Published in final edited form as:
J Biol Chem. 2007 November 23; 282(47): 34346–34355. doi:10.1074/jbc.M706565200.

Convergent Evolution of a New Arsenic Binding Site in the ArsR/SmtB Family of Metalloregulators[,*,S]

Jie Qin[‡], **Hsueh-Liang Fu**[‡], **Jun Ye**[‡], **Krisztina Z. Bencze**[‡], **Timothy L. Stemmler**[‡], **Douglas E. Rawlings**[§], and **Barry P. Rosen**[‡,1]

[‡] Department of Biochemistry and Molecular Biology, Wayne State University, School of Medicine, Detroit, Michigan 48201

[§] Department of Microbiology, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa

Abstract

Acidithiobacillus ferrooxidans has an arsenic resistance operon that is controlled by an As(III)-responsive transcriptional repressor, AfArsR, a member of the ArsR/SmtB family of metalloregulators. AfArsR lacks the As(III) binding site of the ArsRs from plasmid R773 and *Escherichia coli*, which have a Cys^{32}-Val-Cys^{34}-Asp-Leu-Cys^{37} sequence in the DNA binding site. In contrast, it has three cysteine residues, Cys^{95}, Cys^{96}, and Cys^{102}, that are not present in the R773 and *E. coli* ArsRs. The results of direct As(III) binding measurements and x-ray absorption spectroscopy show that these three cysteine residues form a 3-coordinate As(III) binding site. DNA binding studies indicate that binding of As(III) to these cysteine residues produces derepression. Homology modeling indicates that As(III) binding sites in AfArsR are located at the ends of antiparallel C-terminal helices in each monomer that form a dimerization domain. These results suggest that the As(III)-S_3 binding sites in AfArsR and R773 ArsR arose independently at spatially distinct locations in their three-dimensional structures.

The metalloid arsenic is a ubiquitous environmental toxin that is ranked first on Superfund List of Hazardous Substances (www.atsdr.cdc.gov/cercla/05list.html). Nearly every organism has genes for arsenic detoxification (1). In bacteria, arsenic resistance is most frequently conferred by chromosomal or plasmid-encoded *ars* operons that are controlled by members of the ArsR/SmtB family of transcriptional repressors (2,3). The well characterized plasmid R773 and *Escherichia coli* chromosomal ArsRs are homodimers that have a Cys^{32}-Val-Cys^{34}-Asp-Leu-Cys^{37} sequence at the start of the helix-loop-helix DNA binding site (4) (Fig. 1). The three sulfur thiolates of the cysteine residues form an S_3 binding site for As(III), and binding of metalloid is presumed to cause a conformational change in the repressor leading to dissociation from the DNA and hence derepression (5). This site is congruent with the 4-coordinate (S_4) Cd(II) binding site of the CadC repressor (6), and we have termed these Type 1 metal binding site. CadC also has a C-terminal non-regulatory Zn(II) binding site that is formed by Asp^{101} and His^{103} from one monomer and $His^{114'}$ and $Glu^{117'}$ from the other monomer at the dimerization domain. This site corresponds to the

---

[S]The on-line version of this article (available at http://www.jbc.org) contains supplemental Fig. S1.

[*]This work was supported by National Institutes of Health Grants R01 AI043428 (to B. P. R.) and R01 DK068139 (to T. L. S.). A portion of the research was carried out at the Stanford Synchrotron Radiation Laboratory (SSRL), a national user facility operated by Stanford University on behalf of the United States Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program was supported by the Department of Energy, Office of Biological and Environmental Research, and the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program.

1To whom correspondence should be addressed. Tel.: 313-577-1512; Fax: 313-577-2765; brosen@med.wayne.edu.

Zn(II) regulatory site of SmtB, and we have termed these Type 2 metal binding sites. The Type 1 and Type 2 sites have also been called α3N and α5 sites (7), but these designations are imprecise because they were based on an incomplete structure of SmtB.

Recently Butcher *et al*. (8) identified an *ars* operon in *Acidithiobacillus ferrooxidans* in which *arsRC* genes are transcribed divergently from *arsBH* genes. They termed the ArsR homologue "atypical" because it does not contain the $CXC(X)_2C$ motif in the DNA binding site (9) (Fig. 1). Deletion of the last 19 amino acid residues had no effect on arsenic regulation, but deletion of an additional 28 residues resulted in loss of regulation from the *arsBH* promoter. Within that sequence is a vicinal cysteine pair, Cys^{95} and Cys^{96} that is conserved in closely related homologues. Because vicinal cysteine pairs can form the strongest binding site found in nature for trivalent metalloids, we investigated the involvement of that cysteine pair and a third residue, Cys^{102}, in transcriptional regulation by trivalent metalloids. The data strongly support the idea that the three cysteine residues are ligands to As(III) and form an S_3 site in which each sulfur atom is 2.25 Å from the arsenic atom. The results of DNA binding studies indicate that this is the inducer binding site. This As(III) binding site does not correspond to known metal binding sites in other members of the ArsR/SmtB family, showing that As(III) binding sites can arise by convergent evolution, even in otherwise homologous proteins.

MATERIALS AND METHODS

Strains, Plasmids, Media, and Reagents

E. coli cells were grown in Luria-Bertani (LB) medium (10) at 37 °C supplemented with 125 μg/ml ampicillin, 25 μg/ml kanamycin, 12.5 μg/ml tetracycline, or 25 μg/ml chloramphenicol, as required. *E. coli* strains DH5α (Promega, Madison, WI), XL1-Blue (F′ *proA+B+lacI*^{q} Δ(*lacZ*) M15 Tn*10*(*tetR*) *recA1 endA1 gyr96 thi-1 hsdR17 supE44 relA1 lac*), were used for plasmid construction, replication, and mutagenesis. Plasmids pB2*lacZ*, pB3*lacZ*, and *E. coli* strain ACSH501^q (*rspl* Δ(*lac-pro*) (F′ *tra*D36 *pro*AB *lacI*$_q$ ΔM15) Δ*ars::cam*) (9) were used for analysis of AfArsR function *in vivo*. Strain Top10 (Invitrogen) was used for protein expression. Bacterial growth was monitored by measuring the $A_{600\ nm}$. All reagents were obtained from commercial sources.

Plasmid Construction and Mutagenesis

For expression of AfArsR from *A. ferrooxidans* in *E. coli*, plasmid pBAD*arsR*, in which the *arsR* gene is under the control of the arabinose promoter and has the sequence for a C-terminal His tag, was constructed. Two pairs of primers were used to amplify the *arsR* gene from plasmid pB2*lacZ*. The first pair was 5′-GTC-CCGGGTACCGTGGCAGCAATTTTCTG-3′ (changed site underlined) and 5′-AATACCATGGAACCACTACAAGACCCTGC-3′ (NcoI site underlined). The second pair was 5′-GGTTTGTCGACCTGGTTTCCTTCCTGGAC-3′ (SalI site underlined) and 5′-CAGAAAATTGCTGCCACGGTACCCGGGAC-3′ (changed site underlined). Introduction of the CAT codon eliminated the NcoI site in *arsR* without changing the ArsR amino acid sequence. The two PCR fragments generated with the first set of primers that were gel purified (Qiagen Corp., Valencia, CA) were used as template, and the second set of primers were used for the second PCR amplification. A 0.35-kbp PCR fragment was gel purified, digested with NcoI and SalI, and ligated into vector plasmid pBAD/Myc-His-C that had been digested with the same restriction enzymes, generating plasmid pBAD*arsR*.

Cysteine-to-serine mutants were constructed by using a QuikChange XL site-directed mutagenesis kit (Stratagene, La Jolla, CA) using the following forward and reverse primers (mutation sites underlined). The primers were: C95S, 5′-

GCACAGTCCCGGGTACCGTGGCAGGAATTTTCTGTGAGATACGCG-3′ and 5′-CGCGTATCTCACAGAAAATTCCTGCCACGGTACCCGGGACTGTGC-3′; C96S, 5′-GCACAGTCCCGGGTACCGTGGGAGCAATTTTCTGTGAGATACGCG-3′ and 5′-CGCGTATCTCACAGAAAATTGCTCCCACGGTACCCGGGACTGTGC-3′; C96S/C96S, 5′-GCACAGTCCCGGGTACCGTGGGAGGAATTTTCTGTGAGATACGCG-3′ and 5′-CGCGTATCTCACAGAAAATTCCTCCCACGGTACCCGGGACTGTGC-3′; C102S, 5′-GGTTTCACCGGATAGGGCAGAGTCCCGGGTACCGTGG-3′ and 5′-CCACGGTACCCGGGACTCTGCCCTATCCGGTGAAACC-3′.

