PROTEIN STRUCTURE REPORT

Solution structure of a single-domain thiosulfate sulfurtransferase from *Arabidopsis thaliana*

GABRIEL CORNILESCU,[1] DMITRIY A. VINAROV,[2] EJAN M. TYLER,[2] JOHN L. MARKLEY,[1,2] AND CLAUDIA C. CORNILESCU[1,2]

[1]National Magnetic Resonance Facility at Madison, Biochemistry Department, University of Wisconsin–Madison, Madison, Wisconsin 53706-1544, USA
[2]Center for Eukaryotic Structural Genomics, Biochemistry Department, University of Wisconsin–Madison, Madison, Wisconsin 53706-1544, USA

(RECEIVED June 12, 2006; FINAL REVISION September 12, 2006; ACCEPTED September 18, 2006)

Abstract

We describe the three-dimensional structure of the product of *Arabidopsis thaliana* gene *At5g66040.1* as determined by NMR spectroscopy. This protein is categorized as single-domain sulfurtransferase and is annotated as a senescence-associated protein (sen1-like protein) and ketoconazole resistance protein (http://arabidopsis.org/info/genefamily/STR_genefamily.html). The sequence of At5g66040.1 is virtually identical to that of a protein from *Arabidopsis* found by others to confer ketoconazole resistance in yeast. Comparison of the three-dimensional structure with those in the Protein Data Bank revealed that At5g66040.1 contains an additional mobile β-hairpin not found in other rhodaneses that may function in binding specific substrates. This represents the first structure of a single-domain plant sulfurtransferase. The enzymatically active cysteine-containing domain belongs to the CDC25 class of phosphatases, sulfide dehydrogenases, and stress proteins such as senescence specific protein 1 in plants, PspE and GlpE in bacteria, and cyanide and arsenate resistance proteins. Versions of this domain that lack the active site cysteine are found in other proteins, such as phosphatases, ubiquitin hydrolases, and sulfuryltransferases.

Keywords: At5g66040.1; single-domain sulfurtransferase; rhodanese; CESG; Center for Eukaryotic Structural Genomics; NMR

Sulfurtransferases/rhodaneses (Str) are a group of enzymes whose genes have been identified in all organisms studied to date. They catalyze the transfer of a sulfur atom from a suitable sulfur donor to a nucleophilic acceptor. The best-studied member of the family is bovine rhodanese, which catalyzes the transfer of a sulfane sulfur atom from thiosulfate to cyanide to produce sulfite plus thiocyanate (Westley 1973). Other important members of the Str family are 3-mercaptopyruvate sulfurtransferases (3-MSTs), which transfer sulfur from mercaptopyruvate to cyanide. Str proteins are found as single domains or, more frequently, as tandem repeats with the active site always present in the C-terminal domain. The biological functions for most members of the rhodanese superfamily remain to be determined. As suggested in part by their cellular localization, various functions have been proposed: cyanide detoxification for Strs present in liver cells (Vennesland et al. 1982), sulfur insertase for iron-sulfur cluster formation and repair (Pagani et al. 1984), involvement in sulfur metabolism (Nandi and Westley 1998), or interaction with thioredoxin (Ray et al. 2000).

Arabidopsis thaliana At5g66040.1 was chosen as a ''sequence to structure-space target'' by the Center for Eukaryotic Structural Genomics (CESG). At the time the structure was solved, the closest structure in the Protein Data Bank (PDB) had 22% sequence identity (>70 residues)

Reprint requests to: Gabriel Cornilescu, National Magnetic Resonance Facility at Madison, Biochemistry Department, University of Wisconsin–Madison, Madison, Wisconsin 53706-1544, USA; e-mail: gabrielc@nmrfam.wisc.edu; fax: (608) 262-3759.

Article published online ahead of print. Article and publication date are at http://www.proteinscience.org/cgi/doi/10.1110/ps.062395206.

to At5g66040.1. The 129-residue protein (ORF no. At5g66040.1) was cloned from an *A. thaliana* cDNA library, and the protein was produced by CESG's wheat germ cell–free platform (Vinarov et al. 2004).

