Journal of Molecular Modeling (2019) 25: 15
https://doi.org/10.1007/s00894-018-3885-3

ORIGINAL PAPER

Structure and function prediction of arsenate reductase from *Deinococcus indicus* DR1

Deepika Chauhan[1] · Pulkit A. Srivastava[2] · Vidushi Agnihotri[2] · Ragothaman M. Yennamalli[2] · Richa Priyadarshini[1]

Received: 23 July 2018 / Accepted: 28 November 2018 / Published online: 4 January 2019

Abstract

Arsenic prevalence in the environment impelled many organisms to develop resistance over the course of evolution. Tolerance to arsenic, either as the pentavalent [As(V)] form or the trivalent form [As(III)], by bacteria has been well studied in prokaryotes, and the mechanism of action is well defined. However, in the rod-shaped arsenic tolerant *Deinococcus indicus* DR1, the key enzyme, arsenate reductase (ArsC) has not been well studied. ArsC of *D. indicus* belongs to the Grx-linked prokaryotic arsenate reductase family. While it shares homology with the well-studied ArsC of *Escherichia coli* having a catalytic cysteine (Cys 12) and arginine triad (Arg 60, 94, and 107), the active site of *D.indicus* ArsC contains four residues Glu 9, Asp 53, Arg 86, and Glu 100, and with complete absence of structurally equivalent residue for crucial Cys 12. Here, we report that the mechanism of action of ArsC of *D. indicus* is different as a result of convergent evolution and most likely able to detoxify As(V) using a mix of positively- and negatively-charged residues in its active site, unlike the residues of *E. coli*. This suggests toward the possibility of an alternative mechanism of As (V) degradation in bacteria.

Keywords Arsenate reductase · *Deinococcus indicus* DR1 · Mechanism of action · Arsenic tolerance · Molecular modeling

Introduction

Arsenic (As) is a toxic metalloid found in nature. It is released into the environment by natural processes and anthropogenic activities [1]. According to U.S. Environmental Protection Agency (EPA) and World Health Organization (WHO) drinking water standards, arsenic is a primary pollutant with permitted level of 10 μg/l [2]. It is a known human carcinogen and chronic exposure leads to adverse health conditions, such as hypertension, hyperkeratosis, lung, bladder, and skin cancers [3–7]. In nature, two ionic forms of arsenic are found, arsenite [As(III)] and arsenate [As(V)] with tri- and pentavalency, respectively [8]. Among these two forms, arsenite As(III) is considered more toxic but arsenate As(V) is comparatively more prevalent in nature [9]. Pentavalent As(V) causes toxicity by substituting phosphate with As(V), which further inhibits oxidative phosphorylation, whereas trivalent As(III) binds to protein thiols and sulfhydryl groups eliciting toxicity and interrupts normal cell signaling [10, 11]. As(III) also generates reactive oxygen species causing carcinogenicity [9].

It has been previously reported that because of the widespread distribution of arsenic in nature, many microorganisms have acquired resistance genes for heavy metals and metalloids, such as mercury, cadmium, lead, chromium, and arsenic. With the primary goal of arsenic detoxification, organisms like bacteria and fungi have developed pathways to degrade As(V) to As(III) by utilizing enzymes that catalyze As(V) degradation [12]. In some bacteria, arsenic can also act as an electron donor or acceptor in the electron transport chain but these uptake transporters lack specificity. Pentavalent and trivalent forms of arsenic are taken up by phosphate and glycerol transporters [13, 14].

Pulkit A. Srivastava and Vidushi Agnihotri contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00894-018-3885-3) contains supplementary material, which is available to authorized users.

✉ Ragothaman M. Yennamalli
ym.ragothaman@juit.ac.in; ragothaman@gmail.com

✉ Richa Priyadarshini
richa.priyadarshini@snu.edu.in

[1] Department of Life Sciences, School of Natural Sciences, Shiv Nadar University, Gautam Buddha Nagar, Uttar Pradesh, India

[2] Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, Himachal Pradesh 173234, India

Many bacteria, including both gram-negative and gram-positive, possess arsenic resistance determinants (ars) having a transcriptional unit with three to five genes. In *Staphylococcus aureus*, arsenic resistance operon present in plasmid pI258 consists of three genes *arsR*, *arsB*, and *arsC*, whereas in *E. coli* plasmid R773, *ars* operon has five genes (*arsRDABC*). Upstream to the first cistron of this cluster is a promoter sequence, which controls the overall expression of the genes. *arsR* encodes arsenic inducible-repressor, and *arsD* encodes a second *trans*-acting regulatory protein, which functions separately from repressor ArsR. The *arsD* gene is absent in the *staphylococcus* plasmid *ars* operon. *arsA*, present additionally in *E. coli* operon, determines the ATPase subunit, which drives membrane-bound efflux pump ArsB (*arsB* gene) for arsenite efflux from the cell. Arsenate reductase enzyme (transcribed by *arsC* gene) having a cysteine residue in the active site converts As(V) to As(III) using reduced glutathione (GSH). This conversion and expulsion by membrane pump confers resistance to the bacteria. In a few cases, in the absence of ArsA, ArsB compensates and acts as a secondary uniporter imparting partial arsenite resistance [15–18].

We have previously isolated and identified a novel strain *Deinococcus indicus DR1* from the wetland of Dadri, Uttar Pradesh, India. The whole genome sequencing of *D. indicus* DR1 (Accession no. NHMK00000000) showed the presence of *ars* operon genes in addition to mercury (*mer* operon), copper (*cop* operon), and chromate transporter genes [19]. In this study, we characterized the *ars* gene, specifically ArsC, of *D. indicus* DR1 after observing its high tolerance to As(V) and As(III), by using a bioinformatics approach. We report the possibility of an alternate mechanism of action that is utilized by *D. indicus* DR1 in comparison to the well-established mechanism of *E. coli*.