Each mutation was confirmed by sequencing of the entire *arsR* gene. Cysteine-to-alanine mutants were also constructed, and the results were similar to those with serine substitutions, so no data are shown for the alanine substitutions.

A *lacZ* reporter plasmid was constructed from plasmid pMC1403-B3*lacZ* (9), in which *lacZ* was fused to the *A. ferrooxidans arsB* promoter. Plasmid pMC1403-B3*lacZ* was digested with BstBI (Promega Corp., Madison, WI) and vector plasmid pACYC184 was digested with Bsu36I. The fragments were filled in with DNA polymerase I Klenow fragment (Promega Corp., Madison, WI), extracted twice with phenol/chloroform/isoamyl alcohol (25:24:1), and precipitated with 1/10 volume of 3 M sodium acetate and 3 volumes of 95% ethanol at −20 °C. Vacuum-dried DNA was suspended in the appropriate restriction enzyme buffer and digested with EcoRI and gel purified. The *lacZ-arsO* fragment was ligated into pACYC184 to form plasmid pACYC184*lacZ-arsO*. The presence of the insert was verified by digestion with SphI and EcoRI.

To assay *lacZ* activity, a single colony of *E. coli* strain ACSH501^{q} (*rspl* Δ(*lac-pro*) (F′ *tra*D36 *pro*AB lacI^{q} ΔM15) Δ*ars::cam*) bearing plasmids pBAD*arsR* and pACYC184*arsO-lacZ* was inoculated into 5 ml of LB medium supplemented with 125 μg/ml ampicilin, 30 μg/ml chloramphenicol, and 12.5 μg/ml tetracycline and incubated at 37 °C overnight. Late exponential phase cells were diluted 50-fold into 20 ml of LB medium containing the appropriate antibiotics and induced with 0.2% arabinose at 37 °C. For constitutive expression of lacZ while repressing *afarsR* expression, 0.2% glucose was added instead of arabinose. For induction by As(III), 25 μM sodium arsenite and 0.2% arabinose were added together. Growth was monitored as $A_{600\ nm}$ and continued until the cultures reached stationary phase. To assay β-galactosidase activity, a portion of the culture was harvested and suspended in buffer consisting of 40 mM Na_2HPO_4, 60 mM NaH_2PO_4, 10 mM KCl, and 1 mM $MgSO_4$ at an $A_{600\ nm}$ of 0.5. An aliquot (0.2 ml) of cells was vigorously mixed with 10 μl of 1% SDS and incubated for 5 min at room temperature. The reaction was initiated by addition of 0.5 ml of 0.1 mg/ml of *o*-nitrophenyl-β-D-galactoside, and the rate of hydrolysis monitored at 420 nm. Relative activity was calculated as the percentage of the activity of uninduced cells.

AfArsR Purification

Cells of *E. coli* strain Top10 bearing wild type or mutant pBAD*arsR* plasmids were grown at 37 °C in LB medium to an $A_{600\ nm}$ of 0.5, at which point 0.2% arabinose was added as inducer. The cells were grown for another 4 h and harvested by centrifugation at 4 °C, washed once with buffer A (50 mM MOPS[2]-KOH, pH 7.5, containing 20% (w/v) glycerol, 0.5 M NaCl, 20 mM imidazole, and 10 mM 2-mercaptoethanol), and suspended in 5 ml of buffer A/g of wet cells. The cells were lysed by a single pass through a French Press cell at 20,000 p.s.i. and 2.5 μl/g of wet cell of diisopropyl fluorophosphate was added immediately.

[2]The abbreviations used are: MOPS, 4-morpholinepropanesulfonic acid; MMTS, methyl methanethiosulfonate; XAS, x-ray absorption spectroscopy; XANES, x-ray absorption near edge structure; EXAFS, extended X-ray absorption fine structure.

Membranes and unbroken cells were removed by centrifugation at 150,000 × *g* for 1 h, and the supernatant solution was loaded at a flow rate of 0.5 ml/min onto a Ni(II)-nitrilotriacetic acid column pre-equilibrated with buffer A. The column was then washed with 150 ml of buffer A followed by elution with 60 ml of buffer A with the concentration of imidazole increased to 0.2 M. The eluted protein was concentrated with a 10-kDa cut-off Amicon Ultracentrifugal filter (Millipore Corp., Billerica, MA). The concentrated protein was further purified by gel filtration using Superdex 75 (Amersham Biosciences) in a 45 × 1.5-cm column with a total bed volume of 80 ml. The protein was eluted with buffer A containing 0.2 mM EDTA at a flow rate of 0.3 ml/min. AfArsR was identified by SDS-PAGE (11). Fractions containing AfArsR were concentrated by ultrafiltration centrifugation. Protein concentrations were estimated using the method of Bradford (12) or from the $A_{280\ nm}$ using a calculated extinction coefficient of 4,200 cm^{-1} (13).

AfArsR Molecular Mass Determination

The molecular mass of wild type and mutant AfArsRs was measured using a high pressure liquid chromatograph (PerkinElmer Series 200) with synchropak GPC300 column and monitored at $A_{280\ nm}$. Protein was eluted with a buffer consisting of 50 mM MOPS-KOH, pH 7.5, containing 20% glycerol and 0.5 M NaCl at a flow rate of 0.5 ml/min. The column was calibrated with 20 μl of a mixture of gel filtration calibration standards 17-0442-01 (Amersham Biosciences) of blue dextran 2000, ribonuclease, chymotrypsinogen A, ovalbumin, and albumin (5 mg/ml of each). Wild type and mutant AfArsRs (20 μl of 5 mg/ml) were injected, and elution was monitored at $A_{280\ nm}$. For each protein, K_{av} was calculated as $K_{av} = (V_e - V_0)/(V_t - V_0)$.

Methyl Methanethiosulfonate (MMTS) Modification

MMTS modification was performed as described (14). Reducing agents were removed from AfArsR prior to reaction with MMTS by two passages through a Bio-Gel P-6 Micro Bio-Spin column (Bio-Rad) pre-equilibrated with degassed 50 mM MOPS-KOH, pH 7.5, 20% glycerol, and 0.5 M NaCl. AfArsR was then incubated with MMTS at a ratio of 6 eq per cysteine residue. The concentration of free cysteine thiolate was determined by titration with 0.1 mM 5,5′-dithiobis-(2-nitrobenzoic acid) in a buffer consisting of 0.1 M Tris-HCl, pH 8, containing 4 M guanidine-HCl and 0.2–2 mg/ml of AfArsR. After 5 min, the amount of free –SH was calculated from $A_{412\ nm}$ using a molar extinction coefficient of 13,600 $M^{-1}\ cm^{-1}$ (15).

Measurement of Metalloid Binding

The buffer used for purification of AfArsR was exchanged with a buffer containing 50 mM MOPS-KOH, pH 7, using a Bio-Gel P-6 Micro Bio-Spin column (Bio-Rad). Buffers were degassed by bubbling with argon. Purified protein (10 μM) was incubated at 4 °C with the indicated concentrations of sodium arsenite or potassium anti-monyl tartrate. After 1 h, each sample was passed through a Bio-Gel P-6 column pre-exchanged with the same buffer. Portions (25 μl) were diluted with 2% HNO_3, and the quantity of metalloid measured by inductively coupled mass spectrometry with a PerkinElmer ELAN 9000. Antimony standard solutions in the range of 0.5–10 ppb in 2% HNO_3 were obtained from Ultra Scientific, Inc. (North Kingstown, RI).

X-ray Absorption Spectroscopy (XAS)

XAS samples of wild type and C102S AfArsR were prepared in 50 mM MOPS, 0.5 M NaCl, and 30% glycerol, pH 7.5. Multiple reproducible samples were prepared at substoichiometric As(III):protein ratios at an As(III) concentration of 1 mM. Solution

samples were loaded in Lucite cells, wrapped in Kapton tape, and flash frozen in liquid nitrogen.

