Microbial arsenic: from geocycles to genes and enzymes

Rita Mukhopadhyay a, Barry P. Rosen a, Le T. Phung b, Simon Silver b,*

a Department of Biochemistry and Molecular Biology, Wayne State University School of Medicine, Detroit, MI 48201, USA
b Department of Microbiology and Immunology, University of Illinois, Chicago, IL 60612-7344, USA

Received 13 March 2002; received in revised form 30 May 2002; accepted 31 May 2002
First published online 18 July 2002

Abstract

Arsenic compounds have been abundant at near toxic levels in the environment since the origin of life. In response, microbes have evolved mechanisms for arsenic resistance and enzymes that oxidize As(III) to As(V) or reduce As(V) to As(III). Formation and degradation of organoarsenicals, for example methylarsenic compounds, occur. There is a global arsenic geocycle, where microbial metabolism and mobilization (or immobilization) are important processes. Recent progress in studies of the ars operon (conferring resistance to As(III) and As(V)) in many bacterial types (and related systems in Archaea and yeast) and new understanding of arsenite oxidation and arsenate reduction by respiratory-chain-linked enzyme complexes has been substantial. The DNA sequencing and protein crystal structures have established the convergent evolution of three classes of arsenate reductases (that is classes of arsenate reductases are not of common evolutionary origin). Proposed reaction mechanisms in each case involve three cysteine thiols and S-As bond intermediates, so convergent evolution to similar mechanisms has taken place. © 2002 Federation of European Microbiological Societies. Published by Elsevier Science B.V. All rights reserved.

Keywords: Arsenic resistance; Arsenite oxidase; Arsenate reductase; Plasmid resistance; Geocycle; Methylation

1. Introduction: arsenic resistance in natural and unnatural environments

1.1. The arsenic global geocycle

1.2. Methylation of arsenic

2. Genes for arsenic resistance

2.1. Varieties of genes and operons

3. Arsenate reductases: convergent evolution to similar chemistry

3.1. The E. coli GSH/Grx ArsC family

3.2. The Staphylococcus Trx ArsC family

3.3. The Saccharomyces family, a third clade

4. Arsenite oxidation, microbial and enzymatic

Acknowledgements

References
senic toxicity, (c) release of arsenic from burning coal into foods in Southwest China, and (d) within the USA [3], where arsenic in drinking water from residential wells in Michigan and Wisconsin [4,5], as well as industrially impacted mining waters in the western USA, and recreational waters north of Boston [6] are far above WHO recommended levels. Useful URLs for information about global arsenic pollution problems at the time of writing include http://www.who.int/inf-fs/en/fact210.html (WHO); http://co.water.usgs.gov/trace/arsenic/ (USA); and http://bicn.com/acic/ (Bangladesh). The widespread occurrence of high arsenic in water caused the US President to state “Arsenic is a natural substance that sometimes causes problems” and to reverse the previous government’s decision to accept five times lower WHO standards [7].

The ultimate source of arsenic on the Earth’s surface is igneous activity [8]. Arsenic is widely spread in the upper crust of the Earth, although mainly at very low concentrations, with arsenic concentrations in soil ranging from 0.1 to more than 1000 ppm (mg kg$^{-1}$). In atmospheric dust, the range is 50–400 ppm. In seawater, the average arsenic level may be 2.6 ppb (µg L$^{-1}$) and in fresh water about 0.4 ppb. Arsenic at significant levels is all around us (Fig. 1).

Acidic mining environments are of particular concern with regard to arsenic pollution and microbial activities result in very low pH (sometimes approaching pH 1), since bacterial energy metabolism and growth oxidizes reduced sulfur to H$_2$SO$_4$. Arsenic is oxidized in parallel to AsO$_3^{3-}$. Both bacteria and Archaea are found in these environments and are involved in arsenic transformations [9–11]. Arsenic compounds occur naturally at significant levels especially in marine foodstuffs, for example prawns contain levels approaching 200 ppm [8]. In addition, arsenical compounds have been used widely in medicine (especially for protozoan diseases), in agriculture (as herbicides and animal feed additives), and in crime (as a poison) [12–14]. Arsenic is itself a carcinogen ([15,16], http://www.who.int/inf-fs/en/fact210.html). Indeed, Paul Ehrlich won the Nobel Prize in medicine in 1908 for use of arsenicals as chemotherapeutic agents, culminating with the famous arsenical compound ‘Salvarsan’, which was the best of ‘early modern’ antimicrobial agents. In North America and elsewhere in the developed world, organoarsenicals are added at levels of 20–40 g ton$^{-1}$ to feed for boiler chickens [17]. It is thought that the aromatic organoarsenicals Roxarsone and p-arsanilic acid enhance chicken growth, perhaps by limiting diseases such as coccidiosis. Considering the number of chickens consumed annually in North America and Europe, the amount of arsenic used and then released to the environment as chicken effluent is substantial [17]. The microbial degradation of organoarsenicals is known [18], but barely studied.

![Fig. 1. The global arsenic geocycle (see text).](image-url)
1.1. The arsenic global geocycle

Just as there are well-studied geocycles for carbon, nitrogen, oxygen, sulfur and other elements that are components of all living cells, there are also geocycles for toxic elements including arsenic (Fig. 1). Living cells (especially microbes) carry out redox and covalent bond chemistry and are important players in the arsenic geocycle. Higher plants and animals bio-accumulate compounds to levels far above those of the environments in which they live.

The major sources of human contamination and occupational exposure with arsenic are the burning of coal and industrial metal smelting (Fig. 1), and more recently the semiconductor industry, as well as release from arsenic-rich ores during mining. Bio-mining releasing soluble arsenic from rock produces local environments of high and toxic arsenic levels.

This arsenic geocycle is presented here for the first time (a more limited version was in Cervantes et al. [19]) and is patterned after a similar geocycle for mercury that we first published over 25 years ago. That geocycle figure has since been used in reviews and textbooks, usually without attribution. A similar experience is expected with greater understanding of the arsenic geocycle. To paraphrase and expand on the US President George W. Bush: arsenic occurs so broadly in nature at subtoxic and near-toxic levels that human exposure from perturbation of natural arsenic should always be of concern [7].

