Sequence based homology study of metalloregulatory protein ArsR and cloning of *arsR* gene from *Enterobacter cloacae* in *E. coli* for arsenic bioremediation

Mariyaraj Sivakumar Selvi[a#], Rajkumar Prabhakaran[a#], Prakash Shyam Karuppiah[b], Eldho Paul[a],

Abhishek Albert[a] & Govindan Sadasivam Selvam [a]*

[a]*Department of Biochemistry, Centre for Advanced Studies in Functional Genomics, School of Biological Sciences, Madurai Kamaraj University, Madurai TN, 625021, India.*

[b]*Lead Discovery Group, Discovery Biologics, Chigene Biotech Private Limited, Sivagangai Tamil Nadu 630211, India*

#both authors contributed

Abstract : Arsenic (As) toxicity is a serious health problem that affects millions of population around the world. Particularly, monitoring and removal of arsenite [As (III)] in the environment becomes inevitable. The existing removing techniques having some disadvantages, the reseach revealed the best alternative is microbial As (III) remediation. Microbial As (III) detoxification pathway involving the *ars* operon system offers a cost-effective and efficient opportunity in both detection and biodegradation. Hence, the molecular mechanism and regulation of *ars* operon genes and proteins for arsenic detoxicification is still unclear. Nowadays, *in silico* studies provide better understanding of the mechanism of metal binding to target proteins. In this context, the plasmid encoded 353bp length of *arsR* gene was identified from *E. cloacae* BC2 and its homodimeric 3D structure was predicted. *In silico* study exhibited ArsR is a cytoplasmic, soluble protein and its metal binding sites are H3, C30, C32, C35, H48, C89 and C106. This result implicates the N-terminal cysteine residues has vital role for As (III) removal. Docking indicated that As (III) binding with homodimeric ArsR was more selective and efficient than other heavy metals. The order of metals binding with ArsR are $As^{(III)}$>$As^{(V)}$>Co>Cd>Cr>Pb>Zn>Ag. In addition, the recombinant *E.coli* strain bearing *arsR* gene have efficient ArsR (12kDa) expression and significantly As removal rate at the concentration of 1000ppb of As (III). The present study provides us the structural and functional view of ArsR and its possibility of using engineered recombinant *E. coli* expressing ArsR protein as an inexpensive alternative solution for As (III) removal from contaminated environment.

Keywords: *Arsenic toxicity; ars operon; arsenic resistance; trans-acting regulatory protein; bioremediation.*

I. Introduction

Heavy metal contamination in the environment is a serious problem resulting in the deterioration of human health [1]. In particular, As is an extremely toxic heavy metal compound present in the environment. As usually exists as organic and inorganic forms in nature; the inorganic arsenate [As (V)] and arsenite [As (III)] are toxic to human health and that causes black foot disease, cancers, osteoporosis, lung disease and atherosclerosis in human [2,3]. As is released into the environment by natural and anthropogenic sources. There are several reports on increased levels of As in soil, drinking, irrigation water and presence of As in variety of foodstuffs like cereals, vegetables and ani-mal food such as fish, meat, and milk [4-7]. In consequence, controlling the level of As in environment is gaining primary importance.

Reports indicate As contamination in soil and drinking water has increased several times in many parts of the world [6]; WHO 2011 and ATSDR 2015). In order to control and remediate As contamination, a number of conventional chemical methods are used at present. As (III) removal from contaminated soil and ground water can be attained by precipitation-coagulation, membrane separation, co-precipitation, adsorption, and ion-exchange methods. But the demerits of currently available methods include high cost on processing the sample for detection and instrumental cost, time consuming [8,9]. The major disadvantages of these methods are more expensive and they produce secondary compounds that are more toxic than As *per se*. Moreover, at present As (III) remediation needs sample manipulation and extraction process, thus, As (III) removal is a highly complex procedure and roadblocks exist at the

[a.] *Department of Biochemistry, Centre for Advanced Studies in Functional Genomics, School of Biological Sciences, Madurai Kamaraj University, Madurai 625 021, India.*

*# **Equal contribution.** *Corresponding Author. Tel: +91 452 2458213, E-mail: drselvamgsbiochem@rediffmail.com*

Electronic Supplementary Information (ESI) available: [details of supplementary information available in separately]. See http://doi.one/10.1729/IJCRT.17414

detection stage itself. Nevertheless quantification of As (III) in the environment is further tedious owing to availability of sophisticated instruments and laboratory facilities. Therefore, an alternative method of As bioremediation is needed [10]. Nowadays, microbial bioremediation gaining more importance for the removal of As from environment.

Bacterial systems offer viable solution to meet the demands of cost effective and simple As biodegradation method and also offer us with an option to use them as biosensor. Briefly, bacteria's have evolved to resist As (III) toxicity and maintain cellular response in terms of metabolic processes to remediate As [11,12]. Various bacterial genera have developed different resistance mechanism for As compounds due to the natural abundance of As in the environment. As removal capability has been reported in different bacterial species such as *Alcaligenes faecalis, Bacillus subtilis, Pseudomonas putida* and *Pseudomonas arsenitoxidans* etc., Bacteria from genus*, Citrobacter, Enterobacter, Psychrobacter, Vibrio and Bosea* are also reported as As resistant as well as As transforming genera [13,14].

As resistance in bacteria is conferred by a variety of mechanisms: (a) minimizing the uptake of As (V) through the system for phosphate uptake, bacterial plasmids encode specific efflux pumps that are able to force out As (III) from the cytoplasm thereby the cells maintain the intracellular concentration As(III) [15] (b) by peroxidation reactions with membrane lipids, and (c) using As (III) detoxification pathway involving the *ars* operon. *ars* operon system has either three (*arsRBC*) or five (*arsRDABC*) genes are organized into a single transcriptional unit [16]. The three-gene system comprises of *arsRBC*, *arsR* encodes As transcriptional repressor, *arsB* encodes As (V) permease protein, *arsC* encodes As reductase enzyme [14]. This system is found to be present in the chromosome of *Escherichia coli*, *Pseudomonas aeruginosa*, and other *Enterobacteria* and also in plasmids pI258 and pSX267 of *Staphylococcus*. In the five-gene operon system *arsRDABC, arsR* encodes an As (III)-inducible trans-acting metalloregulatory protein, *arsD* encodes negative regulatory protein, *arsA* encodes ATPase enzyme, *arsB* encodes a membrane located As (III) efflux pump protein, and *arsC* encodes As (V) reductase enzyme. This five-gene operon system *arsRDABC* was initially discovered in the *E. coli* plasmids R773 and R46 and then on plasmid pKW301 from *Acidiphilium multivorum*.

Studies suggest that recombinant bacteria capable of removing As (III) from their environment could be an ideal alternative for existing physicochemical method [17]. Reports put forward that the genetically modified bacteria overexpressing ArsR protein acts as an effective tool for treatment of As (III) contaminated environment [18]. *Enterobacter* sp. having the characteristic features of multiple antibiotics resistance confer *ars* operon (either of chromosomal or plasmid origin) and present a transport system to expel toxic As (III) out of the cell [19]. We have previously reported that As (III) resistance in *Enterobacter* sp. BC1 and *Enterobacter.* sp BC2 is plasmid encoded and is conferred by the *aoxA/arsC* genes. Herein, we report the structural features of ArsR protein in *in silico* analysis and propose an effective strategy to use engineered *E. coli* inserted with *Enterobacter cloacae* (*E. cloacae*) *arsR* gene for As bioremediation. In the present work, the *arsR* gene was identified from As resistant bacteria *E. cloacae* BC2 and evaluated the protein structure, molecular docking and metal binding efficiency of ArsR using computational approach. Furthermore, the genetically engineered *E. coli* cells over-expressing *arsR* that tested for selective removal of As (III) from the contaminated water. Therefore, the present work will provide solution for cost effective removal of As from contaminated environment and will pave a way in successful development of a cost-effective As biosensor.

Materials and methods

Chemicals and Stock Solutions

Chemicals used in this study were purchased from Himedia, India. Stock solutions of As (III) and As (V) were prepared freshly before use from sodium arsenite ($NaAsO_2$) and sodium arsenate ($Na_2HAsO_4.7H_2O$), respectively.

Strains and Plasmids

Bacterial strains and primers used in the study are listed in Supplementary Table S1 and S2 respectively. *E. coli* DH5α (Novagen, USA) and *E. coli* BL21 (Novagen, USA) were used for plasmids multiplication and for expression of heterologous protein. Plasmid pUCP18 (donated by Prof. Pradeep Singh, University of Washington) was employed to clone the *arsR* gene for protein expression. As resistant *E. cloacae* (*arsR*) were isolated from agricultural soil from Madurai, India.

