Published in final edited form as:

Biochemistry. 2012 July 10; 51(27): 5476–5485. doi:10.1021/bi3004632.

The structure of an As(III) S-adenosylmethionine methyltransferase: insights into the mechanism of arsenic biotransformation

A. Abdul Ajees^{a,b,1}, Kavitha Marapakala^{a,1}, Charles Packianathan^a, Banumathi Sankaran^c, and Barry P. Rosen^{a,†}

^aDepartment of Cellular Biology and Pharmacology, Herbert Wertheim College of Medicine, Florida International University, Miami, FL 33199 USA

^bCentre for Atomic and Molecular Physics, Manipal Institute of Technology Campus, Manipal University, Manipal - 576 104, Karnataka, India

^cBerkeley Center for Structural Biology, Lawrence Berkeley Laboratory, 1 Cyclotron Road, BLDG 6R2100, Berkeley, CA 94720

Abstract

Enzymatic methylation of arsenic is a detoxification process in microorganisms but in humans may activate the metalloid to more carcinogenic species. We describe the first structure of an As(III) S-adenosylmethionine methyltransferase by x-ray crystallography that reveals a novel As(III) binding domain. The structure of the methyltransferase from the thermophilic eukarotic alga *Cyanidioschyzon merolae* reveals the relationship of the arsenic and S-adenosylmethionine binding sites to a final resolution of ~1.6 Å. As(III) binding causes little change in conformation, but binding of SAM reorients helix α4 and a loop (residues 49 to 80) toward the As(III) binding domain, positioning the methyl group for transfer to the metalloid. There is no evidence for a reductase domain. These results are consistent with previous suggestions that arsenic remains trivalent during the catalytic cycle. A homology model of human As(III) S-adenosylmethionine methyltransferase with the location of known polymorphisms was constructed. The structure provides insights into the mechanism of substrate binding and catalysis.

Arsenic is a ubiquitous environmental toxins and human carcinogen that poses a serious threat to human health and, consequently, ranks first on the Environmental Protection Agency's 2011 Comprehensive Environmental Response, Compensation, and Liability Act List of Hazardous Substances <<http://www.atsdr.cdc.gov/spl/>>. It is introduced primarily from geochemical sources and is acted on biologically, creating an arsenic biogeochemical cycle (1). Members of every kingdom, from bacteria to humans, methylate arsenite, producing the trivalent species methylarsenite (MAs(III)), dimethylarsenite (DMAs(III)) and volatile trimethylarsine (TMAs(III)) (2–4). The mammalian enzyme that catalyzes transfer of the methyl group of S-adenosylmethionine (SAM) to As(III) is AS3MT (5). The trivalent intermediates MAs(III) and DMAs(III) are formed during liver biotransformation of inorganic arsenate and arsenite, so, in humans, methylation has been proposed to activate inorganic arsenic to more carcinogenic species (6). Humans and other mammals excrete in urine dimethylarsenate (DMAs(V)) and, to a lesser extent, methylarsenate (MAs(V)) (3).

[†]To whom correspondence should be addressed. Barry P. Rosen. Department of Cellular Biology and Pharmacology, Herbert Wertheim College of Medicine, Florida International University, Miami, FL 33199 USA. Ph: 305-348-0657, Fax: 305-348-0651, prosen@fiu.edu.

¹These authors contributed equally to this study

Whether the oxidized species are products of AS3MT or are the result of nonenzymatic oxidation of the unstable trivalent species is controversial (7, 8). 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. Thus resolution of this uncertainty is of considerable consequence in our understanding of the health effects of arsenic.

In contrast, microbial arsenic methylation is an established detoxification process that is proposed to have an impact on the global arsenic cycle (9, 10). Genes for ArsM (arsenite S-adenosylmethyltransferase) orthologues of the human AS3MT are widespread in the genomes of bacteria, archaea, fungi and lower plants. CmArsM (accession number ACN39191) from an environmental isolate of the acidothermoacidophilic eukaryotic red alga *Cyanidioschyzon merolae* from Yellowstone National Park catalyzes arsenic methylation and volatilization, leading to resistance (9). CmArsM is a 400-residue thermostable enzyme (44,980 Da) that methylates As(III) to a final product of volatile TMAs(III).

To understand, on the one hand, how arsenic methylation is involved in carcinogenesis and, on the other hand, how microorganisms remodel the environment in arsenic-rich regions, we set the goal of elucidating the CmArsM catalytic cycle by a combination of biochemical (11) and structural (12) analyses. Here we report crystal structures of CmArsM with or without bound SAM or As(III), with a final resolution of 1.6 Å. The structure shows that CmArsM is a multimodular protein composed of a short N-terminal domain, a SAM binding domain, a novel arsenic-binding site and a C-terminal domain of unknown function. Three cysteine residues are conserved in ArsM orthologues; from the results of mutagenesis, substitution of any of the three leads to loss of As(III) methylation, indicating a role for these three residues in binding and catalysis (11). In contrast, Cys72 is not required for As(III) methylation, suggesting that it might have a different role in catalysis from the other two conserved cysteines. Consistent with this proposal, in the structure As(III) is bound by Cys174 and Cys224, while Cys72 moves toward the other two As(III)-binding cysteine residues when SAM is bound. Finally, by homology modeling the structure of the human AS3MT was modeled, and the location of the three known exonic single nucleotide polymorphisms identified.