At5g66040.1 was one of 18 rhodanese domain-containing genes identified by detailed screening of the *Arabidopsis* genome (Bauer and Papenbrock 2002); it was classified, along with four other genes as a ''group VI'' Str, on the basis of conservation of the active site cysteine, the rhodanese signature sequence, and 12 further conserved amino acids. In vitro activity studies of At5g66040.1 showed high rhodanese activity but nearly undetectable 3-MST activity (Bauer and Papenbrock 2002). A subsequent study showed that At5g66040.1 is localized in chloroplasts (Bauer et al. 2004). Because two other *Arabidopsis* Strs were also shown to be localized in plastids, a putative role in iron-sulfur cluster biosynthesis was suggested, similar to that of plastidic NifS-like Cys desulfhydrase (Leon et al. 2003).

The sequence of At5g66040.1 is virtually identical (95% sequence identity) to that of a ketoconazole-resistant protein (GI:928938) obtained by transformation of a yeast sterol mutant erg3 with an *Arabidopsis* cDNA library and subsequent selection with ketoconazole. The At5g66040.1 sequence also is very similar (sequence identity >60%) to those of several members of the senescence-associated family proteins.

Results

Resonance assignments and secondary structure

The protein studied (here denoted as At5g66040.1) has the sequence of the ORF plus an N-terminal purification tag: $MG(H)_6LE-$. ^{15}N T_2 measurements with ~0.7 mM [*U*-^{15}N]-At5g66040.1 (Fig. 1B) yielded uniform values ~100 msec for the rigid part of the molecule; this suggested that the protein is monomeric in solution under these conditions. Average ^{15}N T_2 values and the line widths of [^1H-^{15}N] HSQC cross peaks determined for [*U*-^{15}N]-At5g66040.1 showed no concentration dependence in the range 0.1−0.9 mM protein; this indicated that the protein remains monomeric at higher concentrations.

The limited stability of the sample over time made it difficult to use automated peak assignment methods. Instead, ~87% of the backbone and side-chain resonances were assigned manually using the PIPP/STAPP software (Garrett et al. 1991).

The TALOS program (Cornilescu et al. 1999) was used to provide 85 pairs of φ/ψ backbone torsion angle restraints and to identify the secondary structural elements (confirmed by local NOEs) (Fig. 1A). The high mobility of residues 11−15 (as suggested by the T_2 values) (Fig. 1B) was confirmed by their backbone chemical shifts appearing close to the random coil values in the TALOS graphical output.

Hydrogen bond restraints were inferred initially for α-helices and later for β-strands, when the level of structural refinement allowed their unambiguous alignment within the β-sheet. Two distance restraints of 1.9 Å and 2.9 Å per involved pair of residues were used to represent hydrogen bonds for H^N-O and N-O, respectively (Wüthrich 1986).

Peak intensities in two 3D NOESY spectra were assigned using the PIPP/STAPP package and converted into a continuous distribution of 2212 approximate interproton distance restraints, with a uniform 40% distance error applied to take into account spin diffusion.

Structure determination

We attempted to solve the structure by automatic NOE assignments using the CANDID iterative protocol of CYANA (Herrmann et al. 2002), but the resulting structures did not converge, apparently due to minute peak shifting caused by the sample instability. After manual NOE assignments, a CYANA calculated structure served as an initial fold for a subsequent XPLOR simulated annealing protocol with a starting temperature of 3000 K.

Further structure calculation and refinement made use of the torsion angle molecular dynamics and the internal variable dynamics modules (Schwieters and Clore 2001) of Xplor-NIH (Schwieters et al. 2003).

At5g66040.1 adopts a βαβαββββαβαβαβ fold (Fig. 2A) that contains a characteristic rhodanese domain, with a central β-sheet flanked on both sides by α-helices. A summary of the agreement between experimental constraint and calculated structures is provided in Table 1, along with structural statistics and independent measures of structural quality. The refinement protocol resulted in a subset of five lowest energy structures (out of 20) with a backbone root mean square deviation (RMSD) from the average of 0.87 Å, excepting the disordered N terminus and residues 56−84. The latter excluded region, connecting strands β4 and β7 of the central β-sheet, contains a long, irregular β-hairpin loop followed by a short α-helix. This loop, not found in any of the similar rhodanese folds, shows a slight indication of mobility by ^{15}N T_2 relaxation criteria (Fig. 1B) and has no NOE contacts with the rest of the domain; this resulted in a wide conformational sampling of this loop in the calculated structures (Fig. 2A)

The final set of structures did not show any consistent (i.e., in >35% of the calculated structures) NOE violations larger than 0.5 Å. The result of a PROCHECK analysis showed 94% of the residues in the most favored and 6% in the allowed regions of the Ramachandran map (Table 1).