Isolation and identification of *arsR* gene

The plasmid DNA was isolated from As resistant bacterial *E. cloacae* BC2 using Himedia - HiPurA™ plasmid DNA Miniprep Purification Kit method. The *arsR* gene amplifications was carried out by PCR (Gene Amp PCR system 2700, Applied Biosystems) using plasmid DNA of *E. cloacae* BC2 as the template and primers *arsR* F and *arsR* R with the following cycling conditions. The PCR

program consisted of an initial denaturation at 94 °C for 3 min and 35 cycles of 94 °C for 30 s and 52.7 °C for 30 s and extension 72 °C for 30 s followed by final extension at 72 °C for 5 min. The *arsR* gene amplicon was eluted and sequenced. The obtained gene sequences were confirmed based on homology analysis by using NCBI BLAST software [20]. The *arsR* gene sequences were submitted to GenBank.

In silico analysis

Retrieval of target sequence and phylogenetic analysis

The nucleotide sequence of the *arsR* gene was retrieved from the NCBI (National center for biotechnology information) database (http://www.ncbi.nlm.nih.gov/) by hitting the accession number KT251200 in which the ArsR AKQ21142 (Uniprot ID) protein sequences were obtained from SWISS-PROT data bank for structural and functional analysis. A total of 74 ArsR protein sequences of *E. cloacae* were collected from SWISS-PROT (Supplementary Table S3). Sequences with significant identity were aligned with ClustalW algorithm implemented in Molecular Evolutionary Genetic Analysis (MEGA 6) (http://www.megasoftware.net) by using distance matrix and it was trimmed to consensus [21]. Neighbour Joining (NJ) trees were constructed with 1000 bootstraps at uniform divergence rates with distance '*p*' as the evolutionary model and with a data subset to use with gaps/ missing data treatment complete deletion [22]. Posterior probability and conserved regions among the closely related sequences were done with MEGA 6 [23].

Physicochemical characterization

Physicochemical properties of ArsR protein in *Enterobacter* sp. such as isoelectric point (pI), molecular weight, number of atoms present, aliphatic index, and grand average of hydropathicity (GRAVY) were computed using ProtParam tool (http://web.expasy.org/protparam/) [24]. Prediction of sub cellular localization of the protein was carried out by CELLOv.2.5 [25]. The secondary structure of the protein was studied using PSIPRED server and BLAST from NCBI to compare the query sequence with the database homologues sequences [26,27]. Conserved domains were analyzed from CDD of BLAST. The query protein domain study was performed using Interproscan analysis [28]. In order to study the tentative functional assignments the ArsR protein fingerprints was analyzed by PRINTS server [29]. The metal binding sites were identified by MetalDetector v2.0 (http://metaldetector.dsi.unifi.it/v2.0/).

Conserve domain and motif analysis

The conserved pattern of *arsR* gene sequence was identified using motif search (http://www.genome.jp/tools/motif/). Target sequence domain annotation, precise locations of domain boundaries and functional sites were analyzed by conserved domain database (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). The target protein secondary structure was predicted by PDBsum (www.ebi.ac.uk/pdbsum), Phyre2 and SOPMA server.

Three dimensional structure modeling, energy minimization and validation of the model

Comparative 3D modelling has been performed in the following stages: *arsR* gene encoded target ArsR sequences of *E. cloacae* BC2 was identified and taken from *E. cloacae* AKQ21142 (Uniprot ID) protein database. The target protein sequence was retrieved in fasta format and subjected to BLASTp against PDB database for obtaining most similar crystal structure, used as a template for the target protein modeling, which resulted as the identification of 3F6O PDBid (Homodimeric ArsR protein) as suitable template. The three dimensional structure of target protein was predicted by Swiss-Model tool and RaptorX server (http://raptorx.uchicago.edu/) [30]. The modelled protein structures were viewed using Swiss-Model tool (http://www.expasy.org/spdbv/) and the individual residues were collected using 100 cycles of steepest descent algorithm in GROMOS96 [31]. The value of the predicted ArsR protein model was analyzed by PyMOL [32]. Energy minimized models were assessed for analyzing the stereo chemical quality and residue by residue geometry. RAMPAGE (http://mordred.bioc.cam.ac.uk/~rapper/rampage) was applied for validating the modeled target protein structure by Ramachandran Plot [33]. The QMEAN Z scores were calculated for the complete assessment and refinement of the predicted model.

Docking studies

The protein docking study was performed to identify the interaction between metal protein complexes. The 2D structure of heavy metals (Arsenate, Arsenic, Cadmium, Chromium, Cobalt, Copper, Nickel, Lead, Silver and Zinc ion) were obtained from PUBCHEM database in SDF format and converted to PDB format using online molecular converter tool (http://cactus.nci.nih.gov/translate/). The docking analysis of 3D structure of ArsR and heavy metal ions was done in AutoDock4.2.1, graphical user interface AutoDock Tools which is molecular modelling simulation software, effective for protein-ligand docking and Patch Dock [34,35]. In general

the docking parameters were kept to default values. The 100 independent runs were processed using the built in clustering analysis with a 2.0 Å cut-off. The binding patterns of the docked complexes were analyzed on UCSF Chimera, an extensible molecular modeling system for the visualization and analysis [36].

In vitro assays

PCR analysis of the recombinant strains

The *E. coli* BL21 transformants bearing plasmid pUCP-P_{lac}-gfp^{mut2} were selected and cultured in LB medium with 100 µg/mL of ampicillin. The plasmids were isolated using GenElute plasmid Miniprep Kit (Sigma), according to the manufacturer's protocol. Then PCR amplification was performed using plasmid DNA of transformant as the template and primers *arsR* F and *arsR* R. The PCR program consisted of an initial denaturation at 94 °C for 3 min and 35 cycles of 94 °C for 30 s and 52.7 °C for 30 s and extension 72 °C for 30 s followed by final extension at 72 °C for 5 min. The PCR fragments were analyzed using 2% agarose gel electrophoresis.

Strain cultivation and protein expression

E. coli BL21 harboring plasmid pUCP-P_{lac}-gfp^{mut2} was grown overnight at 37 °C with shaking at 120 rpm. Then 100µl of overnight culture was transferred into 100 mL of LB medium with 100 µg/mL of ampicillin were grown until optical density of the cells reached 0.6 at wavelength of 600 nm. In order to induce the protein expression, 1mM of isopropyl β-D-1-thiogalactopyranoside (IPTG) was added. Then the cells were further incubated for a period of 12 h at 37 °C and subsequently harvested by centrifugation at 5000 rpm for 5 min. The obtained cell pellet was sonicated and analyzed using 12% SDS-PAGE with Coomassie Blue R-250 staining.

Atomic absorption spectro-photometry analysis

As removal ability of recombinant *E.coli* harboring *arsR* gene was carried out in LB broth supplemented with 1000ppb As (III). Bacterial cultures were grown aerobically for 24 h in 25 ml medium at 37 °C. The cells were separated from the mixture by centrifuging at 4000 rpm for 5 min, followed by carefully removal of the supernatant at different time intervals (0,4,8,12,16,20 and 24h). Supernatants of bacterial cultures were subjected to AAS (GBC-SensAA, Australia). *E. coli and E. cloacae* BC2 grown in broth supplemented in 1000ppb As (III) were treated as controls. As (III) content in the culture supernatant was analyzed by atomic absorption spectrophotometer (AAS) (GBC-SensAA, Australia). For As (III) determination, the soil samples were pre-reduced prior to analysis. This was accomplished by the addition of 1 mL of concentrated hydrochloric acid (HCl) and 1 mL of reducing solution (10% (w/v) potassium iodide) to 1 mL of the sample. The reduction rate was improved by increasing the acid concentration. The solution was left to stand for 30 - 45 min at room temperature thus allowing the conversion of $As^{(V)}$ to $As^{(III)}$ in the sample in order to provide increased sensitivity. All solutions were filtered using Whatman No 1 filter paper to remove any fine suspended particulates. Standard As (III) solution was prepared in the range 10-60 mg L^{-1}. The As (III) containing samples were analyzed with an As (III) hollow cathode lamp (190 nm) being used as a light source operated at 12 mA with a 0.7 nm slit width. Pyrolytically coated graphite tubes with Lvov platforms were used with the temperature program for the detection of As (III) as follows: drying at 300 °C (10 s/10 s ramp/holding time), pyrolysis at 1100 °C (10 s/15 s), atomization at 2100 °C (0 s/3 s), and cleaning at 2700 °C (0 s/3 s). Each sample was analyzed in triplicate and the values were presented mean ±SD

Results and discussion

Identification, PCR amplification and Phylogenetic distribution of arsR gene

The plasmid DNA was isolated from As resistant bacteria *E. cloacae* BC2 which has the As resistant *ars* operon [37]. The As transacting regulatory gene *arsR* was amplified using plasmid DNA as a template. Further, the amplified *arsR* gene product shows 353bp length in agarose gel electrophoresis (Fig. 1). The plasmid encoded 353bp length of *arsR* gene from *E. cloacae* was identified and sequenced for evolutionary relationship analysis (Fig. 1). The local alignment of *arsR* gene sequence with NCBI database using BLAST showed 100% homology with *E. cloacae* strain 34399, 98% with *E. cloacae* strain 34978, 96% with *E. cloacae* strain ECNIH5, *E. cloacae* ECR091 and *E. cloacae* ECNIH3. It became evident that *arsR* gene sequence isolated from the *E. cloacae* BC2 has 100% homology with *E. cloacae* strain 34399 that is available with NCBI database. Furthermore, a phylogenetic tree was constructed for *arsR* gene sequences using Neighbour-Joining (NJ) method by MEGA 6.0 software (Fig. 1B). The sequence of plasmid that encodes 353bp length of *arsR* gene was submitted to the NCBI database (accession no KT251200). The result showed that *arsR* gene from *E. cloacae* has highest homology with reported As (III) resistance genes. In concordance with our results many reports suggested that the *ars* operon As (III) detoxification regulated by plasmid encoded *arsR* gene that encodes the trans-acting metalloregulatory protein or regulatory protein ArsR [38,39].

