Identification of steps in the pathway of arsenosugar biosynthesis

Xi-Mei Xue,¹ Jun Ye,¹ Georg Raber,² Barry P. Rosen,³ Kevin Francesconi,² Chan Xiong,² Zhe Zhu,⁴ Christopher Rensing,⁵ Yong-Guan Zhu ¹,⁶*

¹ Key Laboratory of Urban Environment and Health, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021, China
² Institute of Chemistry, NAWI Graz, University of Graz, Graz 8010, Austria
³ Department of Cellular Biology and Pharmacology, Herbert Wertheim College of Medicine, Florida International University, Miami, FL 33199
⁴ Department of Chemical and Environmental Engineering, Faculty of Science and Engineering, University of Nottingham, Ningbo 315100, China
⁵ College of Resources and Environment, Fujian Agriculture and Forestry University, Fuzhou 350002, China
⁶ State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China

*Address Correspondence to Yong-Guan Zhu,
Address: Institute of Urban Environment, Chinese Academy of Sciences, 1799 Jimei Road, Xiamen 361021, China
Phone number: +86(0)592 6190997

Fax number: +86(0)592 6190977

Email address: ygzhu@iue.ac.cn
Abstract

Arsenosugars are arsenic-containing ribosides that play a substantial role in arsenic biogeochemical cycles. Arsenosugars were identified more than thirty years ago, and yet their mechanism of biosynthesis remains unknown. In this study we report identification of the \textit{arsS} gene from the cyanobacterium \textit{Synechocystis} sp. PCC 6803 and show that it is involved in arsenosugar biosynthesis. In the \textit{Synechocystis} sp. PCC 6803 \textit{ars} operon, \textit{arsS} is adjacent to the \textit{arsM} gene that encodes an As(III) S-adenosylmethionine (SAM) methyltransferase. The gene product, ArsS, contains a characteristic CX$_3$CX$_2$C motif which is typical for the radical SAM superfamily. The function of ArsS was identified from a combination of \textit{arsS} disruption in \textit{Synechocystis} sp. PCC 6803 and heterologous expression of \textit{arsM} and \textit{arsS} in \textit{Escherichia coli}. Both genes are necessary, indicating a multi-step pathway of arsenosugar biosynthesis. In addition, we demonstrate that ArsS orthologs from three other freshwater cyanobacteria and one picocyanobacterium are involved in arsenosugar biosynthesis in those microbes. This study represents the identification of the first two steps in the pathway of arsenosugar biosynthesis. Our discovery expands the catalytic repertoire of the diverse radical SAM enzyme superfamily and provides a basis for studying the biogeochemistry of complex organoarsenicals.
Introduction

Arsenoribosides, commonly known as arsenosugars, were first isolated from the brown kelp, *Ecklonia radiata*, and chemically identified in 1981. More than 20 arsenosugars consisting of dimethylated arsenosugars, dimethylated thio-arsenosugars, and trimethylated arsenosugars are known to occur primarily in marine organisms but also in some terrestrial organisms (Figure S1). They are natural constituents of marine organisms and not toxic to marine plants or animals. Thus the synthesis of arsenosugars represents a mechanism of arsenic detoxification. Arsenosugars also appear to play a central role in the pathway of arsenic biotransformation in marine systems. First, they are likely intermediates in the biosynthesis of the osmolyte arsenobetaine, which is the dominant arsenical in marine organisms and is nontoxic in humans. Edmonds & Francesconi proposed that arsenosugars can be converted to arsenochochine by the cleavage of the C3-C4 bond of the sugar residue and methylation of the arsenic atom. Further oxidation at the C-4 of the sugar residue results in the production of arsenobetaine. The enzymes for glycine betaine synthesis (GbsB/GbsA) have been shown to be involved in the biosynthesis of arsenobetaine from arsenochochine. Second, arsenosugars are also the likely precursors of arsenosugar phospholipids, which may be components of the algal cell membrane. Incorporation of arsenosugars increases under low phosphate conditions and may represent a phosphate-sparing mechanism under those conditions.

Dimethylated arsenosugars consist of a dimethylarsinoyl moiety, in which
pentavalent arsenic is bound to one oxygen, to two methyl groups and to a 5'-deoxyribose (Figure S1). The dimethylated arsenosugars differ from one another only in structure of the side chains at the 5'-deoxyribose C1 position. A proposed pathway for arsenic transformation in the marine environment suggests that SAM is the source of both the methyl groups and the 5'-deoxyribose group of arsenosugars.4

In this study, the arsS (hereafter referred to as SsarsS) and arsM (hereafter referred to as SsarsM) genes from Synechocystis were disrupted to examine their function in arsenosugar biosynthesis. Either disruption mutant was unable to synthesize arsenosugars, indicating that both genes are involved in the biosynthetic pathway. The SsarsM and SsarsS genes were heterologously expressed in E. coli. Cells expressing SsarsM could synthesize dimethylarsenate but not arsenosugar, while cells expressing both genes produced dimethylarsenoriboside derivatives. Our results elucidate the first two steps in the pathway of arsenosugar biosynthesis. The first step is addition of methyl groups from SAM to As(III) by the SAM methyltransferase SsArsM. The second step is addition of the deoxyribose moiety by the radical SAM enzyme SsArsS. This study demonstrates the first two steps in the pathway of arsenosugars and puts the synthesis of complex organoarsenical on a firm biochemical basis.

