Secondary Structure and Fold Homology of the ArsC Protein from the *Escherichia coli* Arsenic Resistance Plasmid R773†

Shawn Y. Stevens,‡ Weidong Hu,‡,§ Tatiana Gladysheva,||,⊥ Barry P. Rosen,|| Erik R. P. Zuiderweg,*,‡,○ and Lana Lee*,#

Biophysics Research Division and the Departments of Biological Chemistry and Chemistry, University of Michigan, 930 North University Avenue, Ann Arbor, Michigan 48109-1055, Department of Chemistry and Biochemistry, University of Windsor, Windsor, Ontario, Canada N9B 3P4, and the Department of Biochemistry, Wayne State University School of Medicine, 540 East Canfield Street, Detroit Michigan 48201

Received February 11, 1999; Revised Manuscript Received May 19, 1999

ABSTRACT: Resistance to several toxic anions in *Escherichia coli* is conferred by the *ars* operon carried on plasmid R773. The gene products of this operon catalyze extrusion of antimonials and arsenicals from cells. In this paper, we report the determination of the overall fold for ArsC, a 16 kDa protein of the *ars* operon involved in the reduction of arsenate to arsenite, using multidimensional, multinuclear NMR. The protein is found to contain large regions of extensive mobility, particularly in the active site. A model fold, computed on the basis of a preliminary set of NOEs, was found to be structurally homologous to *E. coli* glutaredoxin, thiol transferases, and glutathione S-transferase. Some kinship to the structure of low molecular weight tyrosine phosphatases, based on rough topological similarity but more so on the basis of a common anion-binding-loop motif H−CX_nR, was also detected. Although functional, secondary, and tertiary structural homology is observed with these molecules, no significant homology in primary structure was detected. The mobilities of the active site of ArsC and of other enzymes are discussed.

The *ars* operon of *E. coli*[1] R-factor 773 confers resistance to a variety of heavy metals, including arsenate, arsenite, antimonite, and tellurite (*1*, *2*). The operon encodes five proteins, A, B, C, D, and R, that collectively mediate this resistance. ArsR and ArsD are DNA-binding proteins that regulate expression of the *ars* operon. The ArsA protein is a heavy-metal-stimulated ATPase that associates as a dimer with the membrane-binding protein ArsB. Together, the two proteins act as an oxyanion pump, extruding heavy metals from the cell, to maintain subtoxic levels. These two proteins alone mediate resistance to arsenite and antimonite (*3*). Resistance to arsenate requires another protein, ArsC (*3*). This 16 kDa protein has been implicated in the reduction of arsenate (As^V) to arsenite (As^{III}), the only form of arsenic that is accepted by the oxyanion pump (*4*).

Proteins homologous to the *ars* operon proteins have been identified throughout the animal kingdom. However, while homologues to the proteins involved in the anion pump and regulatory mechanism are found in higher organisms, homologues to ArsC have only been found in selected bacterial strains. Several strains, including *E. coli* R46 and *Yersinia enterocolitica*, express proteins with ≥84% sequence identity to R773 ArsC. In contrast, ArsC proteins from other strains have little sequence identity with the R773 protein, although they are also involved in arsenate reduction. For example, the ArsC protein of the *ars* operon of pI258 of *Staphylococcus aureus*, which has been implicated in cell resistance to arsenate as an arsenate reductase, exhibits <20% sequence identity to the R773 ArsC protein (*4*, *5*). While this sequence identity is low, the two proteins share an anion binding motif, H−CX_nR, that is commonly observed in proteins, such as phosphatases and sulfatases (*6*). The disparity of sequence, and most likely structure, but similarity of function between these two ArsC proteins in combination with the absence of similar proteins in higher organisms poses an intriguing evolutionary question.

In vitro reduction of arsenate ($As^VO_4^{3-}$) to arsenite ($As^{III}O_2^-$) requires ArsC, glutaredoxin, and reduced glutathione with a pH optimum between 6.3 and 6.8. ArsC has two cysteinyl residues at positions 12 and 106. Mutagenesis experiments have shown that of those cysteines, only residue Cys-12 is indispensable for protein function (*7*). The cysteinyl side chain of residue 12 has an unusually low pK_a of 6.4 that is attributed to an electrostatic interaction with residue His-8 (*8*), corroborating the pH dependence of the in vitro activity assays. Similar observations were made for Cys-403 and His-402 of a tyrosine phosphatase of genus *Yersinia*. In this case, the Cys is the nucleophile that attacks

† This work was supported in part by the Natural Sciences and Engineering Research Council of Canada, and NIH Grants 5R01 GM 52406 and GM 52216.

* To whom correspondence should be addressed. L.L.: Fax 519-978-7098. E-mail llee@server.uwindsor.ca. For E.R.P.Z.: Fax 734-764-3323. E-mail zuiderwe@umich.edu.

‡ Biophysics Research Division, University of Michigan.

§ Present Address: Memorial Sloan-Kettering Cancer Center, Box 557, New York, New York 10021.

|| Wayne State University School of Medicine.

⊥ Present Address: Proscript, Inc., 38 Sydney Street, Cambridge, Massachusetts 02139.

○ Departments of Biological Chemistry and Chemistry, University of Michigan.

University of Windsor.

[1] Abbreviations: *Escherichia coli*, *E. coli*; NMR, nuclear magnetic resonance; HSQC, heteronuclear single-quantum coherence; HMQC, heteronuclear multiple-quantum coherence; NOE nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; TOCSY, total correlation spectroscopy; WATERGATE, water suppression by gradient-tailored excitation; DTT, dithiothreitol.

10.1021/bi990333c CCC: $18.00
Published on Web 07/02/1999

the scissile tyrosyl−phosphorus bond and becomes the phosphate anion binding site (*9*).

No three-dimensional structures are known for any of the ArsC proteins. While three-dimensional structures have been determined for several phosphatases and sulfatases containing the signature anion-binding-loop H−CX_nR (*6*, *10*−*12*), no significant amino acid homology could be detected for ArsC with these or any other proteins of known three-dimensional structure. Experimental structural information for ArsC should provide insight into the interaction between the protein and arsenite (or arsenate) and describe how ArsC relates to other anion-binding proteins. In addition, these studies may help elucidate the exact function of ArsC in the arsenate resistance pathway and explain why it is unnecessary in higher organisms. Together, this information may contribute to the understanding of the evolution of bacterial resistance to arsenic and heavy metals in general.

In this paper, we report the determination of the secondary structure and overall protein fold for ArsC, using multidimensional, multinuclear NMR. The protein is found to contain large regions of mobility on the millisecond time scale, as evidenced by line broadening and multiple resonances for single residues, most notably in the active-site area. Nevertheless, NMR assignments and secondary structure could be obtained for 128 of the 141 residues. A model fold, computed on the basis of a preliminary set of 300 interbackbone NOEs, was found to be structurally homologous to *E. coli* glutaredoxin (*13*), thiol transferases (*14*), and glutathione S-transferase (*15*). Some similarity to the structure of low molecular weight tyrosine phosphatases was also detected (*11*) on the basis of the presence of a common anion-binding-loop motif, H−CX_nR, in addition to rough topological homology. While ArsC possesses functional as well as secondary and tertiary structural homology to several proteins, significant homology in the primary structure, even in the structurally homologous elements of these molecules, was absent.