Phylogenetic distribution of ArsR protein

In order to have an enhanced overview of the *arsR* gene encoded ArsR protein distribution in different *E. cloacae*, the seventy four ArsR protein sequences we

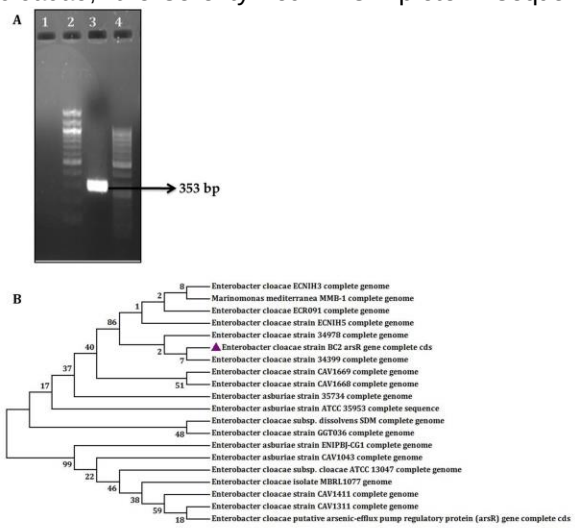


Fig. 1 (A) PCR amplification of *arsR* gene from *E. cloacae:* M – 50bp ladder, Lane 3 *arsR* gene PCR amplified product, (B) Phylogenetic tree of the relatedness of *arsR* gene from *E. cloacae* with other *Enterobacter* sp. The un-rooted tree was constructed from NJ analysis.

-re retrieved and that have homologous amino acid sequences using NCBI and tBLASTn from Uniprot Knowledge database (Supplementary Table S3). The phylogenetic analysis for ArsR protein was performed using MEGA 6.0 software. The value of 90 (highest bootstrap value) in the phylogenetic tree clears the ArsR protein has identical support and by the use of p-distance the computed genetic distances indicates that ArsR from *E. cloacae* BC2 has close relationship among different ArsR protein sequences of *Enterobacter* sp. and other closely related groups of *Enterobacter* sp. Thus, we understood that the gene product ArsR protein is coded by *arsR* of *E. cloacae* BC2 is closely related to the ArsR protein of *Enterobacter* genera. Previous reports indicated that ArsR is a trans-acting regulatory protein that acts as repressor on the *arsRDABC* operon in the absence of As inside the bacteria [40-42]. Thus expression of *arsR* is controlled by As concentration inside the cell [42].

Computational studies reveal the presence cysteine at As binding pocket in N-terminal of ArsR as key factor in As sensing

Studies suggested that MerR family regulatory proteins bind with DNA as a homodimeric form; accordingly, the homodimeric ArsR protein 3D structure was predicted (Fig. 2D). Xu and Rosen [43] demonstrated that As resistance in *E. coli* is conferred through a specific efflux pump controlled by *arsR*. Here, the computational analysis to enhance the understanding of the structural functionality of transacting As regulatory protein ArsR to exploit in As (III) bioremediation. The high specificity of regulatory protein against each individual metal is more important while employing them in industrial application such as detoxification of hazardous heavy metal ions. Hence understanding the physicochemical parameters is important to more precisely engineer the protein into an organism that could potentially be used as either biodegrader or biosensor or both.

In order to gain more knowledge on the ArsR protein, the physicochemical parameters of ArsR protein was analyzed using ProtParam computational tool. In addition, the understanding the physicochemical parameters is important to more precisely engineer the protein into an organism that could potentially be used as either biodegrader or biosensor or both. ArsR protein was predicted to have 106 amino acids with a molecular weight of 12.0741 kDa and the theoretical Isoelectric point (PI) value is 8.70 (Table 1). The instability index (II) was computed to be 44.66 which imply the sequence of ArsR is unstable. The 12 kDa transacting regulatory protein ArsR consists of 14 amino acid residues, among them the major amino acids are leucine (13.3%), arginine (9.3%), serine and alanine (6.7%).Among the total amino acids, the sequence has about twelve (12) negatively charged and fifteen (15) positively charged residues respectively. The subcellular localization of ArsR using CELLO predicted that the target ArsR protein as a cytoplasmic protein and SOSUI result showed that the soluble protein have hydrophobicity of -0.485000. The aliphatic index was calculated as 90.10. The hydropathicity of grand average (GRAVY) was found to be -0.278 which indicates that the possibility of protein has better interaction with water. The DNA binding sites were denoted in Table 1. Likewise, Bose et al. reported that ArsR protein has the conserved DNA binding sites ELCVCDLC found in typical ArsR/SmtB family members [40]. The ArsR protein metal binding sites H3, C30, C32, C35, H48, C89, C106 were identified using MetalDetector v2.0. Similarly, Shi et al. reported that cysteines in the ArsR protein consist of portion of a metal binding motif present in members of the ArsR family of metalloregulatory proteins [44].

Conserved domain database basic local alignment searching tool for transacting regulatory protein result revealed that the target ArsR protein comes under HTH (Helix turn Helix) superfamily ArsR- type DNA-binding domain, Arsenical Resistance operon Repressor and similar prokaryotic, metal regulated homodimeric repressors (Supplementary Fig. S1B). Additionally, two motifs were recognized by the PRINTS server (Supplementary Table 3). The trans-acting regulatory protein contains two fingerprints, such as HTHARSR which has four of four motifs and HTHASNC has two of three motifs. The secondary structural analysis of the protein by PSIPRED indicates the random coil was more frequent

(48.72%) followed by extended strand found to be the least frequent (21.03) and α helix to be 30.26% (Supplementary Fig. S1A).

Table 1 ProtParam results exhibiting physicochemical characters, ArsR of *Enterobacter cloacae*

	ArsR of *Enterobacter cloacae*		
	Amino acids	*Residues*	*%*
Amino acid composition	Ala (A) , Ser (S)	7	6.7%
	Arg (R)	10	9.5%
	Glu (E)	8	7.6%
	Gly (G), Ile (I), Asp (D)	4	3.8%
	His (H), Tyr (Y), Gln (Q)	2	1.9%
	Leu (L)	14	13.3%
	Lys (K), Met (M), Cys (C) , Asn (N),Thr (T)	5	4.8%
	Phe (F)	1	1.0%
	Pro (P), Val (V)	6	5.7%
	Trp (W)	3	2.9%
	Pyl (O), Sec (U)	0	0.0%
Molecular weight	12074.1		
Theoretical Pi	8.70		
Instability Index (II)	44.66		
Aliphatic index	90.10		
Grand Average of hydropathicity (GRAVY)	-0.278		

Putative DNA binding site_238042 was analyzed by Phyre2
Q13,A14,P18,T19,R20,T33,V34,S35,A44,L45,P46,S47,M49,K50,H53,F54,E56,D57,T63,H64,K65,R70

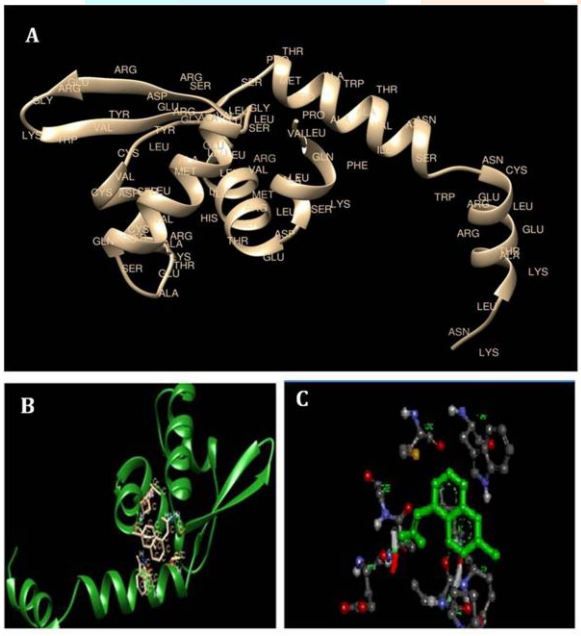


Fig. 2 ***Docking study carried out using Auto Dock to identify the binding potential of ArsR protein with As (III) and As (V).*** (A) The target protein 3D structure of ArsR bound arsenate complex of *E. cloacae* was visualized in UCSF Chimera. (B) Protein-ligand interaction of ArsR and arsenate complexes (C) Enhanced view of As (III) binding to the cysteine residue at the catalytic site, the N-terminal of the protein undergoes a conformational change. D) The target protein 3D dimeric structure of ArsR was visualized in PyMOL.

