BIOLOGICAL CRYSTALLOGRAPHY

ISSN 1399-0047

A disulfide-bond cascade mechanism for arsenic(III) *S*-adenosylmethionine methyltransferase

Kavitha Marapakala,[a]‡ Charles Packianathan,[b]‡ A. Abdul Ajees,[c] Dharmendra S. Dheeman,[b] Banumathi Sankaran,[d] Palani Kandavelu[e] and Barry P. Rosen[b]*

[a]Department of Chemistry, Osmania University College for Women, Osmania University, Hyderabad 500 095, India, [b]Department of Cellular Biology and Pharmacology, Florida International University Herbert Wertheim College of Medicine, 11200 S. W. 8th Street, Miami, FL 33199, USA, [c]Department of Atomic and Molecular Physics, Manipal Institute of Technology, Manipal University, Manipal, Karnataka 576 104, India, [d]Physical Biosciences Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94702, USA, and [e]SER-CAT and the Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602, USA. *Correspondence e-mail: brosen@fiu.edu

Received 2 September 2014
Accepted 17 December 2014

‡ These authors contributed equally to the results and are equal co-first authors.

Keywords: arsenic(III) *S*-adenosylmethionine methyltransferase; AS3MT; phenylarsenite; roxarsone; disulfide-bond cascade.

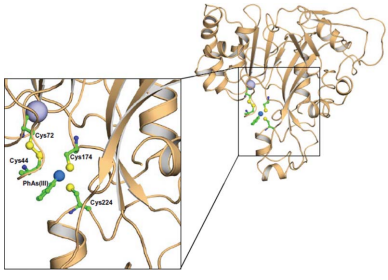
PDB references: CmArsM, with bound phenylarsenite, 4kw7; with bound roxarsone, 4rsr

Supporting information: this article has supporting information at journals.iucr.org/d

Methylation of the toxic metalloid arsenic is widespread in nature. Members of every kingdom have arsenic(III) *S*-adenosylmethionine (SAM) methyltransferase enzymes, which are termed ArsM in microbes and AS3MT in animals, including humans. Trivalent arsenic(III) is methylated up to three times to form methylarsenite [MAs(III)], dimethylarsenite [DMAs(III)] and the volatile trimethylarsine [TMAs(III)]. In microbes, arsenic methylation is a detoxification process. In humans, MAs(III) and DMAs(III) are more toxic and carcinogenic than either inorganic arsenate or arsenite. Here, new crystal structures are reported of ArsM from the thermophilic eukaryotic alga *Cyanidioschyzon* sp. 5508 (CmArsM) with the bound aromatic arsenicals phenylarsenite [PhAs(III)] at 1.80 Å resolution and reduced roxarsone [Rox(III)] at 2.25 Å resolution. These organoarsenicals are bound to two of four conserved cysteine residues: Cys174 and Cys224. The electron density extends the structure to include a newly identified conserved cysteine residue, Cys44, which is disulfide-bonded to the fourth conserved cysteine residue, Cys72. A second disulfide bond between Cys72 and Cys174 had been observed previously in a structure with bound SAM. The loop containing Cys44 and Cys72 shifts by nearly 6.5 Å in the arsenic(III)-bound structures compared with the SAM-bound structure, which suggests that this movement leads to formation of the Cys72–Cys174 disulfide bond. A model is proposed for the catalytic mechanism of arsenic(III) SAM methyltransferases in which a disulfide-bond cascade maintains the products in the trivalent state.

1. Introduction

Arsenic is one of the most ubiquitous environmental toxins and is a human carcinogen. It poses a serious threat to human health, and consequently ranks first on the 2013 Environmental Protection Agency's Comprehensive Environmental Response, Compensation and Liability Act List of Hazardous Substances (http://www.atsdr.cdc.gov/spl/). This toxic metalloid is introduced primarily from geochemical sources and is acted on biologically, creating an arsenic biogeocycle (Zhu *et al.*, 2014). Members of every kingdom, from bacteria to humans, methylate arsenite, producing the trivalent species methylarsenite [MAs(III)], dimethylarsenite [DMAs(III)] and the volatile trimethylarsine [TMAs(III)] (Thomas & Rosen, 2013; Ye *et al.*, 2012; Liu *et al.*, 2013). The enzymes that catalyze this reaction are arsenic(III) *S*-adenosylmethionine (SAM) methyltransferases (AsMTs; EC 2.1.1.137), which have been termed ArsM in microbes and AS3MT in animals. In microbes arsenic methylation is a detoxification process (Qin *et al.*, 2006, 2009), but in humans the production of MAs(III)



and DMAs(III) has been proposed to increase reactivity and toxicity (Thomas *et al.*, 2007).

Whether AsMTs detoxify arsenic or transform it into more toxic products depends on their enzymatic mechanism. Challenger proposed that the mechanism is an alternating series of oxidative methylations and reductions, with pentavalent species as products (Challenger, 1947). This hypothesis is supported by the fact that humans and other mammals excrete dimethylarsenate [DMAs(V)] and, to a lesser extent, methylarsenate [MAs(V)] in urine. An alternate proposal is that the products are all trivalent methylarsenicals (Hayakawa *et al.*, 2005). If the primary intracellular products of methylation are the pentavalent species, then arsenic would have limited carcinogenic potential. On the other hand, if the trivalent species are the major methylated intracellular products, then methylation would increase the carcinogenicity of arsenic. Whether the oxidized species found in urine of mammals or the growth medium of microbes are the products of the AsMTs or are the result of nonenzymatic oxidation of the unstable trivalent species is controversial (Cullen, 2014), but, with careful handling, the trivalent forms have been detected in urine (Le *et al.*, 2000; Drobná *et al.*, 2012). Thus, resolution of this uncertainty is of considerable consequence in our understanding of the health effects of arsenic.

In contrast to the uncertainty in the human pathway of methylation, microbial arsenic methylation is an established detoxification process that is proposed to have an impact on the global arsenic cycle (Zhu *et al.*, 2014). Genes for ArsM orthologues of the human AS3MT are widespread in the genomes of bacteria, archaea, fungi and lower plants (Ye *et al.*, 2012). ArsM from the thermoacidophilic eukaryotic red alga *Cyanidioschyzon* sp. 5508 from Yellowstone National Park (CmArsM) catalyzes arsenic methylation and volatilization, leading to resistance (Qin *et al.*, 2009). CmArsM is a 400-residue thermostable enzyme (44 980 Da; accession No. ACN39191) that methylates arsenic(III) to a final product of volatile TMAs(III). CmArsM has been crystallized (Marapakala *et al.*, 2010) and its structure solved to 1.6 Å resolution with and without arsenic(III) or SAM (Ajees *et al.*, 2012). To understand on the one hand how microorganisms remodel the environment in arsenic-rich regions and on the other hand how arsenic methylation is involved in carcinogenesis, our goal is to elucidate the catalytic cycle and determine the X-ray crystal structure of CmArsM with various ligands.

Three conserved cysteine residues have previously been recognized in ArsM orthologs, which are Cys72, Cys174 and Cys224 in CmArsM (Marapakala *et al.*, 2012; Supplementary Fig. S1). Here, we report crystal structures of CmArsM with bound phenylarsenite [PhAs(III)] and roxarsone [Rox(III)] with final resolutions of 1.80 and 2.25 Å, respectively. Both aromatic arsenicals, which are the trivalent forms of poultry antimicrobial growth enhancers, were found to be substrates of CmArsM and to bind with high affinity to Cys174 and Cys224. We now identify a fourth conserved cysteine, Cys44, in CmArsM which, like Cys72, is required for the first methylation step but not the second. In previous structures, Cys44 was not resolved (Ajees *et al.*, 2012). In the new structures, Cys44 and Cys72 are found to form a disulfide bond. In the previously reported SAM-bound structure Cys72 and Cys174 were observed to be disulfide-bonded to each other (Ajees *et al.*, 2012). We propose a new model for the catalytic cycle of arsenic(III) SAM methyltransferases that utilizes a disulfide-bond cascade to successively reduce pentavalent arsenical intermediates. In this model the substrates and products are trivalent, but there are pentavalent enzyme-bound intermediates that are reduced by cysteine residues, creating a disulfide-bond cascade mechanism of alternating oxidations and reductions of the bound arsenic.

2. Materials and methods

2.1. Preparation and purification of CmArsM and derivatives

CmArsM 7B lacking 31 residues from the N-terminus and 28 residues from the C-terminus with a C-terminal histidine tag (termed simply CmArsM) was expressed and purified as described previously (Marapakala *et al.*, 2010). The purified enzyme was stored at −80°C until use. Protein concentrations were estimated from the absorbance at 280 nm (Gill & von Hippel, 1989).

