Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

The structure at 2.4 Å resolution of the protein from gene locus At3g21360, a putative Fe^{II}/2-oxo-glutarate-dependent enzyme from *Arabidopsis thaliana*

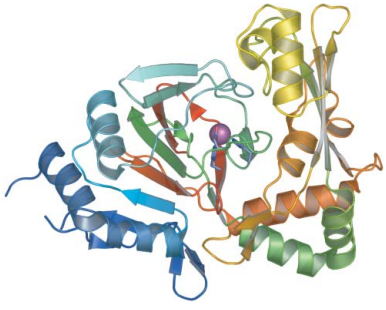
Eduard Bitto, Craig A. Bingman, Simon T. M. Allard, Gary E. Wesenberg, David J. Aceti, Russell L. Wrobel, Ronnie O. Frederick, Hassan Sreenath, Frank C. Vojtik, Won Bae Jeon, Craig S. Newman, John Primm, Michael R. Sussman, Brian G. Fox, John L. Markley and George N. Phillips Jr*

Center for Eukaryotic Structural Genomics, Department of Biochemistry, University of Wisconsin-Madison, USA

Correspondence e-mail: phillips@biochem.wisc.edu

Received 7 March 2005
Accepted 13 April 2005
Online 26 April 2005

PDB Reference: At3g21360, 1y0z, r1y0zsf.



The crystal structure of the gene product of At3g21360 from *Arabidopsis thaliana* was determined by the single-wavelength anomalous dispersion method and refined to an R factor of 19.3% (R_{free} = 24.1%) at 2.4 Å resolution. The crystal structure includes two monomers in the asymmetric unit that differ in the conformation of a flexible domain that spans residues 178–230. The crystal structure confirmed that At3g21360 encodes a protein belonging to the clavaminate synthase-like superfamily of iron(II) and 2-oxoglutarate-dependent enzymes. The metal-binding site was defined and is similar to the iron(II) binding sites found in other members of the superfamily.

1. Introduction

The At3g21360 gene of *Arabidopsis thaliana* encodes a protein with molecular weight of 37.2 kDa (residues 1–330) and a calculated isoelectric point of 5.97. Although the enzymatic function of this protein product in *A. thaliana* is not yet established, At3g21360 shares distant sequence homology with clavaminate synthase-like enzymes based on SUPERFAMILY server prediction (Gough *et al.*, 2001). Enzymes of this superfamily bind iron(II) ions and typically use 2-oxoglutarate (2OG) as a cofactor. The iron(II) ion is coordinated by the histidine and glutamate/aspartate residues of a His-X-Asp/Glu-X_n-His sequence motifs (Koehntop *et al.*, 2005). They catalyze a wide range of two-electron oxidation reactions; most representatives couple the oxidative decomposition of 2OG (forming CO_2 and succinate) to the hydroxylation of a co-substrate (Hausinger, 2004). Some members of the superfamily catalyze hydroxylation, desaturation, ring-expansion, ring-formation and other types of oxidative reactions. Here, we report the three-dimensional structure of At3g21360 protein at 2.4 Å determined by the single-wavelength anomalous dispersion (SAD) method. The structure was determined under the National Institutes of Health NIGMS Protein Structure Initiative.

2. Materials and methods

Native and SeMet-labeled At2g21360 proteins were cloned and purified following the standard CESG pipeline protocol described in detail elsewhere (Thao *et al.*, 2004; Sreenath *et al.*, 2005; Jeon *et al.*, 2005; Zolnai *et al.*, 2003). Crystals of At3g21360 were grown by the hanging-drop method from 10 mg ml^{-1} protein solution in buffer (50 mM NaCl, 3 mM NaN_3, 0.3 mM TCEP, 5 mM bis-Tris pH 6.0) mixed with an equal amount of well solutions containing 1.8 M ammonium sulfate, 16% glycerol, 100 mM HEPES pH 7.5 at 277 K. Crystals typically grew as tiny thin half-moon-shaped plates (100 × 30 × 10 µm). They belong to space group $C2$, with unit-cell parameters a = 145.3, b = 61.1, c = 114.7 Å, β = 121.7°. Crystals were cryoprotected by soaking in well solution supplemented with increasing concentrations of glycerol up to a final concentration of 30%. X-ray diffraction data for native and selenomethionine crystals were collected at the SBC 19-BM and BioCARS 14-ID beamlines at the Argonne National Laboratory Advanced Photon Source, respectively. The selenium substructure of SeMet-labeled At3g21360 crystals was determined using *HySS* (Grosse-Kunstleve & Adams, 2003). The protein structure was phased using SAD data in *CNS* (Brünger *et al.*, 1998). Initial phase information obtained from *CNS*