Materials and methods

Bacterial strain and growth conditions

D. indicus strain DR1 (Accession number NHMK00000000) was grown in tryptic soy broth (TSB) containing pancreatic digest of casein (17.0 g/L), papaic digest of soybean meal (3.0 g/L), sodium chloride (5.0 g/L), dipotassium hydrogen phosphate (2.5 g/L), and dextrose (2.5 g/L) under aerobic conditions at 30 °C (200 rpm). For solid media preparation, 1.5% agar was used with TSB. All chemicals were procured from Himedia Pvt. Ltd., Mumbai, India.

Arsenic tolerance assay

Overnight grown cultures of *D. indicus* DR1 in TSB were used (1:100 dilution) to test against different concentrations of $Na_2HAsO_4.7H_2O$ [As(V)] (Central Drug House Pvt. Ltd., New Delhi, India) and AsO_4 [As(III)] (Loba Chemie Pvt. Ltd., Mumbai, India) in a 24-well plates. After 24 h of incubation at 30 °C under shaking condition, absorbance (OD_{600}) was measured using Bio-Rad iMark microplate absorbance reader.

Multiple sequence alignment

Two sets of protein sequences were downloaded: a) 19 sequences of ArsC from available genomes of *Deinococcus*, including the ArsC of *D. indicus* DR1 (OWL94580.1), b) 11 sequences belonging to Grx- and Trx-linked arsenate reductases. The accession ids of 19 sequences mined from whole genome sequences are as follows: *Deinococcus peraridilitoris* (WP_015237143.1), *Deinococcus radiodurans* R1 (NP_285447.1), *Deinococcus actinosclerus* (WP_062158691.1), *Deinococcus maricopensis* (WP_013555254.1), *Deinococcus geothermalis* (WP_011529899.1), *Deinococcus puniceus* (WP_064015050.1), *Deinococcus proteolyticus* (WP_049775260.1), *Deinococcus peraridilitoris* (WP_015236500.1), *Deinococcus maricopensis* (WP_043817366.1), *Deinococcus gobiensis* (WP_014686224.1), *Deinococcus puniceus* (WP_064015797.1), *Deinococcus geothermalis* (WP_041220936.1), *Deinococcus hopiensis* (WP_084050784.1), *Deinococcus deserti* (WP_041227580.1), *Deinococcus swuensis* (WP_039682397.1), *Deinococcus actinosclerus* (WP_062159886.1), *Deinococcus indicus* (OWL94580.1), *Deinococcus soli* (WP_046845037.1), *Deinococcus grandis* (WP_058978479.1). For the second set, sequences were downloaded from PDB [20] and NCBI for the members of both families. The accession id for 11 sequences belonging to Grx-linked arsenate reductases are as follows: *Escherichia coli* O104:H4 str. C227–11 (EGT69533.1), Plasmid R773 (AAA21096.1), Plasmid R46 (AAB09628.1), *Yersinia enterocolitica* (AJJ28928.1), *Haemophilus influenzae* (KMZ32737.1), *Neisseria gonorrhoeae* NCCP11945 (ACF30932.1) and (WP_061288087.1). The accession id of Trx-linked arsenate reductases are as follows: Plasmid pI258 (AAA25638.1), Plasmid pSX267 (AAA27589.1), *Bacillus subtilis* (KIX84063.1) and Human B-Form lmwPTPase (1XWW). Multiple sequence alignment was performed for all 19 sequences from various *Deinococcus* genomes (Fig. 2b), and for 11 sequences (Fig. 2c) from two arsenate reductase families with ArsC of *D.indicus* DR1 using MUSCLE alignment tool [21] in Molecular Evolutionary Genetics Analysis (MEGA v7.0) software package [22]. Parameters like gap open, gap extend hydrophobicity multiplier, maximum iterations, and clustering method were set to −2.9, 0, 1.2, 8, and UPGMB, respectively.

Phylogenetic analysis

Phylogenetic analysis of two sequence sets of arsenate reductases was performed. The sequence set consisting of 12 sequences was used to identify ArsC *D. indicus*'s evolutionary relationship with two arsenate reductase families: Grx-linked and Trx-linked prokaryotic ArsC reductases [22]. The second sequence set consisting of 19 sequences from various *Deinococcus* genomes was analyzed for its sequence conservation and evolution.

The alignment generated by MUSCLE tool was subjected for construction of phylogenetic tree using the UPGMA method in MEGAv7.0 software package with the following parameters: statistical method (UPGMA), substitutions type (amino acid), model (Poisson model), rate among sites (uniform rates), and pattern among lineages (homogenous) [22].

De novo motif discovery

The Multiple EM for Motif Elicitation (MEME) [23] tool was used to search for previously unidentified sequence motifs, e.g., short patterns representing significant structural or functional characteristics of a protein and MEME motifs differ from random patterns overrepresented in a sequence. The sequence of ArsC *D.indicus* DR1 was given as query sequence in the MEME tool with parameters like motif discovery tool set to normal mode, motif distribution in sequence to any number of repetitions, number of motifs to be found by algorithm to 3, value of minimum and maximum width of motifs was set to 6 and 50, respectively; and minimum and maximum number of sites for each motif to 2 and 600, respectively.

Molecular modeling of ArsC

In order to decipher the structural aspect of arsenate reductases, the 3D structure was predicted using the ab initio method of Robetta server [24]. Amino acid sequence of ArsC of *D. indicus* DR1 (OWL94580.1) was given as input into the Robetta server for computational structure prediction. For a given sequence, Robetta breaks the query sequence into putative domains using the Ginzu domain prediction algorithm. The method used by Robetta is discussed in detail in [24].

