Title page

Protein Sequences Insight into Heavy Metals Tolerance in *Cronobacter sakazakii* BAA-894 encoded by Plasmid pESA3

Authors: Navaneet Chaturvedi

Department of Biochemistry and Molecular Biology, Tel Aviv University, Tel Aviv, Israel.

Michal Kajsik and Stephen Forsythe

Pathogen Research Group, School of Science and Technology, Nottingham Trent University, Clifton Lane, Nottingham, NG11 8NS, UK.

Paras Nath Pandey

Department of Mathematics, University of Allahabad, 211002, India.

***Corresponding author:** Navaneet Chaturvedi,

Department of Biochemistry and Molecular Biology, Tel Aviv University

Tel Aviv, Israel. 69978

Mob: +972-547387142. Email: bioinfonavneet@gmail.com, navaneettau@post.tau.ac.il

Abstract The recently annotated genome of the bacterium *Cronobacter sakazakii* BAA-894 suggests the organism has the ability to bind heavy metals. This study demonstrates heavy metal tolerance in *Cronobacter sakazakii*, in which proteins with the heavy metal interaction were recognized by computational and experimental study. As the result, approximately one fourth of proteins encoded on the plasmid pESA3 are proposed to have potential interaction with heavy metals. Interaction between heavy metals and predicted proteins was further corroborated using protein crystal structures from protein data bank database and comparison of metal-binding ligands. In addition with, a phylogenetic study was undertaken for most toxic heavy metals, like arsenic, cadmium, lead and mercury and obtained related tree pattern for lead, cadmium and arsenic. Laboratory studies confirmed the organism's tolerance to tellurite, copper and silver. These experimental and computational study data extend our understanding of the genes encoding for proteins of this important neonatal pathogen and provides further insights into the genotypes associated with features that can contribute to its persistence in the environment. The information will be of value for future environmental protection from heavy toxic metals.

Keywords Bioremediation; *Cronobacter sakazakii*; Heavy metal protein; Heavy metal resistance; Plasmidborne genes; Toxic heavy metals

Introduction

Bacteria and higher organisms have developed resistance mechanisms to toxic metals. Consequently, phytoremediation and bioremediation processes, using plants and bacteria respectively, are very effective in comparison with other physical and chemical process for heavy metal removal. These processes are inexpensive and efficient at low metal concentrations for enhanced reduction of such contaminants in wastes, sediments or soils contaminated with toxic heavy metals (Pilon-Smits 2005; Valls and de Lorenzo 2002; White et al. 1998). Currently, heavy metal toxicity is one of the most hazardous environmental problems and constitutes a global issue. One means of finding novel solutions is to use bioinformatics to search the genomes of new microorganisms which possess capacity to resist heavy metals. Bacterial genes are mainly chromosomal and to a lesser extent plasmid encoded. However, heavy metal binding phenomenon is reportedly mostly governed by plasmid borne genes and include resistance to many toxic metal ions, including AsO_2^{3-}, $AsO_3^{3-,}$ TeO_3^{2-}, Cd^{2+}, Co^{2+}, Cu^{2+}, Hg^{2+}, Ni^{2+}, Pb^{2+}, Sb^{3+}, TeO_2^{3-}, Ti^{+} and Zn^{2+} (Silver et al. 1996; Silver and Phung 1996). Many workers have studied the effect of heavy metals on microorganisms and described various metal-resistance systems (Nies 1999). Plasmid borne metal resistance and genomic islands have been reported (Mergea et al. 2007).

Cronobacter is member of the family *Enterobacteriaceae* and is ubiquitous in the environment (Iversen & Forsythe 2003). Earlier, phylogenetic analysis, using multilocus sequence analysis, predicts that the *Cronobacter* genus split from its closest ancestor in the *Enterobacteriaceae* family approximately 45-68 million years ago (MYA) with *C. sakazakii*

further emerging ~15-23 MYA (Joseph et al. 2012). The bacterium is tolerant of many environmental stresses such as prolonged desiccation and heat (Osaili and Forsythe 2009). However metal resistance in this organism has not previously been investigated. Recently the whole genome was released for *C. sakazakii* strain BAA-894 (Kucerova et al. 2010). This included the partial annotation of the plasmid pESA3 (131 kb, 127 genes) with predicted arsenic resistance genes, as well as copper, silver and tellurium resistance genes encoded on the chromosome (Kucerova et al. 2010, 2011). The copper, silver and tellurium resistance genes are homologs of the genes found on plasmids in other *Enterobacteriaceae*; IncII plasmid R478 of *Serratia marcescens*, pK29 of *Klebsiella pneumoniae* NK29, and pAPEC-O1-R of *E. coli* APEC O1 (Gilmour et al. 2004).