was improved and extended to the higher resolution of a native data set by averaging and electron-density modification as implemented in *RESOLVE* (Terwilliger, 2000). The automatic tracing procedure of *ARP/wARP* (Perrakis *et al.*, 1999) produced an inital model with approximately 80% residues placed, of which 60% had side chains assigned. The structure was completed using alternate cycles of manual building in *Xfit* (McRee, 1999) and refinement in *CNS*. All refinement steps were monitored using an $R_{\rm free}$ value based on 5.3% of the independent reflections. The stereochemical quality of the final model was assessed using *PROCHECK* (Laskowski *et al.*, 1993) and *MolProbity* (Lovell *et al.*, 2003).

3. Results and discussion

The structure of At3g21360 has been solved to a resolution of 2.4 Å. Data collection, refinement and model statistics are summarized in Table 1. The final model includes two independent monomers representing residues 4–29 and 36–330 from monomer *A* and residues 2–91 and 99–330 from monomer *B*. In addition, two putative Fe^{2+} ions, four sulfate ions and 780 water molecules were built into the final structure. To classify the fold of At3g21360, a structural homology search was conducted using the DALI server (Holm & Sander, 1993). At the time of deposition to the PDB, analysis of the At3g21360 sequence by FFAS (Jaroszewski *et al.*, 2000) showed an overall 12–16% sequence identity to several prior PDB entries from a clavaminate synthase-like superfamily (CLS; Jaroszewski *et al.*, 2000). Close structural homologs ($Z > 15$) of At3g21360 produced by DALI include several established members of the CLS superfamily of iron(II)/2OG-binding enzymes: specifically, taurine dioxygenase with $Z = 19.2$, r.m.s.d. 3.2 Å over 242 aligned C^{α} residues, 15% sequence identity (PDB code 1gqw; Elkins *et al.*, 2002); carbapenem synthase with $Z = 17.1$, r.m.s.d. 2.8 Å over 216 aligned residues, 17% sequence identity (PDB code 1nx4; Clifton *et al.*, 2003); hypothetical 37.4 kDa protein with $Z = 16.7$, r.m.s.d. 3.3 Å over 236 aligned residues, 17% sequence identity (PDB code 1jr7); clavaminate synthase with $Z = 15.1$, r.m.s.d. 3.0 Å over 217 aligned residues, 19% identity (PDB code 1drt; Zhang *et al.*, 2000). In addition to close structural homologs of At3g21360 revealed by DALI, the profile–profile sequence tool FFAS predicted another close homolog based on the primary sequence, namely alkylsulfatase AtsK (PDB code 1oih), with an FFAS score of −84.2 and 12% sequence identity over the aligned range (Muller *et al.*, 2004). Structural superposition of At3g21360 and alkylsulfatase AtsK confirmed this prediction.

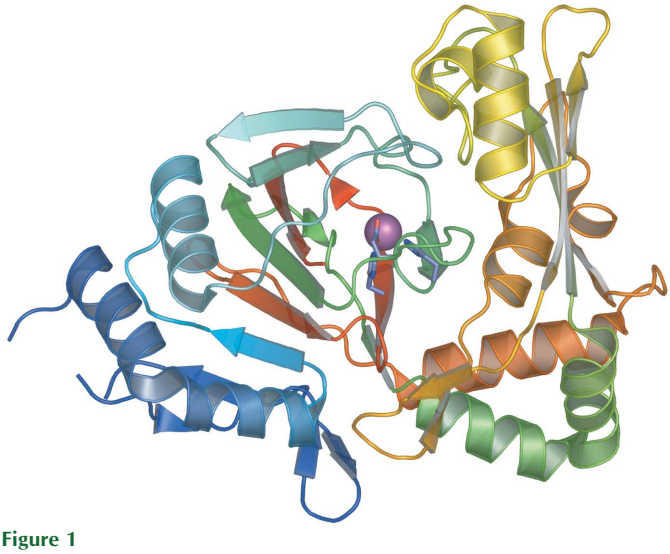


Figure 1
A ribbon diagram of the crystal structure of At3g21360 color-coded from the amino-terminus (blue) to carboxy-terminus (red). The putative Fe_{2+} ion is depicted as a purple sphere; three protein residues (His120, Glu122, His313) involved in its coordination are shown as blue sticks. The figure was generated using *PyMol* (DeLano, 2002).