SOPMA server also used for the secondary structure prediction that showed 47.17% alpha helix (50/106 amino acid residues), followed by random coil 25.47 % (27/106), extended strand 16.04 % (17/106) and beta turn 11.32 % (12/106).The presence of high percentage helix, coils and extended strand indicates that the secondary structure of transacting regulatory protein was successfully predicted. The Ramachandran Plot of the target protein secondary structure showed phi-psi torsion angle for 96.8% (102 out of 106 residues) of residues that have the most favorable region in plot and 2.1% and 1.1% of residues are present in the allowed, and outlier region respectively. These result revealed that the derived structure of ArsR is stereo chemically feasible and has higher quality in terms of protein folding (Supplementary Fig. S1C).

The target protein 3D structure was predicted using RaptorX and visualized using UCSF Chimera. A RaptorX result indicates that the predicted model has better quality in terms of protein folding. These results indicated that the ArsR protein 3D model is stereo chemically good. Chaturvedi and Pandey reported that the RAMPAGE result of As (III) resistance protein (ArsC) model that showed 91.7% of residues are present in the most allowed and favored region [45], therefore the present resultant model in our study with 96.8% present in the most favorable region is feasible for secondary structure prediction and modeling [46,47]. Based on the *P*-value which denotes an index for relative quality and uGDT (unnormalized Global Distance Test), RaptorX assigned the confidence score for target protein 3D structure. Based on QMEAN global scores, the Z-score indicates overall model quality and deviation of the total energy for the structure with respect to an energy distribution derived from random conformation comparison with non-redundant set of PDB structures (Supplementary Fig. S1D). The results indicate that complete refinement and assessment of the predicted model quality was high and reliable (Fig. 2).

The resulted model was docked on to the heavy metals including As (III) to identify the binding potential of homodimeric ArsR protein with As (III). AutoDock result shows possible complexes which were sorted by shape complementarily criteria and the ArsR-As (III) complexes are clustered accordingly. The target protein 3D structure of ArsR (Fig 2A) bound As (V) complex of *E. cloacae* was visualized in UCSF Chimera. ArsR binds with heavy metals through the Cys-32 and Cys-34. Figure 2B and C shows the protein-ligand interaction of ArsR and As (V) complexes. Figure 2D shows the protein-ligand interaction of dimeric ArsR and As (III) complexes. The binding of As (III) with the cysteine residues in the N-terminal indicates that the N-terminal cysteine residue in the ArsR protein plays vital role in the As sensing and regulatory

mechanism in *E. cloacae*. Upon As (III) binding to the cysteine residue, the N-terminal of the protein undergoes a conformational change that can be identified by overlaying the apo-ArsR and holoprotein (As-ArsR complex). We performed docking studies using PatchDock to predict the binding ability of the target ArsR protein with different heavy metals that include arsenite, arsenate, arsenic, zinc, cadmium, copper, lead, cobalt and silver. The PatchDock results analysis (score, area and ACE) reveals that the metal binding with target protein showed higher affinity to As when compared to other metal ions. The result of the PatchDock analysis predictable that the homodimeric ArsR has higher affinity binding to As (III) than As (V) and other heavy metals based on -11.32Kcal/Mol (ACE) with highest binding score 1212 and the area values (141.00). The PatchDock analysis documented that ArsR has binding affinity to As (V) than other heavy metals based on -23.25Kcal/Mol (ACE) with highest binding score 1182 and the area values (127.10). Figure 3 shows the PatchDock values of other heavy metals binding with ArsR. The order of arsR binds with heavy metals are $As^{(III)}$>$As^{(V)}$>Co>Cd>Cr>Pb>Zn>Ag. Thus the identified ArsR protein seems to be a potential candidate to exploit them industrially for both the removal of As from contaminated environment.

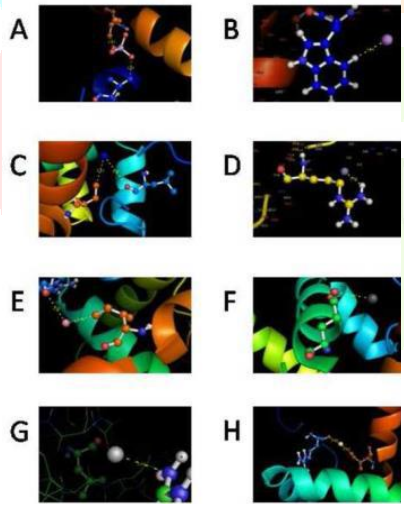


Fig. 3 ***Docking result showing the binding efficiency of arsenite, arsenate, arsenic, chromium, zinc, cobalt, lead, silver and cadmium (ligands) with E. cloacae ArsR protein***. Docking was performed using PatchDock software. The increasing binding efficiency is shown by high score. The high binding area indicated the strong binding of ligand to protein. (A) As (III), (Score = 1212; Area = 141.00; Atomic contact energy = -11.32): (B) As (V), (Score = 1182; Area = 127.10; Atomic contact energy = 23.25): (C) Arsenic (Score = 340 Area = 40.30; Atomic contact energy = 0.00): (D) Chromium (Score = 894; Area = 97.70; atomic contact energy = -11.70): (E) Zinc (Score = 402; Area - 54.20; Atomic contact energy = 0.00): (F) Cobalt (Score = 894;Area = 97.70; Atomic contact energy = -11.70); (G) Lead (Score = 658; Area = 79.30; Atomic contact energy = 10.07); (H) Silver (Score = 402; Area = 54.20; Atomic contact energy = 0.00); and Cadmium (Score = 894; Area = 97.70; Atomic contact energy = -11.70). It can be observed that As (III) binds with ArsR strongly.

As removal efficiency is higher in recombinant E. coli cells bearing pUCP-P_{lac}-arsR compared to wild type E. coli cells

Based on the results obtained from *in silico* approach, the ArsR protein is highly specific to As (III) than any other metal ion thus engineering them as biodegrader is possible. In context of *supra*, the 353bp *arsR* gene segment was inserted into *E. coli.* The copy number was calculated based on the assumption that the average weight of a base pair (bp) is 650 Daltons. The number of template copies was calculated to be ~$2.1x10^{11}$ copies. The successful gene insertion was confirmed by PCR, restriction digestion and gene sequencing analysis. Figure S2 (Supplementary information) depicts the schematic representation of resultant vector *E. coli-arsR*. Now, the *E. coli* cells bearing pUCP-P_{lac}-*arsR* and pUCP18 were used for protein expression analysis with IPTG as an inducer. Figure S3 (Supplementary information) shows the intracellular protein expression pattern of ArsR in *E. coli*. The molecular mass of the ArsR, was found to be 12 kDa on SDS-PAGE that coincides with our earlier prediction (12.0741 kDa using ProtPram) and is concordant with the previous report of Yang et al. [48]

Over-expression of ArsR protein using IPTG enhances the resistance of recombinant bacterial cells towards As (III). The recombinant *E. coli-arsR* exhibited an optimum growth in As (III) containing medium when compared with wild type *E. coli*. The concentration of As (III) in the medium was considerably reduced (70%) by recombinant *E. coli-arsR* when compared to wild type *E. coli* (Fig 4). Thus recombinant *E. coli-arsR* exhibits great potential for As (III) remediation.

Previously, Kostal et al. reported that recombinant *E. coli* harboring ArsR efficiently removed As (III) (100%) from 50 ppb As (III) contaminated water [18]. In our study, on using 50ppb of As (III), the engineered *E. coli* removed approximately 93% of As (III) (data not shown). Kostal et al. observed plateau stage while increasing the As (III) concentration higher than 50ppb, whereas in our study we observed increased efficiency of As (III) removal until 500 ppb; however, after 500 ppb the As (III) removal efficiency substantially decreased and reached 70% at 1000 ppb.

The decreased levels of As (III) removing ability at higher concentrations can be due to increase As toxicity that the recombinant *E. coli-arsR* cells could not withstand. Kostal et al. studied the bioaccumulation of As (III) with recombinant *E.coli* cells using resting cells, whereas in our study we used cultured cells at log phase. The cellular multiplication and cell load might be a possible explanation in cell tolerance and biodegradation ability until 500 ppb but the decrease in the As (III) removal efficiency is suspected to be linked to *arsRDABC* operon and is not properly understood

here, further studies are required to elucidate the exact mechanism. We have reported that the plasmid-borne operon confer increased resistance to toxic metalloid than chromosomal encoded As (III) resistant gene [49,42,37].

