# A structural and functional investigation of the periplasmic arsenate-binding protein, ArrX from *Chrysiogenes arsenatis*

Nilakhi Poddar[1], Consuelo Badilla[2], Shadi Maghool[1], Thomas H. Osborne[2], Joanne M. Santini[2*], Megan J. Maher[1,3*]

[1]School of Chemistry and The Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, Parkville, Australia.

[2]Institute of Structural and Molecular Biology, Division of Biosciences, University College London, London WC1E 6BT, United Kingdom.

[3]Department of Biochemistry and Genetics, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, Australia.

*Corresponding authors: Megan J. Maher and Joanne M. Santini

E-mail: megan.maher@unimelb.edu.au and j.santini@ucl.ac.uk

**Running Title**: Structure and function of the ArrX protein.

**Keywords:** arsenate, *Chrysiogenes arsenatis*, periplasmic binding protein (PBP), isothermal titration calorimetry (ITC), X-ray crystallography

**Abstract**

The anaerobic bacterium *Chrysiogenes arsenatis* respires using the oxyanion arsenate ($AsO_4^{3-}$) as the terminal electron acceptor, where it is reduced to arsenite ($AsO_3^{3-}$) while concomitantly oxidizing various organic (e.g. acetate and formate) or inorganic (e.g. hydrogen) electron donors. This respiratory activity is catalyzed in the periplasm of the bacterium by the enzyme arsenate reductase (Arr), with expression of the enzyme controlled by a sensor histidine kinase (ArrS) and a periplasmic binding protein (PBP), ArrX. Here, we report for the first time, the molecular structure of ArrX in the absence and presence of bound ligand, arsenate. Comparison of the ligand-bound structure of ArrX with other PBPs shows a high level of conservation of critical residues for ligand binding by these proteins, however this suite of PBPs show different structural alterations upon ligand binding. For ArrX and its homologue AioX (from *Rhizobium* sp. str. NT-26), which specifically binds arsenite, the structures of the substrate binding sites in the vicinity of a conserved and critical cysteine residue, contribute to the discrimination of binding for these chemically similar ligands.

## Introduction

Arsenic is a naturally occurring toxic metalloid that is widely distributed throughout the Earth's crust, commonly found in minerals associated with other elements such as iron and sulfur [1]. The presence of arsenic in water reservoirs originates either naturally from underlying mineralogy or from anthropogenic processes such as smelting, mining, the use of arsenical insecticides and pesticides and the release of industrial waste [2, 3]. The contamination of drinking water with arsenic poses a significant health threat, with levels of arsenic ~50 µg/L affecting hundreds of million people worldwide. The World Health Organization (WHO) recommended maximum contaminant level (MCL) for arsenic (i.e., the inorganic soluble forms, arsenite +3 and arsenate +5) in drinking water is 10 µg/L [4]. However, a 2009 survey found elevated levels of arsenic (~ 200 µg/L) in the drinking water of areas of Bangladesh, with 22-35 million people estimated to be affected [5] and the mortality rate due to arsenic poisoning higher than 20% [6]. In aqueous environments, arsenic is present in two forms: arsenite ($AsO_3^{3-}$) and arsenate ($AsO_4^{3-}$), with arsenite being more soluble and more toxic than arsenate [7]. Arsenate is strongly adsorbed onto the surfaces of minerals whereas arsenite readily leaches from these sources, contaminating water reservoirs. Arsenate is able to enter cells through the phosphate transport system and is able to mimic inorganic phosphate into cellular or glycolytic respiration pathways [8, 9]. Arsenite enters the cells through membrane bound aquaglyceroporins [10, 11] and binds to sulfhydryl groups of proteins, thus inactivating essential enzymes [12]. Chronic arsenic exposure, mainly through ingestion of contaminated water, is associated with pathologies such as cardiomyopathic disease, keratosis, hepatitis, respiratory disease, various types of cancers and neurological effects as peripheral neuropathy [13].

The conversion between the two soluble forms of arsenic leads to biogeochemical cycling in the environment which is mediated by various prokaryotes [14]. Although arsenic is toxic to most

living organisms, a variety of phylogenetically diverse microorganisms are able to use arsenic oxyanions for respiration [14, 15]. For example, *Bacillus selenitireducens* [16], *Shewanella trabarsenatis* strain ANA-3 [17] and *Chrysiogenes arsenatis* [18, 19] under anoxic conditions, respire using $AsO_4^{3-}$ as the terminal electron acceptor (where it is reduced to $AsO_3^{3-}$) while concomitantly oxidizing various organic (e.g. acetate and formate) or inorganic (e.g. hydrogen) electron donors [20, 21].

Bacterial respiration with arsenic oxyanions is facilitated by members of the DMSO reductase family of molybdenum-containing enzymes [17, 18, 22, 23]. Respiratory arsenate reductase (Arr), has been purified and characterized from *Chrysiogenes arsenatis* [17, 18] and the crystal structure of Arr from *Shewanella* sp. ANA-3 has been reported [24]. The Arr complex consists of two subunits, a large subunit ArrA and a smaller subunit ArrB. The ArrA subunit contains a molybdopterin guanine dinucleotide (Mo-bisPGD) cofactor with a 4Fe–4S cluster [18, 24] and the ArrB subunit contains four 4Fe–4S clusters [17, 18, 25]. In *Chrysiogenes arsenatis*, the expression of the Arr complex is regulated by three genes, *arrR*, *arrS* and *arrX*, which encode a response regulator, a sensor histidine kinase and periplasmic-binding protein (PBP), respectively [18, 26].

PBPs play essential roles in solute uptake, chemotaxis and signaling [27]. PBPs are located in the periplasm of Gram-negative bacteria and bind to a variety of ligands such as vitamins [28], carbohydrates [29], oligopeptides [30], amino acids [31] and transition metal ions [32]. The PBPs share a common fold, despite the relatively low sequence identities across the family, with two (or more, in rare cases) globular domains linked by a hinge region, which forms a groove-like cavity representing the ligand-binding site [33]. The binding of a ligand, into a cleft between the two domains, is generally accompanied by a change in the structure of the protein [34, 35]. In the case of PBPs that interact with ABC transporters [36], this change is typically a reorganization of

the relative positioning of the two domains, so that on substrate binding, the domains lie in close proximity to each other and essentially 'close' the substrate binding cleft. These rearrangements can be quite significant, with rotations of domains by between 25-72° observed [37]. It is the 'closed' conformation of the PBP that interacts with the transporter for substrate handover and translocation across the membrane [38-40]. For PBPs that interact with sensory histidine kinases (SHKs) [41, 42] to facilitate the activation of gene expression, the structural reorganization that occurs on substrate binding seems more subtle, with small loop movements observed on substrate binding [43]. It has been suggested that these minor rearrangements change the surface structures and/or charge distributions of the proteins that allow a productive interaction with the periplasmic domain of the corresponding SHK [44].

We have previously reported the structural analysis of the AioX protein from *Rhizobium* sp. str. NT-26 [45, 46], which showed two globular domains with a hinge region accommodating the ligand binding site. Structures determined in the absence and presence of arsenite, revealed that a loop, in proximity to the substrate binding cleft, shows different conformations in the two forms (PDB 6ESK and 6EU7, respectively) [46]. This loop movement upon arsenite binding was proposed to facilitate the interaction of AioX with AioS (the sensory histidine kinase), leading to the activation of the expression of the arsenite oxidase Aio. This study also reported the overexpression and purification of the ArrX protein from *Chryisogenes arsenatis* and the binding affinities of AioX and ArrX for substrate (for $AsO_3^{3-}$ and $AsO_4^{3-}$, respectively). Crucially, the binding of $AsO_4^{3-}$ to the AioX protein and $AsO_3^{3-}$ to the ArrX protein could not be detected [46], indicating significant specificity of these two PBPs for their respective substrates.

To explore the molecular basis of the substrate specificities of the AioX and ArrX proteins, we have determined the crystal structures of ArrX in complex with a variety of oxyanions, including arsenate, by X-ray crystallography. We have investigated the thermodynamics of oxyanion binding to ArrX and AioX variant proteins, with mutations to conserved residues at the substrate binding sites. These studies allow us to propose the mechanisms by which these proteins recognize and specifically bind chemically similar ligands.

## Materials and Methods

The *arrX* and *aioX* genes from *Chyrisogenes arsenatis* (UniprotK: UPI00041882C4) and *Rhizobium* sp str. NT-26 (UniprotKB: L0NML6) were amplified from genomic DNA (minus their TAT leader signal sequences) *via* PCR with forward and reverse primers listed in Table S1, and subcloned into pPROEX-HTb (Invitrogen) TEV-cleavable $His_6$ tagged fusion vector [46]. The point mutations $^{C112A}$ArrX, $^{C112S}$ArrX, $^{Y138A}$ArrX, $^{C106A}$AioX, $^{C106S}$AioX and $^{Y131A}$AioX were introduced into the respective ORFs using a Q5 Site directed Mutagenesis Kit (New England Biolabs), according to the manufacturer's instructions. Forward and reverse primers for all mutant proteins were designed using NEBaseChanger online tool (https://nebasechanger.neb.com/) as listed in Table S1. In this study, ArrX is composed of residues 30-310. Residue numbering for the structurally characterized proteins described here follow the PDB convention, being numbered according to the full-length proteins.