Materials and Methods

Strains and plasmids

Unless otherwise indicated, E. coli was cultured aerobically in Luria–Bertani (LB) medium at 37°C containing 100 μg mL⁻¹ ampicillin, 50 μg mL⁻¹ kanamycin, or 30 μg mL⁻¹ chloramphenicol, as required. E. coli AW3110 (DE3) (ΔarsRBC) 10 was used for
resistance assays. Plasmids pET-28a-SsarsS, pET-22b-SsarsS, pETDuet-SsarsS or/and pET-28a-SsarsM were transformed into *E. coli* BL21 (DE3) or Rosetta (DE3) for expression of SsArsM and SsArsS and for production of arsenical products. All strains and plasmids are shown in Table S1.

**Disruption of SsarsS and SsarsM in Synechocystis sp. PCC 6803**

A mutant of *Synechocystis* with a disruption of the *SsarsM* gene has been described.\(^{11}\) Mutagenesis of *SsarsS* (*Synechocystis ΔarsS*) was performed (Figure S3A) as follows:

(a) *SsarsS* was amplified with primers SsarsSF and SsarsSR (Table S2); (b) the purified *SsarsS* fragment after PCR was cloned into pMD19T simple vector (TaKaRa, Dalian, China) to yield plasmid pMD-arsS; (c) a kanamycin resistance (*Kan*\(^B\)) cassette cut from pKW1188 \(^{12}\) with EcoRI was inserted into the EcoRV site of pMD-arsS to create gene-disruption *Kan*\(^B\)-insertion cassette, pMDarsSkan; (d) the plasmid was transformed into *Synechocystis* WT according to a previous description;\(^{13}\) (e) *Synechocystis ΔarsS* was obtained after three serial isolation of single colonies on BG11 agar plates containing 30 µg mL\(^{-1}\) kanamycin; (f) PCR with diagnostic primers (Table S2) was used to confirm that the *SsarsS* gene was completely replaced. *Synechocystis ΔarsS* was cultivated in 100 mL of BG11 medium supplemented with 50 µg mL\(^{-1}\) kanamycin at 30\(^o\)C.

**Complementation of SsarsS in Synechocystis sp. PCC 6803**

Complementation of the *Synechocystis ΔarsS* mutant (Synechocystis ΔarsS::arsS) (Figure S3A) was achieved by “knock-in” of intact *SsarsS* containing the rbcL2A promoter with a chloramphenicol resistance (*Cam*\(^B\)) gene. An intact *SsarsS* was amplified using primers arsSc-F and arsSc-R (Table S2). The rbcL2A promoter, P\(_{rbcL2A}\),
was produced with primers FP_CO2noEcoRI and RP_BBRBS. The above PCR 
products were cloned into pMD19T simple vector to produce plasmid pMarsSc and 
pMpromoter respectively, sequenced to confirm the integrity of the cloned flanking 
DNA. The plasmid containing SsarsS with the promoter was constructed by inserting 
PrbcL2A into the EcoRI and SpeI sites of the plasmid pMarsSc. A Cam\(^R\) gene amplified 
from pACYCDeut-1 (Novagen, Madison, WI) using primers cmr-F and cmr-R was 
digested with NheI and HindIII, cloned into the upstream of PrbcL2A. Finally, the 
fragment consisting of a Cam\(^R\) gene, P\(_{rbcL2A}\), and SsarsS was linked with a plasmid 
pKW1188 without Kan\(^R\) by T4 DNA ligase (MBI Fermentas, Flamborough, ON, 
Canada) to prepare plasmid pKcmrparsS. This plasmid was used to transform 
Synechocystis \(\Delta\)arsS, and the above fragment was inserted into the neutral site slr0168. 
In all cases, Synechocystis \(\Delta\)arsS::arsS was selected on BG11 agar plates containing 30 
\(\mu\)g mL\(^{-1}\) kanamycin and 15 \(\mu\)g mL\(^{-1}\) chloramphenicol. PCR amplification with the 
genomic DNA from Synechocystis \(\Delta\)arsS::arsS as templates and the primers gp267-9 
and gp267-10 \(^{15}\) was carried out to confirm that an intact SsarsS was present in 
Synechocystis \(\Delta\)arsS::arsS (Figure S3B). The resulting Synechocystis \(\Delta\)arsS::arsS 
strain was cultivated in 100 mL of BG11 medium with 50 \(\mu\)g mL\(^{-1}\) kanamycin and 30 
\(\mu\)g mL\(^{-1}\) chloramphenicol. 

**As(III) transformation assays**

After one or two weeks of continuous illumination culture, 50 or 100 mL of 
cyanobacteria (Synechocystis WT, Synechocystis \(\Delta\)arsS, and Synechocystis \(\Delta\)arsS::arsS) 
were collected by centrifuging and washing twice with cold MES buffer. \(^{16}\) About 20
mg of lyophilized cells were added into 1.5 mL Eppendorf tubes with 1 mL of a mixture of CH$_3$OH(MeOH)/H$_2$O (1+1, v/v). The cells were extracted by placing the tubes in an ultrasonic bath for 20 minutes followed by rotating on a rotary wheel overnight. The extracts were centrifuged for 10 min at 13520 g. The supernatant was purged with N$_2$ to remove MeOH, and then filtered through a 0.22 µm membrane filter into 1 mL crimp/snap polypropylene vials (Agilent Technologies, Palo Alto, CA, USA) for arsenic species analysis.