Table 1
Summary of crystal parameters, data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

	Native	Se-peak
Space group	*C*2	
Unit-cell parameters (Å, °)	*a* = 145.3, *b* = 61.1, *c* = 114.7, $\alpha = 90$, $\beta = 121.68$, $\gamma = 90$	
Data-collection and phasing statistics		
Energy (keV)	12.658	12.661
Wavelength (Å)	0.97954	0.97927
Resolution range (Å)	43.63–2.4 (2.46–2.4)	43.51–2.85 (2.92–2.85)
No. of reflections (measured/unique)	123132/34020	151123/20177
Completeness (%)	99.8 (99.9)	99.8 (97.7)
$R_{\rm merge}$†	0.079 (0.283)	0.069 (0.268)
Redundancy	3.6 (3.5)	7.5 (6.4)
Mean $I/\sigma(I)$	12.57 (4.54)	15.99 (4.90)
Sigma cutoff	0.0	0.0
Mean FOM of phasing		0.3509
Refinement and model statistics		
Resolution range (Å)	43.63–2.40	
Data set used in refinement	Native	
No. of reflections (total/test)	32733/1628	
$R_{\rm cryst}$‡	0.193	
$R_{\rm free}$§	0.241	
R.m.s.d. bonds (Å)	0.006	
R.m.s.d. angles (°)	1.30	
E.s.d. from σ_A¶ (Å)	0.35	
Average *B* factor (Å^2)	35.60	
No. of water molecules	780	
Ramachandran plot, residues in		
Most favorable region (%)	89.6	
Additional allowed region (%)	10.2	
Generously allowed region (%)	0.2	
Disallowed region (%)	0.0	

† $R_{\rm merge} = \sum_h \sum_i |I_i(h) - \langle I(h)\rangle| / \sum_h \sum_i I_i(h)$, where $I_i(h)$ is the intensity of an individual measurement of the reflection and $\langle I(h)\rangle$ is the mean intensity of the reflection. ‡ $R_{\rm cryst} = \sum_h ||F_{\rm obs}| - |F_{\rm calc}|| / \sum_h |F_{\rm obs}|$, where $F_{\rm obs}$ and $F_{\rm calc}$ are the observed and calculated structure-factor amplitudes, respectively. § $R_{\rm free}$ was calculated as $R_{\rm cryst}$ using 5.3% of the randomly selected unique reflections that were omitted from structure refinement ¶ Cross-validated estimated coordinate error from cross-validated σ_A (Brünger *et al.*, 1998).

The central structural feature of At3g21360 is a double-stranded β-helix fold with a jelly-roll topology (see Fig. 1). This is the most structurally conserved part of the CSL superfamily (Hausinger, 2004). When the backbone of At3g21360 is compared in detail with that of taurine dioxygenase (Elkins *et al.*, 2002), the closest At3g21360 structural homolog as revealed by DALI, several striking differences can be observed (see Fig. 2*a*). An internal domain spanning residues 150–292 in At3g21360 (129–234 in taurine dioxygenase) shows a large difference in backbone conformation. Major differences are found in a subregion spanning residues 177–230 of At3g21360 (154–183 in taurine dioxygenase). These regions differ in length by as much as 24 residues in the respective proteins. While taurine dioxygenase has a helix followed by an extended coil returning to the same area to form a complementary β-strand, At3g21360 folds into a small domain with three antiparallel β-strands, a helix and an extended loop in an approximately perpendicular direction to the direction of the antiparallel β-sheet. Residues 254–268 of At3g21360 also form a loop not observed in the taurine dioxygenase structure. This loop seems to stabilize the subdomain 177–230 through a salt bridge between interloop residues Asp186 and Arg262. Further notable differences are also found in the amino-terminal part of the two proteins, where a region spanning residues 29–49 in At3g21360 forms an extended loop incorporating a three-turn helix. Finally, residues 68–116 in

At3g21360 show a dramatic displacement from the corresponding residues 49–95 of taurine dioxygenase.