### *Protein overexpression and purification*

The pPROEX-HTb plasmids including DNA sequences encoding ArrX [46], $^{C112A}$ArrX, $^{C112S}$ArrX, $^{Y138A}$ArrX, $^{C106A}$AioX, $^{C106S}$AioX and $^{Y131A}$AioX were individually transformed into *E. coli* strain JM109 (DE3) (New England, Biolabs). Cultures were grown at 37°C in Luria

Broth (LB) supplemented with ampicillin (100 μg/mL) to an $OD_{600}$ value between 0.5 - 0.8, induced with IPTG (0.2 mM) and harvested after 16 h with shaking at 16°C.

His-ArrX, His-$^{C112A}$ArrX, His-$^{C112S}$ArrX, His-$^{Y138A}$ArrX, His-$^{C106A}$AioX, His-$^{C106S}$AioX and His-$^{Y131A}$AioX were purified by $Ni^{2+}$-NTA affinity chromatography followed by size exclusion chromatography. Frozen cell pellets were thawed and resuspended in binding buffer (20 mM phosphate, 500 mM NaCl, 40 mM imidazole, pH 7.2). The cells were lysed using a TS series bench top cell disruptor (Constant Systems Ltd) at 35 kpsi and insoluble debris were removed by centrifugation (Beckman JLA-25.50, 30,000*g*, 20 min; 4°C). The soluble fraction was incubated with Ni-Sepharose™ 6 Fast Flow resin (5 mL; GE Healthcare Life Sciences; 4°C; 1hr stirring), which was pre-equilibrated with binding buffer (20 mM potassium phosphate, 500 mM NaCl, 40 mM imidazole, pH 7.2). The resin was washed with (10 column volumes, CV) of binding buffer followed by elution of bound proteins with elution buffer (20 mM phosphate, 500 mM NaCl; 200 mM imidazole; pH 7.2, 5 CV). The eluent was concentrated by centrifugal ultrafiltration (10K molecular weight cut-off, Millipore Amicon Ultra) followed by size exclusion chromatography (Hi-Load 16/600 Superdex 75 pg; GE Healthcare Life Sciences; 4°C) in 50 mM Tris-HCl, pH 8.0 (His-ArrX, His-$^{C112A}$ArrX, His-$^{C112S}$ArrX and His-$^{Y138A}$ArrX) or 20 mM Tricine, pH 7.5 (His-$^{C106A}$AioX, His-$^{C106S}$AioX and His-$^{Y131A}$AioX). The fractions showing the highest purities as determined by SDS-PAGE analyses were pooled and concentrated to approximately 20 mg/mL by centrifugal ultrafiltration. Protein aliquots were snap frozen and stored at -80°C until further use. The His-ArrX protein generated by this method was used for crystallization trials.

For isothermal titration calorimetry (ITC) analyses, cleaved proteins ArrX, $^{C112A}$ArrX, $^{C112S}$ArrX ,$^{Y138A}$ArrX, $^{C106A}$AioX, $^{C106S}$AioX and $^{Y131A}$AioX were prepared by incubation with Tobacco Etch Virus (TEV) protease (1:100 (w/w) TEV protease protein) with dialysis (3,000

Da MW cut-off SnakeSkin® Dialysis Tubing, ThermoScientific) against dialysis buffer (20 mM phosphate, 500 mM NaCl, pH 7.2; 2.0 L) overnight. Subsequent separation from the TEV protease using Ni-Sepharose™ 6 Fast Flow resin (GE Healthcare Life Sciences), followed by SEC was performed as described above. The $^{Y138A}$ArrX protein prepared in this manner was also used for crystallization trials.

***Analyses of substrate binding thermodynamics by ITC***

Binding affinities were measured using a MicroCal200 ITC (Malvern) at 25°C with constant stirring at 750 rpm. Protein concentrations were determined using a Nanodrop UV spectrophotometer (Thermo Scientific) at a wavelength of 280 nm. Sodium arsenate ($Na_2HAsO_4$), sodium (meta) arsenite ($NaAsO_2$), lithium sulphate ($Li_2SO_4$), sodium malonate ($Na_2C_3H_2O_4$), sodium acetate ($NaCH_3CO_2$), sodium formate ($NaHCO_2$) (Sigma Aldrich; 2 mM) and potassium phosphate ($KH_2PO_4$; 3 mM), were prepared at these concentrations in buffers matching those of the respective protein samples and titrated using 25 injections of volume 1.5-2 µL into the protein samples (200 µM; 50 mM Tris-HCl, pH 8.0 for $^{C112A}$ArrX, $^{C112S}$ArrX, $^{Y138A}$ArrX and 20 mM Tricine, pH 7.5 for $^{C106A}$AioX, $^{C106S}$AioX and $^{Y131A}$AioX). All measurements were performed in triplicate. Binding parameters were determined using Microcal Origin software (Origin7, OriginLab Corporation) and fitted to a one-site model to yield values of binding stoichiometry (N), association constants ($K_a$), ΔH and ΔS. The dissociation constant ($K_D$) was calculated as the reciprocal of the association constant: that is, $K_D = 1/K_a$.

***Protein crystallization and data collection***

Initial crystallization trials for His-ArrX were conducted using the Index HT™ (Hampton Research), Structure, PACT and JCSG (Molecular Dimensions) screens. For the $^{Y138A}$ArrX

protein, only the Index HT™ (Hampton Research) was used. Crystallization experiments were carried out by sitting drop vapor diffusion in 96 well plates (Molecular Dimensions) using purified His-ArrX and $^{Y138A}$ArrX at two different concentrations (10 and 20 mg/mL, in 50 mM Tris-HCl, pH 8.0). Drops consisting of equal volumes (0.2 µL) of reservoir and protein solutions, were dispensed by a Crystal Gryphon Liquid Handling System (Art Robins Instruments) and were equilibrated against a reservoir of screen solution (50 µL) at 20°C. Multiple rod-like crystals of His-ArrX were observed within one week in conditions G5, G9, H3 and H8 of the Index HT™ screen (0.2 M lithium sulphate, 0.1 M Tris-HCl pH 8.5, 25% (w/v) PEG 3350; 0.2 M ammonium acetate, 0.1 M Tris-HCl pH 8.5, 25% (w/v) PEG 3350; 0.2 M sodium malonate-malonic acid pH 7.0, 20% (w/v) PEG 3350; 0.1 M magnesium formate, 15% (w/v) PEG 3350 respectively). Multiple small $^{Y138A}$ArrX crystals were obtained after two days in condition G5 of the Index HT™ screen (0.2 M lithium sulphate, 0.1 M Tris-HCl pH 8.5, 25% (w/v) PEG 3350).

Optimization of these conditions were carried out by hanging drop vapor diffusion in 24-well VDX plates (Hampton Research). Diffraction-quality crystals of His-ArrX were grown in a number of conditions, in the presence of different buffers (Table S2). To prepare crystals of His-ArrX in complex with $AsO_4^{3-}$, His-ArrX crystals (grown in 0.2 M lithium sulphate, 0.1 M Tris-HCl pH 8.5, 25% (w/v) PEG 3350) were soaked (30 sec) in 5 mM $Na_2HAsO_4$, 0.1 M Tris-HCl pH 8.5, 27% (w/v) PEG 3350, 25% (v/v) glycerol immediately prior to cryo cooling and data collection. In an attempt to prepare His-ArrX crystals without ligand in the substrate binding site, the crystals were soaked (30 sec) in 0.2 M lithium chloride, 0.1 M Tris-HCl pH 8.5, 27% (w/v) PEG 3350, 25% (v/v) glycerol immediately prior to cryo cooling and data collection.

$^{Y138A}$ArrX was crystallized at 20°C by hanging drop vapor diffusion with drops consisting of equal volumes (1.0 µL) of protein (5 mg/mL in 50 mM Tris-HCl, pH 8.0) and reservoir solution (0.2 M lithium sulphate, 0.1 M Tris-HCl pH 8.7, 27% (w/v) PEG 3350). All crystals were harvested and immersed in cryoprotectant (reservoir solution (Table S2) with 25% (v/v) glycerol) before flash cooling in liquid nitrogen.

Diffraction data were collected from the His-ArrX crystals on a Pilatus 6M-F detector at the Diamond Light source on I04 beamline at 12,664 eV. For the remaining His-ArrX and $^{Y138A}$ArrX structures, diffraction data were collected on an Eiger 16M detector at the Australia Synchrotron on beamline MX2 at 13,000 eV. For the $Na_2HAsO_4$-soaked His-ArrX crystals, datasets were collected at 11,604, 12,110 and 13,000 eV ('low energy', 'high energy' and 'native', respectively; Table 1). All data were collected at 100K, processed with XDS [47] and merged and scaled with AIMLESS [48].

***Structure solution and refinement***

All structures were solved by molecular replacement with PHASER [49] from the CCP4 suite [50]. The His-ArrX was solved with a search model generated with CHAINSAW [51] from the coordinates of the *apo*-AioX structure (PDB 6ESK) [46]. The structures of $^{Y138A}$ArrX, arsenate-His-ArrX and His-ArrX in complex with buffer molecules were determined with the structure of His-ArrX as the search model, after removal of all non-protein atoms. Manual model building and the addition of water molecules for all data sets were carried out in COOT [52] with iterative cycles of refinement using REFMAC5 [53]. The quality of all structures was determined using MOLPROBITY [54].