XAS data were collected at the Stanford Synchrotron Radiation Laboratory (SSRL) beamlines 9-3 and 10-2. Harmonic rejection was achieved by utilizing a harmonic rejection mirror on beamline 9-3 and detuning the Si (220) double-crystal monochromator on beamline 10-2. During data collection, samples were maintained at 10 K using Oxford Instrument continuous flow liquid helium cryostats at both locations. Fluorescence excitation spectra were collected using a 30-element and a 13-element germanium solid-state array detector at beam-lines 9-3 and 10-2, respectively. XAS spectra were measured using 5 eV steps in the pre-edge region (11,625–11,825 eV), 0.25 eV steps in the edge region (11,850–11,900 eV), and 0.05 $Å^{-1}$ increments in the extended x-ray absorption fine structure (EXAFS) region out to a k range of 13 $Å^{-1}$, integrating from 2 to 20 s in a k-weighted manner in the EXAFS region for a total scan length of ~35 min. X-ray energies were calibrated using an arsenic foil absorption spectrum collected simultaneously with the protein data. The first inflection point for the arsenic foil edge was assigned to 11,867 eV.

Data reduction and analysis followed previously published protocols (16). Prior to data reduction and averaging, each fluorescence channel of each scan was examined for spectral anomalies. Published XAS spectra represent the average of 5 scans. Spectra were processed using the Mac OS X version of the EXAFSPAK program suite integrated with Feff 7.2 software for theoretical model generation. Data reduction involved fitting the pre-edge spectral region with a second order polynomial and the EXAFS region with a cubic spline constructed of three regions. Following spline subtraction, EXAFS data were converted into k space using an E_0 of 11,885 eV. The k^3 weighted EXAFS were Fourier transformed and simulated. The final published fitting results are, however, from analysis of the raw unfiltered data. Data were fit using single scattering amplitude and phase functions generated using Feff. Single scattering Feff models were calculated for oxygen, sulfur, and arsenic coordination to simulate arsenic-ligand nearest neighbor environments. Values for the scale factor and E_0, obtained previously from fitting crystallographically characterized As(III) model compounds, were equal to 0.98 and −10, respectively (16). During fitting analysis, the scale factor and E_0 were fixed, the coordination number was fixed but manually varied, and values for the bond length and Debye-Waller factor were allowed to vary freely. The lowest mean square deviation between data and fit, corrected for the number of degrees of freedom (F'), was used to judge best-fit EXAFS simulations and to justify addition of supplementary interactions within the simulations.

Circular Dichroism Measurements

Circular dichroism (CD) spectra from 190 to 260 nm were acquired with a spectrometer from Olis Inc. (Bogart, GA) at 20 °C using a 0.2-cm cell at 1.8-nm intervals. Three scans were averaged for each spectrum. Wild and mutant AfArsRs were assayed at 10 μM in 10 mM potassium phosphate buffer, pH 7.4.

DNA Mobility Shift Assays

DNA mobility shift assays were performed as follows. A 199-bp DNA fragment was amplified by PCR from plasmid pB2*lacZ* by using the primers 5′-GGCGAGGCCAGGGCCTCCAGGC-3′ and 5′-AAAACGACGGGATCCAGGG-3′. The PCR fragment was gel purified, and 40 ng was mixed with the indicated concentrations of purified wild type or mutant AfArsRs, with or without As(III), in a buffer consisting of 10 mM Tris-Cl, pH 7.6, 80 mM KCl, 0.2 mM EDTA, 0.2 mM dithiothreitol, 10% glycerol, and 0.75 μM bovine serum albumin for 30 min on ice in a total volume 20 μl. The DNA/AfArsR

mixture was separated by PAGE on a 6% nondenaturating gel at 4 °C for 30 min and stained with SYBR Green (Invitrogen).

DNA Footprinting

Complementary DNA primers in which one was labeled at the 5′ end with light-sabre green fluorescence (Integrated DNA Technologies, Coralville, IA), as indicated, were used to amplify the *A. ferrooxidans ars* operator region. Two pairs of primers were used, one for determining the *arsRC* direction, and the other for the *arsBH* direction. The primer pairs used were: *arsRC*, 5′-GGCGAGGCCAGGGCCTCCAGGC-3′ (labeled) and 5′-AAAACGACGGGATCCAGGG-3′; *arsBH*, 5′-GGCGAGGCCAGGGCCTCCAGGC-3′ and 5′-GTAACGCCAGGGTTTTCCCAGTCACGACG-3′ (labeled). The PCR products were purified with a NucleoSpin 635961 kit (Clontech) and digested with DNase I (Sigma) in a buffer consisting of 10 mM Tris-HCl, pH 7.6, containing 80 mM KCl, 0.2 mM EDTA, 0.2 mM dithiothreitol, 10% glycerol, 0.75 μM bovine serum albumin, and 5 mM $MgCl_2$. After optimization, 0.45 μM DNA and 0.16 Kunitz units of DNase I were used in a 20-μl reaction volume at room temperature for 30 min. As a negative control 5 μM bovine serum albumin was used in place of AfArsR. The reaction was terminated by heating at 95 °C for 5 min. The DNA was precipitated by addition of 2 μl of 3 M sodium acetate, 2 μl of 0.1 M EDTA, 1 μl of glycogen, and 60 μl of 95% ethanol (at −20 °C) and pelleted by centrifugation at 13,000 × *g* at 4 °C for 20 min. The pellet was washed once with 0.2 ml of 70% ethanol. The DNA was vacuum dried for 20 min and suspended in sequencing buffer (Beckman Coulter, Fullerton, CA) to which 0.5 μl of DNA standard was added (CEQ DNA Size Standard-400 P/N 608098, Beckman Coulter). Then, samples were loaded into the 96-well sample plate and assayed with a CEQ 2000XL DNA sequencer (Beckman Coulter). The running conditions were: capillary temperature, 58 °C; denaturation at 90 °C for 2 min; injection at 2 kV for 30 s; and separation at 6 kV for 35 min. The data were analyzed with CEQ 2000 fragment analysis software.

Fluorescence Anisotropy

DNA binding studies by fluorescence anisotropy using a Photon Technology International spectrofluorometer were fitted with polarizers in the L format. Changes in anisotropy were calculated after each addition using the supplied Felix32 software. Complementary 30-mer oligonucleotides, one of which was labeled at the 5′ end with fluorescein, were synthesized (Integrated DNA Technologies, Inc., Coralville, IA) based on the DNA footprinting results: 6-FAM-5′-ATCCACGAATATTTCTTGCAGTATTGACAA-3′ and 5-TAGGTGCTTATAAAGAACGTCATAACTGTT-3′. Desalted DNA was annealed at 94 °C for 5 min, cooled to room temperature, and stored in aliquots at −20 °C. Anisotropy measurements were performed in a buffer consisting of 10 mM MOPS-NaOH, pH 7.5, containing 0.1 M NaCl, 15% glycerol, and 20 nM DNA. Wild type or mutant AfArsRs were titrated with 1.7 ml of 50 nM fluorescein-labeled *ars* O/P DNA. AfArsR was metallated by mixing 1 M eq of AfArsR monomer with the indicated concentrations of As(III) or Sb(III) under anaerobic conditions.

RESULTS

Contribution of AfArsR Cysteine Residues to Metalloregulation in Vivo

An alignment of the primary sequence of homologues of *A. ferroxidans* ArsR shows that the six closest homologues have a vicinal cysteine pair corresponding to Cys^{95}–Cys^{96} in AfArsR, and a single additional cysteine, which is Cys^{102} in AfArsR (Fig. 1). The codons for these three cysteine residues were altered alone and in combination with serine or alanine codons in plasmid pBAD-RT-*arsR*, with expression controlled by the *ars* promoter. (Both types of substitutions at each of the three positions had similar effects, so only results

with serine mutants are shown.) The mutated *arsR* genes were co-transformed into *E. coli* strain ACSH501^{q}, in which the chromosomal *arsRBC* operon was deleted (17), with plasmid pACYC184-*lacZ-arsO*, in which a *lacZ* gene was introduced under control of the *A. ferroxidans arsBH* promoter. Expression of *lacZ* was constitutive when wild type *afarsR* was repressed by growth on glucose (Fig. 2). When wild type *afarsR* was expressed by addition of arabinose, *lacZ* was repressed, and addition of 25 μM sodium arsenite resulted in derepression of *lacZ*.

Mutagenesis of the codons for either Cys^{95} or Cys^{96} to serine codons (or both) resulted in constitutive *lacZ* expression, suggesting that these mutants are unable to bind to the *ars* promoter. Curiously, addition of sodium arsenite resulted in lower expression. On the other hand, a C102S mutant was able to repress in the absence of metalloid and exhibited partial derepression in its presence, suggesting that Cys^{102} may contribute to but is not required for metalloregulation.

