Computational and Structural Biotechnology Journal 21 (2023) 519–534

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Structural and functional mapping of *ars* gene cluster in *Deinococcus indicus* DR1

Shrivaishnavi Ranganathan [a,1], Deepa Sethi [b,1], Sandhya Kasiviswaran [b,2], L. Ramya [c], Richa Priyadarshini [b,*], Ragothaman M. Yennamalli [d,*]

[a] Department of Biotechnology, School of Chemical and Biotechnology, SASTRA Deemed to be University, Thanjavur, Tamil Nadu 613401, India
[b] Department of Life Sciences, School of Natural Sciences, Shiv Nadar University, Gautam Buddha Nagar, Uttar Pradesh, India
[c] Computational and Molecular Biophysics Laboratory, School of Chemical and Biotechnology, SASTRA Deemed to be University, Thanjavur, Tamil Nadu 613401, India
[d] Department of Bioinformatics, School of Chemical and Biotechnology, SASTRA Deemed to be University, Thanjavur, Tamil Nadu 613401, India

ARTICLE INFO

Article history:
Received 22 July 2022
Received in revised form 8 December 2022
Accepted 8 December 2022
Available online 11 December 2022

Keywords:
Arsenic tolerance
Deinococcus indicus
Structural characterization
Molecular modeling

ABSTRACT

Deinococcus indicus DR1 is a novel Gram-negative bacterium, isolated from the Dadri wetlands in Uttar Pradesh, India. In addition to being radiation-resistant, the rod-shaped, red-pigmented organism shows extraordinary resistance to arsenic. The proteins of the corresponding *ars* gene cluster involved in arsenic extrusion in *D. indicus* DR1 have not yet been characterized. Additionally, how these proteins regulate each other providing arsenic resistance is still unclear. Here, we present a computational model of the operonic structure and the corresponding characterization of the six proteins of the *ars* gene cluster in *D. indicus* DR1. Additionally, we show the expression of the genes in the presence of arsenic using qRT-PCR. The *ars* gene cluster consists of two transcriptional regulators (ArsR1, ArsR2), two arsenate reductases (ArsC2, ArsC3), one metallophosphatase family protein (MPase), and a transmembrane arsenite efflux pump (ArsB). The transcriptional regulators are trans-acting repressors, and the reductases reduce arsenate (As^{5+}) ions to arsenite (As^{3+}) ions for favourable extrusion. The proteins modelled using RoseTTAFold, and their conformationally stable coordinates obtained after MD simulation indicate their various functional roles with respect to arsenic. Excluding ArsB, all the proteins belong to the α + β class of proteins. ArsB, being a membrane protein, is fully α-helical, with 12 transmembrane helices. The results show the degree of similarity or divergence of the mechanism utilized by these proteins of *ars* gene cluster in *D. indicus* DR1 to confer high levels of arsenic tolerance. This structural characterization study of the *ars* genes will enable new and deeper insights of arsenic tolerance.

1. Introduction

Arsenic is a toxic metalloid that is released into the environment through natural processes like leaching, and anthropogenic activity like fossil fuel combustion and mining [1]. Arsenic exists primarily in two oxidation states- a pentavalent arsenate ion (As^{5+}) and a trivalent arsenite ion (As^{3+}). Despite being an essential micronutrient with a daily requirement of upto 12 µg per day in humans, arsenic is highly toxic to all living cells [2]. The phosphate transporters and glyceroporin membrane proteins are involved in the uptake of arsenate, As(V); and arsenite, As(III), respectively. The phosphate transporters cannot discriminate between phosphate and arsenate due to structural similarity. Arsenate is a structural analogue of organic phosphate, making it a potent inhibitor of oxidative phosphorylation [3]. Arsenite has a strong affinity towards sulfhydryl functional groups of proteins, disrupting the function of many essential enzymes including those involved in DNA repair and glycolysis [4]. At an organism level, arsenic is a known carcinogen with chronic exposure causing neurological, respiratory, endocrinal, and developmental disorders [5–7].

The prevalence of heavy metals in the environment has caused microorganisms to acquire and select for variants possessing genetic determinants encoding heavy metal tolerance, and in this case,

* Corresponding authors.
E-mail addresses: richa.priyadarshini@snu.edu.in (R. Priyadarshini), ragothaman@scbt.sastra.edu, ragothaman@scbt.sastra.ac.in (R.M. Yennamalli).
[1] Equal Contribution
[2] **Current Address:** Department of Biological Sciences, Mellon Institute, Carnegie Mellon University, 4400 Fifth Avenue, Pittsburgh, PA 15213

https://doi.org/10.1016/j.csbj.2022.12.015

towards arsenic. Arsenic resistance genes are often found on plasmids and confer tolerance through extracellular immobilization, chelation, and transformation into less toxic analogues that are either metabolized or extruded by the cell [8]. These genes are usually encoded within the arsenite inducible transcription unit known as the arsenic operon. Some bacteria have evolved to contain multiple copies of *ars* operon which may be regulated under different conditions. A few examples are *Cornybacterium glutamicum, Pseudomonas putida* KT2440, *Rhodopseudomonas palustris* CGA009, *Bacillus* sp., etc [9–12]. *Herminiimonas arsenicoxydans* is a classic example which has four copies of *ars* operons on the chromosome and is believed to be one of the ancient colonizers in arsenic-rich environment [13]. The number of genes within the operon varies across organisms, where the *ars* operon in *Staphylococcus aureus* contains three genes (*arsRBC*) encoded in pI258 plasmid, while that of *Escherichia coli* has five genes (*arsRDABC*) on the R773 plasmid [14]. Overall operon expression is controlled by a trans-acting repressor element ArsR, while additional control is provided by a second trans-acting protein ArsD [15]. ArsR binds to an operator site proximal to the promoter of the operon and acts as a repressor [16]. Binding of arsenite to the cysteine residues within α-helix of the repressor leads to conformational change in the repressor molecule; helix unwinding and repressor dissociation from the operator site on DNA, thus, inducing arsenic tolerance gene expression [17]. *In vitro*, DNA binding experiments show that arsenite (As[III]) is the true inducer of the operon, that interacts with ArsR, releasing it from its bound state [18]. ArsD is an inducer-independent regulatory protein, which is transcribed along with *arsABC* in the presence of the operon inducer. When intracellular ArsD levels reach a critical concentration, ArsD prevents further *ars* operon expression [15]. Other structural products of the operon include ArsA, an ATPase which associates with ArsB, an integral membrane protein to form a membrane-bound efflux pump ArsAB [19]. In the absence of ArsA, as in the case of *S. aureus*, ArsB acts as a membrane potential driven secondary pump that confers partial resistance by preventing arsenite accumulation in the cell [20]. ArsC is an oxidoreductase that gains reducing power either directly from thiol transfer protein thioredoxin or indirectly from glutaredoxin through glutathione. In *E. coli* R773, Cys-12 in ArsC active site has been shown to be essential for the disulfide bond formation between ArsC and reduced glutathione [21,22]. Multiple *arsC* genes have been identified in the case of *Cornybacterium efficiens* [3]. Additionally, ArsD has a secondary role as a metallochaperone that transfers reduced arsenite from the reductase to the efflux pump, where it is then expelled from the cell [23]. Recently, the function of some more arsenic tolerance genes have been identified such as *arsH* (methylarsenite oxidase) [24], *arsM* (arsenite S-adenosylmethionine methyltransferase), [25,26] *arsI* (C-As lyase) [27] for providing organoarsenical tolerance.

The *Deinococcaceae* family comprises of aerobic, non-spore forming microbes that exhibit extremophilic traits. Consequently, members of this family have been isolated from a range of locations including hot springs, Antarctic soils, desert soil, radiation sites, heavy metal contaminated sites, and the human stomach [28,29]. *Deinococcus* is one of the three primary genera within the family and comprises of members like the radiation and desiccation tolerant *D. radiodurans*, *D. geothermalis,* and *D. deserti* [30–32]. *Deinococcus indicus* DR1, a novel bacterium, was isolated from wetlands in Dadri, Uttar Pradesh, India and studied for its arsenic tolerance potential [33]. *D. indicus* DR1 contains genes for arsenic (*ars* operon), mercuric (*mer* operon), and copper (*cop* operon) tolerance [33]. In this study, we have utilized homology-based approaches to identify the genetic elements present in *D. indicus* DR1 that encode for arsenic tolerance. The arsenic gene cluster present on the chromosome, is flanked on either side by ArsR regulatory proteins, and contains two arsenate reductases ArsC, an efflux pump protein ArsB, and a metallophosphatase family protein. *D. indicus* DR1 genome encodes three arsenate reductases; *arsC1* encoded outside the *ars* gene cluster, *arsC2* and *arsC3* encoded within the cluster. In this study, we use computational methods including *ab initio* molecular modelling, molecular dynamics simulations, docking, and active site prediction tools to study the genes involved in arsenic tolerance in *D. indicus* DR1. This study aims to fill the gap of arsenic utilization in *D. indicus* DR1 through structural models and their functional roles.

2. Materials and methods

2.1. Bacterial strain and growth conditions

Deinococcus indicus strain DR1 was grown in peptone yeast extract (PYE) media at 30 °C under aerobic conditions as described previously [34]. Arsenate and arsenite were added to the growth media as required. All chemicals were obtained from Sigma-Aldrich (USA) and Himedia Pvt. Ltd. (India).

2.2. RNA isolation, cDNA synthesis and real-time quantitative RT-PCR of arsenic resistance genes in D. indicus DR1

To investigate the levels of expression of *ars* genes in *D. indicus* DR1, RT-qPCR was performed for all the 8 genes in the *ars* gene cluster in presence and absence of both arsenate and arsenite. The cells were grown in PYE media and induced at mid exponential phase with 2.5 mM $Na_2HAsO_4 0.7\,H_2O$ [As(V)] (Sigma-Aldrich) and 0.25 mM $NaAsO_2$ [As(III)] (Sigma-Aldrich, USA) for 3 h. The cells without arsenic treatment were used as control. RNA was isolated by TRIzol method [35], with some modifications. Briefly, the cells were pelleted and heated with TRIzol (Ambion, Life Technologies) at 65 °C for 15 mins. Genomic DNA contamination was determined by PCR using Taq polymerase. DNase I (NEB, USA) was used to remove residual DNA if required. cDNA synthesis was done from 1 μg of the extracted RNA with Verso cDNA synthesis kit using random hexamer primers following the kit's protocol (Thermo Scientific, USA). The resulting cDNA was used for qPCR after 1:3 dilution in nuclease free water.

qPCR was performed on the BioRad CFX 96 Real Time PCR system. The 10 μl reaction mixture consisted of 2 μl of cDNA, 300 nM forward and reverse primers, and iTaq Universal SYBR Green Supermix (BioRad, USA) according to the manufacturer's instructions. The primers used for qPCR are listed in Table 1. The conditions for qPCR consisted of an initial denaturation step at 95 °C for 5 mins, followed by 40 cycles of 20 s at 95 °C, 1 min at 60 °C, and 20 s at 72 °C. Each reaction was performed in duplicate. Expression of the *ars* genes was normalized to 16S rRNA expression [36]. Fold change was calculated using the ΔΔct method, or the Livak method [37]. The data is shown as mean fold change over untreated.(Table 2).

2.3. Sequence retrieval

Sequences of the six proteins comprising the *ars* gene cluster were downloaded from NCBI: ArsR1 family transcriptional regulator (OWL98583.1), Metallophosphatase family protein (OWL98584.1), arsenate reductase ArsC2 (OWL98585.1), Arsenical efflux pump membrane protein (OWL98586.1), arsenate reductase ArsC3 (OWL98587.1), and ArsR2 transcriptional regulator (OWL98588.1).

2.4. Sequence alignment and phylogenetic tree construction

Multiple Sequence Alignment (MSA) was performed for ArsB and the arsenite efflux pumps present in all 62 *Deinococcus* species on Mega X software [38] using the MUSCLE alignment tool [39]. A phylogenetic tree was constructed using the Maximum Likelihood method of analysis and the Jones-Taylor-Thornton substitution model, with a bootstrap of 500.

Table 1
List of primers used in qRT-PCR.

Gene	Primer	5'-sequence-3'
arsC1	ArsC1-FP	CTACGAACGCAGCAACCTC
	ArsC1-RP	CGACGGTCAGGTGCTTCC
arsC2	ArsC2-FP	CTCCGCACGGTCACACA
	ArsC2-RP	GCTGCAGGAGTCCGGT
arsC3	ArsC3-FP	TGCACGAAGGCCTCGAA
	ArsC3-RP	GTGACCGTGTGCGACAG
arsR1	ArsR1-FP	GTGTGCACCTGCGACG
	ArsR1-RP	TTCTGCGCTGTGACCAGT
arsR2	ArsR2-FP	TGCGTCTGCGATCTGGAAG
	ArsR2-RP	GGCAACAGGTCGGTCAGC
arsB	ArsB-FP	GCACTGGTGGTGCTGCTG
	ArsB-RP	ACGATGTTCACGAGGTTGCT
Hypothetical protein (CBQ26_01570)	CBQ26_01570-FP	TGACCAACCGCAAGTTCCTG
	CBQ26_01570-RP	GGTGACTGACGTGGTACTGA
Metallophosphatase family protein	Metallophosphatase-FP	TGACCGAGACGACCGAGAA
	Metallophosphatase-RP	GGTACGTCCACGCGGAAT
16 S rRNA	16 S rRNA-FP	CCTACGGGAGGCAGCAGTAG
	16 S rRNA-RP	CAACAGAGCTTTACGATCCGAAA

Table 2
Residues comprising the transmembrane helices of ArsB of *E. coli* and *D. indicus* DR1.

Transmembrane Helix	*E. coli*	*D. indicus* DR1
1	Ala6 - Trp26	Ala4 - Leu21
2	Phe60 - Ala80	Leu28 - Ala50
3	Leu93 - Asn111	Val54 - Leu76
4	Ile121 - Met141	Gly97 - Ala116
5	Thr149 - Phe168	Thr120 - Leu137
6	Val178 - Phe198	Ala144 - Val166
7	Pro221 - Leu241	Tyr181 - Leu203
8	Ile246 - Ala263	Gly233 - Val252
9	Trp280 - Ala299	Val256 - Ile275
10	Leu317 - Asn336	Val280 - Leu302
11	Ser351 - Leu371	Gly322 - Met341
12	Tyr401 - Leu421	Val411 - Val433

Amino acid sequence of ArsB was also aligned with ArsB sequence found in the plasmid R773 of *E. coli* strain K12, since it is a known structural homolog.

Sequence alignment was performed for the three arsenate reductases (ArsC1, ArsC2, and ArsC3) as well, to see the degree of similarity among them, and to identify the family they can be categorized into.