Mutations were introduced into the *CmarsM* 7B gene by site-directed mutagenesis using a QuikChange mutagenesis kit (Stratagene, La Jolla, California, USA). The forward and reverse primer oligonucleotides used for mutagenesis were 5′-CTAAAGACGAGCGCTGCCAAGCTTGCCGCGGC and 5′-GCCGCGGCAAGCTTGGCAGCGCTCGTCTTTAG, respectively, for Cys44 mutagenesis and 5′-GTCCTGGAA-AAGTTCTACGGTGCCGGGTCTACGC and 5′-GCGTAG-ACCCGGCACCGTAGAACTTTTCCAGGAC, respectively, for Cys72 mutagenesis. Using these primers, the codons for the conserved residues Cys44 and Cys72 were changed to alanine codons, generating C44A and C72A derivatives. The double C44A/C72A mutant was generated by mutating codon 44 in the C72A derivative to an alanine codon using the same forward and reverse primers. Each mutation was confirmed by sequencing the entire gene. When the gene for each derivative was expressed in *Escherichia coli*, each protein was produced in a soluble form in amounts comparable to the parental CmArsM. As described below, both enzymes retained the ability to catalyze the methylation of MAs(III). These results indicate that the substitutions did not have a major effect on the overall folding of the proteins.

2.2. Crystallization of ArsM in the presence of aromatic arsenical ligands

CmArsM was co-crystallized with the aromatic arsenicals PhAs(III) or Rox(III) by the hanging-drop vapor-diffusion method as described previously (Marapakala *et al.*, 2010; Ajees *et al.*, 2012). CmArsM (15 mg ml^{-1}) was incubated with 1 m*M* PhAs(III) or Rox(III) for 30 min at room temperature and 2.5 µl portions of protein solution were mixed with equal volumes of reservoir solution consisting of 18% PEG 3350, 0.2 *M* calcium acetate, 0.1 *M* Tris–HCl pH 7.0. The hanging drops were equilibrated with 0.5 ml well solution and crystals

Table 1
Data-collection and refinement statistics for CmArsM with bound PhAs(III) or Rox(III).

Values in parentheses are for the highest resolution bin.

	PhAs(III)-bound	Rox(III)-bound
Data collection		
X-ray source	ALS beamline 5.0.2	APS beamline 22-ID
Wavelength (Å)	1.000	1.000
Detector	ADSC Quantum 315r CCD	MAR 300 CCD
Resolution (Å)	50.0–1.80 (1.86–1.80)	50.0–2.25 (2.29–2.25)
Space group	*C*2	*C*2
Unit-cell parameters (Å, °)	a = 85.10, b = 47.41, c = 100.34, β = 113.4	a = 85.16, b = 46.41, c = 100.26, β = 114.1
Observed reflections	135583 (11104)	120158 (4420)
Unique reflections	33659 (3266)	16680 (762)
Multiplicity	4.0 (3.4)	7.2 (5.8)
$\langle I\rangle/\langle\sigma(I)\rangle$	22.4 (1.7)	39.9 (6.2)
Completeness (%)	98.2 (96.8)	96.8 (88.0)
$R_{\rm merge}$† (%)	5.9 (52.5)	4.4 (24.8)
$R_{\rm r.i.m.}$‡ (%)	6.2 (17.6)	4.8 (27.0)
$R_{\rm p.i.m.}$§ (%)	3.6 (10.6)	1.8 (10.6)
B factor, Wilson plot (Å^2)	36.9	34.8
Refinement¶		
$R_{\rm work}/R_{\rm free}$†† (%)	20.5/25.4	17.3/22.4
No. of protein atoms	2569	2579
No. of water molecules	215	124
No. of ligand atoms	9	23
Mean B value (Å^2)	36.3	32.4
R.m.s. deviation from ideal		
Bond lengths (Å)	0.019	0.019
Bond angles (°)	1.96	1.97
Ramachandran plot‡‡ (%)		
Most favored	91.3	91.7
Additionally allowed	8.7	7.6
Generously allowed	0	0.7
PDB code	4kw7	4rsr

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity and $\langle I(hkl)\rangle$ is the average intensity over symmetry-equivalent measurements. ‡ $R_{\rm r.i.m.} = \sum_{hkl}\{N(hkl)/[N(hkl)-1]\}^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$. § $R_{\rm p.i.m.} = \sum_{hkl}\{1/[N(hkl)-1]\}^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$. ¶ Refinement using *REFMAC* (Murshudov *et al.*, 2011). †† $R_{\rm work} = \sum_{hkl}||F_{\rm obs}| - |F_{\rm calc}||/\sum_{hkl}|F_{\rm obs}|$, where $R_{\rm free}$ is calculated for a random chosen 5% of reflections which were not used for structure refinement and $R_{\rm work}$ is calculated for the remaining reflections. ‡‡ Ramachandran plot calculated by *PROCHECK* (Laskowski *et al.*, 1993).

were obtained at 293 K using Linbro 24-well plates (catalog No. HR3-110; Hampton Research, Aliso Viejo, California, USA). Crystals were transferred to a cryoprotectant solution (25% PEG 3350, 0.2 *M* calcium acetate, 0.1 *M* Tris–HCl pH 7.0) and flash-cooled in liquid nitrogen for data collection.

2.3. X-ray data collection and processing

Data sets were collected from crystals of CmArsM with bound PhAs(III) using an ADSC Quantum 315r (3 × 3 CCD array) detector at 100 K under a liquid-nitrogen stream on beamline 5.02 of the Advanced Light Source (ALS), Lawrence Berkeley National Laboratory. Crystallographic data for co-crystals of CmArsM with bound Rox(III) were collected at the Southeast Regional Collaborative Access Team (SER-CAT) facility at the Advanced Photon Source (APS), Argonne National Laboratory. 360 image frames with a 1° rotation angle about φ were collected using a MAR 300 CCD detector under standard cryogenic conditions (100 K) on synchrotron beamline 22-ID. Both data sets were indexed, integrated and scaled using the *HKL*-2000 suite (Otwinowski & Minor, 1997). Evaluation of crystal-packing parameters indicated that the lattice can accommodate one molecule in the asymmetric unit. The Matthews coefficients for CmArsM with bound PhAs(III) or Rox(III) are 2.42 or 2.26 Å^3 Da^{-1}, respectively, with solvent contents of 42.0 and 45.6%, respectively (Matthews, 1968).

Table 2
Ligand interactions with amino-acid residues.

Ligand atom	Amino-acid atom	Distance (Å)
PhAs(III)		
As	Cys174 O	3.1†
As	Cys174 C$^\beta$	3.1†
C6	Cys174 S	3.1†
C1	Cys174 S	3.3†
C2	Gly222 O	3.4†
C5	Glu223 O	3.7‡
C6	Glu223 O	3.8‡
Rox(III)		
As	Cys174 C$^\beta$	3.0†
C4	Glu223 OE2	2.9†
O3	Glu223 OE2	3.0†
As	Cys174 O	3.4†
O1	Cys44 O	3.5‡
C3	Glu223 O	3.7‡
C2	Glu223 O	3.7‡
C5	Gly222 O	3.7‡
C6	Gly222 O	3.7‡

† Moderate hydrogen bond. ‡ Nonbonded interaction between atoms of PhAs(III) or Rox(III) with amino-acid atoms.

2.4. Structure solution and refinement

The crystals of CmArsM with bound PhAs(III) or Rox(III) belonged to space group *C*2. The unit-cell parameters for the PhAs(III)-bound structure are a = 85.10, b = 47.41, c = 100.34 Å, β = 113.4° and those for the Rox(III)-bound structure are a = 85.16, b = 46.41, c = 100.26 Å, β = 114.1°. The structures were determined by molecular replacement with *Phaser* (McCoy, 2007). The atomic coordinates of the ligand-free crystal structure of CmArsM (PDB entry 4fs8; Ajees *et al.*, 2012) were used as the starting model. Structure refinement of each data set was performed using *REFMAC*5 (Murshudov *et al.*, 2011) as implemented in the *CCP*4 suite (Winn *et al.*, 2011). About 5% of the reflections were used for the test set. The refined model required manual adjustment to improve the fit to the experimental electron density using *Coot* (Emsley & Cowtan, 2004). At this stage restrained refinement was carried out and the final R factors converged to 20.5% ($R_{\rm free}$ = 25.4%) for the PhAs(III)-bound structure and 17.3% ($R_{\rm free}$ = 22.4) for the RoxAs(III)-bound structure.