Arsenate (the main arsenic compound in seawater) is taken up by marine organisms, ranging from phytoplankton, algae, crustaceans, mollusks and fish [20,21], and converted to small organic compounds (such as methylarsonic acid or dimethylarsinic acid; Figs. 1 and 2), or is converted to organic storage forms that are then secreted into the environment. However, some arsenic is retained by phytoplankton and metabolized into complex organic compounds. The transformation of inorganic arsenic into lipid-soluble compounds might be an adaptive mechanism for marine phytoplankton to compensate for limited nitrate availability [20]. More complex algal organoarsenical compounds include water-soluble arsenosugars (i.e. dimethylarsenosugars) and lipid-soluble compounds (arsenolipids). While phytoplankton and macroalgae are the primary producers of complex organoarsenic compounds in the sea (Fig. 1), these organisms are themselves consumed and metabolized by marine animals. Fish and marine invertebrates retain 99% of accumulated arsenic in organic form; and crustacean and mollusk tissues contain higher concentrations of arsenic than fish. The major organoarsenic compound isolated from marine organisms is arsenobetaine. It occurs in algae, clams, flounder, lobsters, sharks, and shrimp. It is not known how arsenosugars and arsenolipids are converted into arsenobetaine within the higher animals in the marine environment. Arsenobetaine is degraded by microbial metabolism in coastal seawater sediments to methylarsonic acid and to inorganic arsenic. This process completes the biological cycling of arsenic in marine biosystems (Fig. 1).

1.2. Methylation of arsenic

Microbial conversion of arsenic to methylarsenic was

![Microbial formation of trimethylarsine from inorganic arsenate. Summarizing work of 19-23,27. The nomenclature for inorganic and organoarsenical compounds is as in 13.](image-url)
first observed over 150 years ago. It has been understood at the level of products formed from the work of Challenger and coworkers before World War II [13,22-24]. Fungi dominate the microbes that produce volatile, garlic-smelling trimethylarsine [23], although bacteria and animal tissues also have this potential [25]. Hall et al. [26] showed that the microbial content of the mouse intestinal cecum (mostly anaerobic bacteria) methylates inorganic arsenic. Up to 40% of low levels of As(III) and As(V) was methylated in vitro by cecal contents in less than 24 h. Both monomethyl- and dimethyl-arsenic compounds were formed, and addition of potential methyl donors (produced by microbes) increased the yield of methylarsonic [26]. The conversion of arsenate to methylarsonic acid or to dimethylarsinic acid (Fig. 2) is a possible mechanism for detoxification.

Since microbial methylation has been reviewed recently [21,27], this topic will not be considered here in depth. However, Fig. 2 shows the proposed reaction pathway for microbial methylation (mostly derived from studies of whole cell transformations and crude enzyme preparations) [23,24]. The purpose here is to introduce a topic that needs the movement from geocycles to purified enzymes and genes, much as presented below for arsenite oxidation and arsenate reduction.

The conversion of water-soluble arsenate to volatile trimethylarsine is a multistep process (Fig. 2), with arsenate initially being reduced to arsenite. There then follows a sequence of methylation and reoxidation, followed by reduction of the organoarsenical intermediates (Fig. 2). Dimethylarsinic acid is the substrate for fungal conversion to arsenobetaine and arsenolipids found in marine animals [19,23].

2. Genes for arsenic resistance

Toxic metals in the environment select and maintain microbes possessing genetic determinants which confer resistance to the toxic compounds. In bacteria heavy metal resistance genes are frequently located on plasmids. While bacterial resistance mechanisms to metallic ions, including those of arsenic, have been studied with molecular detail [28-30], there is much less information on this subject available for algae and fungi. Here, we attempt to put together current knowledge on resistance to, and microbial metabolism of, arsenic compounds, with emphasis on molecular mechanisms of reduction and oxidation coming from recent crystal structures of the enzymes. Other aspects of microbial interactions with arsenicals, especially the genes for plasmid and chromosomal resistance systems, were summarized recently [31] and here the emphasis will be different.

The transport of arsenate into bacterial cells as oxyanions comparable to those of phosphate is carried out by phosphate transport membrane systems, for example the Pit and Pst systems of Escherichia coli (Fig. 3). Arsenate oxyanions in water show three pKₐ values (2.2, 7.0, and 11.5) [13], comparable to 2.1, 7.2, and 12.7 for phosphate. This means that approximately equal amounts of H₃AsO₄⁻²⁻ and H₂AsO₄⁻ occur at pH 7, whereas H₃AsO₄ and H₂AsO₄⁻ predominate in acidic environments. Arsenite (As(III)), in contrast, appears mostly un-ionized as As(OH)₃ at neutral pH, with a pKₐ of 9.3 for dissociation to H₂AsO₄⁻ [13]. As(OH)₃ is transported into cells at neutral pH by aqua-glyceroporins (glycerol transport proteins) [32] in bacteria [33], yeast [34] (Fig. 3) and mammals [35], since As(OH)₃ resembles the inorganic equivalent of a polyol. GlpF, which was originally identified as the glycerol facilitator in E. coli, is a member of the aqua-glyceroporin family. Disruption of the glpF gene confers resistance to antimonite, consistent with the GlpF channel allowing translocation of Sb(OH)₃ [33]. It seems likely that As(OH)₃ is a GlpF substrate also, but arsenite also may enter E. coli by another system. The structure of the E. coli GlpF protein has been solved at 6.9 Å resolution [32] and shows a tetrameric assembly of subunits, comparable to aquaporins, but with a larger central channel in each subunit. The problem for a glyceroporin membrane protein is how to allow glycerol (and arsenite) movement across the membrane while excluding the smaller water molecules.

Deletion of the gene for yeast aqua-glyceroporin Fps1p similarly confers resistance to Sb(III) and to As(III) [36,37]. However, the fpsΔ strain is still relatively sensitive to As(III), suggesting that the yeast cell also has an additional As(III) uptake pathway. High osmolarity, which closes the Fps1p aqua-glyceroporin channel, provides resistance to both As(III) and Sb(III); and mutants with a constitutively open glycerol channel protein are hypersensitive to As(III) and Sb(III). Recently, it was shown that the mammalian homologues of Fps1p, AQP7 and AQP9, functionally substitute for Fps1p in an fpsΔ yeast strain [35]. Thus, it is probable that As(III) also enters mammalian cells via aqua-glyceroporins.