Molecular dynamics simulation

Among the five models predicted by Robetta, we selected the first model, called Model1 throughout this study, as the structural variations among the five structures were negligible. Molecular dynamics (MD) simulation of Model1 was performed in GROMACS v5.1.2 [25–27] for 100 ns to validate the stability of the homology modeled structure of ArsC. The use of MD on homology modeled structures with longer timescales, recently, is a necessary step to evaluate stability and observe time-dependent trajectories of protein in explicit solvent [28–31].The coordinates of Model1 obtained from Robetta were solvated using SPC water model [32, 33] in a cubic box and charge neutralized using six chloride ions. The resulting structure was energy minimized for 50,000 steps using the steepest descent algorithm, followed by NVT (constant number of particles, volume, and temperature) and NPT (constant number of particles, pressure, and temperature) equilibration for 100 ps, each using the leap-frog integrator algorithm [34]. The modified Brendensen thermostat [35] was used for controlling temperature, and Parrinello-Rahman isotropic pressure coupling [36] was used for controlling pressure. The force field used for the MD simulation was OPLS-AA/L all-atom force field [37]. The production MD was reset to time of 0 ps and the total run was for 100 ns, with coordinates, velocities, energies stored every 10 ps. The timestep for integration was 2 fs, with all atoms treated explicitly and coupled with LINCS constraint algorithm [38] for constraining the hydrogen and heavy atom bonds, the Verlet cutoff scheme was used with Coloumb, and van der Waal cutoff was kept at 1.0, with particle mesh Ewald long range electrostatics [39], initial velocities were assigned based on Maxwell distribution at 300 K with a minimum distance of 1 nm in the periodic boundary conditions applied in all three directions. The root mean square deviation (RMSD) and root mean square fluctuation (RMSF) calculations were performed using the inbuilt g_rms and g_rmsf tools, respectively.

Structural homologs

DALI was used to find structural neighbors of ArsC in order to functionally annotate from the structure predicted through Rosetta. The DALI server enlists ortholog proteins for a query 3D protein structure through pairwise comparison based on a specific search called branch-and-bound search and gives nonoverlapping solutions as output in decreasing order of alignment score [40].

Active site prediction

Protein possesses global and local physicochemical properties that are created either by specific amino acid positioning or by the cavities present on the surface of proteins, which translates to its function. Mapping functionally important residues highlights the structural basis of protein function. In ArsC *D.indicus* DR1 the active site residues were mapped through CASTp (http://sts.bioe.uic.edu/castp/), an online tool that helps to locate voids and pockets present on 3D protein structures using annotations derived from three sources namely, Protein Data Bank (PDB), Online Mendelian Inheritance in Man (OMIM), and Swiss prot [40].

Active site validation

The canonical Simplified Molecular-Input Line Entry System (SMILES) format for As(V) (CID:27401) and As(III) (CID: 5460562) were obtained from PubChem (https://pubchem.ncbi.nlm.nih.gov/) database. Their 3-D structure was obtained using Corina (https://www.mn-am.com/online_demos/corina_demo), and the coordinates were downloaded in PDB format. The PDB files of ArsC of *E.coli* (pdb id: 1I9D) (closest homolog) and *D. indicus* DR1 (Model1) were superposed in PyMOL (http://www.pymol.org) to identify the location of possible arsenate binding sites. Manual docking was performed using the SO_3 ion from the *E.coli* structure as reference, where the substrate As(V) and product As(III) were oriented in a similar spatial position as SO_3. The manual method was used as arsenic is not identified by other commercial/academically free molecular docking tools. The As(V) was mapped and oriented to the active site and possible interacting residues/atoms were identified by the measurement wizard of PyMOL.

Results

D indicus DR1, a radiation resistant, Gram-negative, rod shape bacterium has additional resistance toward highly toxic forms of arsenic [19]. The strain *D. indicus* Wt/1a^T, originally isolated from an aquifer in West Bengal, India, can survive very high concentration of arsenate As(V) (10 mM) as well as arsenite As(III) (0.2 mM) [41]. Compared to the previous study, the newly identified *D. indicus* DR1 exhibited higher resistance to arsenic [19]. *D. indicus* DR1 was able to grow in As(V) (up to a concentration of 30 mM) arsenate and in As(III) (up to a concentration of 0.5 mM) (Fig. 1a and b), which is threefold and 2.5-fold more than what is reported for *D. indicus* Wt/1a^T. At 40 mM arsenate concentration, *D. indicus* DR1 was able to survive with slightly lower cell density (~0.33 OD_{600}), which further declined at 50 mM concentration (Fig. 1a). In the case of arsenite, concentrations above 1 mM drastically reduced cell density.

To elucidate the mechanism of action of ArsC in *D. indicus* DR1, we looked at ArsC of available *Deinococcus* genomes and also to categorize the ArsC to the reported three families of arsenate reductases. The multiple sequence alignment (MSA) for the available 19 ArsC protein sequences from *Deinococcus* genomes showed that a major part of their sequence is conserved. The red highlighted regions signify the amino acids that are conserved in all the sequences considered (Fig. 2a), which indicates the evolutionary relationship that most likely exists between the different species of this genus. However, in some *Deinococcus* genomes (*D. geothermalis*, *D. puniceus, D. actinosclerus, D. maricopensis,* and *D. peraridilitoris*) there is more than one sequence annotated as arsenate reductase. Also, six ArsC sequences (Accession id: WP_015237143.1, NP_285447.1, WP_062158691.1, WP_013555254.1, WP_011529899.1, WP_064015050.1) have varying length compared to the rest of the 13 ArsC sequences.