The putative iron(II) metal-binding site was clearly identified in At3g21360 during structure refinement based on side-chain coordination typical of a metal-binding site and a strong anomalous diffraction peak located at the metal position. In At3g21360, residues His120, Glu122 and His313 are involved in the coordination of the ion. The presence of a glutamate instead of an aspartate as a ligand is quite unusual in members of the CSL superfamily (Hausinger, 2004). To the best of our knowledge, the only previously structurally documented member with an active-site glutamate is that of *Streptomyces clavuligerus* clavaminate synthase (Zhang *et al.*, 2000). Based on the crystal structures of several CSL superfamily members, 2OG typicaly chelates iron(II) using its C2-keto group and C1-carboxylate group, while its C5-carboxylate group usually interacts electrostatically with a conserved arginine residue. Inductively coupled plasma metal analysis of protein samples used for crystallization revealed the presence of submillimolar concentrations of nickel and copper. However, no detectable amount of iron was found. We suspect that iron(II) may be the natural metal ion, but have not pursued functional studies to prove the metal identity. The high content of nickel in the analyzed samples is likely to be an artefact caused by use of nickel-affinity chromatography during protein purification. To the best of our knowledge, it has not yet been established whether At3g21360 uses 2OG as a co-substrate. The arginine residue expected to be involved in salt-bridge formation with 2OG C5-carboxylate is conserved in At3g21360 (Arg332) and is found facing a cavity occupied by a sulfate ion. This indicates that the cavity is accessible to small molecules and stabilization of 2OG through electrostatic interaction could indeed occur.

Interesting differences were observed between the two monomers in the asymmetric unit of the At3g21360 crystals. A domain formed by residues 178–230 rotates by several degrees and closes the catalytic cavity in monomer *A* (see Fig. 2*b*). This conformational change is coupled with a partial ordering of a loop spanning residues 91–99 located directly above the catalytic site. A comparison of the At3g21360 structure to those of several other superfamily members co-crystallized in the presence of their respective substrates suggests that residues 91–99 are located directly above the active site and are likely to be responsible for modulating the enzyme's substrate specificity.

A. thaliana encodes around 125 members of the CSL superfamily. At this time, the only structurally characterized member is an anthocyanidin synthase (ANS; Wilmouth *et al.*, 2002). This enzyme is involved in the oxidation of leucocyanidin to cyanidin. Both of these compounds belong to an anthocyanin class of flavonoids, common plant water-soluble red, blue and violet pigments responsible for flower coloring.

Another example of an *A. thaliana* CSL superfamily protein is a gibberellin 20-oxidase (GA) involved in the synthetic pathway of gibberellins (Xu *et al.*, 1995). The gibberellins are tetracyclic diterpenoid plant products involved in many aspects of plant growth and development, with a definitive role in the mediation of photoperiodic control of stem elongation. Although GA is not yet structurally characterized, it shows sequence homology (15–26% sequence identity) to several other structurally characterized members of the CSL superfamily, including anthocyanidin synthase (PDB code 1gp4; Wilmouth *et al.*, 2002), isopenicillin N synthase (PDB code 1bk0; Roach *et al.*, 1997) and deacetoxycephalosporin C synthase (PDB code 1dsc; Valegard *et al.*, 1998).

The family relationship of ANS and GA to At3g21360 is much weaker and is nearly undetectable by sequence-analysis methods (around 9% identity). However, ANS and At3g21360 share similar structural folds; a DALI search with At3g21360 as a structural template produced a score for ANS of Z = 5.1, r.m.s.d. 3.8 Å over 123 aligned residues, with 12% sequence identity.

In conclusion, the presented crystal structure of At3g21360 establishes that this protein is a member of a clavaminate synthase-like superfamily in *A. thaliana*. Although the identity of the At3g21360 subtrate is not yet defined, one would expect At3g21360 to be involved in an 2OG-dependent oxidative reaction in the synthetic pathway of a plant product.