In addition to chromosomal genes that function for uptake of inorganic arsenic as alternative substrates to useful nutrients, many microbes possess genes that specifically confer resistance to inorganic arsenic, both arsenate (As(V)) and arsenite (As(III)), as their natural primary substrates [28,38,39]. As this has been reviewed repeatedly [31], these genes and the proteins will be only briefly considered. In bacteria, these resistance determinants are often found on plasmids, which has facilitated their study at the molecular level. However, basically the same determinants are found also on bacterial chromosomes, including those of E. coli, Pseudomonas aeruginosa, Bacillus subtilis and Mycobacterium tuberculosis. As more and more bacterial genomes are sequenced, it has become clear that arsenic resistance operons are ubiquitous. The chromosomal systems are functional and provide arsenic tolerance; inactivating these ars operons leads to ‘hypersensi-
tivity’ to arsenic compounds. Bacterial *ars* systems confer arsenic resistance primarily by encoding a specific efflux pump that extrudes As(III) from the cytoplasm [39], thus lowering the intracellular concentration of the toxic arsenic.

In some plasmid-determined systems of Gram-negative bacteria, the efflux pump consists of a two-component ATPase complex. The *arsA* gene product is a soluble ATPase subunit [40], which physically associates with an integral membrane protein, the product of the *arsB* gene [41,42]. Prior to efflux, arsenate is enzymatically reduced to arsenite (the substrate of ArsB and the activator of ArsA ATPase activity) by the small cytoplasmic arsenate reductase, the product of the *arsC* gene [31] (see below). In most chromosomal arsenic resistance systems of Gram-negative bacteria and the plasmids and chromosomes of Gram-positive bacteria, contiguous *arsB* and *arsC* genes are found, but there is no *arsA* gene.

Little is known about this subject in algae and fungi. An arsenic resistance gene cluster similar to that of bacteria is found in the yeast *Saccharomyces cerevisiae* [43]. There are three contiguous genes in the cluster, ARR1, ARR2 and ARR3 (previously called ACR1, ACR2 and ACR3). The first gene, ARR1, appears to produce a yeast transcriptional regulator (so it presumably functions as the equivalent of the unrelated bacterial *arsR* gene). Disruption of the chromosomal ARR1 gene leads to hypersensitivity to arsenite and arsenate. Deletion of the ARR3 gene (determinant of the membrane efflux protein) also eliminates both resistances [37], whereas disruption of ARR2 (equivalent to the bacterial *arsC*) eliminates arsenate resistance alone, as expected [44]. The yeast ARR2 arsenate reductase gene also functions in *E. coli* [45]. Arr3p (the protein product of ARR3) is a member of a family of arsenite carrier efflux that includes bacterial and archaeal members such as the *B. subtilis* SKIN (Sigma K insertion) element Yqcl protein. This family is unrelated to the larger family of ArsB proteins found in many bacterial *ars* operons (including in *E. coli* and *Staphylococcus aureus*) and appears to be the result of convergent evolution [46]. Thus far, no other yeast or other eukaryote orthologs of Arr2p and Arr3p have been identified.

In addition to the three ARR gene products, another yeast protein, Ycf1p, which is an ABC ATPase, also contributes to resistance to As(III) and Sb(III), by being located in the vacuolar membrane and by pumping glutathione adducts, As(GS)₃ (and presumably Sb(GS)₃) from the cytoplasm into the vacuole compartment, functioning as an ATPase. GSH, glutathione.

Fig. 3. Arsenite and arsenate transport in *E. coli* and *S. cerevisiae*. ArsC and Arr2p, the cytoplasmic arsenate reductases. ArsB and Arr3p, the potential-driven membrane arsenite efflux proteins (in bacteria coupled with ArsA to form an ATPase). GlpF and Fps1p, the glycerol transport proteins that also transport arsenite. Pit and Pho87p, the potential-coupled phosphate (and arsenate) uptake transporters. PstB, PstC, PstS, the three-component Pst ATP-coupled phosphate (and arsenate) uptake system. Ycf1p, the As(III)-3 GSH adduct carrier that transports the adduct complex into the yeast cell vacuole compartment, functioning as an ATPase. GSH, glutathione.
in the same microbe, for example the *P. aeruginosa* genome contains two *arsC* genes that are unrelated in sequence and probably function with different coupling proteins (see below). An analysis at the National Center for Bioinformatics ‘Cluster of Orthologous Genes’ (COG) site finds 47 presumed *ArsC* protein sequences at http://www.ncbi.nlm.nih.gov/cgi-bin/COG/palox?seq=ArsC and 29 presumed *ArsB* sequences at http://www.ncbi.nlm.nih.gov/cgi-bin/COG/palox?seq=ArsB

Regardless of which family they belong to, an *arsB* gene is followed by and cotranscribed with an *arsC* gene, with few exceptions [28,39]. It is thought that in addition to separate evolutionary origins for different *ArsB* and *ArsC* families, lateral transfer of either or both genes occurred later in microbial evolution, although sufficiently long ago that the G+C content of *ars* genes in each organism closely resembles that of other genes. Since the early Earth atmosphere was not oxidizing, there was probably little arsenate present, and little pressure to evolve an enzyme to reduce arsenate. Therefore, we favor the hypothesis that arsenite efflux proteins evolved earlier than the arsenate reductases, which appeared only after the introduction of oxygen into the atmosphere 3.8 billion years ago by the first cyanobacteria. One can conclude that arsenic resistance including arsenite oxidation and arsenate reduction arose early in the evolution of cellular life on Earth in response to high environmental levels. The subsequent fixation of different families of arsenate reductase (*ArsC*) and arsenite efflux pumps (*ArsB*) in different bacterial lineages may have occurred later, but nevertheless prior to the origin of eukaryotic cells some 850 million years ago.

Genomic sequencing of a more diverse range prokaryotes plus associated physiological studies are needed to phylogenetic questions. A member of one of the two families of bacterial *ars* operons and *arsC* genes is found in most of the more than 100 bacterial *ars* operons that have now been sequenced and *arsC* usually follows the *arsB* gene, even when the families are ‘mixed’. There are a few variants from this pattern: for example, in *M. tuberculosis*, the *arsB* and *arsC* genes are fused into a single gene, encoding a 498-residue protein; other mycobacteria have separate genes. The mycobacterial fusion protein is related to the *ArsB* and *ArsC* proteins of the *B. subtilis* *ars* operon, with 50% (*ArsB*) and 37% (*ArsC*) identical amino acids, and only a three-amino acid linker between regions. Other exceptions include the *arsC* genes in the chromosomes of *P. aeruginosa* (accession number gi15596147), *Haemophilus influenzae* and *Neisseria gonorrhoeae* that are not associated with *arsB* genes. The *P. aeruginosa* genome contains a second copy of the *arsC* gene (accession numbers gi15597475 and AAC69644) within a canonical *arsRBC* operon. Although the two sequences are unrelated, it is thought that both are functional.