Cys^{95}, Cys^{96}, and Cys^{102} Form a Three-coordinate As(III) Binding Site

The three single cysteine mutants and the C95A/C96A double mutants were expressed and purified as His_6-tagged proteins. Each eluted from a gel filtration column at a position corresponding to 31 kDa (Fig. 3). Thus the substitutions did not interfere with dimerization, indicating that the cysteine residues are not required for dimer formation. Circular dichroism spectroscopy in the far UV showed that each protein is highly helical, with negative maxima peaks at 208 and 220 nm (Fig. 4). The cysteine substitutions did not affect the CD spectra significantly, indicating that the overall helical content was not affected by the mutations. Addition of As(III) to the wild type protein also did not change the spectrum, indicating that binding of inducer does not produce a substantial alteration in the overall structure.

Binding of As(III) and Sb(III) by the wild type and mutant AfArsRs was assayed (Fig. 5). Wild type AfArsR bound 1 mol of As(III) or Sb(III) per monomer. The C95S, C96S, and C95S/C96S mutants did not bind either trivalent metalloid with high affinity. The C102S mutant bound both As(III) and Sb(III), but the affinity was lower than the wild type and too low to calculate accurately. When either the wild type or C102S proteins was treated with MMTS, which specifically forms a methyl disulfide with cysteine residues, binding of both metalloids was reduced to background levels. These results are consistent with a low affinity metalloid binding site composed of Cys^{95} and Cys^{96}, and that the presence of Cys^{102} converts it into a high affinity site.

The nature of the binding site was investigated in structural detail using XAS. XANES analysis indicates arsenic in both the wild type and C102S proteins is stably bound as As(III) (supplemental Fig. S1). The observed values for the first inflection point energies are 11868.87 and 11869.33 eV for wild type and C102S, respectively, consistent with observed values for As(III) systems (16). The EXAFS fitting results are most consistent with trivalent arsenic being maintained in a 3-coordinate structural environment in both wild type and mutant proteins (Fig. 6). Simulation analysis for wild type AfArsR were best fit with a single nearest neighbor ligand environment constructed only with sulfur atoms (Table 1). Attempts to simulate these data with a mixed sulfur and oxygen/nitrogen nearest neighbor environment were unsuccessful. In contrast, simulation analysis for C102S showed the nearest neighbor environment in the mutant was constructed by one oxygen and two sulfur atoms, for an averaged environment of As-S_2O. The average As-S bond length of 2.24 Å in the mutant is identical to that seen in the wild type, whereas the As-O bond length of 1.76 Å is equal to that obtained for a hydroxyl bound to As(III) (16). Attempts to simulate long-range scattering (R > 2.5 Å) were unsuccessful in both samples. The most reasonable interpretation of these data is that in wild type AfArsR, Cys^{95}, Cys^{96}, and Cys^{102} form an S_3 binding site for As(III), with two sites in the homodimer. In C102S, one of the coordinated

sulfur atoms is replaced by an oxygen-based ligand for an S_2O binding site for As(III). The short As-O bond length in C102S matches that expected for a bound hydroxyl and is about 0.1 Å shorter than that expected for a bound water or a serine hydroxyl group (16).

As(III) Binding Induces Dissociation of AfArsR from DNA

The binding of AfArsR to DNA was assayed by gel mobility shift assay using a 199-bp double-stranded fragment containing the operator-promoter region (Fig. 7). Wild type AfArsR and the C102S mutant retarded the mobility of the DNA. The mobility shift was abolished by addition of As(III), but the C102S mutant required considerably more metalloid than the wild type. In contrast, neither the C95S nor C96S mutant affected the migration of the DNA. These results are consistent with the data from the reporter gene assays (Fig. 2) and indicate that the wild type protein binds to the *ars* promoter and dissociates upon binding of As(III). The C102S mutant likewise responds to As(III), both *in vivo* and *in vitro*, but with lower affinity for metalloid. Alteration of Cys^{95} or Cys^{96} produces an AfArsR that no longer binds to DNA.

DNA binding was examined in more detail by a fluorescence anisotropy assay. To design an oligonucleotide sequence for this assay, the binding site for AfArsR was first determined by DNA footprinting of the region between the two divergent transcripts. AfArsR protected the 28-nucleotide sequence 5′-ATCCACGAATATTTCTTGCAGTATTGAC-3′, which is located at −60 to −86 nucleotides relative to the start of the *arsB* gene (Fig. 8). Complementary 30-mer oligonucleotides were synthesized based on this sequence (the 28 nucleotides plus two additional upstream nucleotides). One oligonucleotide was labeled at the 5-end with 6-FAM to provide a fluorescein fluorophore for fluorescence anisotropy measurements. In this assay, anisotropy increases as the rotation of the DNA molecules slow as a consequence of AfArsR binding. Anisotropy increased as a function of the concentration of either wild type or C102S AfArsR (Fig. 9*A*). Addition of either As(III) or Sb(III) reversed the increase in a concentration-dependent manner (Fig. 9*B*). Again, the C102S ArsR also exhibited lower affinity for metalloid than wild type repressor. From three separate experiments measuring As(III)-induced release, K_d values of 12 ±2 μM for the wild type and 73 ± 32 for C102S were calculated. Thus, in these assays, the affinity of wild type AfArsR for Sb(III) was ~6-fold higher than for As(III). Cd(II), which is not an inducer, did not reduce anisotropy. Modification of the cysteine residues in wild type AfArsR with MMTS did not affect its ability to bind to DNA (Fig. 9A) but eliminated dissociation by As(III) (Fig. 9*C*). The C95S, C96S, and C95S/C96S mutants did not produce an increase in anisotropy, again indicating that mutagenesis of either of these two cysteine residues results in a repressor that is unable to bind to DNA (data not shown). However, when the two cysteine residues were modified by MMTS in the C102S mutant, the protein was still able to bind to DNA (Fig. 9A) even though it lost As(III) binding (Fig. 5), and As(III) did not induce dissociation from the DNA (Fig. 9*C*). Thus, the thiolates of Cys^{95} and Cys^{96} are not required DNA binding.

DISCUSSION

The ArsR/SmtB family of metal-responsive repressors includes members that respond to transition metals, heavy metals, and metalloids (2). Only a few members of the family have been characterized at the biochemical level, and only three (SmtB, CadC, and CmtR) at the structural level (6,18,20). Each has a DNA binding helix-loop-helix at each end of the elongated dimer, and it appears that they all evolved from a common ancestral winged helix protein. On the other hand, the inducer binding sites are diverse. Not only do they have selectivity for different metals and metalloids, but the location and nature of the binding sites are equally diverse, suggesting that, in at least some cases, they evolved independently from each other on the shared repressor structure. The Type 1 As(III) binding site in the

R773 ArsR consists of three cysteine residues, Cys^{32}, Cys^{34}, and Cys^{37}, in or near the first helix of the DNA binding site. Two of the cysteine residues, Cys^{58} and Cys^{60}, in the CadC Cd(II) binding site correspond to Cys^{32} and Cys^{34} of the R773 ArsR, but the other two cysteine residues, Cys^{7} and Cys^{11}, do not correspond to anything in the R773 ArsR. In addition, CadC has a non-regulatory Type 2 harder metal binding site for Zn(II) in the antiparallel C-terminal helices that contribute to the dimer interface, and this site is the same as the regulatory Zn(II) site in SmtB. Thus the Type 1 binding sites of R773 ArsR and CadC are at least partially related to each other, and the Type 2 sites of CadC and SmtB shared a common ancestor, but the Type 1 and Type 2 metal binding sites arose independently. The S_3 Cd(II) binding site in CmtR is even more diverse, bearing no relatedness to the S_4 Cd(II) binding site in CadC.

The As(III)-responsive AfArsR repressor has been described as atypical because it does not contain the C*X*C*XX*C As(III) binding site of the R773 ArsR in the DNA binding site (9). From inspection of the primary sequence of AfArsR and its six closest homologues (Fig. 1), Cys^{95}, Cys^{96}, and Cys^{102}, the only cysteine residues in the protein, are candidate ligands for As(III). In a model of AfArsR built on the structure of CadC (6), Cys^{95} and Cys^{96} are predicted to be at the end of the α5 helices of each monomer that forms a dimerization domain between the two subunits (Fig. 10*A*). The corresponding residues in CadC, SmtB, and CmtR are each in a similar position in helices. In a helix, the minimum distance between the sulfur atoms of a pair of adjacent cysteine residues would be ~4 Å. From the EXAFS results, the distance between the two sulfur atoms and the arsenic atom is 2.25 Å (Fig. 6). Assuming bond angles of 92° typically found in crystal structures of small molecules with AsS_3 sites, the sulfur atoms must be 3.2 Å from each other in the metallated repressor (Fig. 6C). This would require local unraveling of the ends of the α5 helices.