A Ramachandran plot for Rox(III)-bound CmArsM calculated using *PROCHECK* (Laskowski *et al.*, 1993) indicated that 91.7% of the residues are in the most favoured region, 7.6% of the residues are in the additionally allowed region and 0.7% of the residues are in the generously allowed region. Final data-collection and refinement statistics and Protein Data Bank accession codes are given in Table 1. Figures were prepared using *PyMOL* (Peters *et al.*, 2006). *LigPlot*+ was used to illustrate hydrogen bonds and

nonbonded interactions of CmArsM with Rox(III) and PhAs(III) (Laskowski & Swindells, 2011) and these are listed in Table 2.

2.5. Assay of CmArsM activity

The methylation activity of purified CmArsM was assayed at 60°C in 50 m*M* MOPS pH 7.5, 0.5 *M* NaCl containing 5 m*M* GSH and 1 m*M* SAM as described previously (Marapakala *et al.*, 2012). Unless otherwise noted, the reactions were terminated by adding 10%(*v*/*v*) H_2O_2 (final concentration) to oxidize all arsenic species. Denatured protein was removed by centrifugation using a 3 kDa cutoff Amicon ultrafilter. To determine the nature of the CmArsM-bound arsenicals, purified CmArsM (5 µ*M*) was incubated at 60°C with 10 µ*M* arsenic(III) as described above for 10 min and terminated by rapid passage through a Bio-Gel P-6 column pre-equilibrated with the same buffer without H_2O_2 oxidation. Portions (25 µl) were immediately diluted with sufficient 6 *M* guanidine–HCl to bring the final concentration to 4 *M* ;to release the bound arsenicals. The products of arsenic(III) or MAs(III) methylation were separated by high-performance liquid chromatography (HPLC) using a Hamilton PRP-X100 C18 reverse-phase column (Hamilton Co., Reno, Nevada, USA) and quantified using a PerkinElmer ELAN DRC-e inductively coupled plasma mass spectrometer (ICP-MS) as described previously (Marapakala *et al.*, 2012). The products of PhAs(III) methylation were separated by HPLC using a Inertsil C4 reverse-phase column (GL Sciences, Rolling Hills Estates, California, USA) with a mobile phase consisting of 80% water at pH 1.5, 15% methanol and 5% acetonitrile at a flow rate of 0.3 ml min^{-1} and with an oven temperature at 40°C (Kinoshita *et al.*, 2008).

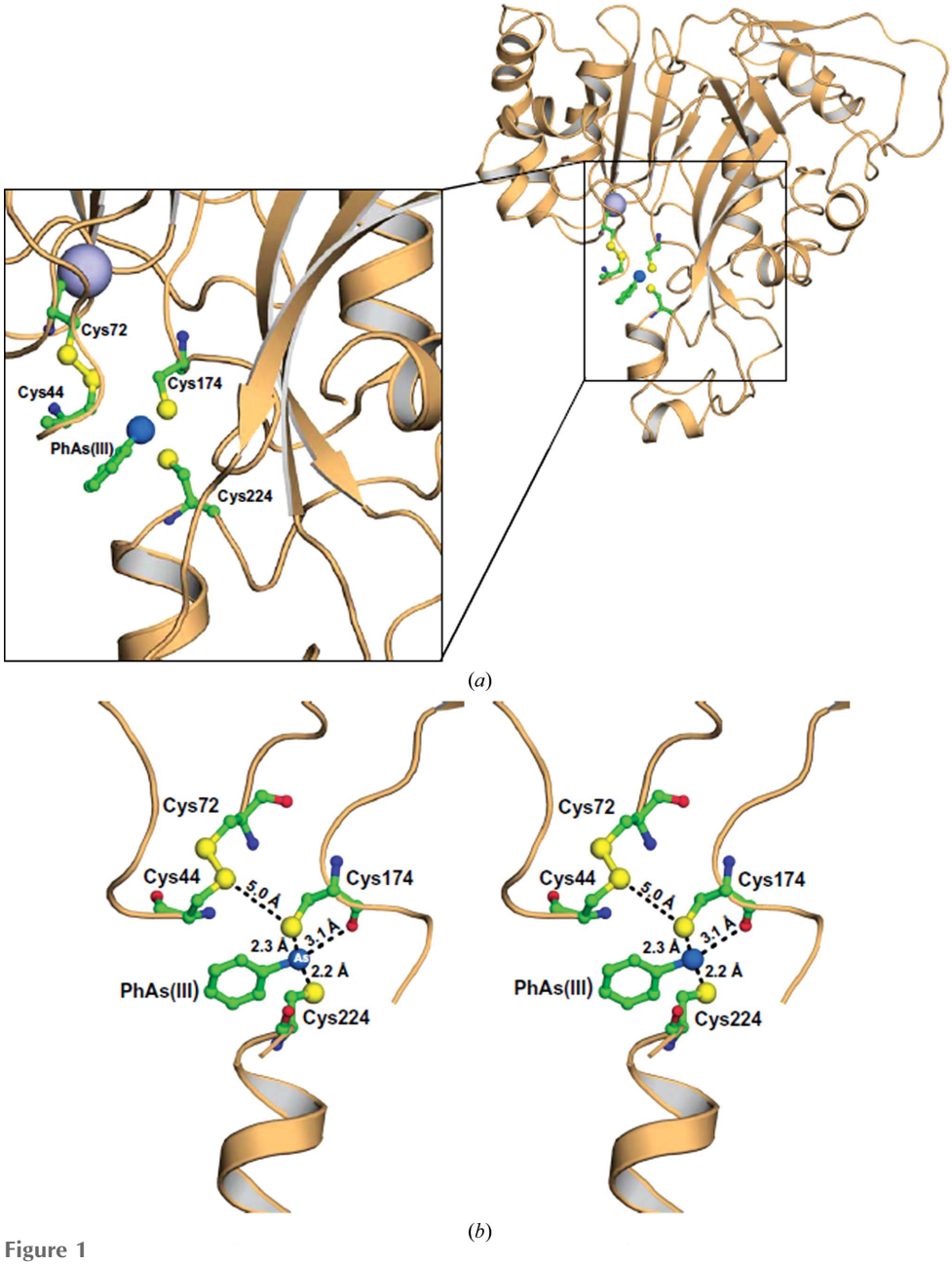


Figure 1
(*a*) Structure of CmArsM with bound PhAs(III). Cartoon diagram (colored in light orange) representation of PhAs(III)-bound CmArsM (PDB entry 4kw7). The conserved cysteine residues are shown in ball-and-stick representation and are colored green (carbon), blue (nitrogen) or yellow (sulfur). The dark blue sphere is the As atom and the light blue sphere is a Ca^{2+} ion in the SAM binding site. PhAs(III) is bound between the conserved residues Cys174 and Cys224. (*b*) A disulfide bond between Cys44 and Cys72. The four conserved cysteine residues are shown in ball-and-stick representation. The length of the disulfide bond is approximately 2.1 Å in the PhAs(III)-bound structure.

3. Results

3.1. Structure of CmArsM with bound PhAs(III)

Crystals of CmArsM with bound PhAs(III) diffracted to 1.80 Å resolution and belonged to space group *C*2, with one molecule in the asymmetric unit and unit-cell parameters *a* = 85.10, *b* = 47.41, *c* = 100.34 Å, *β* = 113.4°. A model of the monomeric structure contains 328 residues from Cys44 to Ser371 (Fig. 1*a*). The structure includes the N-terminal domain, the SAM-binding domain, the arsenic(III)-binding domain and a C-terminal

domain of unknown function (Ajees *et al.*, 2012). Previously reported structures started at residue Val50 in ligand-free (PDB entry 4fs8) and arsenic(III)-bound (PDB entry 4fsd) CmArsM and Ala49 in SAM-bound CmArsM (PDB entry 4fr0). In the PhAs(III)-bound structure, N-terminal electron density is visible starting at Cys44. Residues 44–50 and 70–74 in the final model are well defined in the $2F_o - F_c$ electron-density map contoured at 1.0σ (Supplementary Fig. S3*a*) and residues 44–48 form a loop that was not observed in the previous structures. The N-terminal domain (residues 50–80) starts with two small helices and then a long mobile loop that we propose moves from the SAM binding sites towards the arsenic(III) binding site during the catalytic cycle as described below.