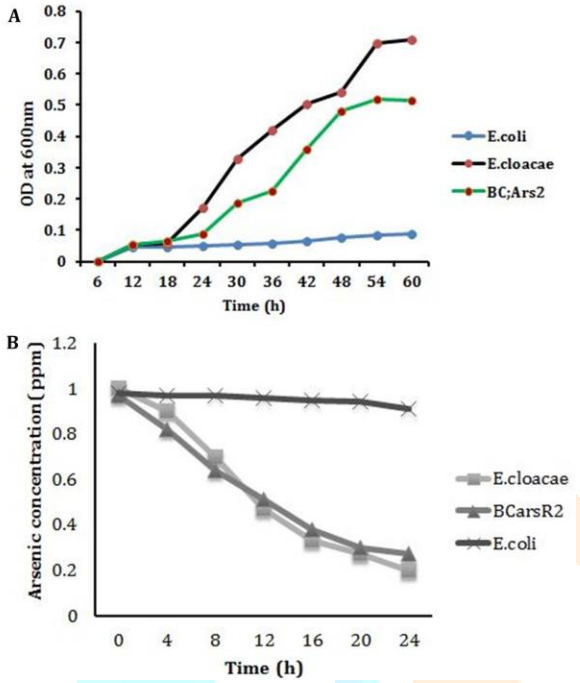


Fig. 4 (A) Growth curve of the *E. coli*, wild type *E. cloacae* and *E. coli-arsR*. Results indicate that *E. cloacae* exhibited an optimum growth in As (III) containing medium when compared with wild type *E. coli*. (B) As removal efficiency of *E. coli*, wild type *E. cloacae* and *E. coli-arsR;* results indicate that the concentration of As (III) in the medium was considerably reduced (70%) by recombinant *E. coli-arsR* when compared to wild type *E. coli*.

Similarly, plasmid-borne operon expressing the *arsRDABC* confers a relatively high resistant to As (III). In the present study, as suspected As resistance in *E. cloacae* is conferred by plasmid encoded *arsR* gene. Yet the intercellular event associated with in the inconsistent behavior of engineered *E. coli-arsR* from *E. cloacae* at higher concentration of As (III) is still under investigation and has not fully understood. We believe identifying the fate of As (III) and As (V) in the cellular environment may shed further light on understanding the As removal mechanism by *E. coli* engineered with *E. cloacae arsR* gene. Nevertheless, the expression of trans-acting regulatory protein in recombinant *E. coli* harboring pUCP-P_{lac}-*arsR* could be a promising strategy to enhance the As (III) removal efficiency and provide a highly selective metal binding target for bioremediation processes and could be potentially exploited in bio-sensing application.

Conclusion

In conclusion the plasmid encoded 353bp length of As (III) resistant *arsR* gene was successfully identified and its protein physicochemical properties as well as 3D structure successfully predicted. Homology modeling and subsequent docking studies indicates that *E. cloacae* ArsR protein has high binding affinity towards As (III) than other metal ions and the N-terminal cysteine residue plays a major role in the As sensing and regulatory mechanism in *E. cloacae* BC2. The comparative physicochemical properties indicated that the functional similarity of ArsR both in *E. cloacae* and *E. coli*. In order to increase the efficiency of bacterial cells for As (III) removal, the *arsR* gene from *E. cloacae* was successfully transformed into *E. coli.* The recombinant *E. coli* harboring pUCP-P_{lac}-*arsR* can effectively remove As (III) (70%) from As (V) medium due to the expression of ArsR. The recombinant *E. coli* harboring pUCP-P_{lac}-*arsR* gene from *E. cloacae* may act as efficient biodegrader in As (III) bioremediation processes and the same can be used in As bio-sensing application upon conjugation with an appropriate transducer.

Acknowledgements This work was financially supported by the University Grant Commission (UGC), New Delhi, India. The authors also thank the central instrumentation facility at SBS, MKU, through CEGS, NRCBS, DST-FIST, DBT-IPLS and DST-PURSE programme.

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

Notes and references

‡Supplementary tables and figures (S1 – S3) are available.

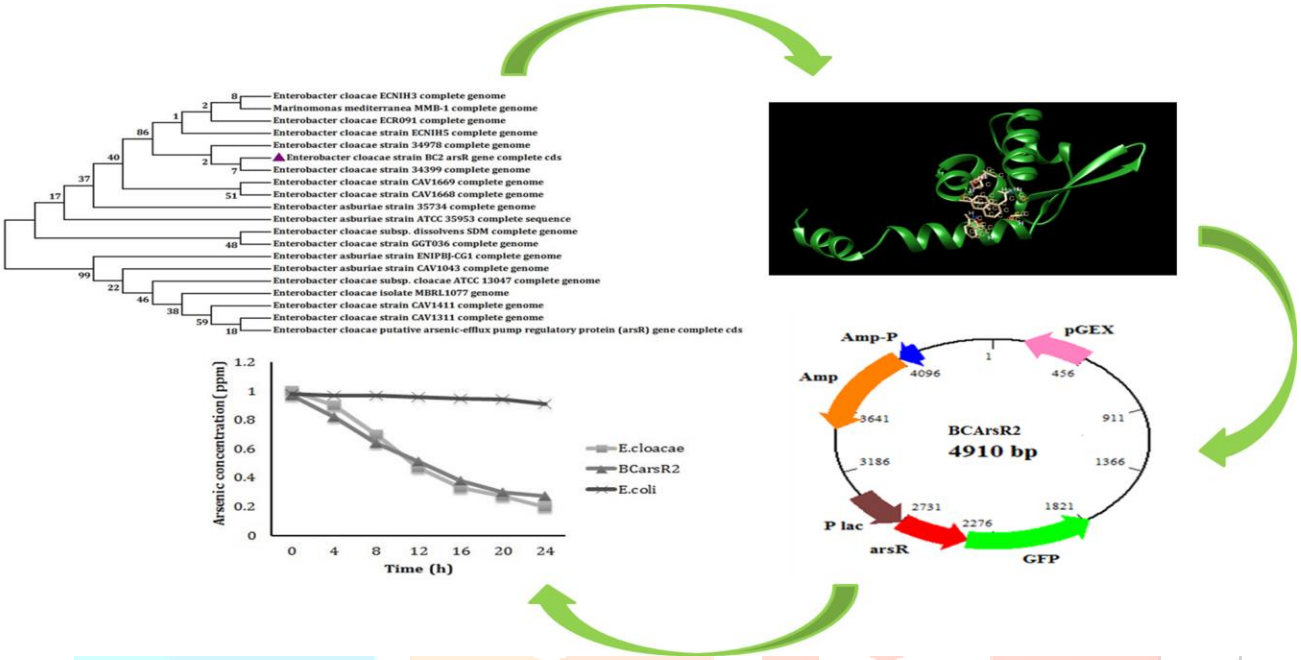
§ Plasmid pUCP18 was donated by Prof. Pradeep Singh at University of Washington, USA and the same was used to clone the *arsR* gene for protein expression.

References

1. Paul S, Bhattacharjee P, Giri AK, Bhattacharjee P (2017) Arsenic toxicity and epimutagenecity: the new LINEage. Biometals : an international journal on the role of metal ions in biology, biochemistry, and medicine. doi:10.1007/s10534-017-0021-2
2. Navas-Acien A, Silbergeld EK, Sharrett R, Calderon-Aranda E, Selvin E, Guallar E (2005) Metals in urine and peripheral arterial disease. Environmental health perspectives 113 (2):164-169
3. Singh P, Dutta SR, Passi D, Bharti J (2017) Benefits of Alcohol on Arsenic Toxicity in Rats. Journal of clinical and diagnostic research : JCDR 11 (1):BF01-BF06. doi:10.7860/JCDR/2017/21700.9146
4. Chowdhury UK, Biswas BK, Chowdhury TR, Samanta G, Mandal BK, Basu GC, Chanda CR, Lodh D, Saha KC, Mukherjee SK, Roy S, Kabir S, Quamruzzaman Q, Chakraborti D (2000) Groundwater arsenic contamination in Bangladesh and West Bengal, India. Environmental health perspectives 108 (5):393-397