The EXAFS results strongly suggest that Cys^{102} is one of the three sulfur ligands to As(III) in the wild type protein, because mutating this residue causes replacement of a direct sulfur ligand with a hydroxyl oxygen (Fig. 6). This implies that there must be a conformational change in segment of the protein including Cys^{102} that brings it into proximity of Cys^{95} and Cys^{96} to allow for three-coordinate As(III) binding. Because the length of an extended polypeptide from Cys^{96} to Cys^{102} is too short to allow Cys^{102} from one monomer to come into proximity of Cys^{96} of the other monomer, the binding site must be formed by the three cysteines of each monomer (intrasubunit) rather than between subunits such as the Type 1 site of CadC. In the C102S mutant, the sulfur atom is replaced by a hydroxyl group, with a change in the geometry of the metalloid binding site (Fig. 6, *C* and *F*).

The recently reported CmtR structure is instructive (18). CmtR binds Cd(II) at an S_3 site that is composed of three cysteine residues, Cys^{57}, Cys^{61}, and Cys^{102}. Like CadC, this Type 1 Cd(II) binding site is intersubunit, with Cys^{57} and Cys^{61} contributed by one monomer, and Cys^{102} is contributed by the other. Cys^{57} and Cys^{61} are in the recognition helix (αR or α4) of the DNA binding site, and Cys^{102} is in the unstructured and flexible C-terminal region. Cd(II) has been proposed to bind first at Cys^{102}, followed by compaction of the C terminus, with presentation of the bound Cd(II) to the αR cysteines. Thus, Cys^{102} is in effect a sensor of cytosolic metal. Cys^{102} in CmtR and AfArsR are in comparable positions (19).

We considered the possibility that AfArsR Cys^{102} might play a similar role in sensing cytosolic As(III), binding the metalloid, and presenting it to Cys^{95} and Cys^{96} at the end of α5. However, from the EXAFS results, direct binding assays and metalloid-induced dissociation from DNA, the C102S mutant still binds As(III) or Sb(III), albeit with lower affinity than the wild type; so Cys^{102} is not required for metalloid binding to Cys^{95} and Cys^{96}. However, because their thiolates are not close enough to bind As(III) two-coordinately while in the helix, the first step must be binding to either Cys^{95} or Cys^{96} (Fig.

10*B*, step 1). This is proposed to be followed by unraveling the end of the helix and exposing the other thiolate, allowing formation of an S_20 site such as that observed in of the C102S mutant (Fig. 10*B*, step 2). Finally, the thiolate of Cys^{102} would interact with the two-coordinately bound As(III) to become the third ligand (Fig. 10, step 3). Thus, the role of Cys^{102} would be to form a stable S_3 binding site, facilitating dissociation of ArsR from the operator/promoter site on the DNA.

What can we conclude from this comparative analysis of members of the ArsR/SmtB family of metal-regulated repressors? First, the inducer binding site can be located either near the DNA binding domain or the dimer interface. Second, it is formed by placement of pairs or triads of cysteine residues for As(III) or Cd(II). Third, metal binding ligands are often at the ends of helices, and unraveling these helices may be involved in dissociation of the repressor from the operator/promoter DNA. Fourth, the binding sites can be formed within a single subunit or between subunits. Fifth, and most interesting, the As(III) binding sites of R773 ArsR and AfArsR, and the Cd(II) binding sites of CadC and CmtR must be the result of independent and recent evolutionary events, building on the same backbone repressor protein by placement of cysteine residues in diverse locations within the protein.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Bhattacharjee, H.; Rosen, BP. Molecular Microbiology of Heavy Metals. Nies, DH.; Simon, S., editors. Springer-Verlag; Heidelberg/New York: 2007. p. 371-406.
2. Pennella MA, Giedroc DP. Biometals 2005;18:413–428. [PubMed: 16158234]
3. Xu, C.; Rosen, BP. Metals and Genetics. Sarkar, B., editor. Plenum Press; New York: 1999. p. 5-19.
4. Shi W, Wu J, Rosen BP. J Biol Chem 1994;269:19826–19829. [PubMed: 8051064]
5. Shi W, Dong J, Scott RA, Ksenzenko MY, Rosen BP. J Biol Chem 1996;271:9291–9297. [PubMed: 8621591]
6. Ye J, Kandegedara A, Martin P, Rosen BP. J Bacteriol 2005;187:4214–4221. [PubMed: 15937183]
7. Busenlehner LS, Weng TC, Penner-Hahn JE, Giedroc DP. J Mol Biol 2002;319:685–701. [PubMed: 12054863]
8. Butcher BG, Deane SM, Rawlings DE. Appl Environ Microbiol 2000;66:1826–1833. [PubMed: 10788346]
9. Butcher BG, Rawlings DE. Microbiology 2002;148:3983–3992. [PubMed: 12480902]
10. Sambrook, J.; Fritsch, EF.; Maniatis, T. Molecular Cloning, A Laboratory Manual. Cold Spring Harbor Laboratory; Cold Spring Harbor, NY: 1989.
11. Laemmli UK. Nature 1970;227:680–685. [PubMed: 5432063]
12. Bradford MM. Anal Biochem 1976;72:248–254. [PubMed: 942051]
13. Gill SC, von Hippel PH. Anal Biochem 1989;182:319–326. [PubMed: 2610349]
14. Bhattacharjee H, Li J, Ksenzenko MY, Rosen BP. J Biol Chem 1995;270:11245–11250. [PubMed: 7744758]
15. Ellman GL. Arch Biochem Biophys 1959;82:70–77. [PubMed: 13650640]
16. Ramirez-Solis A, Mukopadhyay R, Rosen BP, Stemmler TL. Inorg Chem 2004;43:2954–2959. [PubMed: 15106984]
17. Carlin A, Shi W, Dey S, Rosen BP. J Bacteriol 1995;177:981–986. [PubMed: 7860609]
18. Banci L, Bertini I, Cantini F, Ciofi-Baffoni S, Cavet JS, Dennison C, Graham A, Harvie DR, Robinson NJ. J Biol Chem 2007;282:30181–30188. [PubMed: 17599915]
19. Cavet JS, Graham AI, Meng W, Robinson NJ. J Biol Chem 2003;278:44560–44566. [PubMed: 12939264]

20. Cook WJ, Kar SR, Taylor KB, Hall LM. J Mol Biol 1998;275:337–346. [PubMed: 9466913]
21. Thompson JD, Higgins DG, Gibson TJ. Nucleic Acids Res 1994;22:4673–4680. [PubMed: 7984417]
22. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. Annu Rev Biophys Biomol Struct 2000;29:291–325. [PubMed: 10940251]
23. DeLano, WL. The PyMOL User's Manual. DeLano Scientific; San Carlos, CA: 2001.