## Results and Discussion

### *The structure of His-ArrX*

The structure of His-ArrX was solved and refined to 1.67 Å resolution. X-ray data and refinement statistics are presented in Table 1. The structure shows one monomer of His-ArrX per asymmetric unit (asu) and includes residues 52-306. No interpretable electron density was observed for residues 30-51 or 307-310, so they were omitted from the final model. The structure consists of two globular lobe-like domains: a N-terminal (residues 52-131 and 249-306) and a C-terminal domain (residues 141-240), which are linked by a hinge region (residues 132-140 and residues 241-248) (Figure 1A). The cleft between the domains forms a central cavity, which through comparison with the structures of other PBPs, represents the ligand-binding pocket. Both the N- and C-terminal lobes show similar folds, composed of β sheets flanked by α helices, with 6 β strands in the N-terminal domain and 5 β strands in the C-terminal region (Figure 1B, C).

As the refinement of this structure progressed, positive difference Fourier electron density was observed in the substrate binding site, between the two domains, which appeared to have tetrahedral geometry. Given that the crystals were grown in the presence of $Li_2SO_4$, this density was interpreted as a sulfate ($SO_4^{2-}$) ion, occupying the putative substrate-binding site. Residues Tyr138, Ser168 and Ser170 surround this sulfate ion and interact *via* hydrogen bonding. In addition, Thr113 which was modelled in two conformations, shows one conformation which is hydrogen bonded to the sulfate moiety (Figure 1D).

Two notable regions of the structure frame the substrate-binding site. Residues 59-66 constitute a loop (the L1 loop) that caps the substrate binding site. Residues 131-138 (the hinge), which are part of the region that links the N-terminal and C-terminal domains, line the base of this

site and contribute Tyr138, which as mentioned above is one of the residues that interacts with the bound sulfate.

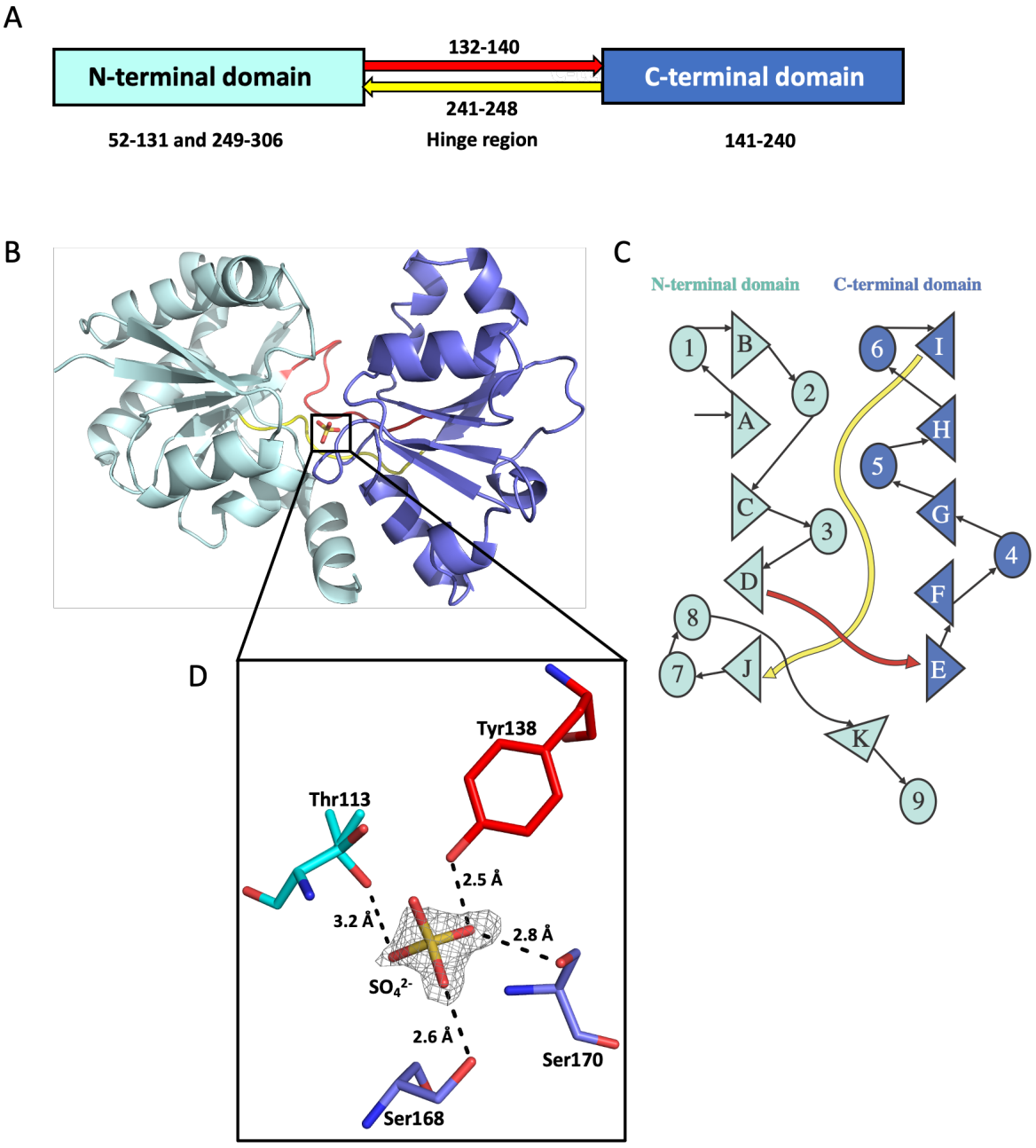



**Figure 1. Overall structure of His-ArrX. (A)** Schematic representation of the structural domains of His-ArrX. The N-terminal domain (cyan), hinge region (red and yellow) and the C-terminal domain (blue) are shown. Residue numbers for each domain are indicated. **(B)** Cartoon representation of the structure of His-ArrX bound to a sulfate ion. Structural domains are colored as in (A). **(C)** Topology of the His-ArrX structure, colored as in (A). Triangles represent β strands and circles represent α helices, with the hinge region represented as curved arrows. **(D)** Insert to (B): the ligand binding site of His-ArrX with a sulfate ion bound. A $F_o$-$F_c$ difference Fourier electron density map (grey) calculated prior to inclusion if the sulfate ion ($SO_4^{2-}$, yellow sticks) in the model,

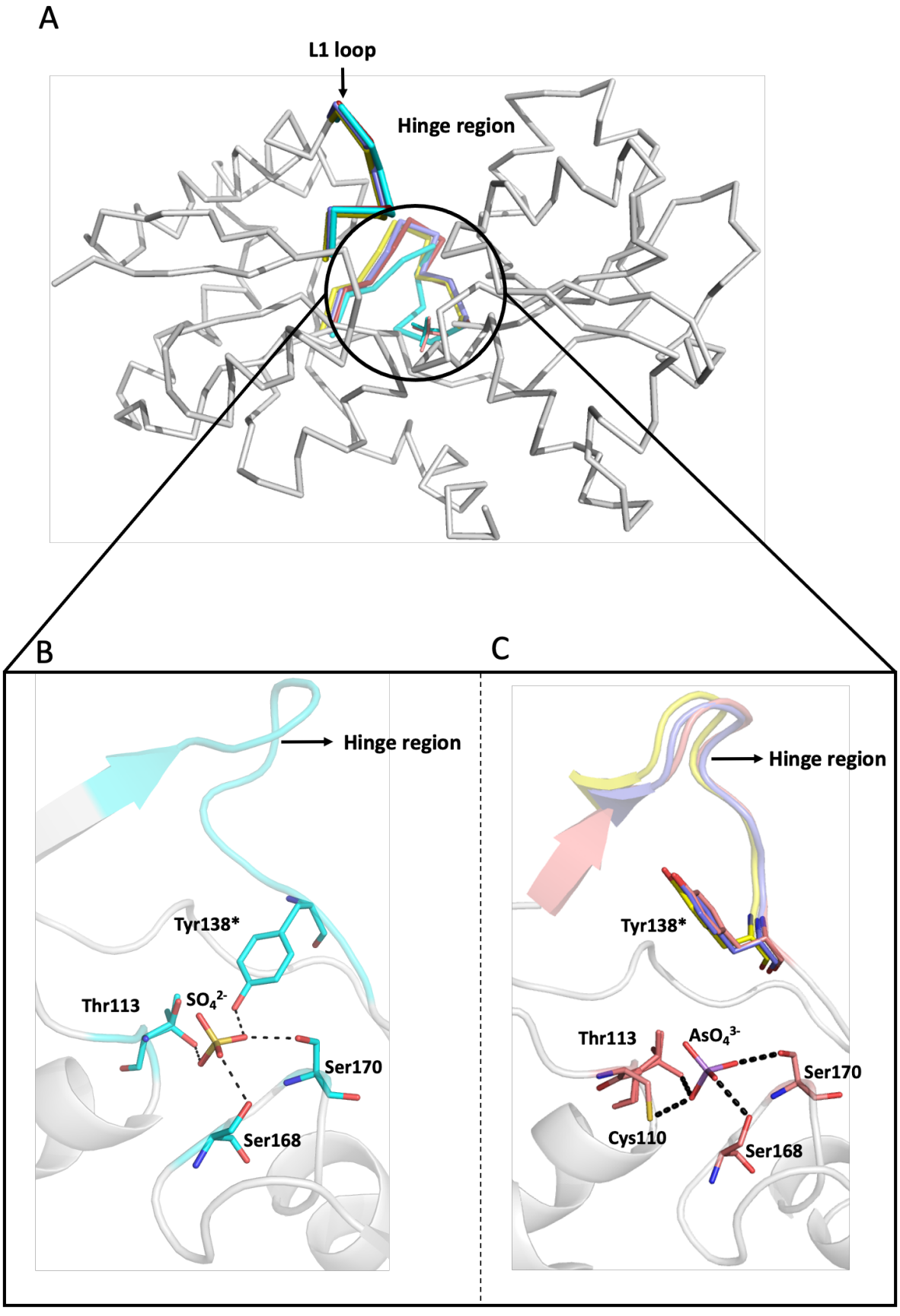
contoured at 4$\sigma$. Hydrogen-bonding interactions between the sulfate ion and surrounding residues are indicated as black, dashed lines.