MATERIALS AND METHODS

Sample Preparation. Uniformly labeled ^{15}N-ArsC was prepared by growing *E. coli* strain JM109 (DE3) in M9 minimal media with $^{15}NH_4Cl$ (1 g/L) as the sole nitrogen source. Uniformly labeled ^{15}N/^{13}C-ArsC was prepared by growing the cells in M9 minimal media with $^{15}NH_4Cl$ (1 g/L) as the sole nitrogen source and ^{13}C6-glucose (3 g/L) as the sole carbon source. Triply labeled ^{2}H/^{13}C/^{15}N-ArsC was prepared using the same media described above, substituting $^{2}H_2O$ for H_2O. All stable isotopes were obtained from Cambridge Isotopes Laboratories. The protein was purified using the procedures described previously (*16*) and was better than 95% pure as judged from gel electrophoresis. The protein was found to be active in the in vitro arsenate reduction assay (*16*). The purified protein was dissolved in 90:10% H_2O/D_2O containing 50 mM NaCl, pH 6.6, 1 mm DTT, and 0.02% sodium azide. The triply labeled protein was exchanged in H_2O; a two-dimensional ^{15}N/^{1}H-HSQC spectrum was used to assess complete amide exchange.

NMR Spectroscopy. All NMR spectra were acquired on Bruker AMX 500 and AMX 600 spectrometers equipped with a B_0-field gradient triple resonance probes. The experiments involving ^{15}N- and ^{13}C/^{15}N-labeled samples at concentrations of 0.8−1.2 mM were carried out using 5 mm probes. Experiments on ^{15}N- and ^{2}H/^{13}C/^{15}N-labeled samples at concentrations of 2 mM were performed using Nalorac 8 mm probes. All spectra were acquired at 20 °C to prevent denaturation of ArsC. Water suppression was achieved with a WATERGATE sequence prior to acquisition (*17*). Where possible, water-up technology was employed. Three-dimensional ^{15}N-resolved NOESY−HSQC (*18*, *19*) experiments were recorded using a mixing time of 130 ms. 3D-HNCA and 3D-HN(CO)CA provided the intra- (n) and interresidue (n − 1) C_α frequencies, respectively (*20*). In the same manner, the 3D-HN(CA)HA (*21*) and HA(CACO)-NH (*22*, *23*) provided the intra- (n) and interresidue (n − 1) H_α frequencies. A double match between an intraresidue C_α/H_α frequency pair and an interresidue (n − 1) C_α/H_α pair established the connection between two residues. A detailed description of this method can be found in ref *22*. In addition, side chain type information was obtained from the 3D-cross-polarization-driven (H)CC(CACO)NH-TOCSY (*24*).

Experiments carried out on the triple-labeled sample using the Bruker AMX 600 spectrometer equipped with a 8 mm Nalorac probe were the quadruple resonance 3D-spectra HNCA, HN(CO)CA, HNCACB, and HN(CO)CACB, using slightly modified versions of the sequences described by Yamazaki et al. (*25*). Deuterium decoupling was achieved using a home-built fourth channel (*26*), using a WALTZ-16 modulation scheme at a rf field of 0.8 kHz. The 3D ^{15}N/^{15}N/^{1}H-HMQC−NOESY−HSQC spectra were recorded using mixing times of 100, 400, and 800 ms on the deuterated sample.

Amide exchange data were acquired from a sample in D_2O containing 50 mM NaCl, pH 6.5 at 20 °C. Thirteen ^{15}N/^{1}H-HSQC spectra were taken between 1 and 36 h after the sample was rapidly exchanged into D_2O. The acquisition parameters for the experiments are summarized in Table 1.

All NMR spectra were processed with Felix 2.0 software (a gift from Hare, Inc.) augmented with in-house-written baseline correction routines. Data analysis was performed using RGO, a quick contour viewing program (a gift from Glaxo-Welcome) and Felixtalk, a program written in our laboratory to interface with FELIX in combination with a suite of UNIX NAWK programs. The program TypeProb aided the assignment of residue type (*27*). The program SeqProb was used to correlate the probabilities for sequential assignments (*27*).

Preliminary Structure Calculation. The distance restraints acquired from NOESY data analysis were used to calculate a low-resolution structure using the distance geometry (DGII) and simulated annealing (Discover) from Molecular Simulations, Inc. The calculations incorporated 49 hydrogen bond restraints, 144 sequential NOEs, 117 medium-range NOEs, and 54 long-range NOEs with 128 dihedral restraints.

RESULTS

ArsC is a stable monomer at high concentrations under conditions of optimal activity (pH 6.5), the most important being the presence of reductants, such as DTT, and maintaining a temperature of ≤20 °C. A variety of line widths were observed throughout the protein for the cross-peaks in the ^{15}N/^{1}H HSQC fingerprint, which is suggestive of regions of varying mobility. In gel filtration experiments, the 141-residue, nominally 16 kDa protein runs similarly to a 22 kDa globular protein, which suggests that the protein is non-

Table 1: Acquisition Parameters for NMR Experiments Performed on ArsC

	acquired data (complex points)			spectral widths (Hz)		
experiment	t1	t2	t3	F1	F2	F3
^{15}N/^{1}H-HSQC	200 (^{15}N)	2048 (^{1}H)			10,000 (^{1}H)	
^{15}N-HSQC−NOESY	80 (^{15}N)	320 (^{1}H)	2048 (^{1}H)	1666.7 (^{15}N)	6666.7 (^{1}H)	8064.5 (^{1}H)
^{15}N/^{15}N/^{1}H-HMQC−NOESY−HSQC	100 (^{15}N)	100 (^{15}N)	2048 (^{1}H)	1666.7 (^{15}N)	1666.7 (^{15}N)	10,000 (^{1}H)
HNCO	60 (^{15}N)	240 (^{13}CO)	2048 (^{1}H)	1666.7 (^{15}N)	4347.8 (^{13}C)	8333.3 (^{1}H)
HNCA	60 (^{15}N)	72 (^{13}CA)	2048 (^{1}H)	1666.7 (^{15}N)	3846 (^{13}CA)	10,000 (^{1}H)
HN(CO)CA	60 (^{15}N)	72 (^{13}CA)	2048 (^{1}H)	1666.7 (^{15}N)	3846 (^{13}CA)	10,000 (^{1}H)
HNCAHA	52 (^{15}N)	72 (^{13}CA)	2048 (^{1}H)	1381.2 (^{15}N)	1851.9 (^{13}CA)	8064.5 (^{1}H)
HNCACB	60 (^{15}N)	72 (^{13}CA,CB)	2048 (^{1}H)	1666.7 (^{15}N)	3846 (^{13}CA,CB)	10,000 (^{1}H)
HN(CO)CACB	60 (^{15}N)	72 (^{13}CA,CB)	2048 (^{1}H)	1666.7 (^{15}N)	3846 (^{13}CA,CB)	10,000 (^{1}H)
HCCCACONH-TOCSY	44 (^{15}N)	150 (^{13}C)	2048 (^{1}H)	1666.7 (^{15}N)	3703.7 (^{13}C)	8333.3 (^{1}H)
CCH-COSY	216 (^{13}C)	216 (^{13}C)	2048 (^{1}H)	3703.7 (^{13}C)	3703.7 (^{13}C)	8333.3 (^{1}H)
CCH-NOESY	124 (^{13}C)	124 (^{13}C)	2048 (^{1}H)	1136.4 (^{13}C)	1136.4 (^{13}C)	6024.1 (^{1}H)
CCH-TOCSY	216 (^{13}C)	216 (^{13}C)	2048 (^{1}H)	3703.7 (^{13}C)	3703.7 (^{13}C)	8333.3 (^{1}H)