Functionally and structurally important regions in a protein family are well conserved across species. Consequently, the detection of functional residues in proteins is important in functional annotation (George et al. 2005; Livingston and Barton 1996; Ouzounis et al. 1998; Pupko et al. 2002). In recent years, the computational studies have been used to predict metal binding motifs and associated genes, along with the binding motif responsible for metals and/or heavy metals interactions (Thilakaraj et al. 2007).

The objective of study was to identify putative proteins for heavy metal binding which were encoded on the plasmid pESA3 of *C. sakazakii* BAA-894 in addition with, a evolutionary study assists for tree pattern detection to most toxic heavy metals. Finally, confirmatory laboratory studies of heavy metals tolerance by *C. sakazakii* BAA-894 was determined. This study also reveals putative heavy metals binding interactions of metal ions such as cadmium, cobalt, iron, zinc, copper, arsenic, mercury, manganese and nickel.

Materials and Methods

Dataset

All 127 pESA3 protein sequences (Accession number NC_009780.1) were considered in this study. These were downloaded in FASTA format from the Genbank database of NCBI (Benson et al. 2000).

Heavy Metal Binding Prediction

Metal ion binding protein sequences for cadmium, lead, cobalt, zinc, arsenic, iron, copper, mercury, manganese and nickel ions were analyzed using the PROSITE tool (Falquet 2002; Nicolas et al. 2004; Sigrist et al. 2002 & 2010) for sequence similarity searching against Swissprot/TrEMBL (Bairoch et al. 2004; Magrane 2011). Hits for all PROSITE motifs on each sequence were obtained and the binding domains were manually inspected. Each heavy metal's binding hits were recorded by assessing the binding statistics associated with full binding description of organic and inorganic legends including heavy metals. These binding statistics record were obtained from the PDBeMotif search tool. PDBeMotif can be used to examine the characteristics of the binding sites of single proteins or classes of proteins either within the same species or across different species (Golovin et al. 2005, 2008 & 2009). Moreover, metal binding sites contain cysteine, histidine, asparatate and glutamate residues because most of the motifs have been designed around these conserved residues (Thilakaraj et al. 2007; Zhang et al. 2000). Therefore amino acid compositions frequencies were also taken into consideration to know the percentage of metal binding amino acids. Furthermore,

the metal ion binding motifs were checked by performing a scan using the ScanProsite tool (De Castro et al. 2006) against the Swiss-Prot/TrEMBL database. Homologous crystal structure for the identified amino acid sequences were investigated by comparision with available PDB database (Bernstein et al. 1977) using BLASTp search (Altschul et al. 1990)**,** to show similarities in the protein crystal structure. Structure knowledge is essential for all areas of protein research such as enzyme kinetics, ligand–protein binding studies, gene characterization and construction, structure based molecule design, and rational designing of proteins. Therefore this concept was used for the validation of potential metal ion binding to proteins encoded on pESA3 in *C. sakazakii* BAA-894.

Phylogenetic Analysis of Protein Sequences for Most Toxic Heavy Metals

The most toxic heavy metals term was applied to cadmium, mercury, lead and arsenic, which are listed by the World Health Organization's as chemicals of major public concern (Brathwaite , Rabone, 1985). In the periodic table the metalloid arsenic (As) is placed in group V and is thus classified as a most toxic heavy metal (Wackett, Dodge, Ellis , 2004). Sequences for all the most heavy toxic metals were recruited from Table 2 and imported into MEGA5 (Tamura K., et al, 2011) for phylogenetic analysis. The Maximum Parsimony (MP) method was used to generate phylogeny tree.

Demonstration of Tellurium, Copper and Silver Metal Tolerance

The whole genome sequenced strain *C. sakazakii* BAA-894 was studied for heavy metal resistance and has been described in earlier publications (Kucerova et al. 2010 & 2011). It is in the *Cronobacter* clonal group sequence type 1 (ST1) as determined using the seven allele *Cronobacter* multilocus sequence typing scheme (Forsythe, S.J., et al. 2014). Further strain details are accessible from the open access *Cronobacter* database hosted by the University of Oxford (UK): www.pubMLST.org/cronobacter.