Discussion

The most similar rhodanese structures in PDB, found using a VAST search (Gibrat et al. 1996), were GlpE, an

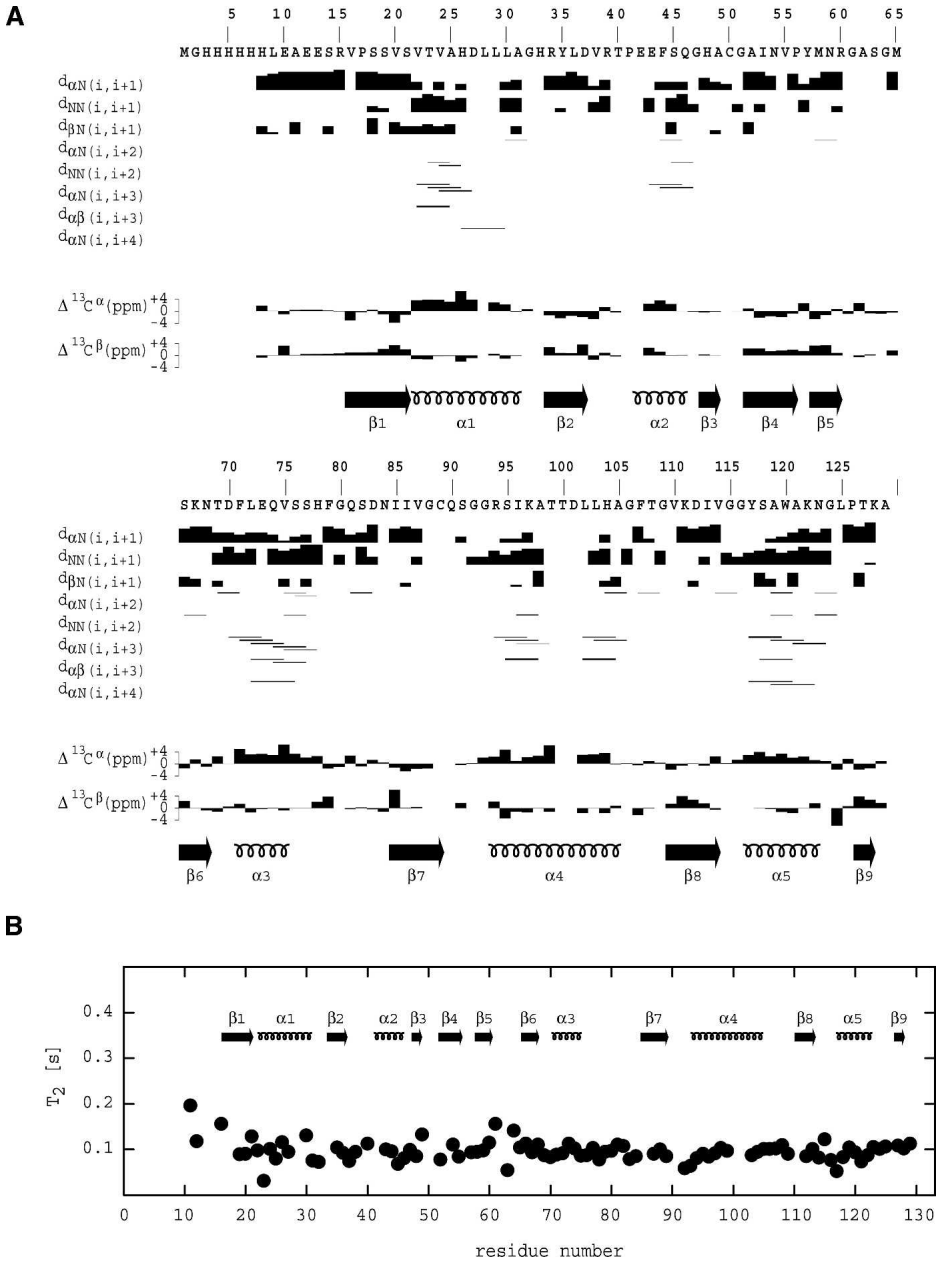

Figure 1. (*A*) Summary of secondary structure elements, local NOE connectivities, and C^{α}/C^{β} chemical shifts vs. the amino acid sequence of At5g66040.1. (*B*) ^{15}N T_2 values plotted as a function of residue number and secondary structure. Protein concentration ~0.7 mM, pH 6.5, and temperature 25°C.

Escherichia coli prototype sulfurtransferase for the single-domain rhodanese homology superfamily (PDB ID 1GMX), with 22% sequence identity over the 77 residues forming the Rhodanese homology domain, the catalytic domain of the human Cdc25 phosphatase (PDB ID 1C25), and a polysulfide-sulfurtransferase (homodimer) from *Wolinella succinogenes* (PDB ID 1QXN). A comparison of these closely related structures (or domains) to At5g66040.1 is shown in Figure 3. Although these domains show minimal sequence identity, they do show common differences from the structures of bovine (Rhobov) or *Azotobacter vinelandii* (RhdA) rhodanese: These differences include shortened loops