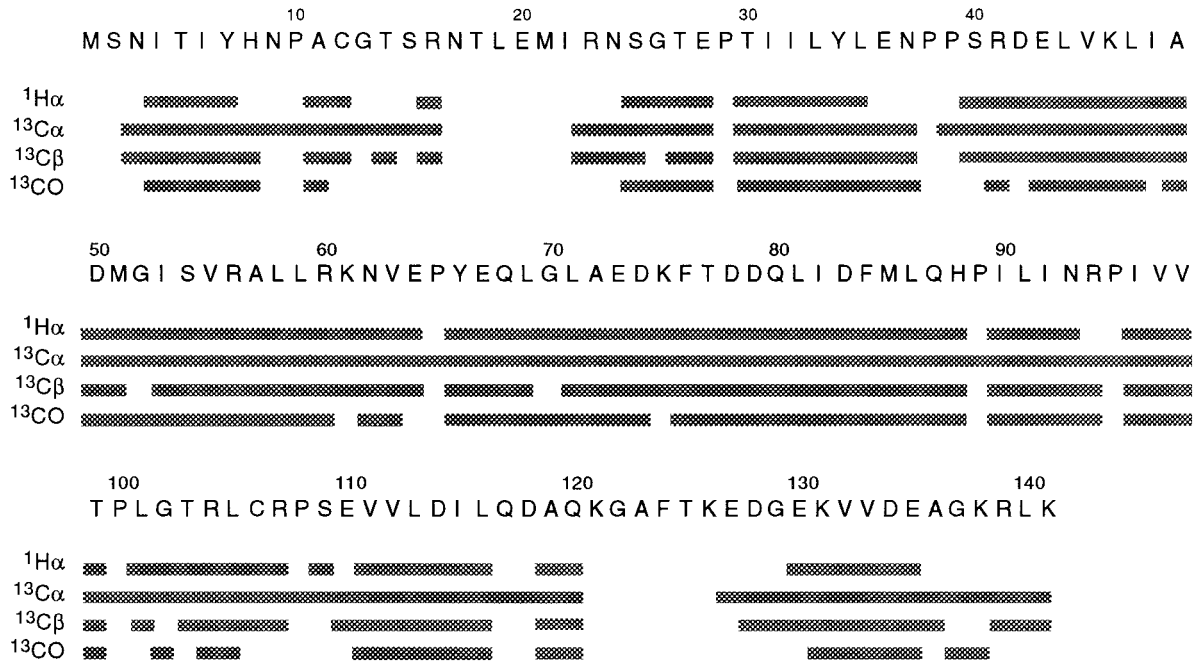

FIGURE 1: Summary of observed scalar connectivities. The first line indicates the sequential H_α derived from the HNCAHA experiment. The second and third lines represent the connectivities between sequential C_α and C_β observed from the HNCA and HN(CO)CA and HNCACB and HN(CO)CACB experiments, respectively. The fourth line indicates assignments from the HNCO spectrum.

globular. The ^{15}N-^{1}H HSQC spectrum at 11.7 T (500 MHz for protons) contains 142 resonances, more than the expected 133 non-proline residues. As all biochemical assays indicate that the protein preparations are better than 95% pure, it follows that multiple conformations, interchanging on very slow time scales, are present. Many of the HSQC crosspeaks are observed in the region between 8 and 8.5 ppm and exhibit significant overlap, which is often a manifestation of poorly structured dynamic regions. Several of the HSQC crosspeaks disappear while using a 14.6 T NMR instrument (600 MHz for protons), indicating severe resonance broadening caused by intermediate conformational exchange processes on fairly slow time scales (0.001−1 s). Together, these results suggest that ArsC is not a rigid, globular protein but exhibits regions of conformational heterogeneity that may give rise to hydrodynamic properties that result in the high apparent molecular weight. Altering temperature or protein concentration afforded no improvement.

Resonance Assignments. A large suite of triple resonance experiments were performed on ^{15}N-, ^{13}C/^{15}N-, and ^{13}C/^{15}N/^{2}H-labeled ArsC. Sequential assignments were initiated on ^{15}N,^{13}C-labeled ArsC using ^{15}N-NOESY−HSQC, HNCA, HN(CO)CA, HN(CA)HA, and HA(CACO)NNH spectra to establish connectivities between consecutive C_α and H_α resonances. Glycines were identified readily on the basis of the C_α chemical shifts. A cross-polarization-driven (H)CC-(CO)NNH-TOCSY spectrum was used to assign the side chain resonances of several sequentially connected residues and allowed residues adjacent to serines, alanines, and threonines to be identified from their characteristic C_β chemical shifts. Other residue type identifications were made on the basis of combined C_α and C_β chemical shifts (*27*). At this stage, only about 60% of the resonances could be unambiguously assigned. The unassigned residues were contained predominantly in two stretches between residues 11 and 29, a region that contains the active site, and the 25 residues of the carboxyl-terminus. Addition of substrate (arsenate) or product (arsenite) resulted in loss of resonances or precipitation, respectively.

There have been several examples that employed deuteration to attain complete assignment of difficult proteins (*28*−*30*). The significant line narrowing of the ^{13}C resonances