5. Zhu YG, Williams PN, Meharg AA (2008) Exposure to inorganic arsenic from rice: a global health issue? Environmental pollution 154 (2):169-171. doi:10.1016/j.envpol.2008.03.015
6. Rebelo FM, Caldas ED (2016) Arsenic, lead, mercury and cadmium: Toxicity, levels in breast milk and the risks for breastfed infants. Environmental research 151:671-688. doi:10.1016/j.envres.2016.08.027
7. Wang Y, Ezemaduka AN, Li Z, Chen Z, Song C (2017) Joint Toxicity of Arsenic, Copper and Glyphosate on Behavior, Reproduction and Heat Shock Protein Response in Caenorhabditis elegans. Bulletin of environmental contamination and toxicology 98 (4):465-471. doi:10.1007/s00128-017-2042-5
8. Pal R, Monroe TO, Palmieri M, Sardiello M, Rodney GG (2014) Rotenone induces neurotoxicity through Rac1-dependent activation of NADPH oxidase in SHSY-5Y cells. FEBS letters 588 (3):472-481. doi:10.1016/j.febslet.2013.12.011
9. Kim M, Lim JW, Kim HJ, Lee SK, Lee SJ, Kim T (2015) Chemostat-like microfluidic platform for highly sensitive detection of heavy metal ions using microbial biosensors. Biosensors & bioelectronics 65:257-264. doi:10.1016/j.bios.2014.10.028
10. Malik AH, Khan ZM, Mahmood Q, Nasreen S, Bhatti ZA (2009) Perspectives of low cost arsenic remediation of drinking water in Pakistan and other countries. Journal of hazardous materials 168 (1):1-12. doi:10.1016/j.jhazmat.2009.02.031
11. Mateos LM, Villadangos AF, de la Rubia AG, Mourenza A, Marcos-Pascual L, Letek M, Pedre B, Messens J, Gil JA (2017) The Arsenic Detoxification System in Corynebacteria: Basis and Application for Bioremediation and Redox Control. Advances in applied microbiology 99:103-137. doi:10.1016/bs.aambs.2017.01.001
12. Dey U, Chatterjee S, Mondal NK (2016) Isolation and characterization of arsenic-resistant bacteria and possible application in bioremediation. Biotechnology reports 10:1-7. doi:10.1016/j.btre.2016.02.002
13. Pepi M, Volterrani M, Renzi M, Marvasi M, Gasperini S, Franchi E, Focardi SE (2007) Arsenic-resistant bacteria isolated from contaminated sediments of the Orbetello Lagoon, Italy, and their characterization. Journal of applied microbiology 103 (6):2299-2308. doi:10.1111/j.1365-2672.2007.03471.x
14. Liao VH, Chu YJ, Su YC, Hsiao SY, Wei CC, Liu CW, Liao CM, Shen WC, Chang FJ (2011) Arsenite-oxidizing and arsenate-reducing bacteria associated with arsenic-rich groundwater in Taiwan. Journal of contaminant hydrology 123 (1-2):20-29. doi:10.1016/j.jconhyd.2010.12.003
15. Cervantes C, Ji G, Ramirez JL, Silver S (1994) Resistance to arsenic compounds in microorganisms. FEMS microbiology reviews 15 (4):355-367
16. Ordonez E, Letek M, Valbuena N, Gil JA, Mateos LM (2005) Analysis of genes involved in arsenic resistance in Corynebacterium glutamicum ATCC 13032. Applied and environmental microbiology 71 (10):6206-6215. doi:10.1128/AEM.71.10.6206-6215.2005
17. Sauge-Merle S, Cuine S, Carrier P, Lecomte-Pradines C, Luu DT, Peltier G (2003) Enhanced toxic metal accumulation in engineered bacterial cells expressing Arabidopsis thaliana phytochelatin synthase. Appl Environ Microbiol 69 (1):490-494
18. Kostal J, Yang R, Wu CH, Mulchandani A, Chen W (2004) Enhanced arsenic accumulation in engineered bacterial cells expressing ArsR. Appl Environ Microbiol 70 (8):4582-4587. doi:10.1128/AEM.70.8.4582-4587.2004
19. Das S, Jean JS, Kar S, Chou ML, Chen CY (2014) Screening of plant growth-promoting traits in arsenic-resistant bacteria isolated from agricultural soil and their potential implication for arsenic bioremediation. Journal of hazardous materials 272:112-120. doi:10.1016/j.jhazmat.2014.03.012
20. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of molecular biology 215 (3):403-410
21. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular biology and evolution 4 (4):406-425
22. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular biology and evolution 4 (4):406-425
23. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular biology and evolution 30 (12):2725-2729. doi:10.1093/molbev/mst197
24. Sigrist CJ, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L, Xenarios I (2013) New and continuing developments at PROSITE. Nucleic acids research 41 (Database issue):D344-347. doi:10.1093/nar/gks1067
25. Yu CS, Lin CJ, Hwang JK (2004) Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. Protein science : a publication of the Protein Society 13 (5):1402-1406. doi:10.1110/ps.03479604
26. Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. Journal of molecular biology 292 (2):195-202. doi:10.1006/jmbi.1999.3091
27. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of molecular biology 215 (3):403-410. doi:10.1016/S0022-2836(05)80360-2
28. Zdobnov EM, Apweiler R (2001) InterProScan--an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17 (9):847-848
29. Attwood TK, Blythe MJ, Flower DR, Gaulton A, Mabey JE, Maudling N, McGregor L, Mitchell AL, Moulton G, Paine K, Scordis P (2002) PRINTS and PRINTS-S shed light on protein ancestry. Nucleic acids research 30 (1):239-241
30. Kallberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, Xu J (2012) Template-based protein structure modeling using the RaptorX web server. Nature protocols 7 (8):1511-1522. doi:10.1038/nprot.2012.085
31. Stocker U, van Gunsteren WF (2000) Molecular dynamics simulation of hen egg white lysozyme: a test of the GROMOS96 force field against nuclear magnetic resonance data. Proteins: Structure, Function, and Bioinformatics 40 (1):145-153
32. DeLano WL (2002) Pymol: An open-source molecular graphics tool. CCP4 Newsletter On Protein Crystallography 40:82-92
33. Lovell SC, Davis IW, Arendall WB, 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins 50 (3):437-450. doi:10.1002/prot.10286
34. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. Journal of computational chemistry 30 (16):2785-2791. doi:10.1002/jcc.21256
35. Schneidman-Duhovny D, Inbar Y, Polak V, Shatsky M, Halperin I, Benyamini H, Barzilai A, Dror O, Haspel N, Nussinov R, Wolfson HJ (2003) Taking geometry to its

edge: fast unbound rigid (and hinge-bent) docking. Proteins 52 (1):107-112. doi:10.1002/prot.10397
36. Peterson AJ, Mallin DR, Francis NJ, Ketel CS, Stamm J, Voeller RK, Kingston RE, Simon JA (2004) Requirement for sex comb on midleg protein interactions in Drosophila polycomb group repression. Genetics 167 (3):1225-1239. doi:10.1534/genetics.104.027474
37. Selvi MS, Sasikumar S, Gomathi S, Rajkumar P, Sasikumar P, Govindan S (2014) Isolation and characterization of arsenic resistant bacteria from agricultural soil, and their potential for arsenic bioremediation. International Journal of Agricultural Policy and Research 2 (11):393-405
38. Silver S, Phung le T (2005) A bacterial view of the periodic table: genes and proteins for toxic inorganic ions. Journal of industrial microbiology & biotechnology 32 (11-12):587-605. doi:10.1007/s10295-005-0019-6
39. Chen PR, He C (2008) Selective recognition of metal ions by metalloregulatory proteins. Current opinion in chemical biology 12 (2):214-221. doi:10.1016/j.cbpa.2007.12.010
40. Bose M, Slick D, Sarto MJ, Murphy P, Roberts D, Roberts J, Barber RD (2006) Identification of SmtB/ArsR cis elements and proteins in archaea using the Prokaryotic InterGenic Exploration Database (PIGED). Archaea 2 (1):39-49
41. Wu J, Rosen BP (1991) The ArsR protein is a trans-acting regulatory protein. Molecular microbiology 5 (6):1331-1336
42. Xu C, Shi W, Rosen BP (1996) The chromosomal arsR gene of Escherichia coli encodes a trans-acting metalloregulatory protein. J Biol Chem 271 (5):2427-2432
43. Xu C, Rosen BP (1997) Dimerization is essential for DNA binding and repression by the ArsR metalloregulatory protein of Escherichia coli. The Journal of biological chemistry 272 (25):15734-15738
44. Shi W, Wu J, Rosen BP (1994) Identification of a putative metal binding site in a new family of metalloregulatory proteins. The Journal of biological chemistry 269 (31):19826-19829
45. Chaturvedi N, Pandey PN (2014) Phylogenetic analysis of gammaproteobacterial arsenate reductase proteins specific to Enterobacteriaceae family, signifying arsenic toxicity. Interdisciplinary sciences, computational life sciences 6 (1):57-62. doi:10.1007/s12539-014-0186-0
46. Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. J Mol Biol 7:95-99
47. Ramachandran GN, Sasisekharan V (1968) Conformation of polypeptides and proteins. Adv Protein Chem 23:283-438
48. Yang T, Chen ML, Liu LH, Wang JH, Dasgupta PK (2012) Iron(III) modification of Bacillus subtilis membranes provides record sorption capacity for arsenic and endows unusual selectivity for As(V). Environmental science & technology 46 (4):2251-2256. doi:10.1021/es204034z
49. 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. Journal of bacteriology 177 (8):2050-2056