The electron densities of the bound PhAs(III) and the four conserved cysteine residues are shown in Fig. 2(*a*). The distance between the As atom and the thiolates of Cys174 and Cys224 are 2.3 and 2.2 Å, respectively (Fig. 1*b*). The distance between the As atom and the C^{β} atoms of Cys224 is 3.3 Å. The distance between the S atoms of Cys44 and Cys174 is 5.0 Å. The thiolates of Cys44 and Cys72 are 5.5 and 7.5 Å from the As atom, respectively.

LigPlot+ was used to identify and display the hydrogen-bond and hydrophobic interactions of the protein with the PhAs(III) ligand (Fig. 3*a*). The ligand is connected to the S atoms of Cys174 and Cys224 and is surrounded by Cys44, Gly222 and Glu223. The Cys44 and Gly222 residues are involved in nonbonded interactions. The hydrogen bonds and nonbonded interactions of bound PhAs(III) with amino-acid residues are shown in Table 2. The CmArsM-bound PhAs(III) structure adopts a pyramidal site in which the As atom of PhAs(III) is coordinated to the S atoms of Cys174 and Cys224. Each of the three liganding atom is at an average distance of 3.3 Å from the others. Superposition of the PhAs(III)-bound structure with ligand-free (PDB entry 4fs8, 319 C^{α} residues), arsenic(III)-bound (PDB entry 4fs8, 318 C^{α} residues) and SAM-bound (PDB entry 4fr0, 317 C^{α} residues) CmArsM structures yields root-mean-square deviation (r.m.s.d.s) of 0.20, 0.41 and 1.27 Å, respectively.

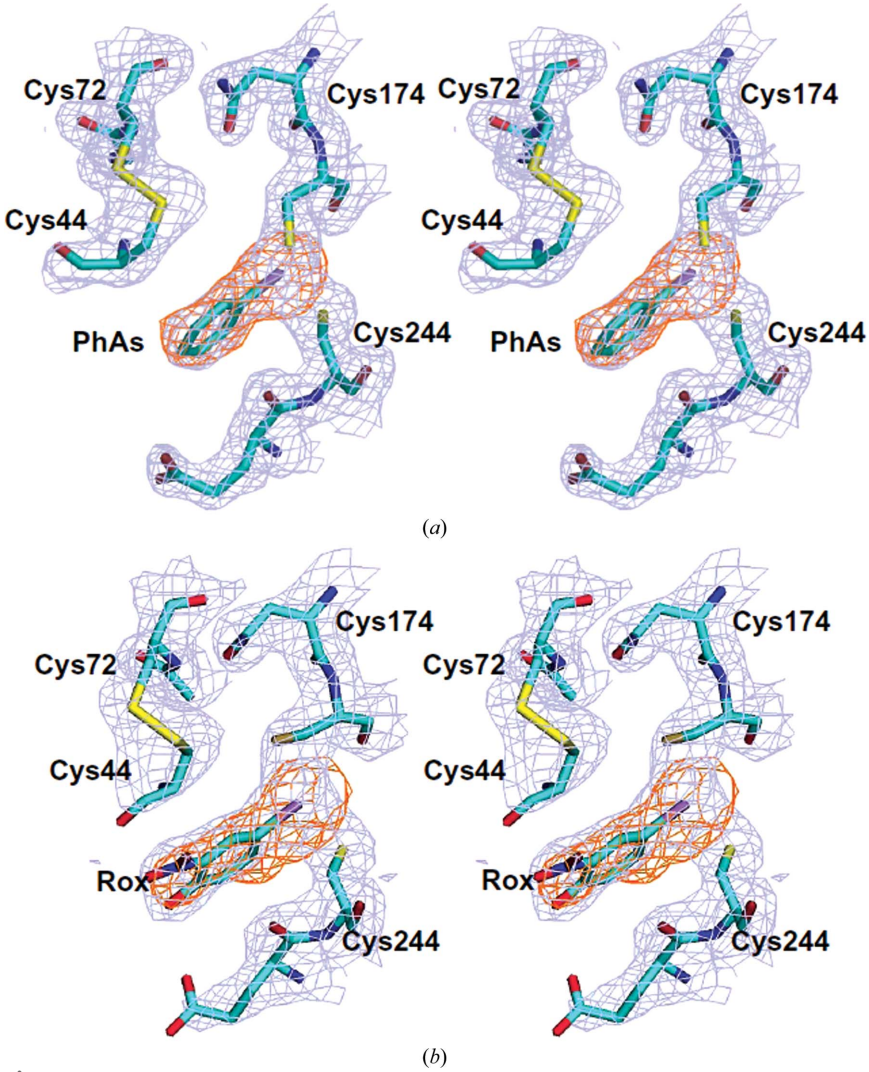


Figure 2
Stereoviews of the electron-density maps of (*a*) PhAs(III) and (*b*) Rox(III) bound to CmArsM. The stereoviews and the $2F_o - F_c$ electron-density maps of the conserved cysteine residues are shown with the bound aromatic arsenicals contoured at 1.0σ (light blue). The cysteine residues, PhAs(III) and Rox(III) are shown as stick models. The As atom is colored purple. The PhAs(III) and Rox(III) electron densities are outlined with the $F_o - F_c$ unbiased OMIT map contoured at 2.5σ (orange).

3.2. Structure of CmArsM with bound Rox(III)

The X-ray crystal structure of CmArsM with bound Rox(III) was determined at 2.25 Å resolution. The structure belonged to space group *C*2, with unit-cell parameters $a = 85.16$, $b = 46.41$, $c = 100.26$ Å, $\beta = 114.1^{\circ}$. CmArsM with bound Rox(III) consists of a mixture of α-helices, β-sheets and 3_{10}-helices similar to the arsenic(III)-bound structure (Ajees *et al.*, 2012). In this structure Rox(III) binds between the conserved cysteines Cys174 and Cys224, similar to PhAs(III)-bound CmArsM (Supplementary Fig. S2(*a*). The final model residues are well defined in the $2F_o - F_c$ electron-density map contoured at 1.0σ (Supplementary Fig. S3*b*). The electron densities of the bound Rox(III) and the four conserved cysteine residues are shown in Fig. 2(*b*). The distances between the As atom and the thiolates of Cys174 and Cys224 are 2.3 and 2.4 Å, respectively (Supplementary Fig. S2*b*). The other two cysteine thiolates of Cys44 and Cys72 are 5.1 and 7.1 Å, respectively, from the As atom. The distance between the S atoms of Cys44 and Cys174 is 5.0 Å. The O and C^{β} atoms of Cys174 are approximately 3.4 and 3.0 Å, respectively,

from the As atom. The distance between the S atoms of Cys44 and Cys174 and the As atom is 5.0 Å. The Rox(III)-bound structure adopts an overall conformation similar to that of the arsenic(III)-bound structure. The Rox(III)-bound CmArsM structure superimposes with the ligand-free, arsenic(III)-bound and SAM-bound structures with r.m.s.d.s of 0.34, 0.33 and 1.32 Å, respectively. This suggests that both the PhAs(III)-bound and the Rox(III)-bound CmArsM structures are similar to the ligand-free and arsenic(III)-bound structures and that the SAM-bound structure adopts a different conformation. Similar to the PhAs(III)-bound structure, Rox(III) is sandwiched between Cys44, Glu223 and Gly222 (Fig. 3*b*). The hydrogen-bond and nonbonded interactions of Rox(III) with nearby residues may serve to stabilize the Rox(III)-bound form (Table 2). The Rox(III)-bound CmArsM structure reveals a pyramidal site similar to that in the PhAs(III)-bound structure, with each of the three liganding atoms at an average distance of 3.4 Å from the others. These organoarsenical derivatives are similar to the product of the first methylation step, MAs(III), with an aromatic group instead of a methyl group. When arsenic(III) accepts a methyl group from SAM, the methyl group must undergo an inversion of configuration to allow a second methylation. The methyl group would then be facing the bulk solvent. In both the PhAs(III) and Rox(III) structures the aromatic *R* groups are exposed to bulk solvent, indicating that they are more similar to product-bound forms than to substrate-bound forms.

The PhAs(III)-bound and Rox(III)-bound structures can be superposed with an r.m.s.d. of 0.36 Å over 324 aligned C^{α} residues (Supplementary Fig. S5*a*). Superposition of the arsenic(III)-bound structure with the two structures with bound aromatic arsenicals show r.m.s.d.s of 0.40 Å over 318 aligned C^{α} residues and 0.32 Å over 322 aligned C^{α} residues, respectively (Supplementary Fig. S5*b*). The C atoms of the aromatic rings of PhAs(III) and Rox(III) interact with residues Gly222 and Glu223 neighboring the arsenic(III)-binding site. These interactions may reduce the access of SAM to its binding site and may explain our inability to obtain crystals with both arsenical and SAM bound.