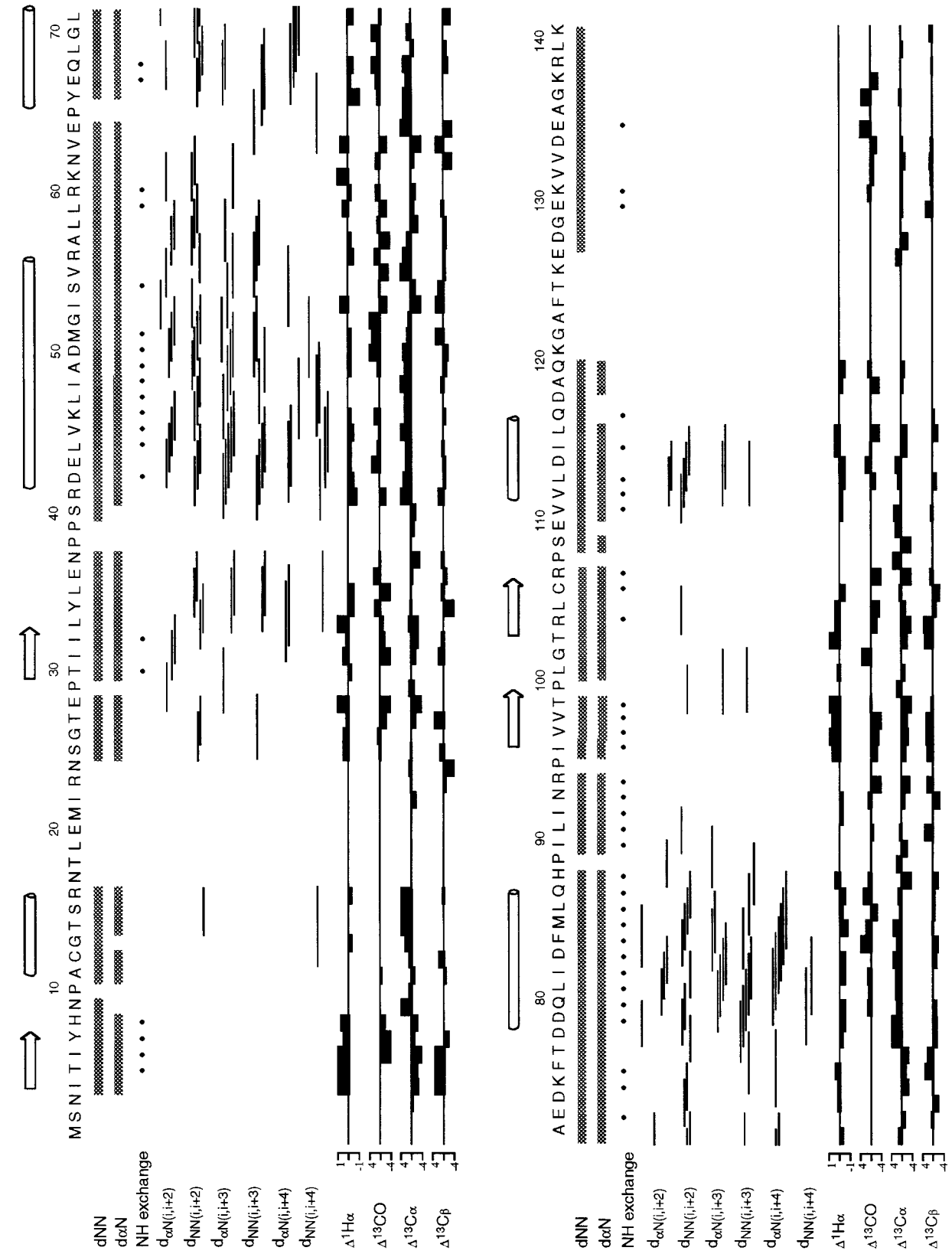

FIGURE 2: Summary of sequential and medium-range NOEs between the $C_\alpha H$ and amide proton and between the backbone amide protons themselves. Moderately slow amide proton exchange was defined to be $t_{1/2} > 1$ h and is represented by filled dots. Chemical shift index analysis for H_α, C_α, C_β, and CO chemical shifts is depicted. The predicted β-strands from residues 25−28 and 73−77 appear to be just extended conformations as there is no NOE data to suggest participation in a β-sheet.

afforded by the deuterated C_α sites allows HNCACB and HN(CO)CACB experiments to be acquired with high sensitivity. This provided the opportunity to confidently assign the intraresidue and sequential C_β resonances in addition to the C_α resonances observed in the HNCA and HN(CO)CA spectra. The previous assignments were confirmed and extended with connectivities obtained through the C_α and C_β resonances. Sequentially linked stretches were specifically

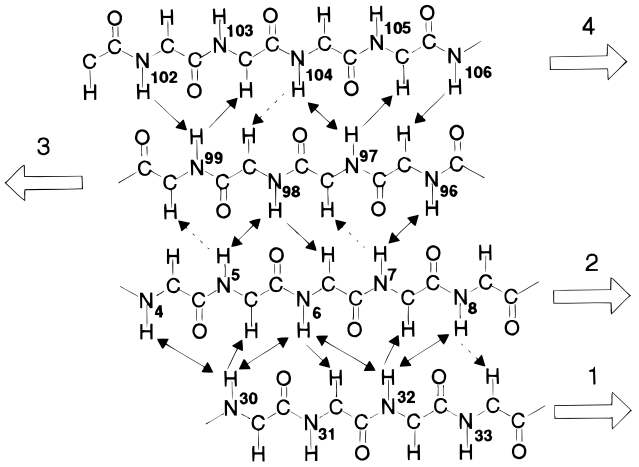

FIGURE 3: Connectivity diagram of the four-stranded mixed β-sheet as determined from long-range NOEs. Observed NOEs are indicated by solid arrows. Dashed arrows indicate that the absence or presence of the NOE could not be ascertained due to overlap in the spectra.

assigned on the basis of C_α and C_β chemical shifts. However, while deuteration significantly improved sensitivity, complete assignment of ArsC could not be obtained. For primarily two regions, N17−I22 and K121−E127, connectivities were not observed and the resonances could not be assigned unambiguously (Figure 1).

Secondary Structure Analysis. A series of 3D $^{15}N/^{15}N/^{1}H$-HMQC−NOESY−HSQC experiments with mixing times ranging from 100 to 800 ms and a 3D $^{1}H/^{1}H/^{15}N$-NOESY-HSQC were acquired on deuterated and protonated ArsC, respectively. The increased sensitivity of these experiments provided extensive information on the secondary and tertiary structure of the protein. The secondary structure of ArsC was derived primarily from the medium-range amide−amide proton [dNN(i,$i\pm n$)] and α-amide-proton [dαN (i,$i+n$)] NOEs. Amide proton exchange rates as determined from a time series of $^{15}N/^{1}H$-HSQC spectra of ArsC freshly exchanged into $^{2}H_2O$ were used to corroborate the interpretation of data derived from the NOESY spectra (Figure 2).

ArsC consists of a relatively rigid core with two large, mobile regions that gave rise to the unassigned resonances. An α/β protein, ArsC contains five α helices and a four-stranded mixed β-sheet. The sheets were identified by relatively strong dαN (i,i+1) NOEs, in addition to diagnostic interstrand NOEs (Figure 3). The interstrand NOEs proved that the β-sheet consists of one parallel strand (**1**) and three antiparallel strands (**2**, **3**, and **4**) (Figure 3). The central strands exhibit moderately and uniformly slow amide proton exchange (see Figure 2). In contrast, the outer strands exhibit alternating fast/slow exchange that is consistent with solvent exposure of every other amide proton. The β-sheet does not display much hydropathic polarity as only strand 2 is amphipathic.

Helices were identified by the presence of overlapping medium range NOEs, dαN(i, $i+3$ or 4) and dNN (i, $i\pm3$ or 4) (Figure 2). A long helix with a small break at position G52−I53 is observed for residues R41 to A57. Two shorter helical fragments (Y66−L71 and D78−M85) separated by a loop lie at the carboxyl-terminus of the long helix. Chemical Shift Index analysis (*31*, *32*) employing the experimentally obtained C_α, H_α, C_β, and CO chemical shifts corroborate the secondary structure determined from the NOE data (Figure 2) and clearly identify a helical fragment in the region of the active site (A11−R16) from which only very little NOE data was obtained. Secondary structure prediction programs (nnPredict (*33*), DSSP, (*34*)) predicted that the helix continued through the unassigned region to residue N23 in addition to the presence of a helix at the carboxyl-terminus (G129−G137). The carboxyl-terminal region exhibits alternating exchange rates of two to three residues in fast exchange, one to two residues in slower exchange, which is consistent with a helix with one side solvent exposed. However, medium-range NOEs were not observed for the putative helix at the carboxyl-terminus and chemical shift index analysis results do not indicate the presence of a helix in this region (Figure 2).