2.5. Molecular modelling of ars gene cluster proteins

The sequences of the five proteins were submitted to RoseTTAFold to predict their tertiary structure. For ArsB, both RoseTTAFold and comparative modelling in the Rosetta server (https://robetta.bakerlab.org/) was used, where comparative modelling involves the usage of sequence homology to experimentally derived structures to predict structure of the proteins. RoseTTAFold uses deep-learning to predict the 3D structure of proteins – the amino acid sequence is broken down into recognizable domains, and these domain sequences are searched against protein structure databases [40]. It provides five models for each query sequence with a confidence score indicating the quality of the models obtained [40]. Similarly, in the comparative modelling approach Rosetta uses a sample of 1000 models by superposing partial threads before performing a hybrid sampling approach.

2.6. Median model and structure validation

To choose the median model among the five predicted structural models for each protein by RoseTTAFold, Theseus software [41] was used, which performs Maximum Likelihood (ML) superposition. In order to validate the 3D structures, the structure files of all the proteins were submitted to PDBsum to conduct PROCHECK analyses [42]. Specifically, structures before and after energy minimization were analyzed using the Ramachandran plot.

2.7. Hydropathy plot

Since ArsB is the only transmembrane protein among the six *ars* gene cluster proteins, it required additional steps before molecular dynamics simulation. The protein structure file was submitted to the TMHMM server that uses a Hidden Markov Model to predict transmembrane residues and construct what is known as a Hydropathy Plot [43]. A hydropathy plot is a graph describing the hydrophobicity of the residues of a protein. When a stretch of amino acids is found to be hydrophobic, they are assumed to form a transmembrane helix. The result obtained from TMHMM was compared to the transmembrane residues of ArsB of *E. coli*, the only known arsenite efflux pump, which is also a structural homolog to ArsB of *D. indicus* DR1.

2.8. Molecular dynamics simulation

Molecular Dynamics (MD) simulation was performed for all the proteins. Five of the *ars* gene cluster proteins (ArsR1, ArsR2, MPase, ArsC2, and ArsC3) were simulated using GROMACS (GROningen MAchine for Chemical Simulations) v5.0.4 to validate the models and obtain an energy minimized structural coordinates [44]. The selected models were kept in a solvent environment using the generalized OPLS forcefield and SPC6 water model. The protein was kept in the center of the cubic box with 1 nm distance from the periodic boundaries to avoid the protein from seeing its own image. The final system used in MD simulations (including solvent and ions) for ArsR1 consisted of 59,364 atoms, 48,521 atoms for ArsR2, 47,525 atoms for MPase, 53,914 atoms for ArsC2, and 29,696 atoms for ArsC3. The system was neutralized by adding appropriate number of Na^+ and Cl^- ions. The system was first energy minimized using the steepest descent method for 50,000 steps or until convergence was achieved earlier than the number of steps. Following this, equilibration was performed in the canonical NVT phase at constant temperature of 300 K and this was followed by NPT equilibration, where the temperature and pressure were stabilized in isothermal-isobaric conditions at 300 K and 1 bar. Both NVT and NPT equilibrations were carried out for 50 ps. The production MD simulation for the five proteins were performed until 100 ns and a frame of velocities and coordinates was dumped every 500 ps. After the simulations, the protein conformation was analyzed by examining the root mean square deviation (RMSD), root mean square fluctuation (RMSF) and radius of gyration (Rg).

ArsB transmembrane protein structural dynamics was studied using NAMD 2.9 in the presence of a lipid bilayer [45]. The membrane bilayer was built using the Membrane Builder tool of CHARMM-GUI server (https://www.charmm-gui.org/) [46,47]. This membrane bilayer consisted of 6 lipid tails (PMPE, POPE, QMPE, PMPG, PSPG, OSPE) in the order of abundance, whose composition is based on the TOP6 membrane model [48].

The system of ArsB dimer in lipid bilayer, solvated in TIP3P water model of cubic water box of dimensions 113.36 Å × 113.36 Å × 109.71 Å and electrically neutralizing ions consists of 1,31,401 atoms. This system was energy minimized using the conjugate gradient method, for 10,000 steps. Following energy minimization, the system was heated to 300 K for 25 ps, after which restrained equilibration was conducted in six steps for 675 ps to relax the system. The system was equilibrated for an additional 5 ns in the NPT ensemble. The subsequent step was the final molecular dynamics production run for 20 ns in the NPT ensemble, whose integration time step was 2 fs. Throughout the simulation, the temperature and pressure of the system were maintained at 300 K using the Langevin thermostat and 1 atm using the Nose-Hoover barostat respectively.

At every 10 ps, the energies and coordinates of the system were recorded, and the trajectory was used for analyses.

After production MD run for all simulations, the trajectory was processed by removing periodic boundary condition effect, removing jumps, and rotating and translating the proteins at the center of the box. For all the six proteins, the last conformation at the end of the trajectory was used for further analysis, such as active site prediction and finding structural homologs.

2.9. Active site prediction

Active site prediction was performed for all the proteins (except ArsB) using the Computed Atlas of Surface Topology of proteins (CASTp) server [49]. CASTp analyzes the structural file of the protein, and all pockets and voids are predicted as an active site. In the event of the presence of multiple such pockets in the protein, the largest pocket is assumed to be the active site.

2.10. Structural homologs

The DALI server [50] was used to obtain the structural homologs for the proteins of *ars* gene cluster (excluding ArsB). DALI searches the Protein Data Bank (PDB) to find homologs using the given structural coordinates. The top 10 structurally similar proteins with their specific chain were noted along with their RMSD values and Z-score. The functionally important residues of each of these homologs were compared to those of *D. indicus* DR1 by performing structural alignment and obtaining the structurally equivalent residues, to identify the active site residues as predicted by CASTp.

3. Results

The arsenic efflux system of *D. indicus* DR1 consists of an *ars* gene cluster and an arsenate reductase, ArsC1, located external to the cluster. The arsenic resistance gene cluster itself is comprised of 6 proteins – two transcriptional regulators flanking the cluster (referred as ArsR1 and ArsR2 henceforth), a metallophosphatase family protein (referred as MPase henceforth), two arsenate reductases (referred as ArsC2 and ArsC3 henceforth) and a transmembrane efflux protein (ArsB). Here, ArsR1 and ArsR2 are trans-acting repressors, ArsC2 and ArsC3 reduce arsenate (As^{5+}) to arsenite (As^{3+}) ions, ArsB is the inner membrane transmembrane protein involved in the extrusion of arsenite, and the function of MPase in this *ars* cluster is unknown.

3.1. Modelling ars gene cluster proteins, selecting a representative model, and their validation using molecular dynamics

The RoseTTAFold prediction statistics are given in the Supplementary Table T1. The ArsR1 and ArsR2 transcriptional regulators had a confidence score of 0.84 and 0.81, respectively, and for the arsenate reductases (ArsC2 and ArsC3) the confidence score was 0.90 and 0.88, respectively. MPase was predicted to have an α/β fold with a confidence score of 0.86, while ArsB was predicted to be a homodimer consisting of 12 α-helices with a confidence score of 0.55 using RoseTTAFold (Supplementary Table T1). At the same time, ArsB was predicted as a monomer consisting of 17 α-helices with a confidence score of 0.83 using the comparative modeling approach of Rosetta. The confidence score for a model represents how structurally close or distant it is to a target structure (which is an experimentally verified structure for a known protein). This number is derived from distances of individual residues of the prediction from that of the target against a threshold of 4 Å. Higher the confidence score, the more accurate the prediction. The lower confidence score for ArsB is due to the presence of multiple sections in the predicted structure containing an error estimate of more than 2 Å. Regions in a protein with higher error estimates are considered to harbor modeling errors.

Using Theseus software [51], the median model for all the proteins were selected. Model 3, Model 1, Model 5, Model 1, Model 5, and Model 3 were selected as the median models of ArsR1, ArsR2, MPase, ArsC2, ArsC3, and ArsB, respectively. These models were used for subsequent analyses. Validation of protein structures was performed by generating Ramachandran plot (Supplementary Figs. S1-S3, S6-S8, and S12) using PROCHECK [52]. The presence of less than 1% of the residues in the disallowed regions for all the proteins was an indicator of well-modelled structures. For the proteins of *ars* gene cluster of *D. indicus* DR1, only one active site was predicted by CASTp for each of them.

Molecular Dynamics (MD) Simulation for the proteins ArsR1, ArsR2, MPase, ArsC2, and ArsC3 was performed in explicit water environment using the software GROMACS for 100 ns, and ArsB was simulated in the lipid bilayer using NAMD program for 20 ns. The structures generated after energy minimization were deposited at ModelArchive.

(ArsR1: https://www.modelarchive.org/doi/10.5452/ma-m92cj;.
ArsR2: https://www.modelarchive.org/doi/10.5452/ma-dkxdu;.
MPase: https://www.modelarchive.org/doi/10.5452/ma-j0xxh;.
ArsC2: https://www.modelarchive.org/doi/10.5452/ma-28bjc;.
ArsC3: https://www.modelarchive.org/doi/10.5452/ma-xpwd6; and.
ArsB: https://www.modelarchive.org/doi/10.5452/ma-yrymn).

The obtained trajectories were utilized to examine the behavior of these proteins in a simulated environment, and many properties were analyzed, which are discussed in the subsequent sub-sections.

3.2. ArsR1 and ArsR2 are diverse

arsR1 (NCBI Accession ID: OWL98583.1) is the first gene and ***arsR2*** transcriptional regulator (NCBI Accession ID: OWL98588.1) is the sixth and last gene located in the arsenic resistance gene cluster of *D. indicus* DR1 (Fig. 1 and Fig. 2). Both genes encode trans-acting repressors of the *ars* gene cluster, referred to as ArsR1 and ArsR2 respectively in this manuscript. Crystal structures of ArsR1 protein from organisms such as *Acidithiobacillus ferrooxidans* and *Corynebacterium glutamicum* show that it is a homodimer in its active state, interacting with up to two arsenite ions [53]. It contains a DNA binding domain which interacts with the *ars* operon promoter and its active site cysteine residues allow it to interact with the As^{3+} oxyanions. In *D. indicus* DR1, the predicted ArsR1 protein structure is a monomer consisting of six α-helices and two β-strands (Fig. 2A), and that of ArsR2 consists of eight α-helices and three β-strands (Fig. 2E). The Ramachandran plot of both proteins' structures show that there are no residues in the disallowed region, indicating that all the residues are sterically stable (Supplementary Fig. S1 and S2). The α-helices are labelled α_1(Thr6-Leu13), α_2(Asp15-Thr27), α_3(Asp43-Gly49), α_4(Ser51-Gly66), α_5(Gly85-Arg90), α_6(Gly94-Ala99) in ArsR1; and α_1(Thr8-Ile13), α_2(Lys14-Leu16), α_3(Gln18-Ala29), α_4(Val35-Asn43), α_5(Pro45-Gly60), α_6(Ser76-Gly84), α_7(Gly85-Leu92), α_8(Ala97-Ser102) in ArsR2. The β-strands in ArsR1 are labelled β_1(Leu67-Gln71), β_2(Tyr78-Ser82), and in ArsR2, they are labelled β_1(His33-Cys34), β_2(Leu61-Arg67) and β_3(Asn70-Leu75). To check the degree of similarity between ArsR1 and ArsR2, sequence and structural alignment was performed for their amino acid sequences. This revealed a lower sequence percentage identity of 36.99% between them as shown by BLAST [54], but the structures were similar as the C-α atoms RMSD was 2.71 Å (Supplementary Table T2).

For both proteins, the conformation obtained at the end of 100 ns of MD simulation was analyzed for potential binding sites. For this, active sites were predicted using CASTp server (Fig. 2B for ArsR1 and Fig. 2F for ArsR2). In ArsR1, this active site is 28.534 $Å^2$ in area, and 13.025 $Å^3$ in volume and in ArsR2, the active site has an area of 71.969 $Å^2$, and a volume of 42.280 $Å^3$. The active site residues in

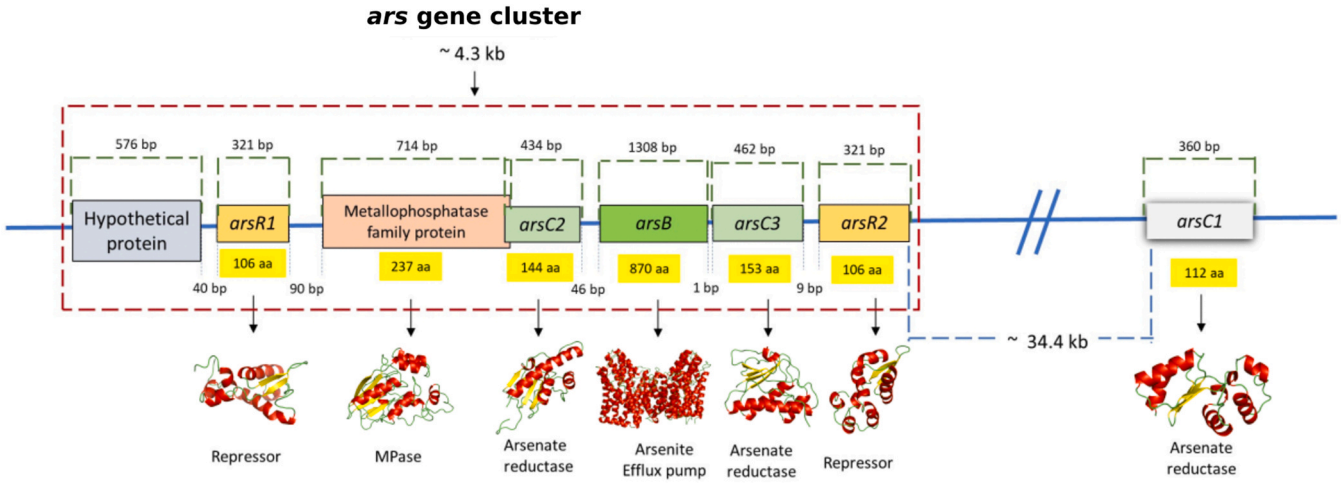

Fig. 1. The organization of *ars* gene cluster in *Deinococcus indicus* DR1 depicting the gene locations. The predicted and validated structures of the *ars* gene cluster are shown below each gene. All the structures except ArsB belong to α/β structural class of proteins. ArsB is completely made of α helices because it is a transmembrane protein involved in the efflux of arsenite ions. As described in the MM section, the validated structures are equilibrated and are the lowest binding energy conformation, obtained after 100 ns of MD Simulation. The subsequent sub-sections are discussed with respect to each protein and suitable comparison-based discussions are elaborated.