Evolutionary trees of protein sequence similarities show three separate clades of arsenate reductases (*ArsC*; Fig. 4), the two bacterial families and the yeast version. The first and best-studied group of arsenate reductases is referred to in Fig. 4A as the glutaredoxin/glutathione (Grx/GSH) clade and has as prototype the *ArsC* arsenate reductase of *E. coli* plasmid R773 [39,48,49]. The plasmid R773 arsenate reductase uses three separate cysteines, one in arsenate reductase and the others in GSH and Grx, in the reaction mechanism (see below). For the nine examples shown in Fig. 4A, six critical amino acids identified by mutational studies in the *ArsC* from plasmid R773 are completely conserved (His8, Cys12, Ser15, Arg60, Arg94 and Arg107 in the R773 numbering). Additional residues are also conserved. All the sequences shown in Fig. 4A are considered orthologs with probable or demonstrated arsenate resistance as phenotype. There are no currently known homologous sequences with different substrates (i.e. paralogs).

The second family of arsenate reductases is referred to as the thioredoxin (Trx) clade in Fig. 4B and has *ArsC* of *S. aureus* plasmid pL258 as prototype [50–52]. As described below, the pL258-type proteins are related in structure and in function [52,53] to the family of low molecular mass protein phosphotyrosine phosphatases (LMW-PTPases), which are widely found in microbes and animals, including humans [53,54]. Many Trx-linked *ArsC* arsenate reductases are from low G+C Gram-positive bacteria, but those from the *ars* operons of *P. aeruginosa* [55] and *Actinobacillus ferrooxidans* [56] are from Gram-negative proteobacteria. For both of these latter two genes, there is evidence that they confer arsenate resistance. The *A. ferrooxidans* *arsC* product has been shown to use Trx and not Grx [56].

The third clade of cytoplasmic arsenic reductase genes is represented by the yeast *Saccharomyces cerevisiae* Arr2p arsenate reductase [43,44], which is related to a different class of larger protein tyrosine phosphatases that includes the eukaryote Cdc25 cell cycle proteins [57]. The closest homolog to Arr2p is the product of another *S. cerevisiae* gene product, the uncharacterized YG4E protein. Arr2p is not found in the second yeast genome, that of *Schizosaccharomyces pombe*, although another *S. pombe* gene product, the thiosulfate sulfurtransferase rhodanase (involved in cysteine biosynthesis [58]), is slightly related and is shown in Fig. 4C. These structurally and functionally related proteins also function with thioredox chemistry. The evolutionary sequence relationships and protein crystal structure similarities suggest that there are common chemical reaction mechanism features between arsenate reductases, phosphatases and sulfurtransferases, each of which uses an oxygen substrate. Finally, two human examples of the Cdc25 class of large protein tyrosine phosphatases are shown as outliers in the Arr2p clade (Fig. 4C). Arr2p and Cdc25 show sequence homology only in the short active site cysteine region [59].

In parallel to the distinct *ArsC* clades, there appear to be two distinct and unrelated families of *ArsB*, the membrane efflux pump (analysis not shown). Several new and
unexpected findings come from this analysis. The total number of known ArsC clades, representing convergent low molecular mass protein phosphosulfurine phosphatases (LMW-PTPase) of bacterial and mammalian origin; and (C) the eukaryotic tree of yeast arsenate reductase Arr2p (NP015526), a second yeast gene product, yeast enzyme rhodanase and two Cdc25A group large PTPases (only comparable lengths of primary sequences were aligned). Reductase determinants for which there are phenotype data are shaded. Dashed lines separate orthologous branches. Accession numbers for sequences used are (A) E. coli plasmids R737 gi78325 and R46 gi17530582; Y. enterocolitica pYVe227 gi10935594 and A. multivorans gi5915690; and genomes of E. coli K12 gi16131375; of Salmonella enterica serovar Typhi gi16761413; of N. gonorrhoeae gi5915691; of P. aeruginosa PAO1 gi15596147 and of H. influenzae gi16272198, (B) Arsenate reductases from staphylococcal plasmids: S. aureus plasmid pI258 gi50729; Staphylococcus xylosus plasmid pSX267 gi31568; and pN315B gi1374157; B. subtilis genome SKIN insertion gi1303778 (listed as YqCM); Bacillus halodurans genome gi1736928; P. aeruginosa PAO1 genome gi15597475; and A. ferrooxidans gi7769068; and LMW-PTPases from B. subtilis genome gi732386 (listed as YwLE) and human PAIF gi1709543; and (C) S. cerevisiae Arr2p arsenate reductase gi2498102; S. cerevisiae YG4E gi1176050 hypothetical protein; S. pombe rhodanese gi5441470; M-phase inducer phosphatases 1 and 2, MPI1 Cdc25A gi266556 and MPI2 gi266558 Cdc25B. Sequence alignments (made with Clustal X 1.81.msw; ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalX/) will be available on http://www.uic.edu/depts/mcmi/home.html. The phylogenetic trees were drawn with Treeview v.3.2 [http://taxonomy.zoology.gla.ac.uk/rod/treeview.html].