The Protein Fold. The 3D $^{15}N/^{15}N/^{1}H$-HMQC−NOESY−HSQC spectra collected for the deuterated protein were analyzed extensively to obtain a connectivity network between amide protons. The data set recorded at 400 ms proved most useful; in the longer mixing time data, the number of NOEs between amide protons was extremely high, suggesting that too much spin-diffusion had occurred. A variety of spectroscopic evidence indicates that the regions between residues 11−29 and 120−141 are mobile in solution. Resonances attributed to the two unassigned regions, N17−I22 and K121−G129, exhibit very narrow line widths with little chemical-shift dispersion. Surrounding these two areas are regions (A11−R16, S25−E28, V111−L116, and E130−G137) that contain multiple crosspeaks and multiple assignment pathways (Figures 4 and 5). We interpret these observations to indicate that under solution conditions ArsC is in dynamic equilibrium between multiple conformations in these areas. Because of this fact, many long range NOEs were difficult to interpret unambiguously as NOEs were often observed to only one member of a pair of "doubled" crosspeaks. A large number of these incidences made it impossible to identify a self-consistent set of NOEs corresponding to only one of the many conformations that exist under these conditions. Therefore, the final number of unambiguously identified long-range NOEs was small; the analysis yielded 144 sequential, 117 medium, and 54 long-range NOEs.

Preliminary distance geometry and simulated annealing calculations carried out using these restraints resulted in a low-resolution model that cannot provide detailed information about the structure in relation to function. This model is, therefore, not shown here. However, the preliminary model structure proved to be adequate to provide information on the uniqueness of the ArsC fold and was used in a variety of Internet-accessible protocols to search for protein structural homology.

DISCUSSION

At a molecular weight of 16 kDa, ArsC should be accessible to structural NMR analysis; several significantly larger proteins have been solved to high resolution by these methods (*24*, *30*, *35*, *36*). However, the highly mobile regions of ArsC preclude complete assignment and, therefore, high-resolution structure determination. When the triple resonance experiments on ^{15}N- and $^{13}C/^{15}N$-labeled protein proved to be inadequate for full resonance assignment, ArsC was fully deuterated to exploit the sensitivity gains that can be derived

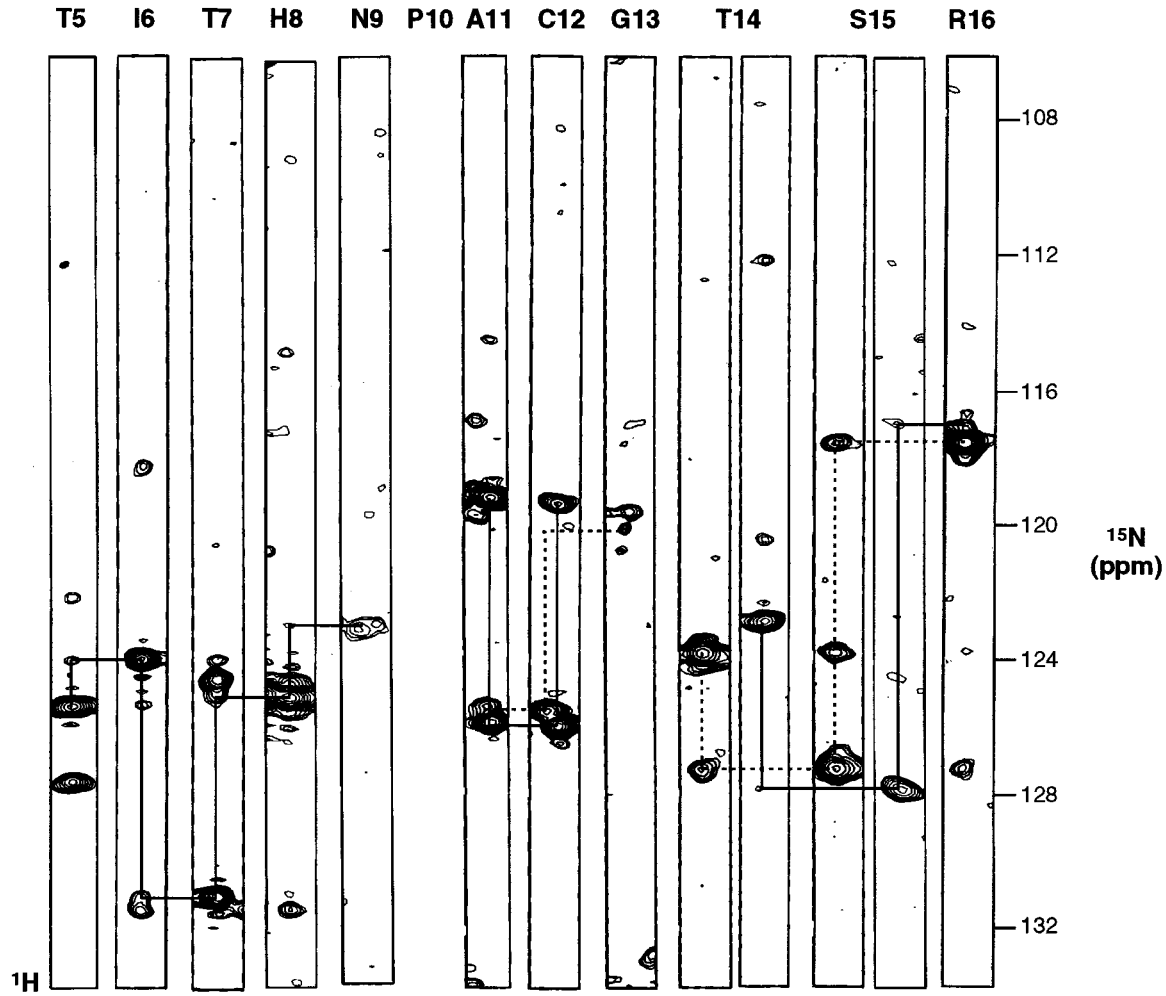

FIGURE 4: Depiction of the multiple connectivity pathways observed in the 3D $^{15}N/^{15}N/^{1}H$-HMQC−NOESY−HSQC spectra in the ArsC active site. Residues 5−9 display the expected single pathway. In contrast, the peak doubling displayed by residues A11−R16 represents slow exchange between multiple conformations, resulting in multiple connectivity pathways.

from a triply labeled sample. The sensitivity of the triple-resonance experiments used for residue identification and the HNCACB and HN(CO)CACB experiments is dependent on magnetic transfer over C_α. It is well-known that the sensitivity of these experiments is limited by the C_α relaxation (*25*, *37*). Since the gyromagnetic ratio of ^{2}H is 7 times less than that of ^{1}H, the heteronuclear dipolar interaction is reduced significantly. Therefore, deuteration increases the sensitivity of the critical HNCACB and HN(CO)CACB experiments, allowing a much more robust assignment protocol that can resolve assignment problems, such as extensive overlap (the motionally averaged areas) and resonance multiplicity.

The dearth of protons also enhances the sensitivity and specificity of NOE crosspeaks between amide protons by minimizing spin diffusion through the side chains (*30*). Consequently, the 3D $^{15}N/^{15}N/^{1}H$-HMQC−NOESY−HSQC spectra provided extensive information about ArsC secondary and tertiary structure of the assigned portions of the molecule. We found that spectra of mixing times up to 400 ms were useful for obtaining distance restraints for simple structure determination protocols, such as distance geometry; NOEs obtained from the spectra with longer mixing times would likely necessitate analysis, using iterative relaxation matrix methodology (*38*). While deuteration of ArsC provided enough sensitivity to assign a much larger portion of the backbone resonances, two highly mobile regions remain unassigned (N17−I22 and K121−G129). These highly mobile regions appear to be in rapid conformational exchange that was manifested in strong resonances exhibiting narrow line widths in the 8.0−8.5 ppm region of the $^{15}N,^{1}H$-HSQC. Each of the unassigned regions is surrounded by areas in slow conformational exchange A11−R16, S25−E28, V111−L116, and E130−G137. The conformational exchange was evident in the NOESY spectra where a variety of peak doubling was observed. Figure 5 illustrates doubling and tripling (V111) of the resonances observed in the helical region formed by residues V111−L116. It is interesting to note that there are no exchange crosspeaks between the conformationally different resonance positions, indicating very slow exchange kinetics (Figure 5). Other regions of the spectra sometimes give a "smear" that may indicate as many as six different conformations. Because of the peak multiplicity, many long-range NOEs could not be assigned to a self-consistent set of data for structure calculation. The limited data could thus provide only a low-resolution three-dimensional picture of the relatively rigid ArsC core.