The His-ArrX protein was also crystallized in the presence of other oxyanions, including malonate, acetate and formate, with the corresponding structures revealing oxyanion binding to the substrate binding site (Figure S1A-D, Table 1). Importantly, despite the observation of these bound oxyanions in the crystal structures, the binding of sulfate, malonate, acetate and formate to ArrX could not be detected by ITC (Table 2). This suggests that the observation of these oxyanions in the crystal structures is due to their presence in the crystallization solutions at significant concentrations (Table S2).

In attempt to solve the structure of ArrX with a vacant substrate binding site, we soaked a crystal grown in the presence of $Li_2SO_4$ in s cryoprotectant solution containing LiCl. However, diffraction data collected from this crystal (chloride-soak-His-ArrX, Figure S1D, Table 1) showed that sulfate binding to the substrate binding site was maintained. The superposition of the His-ArrX structures determined in the presence of the various oxyanions show a variety of differences. The structures of His-ArrX determined in the presence of formate and acetate (in space group $P2_1$, Table 1) are significantly different from the His-ArrX, malonate-His-ArrX and chloride-soak-His-ArrX structures (in space group $P2_12_12_1$), as demonstrated by the overall root mean squared deviations (r.m.s.d.) in common C$\alpha$ positions of 1.54 – 1.84 Å upon their superposition (Table S3). It is possible that these global differences result from crystal packing effects, which would point to significant flexibility in the ArrX structure.

In contrast, the His-ArrX, malonate-His-ArrX and chloride-soak-His-ArrX structures are very similar (r.m.s.d. values of 0.35 – 0.74 Å, Table S3) and show only isolated, local structural

differences. In particular, residue Tyr138 shows different conformations in these structures (Figure 1, S1). In His-ArrX there is a hydrogen bonding interaction between Tyr138 and the bound sulfate ion, whereas in the malonate-His-ArrX and chloride-soak-His-ArrX structures, the side chain of Tyr138 is rotated 'away' from the substrate binding site. The change in the orientation of Tyr138 is accompanied by a rearrangement of part of the hinge region (residues 132-137), which links the two lobe-like domains (Figure 2). The origin of these differences is unclear, particularly since the His-ArrX and chloride-soak-His-ArrX structures show different hinge conformations, despite showing the same entity (sulfate) bound at the substrate binding site. This hinge region in ArrX is presumably flexible and, with soaking of the ArrX yielding alternative possible structures.


A
L1 loop
Hinge region
B
Hinge region
Tyr138*
Thr113
$SO_4^{2-}$
Ser170
Ser168
C
Hinge region
Tyr138*
Thr113
$AsO_4^{3-}$
Ser170
Cys110
Ser168

311

312

**Figure 2. Loop and hinge conformations for the His-ArrX, arsenate-His-ArrX, malonate-His-ArrX and chloride-soak-His-ArrX structures. (A)** The structure of His-ArrX superposed with that of the arsenate-His-ArrX, malonate-His-ArrX and chloride-soak-His-ArrX complexes (the L1 loop and hinge region are colored in cyan, salmon, yellow and blue respectively). **(B)** The substrate binding site and hinge region (cyan) for the His-ArrX structure ($SO_4^{2-}$, yellow). **(C)** The substrate binding site and hinge regions (salmon, yellow and blue) for the arsenate-His-ArrX complex ($AsO_4^{3-}$, purple), malonate-His-ArrX and chloride-soak-His-ArrX structures. * highlights the change in orientation of residue Tyr138 that accompanies arsenate binding to His-ArrX. Hydrogen bonding interactions are shown as black, dashed lines.

### *The arsenate-His-ArrX complex*

The arsenate bound His-ArrX structure was solved and refined to 1.74 Å resolution. X-ray data and refinement statistics are presented in Table 1. The final model includes residues 50-306. Interpretable electron density was not observed for residues 30-49 and 307-310, so these were omitted from the final model. To locate bound arsenate ($AsO_4^{3-}$) in the electron density, anomalous difference Fourier maps were calculated from datasets collected at different energies near the arsenic (As) X-ray absorption edge: 11604 eV ($f''_{As} = 0.53$ e$^-$; 'low E', Table 1) and 12110 eV ($f''_{As} = 3.65$ e$^-$; 'high E', Table 1). An anomalous difference Fourier map calculated with the high-energy (12110 eV) data showed a peak of intensity 6.2 $\sigma$ at the substrate-binding site, indicating the presence of the arsenate ion in this position (Figure 3). This peak was very weak (3.0 $\sigma$) in the corresponding maps calculated from data collected at 11604 eV (Figure 3). Accordingly, an arsenate ion was modelled at this position with an occupancy of 1.0 and a temperature ($B$) factor of 38 Å$^2$.

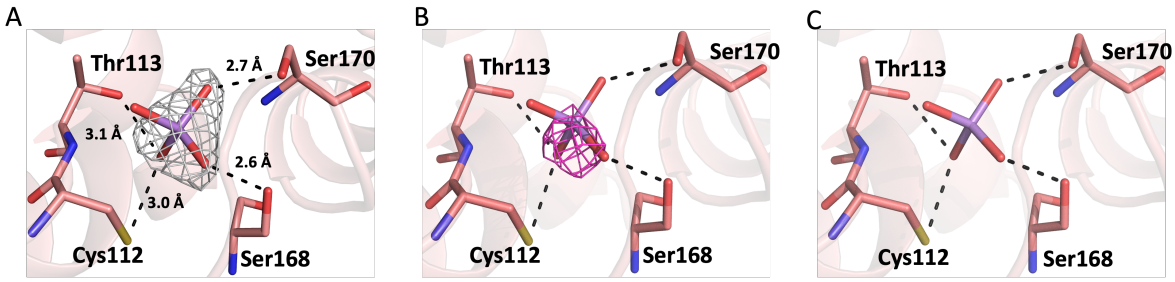



**Figure 3. Identification of the bound arsenate ion in the arsenate-His-ArrX structure. (A)** The ligand binding site of His-ArrX with an arsenate ion bound. A $F_o$-$F_c$ Fourier electron density map (grey) contoured at 4$\sigma$, prior to inclusion of the arsenate ion (purple) in the model coordinates. **(B, C)** Anomalous difference Fourier maps ($|F^+|$- $|F^-|$, contoured at 3 $\sigma$) of the arsenate-His-ArrX complex from the diffraction data collected at energies near

the As X-ray absorption edge: **(B)** 12110 eV ($f''_{As}$ = 3.65 $e^-$, magenta). **(C)** 11604 eV ($f''_{As}$ = 0.53 $e^-$). Hydrogen-bonding interactions between the arsenate ion (purple) and surrounding residues are indicated as black, dashed lines.

Superposition of the arsenate-His-ArrX with the His-ArrX, malonate-His-ArrX and chloride-soak-His-ArrX structures, gives r.m.s.d. for common Cα positions of 0.18 - 0.48 Å, indicating that minimal global conformational changes accompany the binding of arsenate (Figure 2C). The binding position for the arsenate ion overlaps closely with that of the sulfate. Some of the residues forming the ligand binding sites for arsenate-His-ArrX and His-ArrX are conserved, specifically Thr113, Ser168 and Ser170 (Figure 2). The conformation of residue Tyr138 and the hinge region in the arsenate-His-ArrX structure is different from that observed for His-ArrX (Figure 2B, C), but consistent with the malonate-His-ArrX and chloride-soak-His-ArrX structures (Figure 2A, C). Importantly, residue Cys112 forms a hydrogen bonding interaction with the bound arsenate in the arsenate-His-ArrX structure only (Figure 2C).

***Comparison of ArrX with other structurally characterized PBPs***

The two-domain structure observed for ArrX is highly conserved among PBPs which have been structurally characterized [55]. A secondary structure search of the coordinates of arsenate-His-ArrX versus the Protein Data Bank (PDB) [56, 57] revealed similarities with the substrate-bound structures of the arsenite-binding protein AioX from *Rhizobium* sp. str. NT-26 [46] (PDB 6EU7), the phosphonate-binding protein PhnD from *Echerichia coli* (*Ec*PhnD) [36] (PDB 3P7I) and the phosphite-binding proteins from *Prochlorococcus marinus* (*Pm*PtxB) [58] (PDB 5LV1) and *Pseudomonas stutzeri* [58] (*Ps*PtxB, PDB 5O2J, Table S4). Both the ArrX and AioX proteins are PBPs, which interact with two-component histidine signal transduction systems, whereas the *Ec*PhnD, *Pm*PtxB and *Ps*PtxB proteins are substrate-binding proteins associated with ABC transporters and therefore function in nutrient acquisition.