3.3. The N-terminal extension and Cys44–Cys72 disulfide bond

Cysteine residues are frequently evolutionarily conserved and play essential roles in the structure and function of proteins such as the formation of disulfide bonds that stabilize the structure (Fomenko *et al.*, 2007; Beeby *et al.*, 2005). In both the PhAs(III)-bound and Rox(III)-bound structures a disulfide bond is observed between Cys44 and Cys72 with an average distance of 2.1 Å (Fig. 2). The typical S—S bond

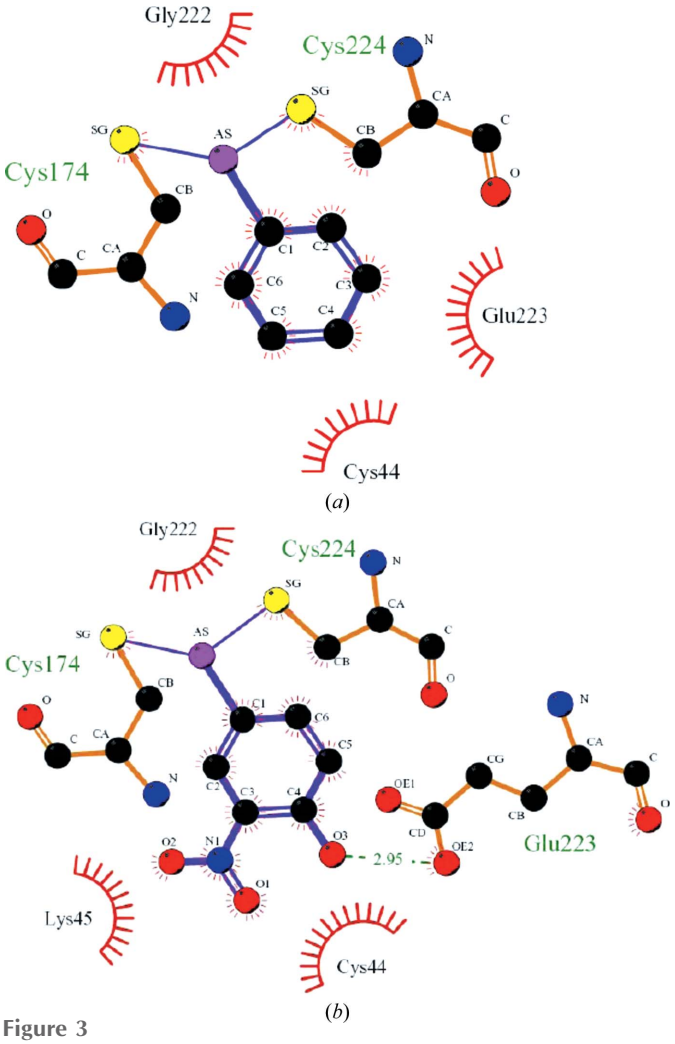


Figure 3
Residues interacting with PhAs(III) or Rox(III). *LigPlot*+ representations illustrating hydrophobic and hydrogen-bond interactions with (*a*) PhAs(III) (PDB entry 4kw7) and (*b*) Rox(III) (PDB entry 4rsr). Ligand bonds are shown in purple and the hydrogen bond is shown in olive green. Nonligand residues involved in hydrophobic contacts are shown as arcs colored brick red. All other atoms are shown as spheres and are colored black (carbon), blue (nitrogen), red (oxygen), yellow (sulfur) and magenta (arsenic).

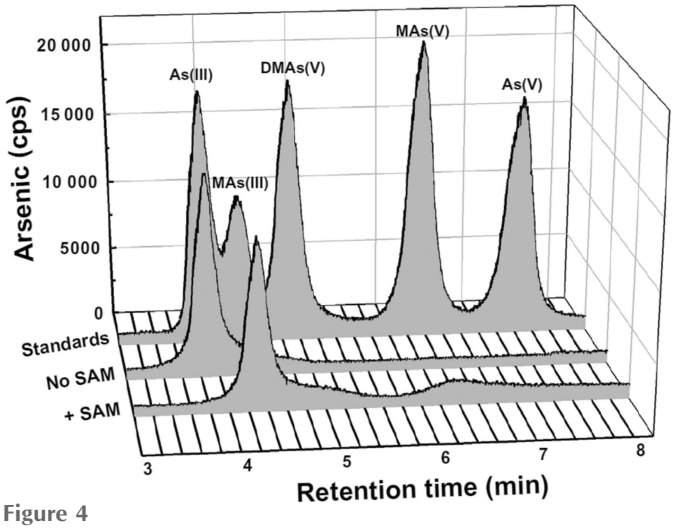


Figure 4
The MAs(III)–enzyme intermediate. Purified CmArsM (5 µ*M*) was incubated at 60°C with 10 µ*M* arsenic(III) containing 5 m*M* GSH with or without 1 m*M* SAM. After 10 min, samples were separated using a Bio-Gel P-6 spin column. Portions (25 µl) were diluted with 6 *M* Gu–HCl to a final concentration of 4 *M* to denature the protein and release the enzyme-bound arsenic. After speciation and quantification by HPLC–ICP-MS, 0.86 mol MAs(III) was bound per mole of CmArsM.

distance in a protein structure is about 2.05 Å, indicating that the two residues are linked by a disulfide bond. Disulfide-bond formation may restrict the movement of the N-terminal loop observed when SAM is bound, which may also contribute to the inability to obtain crystals with both arsenical and SAM bound (Ajees *et al.*, 2012). Why was Cys44 not observed in the arsenic(III)-bound structure reported previously (Ajees *et al.*, 2012)? An additional density was observed near Cys72 in the arsenic(III)-bound structure that was originally interpreted to be a glycerol molecule (Ajees *et al.*, 2012), but is more likely to be a poorly resolved Cys44 in a disulfide bond with Cys72. Why is the region more easily resolved in the PhAs(III)-bound and Rox(III)-bound structures compared with the arsenic(III)-bound structure? During the first round of the refinement process of PhAs(III)-bound and Rox(III)-bound CmArsM, the electron-density map ($2F_o - F_c$) contoured at 1.0σ cutoff traces the residues in the N-terminal region (residues 49–44). In contrast, in the arsenic(III)-bound structure (Ajees *et al.*, 2012) the density traces of the residues are visible below a 0.5σ cutoff after multiple rounds of refinement. This may reflect a more ordered structure with the aromatic trivalent arsenicals, which bind to the arsenic binding sites with higher affinity than inorganic arsenic(III).