Uniqueness of the Fold. To obtain more insight in the fold of ArsC, several methodologies available on the Internet were employed to perform searches for structural homology to proteins of known structure using the secondary and preliminary model tertiary structure. Of the several threading

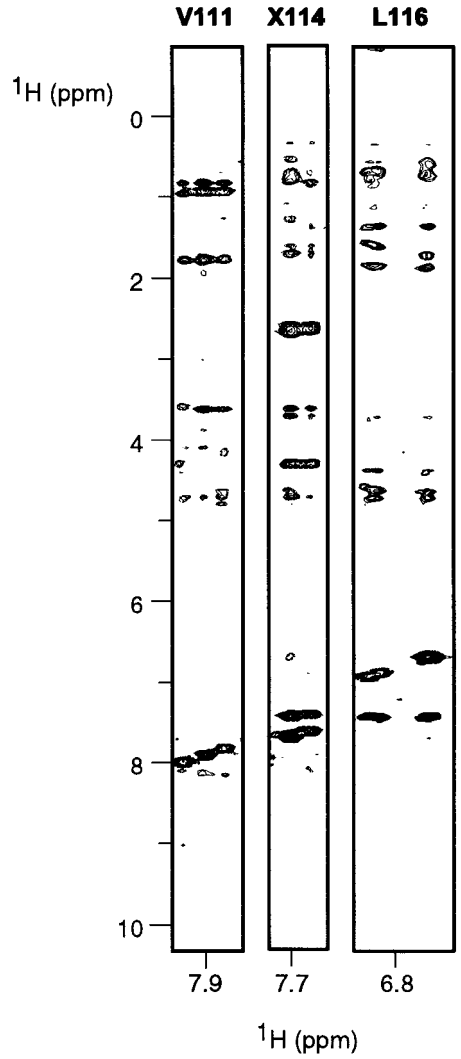

FIGURE 5: Several vectors from the 3D ^{15}N-NOESY−HSQC spectrum display complete sets peak doubling (tripling), which is an indication of slow exchange between multiple conformations.

programs employed to search for similar protein folds, only Eisenberg's protein fold prediction program yielded a fold consistent with the NMR data (*39*). This search pointed to the carboxyl-terminal domain of the porcine class Pi glutathione S-transferase. An expanded search using the protein classification programs SCOP (*40*) and Dali (*41*) was initiated with the glutathione S-transferase fold. The glutaredoxin and thioredoxin folds were found to be structurally related to glutathione S-transferase. Independent searches for structural homology were submitted to the protein classification program CATH (*42*) and the protein topology program TOPS (*43*, *44*). Both programs identified the same group of proteins as structurally related to ArsC: *E. coli* glutaredoxin, the related protein thiol transferase and glutathione transferases.

Like ArsC, thioredoxin, glutaredoxin, thiol transferase, and glutathione S-transferase are α/β proteins. Thioredoxin contains a parallel-stranded β-sheet that in this case contains five, not four, strands, and is, therefore, not considered as a true homologue. In contrast, glutaredoxin, thiol transferase, and glutathione S-transferase all contain a four-stranded mixed β-sheet with the same topology as that of ArsC (Figure 7). In both secondary and tertiary topology, ArsC is similar to these proteins. Accordingly, our present data is sufficient

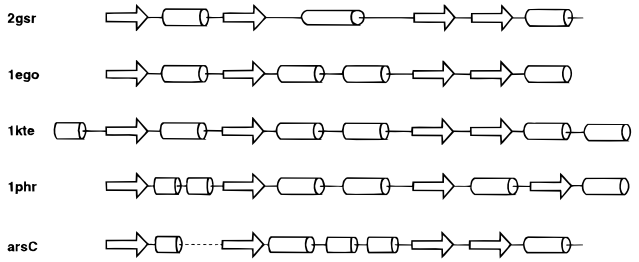

FIGURE 6: A comparison of the secondary structures of several classes of proteins with ArsC. Helices are represented by cylinders and β-strands by arrows. 2gsr is glutathione S-transferase (*15*); 1phr is low molecular weight tyrosine phosphatase (*11*); 1kte is thiol transferase (*14*); 1ego is *E. coli* glutaredoxin (*13*). The dashed line in ArsC figure contains the unassigned, therefore, undefined region of the protein.

to establish that ArsC is structurally related to glutaredoxin, thiol transferases, and glutathione S-transferases.

Interestingly, this group of structurally related proteins are also functionally related. Glutaredoxin and glutathione S-transferase both contain a binding site for reduced glutathione. Glutathione S-transferase catalyzes the addition of reduced glutathione to several organic compounds containing electrophilic centers. In oxidized glutaredoxin, the reduced glutathione regenerates the reduced form. It has been postulated that ArsC may bind to glutathione during the arsenate reduction pathway, but conclusive results have not been attained. However, the active site of ArsC contains H8−C12−R16, which appears to be an alternate, but similar form of a common anion-binding motif. The active site is similar in composition, structural context, and inherent mobility to the H−CX_5R anion-binding sites observed in several phosphatases and sulfatases (*6*, *10*, *45*). This motif has been found to occur as a loop between a β-strand and an α-helix where the cysteine and arginine residues are close in space to one another (*10*, *11*, *46*). The H8 of strand 2 in the β-sheet of ArsC is the relevant histidine in the anion binding motif, forming an ion pair with C12 to lower the pK_a of the active site cysteine to 6.4 (*8*).

Alignment of the secondary structure as well as the active site suggests similarity to the low molecular weight tyrosine phosphatase (PDB accession code 1phr) (*11*). Both proteins are classified as α/β proteins that contain a four-stranded β-sheet. However, while the secondary structure is quite similar in order and content (Figure 6), the phosphatase contains a four-stranded parallel β-sheet in contrast to the mixed β-sheet observed in ArsC (Figure 7). The chemical and structural differences between these two types of β-sheets are significant enough to prohibit consideration as identical folds (*47*). However, the putative functional homology may be an example of convergent evolution.