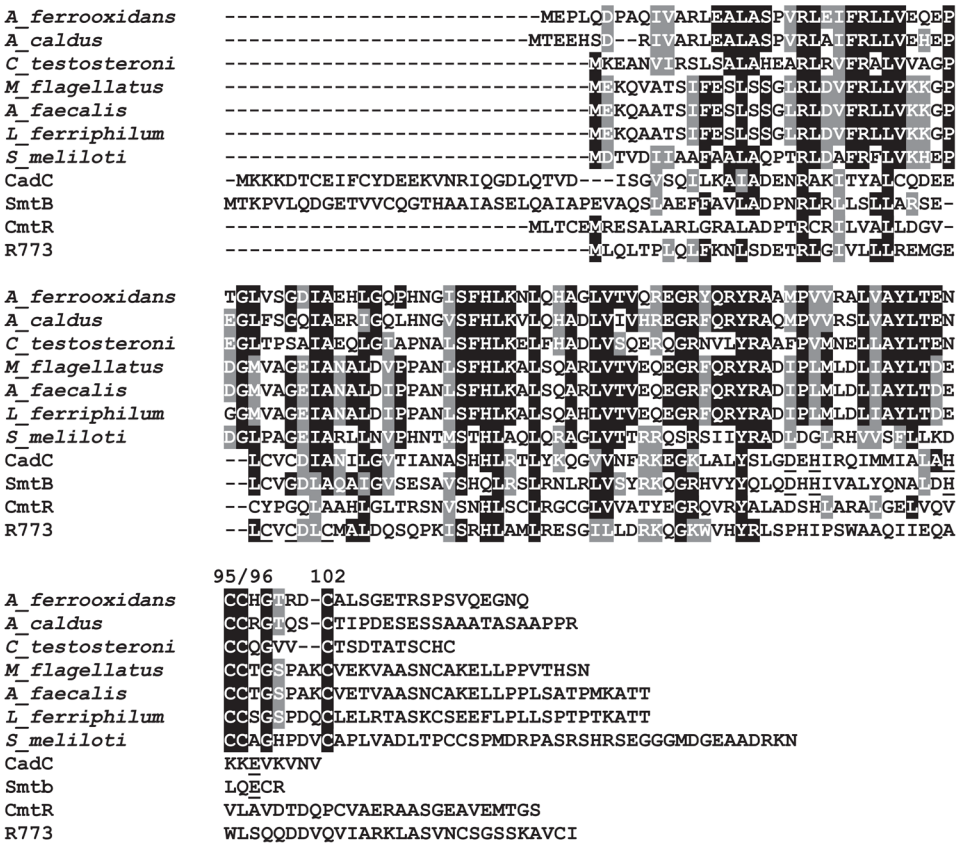

FIGURE 1. Multiple alignment of AfArsR homologues

Representative ArsR homologues (accession numbers in *parentheses*) are from: *A. ferrooxidans* (AAF69241), *Acidothiobacillus caldus* (AAX35682), *Comamonas testosteroni* KF-1 (ZP_01519260), *Methylobacillus flagellatus* KT (ABE49761), *Alcaligenes faecalis* (AAS45114), *Leptospirillum ferriphilum* (AAY85166), *Sinorhizobium meliloti* 1021 (NP_385183), pI258 CadC (P20047), *Synechocystis sp*. PCC 6803 SmtB (BAA10706), *Mycobacterium tuberculosis* H37Rv CmtR (NP_216510), and plasmid R773 ArsR (P15905). Cysteine residues 95, 96, and 102 are indicated. The multiple alignment was calculated with CLUSTAL W (21).

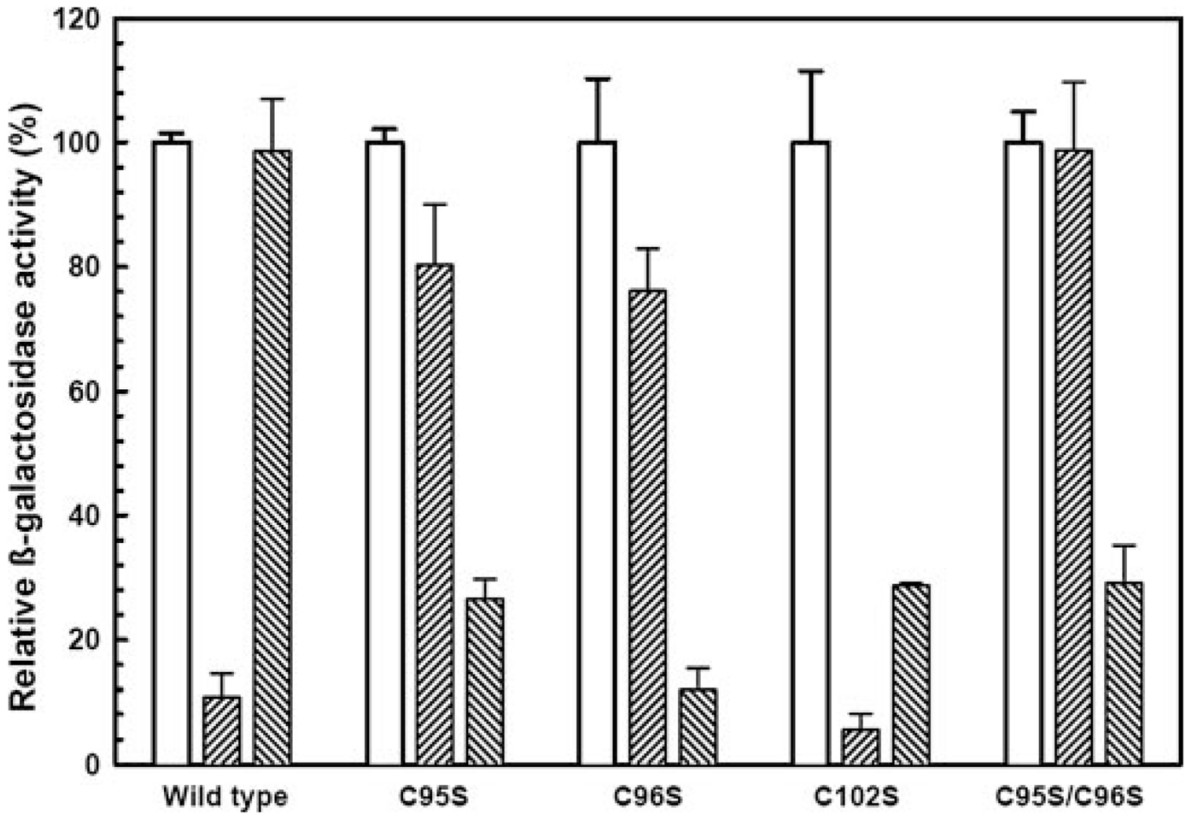

FIGURE 2. Contribution of AfArsR cysteine residues to metalloregulation *in vivo*

Expression of a *lacZ* reporter gene was assayed as described under "Materials and Methods." Cells of *E. coli* strain ACSH501^{q} bearing plasmids with wild type *arsR*, C95S, C96S, C102S, or the C95S/C96S double mutant in *trans* with reporter plasmid pACYC184*lacZ-arsO* were grown with 0.2% glucose (*left bars*), 0.2% arabinose (*middle bars*), or 0.2% arabinose and 25 μM sodium arsenite (*right bars*).

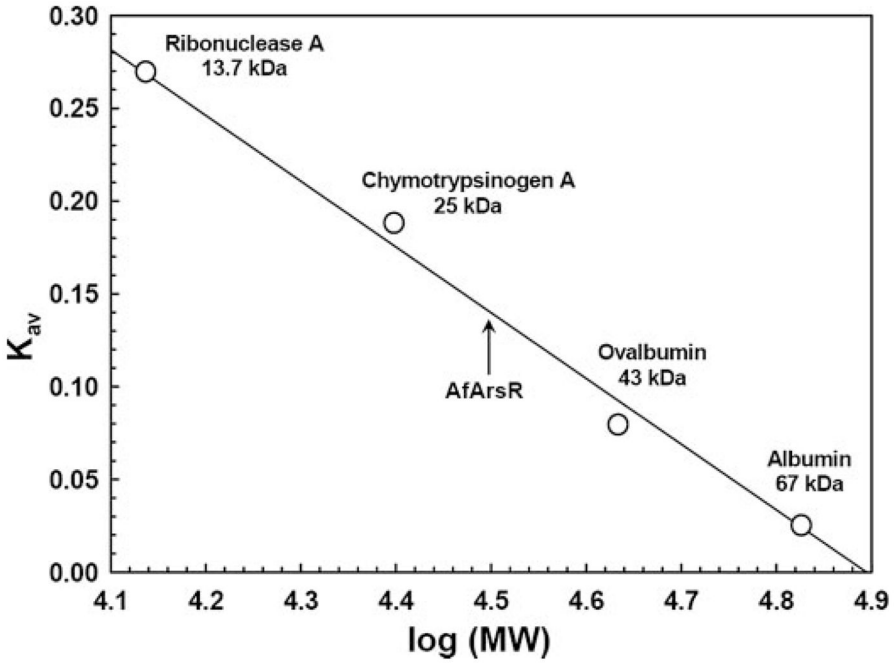

FIGURE 3. Determination of the molecular mass of AfArsR

Proteins were separated by gel filtration, as described under "Materials and Methods." Molecular weight markers were ribonuclease, chymotrypsinogen A, ovalbumin, and albumin. The elution position of wild type and mutant AfArsRs is indicated by the *arrow*.