368 The superposition of the coordinates of the arsenate-His-ArrX complex with arsenite-AioX,
369 the 2-aminoethyl phosphonate complex of *Ec*PhnD (2AEP-*Ec*PhnD), phosphite-*Pm*PtxB and
370 phosphite-*Ps*PtxB yielded r.m.s.d. values of 1.62 - 2.04 Å (Table S4), indicating significant
371 differences between the structures, which is not surprising given their modest sequence
372 identities (Table S4). Despite the overall structural differences however, the residues present
373 at the substrate binding sites are well conserved both structurally and in the primary sequences
374 (aligned and generated with Clustal Omega [59] and ESPript 3.0 [60]) (Figure S2, 4A-D).
375 Specifically, residues Tyr94, Tyr138, Ser168, Ser170 and His199 (all ArrX residue numbering)
376 frame the substrate binding sites and are fully conserved in all four proteins (Figure S2, 4A-
377 D). In both the arsenite-AioX and arsenate-His-ArrX structures, a Cys residue (Cys106 and
378 Cys112, respectively), forms close interactions with the respective bound substrates (Figure
379 4A, B). Finally, in the arsenate-His-ArrX complex, the $AsO_4^{3-}$ ion forms an additional
380 hydrogen bonding interaction with the sidechain of Thr113, which is not conserved (Figure
381 4A, S2).
382
383 The conservation of substrate-interacting residues in these PBP structures does not correlate
384 with the gross conformational changes that are observed on substrate binding, which are
385 different for each of the proteins (Figure S3). A comparison of the *apo*-AioX and arsenite-
386 AioX structures show significantly different loop L1 conformations (Figure 4F), but almost
387 identical structures for the hinge region. As detailed above, the conformation of one of the
388 hinge regions (residues 131-138) is distinct in the His-ArrX structure, compared with those of
389 arsenate-His-ArrX, malonate-His-ArrX and chloride-soak-ArrX, but this does not seem to
390 correlate with the chemical moieties that occupy the substrate binding sites. In contrast to the
391 observations for AioX, in all of these structures the conformation of the L1 loop is the same.
392 Taken together and in the absence of a 'true' *apo*-ArrX structure (that is, with an unoccupied

substrate binding site), it is not possible at this stage to suggest how substrate binding to ArrX might be communicated to its corresponding SHK, ArrS.

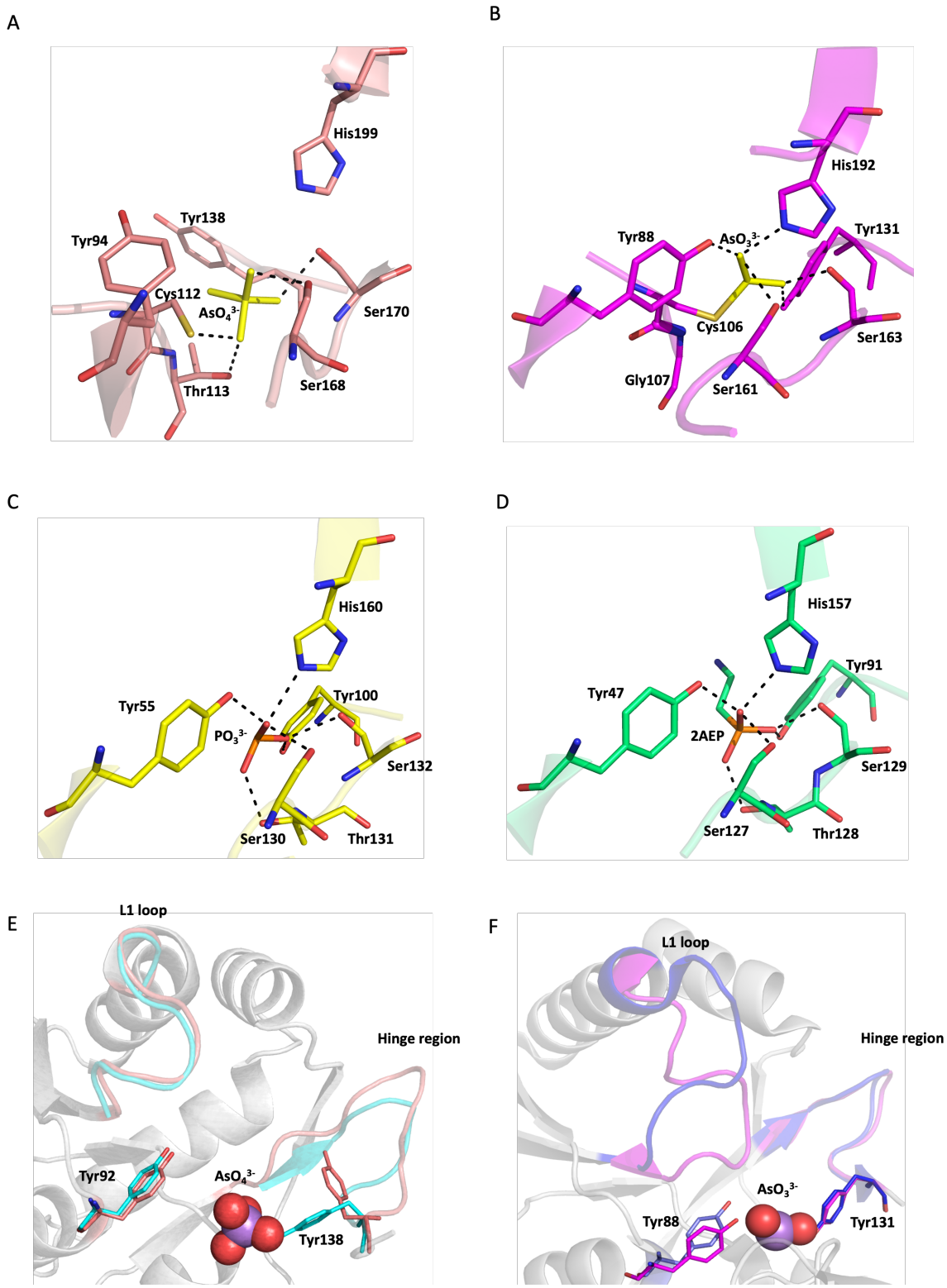


**Figure 4. Comparison of the ligand-binding pockets and loop conformations among PBPs. (A)** The arsenate-His-ArrX complex: the arsenate ion is in yellow ($AsO_4^{3-}$) and surrounding residues in salmon. **(B)** The arsenite-AioX complex (PDB 6EU7 [46]).: the arsenite ion is in yellow ($AsO_3^{3-}$) and surrounding residues are in magenta **(C)** The phosphite-*P*sPtxB complex (PDB 5O2J [58]): the phosphite ion is in orange ($PO_3^{3-}$) and surrounding residues in yellow. **(D)** The 2AEP-PhnD complex (PDB 3P7I [36]): the 2AEP molecule and surrounding residues are shown in green. **(E)** Superposition of arsenate-bound and His-ArrX structures with loops and hinge regions shown in cyan and salmon respectively: the arsenate ion in purple and red spheres ($AsO_4^{3-}$). **(F)** Superposition of the *apo*- and arsenite-bound AioX structures (PDB 6ESK and 6EU7 [46] respectively), with loops and hinge region in indigo and magenta respectively: the arsenite ion in purple and red spheres ($AsO_3^{3-}$). For all representations, hydrogen bonding interactions are indicated with black, dashed lines.

The subtle (or apparently absent) conformational changes observed for the substrate bound AioX and ArrX proteins are in distinct contrast to what has been observed for structurally homologous PBPs that interact with ABC transporters. The structure of substrate-free *Ec*PhnD (PDB 3S4U) was obtained by the generation of a variant protein through mutagenesis to the substrate binding region [36]. The binding of 2AEP to *Ec*PhnD (PDB 3P7I) accompanies a dramatic conformational change, involving a rotation of the C-terminal domain by ~73° at the hinge region to bring the two domains together and close the ligand binding site [36] (Figure S3C, D). The structure of *Pm*PtxB has not been reported in its substrate-free form, however the structures of a homologue, PtxB from *Pseudomonas stutzeri* (*Ps*PtxB) have been described in both *apo*- (PDB 5O2K) and substrate-bound states (PDB 5O2J) [58] (Table S3). Again, for this protein, substrate binding accompanies a significant structural rearrangement with a rotation of the C-terminal domain by ~60° to encapsulate the substrate binding site [58] (Figure S3A, B).

***Probing the roles of residues Cys112 and Tyr138 in the binding of arsenic oxyanions.***

We have previously reported the thermodynamics of substrate binding for the ArrX and AioX proteins and shown that ArrX binds arsenate with $K_D$ = 1.00 ± 0.17 μM and AioX binds to arsenite with $K_D$ = 172 ± 17 nM [46] (Table 2). Given the conservation of the key residues in the substrate binding sites of the two proteins, we sought to investigate the relative contributions of these residues to substrate binding. We mutated ArrX Tyr138 (AioX Tyr131) to Ala and

Cys112 (AioX Cys106) to Ser and Ala. Tyr131 in the AioX structure makes a hydrogen bonding interaction with the bound arsenite, however its equivalent (Tyr 138) faces away from the binding site in the arsenate-His-ArrX complex. AioX Cys106 shows a very close contact with bound arsenite in the arsenite-AioX structure and the equivalent residue Cys112 participates in a hydrogen bonding interaction with arsenate in the arsenate-His-ArrX structure. The determined thermodynamic parameters for arsenate and arsenite binding to the $^{Y138A}$ArrX, $^{C112A}$ArrX, $^{C112S}$ArrX, $^{Y131A}$AioX, $^{C106A}$AioX, $^{C106S}$AioX proteins are given in Table 2.