ArsC *D. indicus* closer to Grx-linked arsenate reductases

Study of all known ArsC reductases in both prokaryotes and eukaryotes reveal that though these enzymes are functionally similar they are thought to evolve independently at least three times by convergent evolution [42]. Therefore, we performed phylogenetic analysis to infer evolutionary relationship of ArsC reductases of *D. indicus* DR1 with the reported ArsC

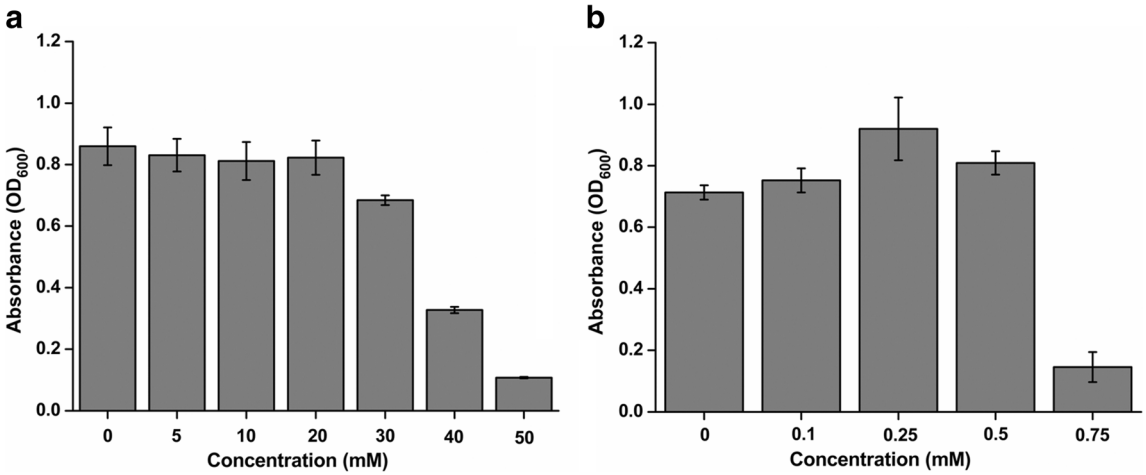

Fig. 1 Arsenate and arsenite resistance in *D. indicus* DR1. Arsenic resistance was determined by inoculating overnight grown cells (TSB) with various concentrations of (**a**) Na2HAsO4.7H2O [As(V)] (0–50 mM range) and (**b**) AsO4 [As(III)] (0–0.75 mM range), followed by 24 h incubation at 30 °C (200 rpm). Absorbance was measured at 600 nm. Error bar represents standard deviation obtained from three independent experiments

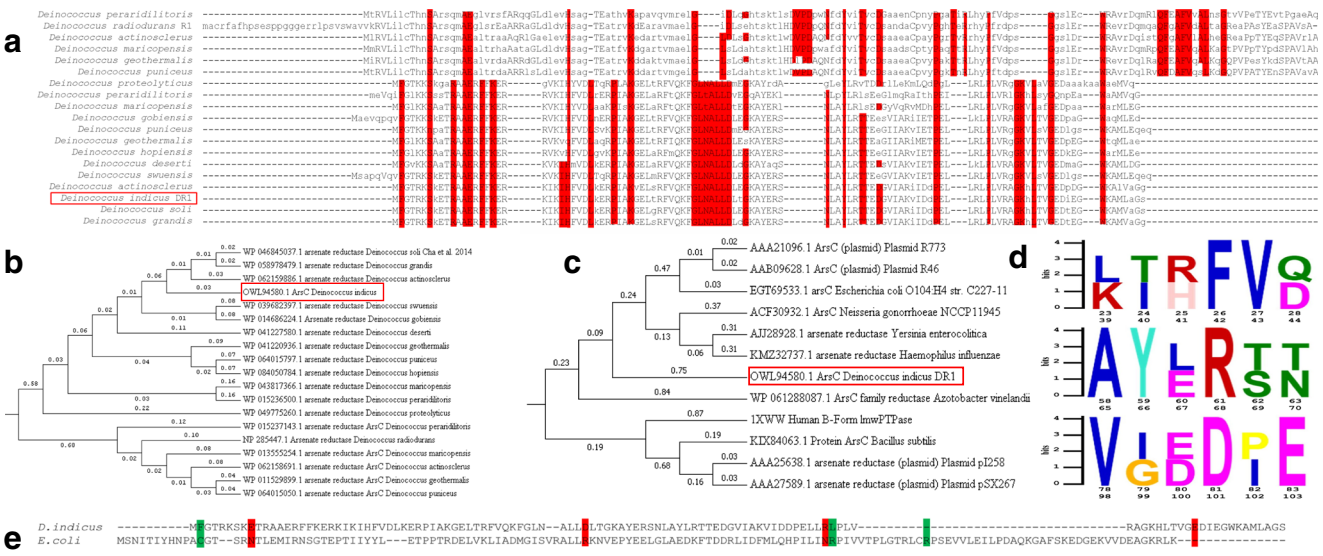

Fig. 2 Sequence analysis of arsenate reductase in *Deinococcus*. (**a**) Multiple sequence alignment of ArsC from published *Deinococcus* genomes, where the red regions indicate conserved residues. (**b**) Phylogenetic tree constructed using the MSA of 3A, where the position of ArsC *D. indicus* is shown in a red box. (**c**) Phylogenetic tree of ArsC *D. indicus* DR1 (shown in red box) with the two families of prokaryotic arsenate reductases, indicating that *D.indicus* DR1 belongs to Grx-linked prokaryotic arsenate reductases. (**d**) MEME motif identification of sequence motifs involved in structural and functional activity in ArsC *D. indicus*, where three motifs were identified occurring in six regions. (**e**) Global sequence alignment of *E. coli* ArsC with *D. indicus* DR1 ArsC showing the location of active site residues, where the *E. coli* residues are shown in green boxes and *D. indicus* DR1 predicted residues from this study are shown in red boxes

of prokaryotic family. Results generated from MEGAv7.0 shows that ArsC of *D. indicus* DR1 has closely co-evolved with the Grx-linked prokaryotic ArsC reductases family (Fig. 2c) and not with the Trx-lined prokaryotic ArsC reductases family.