MATERIALS AND METHODS

Purification and crystallization of CmArsM

CmArsM7B (termed simply CmArsM in this report) from *Cyanidioschyzon* sp. 5508, which has the the N-terminal 31 residues and the C-terminal 28 residues deleted, was prepared from *E coli* strain BL21(DE3) and purified as described previously (9). Selenomethionine labeled protein was prepared and purified as follows. Cells of *E coli* strain BL21(DE3) were grown overnight in 1 ml of LB medium (13), centrifuged at $3000 \times g$ for 10 min at 4 °C and suspended in an equal volume of M9 minimal medium (13) supplemented with 2 mM $MgSO_4$, 0.2% (w/v) glucose and 40 μM kanamycin. One ml of the cell suspension was diluted into 1L of M9 medium and allowed to grow to mid-exponential phase at 37 °C. At that point 100 mg each of lysine, phenylalanine, threonine, and 50 mg each of isoleucine, leucine and valine, and 60 mg of L-(+)-selenomethionine (Sigma) were added. The culture was grown for 15 min at 37 °C, and 0.5 mM isopropyl- β -D-thiogalactopyranoside (final concentration) was added to induce CmArsM. The culture was grown for an additional 6–8 hrs, and CmArsM was purified as described above.

Crystals of CmArsM were obtained as described (12) by the hanging drop vapor diffusion method at 20° C using Wizard II from Emerald Biosystems (20.0% (w/v) polyethylene glycol 3000, 0.1 M Tris–HCl, pH 7.0, containing 0.2 M calcium acetate). The best crystals obtained by microseeding were transferred to a cryoprotectant solution (25% polyethylene glycol 3350, 0.2 M calcium acetate, 0.1 M Tris–HCl, pH 7.0, and 10% glycerol (w/v)) and flash-cooled in liquid nitrogen. Crystals belonging to space group C2 with unit cell dimensions $a = 84.75 \text{ \AA}$, $b = 47.20 \text{ \AA}$, $c = 101.68 \text{ \AA}$ and $\beta = 115.67^\circ$ contained one molecule in the asymmetric unit. SeMet-labeled protein crystallized under the same conditions, and the CmArsM-As(III) complex crystals were obtained by equilibrating 2.5 μl of protein mixed with 2.5 μl reservoir solution (20.0% (w/v) polyethylene glycol 3000, 0.1 M Tris–HCl, pH 7.0, 0.2 M calcium acetate containing 2 mM As(III)). Attempts to obtain crystals of CmArsM by soaking with SAM were not successful, perhaps because a calcium ion occupies the SAM binding site. Instead, CmArsM was purified in the presence of 2 mM SAM, and crystals were obtained by including 10 mM SAM in both the hanging drop and reservoir solution, which contained 20 % polyethylene glycol 8000, 0.2 M NaCl and 0.1 M citrate phosphate, pH 4.2. Several rounds of seeding by mixing equal volumes of protein with reservoir solution were necessary to obtain crystals that diffracted to 2.75 $\text{\AA}$.

X-ray data collection, structure solution and refinement

Diffraction data from SeMet-labeled CmArsM crystals were collected at the selenium anomalous peak (Table 1). The structure was determined by single-wavelength anomalous dispersion at Se peak energy and refined using 1.6 $\text{\AA}$ resolution data. The position of all three SeMet sites were determined, heavy atom parameters refined, and single-wavelength anomalous dispersion (SAD) phases calculated using PHENIX (14). The SeMet crystals had one molecule in the asymmetric unit, and an initial model containing 266 of 322 residues was automatically built with PHENIX. A model with the remaining 56 residues was built with COOT (15) and refined using PHENIX. The final R_{work} and R_{free} values were 17.0 and 19.6, respectively.

Crystals of CmArsM were used to determine the ligand free structure (12). A data set from a crystal of ligand-free CmArsM containing residues 1–372 was collected to 1.78 $\text{\AA}$ resolution. This structure was solved by molecular replacement (Phaser (16)) using the SeMet model and refined using REFMAC5 (17) to R_{work} and R_{free} values of 17.7 and 20.5, respectively. Diffraction data from the crystals of CmArsM with bound SAM were collected to 2.75 $\text{\AA}$ resolution, and the structure was solved by molecular replacement (4) using the structure of the ligand-free protein as a search model and refined to final R_{work} and R_{free} values of 20.6 and 28.3, respectively. The structure of the protein with bound As(III) was refined using data collected to 1.75 $\text{\AA}$, with final R_{work} and R_{free} values of 20.0 and 23.9, respectively. The $R_{\text{free}}/R_{\text{work}}$ ratio for the SAM-bound structure of CmArsM is 1.37, which is in the acceptable range for a normal restrained isotropic refinement (18). The Ramachandran plots produced by MolProbity (18) had only three residues in the SAM-bound structure that were outliers, Asp79, Ser257 and Gly372. These are surface residues located in regions where the electron density poor. The electron density for the methionine moiety, which includes the methyl group, is not as good as the adenine and ribose groups. The SAM moiety was chosen instead of SAH for three reasons. First, crystals with bound SAH could not be obtained. Second, in the absence of the methyl acceptor, As(III), the methyl group should remain on SAM. Third, the temperature factor of the methyl group is similar to the atoms of the methionine moiety of the SAM ligand. However, we cannot rule out the possibility that a SAH impurity in the commercially-obtained SAM or a mixture of SAM/SAH is present in this complex.

All diffraction data sets were collected with an ADSC Quantum 315r (3×3 CCD array) detector at 100° K under a liquid nitrogen stream at beam line 5.01/5.02 of the Lawrence

Berkeley National Laboratory Advanced Light source. Data integration and scaling were performed with HKL2000 (19). Final data collection, refinement statistics, and Protein Data Bank accession codes are given in Table 1. Co-ordinates and structure factors have been deposited for ligand-free (3P7E), SAM- (3QHU) and As(III)-(3QNH) bound structures. Diffractions images and validation reports of the structure were available (<http://public.medicine.fiu.edu/CmArsM/default.aspx>). Electron densities for selected regions for ligand-free, As(III)- and SAM-bound structures are shown in Fig. 1. Molecular models were prepared using PyMOL (<http://www.pymol.org>) (20). Using the ligand-free CmArsM structure, models of the human AS3MT and its polymorphisms were built with the SWISS-MODEL fully automated protein structure homology modeling server (<http://swissmodel.expasy.org/>) (21, 22).