connecting the α-helices and β-sheets, which result in a convex active site region in contrast to the corresponding concave areas of Rhobov and RhdA (Spallarossa et al. 2001).

The active site cysteine is located either at the end of the central (and longest) β-strand of the core five-strand

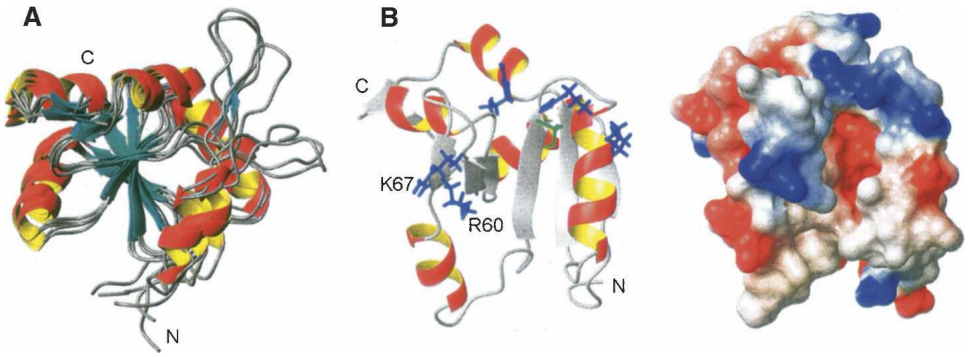

Figure 2. (*A*) Ensemble of the five lowest energy backbone structures of At5g66040.1 with α-helices shown in red/yellow and β-strands in cyan; the N and C termini of the protein are labeled. (*B*) Electrostatic potential surface plot (*right*) with a corresponding view of the lowest energy structure as a ribbon diagram, with labeled N and C termini (*left*). Shown in blue are the positively charged side-chains close to the active site cysteine (shown in green).

β-sheet or as the beginning of the adjacent β-α loop. The side-chains of the residues on this conserved loop point toward the essential cysteine, creating a cradle-like arrangement exposed to a positive charged field of the residues in the loop and the neighboring basic residues (Fig. 2B). Again, even in this essential catalytic cysteine