ArsR1 consist of Met1, Phe3, Arg11, Ala98, Asp104, Ala105, and Ala106 (Fig. 2C), while those in ArsR2 consist of Leu16, Gln18, Asp19, Tyr22, Arg26, Leu92, Pro93, Asp94, and His95 (Fig. 2G). Structural comparison of these proteins with homologs derived from DALI as shown in Fig. 2D (ArsR1) and Fig. 2H (ArsR2) indicate that the top hits are transcriptional regulators from *Acidithiobacillus ferrooxidans*

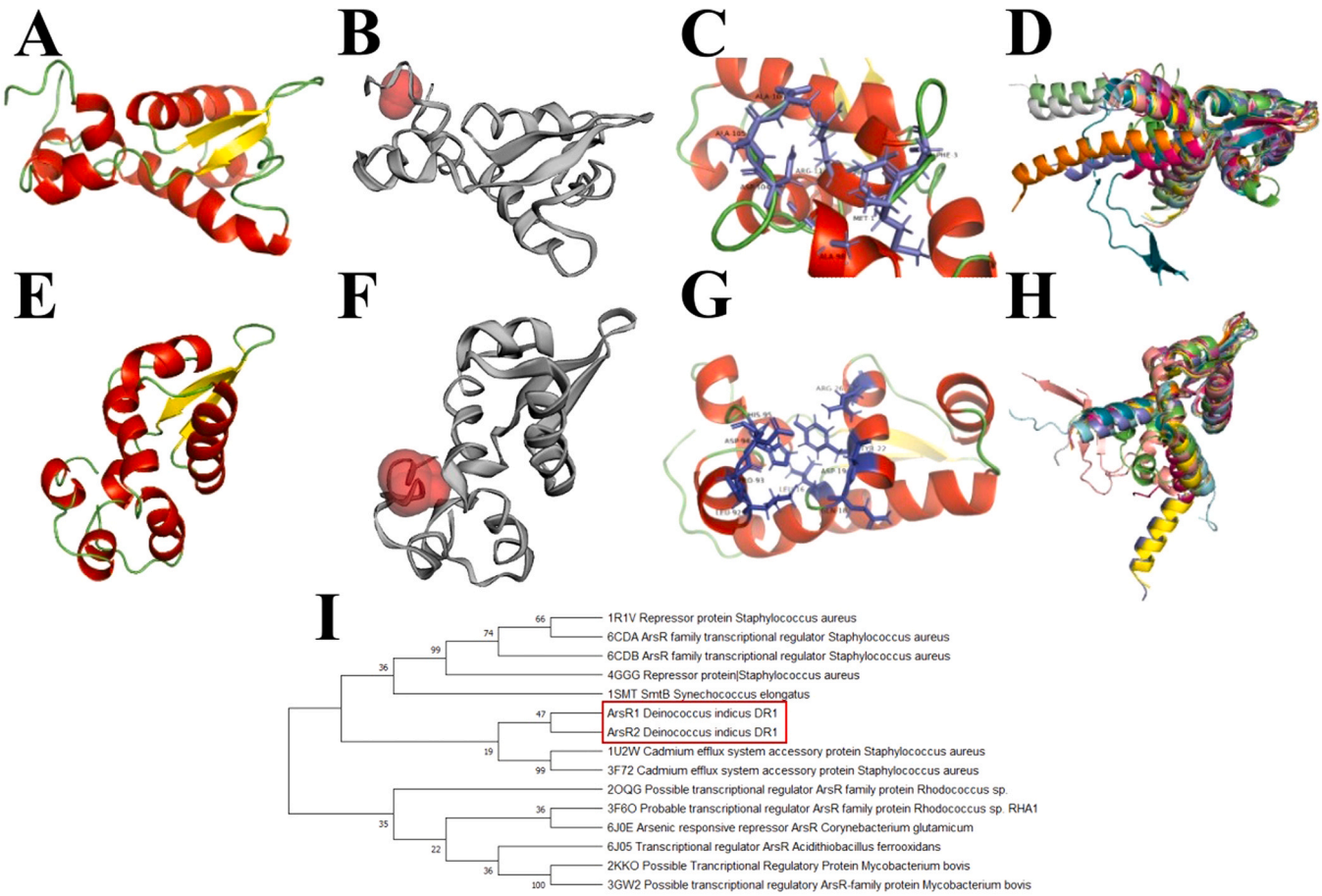

Fig. 2. : *ars* gene cluster transcriptional regulators ArsR1 and ArsR2: (A) ArsR1 after 100 ns of Molecular Dynamics Simulation. (B) Active site pocket (pink) of ArsR1 predicted by CASTp. (C) Active site residues (coloured blue, shown as sticks) of ArsR1 predicted using CASTp. The residues are Met1, Phe3, Arg11, Ala98, Asp104, Ala105, and Ala106. (D) Superposition of ArsR1 with structural homologs identified using DALI (PDB ID: 6J05, 6J0E, 1U2W, 2OQG, 3F6O, 3F72, 1R1U, 4GGG, 3GW2, and 6CDA) showing fold level conservation among various transcriptional regulators. (E) ArsR2 after 100 ns of Molecular Dynamics Simulation. (F) Active site pocket (pink) of ArsR2 predicted by CASTp. (G) Active site residues (coloured blue, shown as sticks) of ArsR2 predicted using CASTp. The residues are Leu16, Gln18, Asp19, Tyr22, Arg26, Leu92, Pro93, Asp94, and His95. (H) Superposition of ArsR2 with structural homologs identified using DALI (PDB ID:1R1U, 1R1V, 1SMT, 1U2W, 2KKO, 3F72, 4GGG, 6CDA, 6CDB, and 6J0E) showing fold level conservation among various transcriptional regulators. (I) Phylogenetic tree of ArsR1 and ArsR2 with their combined structural homologs obtained from the DALI server.

(PDB ID: 6J05) [53], *Corynebacterium alumaticum* (PDB ID: 6J0E) [53], *Staphylococccus aureus* (PDB ID: 1U2W, 3F72, 1R1U, 4GGG, and 6CDA) [55–59], *Rhodococcus jostii* (PDB ID: 2OQG and 3F6O), and *Mycobacterium tuberculosis* (PDB ID: 3GW2) for the former and transcriptional regulators from *Staphylococccus aureus* (PDB ID: 1R1U, 4GGG, 1U2W, 1R1V, 3F72, 6CDB, and 6CDA) [55–59], *Mycobacterium tuberculosis* (PDB ID: 2KKO), *Corynebacterium alumaticum* (PDB ID: 6J0E) [53], and *Synechococcus elongatus* (PDB ID: 1SMT) [60] for the latter. The common structural homologs between the two ArsRs are 6J0E, 1U2W, 3F72, 1R1U, 4GGG, and 6CDA, all of whom are transcriptional repressors in *Staphylococcus aureus*, excluding 6J0E, which is in *Corynebacterium alumaticum* [53]. Upon superposition with these structural homologs, ArsR1 showed high structural similarity where the C-α RMSD ranged between 2.4 and 5.2 Å (Supplementary Table T2). Similarly, for ArsR2, this value was in the range 3–4 Å. When ArsR1 and ArsR2 were aligned with their respective structural homologs, none of the active site residues predicted by CASTp for both proteins showed any consensus among the homologs obtained from the DALI server. Multiple Sequence Alignment (MSA) was performed, and a phylogenetic tree was constructed with the amino acid sequences of both these proteins (Fig. 2I) with their structural homologs obtained from the DALI server to trace their relative evolutionary relationships. It was found that the sequence percentage identity for their alignment ranges between 17.53 % and 31.31 % for ArsR1 and 18.10 % and 29.47% for ArsR2, numbers that indicate that there is evolutionary divergence from these structural homologs. The phylogenetic tree revealed the collective closeness of both ArsRs of *D. indicus* DR1 to the Cadmium efflux system regulators of *Staphylococcus aureus*.

Modeling ArsR1 and ArsR2 using both RoseTTAFold and AlphaFold (using ColabFold) [61] showed highly similar tertiary structures (Fig. 3). The structure of ArsR1 predicted by AlphaFold and RoseTTAFold show the monomeric structure (Fig. 3A) in highly similar conformation (RMSD of 0.797 Å) and with high pLDDT score of 91.4 in AlphaFold and confidence score of 0.84. RoseTTAFold does not predict oligomeric states, hence the oligomeric states was predicted using ColabFold. Comparing the dimeric and tetrameric states prediction, the dimeric conformation (Fig. 3B) had a higher ptm (0.898), and iptm (0.902) scores than the tetrameric conformation, where the ptm score was 0.544, and iptm score was 0.418. Visualizing the tetrameric conformation of ArsR1 also indicated that the tertiary structure most likely prefers to be in dimeric state than tetrameric state. The PAE plots of the monomer, dimer, and tetramer are shown in Supplementary Fig. S4.

In the case of ArsR2, we observed similar trend in terms of tertiary structure prediction, where the structure predicted by RoseTTAFold and AlphaFold are highly similar (RMSD of 0.786 Å) (Fig. 3C). Specifically, the confidence score of RoseTTAFold structure was 0.81 and the pLDDT score for the monomer was 89.2 and a ptm score of 0.722 indicating high quality structure prediction. Similarly, prediction of dimeric (Fig. 3D) and tetrameric (Fig. 3E) states of ArsR2 gave high confidence oligomeric states, where it looks like both dimer and tetramer formation are highly possible. This is evident by the high scores obtained in the three parameters: the pLDDT score for the dimer and tetramer were higher, 88.3 and 88.7, respectively; the ptm score for the dimer and tetramet were 0.818 and 0.874, respectively; and finally the iptm score for the dimer and tetramer were 0.804 and 0.872, respectively. The PAE plots of the monomer, dimer, and tetramer are shown in Supplementary Fig. S5.

Additionally, the Ramachandran plots obtained for the structures by RoseTTAFold for ArsR1 (Supplementary Fig. S1a) and for ArsR2 (Supplementary Fig. S2a) in comparison with the Ramachandran plots for the structures by AlphaFold for ArsR1 (Supplementary Fig. S3a) and for ArsR2 (Supplementary Fig. S3b) shows that both models are highly similar and of high quality.

3.3. MPase could be an ATPase

Mpase (NCBI Accession ID: OWL98584.1) is the second gene in the *ars* gene cluster of *D. indicus* DR1, encoding a metallophosphatase family protein, referred as MPase in this manuscript. Metallophosphatases are a diverse superfamily of proteins with varied functions. They all share a conserved domain of a double β-sheet sandwich, containing metal ions in their active site, that are stabilized by histidine, arginine, and aspartate residues [62]. In the *D. indicus* DR1 *ars* gene cluster, the metallophosphatase could potentially function as an ATPase. We analyzed the structure of the metallophosphatase protein from arsenic gene cluster and found that MPase consists of nine α-helices and 10 β-strands (Fig. 4A). The Ramachandran plot of the structure shows that there are no residues in the disallowed region, indicating that all the residues are sterically stable (Supplementary Fig. S6a and S6b). The alpha helices are labelled α_1(Arg13-Ala20), α_2(Asp22-Arg27), α_3(Gly46-Ala56), α_4(Gly65-Arg70), α_5(Met84-Leu92), α_6(Asp139-Arg144), α_7(Asn210-Trp217), α_8(His224-Ala230), α_9(Gln231-Thr234) and the beta strands are labelled β_1(Arg2-Phe6), β_2(Ala30-Leu34), β_3(Leu61-Val63), β_4(Thr105-Leu109), β_5(Glu113-His118), β_6(Arg152-Val156), β_7(Gln163-Ile167), β_8(Val170-Asn174), β_9(Ala188-Gly196), β_{10}(Ile199-

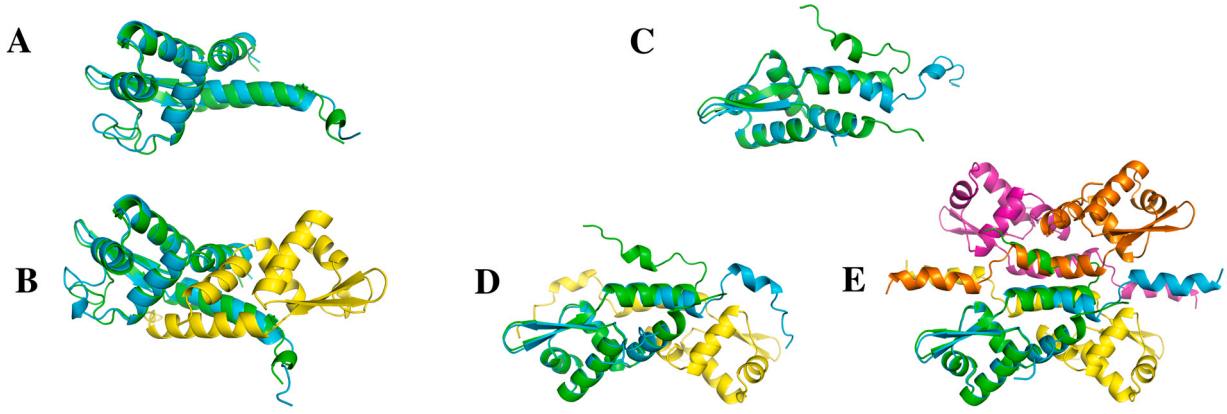

Fig. 3. : AlphaFold prediction of ArsR1 and ArsR2, their oligomeric states, and comparison with RoseTTAFold predicted structures. (A) ArsR1 of RoseTTAFold (colored green) is superposed with ArsR1 predicted by AlphaFold indicating highly similar structures with a RMSD of 0.797 Å. (B) Predicted dimeric state of ArsR1 (colored cyan and yellow) in comparison with the RoseTTAFold predicted ArsR1 (colored green) structure. The tetrameric state is not predicted as a possibility (refer text for discussion) (C) ArsR2 of RoseTTAFold (colored green) is superposed with ArsR2 predicted by AlphaFold indicating highly similar structures with a RMSD of 0.786 Å. (D) Predicted dimeric state of ArsR2 (colored cyan and yellow) in comparison with the RoseTTAFold predicted ArsR2 (colored green) structure. (E) Predicted tetrameric state of ArsR2 (colored cyan, yellow, orange, and magenta) in comparison with the RoseTTAFold predicted ArsR2 (colored green) structure, indicating the possibility of higher order oligomeric states in ArsR2.