Fig. 4. Three families of evolutionarily related arsenate reductase sequences. Selected representatives of (A) the large Grx plus GSH-using clade, including E. coli K12 and plasmid R737; (B) the unrelated Trx-using clade, including three Staphylococcus plasmids, together with paralogous low molecular mass protein phosphosulfurine phosphatases (LMW-PTPase) of bacterial and mammalian origin; and (C) the eukaryotic tree of yeast arsenate reductase Arr2p (NP015526), a second yeast gene product, yeast enzyme rhodanase and two Cdc25A group large PTPases (only comparable lengths of primary sequences were aligned). Reductase determinants for which there are phenotype data are shaded. Dashed lines separate orthologous branches. Accession numbers for sequences used are (A) E. coli plasmids R737 gi78325 and R46 gi17530582; Y. enterocolitica pYVe227 gi10935594 and A. multivorans gi5915690; and genomes of E. coli K12 gi16131375; of Salmonella enterica serovar Typhi gi16761413; of N. gonorrhoeae gi5915691; of P. aeruginosa PAO1 gi15596147 and of H. influenzae gi16272198, (B) Arsenate reductases from staphylococcal plasmids: S. aureus plasmid pI258 gi50729; Staphylococcus xylosus plasmid pSX267 gi31568; and pN315B gi1374157; B. subtilis genome SKIN insertion gi1303778 (listed as YqCM); Bacillus halodurans genome gi1736928; P. aeruginosa PAO1 genome gi15597475; and A. ferrooxidans gi7769068; and LMW-PTPases from B. subtilis genome gi732386 (listed as YwLE) and human PAIF gi1709543; and (C) S. cerevisiae Arr2p arsenate reductase gi2498102; S. cerevisiae YG4E gi1176050 hypothetical protein; S. pombe rhodanese gi5441470; M-phase inducer phosphatases 1 and 2, MPI1 Cdc25A gi266556 and MPI2 gi266558 Cdc25B. Sequence alignments (made with Clustal X 1.81.msw; ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalX/) will be available on http://www.uic.edu/depts/mcmi/home.html. The phylogenetic trees were drawn with Treeview v.3.2 [http://taxonomy.zoology.gla.ac.uk/rod/treeview.html].

unexpected findings come from this analysis. The total number of known ArsC clades, representing convergent evolution and separate families of arsenate reductase, is currently three; but this number may increase with more microbes being analyzed. Although ArsC sequences show clustering in branches (for example for Enterobacteriaceae in Gram-negative bacteria and Staphylococci in low G+C Gram-positive bacteria (Fig. 4), the deeper branchings of these trees are not overall consistent with the ‘universal tree’ of life based on 16S rRNA sequences. For example, the arsB and arsC gene products of staphylococcal plasmids belong to different families. The ArsB sequences of low G+C Gram-positive Bacillus and Staphylococcus are not significantly related. There are in addition Archaeal Ars protein sequences that are not shown in Fig. 4. Halo-
coupled to the general thiol recycling enzymes Grx or Trx [48,51,60]. Nevertheless, they can be subdivided into families whose sequences are unrelated and whose mechanisms differ in detail. These families have common properties that developed by convergent evolution. It is a situation at the enzymatic level somewhat similar to the wings of birds and butterflies: they carry out the same function but have no evolutionary shared ancestor.

The bacterial arsenate reductases [48,50,51,61] are small (131 amino acid residues for S. aureus and 141 for E. coli), monomeric cytoplasmic enzymes. The S. aureus ArsC enzyme uses Trx (a small intracellular protein that functions as a general disulfide reducing agent) in vitro [50,51]. In contrast, the E. coli enzyme requires reduced GSH and Grx, which is similar to Trx but with a different coupling specificity [48,62]. These differences and other major differences [31,48,51] now are understandable in terms of recent studies of protein structure and detailed reaction mechanisms.

3.1. The E. coli GSH/Grx ArsC family

The ArsC reductase from the large resistance plasmid R773 has been well characterized enzymologically and structurally. Purified R773 ArsC exhibits arsenate reductase activity [48], which is dependent on the presence of reduced GSH and Grx. The enzyme has two cysteine residues, Cys12 and Cys106. However, only Cys12 is required for arsenate reduction [63]. There is now crystallographic evidence for both covalent Cys12 thiolate-As(V) and thiolate-As(III) intermediates in the catalytic cycle [49].

The role of Grx in arsenate reduction was investigated [48,62]. In general, Grx can catalyze either intraprotein disulfide bond reduction or reduction of mixed disulfides between a protein Cys thiol and glutathione [64]. The N-proximal cysteine of Grx is required for both reduction reactions, while the other cysteine is required for protein disulfide reduction but not reduction of mixed protein–glutathione disulfides. Single cysteine mutants of each of the three E. coli Grxs were used to distinguish between these two catalytic modes in R773 ArsC-catalyzed arsenate reduction [62]. Grx mutants lacking the second cysteine could still couple E. coli ArsC activity. In contrast, the N-terminal cysteine mutants did not support arsenate reduction. These results indicate strongly that R773 ArsC forms a mixed intramolecular disulfide between ArsC Cys12 and GSH (or, as described below, alternatively the novel intermediate Cys12–S–As(OH)–SG). E. coli has three glutaredoxins, Grx1, Grx2 and Grx3, each of which has a Cys-Pro-Tyr-Cys dithiol consensus sequence [65], and all three can serve as electron donor for the reduction of arsenate by the R773 reductase, with relative efficiencies of Grx2 > Grx3 > Grx1 [62].

Martin et al. [49] reported high-resolution X-ray crystal structures for three forms of R773 ArsC, without bound arsenic and separate crystals with complexed arsenate or arsenite. The overall structures, both secondary structure fold and tertiary structure of ArsC (Fig. 5A), are unrelated to others in protein structure databases, indicating no significant global similarity to other known proteins. This is consistent with the lack of homologous paralogs in sequence libraries (Fig. 4A). There is no relationship between the tertiary structures of the E. coli R773 and the S. aureus and B. subtilis arsenate reductases (Fig. 5), supporting the conclusion that these two classes of enzyme are not related. All three structures have a core of four β-sheet regions, although these are all parallel for the two enzymes from the Trx clade (Fig. 5), but with one antiparallel β-sheet segment for the Grx/GSH enzyme from plasmid R773.