Graphical Abstract

Supplementary Files

Table S1. Bacterial strains used in this study

Strains	Description	Source
E. coli DH5α	Host strain for cloning commercial	Novagen (USA)
E. coli BL21	Host strain for expression	Novagen
Enterobacter cloacae (*arsR*)	Agricultural soil isolate	This study

Table S2. Primers used in this study

Primers used	Nucleotide sequence
ArsR F *Hind* III	5′-TAAGCTTATCCAGCTCTTCAAAACC-3′
ArsR R *Kpn*1	5′-GCCATGGGTTTTTCAGCTTCATAC-3′
ArsR F	5′-ATCCAGCTCTTCAAAACC-3′
ArsR R	5′-GTTTTTCAGCTTCATAC-3′

Table S3. Uniprot search results of ArsR sequences of Enterobacter cloacae used for Phylogenetic analysis

S.No	Organism	Uniprot ID	Protein entry Name	Gene Entry Names	Length
1.	*Enterobacter cloacae*	AKQ21142	AKQ21142	BC21	118
2.	*Enterobacter cloacae*	G8LGS2	G8LGS2_ENTCL	*ars*R, EcWSU1_01406	106
3.	*Enterobacter cloacae*	Q8GGH7	Q8GGH7_ENTCL	*ars*R	106
4.	*Enterobacter cloacae*	C1IUR7	C1IUR7_ENTCL	-	120
5.	*Enterobacter cloacae*	V3ITE1	V3ITE1_ENTCL	L402_00005	106
6.	*Enterobacter cloacae*	A0A0E2JYG1	A0A0E2JYG1_ENTCL	L799_00560	99
7.	*Enterobacter cloacae*	A0A0A6GZX0	A0A0A6GZX0_ENTCL	KV26_22800,	110
8.	*Enterobacter cloacae*	A0A0F1PIX5	A0A0F1PIX5_ENTCL	ECNIH4_23540	65
9.	*Enterobacter cloacae*	A0A0A6G526	A0A0A6G526_ENTCL	KV26_34930	231
10.	*Enterobacter cloacae*	E3G5F5	E3G5F5_ENTLS	EntcI_3234	112
11.	*Enterobacter cloacae*	A0A0A6HC19	A0A0A6HC19_ENTCL	KV26_04295	102
12.	*Enterobacter cloacae*	A0A0A7ZLG1	A0A0A7ZLG1_ENTCL	LI64_23860	116
13.	*Enterobacter cloacae*	A0A0F1ITX2	A0A0F1ITX2_ENTCL	SR91_04735	239
14.	*Enterobacter cloacae*	A0A0E1J437	A0A0E1J437_ENTCL	LI62_00855	107
15.	*Enterobacter cloacae*	G2S7G3	G2S7G3_ENTAL	Entas_3378	99
16.	*Enterobacter cloacae*	A0A0E2JXS5	A0A0E2JXS5_ENTCL	L799_03935	219
17.	*Enterobacter cloacae*	V5CTN7	V5CTN7_ENTCL	EDP2_1768	106
18.	*Enterobacter cloacae*	V3E210	V3E210_ENTCL	L423_01203	106
19.	*Enterobacter cloacae*	V3J6V5	V3J6V5_ENTCL	L402_02894	106
20.	*Enterobacter cloacae*	V5AVB5	V5AVB5_ENTCL	EDP2_2616	231
21.	*Enterobacter cloacae*	V5AZT4	V5AZT4_ENTCL	EDP2_2853	103
22.	*Enterobacter cloacae*	A0A0A7ZKT4	A0A0A7ZKT4_ENTCL	LI64_10910	99
23.	*Enterobacter cloacae*	A0A0F3ZBI7	A0A0F3ZBI7_ENTCL	SG66_18075	213
24.	*Enterobacter cloacae*	A0A0F4BRD5	A0A0F4BRD5_ENTCL	SG71_08205,	120
25.	*Enterobacter cloacae*	A0A0E1JE16	A0A0E1JE16_ENTCL	LI62_20540	119
26.	*Enterobacter cloacae*	A0A0A7ZJP1	A0A0A7ZJP1_ENTCL	LI64_00520	107
27.	*Enterobacter cloacae*	A0A0E3A3L3	A0A0E3A3L3_ENTCL	SS28_22585	107
28.	*Enterobacter cloacae*	A0A0F1JEW4	A0A0F1JEW4_ENTCL	SR86_14485	231
29.	*Enterobacter cloacae*	A0A0F4CIX4	A0A0F4CIX4_ENTCL	SG79_06745	231
30.	*Enterobacter cloacae*	A0A0F1CC64	A0A0F1CC64_ENTCL	SG65_00650	239
31.	*Enterobacter cloacae*	G2S3P7	G2S3P7_ENTAL	Entas_0512	231
32.	*Enterobacter cloacae*	G2S530	G2S530_ENTAL	Entas_2250	99
33.	*Enterobacter cloacae*	V3DSA5	V3DSA5_ENTCL	L423_02058	231
34.	*Enterobacter cloacae*	V3HWQ6	V3HWQ6_ENTCL	L402_04040	100
35.	*Enterobacter cloacae*	V3HIV7	V3HIV7_ENTCL	L402_02063	231
36.	*Enterobacter cloacae*	W0BQM4	W0BQM4_ENTCL	M942_23095	231
37.	*Enterobacter cloacae*	W0BNW3	W0BNW3_ENTCL	M942_13575	99
38.	*Enterobacter cloacae*	W1FHN1	W1FHN1_ENTCL	-	219
39.	*Enterobacter cloacae*	W1FKV6	W1FKV6_ENTCL	-	99
40.	*Enterobacter cloacae*	A0A0F3YQI2	A0A0F3YQI2_ENTCL	RZ87_03175	231
41.	*Enterobacter cloacae*	A0A0F4BL70	A0A0F4BL70_ENTCL	SG71_08815	231
42.	*Enterobacter cloacae*	A0A0F0Z2I4	A0A0F0Z2I4_ENTCL	SS42_13470,	231
43.	*Enterobacter cloacae*	A0A0A7ZGQ0	A0A0A7ZGQ0_ENTCL	LI64_02930	231
44.	*Enterobacter cloacae*	A0A0E1JGF6	A0A0E1JGF6_ENTCL	LI62_03405	222
45.	*Enterobacter cloacae*	A0A0F2AY67	A0A0F2AY67_ENTCL	VE21_14900	231
46.	*Enterobacter cloacae*	A0A0F2B413	A0A0F2B413_ENTCL	VE21_08255	104
47.	*Enterobacter cloacae*	A0A0F3ZL49	A0A0F3ZL49_ENTCL	SG66_11660	99
48.	*Enterobacter cloacae*	A0A0F4BQD9	A0A0F4BQD9_ENTCL	SG71_03700	99
49.	*Enterobacter cloacae*	A0A0F3YJ96	A0A0F3YJ96_ENTCL	RZ87_08705	99
50.	*Enterobacter cloacae*	A0A0F4B827	A0A0F4B827_ENTCL	SG76_11170	231
51.	*Enterobacter cloacae*	A0A0G2MDS4	A0A0G2MDS4_ENTCL	ECNIH4_11245	99
52.	*Enterobacter cloacae*	A0A0G2MIA0	A0A0G2MIA0_ENTCL	ECNIH4_19975	233
53.	*Enterobacter cloacae*	W1FHU9	W1FHU9_ENTCL	-	89
54.	*Enterobacter cloacae*	W1FJG5	W1FJG5_ENTCL	-	75
55.	*Enterobacter cloacae*	G8LP23	G8LP23_ENTCL	ygaV, EcWSU1_03480	99
56.	*Enterobacter cloacae*	A0A0E2JLI4	A0A0E2JLI4_ENTCL	L799_23890	116
57.	*Enterobacter cloacae*	A0A0E2JWD7	A0A0E2JWD7_ENTCL	L799_06430	106
58.	*Enterobacter cloacae*	G8LKC4	G8LKC4_ENTCL	ydfF, EcWSU1_00530	231
59.	*Enterobacter cloacae*	G8LCH1	G8LCH1_ENTCL	yczG, EcWSU1_02368	99
60.	*Enterobacter cloacae*	A0A0E2JRP2	A0A0E2JRP2_ENTCL	L799_14450	99
61.	*Enterobacter cloacae*	A0A0A7ZFE5	A0A0A7ZFE5_ENTCL	LI64_00570	116
62.	*Enterobacter cloacae*	A0A0E2K3T5	A0A0E2K3T5_ENTCL	L799_17450	109
63.	*Enterobacter cloacae*	W0BKR0	W0BKR0_ENTCL	M942_17975	106
64.	*Enterobacter cloacae*	V5CQT6	V5CQT6_ENTCL	M942_07145	103
65.	*Enterobacter cloacae*	W0BVF8	W0BVF8_ENTCL	-	99
66.	*Enterobacter cloacae*	W1FEK4	W1FEK4_ENTCL	-	106
67.	*Enterobacter cloacae*	V3DZN3	V3DZN3_ENTCL	L423_00337	434
68.	*Enterobacter cloacae*	A0A0F3Z790	A0A0F3Z790_ENTCL	SG64_20020	117
69.	*Enterobacter cloacae*	A0A0F1P5M2	A0A0F1P5M2_ENTCL	SR86_11580	116
70.	*Enterobacter cloacae*	A0A0E1J453	A0A0E1J453_ENTCL	LI62_00905	116
71.	*Enterobacter cloacae*	A0A0F4C9M3	A0A0F4C9M3_ENTCL	SG79_14130	116
72.	*Enterobacter cloacae*	A0A0E1J7H6	A0A0E1J7H6_ENTCL	LI62_07860	106
73.	*Enterobacter cloacae*	A0A0F4B827	A0A0F4B827_ENTCL	SG76_11170	230
74.	*Enterobacter cloacae*	A0A0A6HCV2	A0A0A6HCV2_ENTCL	KV26_35635	99