The transformation of As(III) under the control of ArsS or/and ArsM was assayed 
_in vivo._ The assays were performed with Rosetta bearing pET-28a-SsarsM, Rosetta bearing pETDuet-SsarsS, Rosetta bearing empty pETDuet-1 and pET-28a, or Rosetta bearing pET-28a-SsarsM and pETDuet-SsarsS. Cells were grown overnight at 37°C with shaking (200 rpm) in LB medium containing corresponding antibiotic, diluted 100-fold into 5 mL of fresh LB medium. After incubation at an OD$_{600\text{ nm}}$ of 0.5 was carried out with 0.5 mM IPTG for 3 h, cells were treated with 1 µM, 10 µM, 100 µM, 1 mM or 3 µM of either As(III) or As(V). At the indicated times, the cells and culture medium were collected.

**Arsenic species analysis**

High performance liquid chromatography - inductively coupled plasma mass spectrometry (HPLC-ICP-MS) measurements on arsenic species were performed using an Agilent 7500cx ICP-MS (Agilent Technologies, California, USA) for element detection coupled with an Agilent 1200 HPLC system. Anion-exchange chromatography separation was performed using PRP-X100 with a guard column.
(Hamilton, Reno, NV, USA). A Shiseido CAPCELL PAK C18 MGII (Shiseido, Tokyo) with a matching guard column was employed for ion-pair reversed-phase chromatographic separation. All chromatography parameters are shown in Table S3.

Unknown arsenic compounds were identified using HPLC-ICP-MS/electrospray ionization tandem mass spectrometry (ESI-MS-MS, Agilent 6460, Agilent Technologies, Waldbronn, Germany), and higher solution accurate mass (HPLC-ESI-HR-MS-MS) measurements which were performed in positive mode on a Q-Exactive Hybrid Quadrupole-Orbitrap MS after arsenic species were separated on a Dionex Ultimate 3000 series instrument (Thermo Fischer Scientific, Erlangen, Germany).

**Results**

*The arsS gene is involved in arsenosugar biosynthesis.*

Our search for candidate genes involved in arsenosugar biosynthesis was guided by the hypothesis that oxidation of DMAs(III) by addition of the deoxyadenosyl radical group from SAM and enzymatic hydrolytic removal of adenine would result in the formation of arsenosugars in marine algae.⁴ *Synechocystis* produces two species of arsenosugars, indicating that this microbe has the biosynthetic enzymes for arsenosugar synthesis.¹¹ ¹⁷ We previously identified the enzyme SsArsM as an As(III) S-adenosylmethionine methyltransferase in *Synechocystis*.¹⁶ The *SsarsM* gene is adjacent to two other open reading frames, *slr0304* and *slr0305* (Figure 1). A protein BLAST search using SLR0304, which we termed as SsArsS, as a query revealed the presence of regions conserved in the radical SAM superfamily, in particular iron-sulfur cluster (FeS) and SAM binding sites. The *SarsM* and *SsarsS* genes comprise an *ars* operon (Figure S2).
SLR0305 is related to the family of SNARE-associated Golgi membrane proteins, and this gene is co-transcribed with the other two genes, but it is not clear if it has an arsenic-related function. *Nostoc* sp. PCC 7120, which also produces arsenosugars, has a putative *arsMarsS* operon. The *Nostoc* operon lacks a gene corresponding to SLR0305, indicating that the latter does not have an arsenic-related function. We hypothesize from the linkage of the *arsM* and *arsS* genes in more than one arsenosugar producers that these two genes have related functions encoding a biosynthetic pathway for arsenosugar biosynthesis (Figure 1). We previously demonstrated that a strain in which *SsarsM* was disrupted cannot synthesize arsenosugars, supporting our hypothesis that DMAs production by ArsM is the first step in arsenosugar synthesis.11

We hypothesize that ArsS catalyzes the second step in arsenosugar biosynthesis by transfer of a deoxyadenosyl group from SAM to DMAs. This could be the first committed step in the biosynthetic pathway because the product of ArsM, DMAs can have other biosynthetic fates. ArsS is a member of the radical SAM superfamily. Radical SAM enzymes are characterized by a CX$_3$CX$_2$C motif that provides coordinations for binding of a unique four iron-four sulfur ([4Fe-4S]) cluster, although the spacing of the cysteines in this motif can vary.19 Both the alignment of SsArsS with those of functionally characterized radical SAM enzymes and construction of a molecular phylogenetic tree indicate that SsArsS is closely related to MoaA (Figure S4), which harbors an N-terminal [4Fe-4S] cluster involved in a 5’-deoxyadenosyl radical generation and a C-terminal [4Fe-4S] cluster involved in substrate binding and/or activation in conversion of 5’-GTP to an oxygen-sensitive tetrahydropyranopterin.20
We have used UV/visible spectroscopy to examine a visible absorption band at 410 nm for the [4Fe-4S]^{2+} protein (Figure S5).^{21}