The mutation of ArrX residue Cys112 to Ser and Ala lowers the affinity of the variants for arsenate by ~3 and ~170-fold, respectively (Table 2, Figure S4A, B), whereas mutation of the equivalent cysteine residue in AioX (to generate the $^{C106A}$AioX and $^{C106S}$AioX proteins) attenuates binding so that it is not detectable by ITC (Table 2). Taken together, these data indicate different roles for this conserved cysteine residue in substrate binding for the two proteins. The structure of arsenite-AioX reveals that the Cys106-arsenite contact is short at 2.3 Å, suggestive of a covalent interaction. Such direct As-S bonding has previously been observed in the structures of the ArsR arsenic-responsive transcriptional repressor proteins from *Acidithiobacillus ferrooxidans* and *Corynebacterium glutamicum* [61]. In contrast, a hydrogen bonding interaction exists between the bound arsenate and residue Cys112 in the arsenate-His-ArrX structure. It follows then that the mutation of Cys106 in AioX eliminates substrate binding, whereas the mutation of Cys112 to Ser and Ala in ArrX, attenuates the interaction between the protein and the substrate by weakening and eliminating respectively, the hydrogen bond.

Further, examination of the AioX and ArrX structures in the region of this conserved Cys residue, reveals that the neighboring residues are different (Figure S2, 4B). In AioX, residue

Gly107 is adjacent to Cys106, which presumably provides structural flexibility to facilitate the covalent Cys-arsenite interaction. In ArrX, the corresponding residue is Thr113, which is modelled in two conformations, one of which hydrogen-bonds to one of the oxygen atoms on the bound arsenate (Figure 2C). It is possible that the identity of the residues neighboring the conserved cysteine contribute to the binding specificities for arsenate or arsenite.

No binding could be detected for arsenate to the $^{Y138A}$ArrX protein or arsenite to the $^{Y131A}$AioX protein (Figure S4C; Table 2). To confirm the integrity of the $^{Y138A}$ArrX protein, we determined its crystal structure (Table 1; Figure S4D). The structure shows the Y138A mutation clearly, with a r.m.s.d. of 0.28 Å for the superposition of 250 common $C^{\alpha}$ positions between the two structures (His-ArrX and $^{Y138A}$ArrX), indicating minimal structural change in the presence of the mutation (Figure S4D, E). The structure of the substrate binding site is maintained in $^{Y138A}$ArrX, with residues Thr113, Ser168 and Ser170 contacting a sulfate ion at the binding site, *via* hydrogen bonding interactions (Figure S4D). Therefore, despite the fact that Tyr138 (AioX Tyr131) shows different conformations in the structures of the substrate-bound complexes of the two proteins, it is crucial to the substrate-protein interaction for both.

## Conclusion

This study for the first time, reveals the structure of the ArrX PBP in the presence and absence of its substrate arsenate. Unfortunately, despite exhaustive attempts, we were not able to determine a 'true' *apo* (substrate-free) structure of ArrX, therefore the precise structural rearrangements that mediate signaling to the corresponding SHK (ArrS) cannot be determined at this time. However, it seems clear that for both the ArrX and AioX proteins, the conformational communication between these PBPs and their corresponding SHKs is more subtle than the mechanisms employed by PBPs such as the PhnD and PtxB proteins, which

show remarkable structural rearrangements on substrate binding and which interact with ABC transporters.

The arsenate and arsenite complexes of ArrX and AioX, respectively revealed a conserved Cys residue at the substrate binding site that interacts differentially with the bound substrate. This may be linked to the identities of the neighboring residues. In AioX, a neighboring Gly residue facilitates the adjustments required for formation of the covalent Cys-Ar interaction. In ArrX, a neighboring Thr residue provides an additional hydrogen bonding interaction to facilitate arsenate binding. Further investigation is required to reveal whether this short sequence signature is solely prescriptive of the discrimination between the oxyanions arsenite and arsenate.

**Acknowledgements**

This research was funded by an Australian Research Council (ARC) Future Fellowship to MJM (FT180100397) and a Biotechnology and Biological Sciences Research Council (BBSRC) grant (BB/N012674/1) to JMS. NP was supported by a University of Melbourne Research Scholarship and CB by a studentship provided by CONICYT, Chile. Part of this study was carried out using the MX2 beamline at the Australian Synchrotron, with is part of ANSTO, and made use of the ACRF detector, in addition to Diamond Light Source beamline I04. We thank the beamline staff for their enthusiastic and professional support. The ITC data were collected at the Melbourne Protein characterization facility, The Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne.

**Accession codes**

The protein sequences of ArrX and AioX were retrieved from UniprotKB accession IDs: UPI00041882C4 and L0NML6 respectively. The coordinates and structure factors for the His-ArrX, arsenate-His-ArrX, chloride-soak-His-ArrX, formate-His-ArrX, malonate-His-ArrX, acetate-His-ArrX and $^{Y138A}$ArrX structures have been deposited in the Protein Data Bank (PDB) with accession codes 6X6B, 6XL2, 7L22, 6XAD, 6X9G, 6XAB and 6X8W respectively.

**Table 1. Diffraction data and refinement statistics**[a]

| **Data collection** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Crystal | His-ArrX | Arsenate-His-ArrX 'native' | Arsenate-His-ArrX 'high E' | Arsenate-His-ArrX 'low E' | Chloride-soak-His-ArrX | Formate-His-ArrX | Malonate-His-ArrX | Acetate-His-ArrX | $^{Y138A}$ArrX |
| Wavelength (Å) | 0.979 | 0.953 | 1.024 | 1.068 | 0.953 | 1.078 | 0.953 | 0.953 | 0.953 |
| Temperature (K) | 100 | | | | | | | | |
| Diffraction source | Diamond light source | Australian Synchrotron MX2 | | | | | | | |
| Detector | Pilatus 6M-F | Eiger 16M | | | | | | | |
| Space group | $P2_1\,2_12_1$ | $P2_1\,2_12_1$ | $P2_1\,2_12_1$ | $P2_1\,2_12_1$ | $P2_1\,2_12_1$ | $P2_1$ | $P2_1\,2_12_1$ | $P2_1$ | $P2_1\,2_12_1$ |
| *a, b, c* (Å) | 52.8, 53.2, 89.8 | 52.8, 53.5, 90.6 | 52.8, 53.4, 90.6 | 52.8, 53.4, 90.7 | 52.9, 54.4, 90.8 | 46.3, 53.1, 54.8 | 52.4, 55.6, 91.4 | 46.6, 52.9, 54.9 | 53.1, 53.4, 90.2 |

| $\alpha, \beta, \gamma$ (°) | 90, 90, 90 | | | | | 90, 107.9, 90 | 90, 90, 90 | 90, 108.3, 90 | 90, 90, 90 |
|---|---|---|---|---|---|---|---|---|---|
| Resolution (Å) | 45.7 – 1.67 (1.71 – 1.67) | 46.0 – 1.74 (1.77 – 1.74) | 46.1 – 1.86 (1.90 – 1.86) | 46.1 – 1.94 (1.99 – 1.94) | 46.7 – 1.92 (1.97 – 1.92) | 44.1 – 1.89 (1.94 – 1.89) | 47.5 – 1.68 (1.71 – 1.68) | 40.6 – 1.78 (1.82 – 1.78) | 46.0 – 1.97 (2.02 – 1.97) |
| Total no. of reflections | 384838 | 180029 | 147056 | 128170 | 472798 | 71277 | 208040 | 65556 | 123881 |
| No. of unique reflections | 29902 | 27069 | 22208 | 19519 | 20389 | 20100 | 31216 | 23640 | 18755 |
| Completeness (%) | 99.4 (98.4) | 99.9 (98.4) | 99.8 (97.8) | 99.7 (96.5) | 99.0 (85.6) | 98.8 (92.4) | 99.8 (96.1) | 96.7 (89.0) | 99.8 (97.4) |
| Redundancy | 12.9 (12.8) | 6.7 (6.6) | 6.6 (6.6) | 6.6 (6.5) | 23.2 (20.9) | 3.5 (2.9) | 6.7 (6.7) | 2.8 (2.7) | 6.6 (6.4) |
| Average $I/\sigma I$ | 10.1 (1.4) | 14.5 (2.7) | 16.1 (3.9) | 15.8 (4.4) | 20.0 (2.7) | 8.1 (1.6) | 15.6 (2.5) | 4.4 (0.9) | 11.0 (2.0) |
| $R_{merge}$ (%) | 0.212 (1.869) | 0.054 (0.502) | 0.051 (0.327) | 0.053 (0.276) | 0.096 (1.356) | 0.099 (0.753) | 0.052 (0.527) | 0.109 (0.904) | 0.092 (0.708) |
| $R_{pim}$ (%) | 0.061 (0.536) | 0.023 (0.207) | 0.021 (0.136) | 0.022 (0.115) | 0.020 (0.295) | 0.060 (0.503) | 0.030 (0.296) | 0.075 (0.649) | 0.039 (0.298) |
| $CC_{1/2}$ | 0.997 (0.524) | 0.999 (0.899) | 0.999 (0.953) | 0.999 (0.965) | 1.000 (0.857) | 0.995 (0.514) | 0.999 (0.900) | 0.992 (0.374) | 0.999 (0.802) |