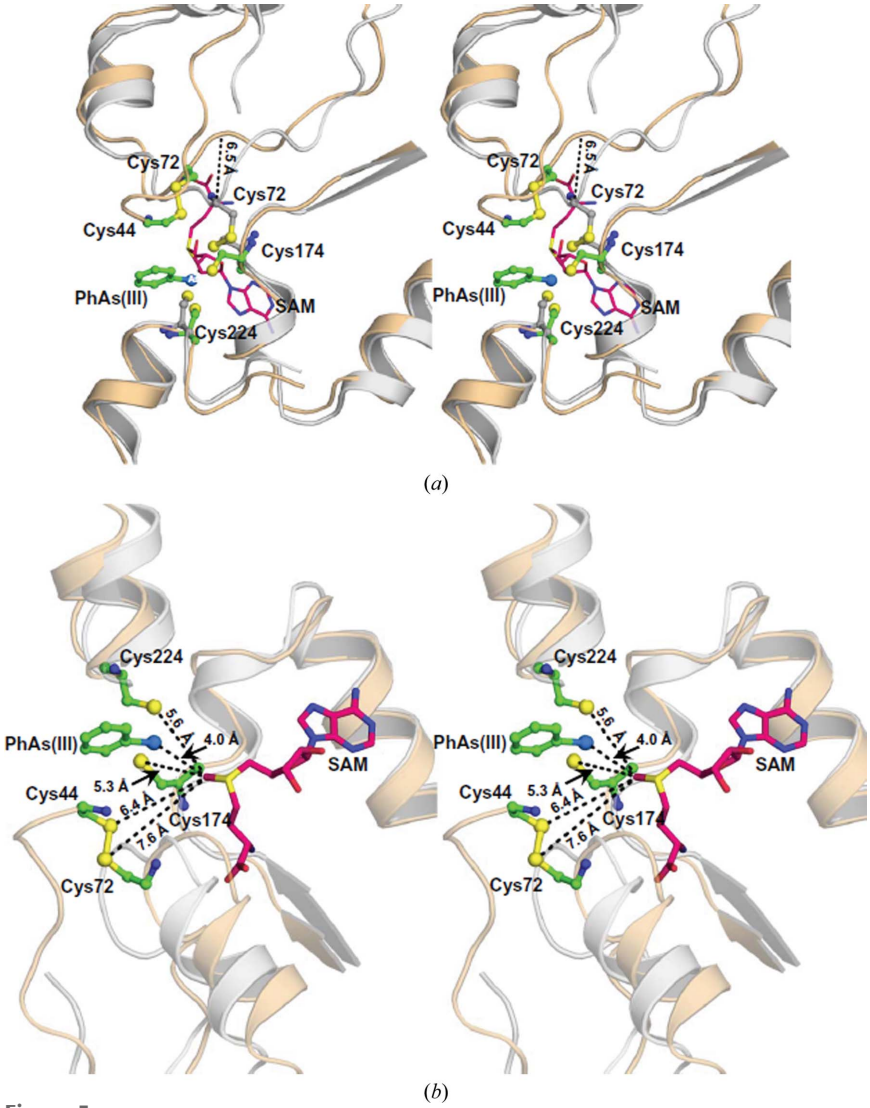


Figure 5
(*a*) Superimposition of PhAs(III)-bound CmArsM (light orange) with the SAM-bound structure (light grey) gives an r.m.s.d. of 1.27 Å. We propose that a 6.5 Å shift in the N-terminal loop from its position in the SAM-bound structure to its position in the aromatic arsenical-bound structures leads to disulfide-bond formation between Cys72 and Cys174. PhAs(III) and cysteine residues are shown in ball-and-stick representation with coloring of atoms as in Fig. 1. SAM is shown in ball-and-stick representation and colored purple (carbon), blue (nitrogen) and red (oxygen). (*b*) Modeling the complex of CmArsM with SAM and PhAs(III). The ternary complex of CmArsM with bound SAM and PhAs(III) was modeled by superposition of their individual structures. Coloring is as in Fig. 5(*a*). The distances from the *S*-methyl group of SAM to the S atoms of conserved cysteine residues and to the As atom of PhAs(III) are indicated.

We previously reported that after 10 min of reaction at 60°C CmArsM methylated arsenic(III) primarily to soluble MAs(III), but that most of the arsenic remained bound to the enzyme at early times (Marapakala *et al.*, 2012). At later times nearly all of the arsenic was transformed into soluble DMAs(V). We considered the possibility that MAs(III), the product of the first methylation step, remains enzyme-bound and is released only after the second round of methylation. To examine the nature of the enzyme-bound arsenic, arsenic(III) methylation was carried out for 10 min at an enzyme:substrate ratio of 1:2 to allow the isolation of sufficient enzyme–product complex. Approximately 43% of the added arsenic was bound to the enzyme, which represents 0.86 mol bound arsenic per mole of enzyme. Unbound arsenic was removed, and the protein with bound arsenicals was denatured with guanidine–HCl to release the protein-bound arsenicals (Fig. 4). Denatured protein was removed by filtration, and the filtrate was speciated by HPLC-ICP-MS. All of the bound arsenic was MAs(III), consistent with our hypothesis that the product of the first round of methylation is enzyme-bound until methylated a second time.

3.4. Movement of the N-terminal domain

Superposition of the SAM-bound structure with the PhAs(III)-bound and Rox(III)-bound structures gave r.m.s.d.s of 1.27 and 1.32 Å, respectively (Supplementary Fig. S5*c*). Comparison of these structures suggests that there is a 6.5 Å movement of the N-terminal residues 50–80 of the N-terminal domain of SAM-bound CmArsM toward the PhAs(III)-binding or Rox(III)-binding domain (Fig. 5*a*, Supplementary Fig. S4*a*). This observed reorientation of the N-terminal loop may reflect the

conformational changes that occur in different steps of the catalytic cycle. The distance between the sulfur thiolate of Cys72 when PhAs(III) is bound compared with Cys72 when SAM is bound is approximately 5.8 Å. The S atoms of Cys44, Cys72, Cys174 and Cys224 in the PhAs(III)-bound structure are 6.4, 7.6, 5.3 and 5.6 Å, respectively, from the *S*-methyl group of SAM. Similarly, the *S*-methyl group of the SAM is 6.1, 7.4, 5.6 and 5.8 Å from the S atoms of Cys44, Cys72, Cys174 and Cys224, respectively, of the Rox(III)-bound structure. The *S*-methyl group is poised for transfer to the As atoms of PhAs(III) or Rox(III), which are 4.0 and 4.2 Å, respectively, from the C atom of the *S*-methyl group (Fig. 5, Supplementary Fig. S4*b*).

3.5. The role of Cys44 in the methylation activity of ArsM and the C72A/C44A double mutant

We previously examined the function of the three conserved residues Cys72, Cys174 and Cys224 (Marapakala *et al.*, 2012). Mutagenesis of any of the three residues eliminated the first round of methylation of arsenic(III) to MAs(III), but a C72A derivative was able to carry out the second step of methylation of MAs(III) to the dimethyl species. Cys44 is here reported as a fourth conserved cysteine, forming a disulfide bond with Cys72 when the arsenic-binding site is filled. A C44A derivative was constructed, expressed and purified. The effect of this cysteine-to-alanine substitution on the methylation of arsenic(III) and MAs(III) was examined (Fig. 6*a*). The C44A protein was unable to methylate arsenic(III), which is similar to the effect of mutagenesis of the other three conserved cysteine residues. Like the C72A derivative, the C44A protein retained the ability to methylate MAs(III). A double C44A/C72A mutant gave similar results, methylating MAs(III) but not arsenic(III) (Fig. 6*b*). These results suggest that Cys44 and Cys72 play a role in catalysis that is different from those of the arsenic-binding residues Cys174 and Cys224.

3.6. CmArsM methylates PhAs(III)

CmArsM has been shown to methylate arsenic(III), MAs(III) and Sb(III) (Marapakala *et al.*, 2012) and is here shown to bind aromatic arsenicals. There are no reports of the methylation of aromatic arsenicals by AsMTs, but a methylated derivative of PhAs(III) has been observed in the urine of patients accidentally exposed to phenyl arsenicals (Kinoshita *et al.*, 2008). CmArsM was incubated with PhAs(III) under catalytic conditions, and the formation of a new species was observed (Supplementary Fig. S6). PhAs(III), PhAs(V) and the putative methylated species PhMAs(V) were speciated by HPLC using a C4 column (Kinoshita *et al.*, 2008). The new species eluted at a position corresponding to the previously reported position of PhMAs(V). PhMAs(V) is likely to be the result of nonenzymatic oxidation of the trivalent product PhMAs(III). In the absence of enzyme, PhAs(III) oxidizes to PhAs(V). This is the equivalent of the nonenzymatic oxidation of DMAs(III) to DMAs(V) in urine (Le *et al.*, 2000). These results are consistent with CmArsM methylating aromatic arsenicals.

4. Discussion

Arsenic is considered the most ubiquitous environmental toxin and carcinogen, and hundreds of millions of people worldwide are exposed to arsenic in their food and drinking-water supplies. In humans, exposure is a major contributor to arsenic-related diseases (Abernathy *et al.*, 2003; Tchounwou *et al.*, 2003), including bladder (Chen *et al.*, 2003), lung (Putila & Guo, 2011) and skin cancers (Rossman *et al.*, 2001). A major question is whether the risk is from exposure to inorganic arsenicals or to the products of methylation by AsMTs. The answer depends in a large part on deducing the enzymatic mechanism of this class of enzymes. In the 1940s Challenger proposed that the pathway of arsenic methylation involved a series of oxidative methylations such that the substrates