**SsArsS is involved in the biosynthesis of arsenosugars**

To examine the involvement of ArsS in arsenosugar biosynthesis, we disrupted the *SsarsS* gene of *Synechocystis* by inserting a Kan^R^ cassette inside *SsarsS*. We evaluated the ability of the parental wild type strains and its derivatives to grow. No difference in growth rate in the absence of As(III) was observed among wild type, *Synechocystis ΔarsS* and *Synechocystis ΔarsS::arsS* strains (Figure S3C), indicating that ArsS is not required for viability under these growth conditions. The production of arsenosugars by wild type and mutant strains in medium containing different arsenic species was determined by HPLC-ICP-MS analysis. Consistent with earlier studies,^{11, 17} wild type *Synechocystis* exposed to As(III) produces As(V) as the predominant soluble arsenic species, and small amounts of MAs(V), DMAs(V) and arsenosugars (Figure 2a). Again, it is not possible to deduce whether the products of the reaction were trivalent species that oxidized in air to their pentavalent forms. In addition, wild type *Synechocystis* transforms MAs(V) or DMAs(V) into arsenosugars (Figure 2c, e). *Synechocystis ΔarsS* was unable to generate arsenosugars in medium containing either As(III) or pentavalent methylated arsenic species (Figure 2b, d, f). The *Synechocystis ΔarsS::arsS* strain regained the ability to produce arsenosugars (Figure S3D). These results clearly showed that SsArsS is likely involved in arsenosugar biosynthesis.

**Heterologous expression of SsArsS and SsArsM**

To demonstrate the linked function of *SsArsM* and *SsArsS* in arsenosugar production,
the *SsarsS* and *SsarsM* genes were heterologously expressed alone or in combination in *E. coli* Rosetta (DE3), which has neither in its genome (Figure 3A and Figure S6A). Heterologous expression of *arsM* genes in *E. coli* strain AW3110 (DE3), in which the chromosomal *ars* operon was deleted, has been shown to confer As(III) resistance. Similarly, expression of *SsarsM* alone in *E. coli* strain AW3110 (DE3) conferred resistance to As(III) (Figure S7). Surprisingly, expression of *SsarsS* alone conferred As(III) resistance (Figure S7). *SsArsS* has 11 cysteine residues that might bind sufficient As(III) to confer moderate resistance or some free radicals generated by the process of As(III) oxidation to As(V) by *SsArsS* when the gene is expressed at high levels.

ICP-MS was also used to quantify total arsenic in *E. coli* Rosetta cells (Figure S8A) treated with 0.5 mM IPTG and 1 μM As(III) for 20 h. Rosetta cells expressing *SsarsS* accumulated the most arsenic, up to 83 mg kg\(^{-1}\). Cells expressing *SsarsM* alone accumulated less (44 mg As kg\(^{-1}\)), which was similar to cells expressing both genes (39 mg As kg\(^{-1}\)). The arsenic content of cells expressing both *SsarsM* and *SsarsS* peaked after exposure for 3 h, and subsequently declined, suggesting that they expelled their arsenic compounds from the cells (Figure S8B).

The arsenic species in medium were analyzed by HPLC-ICP-MS after *E. coli* Rosetta cells were exposed to 3 μM As(III) for 12 h. No arenosugars were identified in Rosetta expressing both *SsarsM* and *SsarsS*. The major arsenic species found in medium of Rosetta cells expressing both *SsarsM* and *SsarsS* or *SsarsM* alone was DMAs(V) (Figure 3). As(III) dominated in medium of Rosetta expressing *SsarsS* alone.
Comparative analysis of the aqueous fraction of Rosetta cells expressing \textit{SsarsM} alone, or \textit{SsarsS} alone, or both \textit{SsarsM} and \textit{SsarsS} revealed that Rosetta cells expressing both \textit{SsarsM} and \textit{SsarsS} generated unknown arsenic compounds (Figure 3B), indicating that biosynthesis of these species requires the concerted action of ArsM and ArsS. These unknown compounds were identified by HPLC-ESI-HR-MS-MS after Rosetta cells expressing both \textit{SsarsM} and \textit{SsarsS} were incubated with 3 \(\mu\)M arsenic for 72 h, and they were a group of dimethylarsinoyl-hydroxycarboxylic acids (DMA-HA, Figure S9 and S10).

None of the unknown arsenic compounds that cells expressing \textit{SsarsM} and \textit{SsarsS} together produced eluted at the same position as 5'-deoxy-5'-dimethylarsinoyl-adenosine, which is a predicted intermediate in the arsenosugar synthesis pathway proposed by Edmonds and Francesconi \(^4\) (Table 1). The results of HPLC-ICP-MS/ESI-MS-MS and HPLC-ESI-HR-MS-MS measurements suggested the presence of additional dimethylarsinoyl-hydroxycarboxylic acids such as 3'-dimethylarsinoyl-2'-hydroxypropionic acid, 4'-dimethylarsinoyl-2', 3'-dihydroxybutyric acid, 5'-dimethylarsinoyl-2'-hydroxypentanonic acid and 5'-dimethylarsinoyl-2', 3', 4'-trihydroxypentanonic acid, as well as two unidentified arsenic compounds which could not be characterized due to low concentrations (Table 1 and Figure S10). A compound with the predicted composition \(C_7H_{16}O_5\text{As}\) (m/z 255.02079) was also detected by HPLC-ESI-HR-MS-MS, but the structure could not be confirmed (Figure S10). Figure S10 shows that in addition to the four dimethylarsinoyl-hydroxycarboxylic acids (except 5-dimethylarsinoyl-2'-hydroxypentanonic acid) identified in incubation
medium, cells expressing *SsarsM* and *SsarsS* also contained the four corresponding thio-arsenic compounds. There are several possible reasons why dimethylarsinoyl-hydroxycarboxylic acids were produced rather than 5′-deoxy-5′-dimethylarsinoyl-adenosine in *E. coli* expressing both *SsarsM* and *SsarsS*. First, there might be additional steps catalyzed by as-yet unidentified *Synechocystis* enzymes not present in *E. coli*. Second, most of the ArsS produced in *E. coli* may be partly active, consistent with our observation that purified ArsS is extremely oxygen sensitive. Oxygen induces the switch from [4Fe-4S]$^{+}$ to [4Fe-4S]$^{2+}$, [3Fe-4S]$^{2+}$, or [2Fe-2S]$^{2+}$ with loss of biological activity. A future goal is to express the genes under anaerobic growth conditions.