RESULTS

Structure of the CmArsM As(III) SAM methyltransferase

For crystallization purposes CmArsM7B residues 377–400 were substituted with a histidine tag (AAALEHHHHH) (12). This active derivative with 376 CmArsM residues was used as the ligand-free protein in this study. The first 49 residues were not visible; residues 50–372 were resolved at 1.7 Å and in selenomethionine-labeled protein at 1.6 Å; both structures were nearly identical, with an RMS deviation of 0.11 Å over 321 aligned Cα residues.

CmArsM assumes a compact globular structure with approximate dimensions of 57 × 56 × 35 Å (Fig. 2A). The N-terminal domain consists of residues 50–84, and the SAM binding domain consists of residues 85–173, 182–206, 231–253 and 270–280. The As(III) binding site consists of residues 174–181, 207–230 and 254–269, which are enclosed by the SAM binding domain, and the C-terminal domain consists of residues 281–372. CmArsM has a mixed structure consisting of α helices (α1–α12), β-strands (β1–β12), 3₁₀ helices (3₁₀1–3₁₀4) and long extended loops (Fig. 2B). The SAM binding domain adopts a class I methyltransferase Rossmann-fold structure (23). This domain has a central, parallel seven-stranded β sheet, β6–β7–β5–β4–β1–β2–β3, where all except β7 are parallel with one another, with four helices (α1–α4) and one 3₁₀ helix (3₁₀1) on one side of the sheet and two helices (α5, α8) and one 3₁₀ helix (3₁₀2) on the other side. Insertion of residues 174–181, 207–230 and 254–269 generates an arsenic binding pocket with three α helices (α6–α7, α9) and one 3₁₀ helix (3₁₀3) extending from the top of the SAM binding domain. SAM methyltransferase structures are well-represented in the Protein Data Bank (PDB), accounting for approximately 1.5% of all entries; a simple search of the PDB for methyltransferase retrieves a nonredundant set of 488 structures. Close structural homologues were identified by submitting the CmArsM structure to the DALI server (24). The top unique 25 structures (Z-score 19.7–15.4), as well as CmArsM, belong to the small molecule methyltransferase family (25).

The S-adenosylmethionine binding domain

The structure of CmArsM with bound SAM was determined at pH 4.2 to 2.75 Å resolution. The SAM cofactor is bound in between the β1/α2 loop containing the highly conserved glycine-rich sequence of D⁸⁹XGXGXG⁹⁵, the hallmark SAM-binding motif of Rossmann fold SAM-dependent methyltransferases, and corresponds well to the location of the cofactor in other SAM-dependant methyltransferases (23, 25) (Fig. 3). Among the 25 closest structural homologues, nine have SAM or its product S-adenosylhomocysteine (SAH), or the inhibitor sinefungin bound, and the SAM in CmArsM superimposes well on them with the RMS deviation of 0.9 Å for all atoms. SAM forms hydrogen bonds and hydrophobic interactions with CmArsM residues (Fig. 3). The adenine ring is sandwiched between Ile151 on one side and Met116 on the other side. The adenine ring forms hydrogen bonds with Ile151 and Glu152. The hydroxyl groups of ribose O2* and O3* are hydrogen bonded to the

side chains of Asp115 and Gln120. An acidic loop between $\beta 2$ and $\alpha 4$ that interacts with ribose hydroxyls is common in SAM-dependent methyltransferases (15). The methionine S-methyl group is bonded to main chain atoms O^{Tyr70} (3.41 Å) and O^{Cys174} (3.64 Å). The sulfur atoms of Cys72, Cys174 and Cys224 are 7.10 Å, 5.18 Å and 4.59 Å, respectively, distance from the S-methyl group. The sulfur atoms of Cys72 and Cys174 are 2.08 Å from each other, consistent with the possibility of a disulfide bond. The carboxyl group forms hydrogen bonds to O^{Cys92} (3.47 Å), O^{Tyr70} (3.66 Å), and the amide group is hydrogen bonded to O^{Gly91} (2.75 Å) (Fig. 3). It should be mentioned that the ligand-free protein crystallized only in the presence of 0.2 M calcium acetate. Under those conditions, a density was observed near Gly91. In contrast, calcium was not required when SAM was bound, and, as mentioned above, Gly91 hydrogen bonds with SAM. This density was best fit with a calcium ion coordinated with O^{Gly91} and six water molecules in a pentagonal bipyramidal configuration. The average bond distance between calcium and the oxygen atoms is 2.4 Å, a common geometry for a calcium ion.

The As(III) binding site

The structure of CmArsM with bound As(III) was determined at pH 7.5 to a resolution of 1.75 Å. The arsenic binding site has three modular components: 1) unit M1 consists of an insertion of eight residues (residues 174–181) between $\beta 4$ and $\alpha 5$, 2) unit M2 consists of an insertion of 24 residues (residues 207–230) between $\beta 5$ and $\alpha 8$, and 3) unit M3 consists of a third insertion of 16 residues (residues 254–269) between $\beta 6$ and $\beta 7$ (Supplementary Figure 1). A DALI server search for structural homologues with similar domains revealed no proteins with a structural similarity, and, compared to the closest 25 methyltransferase structures, the arsenic binding site appears to be novel in methyltransferases. In a comparison between CmArsM and the 25 homologues several differences were noted. First, at the position of M1 insertion in CmArsM (between $\beta 4$ and $\alpha 5$), the others have similar fold, including a 3_{10} helix. Second, at the position of the M2 (between $\beta 5$ and $\alpha 8$) and M3 (between $\beta 6$ and $\beta 7$) insertions in CmArsM, the others have different combinations of α helices and β sheets, and none are similar to CmArsM.