Table 1. *Structural statistics*

	RMSD	
	<Sa> monomer[a]	Lowest energy monomer
Experimental distance restraints (Å) (2212)[a]		
Intraresidue (1168)	0.086 ± 0.008	0.095
Sequential (|i – j| = 1) (563)	0.080 ± 0.019	0.109
Short range (1 < |i – j| ≤ 5) (235)	0.154 ± 0.040	0.184
Long range (|i – j| > 5) (246)	0.101 ± 0.016	0.098
Inferred hydrogen bonds (Å) (84)	0.063 ± 0.020	0.097
Predicted dihedral restraints (°)[b] (166)	0.832 ± 0.119	1.038
Deviations from idealized covalent geometry		
Bonds (Å) (1921)	0.00019 ± 0.00008	0.00020
Angles (°) (3447)	0.324 ± 0.010	0.330
Impropers (°) (1028)	0.339 ± 0.020	0.379
Measures of structure quality		
Lennard-Jones energy (kcal/mol)[c]	−431 ± 19	−441
Ramachandran plot (%)		
Residues 16–55, 85–128 in:		
Most favored regions	93.1	92.9
Allowed regions	6.9	7.1
Disallowed regions	0	0
Residues 16–55, 71–77, 85–128 in:		
Most favored regions	93.8	93.5
Allowed regions	6.2	6.5
Disallowed regions	0	0
Coordinate precision (Å)[d]		
Residues 16–55, 85–128:		
Backbone (N, C^{α}, C′)	0.87 ± 0.13	
All nonhydrogen atoms	1.41 ± 0.20	
Residues 16–55, 71–77, 85–128:		
Backbone (N, C^{α}, C′)	1.02 ± 0.20	
All nonhydrogen atoms	1.60 ± 0.32	

The statistics are for the five structures (of 20 calculated) with the overall lowest energies.
[a] <SA> represents the five lowest-energy structures. For <SA>, the values shown are mean ± SD, with the number of restraints used to calculate these values shown in parentheses.
[b] Error range used for the φ/ψ constraints is twice the standard deviation of the TALOS predicted angles.
[c] The van der Waals energies described by Lennard-Jones potentials were not incorporated into the simulated annealing calculation.
[d] Average RMSD of the five lowest-energy structures from the mean coordinates.

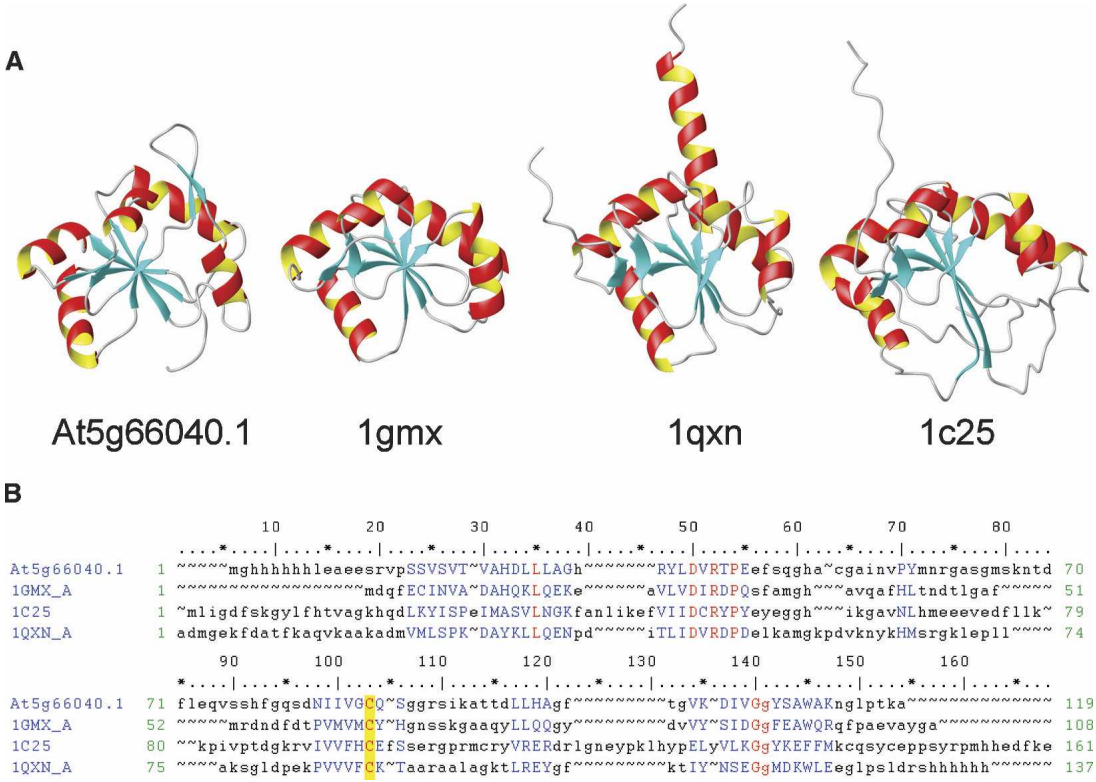

Figure 3. (*A*) Comparison of ribbon diagrams for At5g66040.1; GlpE, an Str from *E. coli* (PDB ID 1GMX); a polysulfide-sulfur transferase (homodimer) from *Wolinella Succinogenes* (PDB ID 1QXN); and the catalytic domain of the human Cdc25 phosphatase (PDB ID 1C25). (*B*) The VAST domain alignment of the above structures, with the active cysteine highlighted.