**The roles of cysteine residues in *SsarsS***

Assuming that the production of dimethylarsinoyl-hydroxycarboxylic acids reflects a partial activity of SsArsS, the role of cysteine residues in that activity was examined. Thiol chemistry plays an important role in arsenic metabolism and transport processes through formation of metalloid-sulfur bonds between arsenic and related proteins. Each of the eleven cysteine residues in SsArsS protein, Cys$^{41}$, Cys$^{45}$, Cys$^{48}$, Cys$^{65}$, Cys$^{129}$, Cys$^{142}$, Cys$^{264}$, Cys$^{279}$, Cys$^{320}$, Cys$^{323}$, Cys$^{331}$, were individually changed to serine residues. Single mutants in *SsarsS* and *SsarsM* were co-transformed into *E. coli* strain Rosetta (DE3) which were treated with As(III) (Figure S11). Three mutant proteins, C65S, C129S, C142S, were able to produce unknown organoarsenicals. Cells expressing the other eight mutants and *SsarsM* lost the ability to produce the unidentified arsenic compounds. The three cysteine residues, Cys$^{41}$, Cys$^{45}$, Cys$^{48}$, in the N-terminal region of SsArsS correspond to the canonical radical SAM cysteine motif.
CX_{3}CX_{2}C, while other three cysteine residues, Cys^{320}, Cys^{323}, Cys^{331}, in the C-terminal region of SsArsS are conserved among ArsSs.

The function of microbial arsS genes in arsenic transformation.

We identified putative ArsS orthologs with eight conserved cysteine residues and a radical SAM domain in other microbes. Genes encoding for ArsS from four other cyanobacteria and one eukaryotic green alga (Table S4) were co-expressed with SsarsM in E. coli. The cyanobacteria Nostoc sp. PCC 7120^{18, 27} and the eukaryotic alga Chlorella^{28} have been reported to produce arsenosugars. Recently, a putative ArsM was identified in the picocyanobacterium P. marinus str. MIT 9313^{29} although it has not been demonstrated if this species generates arsenosugars. Cyanobacteria A. platensis NIES-39 and S. elongatus PCC 6301, which have genes encoding ArsS orthologs were chosen as representative cyanobacteria. Arsenic transformation was analyzed in cells of E. coli co-expressing genes for orthologs of SsArsS in pETDuet-arsS together with pET-28a-SsarsM. E. coli co-expressing SsarsM and NsarsS from Nostoc sp. PCC 7120 produced the dimethylarsinoyl-hydroxycarboxylic acids observed in E. coli expressing SsarsM and SsarsS (Figure 4 and Figure S6B). ArsSs from cyanobacteria A. platensis NIES-39, S. elongatus PCC 6301, and P. marinus str. MIT 9313, organisms that have not been previously reported to generate arsenosugars, also synthesized dimethylarsinoyl-hydroxycarboxylic acids (Figure 4). Only SsArsM and CvArsS from the eukaryotic algae C. variabilis did not synthesize 5’-dimethylarsenoriboside derivatives when co-expressed in E. coli. The reason for this negative result is unknown, but perhaps CvArsS, which undergoes extensive post-
translational processing in *C. variabilis*, is not modified to an active enzyme in *E. coli*.
or perhaps those gene products did not fold properly in *E. coli*.

**Discussion**

The most abundant arsenic compounds identified in marine algae are arsenosugars. Prokaryotic cyanobacteria, eukaryotic green algae and fungi all have been reported to produce arsenosugars. Arsenosugars have also been found in underwater plants as well; their source, however, is likely to be epiphytic algae or bacteria because plants cannot methylate arsenic and dimethylarsenic has been postulated to be the immediate precursor of arsenosugars. Moreover, the physiological role of arsenosugars is still unclear. For the oxo-arsenosugars, it’s known to be non-toxic to humans, one arsenosugar (Oxo-Gly) was shown to be nontoxic to murine peritoneal macrophages and alveolar macrophages at micromolar levels. The toxicity of thio-arsenosugars whose bioaccessibility are much higher than the oxo-analogs, and trimethylated arsenosugars haven’t been proved yet. Arsenosugar biosynthesis may serve both as a detoxification process for the organisms which generate them, and as a precursor for more complex organoarsenicals. Unraveling the mechanism of arsenosugar biosynthesis will help determine toxicity of arsenosugars to human and the physiological role of arsenosugars in arsenic biogeochemical cycling.