The crystal structure of CmArsM with bound As(III) reveals a pyramidal site in which the central arsenic atom is coordinated with the thiolates of Cys174 and Cys224 at an average distance of 2.21 Å and a third non-protein ligand at a distance of 2.26 Å. Each of the three liganding atoms is at an average distance of 3.34 Å (Fig. 4). Two possibilities for the atom at the observed distance of 2.26 Å are chloride or sulfur. It is unlikely to be either water or hydroxide ion, which left an extra density during structure refinement. Since the protein was purified in the presence of 0.5 M NaCl, and the protein was not exposed to inorganic sulfide during purification or crystallization, a reasonable deduction is that the inorganic ligand is chloride. During refinement of the CmArsM-As(III) structure, an additional density was observed near the conserved residues Cys72. A glycerol moiety fit well into the density. This density was present only when As(III) was bound to Cys174 and Cys224.

The C-terminal domain

A major difference between the pathways of Challenger (7) and Hayakawa et al. (8) is whether the product of each partial reaction is trivalent or pentavalent. AS3MT has been proposed to catalyze reduction of pentavalent products back to trivalency using exogenous reductants (26). However, neither the SAM nor As(III) binding site has the hallmark of a reductase. The C-terminal domain (residues 281–372) appears to be novel and has no structural homology with reductase structures. This domain has a mixed structure consisting of α helices ($\alpha 10$ – $\alpha 12$), β -strands ($\beta 8$ – $\beta 12$), 3_{10} helix ($3_{10}4$) and long extended loops. Note that residues 373–400, containing two Cys-Cys pairs that could be additional As(III) binding sites, are missing in the CmArsM structure.

Relationship of the As(III)- and SAM-binding structures

Superposition of ligand-free CmArsM with the SAM-bound structure (RMSD 1.27 Å) revealed that residues 50 to 79 moved by 2.0 – 4.0 Å when SAM is bound (Supplemental Figure 2A). We predict that the SAM-bound form is an intermediate conformation of the catalytic cycle, and that Cys72 may reorient to become the third ligand to As(III) during another step of the cycle. In contrast, arsenic binding does not lead to global structural changes when the As(III)-bound structure was compared with the ligand-free protein (RMSD 0.22 Å) (Supplemental Figure 2B). The side chain of Cys174 reorients to allow formation of the bond with As(III). Finally, an overlay of the ligand-free structure with both the As(III)-bound and SAM-bound structures was constructed by superposition of the C α atoms of all three structures (Supplemental Figure 2C). The ligand-free structure can be superimposed with the arsenic-bound and the SAM-bound structures with RMSDs of 0.22 Å over 322 aligned C α residues and 1.27 Å over 320 aligned C α residues, respectively. As mentioned, the ligand-free and arsenic-bound complexes adopt similar conformations compared to the SAM-bound complex. In the SAM-bound complex, helix α 4 and residues 50–80, which include conserved residue Cys72, exhibit substantial movement compared to the ligand-free structure by 2.0 – 4.0 Å (Fig. 5 and Supplemental Figure 3A).

Attempts to crystallize CmArsM in complex with both As(III) and SAM or SAH (or with SAH alone) were unsuccessful, so the ternary complex was modeled by superimposition of the As(III)- and SAM-bound structures (Fig. 6 and Supplemental Figure 3B). The model shows that the methyl group of the SAM moiety is directed towards the arsenic binding pocket. Cys174 and Cys224 are 5.0 Å and 4.50 Å, respectively, from the methyl group of SAM, and the arsenic-SAM methyl distance is 4.22 Å. The chlorine atom is exposed to solvent on the opposite site of the arsenic atom from the SAM cofactor (5.80 Å from the methyl group), and the arsenic lone pair would be oriented toward the carbon of the methyl group, facilitating methyltransfer in an S_N2 reaction.

Conserved residue Cys72 is not a ligand to As(III) in this structure. From the results of mutagenesis, Cys72 appears to be required for methylation of As(III) but not MAs(III), suggesting that it might be a third ligand to As(III) (11). In the ligand-free and As(III)-bound structures, the C α of Cys72 is 8.23 Å from the arsenic atom, but, in the SAM-bound structure, the loop α 2– β 101 (residues 68–80) containing Cys72 moves toward the As(III)-binding site, and the C α of Cys72 closes to 6.58 Å from the arsenic atom, suggesting that Cys72 might participate in As(III) binding during a step in the catalytic cycle. The data suggest that Cys72 and Cys174 might form a disulfide bond when SAM is bound, but it is not clear at this time whether an oxidized form of the enzyme might be an intermediate in the pathway of methyltransfer or represents a crystallization artifact. What other CmArsM residues participate in catalysis? The methyl group of SAM makes van der Waals contacts with main chain carbonyl oxygen atoms of Tyr70 (3.41 Å) and Cys174 (3.64 Å). The carbonyl oxygen of these two residues may help to orient the methyl group of SAM during its approach to the arsenic lone pair.

Finally, in all structures a positive density was observed in the location of the β 102– β 4 loop and side chains of Lys110, Arg145 and Glu160. Analysis using the PHENIX automated ligand search program suggested either a polyethylene glycol or glycerol moiety, both of which were present during crystallization and freezing of the crystals. Attempts to build these moieties with either of full or low occupancy only resulted in high temperature factors and excessive positive or negative difference electron density. Due to the ambiguity of this unknown density, it was not modeled in the data submitted to the PDB.

Modeling human AS3MT

One goal of structural analysis of this As(III) SAM-methyltransferase is to gain an understanding of the metabolism of arsenic in human liver and its relationship to carcinogenesis. Alignment of CmArsM with AS3MT from human, rat and zebra fish shows sequence identity of approximately 42% and overall similarity of 60% with each of the three (Fig. 7). Full-length CmArsM has 400 residues, while human AS3MT has 375. Twelve of those represent an N-terminal extension in CmArsM, and the other additional residues are the result of small insertions. The three conserved cysteine residues proposed to participate in catalysis in CmArsM, Cys72, Cys174 and Cys224 (11), are Cys61, Cys156 and Cys206 in human AS3MT (yellow in Supplemental Figure 4).