Mobility of the Active Site. Different functional states of a protein may be manifested in alternative conformational states. Changes in active site side chain orientation, loop positions, and relative subunit orientations in oligomeric proteins have been observed. This mobility and flexibility of proteins can serve to regulate specificity and affinity of ligand recognition. For example, enzymes, such as proteases and nucleases, must be able to accommodate ligands with a variety of compositions. To accomplish this function, the active sites must be dynamic. Indeed, active site mobility has been observed in the proteases stromelysin (Schurter,

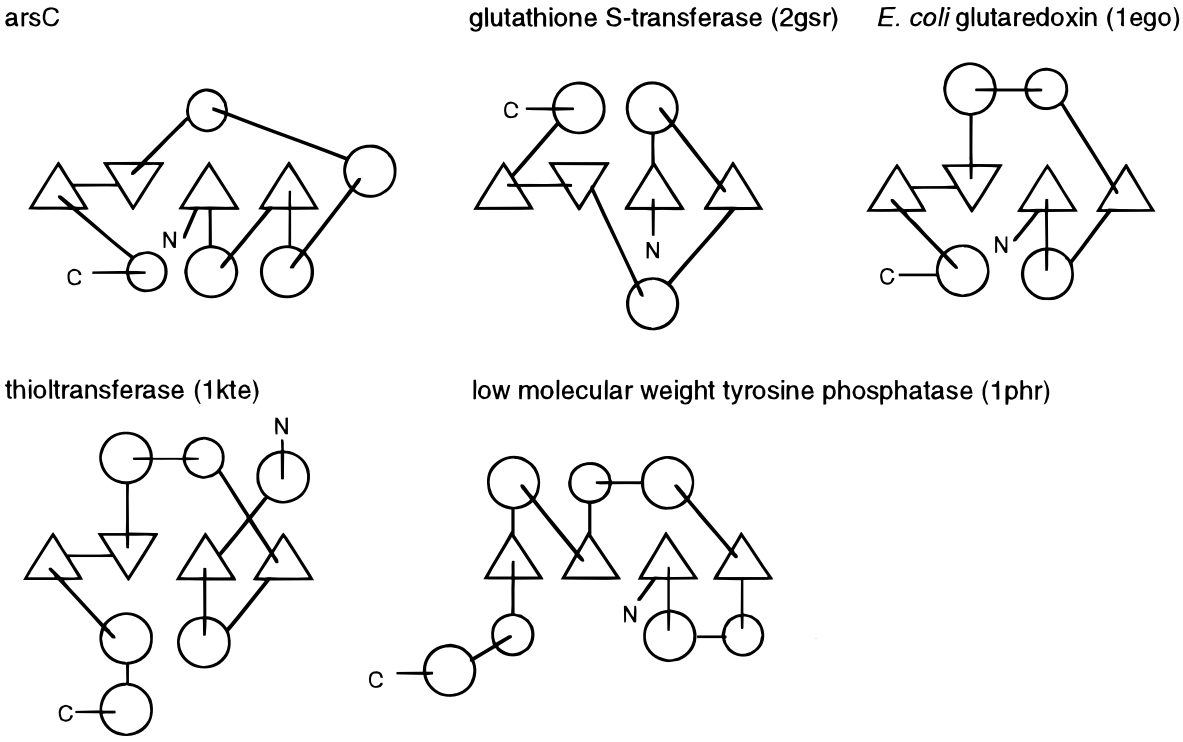

FIGURE 7: The same structures that were compared in Figure 6 are depicted here using TOPS formalism to display the tertiary folds. The circles represent helices and the triangles β-strands. The direction of the strands is illustrated by the direction of the triangles. Out of the plane is defined as the triangle pointing up. Small circles represent helices with a length of less than five residues.

E. J., and Zuiderweg, E. R. P. unpublished results), α-lytic protease (*48*), and the ribonuclease binase [(*49*), Pang, Y., and Zuiderweg, E. R. P. unpublished results]. In addition, the motion of α helix 2 of glutathione S-transferase P1−1 regulates the affinity of the enzyme for reduced glutathione by assisting in the search for optimal substrate orientation (*50*). Residues in this helix have been shown to move 4 Å upon substrate binding (*51*). Moreover, substrate-induced conformational changes in enzymes, such as hexokinase, contribute to the specificity of the enzyme (*52*).

ArsC is an example of a protein containing a mobile active site. The mobility of this region is well-illustrated in the 3D $^{15}N/^{15}N/^{1}H$-HMQC−NOESY−HSQC spectra, which contains parallel connectivities for two conformations that are in slow exchange under these conditions (Figure 4). In an attempt to decrease the mobility of the protein, ArsC was complexed with product (AsO_2^-), which resulted in precipitation and substrate (AsO_4^{3-}). A few resonances appear to shift or broaden, most notably, the C12 and R16 resonances of the active site, upon binding to substrate. In addition, the resonances associated with the dynamic (unassigned) regions of the protein (N17−N24 and the carboxyl-terminus) are broadened beyond observation, indicating that the dynamics is attenuated but not arrested by ligand binding. Therefore, while ligand binding appears to decrease the motional freedom of both the active site and carboxyl-terminal region by slowing them to the intermediate-exchange limit on the chemical shift time scale, it is not sufficient to allow for resonance assignment and structure determination of these areas. However, the results indicate that the active-site area and carboxyl-terminal area of ArsC are dynamically and presumably functionally linked.

The structure of a low molecular weight protein tyrosine phosphatase, a protein that has the H−CX_5R active site, was also studied by NMR (*45*). Similar to ArsC, the active-site residues of the phosphatase exhibited line widths that were too broad to be observed without inhibitor bound. While some of the residues became visible upon saturation with ligand, several remained exchange broadened. In addition, enhanced temperature factors were observed for the anion-binding site area in the crystallographic electron-density maps of the related protein YG3E (Fauman, E., and Saper, M., personal communication). Crystallographic studies of a different protein tyrosine phosphatase from *Yersinia* suggest that a second conserved loop in the active-site region possesses two conformational states (*53*). Studies by ultraviolet resonance Raman spectroscopy indicate that this WpD loop moves rapidly between an open and closed conformation in the absence of substrate (*54*). Ligand binding alters the equilibrium between the two states to favor the closed loop conformation (*54*). Thus, while ArsC tends toward the dynamic extreme, high mobility appears to be inherent in the anion-binding motif.

The deduced homology between the common anion binding motif in ArsC, phosphatases, and sulfatases raises the possibility that the ArsC function may be more involved with essential arsenate or arsenite binding in this complex than with the actual reduction process. Determination of whether a glutathione binding site in ArsC exists will establish whether *E. coli* glutaredoxin, thioltransferase, and glutathione S-transferase are functionally related as well as structurally related. It is evident that internal motions play an important role in anion binding for proteins containing the H−CX_nR motif. Results reported herein suggest a similar importance in ArsC function. Therefore, a thorough understanding of the activity of these anion binding proteins should include characterization, and comparison of the dynamic motion in the active sites.

ACKNOWLEDGMENT

We thank Drs. Jiyang Liu, Christopher Talpas, Peter Sandusky, and Alex Kurochkin, and Dominic Perri for assistance at various stages of this work.

SUPPORTING INFORMATION AVAILABLE

The $^{15}N/^{1}H$-HSQC spectrum of ArsC and a table of the backbone assignments. This material is available free of charge via the intermat at http://pubs.acs.org.