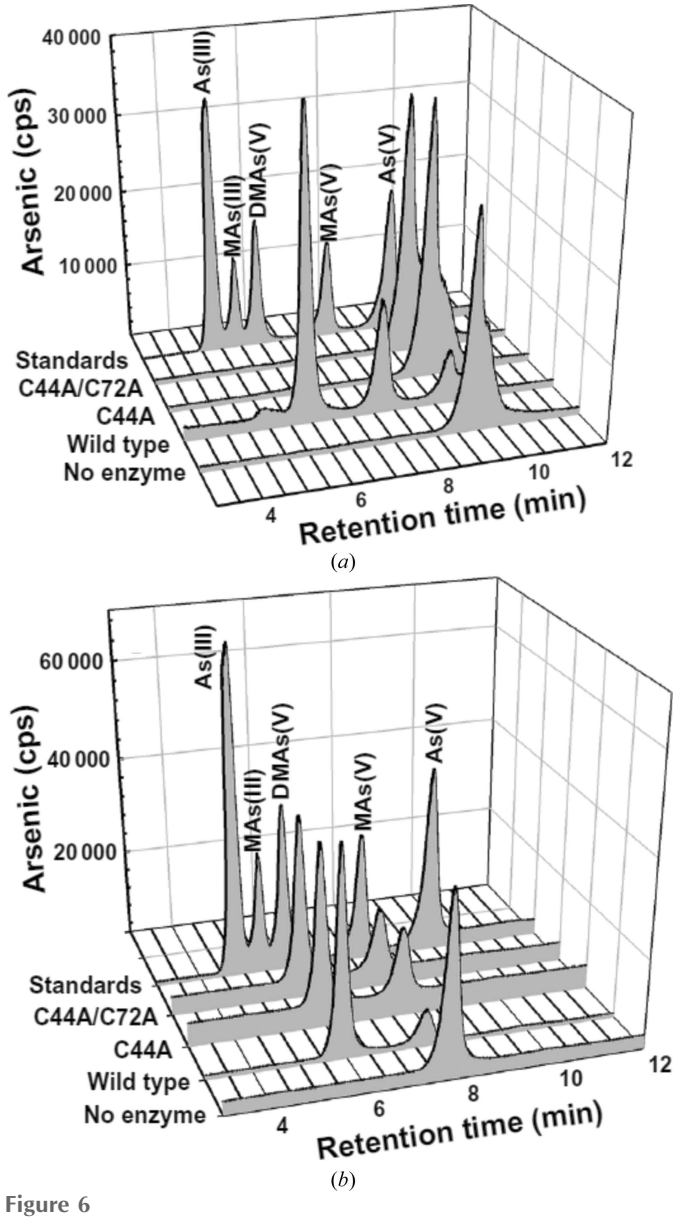


Figure 6
Cys44 is required for the methylation of arsenic(III) but not of MAs(III). Methylation of arsenic(III) or MAs(III) was assayed for 2 h at 60°C as described in §2 with either (*a*) arsenic(III) or (*b*) MAs(III) at 10 µ*M*. Samples were analyzed by reverse-phase HPLC–ICP-MS.

were trivalent arsenicals and the products were pentavalent (Challenger, 1947). Hayakawa *et al.* (2005) proposed that both the substrates and the products are trivalent. This difference between the two pathways is key to whether the methylated products are the relatively nontoxic and noncarcinogenic pentavalent species or the more toxic and carcinogenic trivalent species. Cullen recently reviewed the chemistry of methylation and concluded that 'the Challenger pathway with SAM as a donor of CH_3^+ remains the most rational option to describe the biological methylation of arsenic' which might suggest a positively charged pentavalent methylarsenic enzyme-bound intermediate (Cullen, 2014). However, no one has demonstrated the presence of a pentavalent arsenic intermediate bound to an arsenic(III) SAM methyltransferase. This question remains controversial and will require detailed structure–function and enzymological analysis to settle.

What can be learned from over 50 years of study of other members of the methyltransferase superfamily? They all append a methyl group from SAM to acceptor groups in S_N2 displacement reactions involving the attack of an electron-rich, methyl-accepting atom, usually O, N, C or S, on the SAM methyl group, with the release of SAH and inversion of the configuration of the methylated intermediate (Liscombe *et al.*, 2012). The SAM-binding fold is conserved, and they utilize a common enzymatic mechanism. In contrast to the proposed oxidation of As(III) in the Challenger pathway, no O-, N-, C- or S-SAM methyltransferase catalyzes substrate oxidation. Lysine and arginine methyltransferases, which, like AsMTs, also methylate their substrate three times, do so without oxidation of the nitrogen. Thus, as a class, these enzymes methylate their substrates without substrate oxidation and, by application of the principle of parsimony, there is no reason to consider that AsMTs would have evolved a mechanism

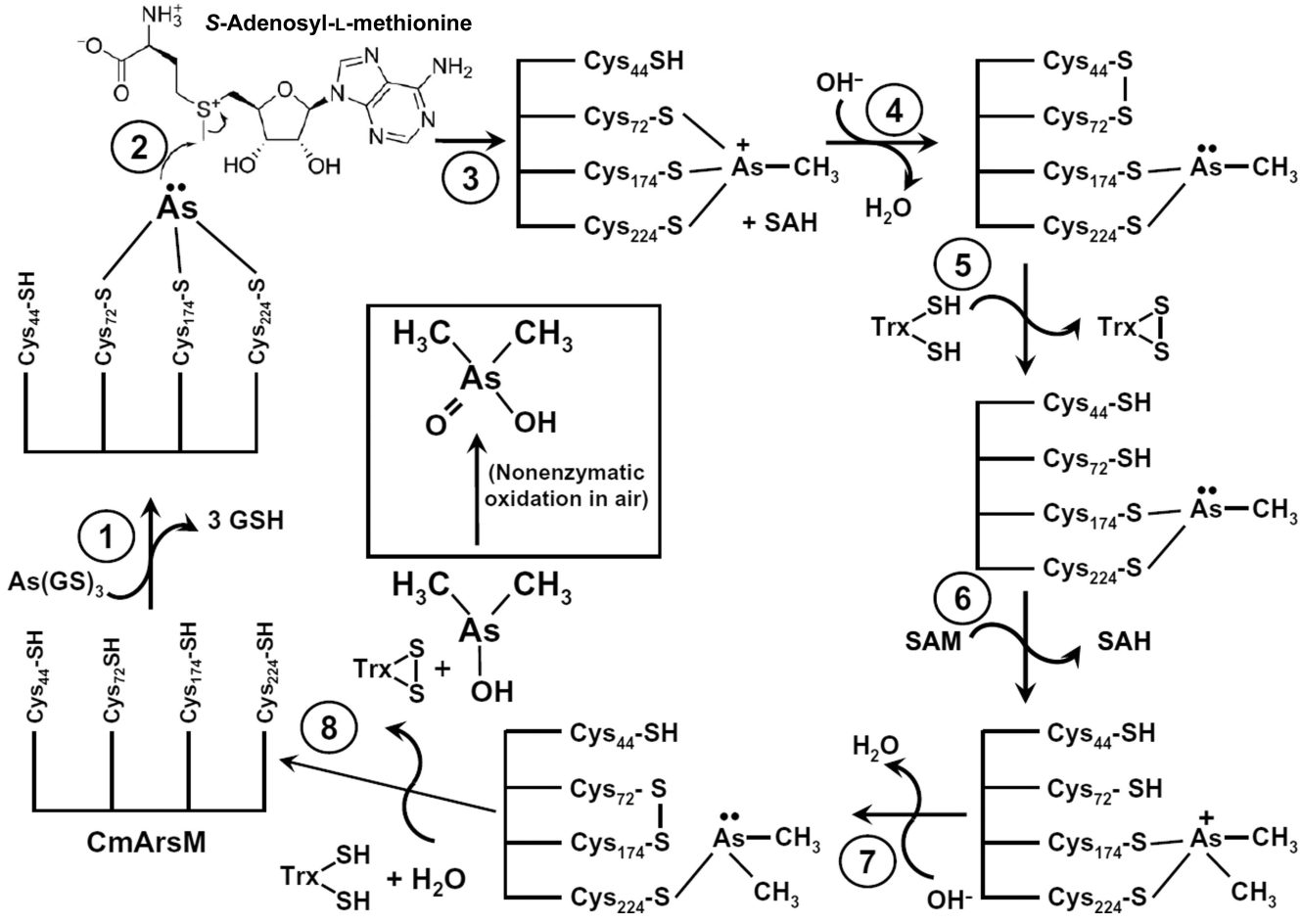


Figure 7
Proposed reaction scheme for AsMTs. The proposed reaction pathway is a general one for AsMTs, but the numbering of the cysteine residues follows the sequence of CmArsM. (1) In the first round of methylation AsMT binds arsenic(III) in a series of three thiol-transfer reactions. (2) The methyl group of SAM is attacked by the arsenic lone pair. (3) A positively charged pentavalent MAs(V) intermediate is formed and (4) reduced to an enzyme-bound MAs(III) intermediate by Cys44, with the formation of a Cys44–Cys72 disulfide. (5) The disulfide is reduced by Trx, regenerating the enzyme for the second round of methylation. (6) MAs(III), which is tightly bound to Cys174–Cys224, is methylated, producing an enzyme-bound positively charged pentavalent DMAs(V) intermediate (7) which is reduced to DMAs(III) by Cys72 with the formation of a Cys72–Cys174 disulfide. (8) DMAs(III), which is weakly bound to the single Cys224, dissociates from the enzyme and is oxidized nonenzymatically to DMAs(V). Finally, the Cys72–Cys174 disulfide is reduced by Trx, allowing the cycle to begin over.

different from other members of the superfamily. However, unlike the nonmetal (O, N, C and S) substrates, the metalloid arsenic is redox-active under physiological conditions, so a positively charged enzyme-bound intermediate is a possibility, either as an obligatory step in the catalytic cycle or as a side reaction owing to oxidation.