A model of human AS3MT was built on the ligand-free CmArsM structure using the SWISS-MODEL fully automated protein structure homology modeling server (<http://swissmodel.expasy.org/>) (21, 22) (Fig. 7A). The model quality was estimated based on the QMEAN scoring function of 0.63, which is within the acceptable range (27). Residues 50 – 363 from CmArsM were used as template, and the final homology model incorporated 291 of those 323 residues. The secondary structural arrangement of human model is nearly the same as the CmArsM structure, with equivalent As(III) and SAM binding elements. The C α backbones of CmArsM and human AS3MT are nearly superimposable, with an RMSD value of 0.63Å over 291 residues. This structural similarity gives confidence that these two orthologues employ equivalent reaction mechanisms.

The human AS3MT gene is approximately 32-kilobase nucleotide base pairs and composed of 11 exons (28). In recent years a number of intronic single nucleotide polymorphisms (SNPs) have been identified. Three exonic SNPs, R173W, M287T and T306I (cyan in Fig. 7A), have been identified in the AS3MT coding region of African-American and Caucasian-American subjects. Arg173 and Thr306 are conserved in CmArsM as Arg191 and Thr327. Residue Lys305 is in the position corresponding to hAS3MT residue Met287. Met287 and Thr306 are in the C-terminal domain, and Arg173 is in helix α 5 of the arsenic binding domain. Each of the three appears to have an effect on enzyme stability, with the M287T polymorphism stabilizing and the other two decreasing stability (28). Individuals with the M287T SNP displayed increased urinary production of DMAs and might be at higher risk for toxic and genotoxic effects of arsenic exposure (29). Little difference was observed between the models of the normal AS3MT, the M287T and the T306I proteins, so it is difficult to relate those polymorphisms to their phenotypes. However, the fact that the M287T polymorphism leads to increased protein stability suggests a role for the C-terminal domain in AS3MT folding and/or function.

Polymorphism R173W is predicted to have a conformational change in the As(III) binding domain. There is a putative ~3.0 Å hydrogen bond between Arg173 NE with OE1 of residue Glu170, which is one turn of the helix distant. The polymorphic residue Trp173 NE1 has a predicted interaction with Glu70 OE1 of ~4.5 Å. A predicted consequence of loss of the Arg173-Glu170 hydrogen bond is movement of the loop connecting to helix α 5 and then to putative As(III) binding residue Cys156 (Fig. 7B). This conformational change could lead to the observed instability of the polymorphic enzyme (28).

DISCUSSION

Arsenic methylation is a biotransformation carried out by many organisms, from bacteria to humans. The process is catalyzed by the enzyme As(III) S-adenosylmethyltransferase (ArsM in microorganisms and AS3MT in vertebrates). ArsM activity in bacteria and eukaryotic algae is sufficient to detoxify arsenic and acts in parallel to other detoxification mechanisms such as efflux systems (9, 10). The physiological function of human AS3MT is less clear.

AS3MT was originally proposed to detoxify arsenic, but more recently has been postulated to transform inorganic arsenic into the more carcinogenic methylated species MAs(III) and DMAs(III) (5). The enzyme has a complicated pathway that produces mono-, di- and trimethylated arsenicals. There is no agreement on the chemical nature of the substrates and products or on the pathway itself. One the one hand, As(III) has been proposed to undergo oxidative methylation to MAs(V), which would then be reduced to MAs(III), the substrate of the next methylation reaction. The next two steps would go through similar oxidative methylations. This pathway requires the enzyme to have binding sites for a series of both trivalent and pentavalent arsenicals and to carry out two quite different types of reactions, methylation and reduction (26). There are no data that exclude binding of pentavalent species to CmArsM, so this remains a possibility. On the other hand, a simpler pathway has been proposed in which arsenic remained trivalent throughout, with the products being MAs(III), DMAs(III) and TMAs(III) (8). Our recent characterization of CmArsM is consistent with the latter pathway (11). The results of mutagenesis suggest that the three conserved cysteine residues, Cys72, Cys174 and Cys224, are required for As(III) binding and methylation, while only two, Cys174 and Cys224 are involved in MAs(III) binding and methylation. The rate limiting step in the proposed catalytic mechanism is the final methylation reaction (11), which explains why the primary product is usually DMAs(III) and not TMAs(III).

To understand the relationship of the proposed catalytic residues to function, we examined and compared the structure of CmArsM in three states: ligand-free, with bound As(III) and with bound SAM. As(III) binding sites invariably involve cysteine residues. There are 17 cysteine residues in full-length CmArsM, and 11 in the C-terminally truncated construct used in this study, of which only Cys72, Cys174 and Cys224 are conserved in the 100 closest homologues in the BLINK database. In all reported As(III) SAM methyltransferases orthologues there are multiple cysteine residues, usually cysteine pairs, near the C-terminus. The exact sequences are not conserved, but the ubiquitous presence of cysteine pairs suggests an arsenic-related function. At the C-terminus of CmArsM there are six cysteine residues, including two cysteine pairs that are not required for catalysis (11). Although there are no data that speak to a function for these cysteines, we predict that they serve as arsenic sensors or chaperones to transfer As(III) to the active site cysteines. The N-terminal region, which was not visible in the structure, has four non-conserved cysteine residues. The remaining seven cysteine residues were identified in the structure. Of those, four (Cys72, Cys174, Cys176 and Cys224) are in the arsenic binding pocket, one in SAM binding pocket (Cys92) and other two are in β 7-strand (Cys273 and Cys277). Cys44, Cys92 and Cys273 are found in 83, 72 and 6, respectively, of the 100 closest homologues. Cys176, which is near the arsenic binding pocket, and Cys277 are found only in CmArsM but not in other close homologues. Even though As(III) is observed bound to only Cys174 and Cys224, there are three reasons to believe that As(III) is bound to the thiolate of Cys72 as well during the catalytic cycle. First, none of the other cysteine residues are conserved in other orthologues. Second, Cys72 is observed to move toward Cys174 and Cys224 when SAM is bound. Third, when Cys72 was changed to an alanine residue, the resulting mutant neither bound nor methylated As(III) (11). While participation of other non-conserved cysteines cannot be ruled out, the results support participation of Cys72 as an arsenic ligand during the catalytic cycle. The fact that it is not a ligand in this As(III)-bound structure emphasizes the need to obtain more crystal forms.