Table S4: PRINT results of ArsR

Fingerprint	No. Motifs	Sum Id	Ave Id	Prof Score	P value
HTHARSR	4 of 4	193.19	48.30	1495	1.3e-31
HTHASNC	2 of 3	41.85	20.93	233	0.0011

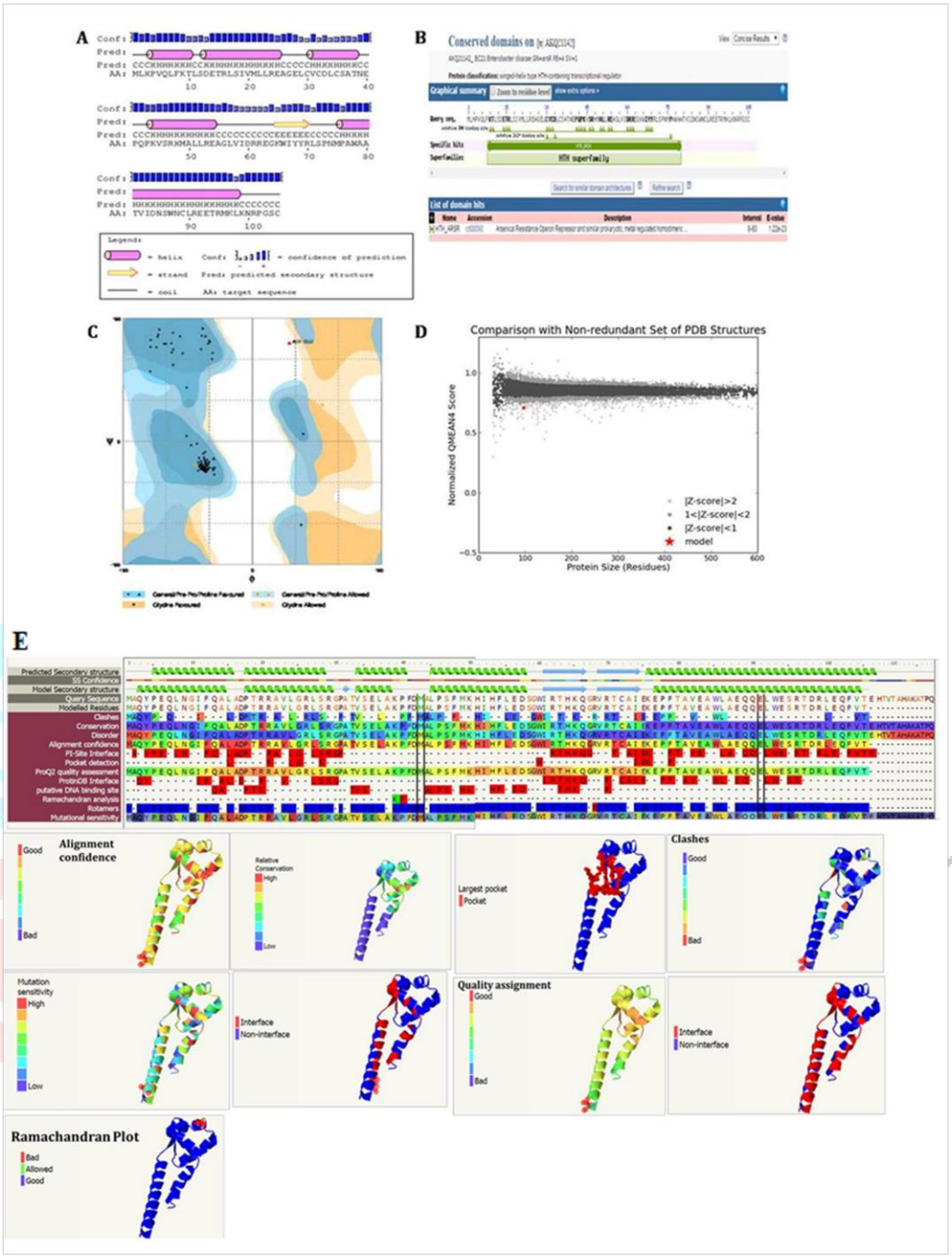


Fig S1. (A) Secondary structure elements of ArsR, showing the alpha helix, beta strand and coils of the amino acid target sequence (B) The CDD BLAST results showing the query protein comes under HTH_ArsR like super family and Arsenical Resistance Operon Repressor and similar prokaryotic, metal regulated homodimeric repressors. ArsR subfamily of helix-turn-helix bacterial transcription regulatory proteins (winged helix topology). Includes several proteins that appear to dissociate from DNA in the presence of metal ions (C) Ramachandran plot displays residues found in blue shade as Glycine favored and allowed regions; residues found in orange shade as Proline favored and allowed regions, (D) Graph showing the comparison with non -redundant set of PDB structures with the protein size and the normalized QMEAN4 score indicates the quality of the model with an indication of the red star (E) Structure prediction by Phyre2 server.

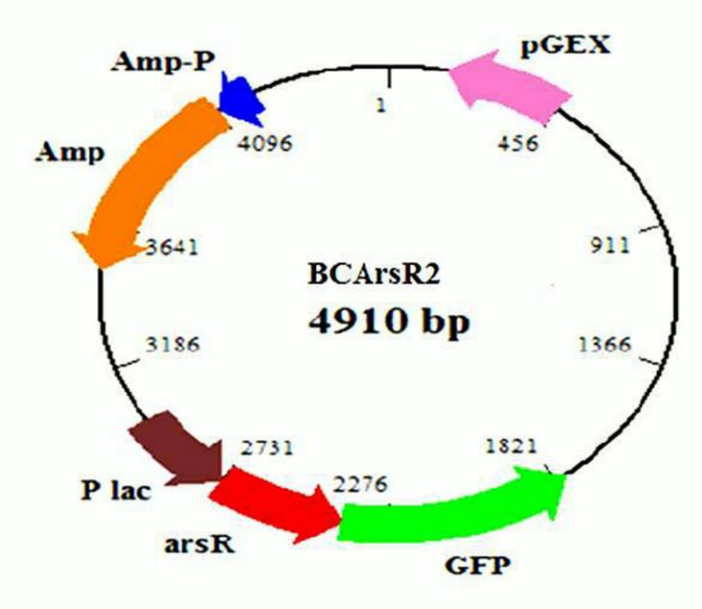


Fig S2. Schematic diagram of the final construct

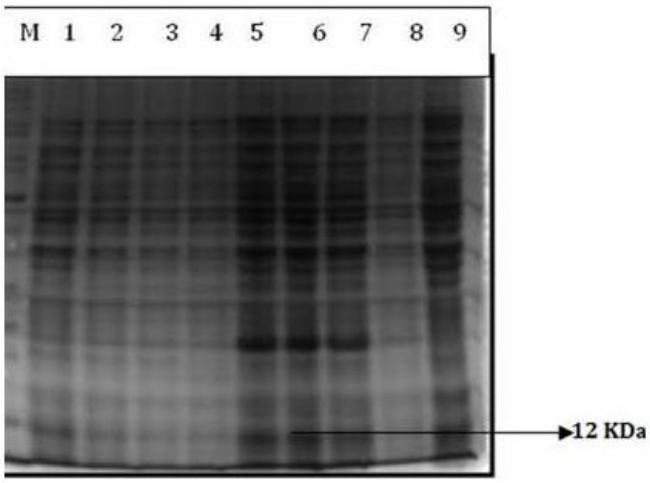


Fig S3. Expression of *arsR* in *E. coli* BC2*arsR*. Lane M protein Ladder (10-200 kDa): Lane 1-7 expression of *arsR* gene encoding for ArsR protein in different concentrations of IPTG- (0.2-1.4mM): Lane 8 –*E. coli* without *arsR* gene; Lane 9: Lysate of *E. cloacae* BC2.