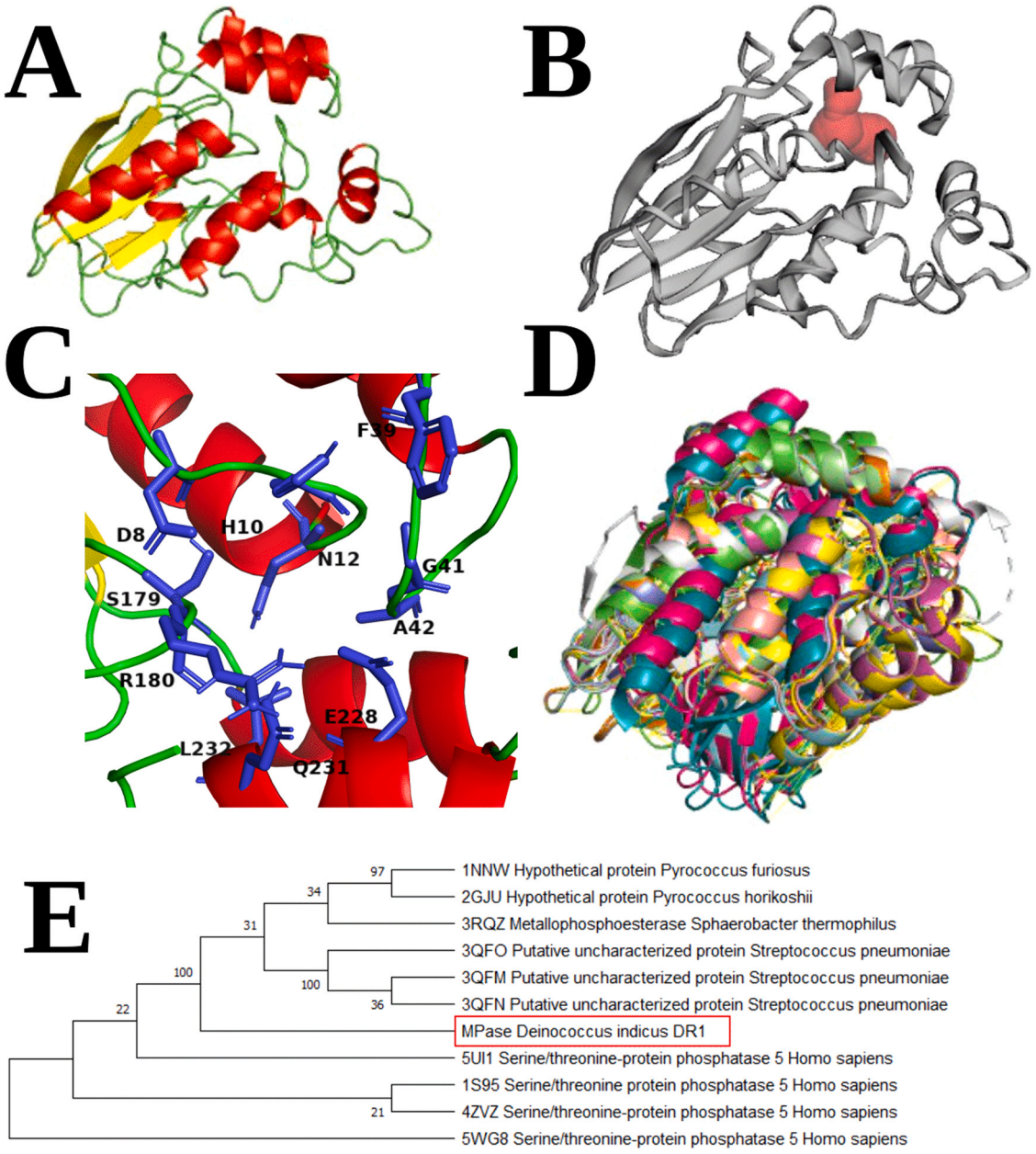

Fig. 4. : Metallophosphatase Family protein MPase: (A) MPase after 100 ns of Molecular Dynamics Simulation. (B) Active site pocket (pink) of MPase predicted by CASTp. (C)Active site residues (coloured blue, shown as sticks) of MPase predicted using CASTp. The residues are Asp8, His10, Asn12, Phe39, Gly41, Ala42, Ser179, Arg180, Glu228, Gln231 and Leu232. (D) Superposition of MPase with structural homologs identified using DALI (PDB ID: 1NNW, 2GJU, 3QFO, 3QFM, 3QFN, 3RQZ, 1S95, 5UI1, 5WG8, and 4ZVZ) showing fold level conservation among various phosphatases. (E) Phylogenetic tree of MPase with its structural homologs as given by the DALI server.

Val207). Analyzing the conformation obtained at the end of 100 ns of MD simulation of MPase for potential binding sites, active site prediction was obtained from CASTp, as shown in Fig. 4B. Specifically, the active-site residues are Asp8, His10, Asn12, Phe39, Gly41, Ala42, Ser179, Arg180, Glu228, Gln231, and Leu232 (Fig. 4C). Structural comparison of MPase with homologs derived from DALI as shown in Fig. 4D indicates that the top hits are Phosphatase enzymes from *Pyrococcus furiosus*(PDB ID: 1NNW), *Pyrococcus horikoshii* (PDB ID: 2GJU), *Streptococccus pneumoniae* (PDB ID: 3QFO, 3QFM, and 3QFN), *Sphaerobacter thermophilus* (PDB ID: 3RQZ) and *Homo sapiens* (PDB ID: 1S95, 5UI1, 5WG8, and 4ZVZ) [63–65]. The structure is highly similar to these known structural homologs, where the C-α atoms RMSD ranges from 3.4 to 4.8 Å (Supplementary Table T2), MSA with its structural homologs reveals a sequence identity range of 17.89 %– 28.36 %, and a phylogenetic tree was constructed with this set (Fig. 4E). The results of the superposition and MSA suggest evolutionary divergence from the known structural homologs.

When MPase was aligned with its structural homologs, three active site residues, namely Asp8, His10 and Phe39 had structural equivalence with Asp242, His244, and Asp274 respectively from all identified homologs. While Asp242 and His244 are involved in metal coordinating, Asp274 stabilizes the active site with the help of hydrogen bond with His304, with another hydrogen bond between Arg275 and Tyr451 [64]. We hypothesize that this function can be extrapolated to the structurally equivalent residues in the *ars* gene cluster MPase.

3.4. ArsC2 and ArsC3 arsenate reductases are different from ArsC1

arsC2 (NCBI Accession ID: OWL98585.1) is the third gene in the *ars* gene cluster of *D. indicus* DR1. This gene's translated product is a

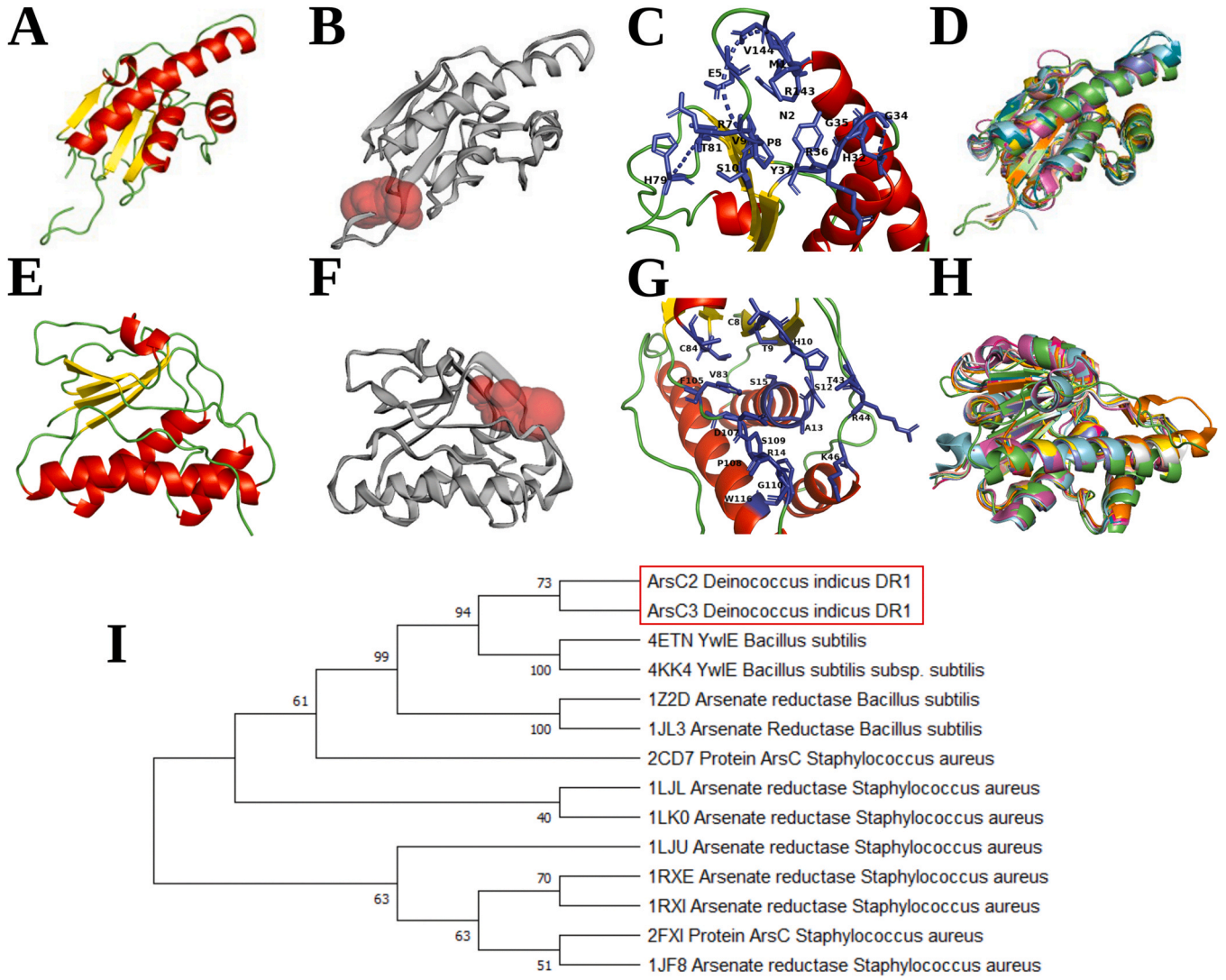

Fig. 5. : Arsenate reducatses ArsC2 and ArsC3: (A) ArsC2 after 100 ns of Molecular Dynamics Simulation. (B) Active site pocket (pink) of ArsC2 predicted by CASTp. (C) Active site residues (coloured gray, shown as sticks) of ArsC2 predicted using CASTp. The residues are Met1, Asn2, Glu5, Arg7, Pro8, Val9, Ser10, His32, Gly34, Gly35, Arg36, Tyr37, His79, Thr81, Arg143, and Val144. (D) Superposition of ArsC2 with structural homologs identified using DALI (PDB ID: 1JF8, 1JL3, 1LJL, 2CD7, 1LK0, 2FXI, 1RXI, 1RXE, 1LJU, and 1Z2D) showing fold level conservation among various transcriptional regulators. (E) ArsC3 after 100 ns of Molecular Dynamics Simulation. (F) Active site pocket (pink) of ArsC3 predicted by CASTp. (G) Active site residues (coloured blue, shown as sticks) of ArsC3 predicted using CASTp. The residues are Cys8, Thr9, His10, Ser12, Ala13, Arg14, Ser15, Thr43, Arg44, Lys46, Val83, Cys84, Phe105, Asp107, Pro108, Ser109, Gly110, and Trp116. (H) Superposition of ArsC3 with structural homologs identified using DALI (PDB ID: 1RXE, 2FXI, 1RXI, 1JL3, 1JF8, 1LK0, 2CD7, 1LJL, 4ETN, and 4KK4) showing fold level conservation among various transcriptional regulators. (I) Phylogenetic tree of ArsC2 and ArsC3 with their collective structural homologs, which were listed using the DALI server.

low molecular weight phosphatase family protein, as annotated by NCBI, labeled as ArsC2 in this manuscript. The predicted and validated ArsC2 protein structure consists of six α-helices and four β-strands (Fig. 5A). The Ramachandran plot of the structure shows that there are no residues in the disallowed region, indicating that all the residues are sterically stable (Supplementary Fig. S7a and S7b). The α-helices are labelled α_1(Ala20-His31), α_2(Asn50-Val54), α_3(Gln55-Ser60), α_4(Pro74-Glu78), α_5(Cys87-Ala92), and α_6(Ser117-Arg143) and the β-strands are labelled β_1(Val9-Cys15), β_2(Ala38-Leu44), β_3(Thr81-Val86), and β_4(Tyr102-Trp106). Analyzing the conformation obtained at the end of 100 ns of MD simulation for potential binding sites, active site was obtained from CASTp, as shown in Fig. 5B. Specifically, the active-site residues are Met1, Asn2, Glu5, Arg7, Pro8, Val9, Ser10, His32, Gly34, Gly35, Arg36, Tyr37, His79, Thr81, Arg143, and Val144 (Fig. 5C). Structural comparison of ArsC2 with other homologs derived from DALI as shown in Fig. 5D indicates that the top hits are arsenate reductases from *Staphylococcus aureus* (PDB ID: 1JF8, 1LJL, 2CD7, 1LK0, 2FXI, 1RXI, 1RXE and 1LJU) [66–69], and *Bacillus subtilis* (PDB ID: 1JL3 and 1Z2D) [70,71]. ArsC2 structure is highly similar to other known structural homologs, where the C-α atoms RMSD ranges from 2.1 to 2.9 Å (Supplementary Table T2). Superposition with structural homologs obtained from DALI server was done to identify structurally equivalent active residues.