The structure–function study reported here was designed to address this issue. Based in part on the two different disulfide bonds observed in CmArsM structures, we propose a new and novel reaction scheme involving the formation of two sequential disulfide bonds between Cys44 and Cys72 and then between Cys72 and Cys174 (Fig. 7). Since we have no data about the third methylation, the proposed pathway considers only the first two methylations in eight steps. In the first round of methylation (1) CmArsM binds arsenic(III) in a series of three thiol-transfer reactions from $As(GS)_3$, the preferred substrate (Marapakala *et al.*, 2012). (2) The methyl group of SAM is attacked by the arsenic lone pair. (3) A positively charged pentavalent MAs(V) intermediate is formed. (4) Cys44 provides electrons to reduce the enzyme-bound MAs(V) intermediate to MAs(III), allowing the second round of methylation. Consistent with the postulation of an enzyme-bound intermediate, most of the arsenic is enzyme-bound as MAs(III) (Fig. 4) and little is released into the medium (Marapakala *et al.*, 2012) within the first 10 min of methylation. (5) Oxidized Cys44 forms a disulfide bond to Cys72, which must be reduced before the next round of methylation can occur. *In vivo*, Trx has been suggested to be the reductant for AsMTs (Thomas *et al.*, 2004), but in these studies GSH was used as the reductant because thermostable thioredoxin and thioredoxin reductase were not available. MAs(III) remains strongly bound by the thiol pair Cys174–Cys224. (6) A second methylation forms a positively charged pentavalent DMAs(V) intermediate which is (7) reduced to DMAs(III) by Cys72, forming a Cys72–Cys174 disulfide. (8) This disulfide is reduced by Trx, regenerating the enzyme and releasing DMAs(III), which nonenzymatically oxidizes in air to DMAs(V). Thus, the four conserved cysteine residues play two distinct roles, firstly as the binding site for arsenicals and secondly as a source of electrons to maintain arsenic in the reduced form.

This pathway explains nearly all of the current results with AsMTs. It resolve the differences between the Challenger hypothesis of oxidative methylations (Challenger, 1947), but with enzyme-bound pentavalent intermediates (Cullen, 2014), and the idea of Hayakawa and coworkers that the substrates and products are trivalent (Hayakawa *et al.*, 2005). It is not clear whether pentavalent intermediates are obligatory intermediates or side products. In either case, the arsenic must be reduced to trivalency before the reaction can proceed. The electrons to reduce the pentavalent intermediates come from conserved cysteine residues, which form disulfide bonds. Trx is proposed to be involved (Thomas *et al.*, 2004), but in this new model it serves in the classical role of thioredoxins in disulfide-bond reduction (Holmgren, 1989). Thus, AsMTs employ a basic catalytic mechanism similar to that of *O*-, *N*-, *C*- and *S*-methyltransferases. What differentiates AsMTs from other members of the methyltransferase superfamily is the necessity to bind trivalent arsenicals and to maintain them in the reduced form, for which they utilize four conserved cysteine residues.

Acknowledgements

This work was supported by NIH grant R37 GM55425. This project utilized the Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamline of the Advanced Photon Source, Argonne National Laboratory. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under contract No. W-31-109-Eng-38. The Berkeley Center for Structural Biology is supported in part by the National Institutes of Health, National Institute of General Medical Sciences and the Howard Hughes Medical Institute. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences of the US Department of Energy under Contract No. DE-AC02-05CH11231.

References

Abernathy, C. O., Thomas, D. J. & Calderon, R. L. (2003). *J. Nutr.* **133**, 1536S–1538S.
Ajees, A. A., Marapakala, K., Packianathan, C., Sankaran, B. & Rosen, B. P. (2012). *Biochemistry*, **51**, 5476–5485.
Beeby, M., O'Connor, B. D., Ryttersgaard, C., Boutz, D. R., Perry, L. J. & Yeates, T. O. (2005). *PLoS Biol.* **3**, e309.
Challenger, F. (1947). *Sci. Prog.* **35**, 396–416.
Chen, Y.-C., Su, H.-J., Guo, Y.-L. L, Hsueh, Y.-M., Smith, T. J., Ryan, L. M., Lee, M.-S. & Christiani, D. C. (2003). *Cancer Causes Control*, **14**, 303–310.
Cullen, W. R. (2014). *Chem. Res. Toxicol.* **27**, 457–461.
Drobná, Z., Del Razo, L. M., García-Vargas, G. G., Sánchez-Peña, L. C., Barrera-Hernández, A., Stýblo, M. & Loomis, D. (2012). *J. Expo. Sci. Environ. Epidemiol.* **23**, 151–155.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Fomenko, D. E., Xing, W., Adair, B. M., Thomas, D. J. & Gladyshev, V. N. (2007). *Science*, **315**, 387–389.
Gill, S. C. & von Hippel, P. H. (1989). *Anal. Biochem.* **182**, 319–326.
Hayakawa, T., Kobayashi, Y., Cui, X. & Hirano, S. (2005). *Arch. Toxicol.* **79**, 183–191.
Holmgren, A. (1989). *J. Biol. Chem.* **264**, 13963–13966.
Kinoshita, K., Noguchi, A., Ishii, K., Tamaoka, A., Ochi, T. & Kaise, T. (2008). *J. Chromatogr. B*, **867**, 179–188.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Laskowski, R. A. & Swindells, M. B. (2011). *J. Chem. Inf. Model.* **51**, 2778–2786.
Le, X. C., Lu, X., Ma, M., Cullen, W. R., Aposhian, H. V. & Zheng, B. (2000). *Anal. Chem.* **72**, 5172–5177.
Liscombe, D. K., Louie, G. V. & Noel, J. P. (2012). *Nat. Prod. Rep.* **29**, 1238–1250.
Liu, Z., Rensing, C. & Rosen, B. P. (2013). *Metals in Cells*, edited by V. Culotta & R. A. Scott, pp. 429–442. Hoboken: John Wiley & Sons.
Marapakala, K., Ajees, A. A., Qin, J., Sankaran, B. & Rosen, B. P. (2010). *Acta Cryst.* F**66**, 1050–1052.
Marapakala, K., Qin, J. & Rosen, B. P. (2012). *Biochemistry*, **51**, 944–951.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
McCoy, A. J. (2007). *Acta Cryst.* D**63**, 32–41.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Peters, B., Moad, C., Youn, E., Buffington, K., Heiland, R. & Mooney, S. (2006). *BMC Struct. Biol.* **6**, 4.
Putila, J. J. & Guo, N. L. (2011). *PLoS One*, **6**, e25886.
Qin, J., Lehr, C. R., Yuan, C., Le, X. C., McDermott, T. R. & Rosen, B. P. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 5213–5217.
Qin, J., Rosen, B. P., Zhang, Y., Wang, G., Franke, S. & Rensing, C. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 2075–2080.
Rossman, T. G., Uddin, A. N., Burns, F. J. & Bosland, M. C. (2001). *Toxicol. Appl. Pharmacol.* **176**, 64–71.
Tchounwou, P. B., Patlolla, A. K. & Centeno, J. A. (2003). *Toxicol. Pathol.* **31**, 575–588.
Thomas, D. J., Li, J., Waters, S. B., Xing, W., Adair, B. M., Drobna, Z., Devesa, V. & Styblo, M. (2007). *Exp. Biol. Med. (Maywood)*, **232**, 3–13.
Thomas, D. J. & Rosen, B. P. (2013). *Encyclopedia of Metalloproteins*, edited by R. H. Kretsinger, V. N. Uversky & E. A. Permyakov, pp. 138–143. New York: Springer.
Thomas, D. J., Waters, S. B. & Styblo, M. (2004). *Toxicol. Appl. Pharmacol.* **198**, 319–326.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Ye, J., Rensing, C., Rosen, B. P. & Zhu, Y.-G. (2012). *Trends Plant Sci.* **17**, 155–162.
Zhu, Y.-G., Yoshinaga, M., Zhao, F.-J. & Rosen, B. P. (2014). *Annu. Rev. Earth Planet. Sci.* **42**, 443–467.