Since either As(III) or SAM can be bound in the absence of the other, the reaction most likely involves random binding of the two substrates. Once both are bound, the enzyme is poised to catalyze transfer of the methyl group to the arsenic atom. In the model with both As(III) and SAM bound, the methyl group of SAM is 4.22 Å from the arsenic atom, which is somewhat longer than expected, and the two must approach each other more closely

during the methyl transfer step. In the structure As(III) is bound to the two cysteine thiolates at a distance of 2.2 Å, and a third, nonprotein ligand is behind the arsenic atom positioned to be a leaving group in an S_N2 reaction. In the crystal structure this third ligand is likely a chlorine atom, perhaps due to the high chloride concentration in the crystallization buffer. As discussed above, *in vivo* the third ligand is proposed to be the thiolate of Cys72. A comparison of the position of the Cα of Cys72 in the As(III)-bound structure and the model with SAM and As(III) indicates that the Cys72 Cα moves closer to the arsenic when SAM binds (from 8.23 Å to 6.58 Å), consistent with the suggestion that Cys72 is the third ligand during catalysis. Since it appears that the arsenic donor *in vivo* is As(GS)₃ (11), another possibility for the third ligand *in vivo* is glutathione.

We propose that the initial step in catalysis is binding of As(III) to the thiolates of Cys72, Cys174 and Cys224. In an S_N2 reaction the positive charge on the SAM sulfur atom pulls the bonding electron from the carbon of the methyl group, which interacts with the arsenic lone pair to form an As-C bond, producing SAH. The sulfur atom of one of the three conserved cysteines, which we predict is Cys72, becomes the leaving group. The methylarsenic product, which has higher affinity for CmArsM than does the substrate arsenite, remains bound to Cys174 and Cys224 to undergo a second round of methylation, and we predict that either Cys174 or Cys224 becomes the leaving group (11). The dimethylated product becomes the substrate for the third round of methylation, with the remaining cysteine residue becoming the leaving group. Because DMAs(III) can have only a single cysteine coordination, it is bound to the active site with much lower affinity than either As(III) or MAs(III). Thus the products would be in order of prevalence: DMAs(III)>MAs(III)>>TMAs(III). The species of methylated arsenic found in urine are usually, in order of prevalence, DMAs(V)>MAs(V)>>TMAs(V)O. With careful sample preparation, MAs(III) and DMAs(III) are found in human urine (30), suggesting that the pentavalent urinary species may be the result of oxidation during storage and sample preparation (31). In conclusion, the high resolution structure from the thermophilic eukarotic alga *Cyanidioschyzon* is the first structure of an As(III) SAM methyltransferase. It is representative of the structure of other eukaryotic orthologues, including human AS3MT. The structure identifies the arsenic and SAM binding domains, as well as other residues potentially involved in arsenic biomethylation. A number of questions remain, for example, elucidation of the function of the N- and C-terminal domains and the C-terminal cysteines and the participation of GSH is necessary to understand the catalytic mechanism. This structure, together with recent biochemical analysis of CmArsM (11) supports the hypothesis that the products of As(III) SAM methyltransferases are trivalent (8), and not the currently supposed pentavalent forms (5). The distinction is important: generation of trivalent methylated arsenicals poses a much greater risk for diseases such as bladder cancer than would pentavalent methylated products (6). If As(III) SAM methyltransferases catalyze the oxidative methylation of arsenic, producing primarily DMAs(V) and, to a lesser extent, MAs(V) in the cytosol of cells, then the carcinogenic potential of arsenic is less than if the intracellular products are MAs(III) and DMAs(III). Our results imply that the risk of cancer from arsenic exposure may be greater than previously assumed and provide an initial model that may be useful for understanding the relationship of arsenic methylation and carcinogenesis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported by U.S. National Institutes of Health Grant GM55425. 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 U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Abbreviations

ArsM or AS3MT	As(III) S-adenosylmethionine methyltransferase
SAM	S-adenosylmethionine
SAH	S-adenosylhomocysteine
PDB	protein database