Until our study, the pathway of arsenosugar biosynthesis was unknown. Edmonds and Francesconi proposed a pathway in marine alga in which DMAs(V) is reduced to DMAs(III), which is then oxidized by addition of the deoxyadenosyl group from SAM to yield the intermediate 5'-deoxy-5'-dimethylarsinoyl-adenosine. Enzymatic,
hydrolytic removal of adenine would be followed by formation of arenosugars. This proposal is supported by the identification of the proposed intermediate 5'-deoxy-5'-dimethylarsinoethyl-adenosine in the kidney of the giant clam *Tridacna maxima*, which contains large quantities of arenosugars owing to symbiotic algae growing in the mantle of the clam. However, the product of arsenic methylation by ArsM enzymes is almost certainly DMAs(III) and not DMAs(V), so reduction of DMAs(V) is not required. The ArsM reaction is required for producing the substrate of ArsS, but DMAs has other fates such as further methylation to TMAs, so it is not a committed step in the reaction pathway. Moreover, we show that co-expression of ArsM and ArsS in *E. coli* produced many dimethylarsinoyl-hydroxycarboxylic acids but not 5'-deoxy-5'-dimethylarsinoyl-adenosine. We are considering two possible pathways for arenosugar biosynthesis to account for the compounds produced by *E. coli*. Regardless of the actual pathway, our data strongly support our hypothesis that the ArsS reaction catalyzes the first committed step in arenosugar biosynthesis.
The likely reason that arsenosugars are not generated by *E. coli* co-expressing *SsarsS* and *SsarsM* is that the intermediate of arsenosugar biosynthesis cannot be further transformed into arsenosugars because *E. coli* does not have the genes involved in arsenosugar biosynthesis. An alternative fate for this intermediate, however, could be that it is rapidly degraded *in vivo* to small arsenic-containing compounds such as those that we found produced by the transgenic *E. coli* constructed for this study.

It is notable that the *Synechocystis* *ars* operon containing *arsM* and *arsS* is not induced by As(III) or As(V), and the operon does not contain an *arsR* gene for an As(III)-responsive transcription factor. Consistent with our observation of constitutive expression of the *arsM* and *arsS* genes, the amount of arsenosugars and
arsenosugar phospholipids in *Synechocystis* treated with different concentrations of As(V) did not significantly increase with increasing arsenic concentrations. The greater accumulation in *Synechocystis* of arsenosugar phospholipid compared with arsenosugars suggests that arsenosugars are the precursors of arsenosugar phospholipids and not the end products. The absence of arsenosugar intermediates in *Synechocystis* suggests that DMAs is rapidly transformed into arsenosugars in those cells, and that the arsenosugar intermediates, DMAs(III) and MAs(III), are probably transient intermediates that do not accumulate in the cells.

Thus, we hypothesize that SsArsS, a putative radical SAM enzyme, adds a 5'-deoxyadenosyl radical (dAdo·) to DMAs(III) and hydrolyzes the adenine from 5'-deoxyadenosine simultaneously to generate 5'-deoxy-5'-dimethylarsinoyl-ribose. SsArsS has a CX$_3$CX$_2$C motif similar to the assembly site for the [4Fe-4S] cluster found in radical SAM enzymes. In addition, SsArsS has the other three cysteine residues in the C-terminal region of SsArsS, Cys320, Cys323, Cys331, a CX$_2$CX$_7$C motif that binds another putative [4Fe-4S] cluster (Figure S12), which would put it in a subgroup of radical SAM enzymes that bind two [4Fe-4S] clusters, including MoaA, BioB, FbiC, LipA, TYW1. However, this CX$_2$CX$_7$C motif in the C-terminal region of the SsArsS protein, which is not conserved in all radical SAM enzymes, is 5'-deoxy-5'-not similar to other proteins that contained two or more iron-sulfur clusters. Mutations in any of the cysteine residues in the CX$_2$CX$_7$C motif resulted in the loss of the function for arsenosugar synthesis, consistent with a role for that motif in ArsS function. Moreover, none of the cysteine mutants in either motif produced 5'-deoxy-5'-
dimethylarsinoyl-adenosine (Figure S11), indicating that the 5'-deoxy-5'-
dimethylarsinoyl-riboside derivative was formed before 5'-deoxy-5'dimethylarsinoyl-
adenosine.

In summary, ArsS is the only member of the radical SAM superfamily shown to
be involved in arsenosugar synthesis. This study adds a new function to the repertoire
of enzymes of organoarsenical biotransformations and the radical SAM superfamily.
Further biochemical analysis of complex arsenic species and their molecular
mechanisms of biosynthesis and degradation will enable a better understanding for the
ecological roles of these complex organoarsenicals.

Acknowledgments

Our research is supported by the National Natural Science foundation of China
(41877422 and 21507125), Crossing-Group Project of KLUEH (KLUEH-201802), the
Austrian Science Fund (FWF) project number I2412-B21, and National Institutes of
Health Grants R01 GM55425 and ES023779 to bpr.