Sequence motifs in ArsC of D.indicus

Three motifs identified by MEME [23] in ArsC reductases of *D. indicus* DR1 (Fig. 2d and Supplementary Table T1) were found twice in the sequence. Motif 1 was identified at residues 23–28 and 39–44, where two residues (Phe 26/42 and Val 27/43) had 4 bits level of conservation, and others had 2 bits level of conservation. Also, motif 2, ranging from 58 to 63 and 65–70 (Fig. 3d), had three residues (Ala 58/65, Tyr 59/66, and Arg 61/68) with the same 4 bits level of conservation, and other residues with 2 bits level of conservation. However, motif 3, spanning between residue number 78–83 and 98–103 (Fig. 3d), had three residues (Val 78/98, Asp81/101, and Glu83/103) with 4 bits level of conservation and three with 2 bits level of conservation.

Molecular modeling of ArsC

Since ArsC of *D. indicus* does not have a characterized 3D structure available in the PDB database, comparative modeling was performed using Robetta to obtain a high-quality structure, which gave five models (Supplementary Table T2). In the case of ArsC, the ArsC C12S mutant (PDB ID: 1S3C) from *E. coli* was used as the template to model the query sequence. The resultant models had a confidence score of 0.9040, indicating a very high quality model, where an accurate modeled structure has a score of more than 0.8 [43]. The models generated and used in this study have been deposited at ModelArchive (https://doi.org/10.5452/ma-abpir) (http://www.modelarchive.org/doi/10.5452/ma-asp8e). Figure 3a shows the structural superposition of the five models obtained from Robetta. The Ramachandran plot of Model1, considered in further analysis, shows that all the residues are within the favorable regions and there are no outliers (Supplementary Fig. S1).

Molecular dynamics simulation for 100 ns of the Model1 structure was performed to validate the predicted structure as stable, and MD simulation is a necessary step in molecular modeling [28, 29]. The simulation trajectory indicates convergence as indicated by RMSD analysis of the trajectory (Supplementary Fig. S2). Specifically, until 20 ns the structure shows fluctuations, but after that it shows minimal RMS fluctuations, indicating that the structure reached equilibration after 20 ns and the resultant trajectory is a stable conformation of ArsC. Also, the residues that fluctuate the most in the trajectory are at the C-terminus as this part is predicted to form a loop (Supplementary Fig. S2). We did not observe high fluctuations for the residues that are predicted to be in the active site (Supplementary Fig. S3). The radius of gyration analysis shows that the protein does not denature during the simulation

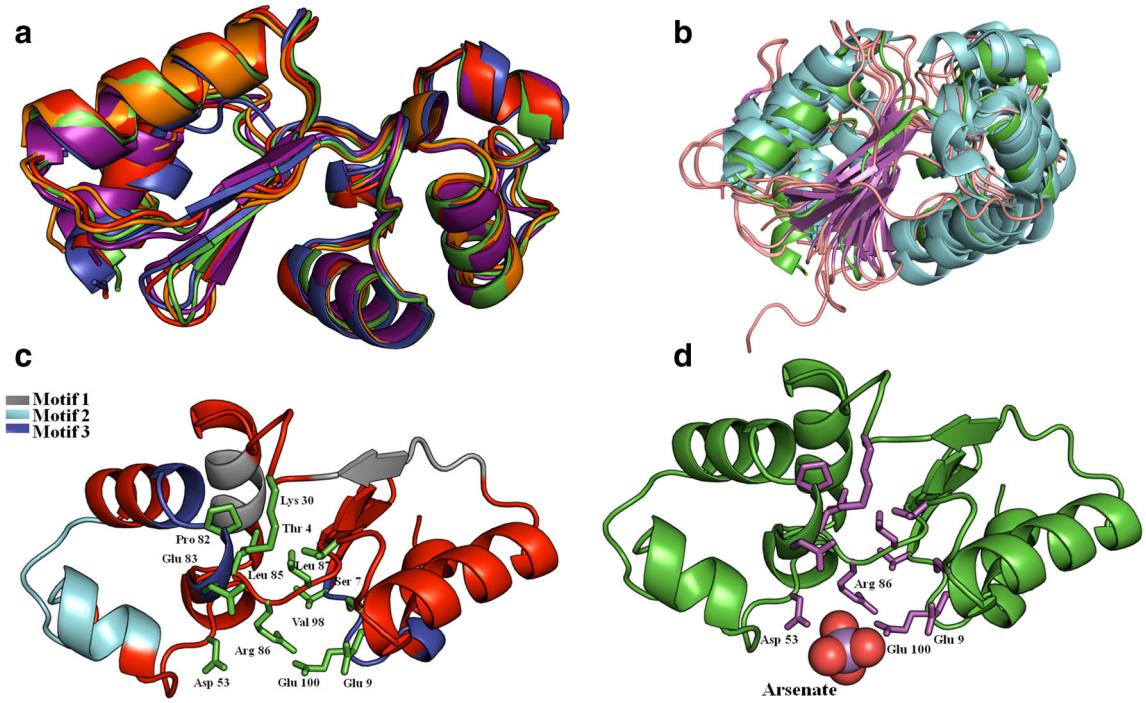

Fig. 3 Molecular modeling and active site prediction in ArsC *D.indicus* DR1. (**a**) Superposition of five models obtained from Robetta showing low variations among each other. (**b**) Superposition of Model1 of ArsC with structural homologs identified using DALI (pdb id: 3RDW, 1I9D, 1J9B, 2KOK, 2 M46, 2MU0, and 3F0I) showing the fold level conservation among the various arsenate reductases. The RMSD difference between Model1 of ArsC with 3RDW is 1.5 Å, 1I9D is 2.2 Å, 1J9B is 2.2 Å, 2KOK is 3.0 Å, 2 M46 is 2.4 Å, 2MU0 is 3.7 Å, and 3F0I is 1.7 Å. (**c**) MEME motifs (colored gray, cyan, and blue) and CASTp active site residues Thr 4, Ser 7, Glu 9, Lys 30, Asp 53, Pro 82, Glu 83, Leu 85, Arg 86, Leu 87, Val 98, and Glu 100(shown as sticks) mapped onto Model1 structure. (**d**) Molecular docking of As(V) substrate with ArsC, indicating the putative residues (Glu 9, Asp 53, Arg 86, and Glu 100) that are proposed in this manuscript as crucial residues

and was stable throughout the simulation (Supplementary Fig. S4).