| **Refinement** | | | | | | | |
|---|---|---|---|---|---|---|---|
| Structure | His-ArrX | Arsenate-His-ArrX | Chloride-soak-His-ArrX | Formate-His-ArrX | Malonate-His-ArrX | Acetate-His-ArrX | $^{Y138A}$ArrX |
| Resolution | 45.7 – 1.67 (1.71 – 1.67) | 46.0 – 1.74 (1.78 – 1.74) | 46.72 – 1.92 (1.97 – 1.92) | 44.1 – 1.89 (1.94 – 1.89) | 47.5 – 1.68 (1.72 – 1.68) | 40.7 – 1.78 (1.82 – 1.78) | 46.0 – 1.97 (2.02 – 1.97) |
| $R_{work}$ (%) | 17.4 (27.2) | 18.8 (26.4) | 16.8 (26.7) | 18.1 (27.7) | 17.0 (23.6) | 18.8 (29.1) | 17.5 (23.4) |
| $R_{free}$ (%) | 20.4 (27.8) | 23.1 (28.9) | 21.3 (27.8) | 22.1 (30.1) | 20.1 (26.7) | 22.1 (25.9) | 20.7 (24.5) |
| No. (%) reflections in test set | 1447 (4.8) | 1371 (5.1) | 1002 (4.9) | 975 (4.9) | 1529 (4.9) | 1164 (4.9) | 970 (5.2) |
| No. of protein molecules per asu | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| No. of other groups per asu | 3 ($SO_4^{2-}$) | 1 ($AsO_4^{3-}$) | 1 ($SO_4^{2-}$) | 1 ($HCO_2^-$) | 1 ($C_3H_2O_4^{2-}$) | 1 ($CH_3COO^-$) | 3 ($SO_4^{2-}$) |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| R.m.s.d bond lengths (Å) | 0.006 | 0.005 | 0.009 | 0.005 | 0.005 | 0.005 | 0.007 |
| R.m.s.d bond angles (°) | 1.36 | 1.27 | 1.48 | 1.31 | 1.27 | 1.31 | 1.43 |
| Average *B*-factors (Å$^2$)[b] | | | | | | | |
| Protein | 31.9 | 26.1 | 42.1 | 26.1 | 22.5 | 30.2 | 30.5 |
| Water | 43.6 | 41.7 | 49.1 | 31.7 | 34.9 | 38.4 | 38.3 |
| Other | 47.2 | 38.5 | 37.9 | 32.7 | 30.1 | 29.8 | 43.0 |
| Ramachandran plot[c] | | | | | | | |
| Most favoured regions (%) | 98.9 | 98.6 | 98.1 | 97.7 | 98.2 | 98.5 | 97.7 |
| Additional allowed regions (%) | 1.1 | 1.4 | 1.9 | 2.3 | 1.8 | 1.5 | 2.3 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Disallowed regions (%) | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 |
| PDB code | 6X6B | 6XL2 | 7L22 | 6XAD | 6X9G | 6XAB | 6X8W |

[a]Values in parentheses are for the highest-resolution shell.

[b]Calculated by BAVERAGE from the CCP4 Suite [50].

[c]Calculated using Molprobity [54]

**Table 2 Thermodynamic parameters for ligand binding to ArrX and AioX.**

| Protein | Ligand | $K_D$ | N | ΔH (kcalmol$^{-1}$) | ΔS (calmoldeg$^{1}$) |
|---|---|---|---|---|---|
| $^{WT}$ArrX [a] *46* | arsenate *46* | 1.00 ± 0.17 μM | 0.82 ± 0.07 | 2050 ± 359 | 34.3 ± 1.3 |
| | sulfate *46* | ND[b] | - | - | - |
| | malonate [a] | ND[b] | - | - | - |
| | acetate [a] | ND[b] | - | - | - |
| | formate [a] | ND[b] | - | - | - |
| $^{C112A}$ArrX[a] | arsenate | 171 ± 12 μM | 1.04 ± 0.14 | 1.04 ± 0.8 | 24.1 |
| $^{C112S}$ArrX[a] | arsenate | 3.26 ± 0.02 μM | 1.11 ± 0.02 | 2613 ± 58 | 33.9 |
| $^{Y138A}$ArrX[a] | arsenate | ND[b] | - | - | - |
| $^{WT}$AioX *46* | arsenite | 172 ± 17 nM | 1.03 ± 0.06 | -1.82E4 ± 0.17E4 | -30.0 ± 5.5 |
| $^{C106A}$AioX[a] | arsenite | ND[b] | - | - | - |
| $^{C106S}$AioX[a] | arsenite | ND[b] | - | - | - |
| $^{Y131A}$AioX[a] | arsenite | ND[b] | - | - | - |

[a] Each parameter represents three independent measurements after background subtraction of the isotherms for buffer only (n=3). Errors are reported as standard error of the mean (± SEM).

[b] Not detected. The range of detection for ITC is from $K_D$ = 1.0 nM – 1.0 mM.

**References:**


[1] Tamaki, S., and Frankenberger, W. T. (1992) Environmental biochemistry of arsenic, In *Reviews of environmental contamination and toxicology*, pp 79-110, Springer.

[2] Oremland, R. S., and Stolz, J. F. (2005) Arsenic, microbes and contaminated aquifers, *Trends in microbiology 13*, 45-49.

[3] Ravenscroft, P., Brammer, H., and Richards, K. (2011) *Arsenic pollution: a global synthesis*, Vol. 94, John Wiley & Sons.

[4] Organization, W. H. (2017) Global status report on water safety plans: a review of proactive risk assessment and risk management practices to ensure the safety of drinking-water, World Health Organization.

[5] BBS, U., and UNICEF. (2011) Bangladesh national drinking water quality survey of 2009, *Dhaka: Bangladesh Bureau of Statistics and UNICEF*.

[6] Argos, M., Kalra, T., Rathouz, P. J., Chen, Y., Pierce, B., Parvez, F., Islam, T., Ahmed, A., Rakibuz-Zaman, M., Hasan, R., Sarwar, G., Slavkovich, V., van Geen, A., Graziano, J., and Ahsan, H. (2010) Arsenic exposure from drinking water, and all-cause and chronic-disease mortalities in Bangladesh (HEALS): a prospective cohort study, *Lancet 376*, 252-258.

[7] Aposhian, H. V., and Aposhian, M. M. (2006) Arsenic toxicology: five questions, *Chem Res Toxicol 19*, 1-15.

[8] Chanpiwat, P., Sthiannopkao, S., Cho, K. H., Kim, K. W., San, V., Suvanthong, B., and Vongthavady, C. (2011) Contamination by arsenic and other trace elements of tube-well water along the Mekong River in Lao PDR, *Environ Pollut 159*, 567-576.

[9] Hughes, M. F. (2002) Arsenic toxicity and potential mechanisms of action, *Toxicology letters 133*, 1-16.

[10] Cohen, S. M., Arnold, L. L., Eldan, M., Lewis, A. S., and Beck, B. D. (2006) Methylated arsenicals: the implications of metabolism and carcinogenicity studies in rodents to human risk assessment, *Crit Rev Toxicol 36*, 99-133.

[11] Jomova, K., Jenisova, Z., Feszterova, M., Baros, S., Liska, J., Hudecova, D., Rhodes, C. J., and Valko, M. (2011) Arsenic: toxicity, oxidative stress and human disease, *J Appl Toxicol 31*, 95-107.

[12] Bissen, M., and Frimmel, F. H. (2003) Arsenic—a review. Part I: occurrence, toxicity, speciation, mobility, *Acta hydrochimica et hydrobiologica 31*, 9-18.

[13] Organization, W. H. Arsenic, Fact sheet N 372. 2012, *Geneva, Switzerland. http://www. who. int/mediacentre/factsheets/fs372/en/(accessed Apr 5, 2016)*.

[14] Stolz, J. F., Basu, P., and Oremland, R. S. (2002) Microbial transformation of elements: the case of arsenic and selenium, *Int Microbiol 5*, 201-207.

[15] Oremland, R. S., and Stolz, J. F. (2003) The ecology of arsenic, *Science 300*, 939-944.

[16] Afkar, E., Lisak, J., Saltikov, C., Basu, P., Oremland, R. S., and Stolz, J. F. (2003) The respiratory arsenate reductase from Bacillus selenitireducens strain MLS10, *FEMS Microbiol Lett 226*, 107-112.

[17] Malasarn, D., Keeffe, J. R., and Newman, D. K. (2008) Characterization of the arsenate respiratory reductase from Shewanella sp. strain ANA-3, *Journal of bacteriology 190*, 135-142.

[18] Krafft, T., and Macy, J. M. (1998) Purification and characterization of the respiratory arsenate reductase of Chrysiogenes arsenatis, *European Journal of Biochemistry 255*, 647-653.

[19] Macy, J. M., Nunan, K., Hagen, K. D., Dixon, D. R., Harbour, P. J., Cahill, M., and Sly, L. I. (1996) Chrysiogenes arsenatis gen. nov., sp. nov., a new arsenate-respiring bacterium isolated from gold mine wastewater, *Int J Syst Bacteriol 46*, 1153-1157.