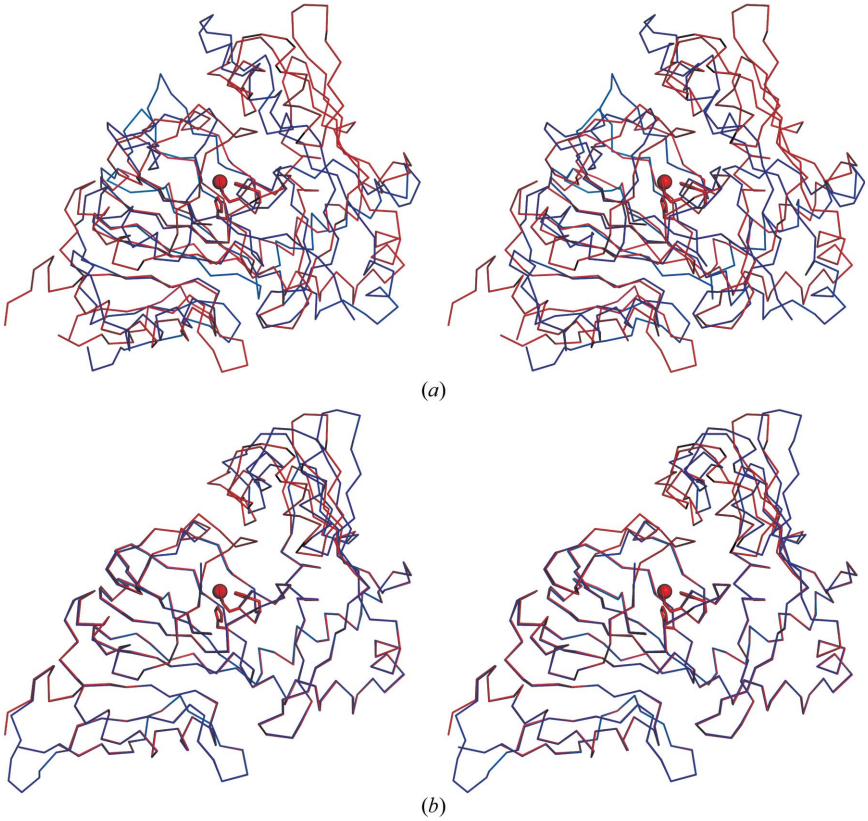


Figure 2
(*a*) A stereo diagram of the main-chain trace of the crystal structure of At3g21360 (red) superimposed onto the crystal structure of its closest structural homolog taurine dioxygenase (blue). The structures were superimposed using the pairwise structure-comparison server SSAP (Taylor *et al.*, 1994). (*b*) A stereo diagram of the main-chain trace of the crystal structure of monomer *A* of At3g21360 (red) superimposed onto the crystal structure of monomer *B* from the asymmetric unit (blue). A domain formed by residues 178–230 (projected to the upper-right quadrant) rotates by several degrees. In both panels, the iron(II) ion is depicted as a red sphere. Three At3g21360 residues (His120, Glu122, His313) involved in its coordination are shown as red sticks. The figure was generated using *PyMol* (DeLano, 2002).

We acknowledge financial support from NIH National Institute for General Medical Sciences grant P50 GM64598. Use of the Advanced Photon Source and the Argonne National

Laboratory Structural Biology Center beamlines at the Advanced Photon Source was supported by the US Department of Energy, Office of Energy Research under contract No. W-31-109-ENG-38. Use of BioCARS Sector 14 was supported by the National Institutes of Health, National Center for Research Resources under grant No. RR07707. We would like to thank 19-BM SBC beamline scientists Norma E. C. Duke PhD and Stephan L. Ginell PhD for help we have received during data collection. Special thanks goes to all members of the CESG team, including Todd Kimball, John Kunert, Nicholas Dillon, Rachel Schiesher, Juhyung Chin, Megan Riters, Andrew C. Olson, Jason M. Ellefson, Janet E. McCombs, Brendan T. Burns, Blake W. Buchan, Holalkere V. Geetha, Zhaohui Sun, Ip Kei Sam, Eldon L. Ulrich, Nathan S. Rosenberg, Janelle Warrick, Zsolt Zolnai, Peter T. Lee and Jianhua Zhang.