Thus, MD simulation demonstrates that the high-quality homology modeled structure obtained from Robetta is energetically stable to perform a molecular docking of arsenate and arsenite to identify the possible active site residues (Supplementary Fig. S5).

Structural homologs and putative active site prediction

Using DALI we identified structural homologs for Model1 of ArsC, where the top hits were arsenate reductases from *E. coli, Vibrio cholerae, Yersinia pestis, Staphylococcus aureus,* and *Bacteroides fragilis*. Comparing the structural homologs we observed that ArsC *D. indicus* DR1 is closest to ArsC *E.coli* (RMSD of 1.5 Angstroms) (Fig. 3b). Also, the results indicate that the fold is conserved among the various arsenate reductases. However, there are differences between Model1 (shown in green in Fig. 3b) with the other structural orthologs (shown in cyan and magenta in Fig. 3b). Specifically, the loop regions, the spatial position of the helices, and the β strands show variations.

Comparing the closest structural homolog from *E.coli*, the three active site residues (Cys 12, Arg 60, Arg 94, and Arg 107) reported by Mukhopadhyay [42] were used to identify structurally equivalent residues in ArsC of *D. indicus* DR1. Surprisingly, the structurally equivalent residues in ArsC were Asp 53 and Leu 87, for Arg 60 and Arg 94, respectively. However, there was no structurally equivalent residue found for Cys 12.

In order to identify key residues that bind to As(V) and convert it to As(III), we used CASTp to predict the active site/substrate binding site region of ArsC. As a general rule, the largest cavity in a protein is the most likely place for an active site [44, 45]. Thus, among the 15 cavities predicted by CASTp, the largest cavity with a surface area of 227.6 $Å^2$ and a volume of 342.6 $Å^3$ was selected for further analysis. The residues that comprise this cavity are Thr 4, Ser 7, Glu 9, Lys 30, Asp 53, Pro 82, Glu 83, Leu 85, Arg 86, Leu 87, Val 98, and Glu 100.

Discussion

Growth of *D. indicus* DR1, isolated from wetland of Dadri, at higher As(V) and As(III) concentration (Fig. 1a and b) indicates that the organism is capable of degrading or detoxifying arsenic more than the strain *D. indicus* Wt/1a^T [41]. Specifically, the 2.5–3 fold higher tolerance is indicative of

efficient cellular machinery that can allow the organism to survive harsh conditions. The inferences of the results obtained using MSA were further extended to understand the phylogenetic tree constructed (Fig. 2b) using MEGA, which shows that *D. indicus* has most likely co-evolved with *D. actinosclerus*, *D. grandis*, and *D. soli* as they form a distinct subclade. This is also in correlation with the output of MSA (Fig. 2a) wherein it is evident that the alignment corresponding to ArsC of *D. indicus* (OWL94580.1) and ArsC of *D. actinosclerus* (WP062159886.1) is significantly similar. However, its evolutionary distance (0.75) (Fig. 2c) with the homolog from *E.coli* (sharing a sequence identity of 31%) indicates that although they belong to the same family (Grx-linked) of arsenate reductase, the mechanism of action to degrade As(V) may be different in ArsC of *D. indicus* compared with the Grx-lined prokaryotic family of ArsC reductases, as the phylogenetic tree indicates that ArsC *D. indicus* DR1 is not that close to Grx-linked arsenate reductases.

Molecular modeling of ArsC *D. indicus* DR1 showed that all five structures belong to a 3-layer sandwich fold, indicating that ArsC of *D. indicus* DR1 has α/β topology, where there are seven helices and four β-strands sandwiched between the helices. Comparing all five models (Fig. 3a), the structural differences between the five models is minimal (average RMSD between all five models was 0.58 Angstroms), and we selected Model1 for further analysis and predictions. A possible explanation to the structural variation in Model1 compared with other ArsC structures (Fig. 3b) could be the diverged sequence in *D.indicus*. While *D.indicus* shows to be part of the Grx-linked prokaryotic ArsC reductase sub-clade (Fig. 2c), it is not closer in terms of sequence similarity to the Grx-linked prokaryotic ArsC reductases.

Comparison of the three motifs (motif 1, motif 2, and motif 3) identified by MEME on Model1 (Fig. 3c) and the predicted residues by CASTp was performed to find any residue(s) that possibly have a function associated with As(V) binding. Accordingly, Pro 82 and Glu 83 residues of motif 3 (78–83 in Fig. 2d) that are part of the helical structure (Fig. 3c), and Val 98 and Glu 100 residues, also of motif 3 (98–103 in Fig. 2d) and part of a loop (Fig. 3c), were also predicted by CASTp as putative active site residues. This indicates that certain residues are likely to be involved in the function of ArsC, while others are possibly involved with providing structural stability to the protein. Specifically, the residues involved in motif 1 and motif 2.

Novel mechanism of action in ArsC *D. indicus* DR1?

