106.
Ko, T.-P., Day, J., Greenwood, A. & McPherson, A. (1994). *Acta Cryst.* D**50**, 813–825.
McPherson, A. & Cudney, B. (2006). *J. Struct. Biol.* **156**, 387–406.
Miyatake, H., Hasegawa, T. & Yamano, A. (2006). *Acta Cryst.* D**62**, 280–289.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Ogata, C. M. (1998). *Nature Struct. Biol.* **5**, 638–640.
Perrakis, A., Morris, R. J. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. (2003). *Acta Cryst.* D**59**, 1131–1137.
Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.
Sheldrick, G. M. (2002). *Z. Kristallogr.* **217**, 644–650.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Sippel, K. H., Robbins, A. H., Reutzel, R., Domsic, J., Boehlein, S. K., Govindasamy, L., Agbandje-McKenna, M., Rosser, C. J. & McKenna, R. (2008). *Acta Cryst.* D**64**, 1172–1178.
Szczepanowski, R. H., Filipek, R. & Bochtler, M. (2005). *J. Biol. Chem.* **280**, 22006–22011.
Weiss, M. S. (2001). *J. Appl. Cryst.* **34**, 130–135.
Weiss, M. S., Palm, G. J. & Hilgenfeld, R. (2000). *Acta Cryst.* D**56**, 952–958.
Wernimont, A. K., Huffman, D. L., Lamb, A. L., O'Halloran, T. V. & Rosenzweig, A. C. (2000). *Nature Struct. Biol.* **7**, 766–771.
Würtele, M., Hahn, M., Hilpert, K. & Höhne, W. (2000). *Acta Cryst.* D**56**, 520–523.
Xie, J., Wang, L., Wu, N., Brock, A., Spraggon, G. & Schultz, P. G. (2004). *Nature Biotechnol.* **22**, 1297–1301.
Yu, S. B. & Watson, A. D. (1999). *Chem. Rev.* **99**, 2353–2378.