References

(1) Edmonds, J. S.; Francesconi, K. A. Arseno-Sugars from Brown Kelp (Ecklonia Radiata) as
(2) Nischwitz, V.; Pergantis, S. A. Mapping of Arsenic Species and Identification of a Novel
Arsenosugar in Giant Clams Tridacna Maxima and Tridacna Derasa Using Advanced Mass
(3) Duncan, E. G.; Maher, W. A.; Foster, S. D. Contribution of Arsenic Species in Unicellular Algae
(5) Hoffmann, T.; Warmbold, B.; Smits, S. H. J.; Tschapek, B.; Ronzheimer, S.; Bashir, A.; Chen,
C.; Rolbetzki, A.; Pittelkow, M.; Jebbar, M.; Seubert, A.; Schmitt, L.; Bremer, E. Arsenobetaine:


ACS Paragon Plus Environment


(38) Marapakala, K.; Packianathan, C.; Ajees, A. A.; Dheeman, D. S.; Sankaran, B.; Kandavelu, P.;


The TOC was created by Dr Xi-Mei Xue
TOC artwork
Table 1 Arsenic species in *E. coli* cells expressing *SsarsM* and *SsarsS* and medium were determined by HPLC-ICP-MS/HR-ES-MS analysis using PRP-X100 when *E. coli* cells expressing *SsarsM* and *SsarsS* were induced with 0.5 mM IPTG and 3 μM As(III) or As(V).

<table>
<thead>
<tr>
<th></th>
<th>m/z</th>
<th>Formula</th>
<th>Delta m [ppm]</th>
<th>Molecular structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>138.97346</td>
<td>C&lt;sub&gt;2&lt;/sub&gt;H&lt;sub&gt;6&lt;/sub&gt;O&lt;sub&gt;2&lt;/sub&gt;As</td>
<td>0.02</td>
<td><img src="image" alt="Structure A" /></td>
</tr>
<tr>
<td>B</td>
<td>210.99459</td>
<td>C&lt;sub&gt;5&lt;/sub&gt;H&lt;sub&gt;12&lt;/sub&gt;O&lt;sub&gt;4&lt;/sub&gt;As</td>
<td>-0.41</td>
<td><img src="image" alt="Structure B" /></td>
</tr>
<tr>
<td>C</td>
<td>239.02589</td>
<td>C&lt;sub&gt;7&lt;/sub&gt;H&lt;sub&gt;16&lt;/sub&gt;O&lt;sub&gt;4&lt;/sub&gt;As</td>
<td>-0.99</td>
<td><img src="image" alt="Structure C" /></td>
</tr>
<tr>
<td>D</td>
<td>241.00514</td>
<td>C&lt;sub&gt;6&lt;/sub&gt;H&lt;sub&gt;14&lt;/sub&gt;O&lt;sub&gt;5&lt;/sub&gt;As</td>
<td>-1.00</td>
<td><img src="image" alt="Structure D" /></td>
</tr>
<tr>
<td>E</td>
<td>271.0157</td>
<td>C&lt;sub&gt;7&lt;/sub&gt;H&lt;sub&gt;16&lt;/sub&gt;O&lt;sub&gt;6&lt;/sub&gt;As</td>
<td>-0.87</td>
<td><img src="image" alt="Structure E" /></td>
</tr>
</tbody>
</table>
Figure 1. Cyanobacterial *ars* operons. Shown are the putative operons of *Synechocystis* sp. PCC 6803 and *Nostoc* sp. PCC 7120. The *arsM* (*slr0303, alr3095*) genes encode As(III) S-adenosylmethionine methyltransferases. *Slr0305* from *Synechocystis* sp. PCC 6803 belongs to SNARE-associated superfamily, *alr3096* of *Nostoc* sp. PCC 7120 encodes PitA (the low affinity inorganic phosphate transporter), *arsS* (*slr0304, alr3097*) encodes a radical SAM superfamily enzyme shown in this report to be involved in arsenosugar biosynthesis.

Figure 2. Anion-exchange HPLC-ICP-MS chromatograms of arsenic species from (a) wild type *Synechocystis* and (b) *Synechocystis* Δ*arsS* after exposure to 1 µM As(III) for two weeks, (c) wild type *Synechocystis* and (d) *Synechocystis* Δ*arsS* incubated with 10 µM MAs(V) for two weeks, (e) wild type *Synechocystis* and (f) *Synechocystis* Δ*arsS* treated with 100 µM DMA(V) for one week.

Figure 3 A. Western blot analysis of SsArsM and/or SsArsS expression in *E. coli* strain Rosetta bearing the indicated expression plasmid constructs. MSR, MR, SR, or CK containing pETDuet-*SsarsS* and pET-28a-*SsarsM*, pET-28a-*SsarsM*, pETDuet-*SsarsS*, or pETDuet and pET-28a, respectively, induced by IPTG. B. Typical chromatograms of arsenic species in *E. coli* strain Rosetta medium incubated with 3 µM As(III) by ion-pairing reverse-phase column HPLC-ICP-MS. DMA-HA indicates the dimethylarsinoyl-hydroxycarboxylic acids.