In the crystal structures of R773 arsenate reductase, the catalytic Cys12 residue is in close proximity to Arg60, Arg94 and Arg107, which are concluded to function by stabilizing the bound substrate arsenate and lowering the pKₐ of the Cys12 thiolate. When arsenate was soaked in to the R773 ArsC crystal, a covalent Cys12–S–As(V) bond was formed [49], which is likely to be an intermediate in the reaction mechanism. This covalent intermediate is analogous to S–P bond intermediate of protein tyrosine phosphatases [66,67] and shows tetrahedral geometry with a sulfur–arsenic distance of 2.18 Å [49]. Arg60 moves closer to the Cys12 thiol when the sulfur–arsenate bond forms and the S–AsO₃ adduct hydrogen bonds with the side chains of Arg60 and Arg107, as well as the amide N of Gly11 [49]. When arsenite was soaked in to the crystal instead of arsenate, a different C12 S–As covalent bond was seen; and this is also considered a likely reaction intermediate [49]. This structure appears to contain C12 S–As–OH⁻, which on addition of H₂O and OH⁻ would form the product As(OH)₃, that is then released. The proposed reaction pathway and intermediates are summarized in Fig. 6A. The covalent S–As–O structure in the electron density map with arsenite has only two apparent atoms linked to the arsenic atom. There are two possible structures for the intermediate. One is Cys–S–As=O, a trivalent arsenical with a double-bonded oxygen similar to that found in compounds such as phenylarsine oxide. However, the arsenic–oxygen bond length of 1.86 Å is too long for a double-bonded oxygen. On the other hand, the distance is consistent with the second possibility, Cys–S–As⁺–OH⁻, a novel covalent enzyme–As(III) intermediate.

3.2. The Staphylococcus Trx ArsC family

The S. aureus plasmid pL258 family of arsenate reductases uses different active site cysteine thioles (Fig. 6B) than does plasmid R773 ArsC. Nevertheless, the reaction mechanisms of the two arsenate reductases have marked similarities (Fig. 6). The pL258 enzyme couples to Trx, not Grx [50,51,68]. Unlike the E. coli arsenate reductase, replacement of any of three cysteine residues (Cys10, which is similarly located to Cys12 of R773 arsenate reductase...
...and also functions as the reaction center; and Cys82 and Cys89, for which equivalent cysteines do not occur in the E. coli enzyme) leads to an inactive enzyme in vitro. The enzyme was proteolytically cleaved after catalysis, and peptide products identified by mass spectroscopy analysis indicated a Cys82^Cys89 oxidized cystine, again an intermediate in the reaction pathway\[68\] (Fig. 6B). Zegers et al. \[52\] solved the crystal structures of both reduced and oxidized S. aureus pI258 ArsC. The secondary structure elements are drawn as ribbon diagrams, with $\beta$-strands colored in yellow and a-helices colored in red. Active site residues are shown as space filling models and colored according to atom type (sulfur, yellow; oxygen, red; nitrogen, blue). The thiolate of the active site Cys12 in the R773 ArsC is facing toward the back and so is not visible. The active site thiol residues in the B. subtilis and S. aureus ArsCs are Cys10, Cys82 and Cys89, with the thiolates in yellow. The pI258 ArsC structures are for reduced and oxidized forms of a Cys10-to-Ser Cys15-to-Ala mutant, with the serine hydroxyl of Ser10 in red. N- and C-termini of the polypeptide are marked when visible in the orientation shown.

Fig. 5. Structures of the ArsC arsenate reductases. Structures were drawn with RasMol V2.7 (see http://www.bernstein-plus-sons.com/software/rasmol) using coordinates from the Protein Data Bank. A: E. coli R773 ArsC PDB 1I7D. B: B. subtilis SKIN ArsC PDB 1JL3. C,D: Cys82/Cys89 reduced (PDB 1JF8) and oxidized (PDB 1JFV) forms respectively of S. aureus pI258 ArsC. The secondary structure elements are drawn as ribbon diagrams, with $\beta$-strands colored in yellow and a-helices colored in red. Active site residues are shown as space filling models and colored according to atom type (sulfur, yellow; oxygen, red; nitrogen, blue). The thiolate of the active site Cys12 in the R773 ArsC is facing toward the back and so is not visible. The active site thiol residues in the B. subtilis and S. aureus ArsCs are Cys10, Cys82 and Cys89, with the thiolates in yellow. The pI258 ArsC structures are for reduced and oxidized forms of a Cys10-to-Ser Cys15-to-Ala mutant, with the serine hydroxyl of Ser10 in red. N- and C-termini of the polypeptide are marked when visible in the orientation shown.

and also functions as the reaction center; and Cys82 and Cys89, for which equivalent cysteines do not occur in the E. coli enzyme) leads to an inactive enzyme in vitro. The enzyme was proteolytically cleaved after catalysis, and peptide products identified by mass spectroscopy analysis indicated a Cys82–Cys89 oxidized cystine, again an intermediate in the reaction pathway [68] (Fig. 6B). Zegers et al. [52] solved the crystal structures of both reduced and oxidized S. aureus pI258 arsenate reductase, using an inactive Cys10Ser Cys15Ala double mutant to stabilize the protein from oxidation during crystallization and irradiation. The secondary and tertiary structures of pI258 arsenate reductase (Fig. 5C,D) [52] are remarkably similar to those of a LMW-PTPase from mammals, a relationship predicted [67] based on overall sequence homology (26% amino acid identities) as well as conservation of key residues in the P loop active site, where the tyrosine-phosphate substrate approaches the active site cysteine (Cys10 for pI258 arsenate reductase). The stability of the P loop structure requires the presence of an oxyanion [69,70].

The overall secondary structure of the reduced (active) and oxidized (Cys82–SS–Cys89; inactive) forms of the pI258 arsenate reductase are basically the same, but there is a major change in conformation, with the Cys89 thiol moving more than 10 Å to form the disulfide bond (compare Fig. 5C and D). A recent report [85] combines further...
NMR spectroscopy and X-ray crystal structures for the *Staphylococcus* arsenate reductase to provide evidence for Cys82 nucleophilically attacking Cys10 to form a Cys10–Cys82 disulfide (Fig. 6B) and providing electrons to reduce arsente to arsenite. This model for the Trx-linked arsenate reductase (Fig. 6B) is quite different from the proposed reaction pathway for GSH-linked *E. coli* arsenate reductase (Fig. 6A). The question as to whether basically similar or fundamentally different chemistry occurs with the two enzymes is still not resolved.