REFERENCES

1. Turner, R. J., Hou, Y., Weiner, J. H., and Taylor, D. E. (1992) *J. Bacteriol. 174*, 3092−3094.
2. Mobley, H. L. T., and Rosen, B. P. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 6119−6122.
3. Chen, C. M., Mobley, H. L. T., and Rosen, B. P. (1985) *J. Bacteriol. 161*, 758−763.
4. Ji, G., and Silver, S. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 9474−9478.
5. Ji, G., Garber, E. A. E., Armes, L. G., Chen, C.-M., Fuchs, J. A., and Silver, S. (1994) *Biochemistry. 33*, 7294−7299.
6. Fauman, E. B., Cogswell, J. P., Lovejoy, B., Rocque, W. J., Holmes, W., Montana, V. G., Piwnica-Worms, H., Rink, M. J., and Saper, M. A. (1998) *Cell 93*, 617−625.
7. Liu, J., Gladysheva, T. B., Lee, L., and Rosen, B. P. (1995) *Biochemistry 34*, 13472−13476.
8. Gladysheva, T., Liu, J., and Rosen, B. P. (1996) *J. Biol. Chem. 271*, 33256−33260.
9. Stuckey, J. A., Schubert, H. L., Fauman, E. B., Zhang, Z.-Y., Dixon, J. E., and Saper, M. A. (1994) *Nature 370*, 571−575.
10. Yuvaniyama, J., Denu, J. M., Dixon, J. E., and Saper, M. A. (1996) *Science 272*, 1328−1331.
11. Su, X.-D., Taddel, N., Stefani, M., Ramponi, G., and Nordlund, P. (1994) *Nature 370*, 575−578.
12. Gliubich, F., Gazerro, M., Zanotti, G., Delbono, S., Bombieri, G., and Berni, R. (1996) *J. Biol. Chem. 271*, 21054−21061.
13. Xia, T.-H., Bushweller, J. H., Sodano, P., Billeter, M., Björnberg, Holmgren, A., and Wüthrich, K. (1992) *Protein Sci. 1*, 310−321.
14. Katti, S. K., Robbins, A. H., Yang, Y., and Wells, W. W. (1995) *Protein Sci. 4*, 1998−2005.
15. Dirr, H., Reinemer, P., and Huber, R. (1994) *J. Mol. Biol. 243*, 72−92.
16. Gladysheva, T. B., Oden, K. L., and Rosen, B. P. (1994) *Biochemistry 33*, 7288−7293.
17. Piotto, M., Saudek, V., and Sklenář, V. (1992) *J. Biomol. NMR 2*, 661−665.
18. Zuiderweg, E. R. P., and Fesik, S. W. (1989) *Biochemistry. 28*, 2387−2391.
19. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Groenenborn, A. M., and Clore, C. M. (1989) *Biochemistry 28*, 6150−6156.
20. Bax, A., and Grzesiek, S. (1993) *Acc. Chem. Res. 26*, 131−138.
21. Clubb, R. T., and Wagner, G. (1992) *J. Biomol. NMR 2*, 389−394.
22. Van Doren, S. R., Kurochkin, A. V., Ye, Q. Z., Johnson, L. L., Hupe, D. J., and Zuiderweg, E. R. P. (1993) *Biochemistry 32*, 13109−13122.
23. Boucher, W., Laue, E. D., Campbell-Burk, S., and Domaille, P. J. (1992) *J. Biomol. NMR 2*, 631−637.
24. Wang, H., Kurochkin, A. V., Pang, Y., Hu, L., Flynn, G. C., and Zuiderweg, E. R. P. (1998) *Biochemistry 37*, 7929−7940.
25. Yamazaki, T., Lee, W., Arrowsmith, C. H., Muhandiram, D. R., and Kay, L. E. (1994) *J. Am. Chem. Soc. 116*, 11655−11666.
26. Van Doren, S. R., and Zuiderweg, E. R. P. (1992) *J. Magn. Reson. Ser. A 104*, 222−225.
27. Grzesiek, S., and Bax, A. (1993) *J. Biomol. NMR 3*, 185−204.
28. Grzesiek, S., Wingfield, P., Stahl, S., Kaufman, J. D., and Bax, A. (1995) *J. Am. Chem. Soc. 117*, 9594−9595.
29. Shan, X., Gardner, K. H., Muhandiram, D. R., Rao, N. S., Arrowsmith, C. H., and Kay, L. E. (1996) *J. Am. Chem. Soc. 118*, 6570−6579.
30. Venters, R. A., Farmer, B. T., II, Fierke, C. A., and Spicer, L. D. (1996) *J. Mol. Biol. 264*, 1101−1116.
31. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1992) *Biochemistry 31*, 1647−1651.
32. Wishart, D. S., and Sykes, B. D. (1994) *J. Biomol. NMR 4*, 171−180.
33. Kneller, D. G., Cohen, F. E., and Langridge, R. (1990) *J. Mol. Biol. 214*, 171−182.
34. Kabsch, W., and Sander, C. (1983) *Biopolymers. 22*, 2577−2637.
35. Martin, J. R., Mulder, F. A. A., Karimi-Nejad, Y., van der Zwan, J., Mariani, M., Schipper, D., and Boelens, R. (1997) *Structure 5*, 521−532.
36. Van Doren, S. W., Kurochkin, A. V., Hu, W., Ye, Q.-Z., Johnson, L. L., Hupe, D. J., and Zuiderweg, E. R. P. (1995) *Protein Sci. 4*, 2487−2498.
37. Grzesiek, S., Anglister, J., Ren, H., and Bax, A. (1993) *J. Am. Chem. Soc. 115*, 4369−4370.
38. Koning, M. M. G. (1990) Ph.D. Thesis, Utrecht University.
39. Fischer, D., and Eisenberg, D. (1996) *Protein Sci. 5*, 947−955.
40. Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995) *J. Mol. Biol. 247*, 536−540.
41. Holm, L., and Sander, C. (1997) *Structure 5*, 165−171.
42. Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B., and Thornton, J. M. (1997) *Structure 5*, 1093−1108.
43. Westhead, D. R., Hatton, D. C., and Thornton, J. M. (1998) *TIBS 23*, 35−36.
44. Flores, T. P., Moss, D. S., and Thornton, J. M. (1994) *Protein Eng. 7*, 31−37.
45. Logan, T. M., Zhou, M.-M., Nettesheim, D. G., Meadows, R. P., Van Etten, R. L., and Fesik, S. W. (1994) *Biochemistry 33*, 11087−11096.
46. Barford, D., Flint, A. J., and Tonks, N. K. (1994) *Science 263*, 1397−1403.
47. Nesloney, C. L., and Kelly, J. W. (1996) *Bioorg. Med. Chem. 4*, 739−766.
48. Davis, J. H., and Agard, D. A. (1998) *Biochemistry 37*, 7696−7707.
49. Reibarkh, M. Y., Nolde, D. E., Vasilieva, L. I., Bocharov, E. V., Shulga, A. A., Kirpichnikov, M. P., and Arseniev, A. S. (1998) *FEBS Lett. 431*, 250−254.
50. Ricci, G., Caccuri, A. M., Lo Bello, M., Rosato, N., Mei, G., Nicotra, M., Chiessi, E., Mazzetti, A. P., and Federici, G. (1996) *J. Biol. Chem. 271*, 16187−16192.
51. Stella, L., Caccuri, A. M., Rosato, N., Nicotra, M., Lo Bello, M., De Mateis, F., Mazzetti, A. P., Federici, G., and Ricci, G. (1998) *J. Biol. Chem. 273*, 23267−23273.
52. Bennett, W. S. Jr., and Steitz, T. A. (1978) *Proc. Natl. Acad. Sci. 75*, 4848−4852.
53. Shubert, H., Fauman, E., Stuckey, J., Dixon, J., and Saper, M. A. (1995) *Protein Sci. 4*, 1904−1913.
54. Juszczak, L. J., Zhang, Z.-Y., Wu, L., Gottfried, D. S., and Eads, D. D. (1997) *Biochemistry 36*, 2227−2236.

BI990333C