References

Brünger, 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.

Clifton, I. J., Doan, L. X., Sleeman, M. C., Topf, M., Suzuki, H., Wilmouth, R. C. & Schofield, C. J. (2003). *J. Biol. Chem.* **278**, 20843–20850.

DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* DeLano Scientific, San Carlos, CA, USA.

Elkins, J. M., Ryle, M. J., Clifton, I. J., Dunning Hotopp, J. C., Lloyd, J. S., Burzlaff, N. I., Baldwin, J. E., Hausinger, R. P. & Roach, P. L. (2002). *Biochemistry*, **41**, 5185–5192.

Gough, J., Karplus, K., Hughey, R. & Chothia, C. (2001). *J. Mol. Biol.* **313**, 903–919.

Grosse-Kunstleve, R. W. & Adams, P. D. (2003). *Acta Cryst.* D**59**, 1966–1973.

Hausinger, R. P. (2004). *Crit. Rev. Biochem. Mol. Biol.* **39**, 21–68.

Holm, L. & Sander, C. (1993). *J. Mol. Biol.* **233**, 123–238.

Jaroszewski, L., Rychlewski, L. & Godzik, A. (2000). *Protein Sci.* **9**, 1487–1496.

Jeon, W., Aceti, D. J., Bingman, C., Vojtik, F., Olson, A., Ellefson, J., McCombs, J., Sreenath, H., Blommel, P., Seder, K., Buchan, B., Burns, B., Geetha, H., Harms, A., Sabat, G., Sussman, M., Fox, B. & Phillips, G. (2005). In the press.

Koehntop, K. D, Emerson J. P. & Que L. Jr (2005). *J. Biol. Inorg. Chem.* **10**, 87–93.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Lovell, S. C., Davis, I. W., Arendell, W. B. III, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003). *Proteins*, **50**, 437–450.

McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.

Muller, I., Kahnert, A., Pape, T., Sheldrick, G. M., Meyer-Klaucke, W., Dierks, T., Kertesz, M. & Uson, I. (2004). *Biochemistry*, **43**, 3075–3088.

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

Roach, P. L. Clifton, I. J., Hensgens, C. M., Shibata, N., Schofield, C. J., Hajdu, J. & Baldwin, J. E. (1997). *Nature (London)*, **387**, 827–830.

Sreenath, H., Bingman, C., Buchan, B., Seder, K., Burns, B., Geetha, H., Jeon, W., Vojtik, F., Aceti, D., Frederick, R., Phillips, G. J. & Fox, B. (2005). In the press.

Taylor, W. R., Flores, T. P. & Orengo, C. A. (1994). *Protein Sci.* **3**, 1858–1870.

Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.

Thao, S., Zhao, Q., Kimball, T., Steffen, E., Blommel, P. G., Riters, M., Newman, C., Fox, B. & Wrobel, R. (2004). *J. Struct. Funct. Genomics*, **5**, 255–265.

Valegard, K., van Scheltinga, A. C., Lloyd, M. D., Hara, T., Ramaswamy, S., Perrakis, A., Thompson, A., Lee, H. J., Baldwin, J. E., Schofield, C. J., Hajdu, J. & Andersson, I. (1998). *Nature (London)*, **394**, 805–809.

Wilmouth, R. C., Turnbull, J. J., Welford, R. W., Clifton, I. J., Prescott, A. G. & Schofield, C. J. (2002). *Structure*, **10**, 93–103.

Xu, Y. L., Li, L., Wu, K., Peeters, A. J., Gage, D. A. & Zeevaart, J. A. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 6640–6644.

Zhang, Z., Ren, J., Stammers, D. K., Baldwin, J. E., Harlos, K. & Schofield, C. J. (2000). *Nature Struct. Biol.* **7**, 127–133.

Zolnai, Z., Lee, P. T., Li, J., Chapman, M. R., Newman, C. S., Phillips, G. N. Jr, Rayment, I., Ulrich, E. L., Volkman, B. F. & Markley, J. L. (2003). *J. Struct. Funct. Genomics*, **4**, 11–23.