The *S. aureus* plasmid pI258 arsenate reductase shows phosphatase activity with the model substrate *p*-nitrophenyl phosphate [52]. The pI258 ArsC had a $K_m$ of 0.5 min$^{-1}$, a very high $K_m$ of 146 mM, and an overall activity far below the range found with enzymatically characterized LMW-PTPases. The oxidized arsenate reductase and mutant Cys10Ser protein were inactive, whereas mutant proteins lacking any of the other three cysteines present in the sequence retained almost full PTPase activity. Arsenate was a competitive inhibitor of phosphatase activity, with a $K_i$ very similar to the $K_m$ for arsenate reductase activity. By all appearances, pI258 arsenate reductase is a dual-function enzyme; and the question can be raised of which activity came first, that toward protein tyrosine-phosphates or that toward arsenate. Some of us favor the latter alternative and suggest that arsenate resistance arose early in cellular life, prior to regulation of cell division by phosphorylation/dephosphorylation [66,67]. Others of us have proposed that arsenate reductases evolved later after the atmosphere became sufficiently oxidizing to allow formation of abundant arsente [39].

Based on the structure of pI258 arsenate reductase and the relationship to protein tyrosine phosphates, Zegers et al. [52] proposed a detailed catalytic mechanism (Fig. 6B) with interesting similarities and differences to that for the R773 arsenate reductase (Fig. 6A). As with the R773 ArsC, the active site Cys10 of the pI258 enzyme nucleophilically attacks the arsenate, with the leaving oxygen protonated from a closely located dicarboxylic acid residue. Asp105. Arsenate is thought to be reduced to arsenite by transfer of electrons from Cys10, which then forms a disulfide bond with neighboring Cys82. Cys89 then attacks Cys82, reforming the Cys10 thiolate. This stage involves the major conformational change in the enzyme. The major difference between the Trx and Grx/GSG classes of arsenate reductases is that GSH plays a role similar to that of pI258 residue Cys82 in the R773 ArsC (Fig. 6A).

A second difference in the catalytic reaction pathways in Fig. 6A,B is the novel Cys12–As-SG intermediate proposed for R773 ArsC [49], which is similar to a bridged S–As–S adduct previously seen with As-trypanothione. In contrast, the Trx arsenate reductases are proposed to release arsenite and form a Cys10–S–S–Cys82 intermediate (Fig. 5D) [52]. An oxidized cystine intermediate was identified by mass spectroscopy [52] with a Cys89Ser mutant enzyme (that can only carry out a single reduction step because the active Cys10-SH cannot be recycled). Further molecular genetic and biochemical studies with both classes of ArsC enzyme will eventually clarify these differences in proposed intermediates (Fig. 6).

A third X-ray crystal structure for arsenate reductase was reported [53], this time for the enzyme encoded by the SKIN element in the chromosome of *B. subtilis* [47]. The *Bacillus* enzyme belongs to the Trx-linked family (Fig. 4) (with 65% amino acid identity to ArsC of pI258). The proposed structures of reduced *Bacillus* and pI258 arsenate reductase are very similar (compare Fig. 5B and C), with only minor differences. As with the pI258 arsenate reductase, Asp105 of the *Bacillus* enzyme is positioned to coordinate with an arsenate bonded to Cys10 [53]. Arg16 is thought to stabilize the active site and to lower the $pK_a$ values of Cys10, Cys82 and Cys89 to facilitate the reaction sequence. The *Bacillus* arsenate reductase also showed phosphatase activity with *p*-nitrophenyl phosphate, and basically the same reaction pathway (Fig. 6B) was proposed [53]. While the *Bacillus* arsc gene has been characterized in vivo [47], the enzyme has not been studied in vitro before, as has the *S. aureus* plasmid pl258 enzyme.

---

**Fig. 6.** The proposed reaction mechanism pathways of ArsC arsenate reductases from (A) plasmid R773 [49] and (B) plasmid pI258 [52]. The figure is highlighted to emphasize the differences in proposed intermediates; see text.)
3.3. The Saccharomyces family, a third clade

The S. cerevisiae Arr2p is the only currently identified eukaryotic arsenate reductase, known to convert arsenate to arsenite [43,44]. It is unrelated to the bacterial arsenate reductases in primary sequence and appears to represent a third convergent evolutionary clade (Fig. 4). Unlike the bacterial ArsC reductases, purified Arr2p is a homodimer of two 130-residue monomers [45]. The yeast Arr2p arsenate reductase shares GSH and Grx substrates with the R773 reductase. The yeast Arr2p arsenate reductase uses GSH and S. cerevisiae Grx1 (or any of the three E. coli Grxs) for recycling [45], similar to the R773 arsenate reductase. Moreover, Grx in which the C-distal cysteine of the CXXC consensus sequence was eliminated by mutation still couples to arsenate reduction [45], suggesting a catalytic pathway similar to that of R773 ArsC (Fig. 5A) and Arr2p is unable to use Trx as a substitute for Grx. The yeast ARR2 gene complements an E. coli strain with a deletion of the chromosomal arsC gene [45].

Arr2p shows significant sequence homology in the active site region to large protein tyrosine phosphatases (PTPases) such as the human cell cycle Cdc25a (Fig. 4C). However, the Cdc25a PTPases are unrelated to the LMW-PTPases that are paralogs of the Trx clade of arsC proteins [54]. The yeast Arr2p arsenate reductase has gained phosphatase activity while losing reductase activity (R. Mukhopadhyay and B.P. Rosen, unpublished). The mutated ARR2 gene no longer confers arsenate resistance but complements a phosphatase-deficient yeast strain. From the ease of mutational engineering of Arr2p, it would seem that protein phosphatases could easily have evolved from arsenate reductases.

In addition to the cytoplasmic ArsC arsenate reductases of arr resistance operons, there are also periplasmic arsenate reductases that are components of respiratory electron transport chains in which arsenate is the terminal electron acceptor [71–73]. Arsenate serves as an anaerobic alternative to oxygen. The respiratory-linked reductases are poorly characterized at the protein and gene levels and will not be described further, except in the context of a hypothesized relationship to arsenite oxidase which functions in the opposite direction (see next below).

4. Arsenite oxidation, microbial and enzymatic

Oxidation of As(III) represents a potential detoxification process that allows microorganisms to tolerate higher levels of arsenite. Several examples of bacterial oxidation of arsenite to arsenate were reported as early as 1918 (reviewed in [21,28,31,74]). Anderson et al. [75] purified and characterized arsenite oxidase from the Alcaligenes faecalis strain of Legge and Turner [76]. This enzyme is located on the outer surface of the inner membrane and exhibits arsenite oxidation activity in the presence of azurin or cytochrome c as electron acceptor. The purified protein was initially thought to be a Mo-pterin monomer containing several metal centers, including both an [4Fe–4S] HiPIP (high potential iron protein) and a Rieske-type [2Fe–2S] center [75,77]. The Mo-pterin cofactor is released on denaturation, like that from other molybdenum proteins [78,79]. The crystal structure of arsenite oxidase was
recently solved by X-ray diffraction analysis [80] and the genes for the two structural subunit polypeptides sequenced (L.T. Phung, unpublished), providing a detailed picture of both structure and reaction pathway (Fig. 7) analogous to those described above for the unrelated cytoplasmic arsenate reductase.