All the known structural homologs had a common set of active site residues, which are Cys10, Arg16, Cys82, Cys89 and Asp105, along with a conserved Asn13, that helps in the proper conformation of the active site. When ArsC2 was aligned with these proteins, structural equivalence was observed from the presence of identical residues at the aligned sites. But these residues were not part of the predicted active site residues from CASTp server.

arsC3 (NCBI Accession ID: OWL98587.1) is the fifth gene in the *ars* gene cluster of *D. indicus* DR1. The protein product of this gene is also annotated as low molecular weight phosphatase family protein, and for the sake of nomenclature and ease of addressal, referred as ArsC3 in this manuscript. The predicted and validated ArsC3 protein structure consists of four α-helices and four β-strands (Fig. 5E). The

Ramachandran plot of the structure shows that there is only one residue in the disallowed region (Ser109), however all the other residues are sterically stable (Supplementary Fig. S8a and S8b). The α-helices are labelled α_1(Ser12-Gly30), α_2(Lys46-Glu55), α_3(Cys84-Glu89), and α_4(Ser112-Gly137), and the β-strands are labelled β_1(Thr2-Cys8), β_2(Pro32-Thr40), β_3(Asp78-Val83), and β_4(Thr99-Tyr103). Analyzing the conformation obtained after 100 ns of MD simulation of ArsC3 for potential binding sites, active site was obtained from CASTp, as shown in Fig. 5F. Specifically, the active-site residues are Cys8, Thr9, His10, Ser12, Ala13 Arg14, Ser15, Thr43, Arg44, Lys46, Val83, Cys84, Phe105, Asp107, Pro108, Ser109, Gly110, and Trp116 (Fig. 5G). Structural comparison of ArsC3 with other homologs derived from DALI as shown in Fig. 5H indicates that the top hits are arsenate reductases from *Staphylococcus aureus* (PDB ID: 1JF8, 1LJL, 2CD7, 1LK0, 2FXI, 1RXI, and 1RXE) [66–69], and *Bacillus subtilis* (PDB ID: 1JL3) [70] and low molecular weight protein tyrosine phosphatases (LMW PTPases) from *Bacillus subtilis* (PDB ID: 4ETN and 4KK4) [72]. ArsC3 structure is highly similar to other known structural homologs, where the C-α atoms RMSD ranges from 2.3 to 3.3 Å (Supplementary Table T2). Among the structural homologs identified from DALI, specifically in the arsenate reductases, Cys10, Asn13, Arg16, Cys82, Cys89, and Asp105 comprise the catalytic site residues and in the LMW PTPases, Cys7, Thr11, Arg13, and Asp118 constitute the residues of the active site. The structurally equivalent residues for the former were found to be Cys8, Arg14, Cys84, Cys91, and Asp107 respectively. For the latter, the structurally equivalent residues in ArsC3 are Cys8, Ser12, Arg14, and Asp107, respectively. In the first case, Cys10 acts as a nucleophile, and attacks arsenate ion, with the help of Asp105. Next, Cys82 and Cys89 get oxidized and form a disulfide bond to donate electrons to the arsenate to reduce it to arsenite and this step is aided by Arg16. In the case of LMW PTPases, while Arg13 helps in coordinating the phosphate group of the substrate, Thr11 and Asp118 orient the incoming phosphate substrate for the two nucleophilic attacks, the first one by Cys7, and the second one by an active site water molecule, where this second step helps in the release of the phosphate from the substrate, regenerating the enzymatic cleft for the next reaction. It is proposed that the corresponding structurally equivalent residues perform similar functions.

Sequence alignment performed using BLAST revealed that these two arsenate reductases (ArsC2 and ArsC3) show a percentage identity of only 38.24 %. When a multiple sequence alignment (MSA) of ArsC1, ArsC2, and ArsC3 was performed using ClustalW, it revealed that ArsC2 and ArsC3 show a significantly low percentage identity with ArsC1 (18.69 % and 19.09 % respectively).

Results of the MSA performed with the collective structural homologs show the sequence identity range is 21.17 %– 35.25 % for ArsC2, and 27.01 %– 38.17 % for ArsC3. For these two reductases as well, these results are suggestive of evolutionary divergence from known structural homologs. Construction of a phylogenetic tree of the sequences of ArsC2 and ArsC3 with their cumulative structural homologs shows that both arsenate reductases of *D. indicus* DR1 are close to those of *Bacillus subtilis* than those of *Staphylococcus aureus* (Fig. 5I).

The presence of three arsenate reductases, ArsC1, ArsC2, and ArsC3, in a single organism prompted the investigation of similarity between their sequences and structures. It was expected that all three would have similar sequences and structures. Surprisingly, it was found that while the reductases inside the cluster show considerable sequence and structural similarity, they do not share the same level of similarity with ArsC1, which is outside the *ars* gene cluster. This was inferred from the MSA and a phylogenetic tree constructed including the sequences of ArsC1, ArsC2, and ArsC3 of *D indicus* DR1 with their respective structural homologs. ArsC1 belongs to the Grx-linked prokaryotic ArsC reductases family of arsenate reductases [34], while ArsC2 and ArsC3 seem to be associated with the Trx-linked prokaryotic ArsC reductases.

3.5. Global properties of molecular dynamics simulation trajectories' analyses

The following analyses were conducted for MD trajectories simulated in GROMACS:

Root Mean Square Deviation (RMSD) of all atoms in each frame, residue-wise Root Mean Square Fluctuation (RMSF), and Radius of Gyration (Rg) of the protein in each frame. RMSD value is the root of the squared value of deviation of atomic coordinates between consecutive frames in a simulation when these frames are superimposed on each other. The RMSD is plotted on the Y-axis against simulation time on the X-axis. This graph tells us the difference between consecutive conformations during the trajectory. Lower the RMSD, higher the stability of the structure in the simulation during the course of the simulation. Backbone RMSD has been calculated for all proteins here. RMSF is the fluctuation of C-α atoms of each residue of the protein with respect to a mean position derived from the trajectory. Residues with higher flexibility tend to show higher values, while rigid residues exhibit lower values. These highly flexible residues are usually part of the loop regions. The Radius of Gyration plot helps us understand the compactness of the polypeptide structure in the simulation. The Rg value is calculated by finding the distribution of the atoms around a central axis of the protein. When this distribution changes drastically (i. e., the value goes up), it indicates rapid conformational change in the protein structure, which, in turn, signifies a reduction in the compactness of the structure, and even unfolding or fold change of the protein. These three plots give a good measure of the performance of the protein model in a simulation, thereby giving insights about its residue placement, steric clashes, and conformational stability. All of the plots were constructed using Xmgrace, the 2-D numerical value plotting tool. The results are discussed protein-wise.

3.5.1. ArsR1

The RMSD plot of ArsR1 plateaus very early into the simulation, at about 5 ns. The RMSD values every 10 ps is never more than 0.7 nm. The RMSF values range between 0.1 nm at its least and 0.65 nm at its most. The high values correspond to loop regions, which is not unexpected. The helical regions show a consistent RMSF of about 0.15–0.2 nm, which is not a high fluctuation value. The Rg plot shows plateauing less than 5 ns into the simulation, after which the difference between the radii of gyration does not exceed 1.5 nm. All three are signs of good energy minimization, therefore deeming 100 ns simulation time sufficient (Supplementary Fig. S9).

3.5.2. ArsR2

The RMSD values for ArsR2 start at 0.6 nm, dip to 0.2 nm, and then stabilize at 0.3 nm after 50 ns. But since the overall difference between the maxima and minima of RMSD is just 0.4 nm, this range is not of concerning significance. The loop regions show relatively high residue fluctuation in this protein as well. But some of the residues in the helical regions flanking the loop regions also exhibit some fluctuation. However, this fluctuation does not go above 0.5 nm, hence, it cannot be a reason for assuming a conformation change. The Rg plot for ArsR2 begin at 1.6 nm, then averages at about 1.45 nm after 35 ns of simulation. These values made us confident about the adequacy of 100 ns simulation time (Supplementary Fig. S9).

3.5.3. MPase

RMSD values were consistent at 0.3 nm. The RMSF values were, expectedly, at the loop regions, and one helix (α_6), which is

punctuated by two highly fluctuating loop regions. Yet, the RMSF values are maintained below 0.2 nm otherwise. The Rg values of this protein are preserved in the range 1.75–1.8 nm, from the beginning to the end of the simulation. These plots provide enough assurance for 100 ns simulation time being enough to study the properties of this protein (Supplementary Fig. S10).

3.5.4. ArsC2

The RMSD curve of ArsC2's MD trajectory initially shows a fluctuation range of 0.2 nm for the first 40 ns of the simulation, after which the curve stabilizes at 0.3 nm. RMSF plot for this protein is maintained low at less than 0.2 nm except at two loop regions and at the termini, which is not unexpected. In the Rg plot, the values fluctuate around a mean of about 1.525 nm, and become nearly constant in the second half of the simulation. These results show that 100 ns of simulation time was enough for ArsC2 for proper energy minimization, and for the protein to converge to a stabilized structure (Supplementary Fig. S11).

3.5.5. ArsC3

The RMSD graph is almost constantly maintained at about 0.275 nm throughout all frames, except at 50 ns, where it is 0.4 nm, indicating a slight conformational change, but not significant enough to destabilize the protein structure. In ArsC3, just like ArsC2, the RMSF values for the polypeptide residues is consistent at about 0.2 nm except the termini. The Rg plot shows consistency at close to 1.47 nm throughout the trajectory. There is a point of increased Rg value at 50 ns, but the trajectory falls back to the average immediately, suggestive of a change in conformation, but one that quickly switches back to the stable conformation. Therefore, 100 ns was sufficient for ArsC3 to remove steric clashes, and stabilize its behaviour in a simulation (Supplementary Fig. S11).

3.6. Arsenical efflux pump

arsB (NCBI Accession ID: OWL98586.1) is the fourth gene in the *ars* gene cluster, encoding an arsenical efflux pump membrane protein , referred to as ArsB in this manuscript. Robetta server, using Comparative Modelling, predicted this protein to be a homodimer. However, RoseTTAFold predicted it as a monomer. While the quality of the model from RoseTTAFold is higher (0.83) than the model obtained from comparative modeling (0.55), the templates used for modeling in the latter have metal transporting function and the biological assembly is a homodimer. Additionally, in *Vibrio cholerae* ArsB's homologs are seen to function as a multimeric transmembrane protein (PDB ids: 5ULD and 4F35) [73,74].

The median model structure (Model 3), validated using a Ramachandran plot, reveals the presence of five residues in the disallowed region of the plot. These five residues out of 870 residues in total, Ile30 in Chain A, Ser163 in both chains, and Ile384 in both chains. Since this number is only 0.7 % (less than 1 % residues, the allowed threshold for number of residues that can be present in the disallowed region), the model was deemed a confident one to take up for further work, (Supplementary Fig. S12a and S12b).

TMHMM server [43] analyzed the amino acid sequence of the homodimeric ArsB, and predicted the transmembrane residues of the 12 α-helices of each monomer of this protein. Fig. 6A and B show the front and top views of the protein, respectively, with its transmembrane helices labelled according to the predictions made by TMHMM server. These residues are indicated in Table 1 along with the transmembrane residues of Arsenical efflux pump ArsB of *E. coli* for comparison. These residues were used as a reference to model and place ArsB inside a lipid membrane using the CHARMM-GUI Membrane Builder tool (Fig. 6C and D) [48].

The focus is mainly on the conformational change of the protein in a lipid bilayer environment and not on the impact of the protein directly on the membrane curvature nor the effect of ionic strength. If there are any conformational changes observed it would happen at an early time in the MD simulation, hence 20 ns was deemed to be sufficient to obtain the coordinates of ArsB. Additionally, the RMSD, RMSF, and Rg plots of MD simulation of ArsB modelled plateau very early in the simulation (at about 5 ns), reinforcing the sufficiency of the simulation time (Supplementary Fig. S11).

A phylogenetic tree was constructed of arsenical membrane pumps from all species of the genus *Deinococcus*, with ArsB of *E. coli* as an outgroup, since that is the only known and the closest structural homolog (Fig. 6E and F). Analysis of the tree showed that ArsB is highly similar to the arsenical pump membrane pump of *D. radiophilus* (80.91 %), which is a radiation-resistant bacterium. Structural superposition of ArsB was performed with ArsB of *E. coli*–the monomeric ArsB of *E. coli* exhibited high degree of closeness with either monomer of the dimeric ArsB of *D. indicus* DR1 with a C-α atoms RMSD of 4.90 Å (Fig. 6G).

The Molecular Dynamics simulation trajectory of 20 ns was used to analyse the curves for RMSD, RMSF, and Rg. The RMSD plot shows the emergence of a plateau by around 2 ns, after which this plateau is maintained well. The RMS fluctuation of the dimer residues shows raised values for specific residues only, and it was found that these residues belong to loop regions. Otherwise, the RMS fluctuation is consistent at an average value of 3.44 Å. In the graph of Rg against time, the values are persistent at about 30.15 Å after around 2.5 ns. This continuance of values in all three plots made us confident of the adequacy of the 20 ns simulation time.

Although literature has always recorded ArsB as a transmembrane α-helical protein, it has remained monomeric – there are no records of an arsenical efflux pump that is dimeric or multimeric. On the contrary, there are previous publications and evidences that point towards other heavy metal anion transporters and drug efflux pumps, such as cadmium efflux systems, and zinc removal operons, that are dimeric and/or multimeric which are also shown to exhibit high level of resistance to the chemical species they remove [75–77]. Establishing the homology between these proteins could reveal the rationale behind the evolution of a dimeric protein and also help in deducing its mechanism of action, which is yet to be elaborated.

3.7. Expression of arsenic resistance genes in D. indicus DR1

All the genes showed significant increase in expression both in presence of 2.5 mM arsenate and 0.25 mM arsenite, suggesting that both arsenate and arsenite can induce gene expression. A two-fold increase in expression of all the three arsenate reductases were observed in presence of arsenate (Fig. 7). *arsB* gene was also found to be inducible by both arsenate and arsenite. Both *arsR1* and *arsR2* also displayed significant increase in expression in presence of arsenate. A fourfold increase in gene expression was observed for hypothetical protein, indicating an important role of this gene in arsenic tolerance in *D. indicus* DR1.