Figure 4. Arsenic species transformed by different combination of ArsS orthologs and SsArsM. A. Western blot analysis of SsArsM and different ArsS orthologs expression in Rosetta bearing the indicated expression plasmid constructs. MNsS, MSeS, MPmS,
MApS, MCvS indicate *E. coli* strain Rosetta cells containing pET-28a-*SsarsM* and pETDuet-*NsarsS*, pETDuet-*SearsS*, pETDuet-*PmarsS*, pETDuet-*AparsS*, or pETDuet-*CvarsS* induced by IPTG respectively. B. Typical chromatograms of arsenic analysis in *E. coli* strain Rosetta medium treated with As(III) by ion-pairing reversed-phase column HPLC-ICP-MS of SsArsM and different ArsS orthologs expressed in *E. coli* strain Rosetta. DMA-HA indicates the dimethylarsinoyl-hydroxycarboxylic acids. The mix consists of As(III), MAs(V), DMAs(V), As(V), and TMAO.
**Associated content**

**Supporting Information:** Supporting methods, Reagents and solutions (pS5), Confirmation of \textit{arsMarsS} operon (pS5), Phylogenetic analyses of SsArsS (pS6), Comparative growth curves for WT, disruption strains and complementary strains (pS6), Cloning arsM and arsS from cyanobacteria (pS7), Resistance assays (pS8), Nucleotide sequence synthesis of ArsS orthologs from other sequenced organisms (pS8), Construction of SsarsS cysteine mutants (pS9), Immunological detection of ArsS and ArsM (pS10), The structures of several common arsenic compounds found in nature (Figure S1), bacterial strains and plasmids (Table S1), Primer sequences used for PCR amplification in the disruption, complementation, gene expression, site-mutant, and \textit{ars} operon identification strategies (Table S2), ICP-MS and chromatographic conditions used for arsenic species analysis (Table S3), Confirmation of \textit{arsMarsS} operon (Figure S2), Gene deletion and complementation (Figure S3), Phylogram of SsArsS and representative enzymes of the radical SAM superfamily with different functions (Figure S4), The UV/visible spectroscopy of SsArsS shows a shoulder centered at 410 nm for the [4Fe-4S]\(^{2+}\) protein (Figure S5), SDS-PAGE analysis (12%, Coomassie blue-stained) of SsArsM and/or ArsS expression in Rosetta bearing the indicated plasmid (Figure S6), SsArsS or /and SsArsM confers resistance to As(III) in \textit{E. coli} (Figure S7), Arsenic concentrations of different \textit{E. coli} Rosetta cells or Rosetta cells expressing \textit{SsarsM} and \textit{SsarsS} (Figure S8), ICP-MS and HR-ESMS measurements of unknown arsenic species in \textit{E. coli} medium (Figure S9), HPLC-ICP-MS (A) and HR-ESMS measurements (B) of unknown arsenic species in \textit{E. coli} cells (Figure S10), Arsenic species transformed by different combination of \textit{SsarsS} mutants and \textit{SsarsM}
(Figure S11), The information of ArsS orthologs studied in this study (Table S4),

Multiple alignment of SsArsS and other five ArsS orthologs (Figure S12).
Cyanobacterial ars operons. Shown are the putative operons of Synechocystis sp. PCC 6803 and Nostoc sp. PCC 7120. The arsM (slr0303, alr3095) genes encode As(III) S-adenosylmethionine methyltransferases. Slr0305 from Synechocystis sp. PCC 6803 belongs to SNARE-associated superfamily, alr3096 of Nostoc sp. PCC 7120 encodes PitA (the low affinity inorganic phosphate transporter), arsS (slr0304, alr3097) encodes a radical SAM superfamily enzyme shown in this report to be involved in arsenosugar biosynthesis.
Anion-exchange HPLC-ICP-MS chromatograms of arsenic species from (a) wild type Synechocystis and (b) Synechocystis ΔarsS after exposure to 1 µM As(III) for two weeks, (c) wild type Synechocystis and (d) Synechocystis ΔarsS incubated with 10 µM MAAs(V) for two weeks, (e) wild type Synechocystis and (f) Synechocystis ΔarsS treated with 100 µM DMAAs(V) for one week.
A. Western blot analysis of SsArsM and/or SsArsS expression in E. coli strain Rosetta bearing the indicated expression plasmid constructs. MSR, MR, SR, or CK containing pETDuet-SsarsS and pET-28a-SsarsM, pET-28a-SsarsM, pETDuet-SsarsS, or pETDuet and pET-28a, respectively, induced by IPTG. B. Typical chromatograms of arsenic species in E. coli strain Rosetta medium incubated with 3 μM As(III) by ion-pairing reverse-phase column HPLC-ICP-MS. DMA-HA indicates the dimethylarsinoyl-hydroxycarboxylic acids.
Arsenic species transformed by different combination of ArsS orthologs and SsArsM. A. Western blot analysis of SsArsM and different ArsS orthologs expression in Rosetta bearing the indicated expression plasmid constructs. MNsS, MSeS, MPmS, MAPS, MCvS indicate E. coli strain Rosetta cells containing pET-28a-SsarsM and pETDuet-NsarsS, pETDuet-SearsS, pETDuet-PmarsS, pETDuet-AparsS, or pETDuet-CvarsS induced by IPTG respectively. B. Typical chromatograms of arsenic analysis in E. coli strain Rosetta medium treated with As(III) by ion-pairing reversed-phase column HPLC-ICP-MS of SsArsM and different ArsS orthologs expressed in E. coli strain Rosetta. DMA-HA indicates the dimethylarsinoyl-hydroxycarboxylic acids. The mix consists of As(III), MAs(V), DMAs(V), As(V), and TMAO.