REFERENCES

1. Bhattacharjee, H.; Rosen, BP. Arsenic metabolism in prokaryotic and eukaryotic microbes. In: Nies, DH.; Simon, S., editors. *Molecular microbiology of heavy metals*. Heidelberg/New York: Springer-Verlag; 2007. p. 371-406.
2. Bentley R, Chasteen TG. Microbial methylation of metalloids: arsenic, antimony, and bismuth. *Microbiol Mol Biol Rev*. 2002; 66:250–271. [PubMed: 12040126]
3. Thomas DJ, Waters SB, Styblo M. Elucidating the pathway for arsenic methylation. *Toxicol Appl Pharmacol*. 2004; 198:319–326. [PubMed: 15276411]
4. Rensing, C.; Rosen, BP. Heavy metals cycles (arsenic, mercury, selenium, others). In: Schaechter, M., editor. *Encyclopedia of Microbiology*. Oxford, U.K.: Elsevier; 2009. p. 205-219.
5. Styblo M, Drobna Z, Jaspers I, Lin S, Thomas DJ. The role of biomethylation in toxicity and carcinogenicity of arsenic: a research update. *Environ Health Perspect*. 2002; 110(Suppl 5):767–771. [PubMed: 12426129]
6. Styblo M, Del Razo LM, Vega L, Germolec DR, LeCluyse EL, Hamilton GA, Reed W, Wang C, Cullen WR, Thomas DJ. Comparative toxicity of trivalent and pentavalent inorganic and methylated arsenicals in rat and human cells. *Arch Toxicol*. 2000; 74:289–299. [PubMed: 11005674]
7. Challenger F. Biological methylation. *Advances in enzymology and related subjects of biochemistry*. 1951; 12:429–491. [PubMed: 14885024]
8. Hayakawa T, Kobayashi Y, Cui X, Hirano S. A new metabolic pathway of arsenite: arsenic-glutathione complexes are substrates for human arsenic methyltransferase Cyt19. *Arch Toxicol*. 2005; 79:183–191. [PubMed: 15526190]
9. Qin J, Lehr CR, Yuan C, Le XC, McDermott TR, Rosen BP. Biotransformation of arsenic by a Yellowstone thermoacidophilic eukaryotic alga. *Proc Natl Acad Sci U S A*. 2009; 106:5213–5217. [PubMed: 19276121]
10. Qin J, Rosen BP, Zhang Y, Wang G, Franke S, Rensing C. Arsenic detoxification and evolution of trimethylarsine gas by a microbial arsenite S-adenosylmethionine methyltransferase. *Proc Natl Acad Sci U S A*. 2006; 103:2075–2080. [PubMed: 16452170]
11. Marapakala K, Qin J, Rosen BP. Identification of catalytic residues in the As(III) S-adenosylmethionine methyltransferase. *Biochemistry*. 2012; 51:944–951. [PubMed: 22257120]
12. Marapakala K, Ajees AA, Qin J, Sankaran B, Rosen BP. Crystallization and preliminary X-ray crystallographic analysis of the ArsM arsenic(III) S-adenosylmethionine methyltransferase. *Acta crystallographica*. 2010; 66:1050–1052. [PubMed: 20823523]
13. Sambrook, J.; Fritsch, EF.; Maniatis, T. *Molecular cloning, a laboratory manual*. New York: Cold Spring Harbor Laboratory; 1989.
14. Adams PD, Gopal K, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Pai RK, Read RJ, Romo TD, Sacchettini JC, Sauter NK, Storoni LC, Terwilliger TC. Recent

- developments in the PHENIX software for automated crystallographic structure determination. *J Synchrotron Radiat*. 2004; 11:53–55. [PubMed: 14646133]
15. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*. 2004; 60:2126–2132. [PubMed: 15572765]
 16. McCoy AJ. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr D Biol Crystallogr*. 2007; 63:32–41. [PubMed: 17164524]
 17. Winn MD, Murshudov GN, Papiz MZ. Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol*. 2003; 374:300–321. [PubMed: 14696379]
 18. Tickle IJ, Laskowski RA, Moss DS. Rfree and the rfree ratio. I. Derivation of expected values of cross-validation residuals used in macromolecular least-squares refinement. *Acta Crystallogr D Biol Crystallogr*. 1998; 54:547–557. [PubMed: 9761849]
 19. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol*. 1997; 276:307–326.
 20. DeLano, WL. The PyMOL user's manual. San Carlos, CA: DeLano Scientific; 2001.
 21. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*. 2006; 22:195–201. [PubMed: 16301204]
 22. Kiefer F, Arnold K, Kunzli M, Bordoli L, Schwede T. The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res*. 2009; 37:D387–D392. [PubMed: 18931379]
 23. Schubert HL, Blumenthal RM, Cheng X. Many paths to methyltransfer: a chronicle of convergence. *Trends in biochemical sciences*. 2003; 28:329–335. [PubMed: 12826405]
 24. Holm L, Kaariainen S, Rosenstrom P, Schenkel A. Searching protein structure databases with DaliLite v.3. *Bioinformatics*. 2008; 24:2780–2781. [PubMed: 18818215]
 25. Martin JL, McMillan FM. SAM (dependent) I AM: the S-adenosylmethionine- dependent methyltransferase fold. *Curr Opin Struct Biol*. 2002; 12:783–793. [PubMed: 12504684]
 26. Waters SB, Devesa V, Del Razo LM, Styblo M, Thomas DJ. Endogenous reductants support the catalytic function of recombinant rat cyt19, an arsenic methyltransferase. *Chemical research in toxicology*. 2004; 17:404–409. [PubMed: 15025511]
 27. Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*. 2011; 27:343–350. [PubMed: 21134891]
 28. Wood TC, Salavagionne OE, Mukherjee B, Wang L, Klumpp AF, Thomae BA, Eckloff BW, Schaid DJ, Wieben ED, Weinshilboum RM. Human arsenic methyltransferase (AS3MT) pharmacogenetics: gene resequencing and functional genomics studies. *J Biol Chem*. 2006; 281:7364–7373. [PubMed: 16407288]
 29. Hernandez A, Xamena N, Sekaran C, Tokunaga H, Sampayo-Reyes A, Quinteros D, Creus A, Marcos R. High arsenic metabolic efficiency in AS3MT287Thr allele carriers. *Pharmacogenet Genomics*. 2008; 18:349–355. [PubMed: 18334919]
 30. McKnight-Whitford A, Chen B, Naranmandura H, Zhu C, Le XC. New method and detection of high concentrations of monomethylarsonous acid detected in contaminated groundwater. *Environ Sci Technol*. 2010; 44:5875–5880. [PubMed: 20583830]
 31. Feldmann J, Lai VW, Cullen WR, Ma M, Lu X, Le XC. Sample preparation and storage can change arsenic speciation in human urine. *Clin Chem*. 1999; 45:1988–1997. [PubMed: 10545070]
 32. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*. 2010; 66:12–21. [PubMed: 20057044]

Fig. 1. _xxx_Electron density maps

The electron density of the unbiased omit map Fo-Fc of ligand-free CmArsM is shown (A) and bound As(III) (B). Both maps are contoured at the 3.0 σ level. C. The SAM cofactor is shown as ball and stick model, and the SAM electron density outlined with the unbiased omit Fo-Fc map contoured at 2.5 σ (grey) or 1.0 σ (orange). The density maps are shown in gray, carbon in green, sulfur yellow, nitrogen blue and oxygen red.