4. Discussion

The shared structural similarity between arsenate and phosphate at neutral pH leads to the arsenate uptake via phosphate uptake system. Though there is a specific pump for arsenite efflux, uptake of arsenite is mediated via aquaglyceroporin due to the similarity between arsenite and glycerol at physiological pH. While the cells cannot differentiate between arsenate and phosphate, the preliminary response to avoid arsenic toxicity is the reduced uptake. This is mediated via membrane bound respiratory arsenate reductase (converts arsenate to arsenite) and respiratory arsenite oxidase (converts arsenite to arsenate). Once arsenic enters the cells, further arsenic toxicity is relieved by *ars* operon which transforms arsenate to arsenite in cytosol (with the aid of cytosolic arsenate

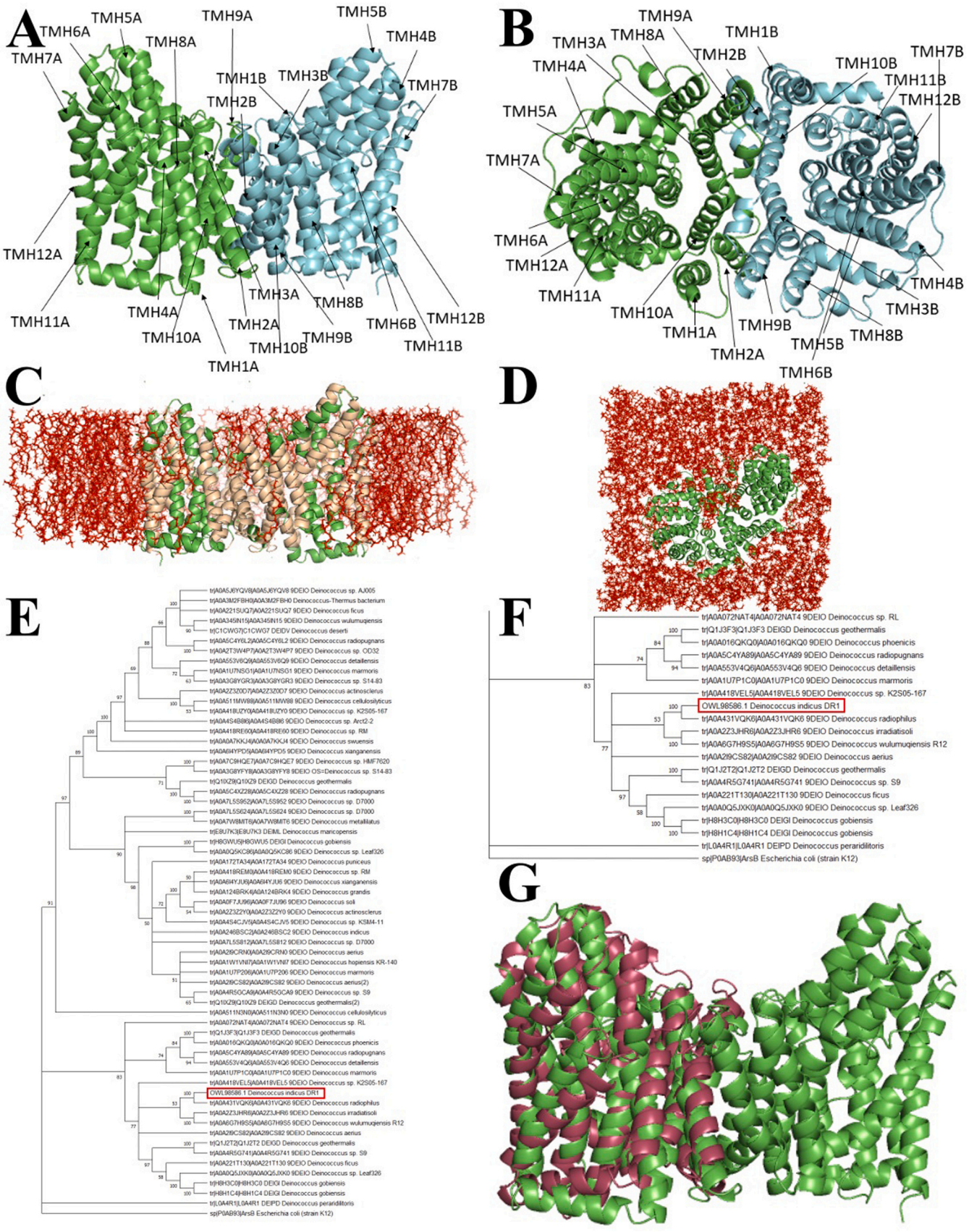

Fig. 6. : Arsenical Membrane Efflux Pump ArsB: ArsB, predicted to be a homodimeric α-helical protein, with its 12 helical chains labelled, and seen from the front (A) and from the top (B). ArsB was modelled inside a lipid membrane, and is visualized from the front, with lipids in front of the protein sliced away in image (C), and from the top in image (D). (E) Phylogenetic tree constructed with amino acid sequences of arsenical efflux pumps from all species in the genus *Deinococcus*. (F) A close-up of the section of the phylogenetic tree highlighting ArsB of *D. indicus* DR1 using a red box. (G) Superposition of ArsB of *D. indicus* DR1 (green) to ArsB of *E. coli* (purple).

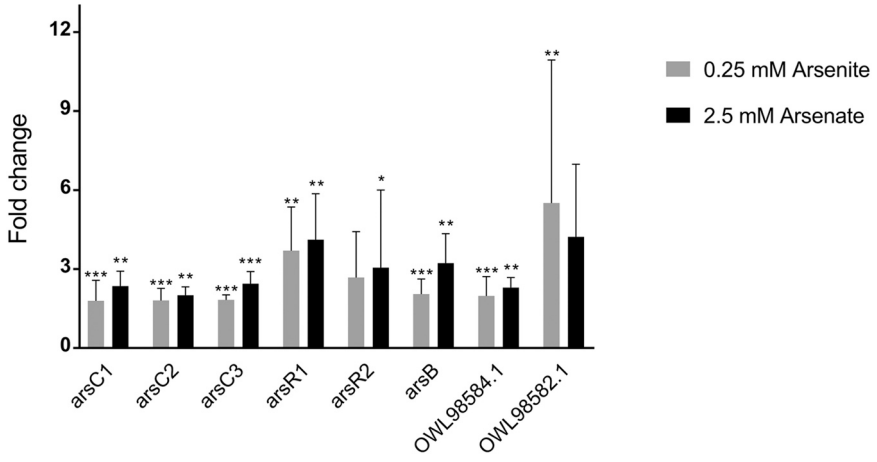

Fig. 7. : Quantitative real-time PCR analysis of *D. indicus* DR1 cells induced with arsenate and arsenite. Bars show mean fold change (+/− standard deviation) over untreated. Asterisks indicate the statistical significance level: P-value < 0.05 (*), < 0.01 (**), and < 0.001 (***), compared with untreated (arbitrarily set as 1; not shown in graph). Acronyms for genes: *arsC1*, arsenate reductase (OWL94580.1); *arsC2*, arsenate reductase (OWL98585.1); *arsC3*, arsenate reductase (OWL98587.1); *arsR1*, ArsR family transcriptional regulator (OWL98583.1); *arsR2*, transcriptional regulator (OWL98588.1); *arsB*, arsenical efflux pump membrane protein (OWL98586.1); metallophosphatase family protein (OWL98584.1); and hypothetical protein CBQ26_01570 (OWL98582.1).

reductase) and exports the more toxic arsenite from the cell. The ubiquity of *ars* operon across different species signifies the need for arsenic detoxification.

Three families of cytosolic arsenate reductase enzyme have been identified so far. The first one known as Grx-linked arsenate reductase utilizes glutaredoxin and reduced glutathione as reducing equivalent for enzymatic activity. It is generally present in gram-negative bacteria, such as *Escherichia coli, Haemophilus influenza*, *Yersinia enterocolitica*, etc [78–80]. The eukaryotic arsenate reductase identified in *Saccharomyces cerevisiae* also require glutaredoxin and reduced glutathione as reductants [78]. The Trx-linked arsenate reductase family is mostly found in gram positive bacteria. The well-studied examples of this family include ArsCs of *Staphylococcus aureus*, *Bacillus subtilis*, etc. [81–84]. They use thioredoxin as the reducing equivalent. This family is related to the low molecular weight protein-tyrosine phosphatases (LMW PTPases) family protein [70]. Arsenate reductase of *B. subtilis, S. aureus* plasmid pI258 and *T. thermophilus* HB27 can also display weak phosphatase activity [66,70,85]. Recently, it was shown in *T. thermophilus* HB27 that both arsenate and arsenite can inhibit the phosphatase activity through non-competitive mode of inhibition [85]. Our results indicate that both the arsenate reductases of *ars* gene cluster i.e., ArsC2 and ArsC3 of *D. indicus* belong to the LMW PTPases family protein and probably utilize thioredoxin.

All the arsenate reductases, irrespective of families or organism, perform the same function i.e., to convert arsenate, As(V) to arsenite, As(III). Arsenate binding with the enzyme is inhibited by phosphate as it is an analog of arsenate and inhibits the reduction of arsenate to arsenite. It has been shown that phosphate competes with arsenate for the active site in ArsC [22,86]. The translated gene products of *arsC2* and *arsC3* share 39 % identity with each other, query coverage being 91 %. It should be noted that the conserved active site signature sequence of phosphate binding loop (P-loop) in LMW PTPases (Cys-$X_{(5)}$-Arg-Ser) (where, X = any amino acid) [87], is present in both ArsC2 (Cys-Thr-Gly-Asn-Thr-Ala-Arg-Ser) and ArsC3 (Cys-Thr-His-Asn-Ser-Ala-Arg-Ser), indicating towards their phosphohydrolase potential. The role of cysteine in active site is to form a covalent phosphoenzyme intermediate, which is further stabilized by arginine [88]. Another role of this conserved cysteine with respect to arsenate reductase feature is formation of thioester bond with arsenate [79]. In *S. aureus* and *B. subtilis* arsenate reductase, the three cysteine residues form a redox relay network [70,82]. The structurally equivalent cysteine residues of *S. aureus* pI258 arsenate reductase (Cys10, Cys82, and Cys89) [82] are Cys15, Cys87 and Cys94 in ArsC2; and Cys8, Cys84 and Cys91 in ArsC3.

In addition, it has been proposed that an aspartate residue in *B. subtilis* (Asp-105) positioned close to the active site is conserved in Gram-positive bacteria and LMW PTPases. It plays a key role in the acid/base catalysis and stabilizes the enzyme-substrate complex by acting as a general acid [70]. An aspartate residue close to the pair of cysteine residues is also present in ArsC2 (Asp-110) and ArsC3 (Asp-116), where it may play the same role. It forms a part of the conserved D-P motif, crucial for phosphate binding. Based on the conserved active site residues in LMW PTPases, and arsenate reductase of other bacteria belonging to the LMW PTPases family, there is a possibility that ArsC2 and ArsC3 of *D. indicus* can also bind phosphate. However, it further needs to be assessed whether ArsC2 and ArsC3 can also display phosphatase activity in vitro.

The ArsR proteins belong to ArsR/SmtB family of transcriptional regulators, which can bind to heavy metals such as cadmium, arsenic, lead, bismuth, zinc, cobalt, nickel, etc. [89]. The repressor proteins have a helix-turn-helix DNA binding domain, metal binding site and a dimerization domain, which is essential for the activity. Binding of metal ions to the repressor leads to its dissociation from the promoter/operator region of DNA due to helix-unwinding. In *E. coli* ArsR, a single domain of nearly 80 amino acids is required for DNA binding, dimerization and recognition of arsenite [90]. ArsR functions mostly as a homodimer. Recently, AioF of *Thiomonas arsenioxidans* belonging to the ArsR/SmtB family was shown to have the ability to form dimer and tetramer [91]. The cadmium and lead sensing CmtR in *M. tuberculosis* can even form octamer under non-reducing conditions and interact with multiple sites in the promoter region [92]. Arsenite binding to the repressor has no effect on the dimerization state. Metal binding produces a conformational change that leads to the release of repressor from promoter/operator region. It should be noted that the binding sites in ArsR/SmtB family proteins can be present together in same monomer such as in AfArsR protein from *Acidithiobacillus ferrooxidans* or distributed between two different subunits, as seen in CgArsR repressor from *Corynebacterium glutamicum* [53], and CmtR repressor from *Mycobacterium tuberculosis* [92]. *D. indicus* DR1 has two arsenic repressors, ArsR1 and ArsR2 that share only 37% sequence identity. However, they both are phylogenetically closer to *S. aureus* cadmium repressor CadC protein and *C. glutamicum* ArsR. The metal binding

sites, dimerization, DNA binding region and residues crucial for repression have been well characterized in CadC [55].

Both arsenic and cadmium bind to the repressor via cysteine residues. The location of these cysteine resides can vary among different organisms. In AfArsR of *Acidithiobacillus ferrooxidans*, the three cysteines are located at 95, 96 and 102 position [93]. In *E. coli* plasmid R773, Cys32, Cys34, and Cys37 of the CXCXXC domain (where, X = any amino acid) interact with arsenite, AsIII+ [17]. However, binding to only Cys32 and Cys34 facilitates de-repression of the operon [94]. The conserved ELCV(C/G)D domain in ArsR repressor located in the HTH region binds with metal [95]. However in *D. indicus* DR1, this metal binding domain is present in ArsR2 (EHCVCDL), but not in ArsR1. ArsR1 (OWL98583.1) has four cysteine residues at positions 32, 33, 40, and 42, which are a part of the putative helix-turn-helix (HTH) region. In ArsR2 (OWL98588.1), the residues are located at 34, 36, and 106 positions. Cys34 and Cys36 are part of the putative DNA binding domain. Note that the Cys106 is also the last amino acid in the ArsR2 protein, located at the C-terminus. Since vicinal cysteine residues are required for binding arsenite, there is a possibility that to complete the triad, the remaining cysteine is contributed by another subunit of ArsR2 homodimer or tetramer.

There are several bacteria which have more than one copy of *ars* operon. The arrangement of genes and copies, operon regulation varies hugely among different organisms. One such example is *Pseudomonas putida* strain KT2440 [10]. It has two *ars* operons located on the chromosome, likely to be the result of a duplication event. The percent identity in amino acid sequence between homologous proteins within the four genes operon (arsRBCH) arrangement ranges from 73% to 87%. One interesting feature of this organism is the crosstalk between ArsR1 and ArsR2 repressors, which can bind to their promoters as well as the non-cognate promoter. This bifan network model is another example of efficient arsenic detoxification system in bacteria [10]. Further investigation is required to reveal DNA binding properties of ArsR1 and ArsR2 and their interactions in *D. indicus* DR1.

5. Conclusions

In this paper, we present computationally validated structures of an entire arsenic resistance gene cluster and correlated them with their functions. To our knowledge, the *ars* gene cluster's structural map for *Deinococcus indicus* DR1 is unknown, due to which the metabolic pathway for arsenic tolerance is unclear. Specifically, for *D. indicus* DR1 the structural mapping using RoseTTAFold, a deep learning-based approach for structure prediction, along with comparative modeling sped the process of filling the knowledge-gap for this organism.

Expression analysis via RT-qPCR showed that all the genes are highly expressed in the presence of arsenic, indicating that the arsenic tolerance genes are working in synergy to metabolize arsenic and enable the organism to thrive in toxic environments. Specifically, ArsC is an oxidoreductase that reduces pentavalent arsenate to trivalent arsenite oxyanions which are then extruded from the cell by the ArsB export system. Traditionally, arsenate reductases contain a cysteine residue in their active site that carries out this reduction. The lack of this residue in ArsC1 of *D. indicus* DR1, which is thought to be imperative for enzyme interaction with the arsenic anion, could be indicative of an alternate mechanism of enzyme action. Previously studied arsenate reductases are known to gain their reducing power from one of two cellular oxidoreductases - glutaredoxin, usually found in gram negative bacteria, or thioredoxin in gram positive bacteria. In a recent study, nine isofunctional groups were identified based on the active site signature, in order to understand the proteins comprising the ArsC superfamily [96]. Homology based analysis of the arsenate reductases from *D. indicus* DR1 show structural similarity with that of *E. coli* and *S. aureus* indicating that its reducing power likely derived from the thioredoxin (Trx)-fold pathway. A study published in 2007 showed the crystallized structure of ArsC with thioredoxin in *Bacillus subtilis*, and found that other than the three main catalytic cysteine residues, Cys29 also had a role in catalysis by forming disulphide bonds with both Cys32 and the active site Cys89. Such additional disulphide linkages seem to help in better reaction progression [83]. Another study that describes the presence of additional buried disulphide linkages between non-active site cysteines and active site cysteines and this interaction seems to help in selectivity of thioredoxin to resolve disulphide linkages in the Trx-ArsC system, and also to bring the oxidized ArsC to its native reduced state [84].