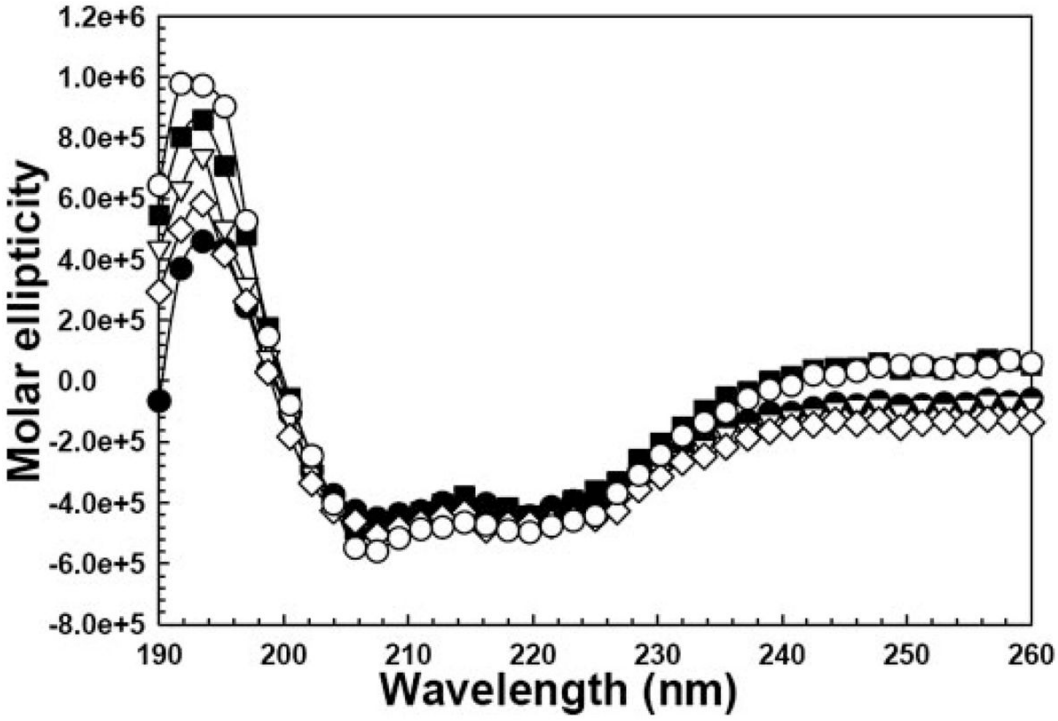

FIGURE 4. Circular dichroism spectroscopy of wild type and mutant AfArsRs

CD spectra were collected with wild type AfArsR (○), or wild type AfArsR with 30 μM sodium arsenite (●), C95S/C96S (▼), C96S (△), and C102S (□), respectively. The differences below 200 nm are likely due to slight differences in protein concentration.

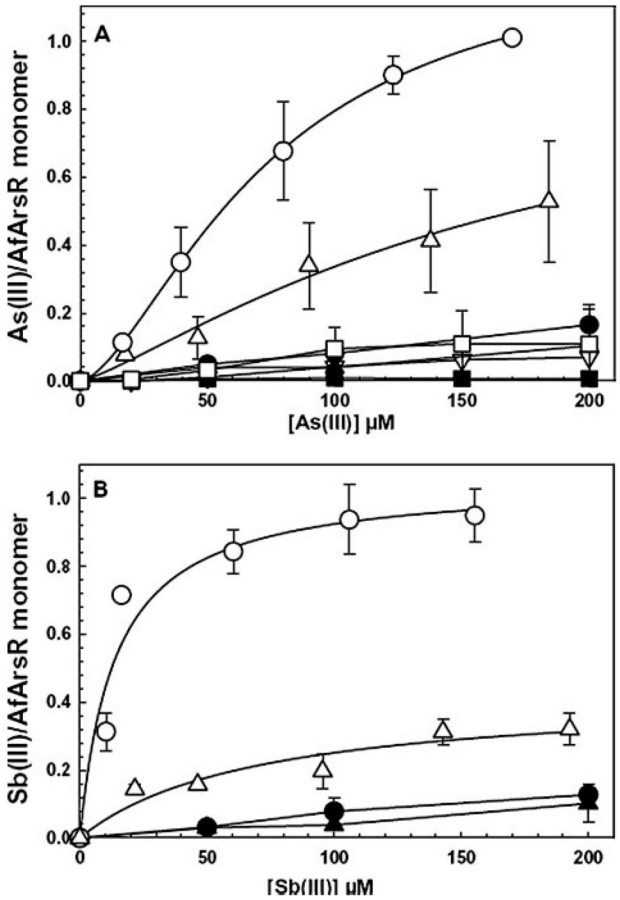

FIGURE 5. Metalloid binding to AfArsR

Binding of (*A*) sodium arsenite or (*B*) potassium antimonyl tartrate was assayed as described under "Materials and Methods" to wild type AfArsR (○), MMTS-modified wild type AfArsR (●), C102S (△), MMTS-modified C102S (▲), C95S (□), C96S (▽), and C95S/C96S double mutant (■).

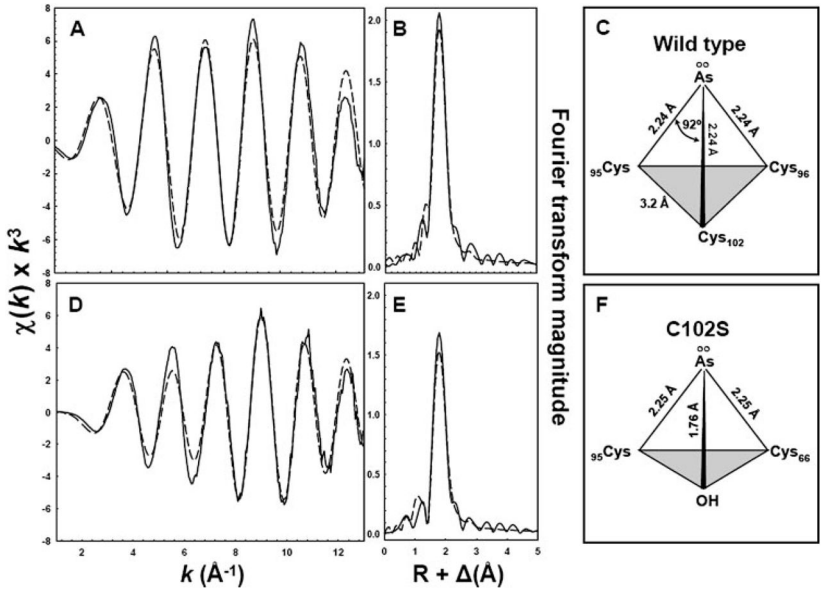

FIGURE 6. EXAFS and Fourier transforms of ArsR XAS data

EXAFS spectra of (*A*) wild type AfArsR and (*D*) C102S, both with bound As(III). The corresponding Fourier transforms (*B* and *E*) are shown as *solid lines*, with the simulations of EXAFS and Fourier transform data shown as *dashed lines*. Sample preparation was as described under "Materials and Methods." A summary of the fitting results is shown in Table 1. *C* and *F* show models of the A_s-S_3 site of the wild type and the A_S-S_2O site of C102S, respectively.

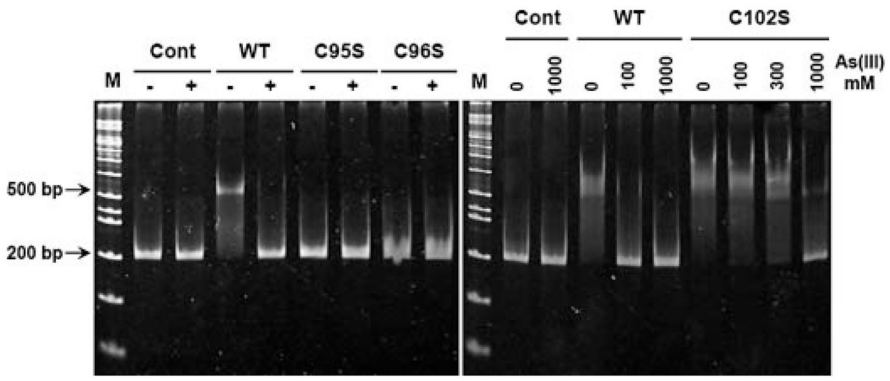

FIGURE 7. AfArsR binding to DNA assayed by mobility shift assays

DNA binding assays were performed as described under "Materials and Methods." *Left*, binding in the presence (+) or absence (−) of 0.1 mM sodium arsenite. *Right*, binding in the presence of varying concentrations (0, 100, 300, or 1000 mM) of sodium arsenite. Added proteins were none (control), wild type AfArsR, C95S, C96S, or C102S. *Lanes M* show a DNA ladder, with representative sizes indicated.