loop, the sequence similarity is negligible, but conservation in structure-based alignment exists for a couple of charged residues corresponding to Asp37 and Arg39 in At5g66040.1. These residues, buried in the hydrophobic core, form a hydrogen-bonded salt bridge in all known rhodanese and Cdc25 phosphatase structures and clearly contribute to fold stability.

The striking distinctive feature of the At5g66040.1 structure, when compared with all similar rhodanese domains, is an extra β-hairpin connecting the β1α1β2α2β3β4 and α3β7α4β8α5β9 structurally conserved elements (Fig. 3A). The ^{15}N T_2 measurement suggests some mobility in this β-hairpin (Fig. 1B), which is confirmed by the lack of NOE contacts to the hydrophobic core of the protein. Since this β-hairpin is located on the same side of the protein as the catalytic cysteine loop, its partial mobility suggests that it may play a role in binding a specific substrate. In addition to the positively charged or polar residues located in or near the catalytic cysteine loop in most single domain Str enzymes, another two charged residues (Arg 60 and Lys67) with side-chains oriented outward (Fig. 2B) are located on this extra β-hairpin, with a possible role in modulating the substrate binding specificity of At5g66040.1.

With the advent of nearly complete sequencing of multiple genomes, multiple genes that encode rhodanese-like domains have been identified within individual genomes, suggesting a diverse biological functionality confirmed by the residue variability in the putative active site region and by the localization in different cellular compartments. Identification of the specific substrates is therefore important for establishing the mechanism of biological functionality.

In conclusion, we solved by NMR the first structure of single-domain thiosulfate sulfurtransferase enzyme from a plant. This structure exhibits a unique structural feature: an additional mobile β-hairpin with putative functional role in binding a specific substrate. The coordinates and NMR structural restraints have been deposited in the PDB (ID 1TQ1), and the NMR assignments and time-domain data have been deposited in the BMRB (accession no. 6240).

Materials and methods

Protein expression and purification

Samples of the gene product of *A. thaliana* gene At5g66040.1 were prepared in [*U*-^{13}C,^{15}N]-labeled form according to CESG

wheat germ cell–free protocols as previously described (Vinarov et al. 2004). Briefly, the protein was expressed with an N-terminal His6 fusion tag in wheat germ extract supplemented with [*U*-^{13}C,^{15}N] amino acids (Cambridge Isotope Labs) and purified by HisTrap HP Chelating chromatography, followed by size-exclusion chromatography.

NMR measurements

All NMR data were acquired at 25°C on 280 μL samples at pH 6.5 containing 0.6–0.9 mM ^{13}C,^{15}N-protein, 10 mM $KHPO_4$, 50 mM NaCl, and 1 mM DTT in 93% H_2O and 7% ^2H_2O.

[^{1}H-^{15}N] HSQC, HNCO, HNCACB, CBCA(CO)NH, HBHA-CONH, CCONH, HCCONH, HCCH-TOCSY, 3D ^{15}N-edited NOESY (t_{mix} = 100 msec, at 600 MHz), and 3D ^{13}C-edited NOESY (t_{mix} = 120 msec, at 750 MHz) spectra were acquired on Varian INOVA 600-MHz, and Bruker AVANCE 500 (equipped with cryogenic probe) and 750-MHz spectrometers. Data were processed using the NMRPipe package (Delaglio et al. 1995).

Our attempts to measure residual dipolar couplings (RDCs) resulted in either unfolding of the protein in 5% (w/v) neutral bicelles (DMPC:DHPC with 3:1 molar ratio) and charged bicelles (Losonczi and Prestegard 1998) (DMPC:DHPC:CTAB and DMPC:DHPC:SDS, both with 15:5:1 molar ratios), or in no quadrupolar splitting of the water lock signal due to interaction with the aligning medium in 12 mg/mL filamentous bacteriophage Pf1 (Hansen et al. 1998).