[20] Ahmann, D., Roberts, A. L., Krumholz, L. R., and Morel, F. M. (1994) Microbe grows by reducing arsenic, *Nature 371*, 750-750.

[21] Slyemi, D., and Bonnefoy, V. (2012) How prokaryotes deal with arsenic(dagger), *Environ Microbiol Rep 4*, 571-586.

[22] Anderson, G., Williams, J., and Hille, R. (1992) The purification and characterization of arsenite oxidase from Alcaligenes faecalis, a molybdenum-containing hydroxylase, *Journal of Biological Chemistry 267*, 23674-23682.

[23] Santini, J. M., and vanden Hoven, R. N. (2004) Molybdenum-containing arsenite oxidase of the chemolithoautotrophic arsenite oxidizer NT-26, *Journal of Bacteriology 186*, 1614-1619.

[24] Glasser, N. R., Oyala, P. H., Osborne, T. H., Santini, J. M., and Newman, D. K. (2018) Structural and mechanistic analysis of the arsenate respiratory reductase provides insight into environmental arsenic transformations, *Proceedings of the National Academy of Sciences 115*, E8614-E8623.

[25] Saltikov, C. W., and Newman, D. K. (2003) Genetic identification of a respiratory arsenate reductase, *Proc Natl Acad Sci U S A 100*, 10983-10988.

[26] Coil, D. A., Lo, J. R., Chen, R., Ward, N., Robb, F. T., and Eisen, J. A. (2013) Draft Genome Sequence of the Arsenate-Respiring Bacterium Chrysiogenes arsenatis Strain DSM 11915, *Genome Announc 1*.

[27] Tam, R., and Saier, M. (1993) Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria, *Microbiology and Molecular Biology Reviews 57*, 320-346.

[28] Karpowich, N. K., Huang, H. H., Smith, P. C., and Hunt, J. F. (2003) Crystal structures of the BtuF periplasmic-binding protein for vitamin B12 suggest a functionally

important reduction in protein mobility upon ligand binding, *Journal of Biological Chemistry 278*, 8429-8434.

[29] Quiocho, F. (1990) Atomic structures of periplasmic binding proteins and the high-affinity active transport systems in bacteria, *Philosophical Transactions of the Royal Society of London. B, Biological Sciences 326*, 341-352.

[30] Ames, G. F.-L. (1986) Bacterial periplasmic transport systems: structure, mechanism, and evolution, *Annual review of biochemistry 55*, 397-425.

[31] Abouhamad, W., Manson, M., Gibson, M., and Higgins, C. (1991) Peptide transport and chemotaxis in Escherichia coli and Salmonella typhimurium: characterization of the dipeptide permease (Dpp) and the dipeptide-binding protein, *Molecular microbiology 5*, 1035-1047.

[32] Chen, C. Y., Berish, S. A., Morse, S. A., and Mietzner, T. A. (1993) The ferric iron-binding protein of pathogenic Neisseria spp. functions as a periplasmic transport protein in iron acquisition from human transferrin, *Molecular microbiology 10*, 311-318.

[33] Quiocho, F. A., and Ledvina, P. S. (1996) Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes, *Molecular microbiology 20*, 17-25.

[34] Lesk, A. M., and Chothia, C. (1984) Mechanisms of domain closure in proteins, *Journal of molecular biology 174*, 175-191.

[35] Sharff, A. J., Rodseth, L. E., Spurlino, J. C., and Quiocho, F. A. (1992) Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis, *Biochemistry 31*, 10657-10663.

[36] Alicea, I., Marvin, J. S., Miklos, A. E., Ellington, A. D., Looger, L. L., and Schreiter, E. R. (2011) Structure of the Escherichia coli phosphonate binding protein PhnD and rationally optimized phosphonate biosensors, *J Mol Biol 414*, 356-369.

[37] Trakhanov, S., Vyas, N. K., Luecke, H., Kristensen, D. M., Ma, J., and Quiocho, F. A. (2005) Ligand-free and-bound structures of the binding protein (LivJ) of the Escherichia coli ABC leucine/isoleucine/valine transport system: trajectory and dynamics of the interdomain rotation and ligand specificity, *Biochemistry 44*, 6597-6608.

[38] Björkman, A. J., and Mowbray, S. L. (1998) Multiple open forms of ribose-binding protein trace the path of its conformational change, *Journal of molecular biology 279*, 651-664.

[39] Magnusson, U., Chaudhuri, B. N., Ko, J., Park, C., Jones, T. A., and Mowbray, S. L. (2002) Hinge-bending motion of D-allose-binding protein from Escherichia coli three open conformations, *Journal of Biological Chemistry 277*, 14077-14084.

[40] Magnusson, U., Salopek-Sondi, B., Luck, L. A., and Mowbray, S. L. (2004) X-ray structures of the leucine-binding protein illustrate conformational changes and the basis of ligand specificity, *Journal of Biological Chemistry 279*, 8747-8752.

[41] Antoine, R., Huvent, I., Chemlal, K., Deray, I., Raze, D., Locht, C., and Jacob-Dubuisson, F. (2005) The periplasmic binding protein of a tripartite tricarboxylate transporter is involved in signal transduction, *Journal of molecular biology 351*, 799-809.

[42] Baraquet, C., Théraulaz, L., Guiral, M., Lafitte, D., Méjean, V., and Jourlin-Castelli, C. (2006) TorT, a member of a new periplasmic binding protein family, triggers induction of the Tor respiratory system upon trimethylamine N-oxide electron-

acceptor binding in Escherichia coli, *Journal of biological chemistry 281*, 38189-38199.

[43] Yeh, J. I., Biemann, H. P., Privé, G. G., Pandit, J., Koshland, D. E., Jr., and Kim, S. H. (1996) High-resolution structures of the ligand binding domain of the wild-type bacterial aspartate receptor, *J Mol Biol 262*, 186-201.

[44] Neiditch, M. B., Federle, M. J., Pompeani, A. J., Kelly, R. C., Swem, D. L., Jeffrey, P. D., Bassler, B. L., and Hughson, F. M. (2006) Ligand-induced asymmetry in histidine sensor kinase complex regulates quorum sensing, *Cell 126*, 1095-1108.

[45] Liu, G., Liu, M., Kim, E. H., Maaty, W. S., Bothner, B., Lei, B., Rensing, C., Wang, G., and McDermott, T. R. (2012) A periplasmic arsenite-binding protein involved in regulating arsenite oxidation, *Environmental microbiology 14*, 1624-1634.

[46] Badilla, C., Osborne, T. H., Cole, A., Watson, C., Djordjevic, S., and Santini, J. M. (2018) A new family of periplasmic-binding proteins that sense arsenic oxyanions, *Scientific reports 8*, 6282.

[47] Kabsch, W. (2010) Integration, scaling, space-group assignment and post-refinement, *Acta Crystallographica Section D: Biological Crystallography 66*, 133-144.

[48] Evans, P. R., and Murshudov, G. N. (2013) How good are my data and what is the resolution?, *Acta Crystallographica Section D: Biological Crystallography 69*, 1204-1214.

[49] McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *Journal of applied crystallography 40*, 658-674.

[50] Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., and McCoy, A. (2011) Overview of the

151 CCP4 suite and current developments, *Acta Crystallographica Section D: Biological*
152 *Crystallography 67*, 235-242.
153 [51] Stein, N. (2008) CHAINSAW: a program for mutating pdb files used as templates in
154 molecular replacement, *Journal of applied crystallography 41*, 641-643.
155 [52] Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics,
156 *Acta Crystallographica Section D: Biological Crystallography 60*, 2126-2132.
157 [53] Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of
158 macromolecular structures by the maximum-likelihood method, *Acta*
159 *Crystallographica Section D: Biological Crystallography 53*, 240-255.
160 [54] Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G.
161 J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-
162 atom structure validation for macromolecular crystallography, *Acta Crystallographica*
163 *Section D: Biological Crystallography 66*, 12-21.
164 [55] Fukami-Kobayashi, K., Tateno, Y., and Nishikawa, K. (1999) Domain dislocation: a
165 change of core structure in periplasmic binding proteins in their evolutionary history,
166 *J Mol Biol 286*, 279-290.
167 [56] Holm, L. (2019) Benchmarking fold detection by DaliLite v.5, *Bioinformatics 35*, 5326-
168 5327.
169 [57] Krissinel, E., and Henrick, K. (2005) Multiple alignment of protein structures in three
170 dimensions, In *International Symposium on Computational Life Science*, pp 67-78,
171 Springer.
172 [58] Bisson, C., Adams, N. B., Stevenson, B., Brindley, A. A., Polyviou, D., Bibby, T. S.,
173 Baker, P. J., Hunter, C. N., and Hitchcock, A. (2017) The molecular basis of
174 phosphite and hypophosphite recognition by ABC-transporters, *Nature*
175 *communications 8*, 1-13.

[59] Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., and Söding, J. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, *Molecular systems biology 7*, 539.

[60] Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server, *Nucleic Acids Research 42*, W320-W324.

[61] Prabaharan, C., Kandavelu, P., Packianathan, C., Rosen, B. P., and Thiyagarajan, S. (2019) Structures of two ArsR As(III)-responsive transcriptional repressors: Implications for the mechanism of derepression, *Journal of Structural Biology 207*, 209-217.