Fig. 2. Structure of CmArsM

A. Ribbon representation of ligand-free CmArsM (PDB ID: 3P7E). N and C indicate the N- and C-terminal domains and are colored blue and red, respectively. The elements in cyan comprise a Rossmann fold. The position of the arsenic domain is marked M1 (forest green), M2 (pale green) and M3 (violet purple). A calcium ion is shown in orange sphere, and cysteine residues are shown in ball and stick and colored green (carbon) and yellow (sulfur). **B.** Topological diagram of secondary structural elements. The numbering and coloring are as in **A**. α helices are drawn as cylinders, β strands as arrows. Cysteine residues are shown in circles, and those in the putative arsenic binding site are shown in yellow. The locations of the arsenic binding site and the SAM-binding domain are indicated.

Fig. 3. Structural overview of CmArsM with bound SAM

SAM cofactor interactions with CmArsM residues (PDB Id: 3QHU) are shown. Bound SAM is shown in stick and colored green (carbon), red (oxygen) and blue (nitrogen), and residues interacting with SAM are shown in stereo view and labeled. Potential hydrogen bond network around the SAM is indicated with dash lines, with distances in Å units.

Fig. 4. Structural overview of CmArsM with bound As(III)

A. The bound arsenic and chlorine atoms are shown as spheres and colored magenta and red, respectively. The arsenic atom is coordinated with thiolates of Cys174, Cys224 and a chlorine atom with distances of 2.2–2.3 Å. Each of the three liganding atoms is at an average distance of 3.34 Å from each other.

Fig. 5. Relationship of the As(III)- and SAM-binding domains

Superposition of ligand-free CmArsM (green) with the SAM-bound (blue) structure. The two structures have an RMSD of 1.268 Å. A calcium ion is shown as an orange sphere. The SAM molecule is shown in sticks and colored in cyan (carbon), blue (nitrogen) and red (oxygen). The thiolates of cysteine residues are colored in yellow. The movement of Cys72 as a result of SAM binding is indicated by an arrow.

Fig. 6. Ternary complex of CmArsM

The ternary complex of CmArsM, SAM and As(III). was modeled by superposition of the As(III)- and SAM-bound structures and colored as in Fig. 1. Selected distances from the methyl group of SAM are shown in Å units. Bound SAM is highlighted in cartoon and colored blue.

Fig. 7. Homology model of human AS3MT

A. Model of human AS3MT. Homology model of human AS3MT based on the 1.78 Å ligand-free CmArsM structure. Coloring is based on secondary structure elements (α helices in green, β strands in blue and loops in salmon). Putative arsenic binding cysteine residues and polymorphic residues are shown in ball and stick. **B.** Comparison of residues 136 to 180 of the normal human AS3MT with the R173W polymorphism. The normal human AS3MT is shown in cyan, and the corresponding region of the R173W polymorphic protein is shown in magenta. The Arg-to-Trp mutation is predicted to produce a conformational change as a result of a shift in the hydrogen bond between Arg173 and Glu170 (3.0 Å) to a longer hydrogen bond between Trp173 and Glu170 (4.5 Å).

Table 1

Summary of diffraction data and structure refinement statistics

	SeMet	Ligand free ^a	SAM-bound	As(III)-bound
Diffraction data				
Wavelength (Å)	0.9795	0.9795	0.9795	0.9795
Space group	C2	C2	C222	C2
Cell parameters				
a, b, c (Å)	84.90, 46.93, 100.40	84.75, 47.20, 101.68	67.18, 128.79, 96.52	85.25, 46.76, 100.53
α, β, γ (°)	90.00, 114.44, 90.00	90.00, 115.67, 90.00	90.00, 90.00, 90.00	90.00, 114.33, 90.00
Resolution (Å)	50.00-1.60 (1.66-1.60) ^b	50.00-1.78 (1.84-1.78)	50.00-2.75 (2.87-2.75)	45.80-1.75 (1.81-1.75)
No. of unique reflections	46076 (4440)	34951 (3450)	11298 (1394)	35321 (3200)
Completeness (%)	96.6 (94.3)	99.9 (99.5)	99.9 (99.9)	96.8 (88.2)
Redundancy	4.5 (4.5)	4.4 (3.5)	5.9 (5.4)	6.2 (4.8)
$\langle I \rangle / \langle \sigma(I) \rangle$	31.06 (5.12)	37.06 (5.83)	17.07 (3.70)	23.93 (2.56)
R _{merge} (%)	5.3 (24.8)	3.4 (19.1)	9.8 (48.5)	6.4 (51.0)
R _{cryst} / R _{free} (%)	17.0 / 19.6	17.7 / 20.5	20.6/28.3	20.0 / 23.9
B factors (Å) ² (# atoms)				
All atoms	28.17 (5520)	25.4 (2856)	52.3 (2582)	34.9 (2742)
Protein	27.40 (5121)	24.3 (2552)	52.1 (2555)	34.4 (2531)
Ligand / ion	CA – 16.8 (1)	CA – 20.5 (1)	SAM – 71.1 (27)	As – 38.2 (1) Cl – 45.8 (1) CA – 35.4 (2)
Water	38.104 (398)	34.8 (303)	---	40.7 (207)
RMSD				
Bond lengths (Å)	0.011	0.012	0.011	0.012
Bond angles (°)	1.305	1.497	1.623	1.489
Ramachandran plot (%) ^c				
Favored regions	99.1	98.8	92.6	98.1
Allowed regions	100.0	100.0	99.1	100.0
Outliers	---	---	3 residues: Asp79, Ser257 and Gly372	---
Missing residues	1-48, 371-383	1-49, 373-383	1-48, 376-383	1-49, 372-383
PDB Codes		3P7E	3QHU	3QNH

^aData for the ligand-free structure are from Marapakala *et al*, 2010 (12)^bValues in parentheses are for the highest resolution shell.^cFrom Molprobit server (<http://molprobit.biochem.duke.edu/>) (32).