Acknowledgments

Work was supported by NIH CESG Grant P50 GM64598 (J.L.M.). We extend special thanks to Marco Tonelli for optimization of Varian pulse programs.

References

Bauer, M. and Papenbrock, J. 2002. Identification and characterization of single-domain thiosulfate sulfurtransferases from *Arabidopsis thaliana*. *FEBS Lett.* **532:** 427–431.

Bauer, M., Dietrich, C., Nowak, K., Sierralta, W.D., and Papenbrock, J. 2004. Intracellular localization of *Arabidopsis* sulfurtransferases. *Plant Physiol.* **135:** 916–926.

Cornilescu, G., Delaglio, F., and Bax, A. 1999. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13:** 289–302.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. 1995. NMRPIPE—A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6:** 277–293.

Garrett, D.S., Powers, R., Gronenborn, A.M., and Clore, G.M. 1991. A common-sense approach to peak picking in two-dimensional, three-dimensional, and four-dimensional spectra using automatic computer-analysis of contour diagrams. *J. Magn. Reson.* **95:** 214–220.

Gibrat, J.F., Madej, T., and Bryant, S.H. 1996. Surprising similarities in structure comparison. *Curr. Opin. Struct. Biol.* **6:** 377–385.

Hansen, M.R., Mueller, L., and Pardi, A. 1998. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat. Struct. Biol.* **5:** 1065–1074.

Herrmann, T., Güntert, P., and Wüthrich, K. 2002. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **319:** 209–227.

Leon, S., Touraine, B., Ribot, C., Briat, J.F., and Lobreaux, S. 2003. Iron-sulphur cluster assembly in plants: Distinct NFU proteins in mitochondria and plastids from *Arabidopsis thaliana*. *Biochem. J.* **371:** 823–830.

Losonczi, J.A. and Prestegard, J.H. 1998. Improved dilute bicelle solutions for high-resolution NMR of biological macromolecules. *J. Biomol. NMR* **12:** 447–451.

Nandi, D.L. and Westley, J. 1998. Reduced thioredoxin as a sulfur-acceptor substrate for rhodanese. *Int. J. Biochem. Cell Biol.* **30:** 973–977.

Pagani, S., Bonomi, F., and Cerletti, P. 1984. Enzymic-synthesis of the iron-sulfur cluster of spinach ferredoxin. *Eur. J. Biochem.* **142:** 361–366.

Ray, W.K., Zeng, G., Potters, M.B., Mansuri, A.M., and Larson, T.J. 2000. Characterization of a 12-kilodalton rhodanese encoded by glpE of *Escherichia coli* and its interaction with thioredoxin. *J. Bacteriol.* **182:** 2277–2284.

Schwieters, C.D. and Clore, G.M. 2001. Internal coordinates for molecular dynamics and minimization in structure determination and refinement. *J. Magn. Reson.* **152:** 288–302.

Schwieters, C.D., Kuszewski, J.J., Tjandra, N., and Clore, G.M. 2003. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160:** 65–73.

Spallarossa, A., Donahue, J.L., Larson, T.J., Bolognesi, M., and Bordo, D. 2001. *Escherichia coli* GlpE is a prototype sulfurtransferase for the single-domain rhodanese homology superfamily. *Structure* **9:** 1117–1125.

Vennesland, B., Castric, P.A., Conn, E.E., Solomonson, L.P., Volini, M., and Westley, J. 1982. Cyanide metabolism. *Fed. Proc.* **41:** 2639–2648.

Vinarov, D.A., Lytle, B.L., Peterson, F.C., Tyler, E.M., Volkman, B.F., and Markley, J.L. 2004. Cell-free protein production and labeling protocol for NMR-based structural proteomics. *Nat. Methods* **1:** 149–153.

Westley, J. 1973. Rhodanese. *Adv. Enzymol. Relat. Areas Mol. Biol.* **39:** 327–368.

Wüthrich, K. 1986. *NMR of proteins and nucleic acids*. Wiley Interscience, New York, NY.