The structure of arsenite oxidase shows two subunits, a larger 88-kDa polypeptide containing the Mo-pterin and a HiPIP 3Fe^4S center and a smaller 14-kDa subunit with the Rieske 2Fe^2S center (Fig. 7) [77,80]. The Mo-pterin cofactor contains two pterins, oriented relative to one another up and down in the protein, similar to other Mo-pterin cofactors in the large dimethyl sulfoxide (DMSO) reductase family of proteins [79]. This is a diverse superfamily of evolutionarily related enzymes that vary in substrates, midpoint electric potential and direction, so that some function physiologically as oxidases (as does arsenite oxidase), while others function as reductases such as DMSO reductase and probably the respiratory arsenate reductase studied by Krafft and Macy [81] (which is a Mo protein with small and large subunits, both containing Fe–S centers). Arsenite oxidase is closest in sequence to the subfamilies nitrate reductase and formate dehydrogenase family of proteins [79]. This is a diverse superfamily of evolutionarily related enzymes that vary in substrates, midpoint electric potential and direction, so that some function physiologically as oxidases (as does arsenite oxidase), while others function as reductases such as DMSO reductase and probably the respiratory arsenate reductase studied by Krafft and Macy [81] (which is a Mo protein with small and large subunits, both containing Fe–S centers). Arsenite oxidase is closest in sequence to the subfamilies nitrate reductase and formate dehydrogenase family of proteins [79].

The DNA sequence (L.T. Phung, unpublished) was obtained with an initial PCR product using oligonucleotide primers designed from the then unpublished primary sequence of Ellis et al. [80]. The amino acid sequence translated from the large asoA gene contains only a single N-terminal residue (methionine) that was not resolved in the crystal structure. Twelve nucleotides before the start of asoA, the asoB gene for the small Rieske subunit ends. The

Fig. 7. The structure and proposed reaction cycle of arsenite oxidase (based on results of [75,80]). Reaction steps are (1) binding of arsenite, AsO_3H_2, to the enzyme, (2) two-electron transfer to Mo, oxidizing As(III) to As(V) and reducing Mo(VI) to Mo(IV), (3) release of arsenate, AsO_4H^2-, (4) two-electron transfer from Mo(IV) to 3Fe–4S center, regenerating Mo(IV) reaction center, (5) two-electron transfer from 3Fe–4S center in large subunit to 2Fe–2S Rieske center of small subunit, and (6) electron transfer from the 2Fe–2S center of arsenite oxidase to the respiratory chain to oxygen.
gene for the small subunit potentially encodes 42 amino acids at its N-terminus that are not found in the crystal structure. These include a canonical signal sequence for a Tat (twin-arginine translocator) [82,83] non-Sec (secretion of polypeptides) protein export signal. Tat-translocated proteins in the periplasm are exported already folded in final tertiary form with bound cofactors (unlike the Sec system which exports unfolded polypeptides that then fold outside the cell membrane). The hypothesis from the preliminary gene sequence is then that a Tat signal sequence of the small subunit suffices for Tat export of the heterodimer, consisting of small and large AsoA and AsoB subunits, with inserted Mo-pterin and Fe–S clusters that were synthesized, assembled and inserted in the cytoplasm. This follows from the primary sequence of the unprocessed protein translated from the *asoAB* genes (L.T. Phung, unpublished) plus extensive studies on other members of the DMSO reductase family [79]. We hypothesize that recently discovered respiratory arsenate reductases of anaerobic bacteria [71,73,84] will be recognized as a new branch of the large DMSO reductase family once primary sequences and structures become available.

The primary amino acid sequences from the DNA sequence and the electron density map of arsenite oxidase differ in a surprising 9% of positions, 78 in AsoA and 12 in AsoB. Those differences available when the article by Ellis et al. [80] was prepared (56 in the center of AsoA and not including the N- and C-termini) were corrected by Ellis et al. [80] from the DNA sequence (L.T. Phung, unpublished). No residues implicated in the reaction pathway or binding of cofactors differ. Thus, the Cys22–X2–Cys25–X1–Cys28–X10–Ser99 positions involved in binding the 3Fe–4S cluster are in both sequences, as are Cys60, His62, Cys78 and His81 binding the Rieske 2Fe–2S cluster in the small subunit [80]. G. Anderson (personal communication) subsequently identified by direct sequencing six additional N-terminal amino acid residues in the Rieske subunit that are not resolved in the crystal structure but found in the gene translation product. In sum, crystallography and gene sequencing have recently placed the arsenite oxidase system ready for detailed molecular genetic analysis.

Acknowledgements

Work on arsenic resistance in our laboratories has been supported by grants from the U.S. National Institutes of Health (NIH Grant GM52216 to B.P.R.) and Department of Energy (to S.S.). Colleagues have freely exchanged information and ideas over the years, especially Gretchen Anderson, Carlos Cervantes, Russ Hille, Guangyung Ji, José C. Martins, Alan McEwan, Joris Messens and Xiao-Dong Su.

References

Liu, Z., Shen, J., Carbrey, J.M., Mukhopadhyay, R., Agre, P. and Silver, S., Budd, K., Leahy, K.M., Shaw, W.V., Hammond, D., Novo-
characterization of ACR2p, the Saccharomyces cerevisiae arsenate reductase. J. Biol. Chem. 275, 21149–21157.
Bobrowicz, P., Wysocki, R., Owsiank, G., Goffau, A. and Ulaszew-

324
311–325
FEMS Microbiology Reviews 26 (2002) 311–325
R. Mukhopadhyay et al.


[80] Ellis, P.J., Conrads, T., Hille, R. and Kuhn, P. (2001) Crystal structure of the 100 kDa arsenite oxidase from Alcaligenes faecalis in two crystal forms at 1.64 Å and 2.03 Å. Structure 9, 125–132.