It has been observed that a conserved Asn13 residue is important for the active site in Trx-linked ArsCs [97]. The structurally equivalent residue for Asn13 in ArsC2 and ArsC3 is observed at the 18th position and 11th position respectively. The behavior of this residue was observed in MD simulations and was seen to stabilize the active site in longer simulation times. This reinforces the observation that ArsC2 and ArsC3 could belong to the Trx-family of arsenate reductases.

In another study, the Trx-type arsenate reductase in *Corynebaterium glutamicum*, uses not three, but just one cysteine residue in its active site, but with a LMW PTPase fold, and a conserved Asn11 [98]. This enzyme could also serve as a link to understand arsenate reductases in bacteria better, and in turn, enable us to elucidate the mechanisms of action of ArsC2 and ArsC3 of *D. indicus* DR1. Additionally, the presence of multiple arsenate reductases within the genome of *D. indicus* DR1 could account for its high level of arsenic tolerance and may be a result of gene duplication or other evolutionary processes.

In conclusion, prevalence of arsenic in the natural environment has led to evolution of arsenic detoxification system in different organisms. The arsenic tolerance genes have been identified across several bacterial species and a few of them have been employed for biosensors and other industrial applications. *D. indicus* DR1 *ars* genes are highly diverse in their sequence to each other as well as the previously characterized proteins.

Further understanding of the gene cluster, corresponding protein structure and gene regulation is required to shed light on the survival mechanism and to address the questions like how some bacteria can withstand high arsenic stress, despite having the same set of homologous proteins. While the exact regulation between the *ars* genes is still unclear, having validated structural models can shine light onto the molecular mechanisms and future studies can be explored that take a systems-level approach to identify gene regulatory mechanisms in the *ars* operon. Additionally, constraint-based modeling such as Flux Balance Analysis or structural systems biology can be explored with the models studied in this work as a starting point.

CRediT authorship contribution statement

Shrivaishnavi Ranganathan: Writing – original draft, Methodology, Data curation, Visualization. **Deepa Sethi:** Writing – original draft, Data curation, Visualization. **Sandhya Kasivisweswaran:** Formal analysis, Investigation, Visualization, Writing – original draft. **L Ramya:** Validation, Formal analysis, Investigation, Supervision, Writing – review & editing, Methodology. **Richa Priyadarshini:** Conceptualization, Writing – review & editing, Supervision, Project administration, Funding acquisition, Methodology. **Ragothaman M Yennamalli:** Conceptualization, Writing – review & editing, Supervision, Project administration, Funding acquisition, Methodology.

Data Availability

The models of the proteins obtained after energy minimization were submitted to ModelArchive and these are available for public use. The hyperlinks for the models are as follows:

ArsR1: https://www.modelarchive.org/doi/10.5452/ma-m92cj;.
ArsR2: https://www.modelarchive.org/doi/10.5452/ma-dkxdu;.
MPase: https://www.modelarchive.org/doi/10.5452/ma-j0xxh;.
ArsC2: https://www.modelarchive.org/doi/10.5452/ma-28bjc;.
ArsC3: https://www.modelarchive.org/doi/10.5452/ma-xpwd6; and.
ArsB: https://www.modelarchive.org/doi/10.5452/ma-yrymn.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgements

Ragothaman M. Yennamalli is supported by the UGC-Basic Science Research Startup Grant, University Grants Commission, Government of India (F.30-561/2021(BSR) and National Agricultural Science Fund-Indian Council of Agricultural Research (NASF-ICAR), Government of India (F. No. NASF/SUTRA-02/2022-23/50). Richa Priyadarshini Lab is supported by CSIR-EMR grant. DS is supported by doctoral fellowship from Shiv Nadar University. The computational infrastructure and support provided by the Bioinformatics Research and Applications Facility (BRAF) funded by the National Supercomputing Mission, Government of India at the Centre for Development of Advanced Computing, Pune are gratefully acknowledged. SR, RL, and RMY thank SASTRA Deemed to be University for infrastructural support and research facilities. RP, DS and SK acknowledge Shiv Nadar University for infrastructural support and research facilities.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2022.12.015.

References

[1] Borah P, Kumar M, Devi P. Types of inorganic pollutants: metals/metalloids, acids, and organic forms. Inorg Pollut Water 2020:17–31. https://doi.org/10.1016/B978-0-12-818965-8.00002-0
[2] Nielsen F.H. Nutritional requirements for boron, silicon, vanadium, nickel, and arsenic: current knowledge and speculation; Nutritional requirements for boron, silicon, vanadium, nickel, and arsenic: current knowledge and speculation n.d. https://doi.org/10.1096/fasebj.5.12.1916090.
[3] Kalia K, Joshi DN. Detoxification of Arsenic. Handb Toxicol Chem Warfare Agents 2009:1083–100. https://doi.org/10.1016/B978-012374484-5.00072-9
[4] Hu Y, Su L, Snow ET. Arsenic toxicity is enzyme specific and its affects on ligation are not caused by the direct inhibition of DNA repair enzymes. Mut Res/DNA Rep 1998;408:203–18. https://doi.org/10.1016/S0921-8777(98)00035-4
[5] Smith AH, Steinmaus CM. Health effects of arsenic and chromium in drinking water: recent human findings. Ann Rev Public Health 2009;30:107. https://doi.org/10.1146/annurev.publhealth.031308.100143
[6] States JC, Barchowsky A, Cartwright IL, Reichard JF, Futscher BW, Lantz RC. Arsenic toxicology: translating between experimental models and human pathology. Environ Health Perspect 2011;119:1356–63. https://doi.org/10.1289/EHP.1103441
[7] Argos M, Ahsan H, Graziano JH. Arsenic and human health: epidemiologic progress and public health implications. Rev Environ Health 2012;27:191–5. https://doi.org/10.1515/reveh-2012-0021
[8] Nies D. Microbial heavy-metal resistance. Appl Microbiol Biotechnol 1999;51:730–50https://doi.org/10.1007/s002530051457.
[9] Ordóñez E, Letek M, Valbuena N, Gil JA, Mateos LM. Analysis of genes involved in arsenic resistance in *Corynebacterium glutamicum* ATCC 13032. Appl Environ Microbiol 2005;71:6206–15. https://doi.org/10.1128/AEM.71.10.6206-6215.2005
[10] Fernández M, Morel B, Ramos JL, Krell T. Paralogous regulators ArsR1 and ArsR2 of *Pseudomonas putida* KT2440 as a basis for arsenic biosensor development. Appl Environ Microbiol 2016;82:4133–44. https://doi.org/10.1128/AEM.00606-16
[11] Zhao C, Zhang Y, Chan Z, Chen S, Yang S. Insights into arsenic multi-operons expression and resistance mechanisms in *Rhodopseudomonas palustris* CGA009. Front Microbiol 2015;6:986. https://doi.org/10.3389/fmicb.2015.00986
[12] Sato T, Kobayashi Y. The ars operon in the skin element of *Bacillus subtilis* confers resistance to arsenate and arsenite. J Bacteriol 1998;180:1655–61. https://doi.org/10.1128/JB.180.7.1655-1661.1998
[13] Muller D, Médigue C, Koechler S, Barbe V, Barakat M, Talla E, et al. A tale of two oxidation states: Bacterial colonization of arsenic-rich environments. PLoS Genet 2007;3:0518–30. https://doi.org/10.1371/journal.pgen.0030053
[14] Silver S, Budd K, Leahy KM, Shaw WV, Hammond D, Novick RP, et al. Inducible plasmid-determined resistance to arsenate, arsenite, and antimony(III) in *Escherichia coli* and *Staphylococcus aureus*. J Bacteriol 1981;146:983–96. https://doi.org/10.1128/JB.146.3.983-996.1981
[15] Wu J, Rosen BP. The arsD gene encodes a second trans-acting regulatory protein of the plasmid-encoded arsenical resistance operon. Mol Microbiol 1993;8:615–23. https://doi.org/10.1111/J.1365-2958.1993.TB01605.X
[16] Wu J, Rosen BP. Metalloregulated expression of the ars operon. J Biol Chem 1993;268:52–8. https://doi.org/10.1016/S0021-9258(18)54113-2
[17] Shen S, Li XF, Cullen WR, Weinfeld M, Le XC. Arsenic binding to proteins. Chem Rev 2013;113:7769–92. https://doi.org/10.1021/CR300015C
[18] Wu J, Rosen BP. The ArsR protein is a trans-acting regulatory protein. Mol Microbiol 1991;5:1331–6. https://doi.org/10.1111/J.1365-2958.1991.TB00779.X
[19] Rosen BP. Transport and detoxification systems for transition metals, heavy metals and metalloids in eukaryotic and prokaryotic microbes. Comp Biochem Physiol Part A: Mol Integr Physiol 2002;133:689–93. https://doi.org/10.1016/S1095-6433(02)00201-5
[20] Ji G, Silver S. Regulation and expression of the arsenic resistance operon from *Staphylococcus aureus* plasmid pI258. J Bacteriol 1992;174:3684–94. https://doi.org/10.1128/JB.174.11.3684-3694.1992
[21] Gladysheva TB, Oden KL, Rosen BP. Properties of the Arsenate Reductase of Plasmid R773. Biochemistry 1994;33:7288–93. https://doi.org/10.1021/bi00189a033
[22] Liu J, Rosen BP. Ligand interactions of the ArsC arsenate reductase. J Biol Chem 1997;272:21084–9. https://doi.org/10.1074/jbc.272.34.21084
[23] Lin YF, Walmsley AR, Rosen BP. An arsenic metallochaperone for an arsenic detoxification pump. Proc Natl Acad Sci USA 2006;103:15617–22. https://doi.org/10.1073/pnas.0603974103
[24] Chen J, Bhattacharjee H, Rosen BP. ArsH is an organoarsenical oxidase that confers resistance to trivalent forms of the herbicide monosodium methylarsenate and the poultry growth promoter roxarsone. Mol Microbiol 2015;96:1042–52. https://doi.org/10.1111/mmi.12988
[25] Qin J, Rosen BP, Zhang Y, Wang G, Franke S, Rensing C. Arsenic detoxification and evolution of trimethylarsine gas by a microbial arsenite S-adenosylmethionine methyltransferase. Proc Natl Acad Sci 2006;103(7):2075–80. https://doi.org/10.1073/pnas.0506836103. U S A.
[26] Qin J, Lehr CR, Yuan C, Le XC, McDermott TR, Rosen BP. Biotransformation of arsenic by a Yellowstone thermoacidophilic eukaryotic alga. Proc Natl Acad Sci 2009;106(13):5213–7. https://doi.org/10.1073/pnas.0900238106. U S A.
[27] Yoshinaga M, Rosen BP. A C·As lyase for degradation of environmental organoarsenical herbicides and animal husbandry growth promoters. Proc Natl Acad Sci 2014;111:7701–6. https://doi.org/10.1073/pnas.1403057111
[28] Gerber E, Bernard R, Castang S, Chabot N, Coze F, Dreux-Zigha A, et al. *Deinococcus* as new chassis for industrial biotechnology: biology, physiology and tools. J Appl Microbiol 2015;119:1–10. https://doi.org/10.1111/JAM.12808
[29] Kang Myung-Suk, Srinivasan Sathiyaraj. A report of 9 unrecorded radiation resistant bacterial species in Korea. Journal of Species Research 2017;6(2):91–100https://doi.org/10.12651/JSR.2017.6.2.091.
[30] Ferreira AC, Nobre MF, Rainey FA, Silva MT, Wait R, Burghardt J, et al. *Deinococcus geothermalis* sp. nov. and *Deinococcus murrayi* sp. nov., two extremely radiation-resistant and slightly thermophilic species from hot springs. Int J Syst Bacteriol 1997;47:939–47. https://doi.org/10.1099/00207713-47-4-939
[31] Gutman PD, Fuchs P, Minton KW. Restoration of the DNA damage resistance of *Deinococcus radiodurans* DNA polymerase mutants by *Escherichia coli* DNA polymerase I and Klenow fragment. Mutation Res/DNA Repair 1994;314:87–97. https://doi.org/10.1016/0921-8777(94)90064-7
[32] de Groot A, Chapon V, Servant P, Christen R, Fischer-Le Saux M, Sommer S, et al. *Deinococcus deserti* sp. nov., a gamma-radiation-tolerant bacterium isolated from the Sahara Desert. Int J Syst Evolut Microbiol 2005;55:2441–6. https://doi.org/10.1099/ijs.0.63717-0
[33] Chauhan D, Srivastava PA, Yennamalli RM, Priyadarshini R. Draft genome sequence of *Deinococcus indicus* DR1, a novel strain isolated from a freshwater wetland. e00754-17 Genome Announcements 2017;5. https://doi.org/10.1128/GENOMEA.00754-17
[34] Chauhan D, Srivastava PA, Agnihotri V, Yennamalli RM, Priyadarshini R. Structure and function prediction of arsenate reductase from *Deinococcus indicus* DR1. J Mol Model 2019;25:1–9. https://doi.org/10.1007/S00894-018-3885-3/FIGURES/8
[35] Rio DC, Ares Jr. M, Hannon GJ, Nilsen TW. Purification of RNA using TRIzol (TRI reagent). Cold Spring Harb Protoc 2010;2010(6):db.prot5439https://doi.org/10.1101/pdb.prot5439
[36] Tam A, Shemesh M, Wormser U, Sintov A, Steinberg D. Effect of different iodine formulations on the expression and activity of *Streptococcus mutans* glucosyltransferase and fructosyltransferase in biofilm and planktonic environments. J Antimicrob Chemother 2006;57:865–71. https://doi.org/10.1093/JAC/DKL085