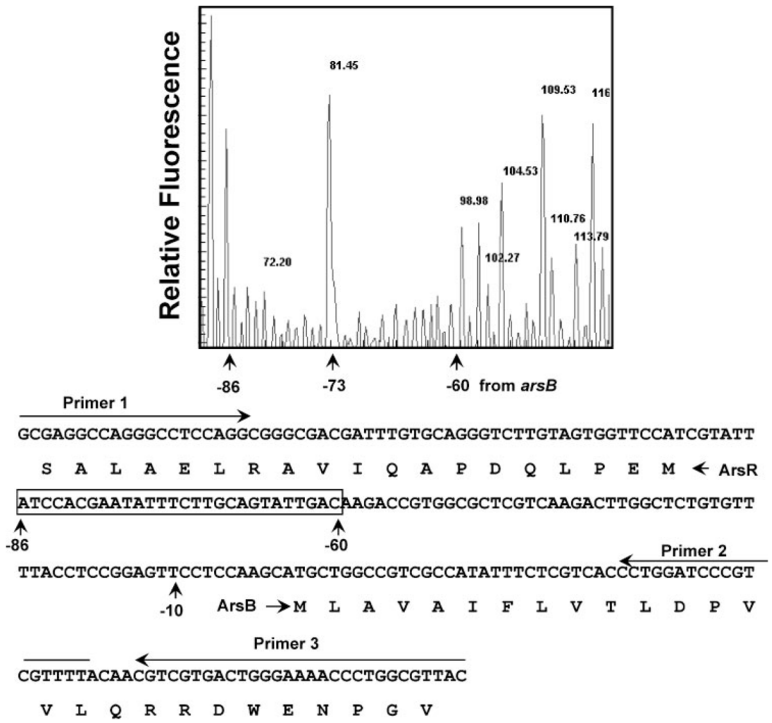

FIGURE 8. DNA footprinting

Top, DNase I footprinting was performed as described under “Materials and Methods” using light-sabre green fluorescently labeled double-stranded DNA. The sizes of major DNA fragments in nucleotides are shown at the *top* of the peaks. The sequence protected by AfArsR is shown *between* the *arrows*, with the positions of −60, −73, and −86 nucleotides relative to the start of the *arsB* gene indicated. *Bottom*, the sequence of the double-stranded DNA is shown, with the region protected by AfArsR *boxed*. The start sites for the *arsR* and *arsB* genes are indicated, with the −10, −60, −73, and −86 nucleotide positions relative to the start of *arsB* indicated.

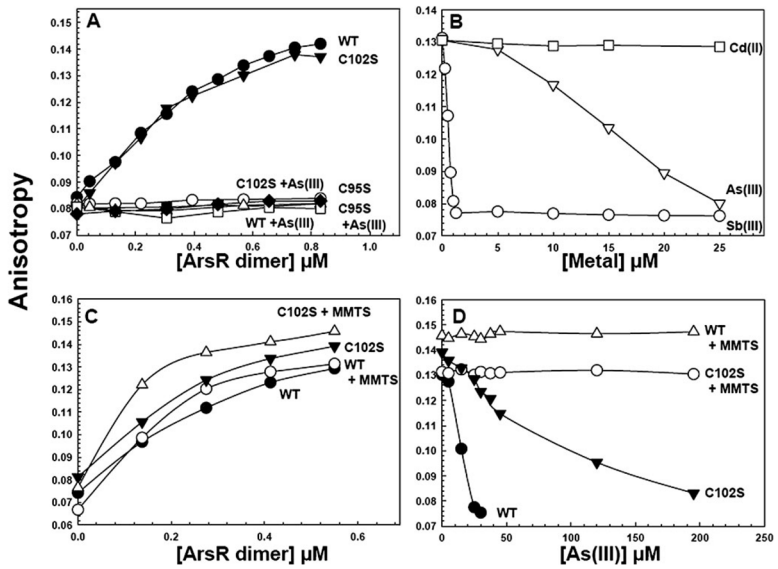

FIGURE 9. AfArsR binding to DNA assayed by fluorescence anisotropy

DNA binding assays were performed as described under "Materials and Methods." *A*, fluorescently labeled double-stranded DNA was titrated with the indicated concentrations of wild type AfArsR (●), wild type AfArsR plus a 6-fold excess of sodium arsenite (○), C102S (▲), C102S plus a 6-fold excess of sodium arsenite (▽), C95S (■), C95S plus a 6-fold excess of sodium arsenite (□). *B*, double-stranded DNA with bound wild type AfArsR was titrated with the indicated concentrations of potassium antimonyl tartrate (○), sodium arsenite (▽), or cadmium chloride (□). *C*, fluorescently labeled double-stranded DNA was titrated with the indicated concentrations of wild type AfArsR (●), MMTS-modified wild type AfArsR (○), C102S (▼), and MMTS-modified C102S (△). *D*, double-stranded DNA with bound wild type AfArsR (●), MMTS-modified wild type AfArsR (△), C102S (▼), or MMTS-modified C102S (○) was titrated with the indicated concentrations of sodium arsenite.

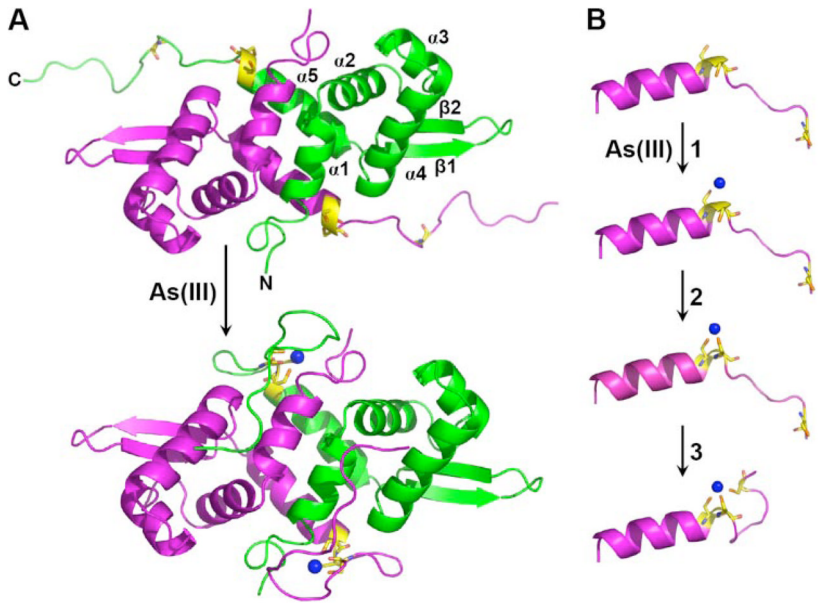

FIGURE 10. As(III) binding model of AfArsR

Modeling of As(III)-free AfArsR was performed with the Modeler 8v1 auto mode (22), using the crystal structure of CadC (6) as template. The model of the As(III)-bound form was built by manually adjusting the C termini so that the geometry of As(III) coordination from Fig. 6 was satisfied. The two monomers are colored *red* and *green*, respectively. The cysteines are shown in *sticks*. The As(III)s are shown as *blue spheres*. The model was drawn using PyMOL (23). *A*, the aporepressor binds two arsenic atoms at the ends of the *α*5 helix. *B*, binding is proposed to occur in three steps: 1) As(III) binds first to the thiolate of either Cys^{95} or Cys^{96} in the *α*5 helix; 2) the end of the helix unravels to allow the adjacent cysteine residue to become a second ligand to As(III), forming a low affinity S_2O site; and 3) Cys^{102} in the flexible C terminus forms the third ligand to the As(III), forming a high affinity S_3 site.

TABLE 1

Summary of AfArsR EXAFS fitting results

		Ligand environment[a]								
Sample	**Fit No.**	**Atom[b]**	**R (Å)[c]**	**C.N.[d]**	**σ^2[e]**	**Atom[b]**	**R (Å)[c]**	**C.N.[d]**	**σ^2[e]**	**F'[f]**
Wild type	1[g]	O/N	2.08	3.0	0.35					4.72
	2[g]	S	2.24	3.0	2.91					0.50
C102S	1[g]	O/N	2.09	3.0	0.85					3.20
	2[g]	S	2.25	2.5	3.03					0.61
	3[g]	O/N	1.76	1.0	3.67	S	2.24	2.0	1.91	0.34

[a] Independent metal-ligand scattering environment.

[b] Scattering atoms: O (oxygen), N (nitrogen), and S (sulfur).

[c] Metal-ligand bond length in A.

[d] Metal-ligand coordination number.

[e] Debye-Waller factor in Å × 10^3.

[f] Number of degrees of freedom weighted mean square deviation between data and fit.

[g] Fit using only single scattering Feff 7 theoretical models.