To identify the putative residues from the above mentioned computational methods to predict the motifs and active site residues, the superposed ArsC structures of *E.coli* and *D. indicus* DR1 were used, and we identified a sulfite ion that is in close proximity to Cys 12. We think that although the sequences of *E.coli* ArsC and *D. indicus* DR1 ArsC diverge to a large extent (31% sequence identity), their structural conservation (1.5 Å RMSD) coupled with the absence of the crucial cysteine (Cys 12 in *E. coli*) in ArsC of *D. indicus* DR1 indicates that there could be an alternate mechanism of action involving different residues. This is supported by comparing the *E.coli* and *D.indicus* sequences (Fig. 2e), where they are of unequal length (*E.coli* has 141 residues and *D.indicus* has 112 residues) and the three active site residues of *E.coli* (green boxes in Fig. 2e) and *D.indicus* are not aligned. The exception is Arg 94 of *E.coli* that aligns with Asp 53 of *D.indicus*. It is interesting to note that the sequence-based alignment of *E.coli* residues is not reflected in the structure based alignment of the modeled strcture of ArsC in terms of the active-site residues. Hence, the manual docking of arsenate was performed using the sulfite ion of *E. coli* as a reference. Calculating the distances between the As(V) ion to the predicted active site residues, we found Glu 9, Asp 53, Arg 86, and Glu 100 within 5 Å radius of As(V) (Fig. 3d). Also, As(V) makes a polar contact with the oxygen atom of Asp 53 suggesting that Asp 53 is indeed one of the crucial active site residues. Surprisingly, Leu 87 (structurally equivalent residue to Arg 94 of *E.coli*) was not in close proximity to As(V). However, we observed that with suitable rotamer change, Arg 86 and Glu 9 can also make polar contacts with As(V). A preliminary analysis of the product, As(III), with ArsC of *D.indicus* shows that Asp 53, Glu 100, and Arg 86 are in close proximity as well (within 5 Å radius) (Supplementary Fig. S5). Thus, we propose that As(V) binds to ArsC of *D. indicus* DR1 to four charged residues and is most likely to have evolved a different mechanism of action for degrading arsenate as compared to the existing literature and reports.

Future work to establish the mechanism as to how *D. indicus* DR1 tolerates conditions of high arsenic concentration needs to be performed, involving site-directed mutagenesis and biochemical characterization of ArsC. Also, the synergistic activity of other *ars* operon genes and their products with ArsC in *D. indicus* DR1 would shed light on the precise mechanism.

Acknowledgments PAS, VA, and RMY acknowledge Jaypee University of Information Technology for providing research facilities to conduct this study. RP acknowledges Shiv Nadar University for providing research facilities and DC is supported by a doctoral research fellowship from Shiv Nadar University.

Data availability The model used in this study has been deposited at ModelArchive (https://www.modelarchive.org/doi/10.5452/ma-abpir) and is available for public use. The trajectory snapshot from MD simulation has been uploaded as Supplementary material.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Nriagu JO, Pacyna JM (1988) Quantitative assessment of worldwide contamination of air, water and soils by trace metals. Nature 333(6169):134–139. https://doi.org/10.1038/333134a0
2. Nordstrom DK (2002) Public health. Worldwide occurrences of arsenic in ground water. Science 296(5576):2143–2145. https://doi.org/10.1126/science.1072375
3. Smith AH, Hopenhayn-Rich C, Bates MN, Goeden HM, Hertz-Picciotto I, Duggan HM, Wood R, Kosnett MJ, Smith MT (1992) Cancer risks from arsenic in drinking water. Environ Health Perspect 97:259–267
4. Naujokas MF, Anderson B, Ahsan H, Aposhian HV, Graziano JH, Thompson C, Suk WA (2013) The broad scope of health effects from chronic arsenic exposure: update on a worldwide public health problem. Environ Health Perspect 121(3):295–302. https://doi.org/10.1289/ehp.1205875
5. Chen CJ, Hsueh YM, Lai MS, Shyu MP, Chen SY, Wu MM, Kuo TL, Tai TY (1995) Increased prevalence of hypertension and long-term arsenic exposure. Hypertension 25(1):53–60
6. Karagas MR, Tosteson TD, Blum J, Morris JS, Baron JA, Klaue B (1998) Design of an epidemiologic study of drinking water arsenic exposure and skin and bladder cancer risk in a U. S. population. Environ Health Perspect 106(Suppl 4):1047–1050
7. Chu HA, Crawford-Brown DJ (2006) Inorganic arsenic in drinking water and bladder cancer: a meta-analysis for dose-response assessment. Int J Environ Res Public Health 3(4):316–322
8. Cullen WR, Reimer KJ (1989) Arsenic speciation in the environment. Chem Rev 89(4):713–764. https://doi.org/10.1021/cr00094a002
9. Liu SX, Athar M, Lippai I, Waldren C, Hei TK (2001) Induction of oxyradicals by arsenic: implication for mechanism of genotoxicity. Proc Natl Acad Sci U S A 98(4):1643–1648. https://doi.org/10.1073/pnas.031482998
10. Klaassen CD, Casarett LJ, Doull J (2013) Casarett and Doull's toxicology: the basic science of poisons, vol xiii, 8th edn. McGraw-Hill, New York
11. Kaltreider RC, Davis AM, Lariviere JP, Hamilton JW (2001) Arsenic alters the function of the glucocorticoid receptor as a transcription factor. Environ Health Perspect 109(3):245–251
12. Silver S, Phung LT (2005) Genes and enzymes involved in bacterial oxidation and reduction of inorganic arsenic. Appl Environ Microbiol 71(2):599–608. https://doi.org/10.1128/AEM.71.2.599-608.2005
13. Rosen BP, Liu Z (2009) Transport pathways for arsenic and selenium: a minireview. Environ Int 35(3):512–515. https://doi.org/10.1016/j.envint.2008.07.023
14. Sanders OI, Rensing C, Kuroda M, Mitra B, Rosen BP (1997) Antimonite is accumulated by the glycerol facilitator GlpF in Escherichia coli. J Bacteriol 179(10):3365–3367
15. Rosen BP (1999) Families of arsenic transporters. Trends Microbiol 7(5):207–212
16. Carlin A, Shi W, Dey S, Rosen BP (1995) The ars operon of Escherichia coli confers arsenical and antimonial resistance. J Bacteriol 177(4):981–986
17. Diorio C, Cai J, Marmor J, Shinder R, DuBow MS (1995) An Escherichia coli chromosomal ars operon homolog is functional in arsenic detoxification and is conserved in gram-negative bacteria. J Bacteriol 177(8):2050–2056
18. Ji G, Silver S (1992) Reduction of arsenate to arsenite by the ArsC protein of the arsenic resistance operon of Staphylococcus aureus plasmid pI258. Proc Natl Acad Sci U S A 89(20):9474–9478
19. Chauhan D, Srivastava PA, Yennamalli RM, Priyadarshini R (2017) Draft genome sequence of Deinococcus indicus DR1, a novel strain isolated from a freshwater wetland. Genome Announce 5(31). https://doi.org/10.1128/genomeA.00754-17
20. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res 28(1):235–242. https://doi.org/10.1093/nar/28.1.235
21. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32(5):1792–1797. https://doi.org/10.1093/nar/gkh340
22. Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33(7):1870–1874. https://doi.org/10.1093/molbev/msw054
23. Bailey TL, Williams N, Misleh C, Li WW (2006) MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res 34(Web Server issue):W369–W373. https://doi.org/10.1093/nar/gkl198
24. Kim DE, Chivian D, Baker D (2004) Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res 32(Web Server issue):W526–W531. https://doi.org/10.1093/nar/gkh468
25. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ (2005) GROMACS: fast, flexible, and free. J Comput Chem 26(16):1701–1718. https://doi.org/10.1002/jcc.20291
26. Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D, Hess B, Lindahl E (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29(7):845–854. https://doi.org/10.1093/bioinformatics/btt055
27. Berendsen HJC, Vanderspoel D, Vandrunen R (1995) Gromacs - a message-passing parallel molecular-dynamics implementation. Comput Phys Commun 91(1–3):43–56. https://doi.org/10.1016/0010-4655(95)00042-E
28. Jabeen A, Mohamedali A, Ranganathan S (2019) Protocol for protein structure modelling. In: Ranganathan S, Gribskov M, Nakai K, Schönbach C (eds) Encyclopedia of bioinformatics and computational biology. Academic, Oxford, pp 252–272. https://doi.org/10.1016/B978-0-12-809633-8.20477-9
29. Dwyer DS (2003) Molecular modeling and molecular dynamics simulations of membrane transporter proteins. In: Yan Q (ed) Membrane transporters: methods and protocols. Humana, Totowa, pp 335–350. https://doi.org/10.1385/1-59259-387-9:335
30. Kokabu Y, Ikeguchi M (2013) Molecular modeling and molecular dynamics simulations of recombinase Rad51. Biophys J 104(7):1556–1565. https://doi.org/10.1016/j.bpj.2013.02.014
31. Zimmermann MT, Urrutia R, Oliver GR, Blackburn PR, Cousin MA, Bozeck NJ, Klee EW (2017) Molecular modeling and molecular dynamic simulation of the effects of variants in the TGFBR2 kinase domain as a paradigm for interpretation of variants obtained by next generation sequencing. PLoS One 12(2):e0170822. https://doi.org/10.1371/journal.pone.0170822
32. Hermans J, Berendsen HJC, Vangunsteren WF, Postma JPM (1984) A consistent empirical potential for water-protein interactions. Biopolymers 23(8):1513–1518. https://doi.org/10.1002/bip.360230807
33. Postma JPM, Berendsen HJC, Straatsma TP (1984) Intramolecular vibrations from molecular-dynamics simulations of liquid water. J Phys-Paris 45(Nc-7):31–40. https://doi.org/10.1051/jphyscol:1984703
34. Buneman O (1983) Computer-simulation using particles - Hockney,Rw, Eastwood,Jw. SIAM Rev 25(3):425–426. https://doi.org/10.1137/1025102

35. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR (1984) Molecular-dynamics with coupling to an external Bath. J Chem Phys 81(8):3684–3690. https://doi.org/10.1063/1.448118
36. Parrinello M, Rahman A (1981) Polymorphic transitions in single-crystals - a new molecular-dynamics method. J Appl Phys 52(12): 7182–7190. https://doi.org/10.1063/1.328693
37. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. J Phys Chem B 105(28):6474–6487. https://doi.org/10.1021/jp003919d
38. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: a linear constraint solver for molecular simulations. J Comput Chem 18(12):1463–1472. https://doi.org/10.1002/(Sici)1096-987x(199709)18:12<1463::Aid-Jcc4>3.0.Co;2-H
39. Wennberg CL, Murtola T, Hess B, Lindahl E (2013) Lennard-Jones lattice summation in bilayer simulations has critical effects on surface tension and lipid properties. J Chem Theory Comput 9(8): 3527–3537. https://doi.org/10.1021/ct400140n
40. Holm L, Kaariainen S, Rosenstrom P, Schenkel A (2008) Searching protein structure databases with DaliLite v.3. Bioinformatics 24(23):2780–2781. https://doi.org/10.1093/bioinformatics/btn507
41. Suresh K, Reddy GS, Sengupta S, Shivaji S (2004) Deinococcus indicus sp. nov., an arsenic-resistant bacterium from an aquifer in West Bengal, India. Int J Syst Evol Microbiol 54(Pt 2):457–461. https://doi.org/10.1099/ijs.0.02758-0
42. Mukhopadhyay R, Rosen BP (2002) Arsenate reductases in prokaryotes and eukaryotes. Environ Health Perspect 110(Suppl 5): 745–748
43. Song Y, DiMaio F, Wang RY, Kim D, Miles C, Brunette T, Thompson J, Baker D (2013) High-resolution comparative modeling with RosettaCM. Structure 21(10):1735–1742. https://doi.org/10.1016/j.str.2013.08.005
44. DesJarlais RL, Sheridan RP, Seibel GL, Dixon JS, Kuntz ID, Venkataraghavan R (1988) Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure. J Med Chem 31(4):722–729
45. Laskowski RA, Luscombe NM, Swindells MB, Thornton JM (1996) Protein clefts in molecular recognition and function. Protein Sci 5(12):2438–2452. https://doi.org/10.1002/pro.5560051206