[37] Schmittgen T, Livak K. Analyzing real-time PCR data by the comparative C_T method. Nat Protoc 2008;3:1101–8https://doi.org/10.1038/nprot.2008.73.
[38] Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 2018;35:1547. https://doi.org/10.1093/MOLBEV/MSY096
[39] Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7. https://doi.org/10.1093/NAR/GKH340
[40] Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. Science 2021;373:871–6. https://doi.org/10.1126/science.abj8754
[41] Theobald DL, Wuttke DS. THESEUS: maximum likelihood superpositioning and analysis of macromolecular structures. Bioinformatics 2006;22:2171–2. https://doi.org/10.1093/BIOINFORMATICS/BTL332
[42] R.A. Laskowski M.W. MacArthur J.M. Thornton PROCHECK: validation of protein-structure coordinates C.P. Brock T. Hahn H. Wondratschek U. Müller U. Shmueli E. Arnold D.M. Himmel M.G. Rossmann In International Tables for Crystallography 2012 doi: 10.1107/97809553602060000882.
[43] Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305(3):567–80. https://doi.org/10.1006/jmbi.2000.4315
[44] Hess B., Kutzner C., Van Der Spoel D., Lindahl E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation 2008. https://doi.org/10.1021/ct700301q.
[45] Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. J Comput Chem 2005;26:1781–802. https://doi.org/10.1002/jcc.20289
[46] Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: a web-based graphical user interface for CHARMM. J Comput Chem 2008;29:1859–65. https://doi.org/10.1002/JCC.20945
[47] Lee J, Cheng X, Swails JM, Yeom MS, Eastman PK, Lemkul JA, et al. CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. J Chem Theory Comput 2016;12:405–13. https://doi.org/10.1021/ACS.JCTC.5B00935/
[48] Wu EL, Cheng X, Jo S, Rui H, Song KC, Dávila-Contreras EM, et al. CHARMM-GUI membrane builder toward realistic biological membrane simulations. J Comput Chem 2014;35:1997–2004. https://doi.org/10.1002/JCC.23702
[49] Binkowski TA, Naghibzadeh S, Liang J. CASTp: computed atlas of surface topography of proteins. Nucleic Acids Res 2003;31:3352–5. https://doi.org/10.1093/NAR/GKG512
[50] Holm L, Rosenström P. Dali server: conservation mapping in 3D. Nucleic Acids Res 2010;38:W545–9. https://doi.org/10.1093/NAR/GKQ366
[51] Theobald DL, Steindel PA. Optimal simultaneous superpositioning of multiple structures with missing data. Bioinformatics 2012;28:1972–9. https://doi.org/10.1093/BIOINFORMATICS/BTS243
[52] Laskowski RA, MacArthur MW, Moss DS, Thornton JM. *PROCHECK*: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. 1993;26:283–91https://doi.org/10.1107/S0021889892009944.
[53] Prabaharan C, Kandavelu P, Packianathan C, Rosen BP, Thiyagarajan S. Structures of two ArsR As(III)-responsive transcriptional repressors: implications for the mechanism of derepression. J Struct Biol 2019;207:209–17. https://doi.org/10.1016/J.JSB.2019.05.009
[54] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403–10. https://doi.org/10.1016/S0022-2836(05)80360-2
[55] Ye J, Kandegedara A, Martin P, Rosen BP. Crystal structure of the *Staphylococcus aureus* pI258 CadC Cd(II)/Pb(II)/Zn(II)-responsive repressor. J Bacteriol 2005;187:4214. https://doi.org/10.1128/JB.187.12.4214-4221.2005
[56] Kandegedara A, Thiyagarajan S, Kondapalli KC, Stemmler TL, Rosen BP. Role of bound Zn(II) in the CadC Cd(II)/Pb(II)/Zn(II)-responsive repressor. J Biol Chem 2009;284:14958–65. https://doi.org/10.1074/jbc.M809179200
[57] Eicken C, Pennella MA, Chen X, Koshlap KM, VanZile ML, Sacchettini JC, et al. A metal-ligand-mediated intersubunit allosteric switch in related SmtB/ArsR zinc sensor proteins. J Mol Biol 2003;333:683–95. https://doi.org/10.1016/j.jmb.2003.09.007
[58] Campanello GC, Ma Z, Grossoehme NE, Guerra AJ, Ward BP, Dimarchi RD, et al. Allosteric inhibition of a zinc-sensing transcriptional repressor: insights into the arsenic repressor (ArsR) family. J Mol Biol 2013;425:1143–57. https://doi.org/10.1016/j.jmb.2013.01.018
[59] Capdevila DA, Edmonds KA, Campanello GC, Wu H, Gonzalez-Gutierrez G, Giedroc DP. Functional role of solvent entropy and conformational entropy of metal binding in a dynamically driven allosteric system. J Am Chem Soc 2018;140:9108–19. https://doi.org/10.1021/jacs.8b02129
[60] Cook WJ, Kar SR, Taylor KB, Hall LM. Crystal structure of the cyanobacterial metallothionein repressor SmtB: a model for metalloregulatory proteins. J Mol Biol 1998;275:337–46. https://doi.org/10.1006/jmbi.1997.1443
[61] Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. Nat Methods 2022;19:679–82. https://doi.org/10.1038/s41592-022-01488-1
[62] Matange N., Podobnik M., Visweswariah S. Metallophosphoesterases: structural fidelity with functional promiscuity. Biochem J 2015; 467 (2): 201–216. doi: https://doi.org/10.1042/BJ20150028.
[63] Swingle MR, Honkanen RE, Ciszak EM. Structural basis for the catalytic activity of human serine/threonine protein phosphatase-5. J Biol Chem 2004;279:33992–9. https://doi.org/10.1074/jbc.M402855200
[64] D'Arcy BM, Swingle MR, Papke CM, Abney KA, Bouska ES, Prakash A, et al. The antitumor drug LB-100 is a catalytic inhibitor of protein phosphatase 2A (PPP2CA) and 5 (PPP5C) coordinating with the active-site catalytic metals in PPP5C. Mol Cancer Therapeutics 2019;18:556–66. https://doi.org/10.1158/1535-7163.MCT-17-1143
[65] Chattopadhyay Debasish, Swingle Mark R, Salter EA, Wood E, D'Arcy Brandon, Zivanov Catherine, et al. Crystal structures and mutagenesis of PPP-family ser/thr protein phosphatases elucidate the selectivity of cantharidin and novel norcantharidin-based inhibitors of PP5C. Biochem Pharmacol 2016:109. https://doi.org/10.1016/j.bcp.2016.03.011
[66] Zegers I, Martins JC, Willem R, Wyns L, Messens J. Arsenate reductase from *S. aureus* plasmid pI258 is a phosphatase drafted for redox duty. Nat Struct Biol 2001;8(10):843–7. https://doi.org/10.1038/nsb1001-843
[67] Messens J, Martins JC, Van Belle K, Brosens E, Desmyter A, De Gieter M, et al. All intermediates of the arsenate reductase mechanism, including an intramolecular dynamic disulfide cascade. Proc Natl Acad Sci USA 2002;99:8506–11. https://doi.org/10.1073/pnas.132142799
[68] Roos G, Buts L, Van Belle K, Brosens E, Geerlings P, Loris R, et al. Interplay between ion binding and catalysis in the thioredoxin-coupled arsenate reductase family. J Mol Biol 2006;360:826–38. https://doi.org/10.1016/j.jmb.2006.05.054
[69] Messens J, Van Molle I, Vanhaesebrouck P, Van Belle K, Wahni K, Martins JC, et al. The structure of a triple mutant of pI258 arsenate reductase from *Staphylococcus aureus* and its 5-thio-2-nitrobenzoic acid adduct. Acta Crystallogr D Biol Crystallogr 2004;60:1180–4. https://doi.org/10.1107/S0907444904007334
[70] Bennett MS, Guan Z, Laurberg M, Su XD. *Bacillus subtilis* arsenate reductase is structurally and functionally similar to low molecular weight protein tyrosine phosphatases. Proc Natl Acad Sci USA 2001;98:13577–82. https://doi.org/10.1073/PNAS.241397198/
[71] Guo X, Li Y, Peng K, Hu Y, Li C, Xia B, et al. Solution structures and backbone dynamics of arsenate reductase from *Bacillus subtilis*: reversible conformational switch associated with arsenate reduction. J Biol Chem 2005;280:39601–8. https://doi.org/10.1074/jbc.M508132200
[72] Fuhrmann J, Mierzwa B, Trentini DB, Spiess S, Lehner A, Charpentier E, et al. Structural basis for recognizing phosphoarginine and evolving residue-specific protein phosphatases in gram-positive bacteria. Cell Rep 2013;3:1832–9. https://doi.org/10.1016/j.celrep.2013.05.023
[73] Nie R, Stark S, Symersky J, Kaplan RS, Lu M. Structure and function of the divalent anion/Na+ symporter from *Vibrio cholerae* and a humanized variant. Nat Commun 2017;8:15009. https://doi.org/10.1038/ncomms15009
[74] Mancusso R, Gregorio GG, Liu Q, Wang D-N. Structure and mechanism of a bacterial sodium-dependent dicarboxylate transporter. Nature 2012;491:622–6. https://doi.org/10.1038/nature11542
[75] Santiago AG, Chen TY, Genova LA, Jung W, Thompson AMG, Mcevoy MM, et al. Adaptor protein mediates dynamic pump assembly for bacterial metal efflux. Proc Natl Acad Sci USA 2017;114:6694–9. https://doi.org/10.1073/PNAS.1704729114
[76] Delmar JA, Su CC, Yu EW. Heavy metal transport by the CusCFBA efflux system. Protein Sci 2015;24:1720–36. https://doi.org/10.1002/PRO.2764
[77] Nikaido H, Takatsuka Y. Mechanisms of RND Multidrug Efflux Pumps. Biochim et Biophys Acta 2009;1794:769. https://doi.org/10.1016/J.BBAPAP.2008.10.004
[78] Mukhopadhyay R, Rosen BP. Arsenate reductases in prokaryotes and eukaryotes. Environ Health Perspect 2002;110:745–8. https://doi.org/10.1289/ehp.02110s5745
[79] Mukhopadhyay R, Rosen BP, Phung LT, Silver S. Microbial arsenic: from geocycles to genes and enzymes. FEMS Microbiol Rev 2002;26:311–25. https://doi.org/10.1111/j.1574-6976.2002.tb00617.x
[80] Stevens SY, Hu W, Gladysheva T, Rosen BP, Zuiderweg ERP, Lee L. Secondary Structure and Fold Homology of the ArsC Protein from the *Escherichia coli* Arsenic Resistance Plasmid R773. Biochemistry 1999;38:10178–86. https://doi.org/10.1021/bi990333c
[81] Ji G, Silver S. Reduction of arsenate to arsenite by the ArsC protein of the arsenic resistance operon of *Staphylococcus aureus* plasmid pI258. Proc Natl Acad Sci USA 1992;89:9474–8. https://doi.org/10.1073/pnas.89.20.9474
[82] Ji G, Garber EAE, Armes LG, Chen C-M, Fuchs JA, Silver S. Arsenate Reductase of *Staphylococcus aureus* Plasmid pI258. ACS Publications; 2002. https://doi.org/10.1021/bi00189a034
[83] Li Y, Hu Y, Zhang X, Xu H, Lescop E, Xia B, et al. Conformational fluctuations coupled to the thiol-disulfide transfer between thioredoxin and arsenate reductase in *Bacillus subtilis*. J Biol Chem 2007;282:11078–83. https://doi.org/10.1074/jbc.M700970200
[84] Messens J., Van Molle I., Vanhaesebrouck P., Limbourg M., Van Belle K., Wahni K., et al. How thioredoxin can reduce a buried disulphide bond. J Mol Biol. 2004 4;339(3):527-37. https://doi.org/10.1016/j.jmb.2004.04.016.
[85] Puopolo R, Gallo G, Limauro D, Contursi P, Fiorentino G. A new strategy for As(V) biosensing based on the inhibition of the phosphatase activity of the arsenate reductase from *Thermus thermophilus*. Int J Mol Sci 2022;23:2942. https://doi.org/10.3390/IJMS23062942/S1
[86] Slaughter DC, Macur RE, Inskeep WP. Inhibition of microbial arsenate reduction by phosphate. Microbiol Res 2012;167:151–6. https://doi.org/10.1016/J.MICRES.2011.05.007
[87] Zhang ZY, Wang Y, Wu L, Fauman EB, Stuckey JA, Schubert HL, et al. The Cys(X)5Arg catalytic motif in phosphoester hydrolysis. Biochemistry 1994;33:15266–70. https://doi.org/10.1021/bi00255a007
[88] Li R, Haile JD, Kennelly PJ. An arsenate reductase from *Synechocystis* sp. strain PCC 6803 exhibits a novel combination of catalytic characteristics. J Bacteriol 2003;185:6780. https://doi.org/10.1128/JB.185.23.6780-6789.2003

[89] Busenlehner LS, Pennella MA, Giedroc DP. The SmtB/ArsR family of metalloregulatory transcriptional repressors: Structural insights into prokaryotic metal resistance. FEMS Microbiol Rev 2003;27:131–43. https://doi.org/10.1016/S0168-6445(03)00054-8

[90] Xu C, Rosen BP. Dimerization is essential for DNA binding and repression by the ArsR metalloregulatory protein of *Escherichia coli*. J Biol Chem 1997;272:15734–8. https://doi.org/10.1074/jbc.272.25.15734

[91] Moinier D, Slyemi D, Byrne D, Lignon S, Lebrun R, Talla E, et al. An ArsR/SmtB family member is involved in the regulation by arsenic of the arsenite oxidase operon in *Thiomonas arsenitoxydans*. Appl Environ Microbiol 2014;80(20):6413–26. https://doi.org/10.1128/AEM.01771-14

[92] Chauhan S, Kumar A, Singhal A, Tyagi JS, Krishna Prasad H. CmtR, a cadmium-sensing ArsR-SmtB repressor, cooperatively interacts with multiple operator sites to autorepress its transcription in. Mycobacterium tuberculosis. FEBS J. 2009;276(13):3428–39. https://doi.org/10.1111/j.1742-4658.2009.07066.x

[93] Qin J, Fu HL, Ye J, Bencze KZ, Stemmler TL, Rawlings DE, et al. Convergent evolution of a new arsenic binding site in the ArsR/SmtB family of metalloregulators. J Biol Chem 2007;282:34346–55. https://doi.org/10.1074/jbc.M706565200

[94] Shi W, Dong J, Scott RA, Ksenzenko MY, Rosen BP. The role of arsenic-thiol interactions in metalloregulation of the ars operon. J Biol Chem. 1996;271(16):9291–7. https://doi.org/10.1074/jbc.271.16.9291W

[95] Shi W, Wu J, Rosen BP. Identification of a putative metal binding site in a new family of metalloregulatory proteins. J Biol Chem 1994;269:19826–9. https://doi.org/10.1016/S0021-9258(17)32094-X

[96] Rosen MR, Leuthaeuser JB, Parish CA, Fetrow JS. Isofunctional clustering and conformational analysis of the arsenate reductase superfamily reveals nine distinct clusters. Biochemistry 2020;59:4262–84. https://doi.org/10.1021/acs.biochem.0c00651

[97] Messens J., Silver S. Arsenate reduction: thiol cascade chemistry with convergent evolution. J Mol Biol. 2006 8;362(1):1-17. https://doi.org/10.1016/j.jmb.2006.07.002.

[98] Villadangos AF, Van Belle K, Wahni K, Tamu Dufe V, Freitas S, Nur H, et al. *Corynebacterium glutamicum* survives arsenic stress with arsenate reductases coupled to two distinct redox mechanisms. Mol Microbiol 2011;82:998–1014. https://doi.org/10.1111/J.1365-2958.2011.07882.X