C. sakazakii BAA-894 was grown overnight in tryptone soya broth at 37°C. An aliquot (200µl) was then added to 5ml fresh Luria broth and shaken for 2.5 hours at 37°C, 200 rpm. The optical density of the culture was measured at 600nm, and diluted to a standard of OD_{600} = 0.5. Then 400µl of the diluted culture was added to 10 ml Top agar (5g tryptone and 7g agar/l water) at 42°C and poured on TSA plates. After solidification, a sterile 1cm diameter filter paper disc was aseptically placed on the agar surface and 7 µl of heavy metal solution added; silver nitrate, copper sulphate and sodium tellurite (10 to 100 mM). The plates were incubated overnight in 37°C, before measuring the radius of the zone of inhibition (ZOI). Plates were then re-incubated for a further 24 hours to observe the formation of black colonies at the zone of inhibition for tellurite (IV).

Results

Experimental Demonstration of Heavy Metal Tolerance

C. sakazakii BAA-894 was resistant to silver nitrate and copper sulphate up to 1mM and 10mM, respectively. The organism was also resistant to tellurite up to 1mM. During

incubation, the bacterial growth at the edge of the zone of inhibition with tellurite (IV) became black which is characteristic of elemental tellurium metal formation.

Putative Heavy Metals Prediction

Although most open reading frames in *C. sakazakii* plasmid pESA3 are annotated, most (98.42%, 125/127) are not characterized and are only given as hypothetical proteins. Almost 26%, (33/127) of these hypothetical proteins have predicted metal interactions. Gene-locus-tag, matched PROSITE-id, matched domain name, putative metal ions and hits obtained through scanprosite are listed in Table 1. Each of these sequences showed one or more metal ion binding site. The binding statistics for each Prosite domain hit for every sequence revealed that only arsenic ion would be bound with Van der Waal interaction, whereas other heavy metals would have covalent bond interactions. The frequency histogram was also represented for highly responsible metal binding amino acid residues (Fig. 1). In this case, aspartate and glutamate residues inferred better interaction with proteins.

All 33 predicted sequences were validated by searching for the most similar protein crystal structure using BLASTp search against the PDB database. The corresponding metal ion bind with the protein crystal structures was collected and most of the predictions were correct. In addition to metal ions, Ca^{2+} and Mg^{2+} ion interactions were also considered. The matching crystal structure and metal ion for each of the 33 predicted sequences are listed in Table 1.

Phylogenetic Analysis of Protein Sequences for Most Toxic Heavy Metals

Mercury, cadmium, lead and arsenic metal-binding protein sequences were collated for phylogenetic analysis. The analysis used 12 amino acid sequences and the consensus tree inferred from 2 most parsimonious trees is shown (Fig 2). Branches corresponding to partitions reproduced in less than 50% trees are collapsed. All positions containing gaps and missing data were eliminated. There were a total of 103 positions in the final dataset. It was possible to analyze the result by clade level phylogeny. Consequently, 3 to 4 clades were obtained after tree construction for 12 amino acid sequences. Although, 3 clades were highlighted by red circle on the basis of similarity (Fig 2). The most interesting fact was observed in study that the Clade A demonstrated most related pattern for lead. Three lead binding putative sequences out of four were revealed evolutionary similarity. Arsenic and cadmium mediated amino acid sequences exhibited similar clustering in clades B and C. Although, mercury mediated sequences were also demonstrated relatedness with clade C with 100% consensus value. One major single branch clad was distantly related which corresponds to two cadmium and one for mercury as well as lead.

Discussion

C. sakazakii is a ubiquitous organism in the environment (Joseph et al. 2012; Kucerova et al. 2011). This bacterium is tolerant of many stresses such as prolonged desiccation and heat (Osaili and Forsythe 2009). However metal resistance has not previously been investigated. In our study, *C. sakazakii* BAA-894 was shown to tolerate silver nitrate and copper sulphate to 1mM and 10mM. Tellurite resistance was notable by the reduction of tellurite (IV) to the

elemental form tellurium which accumulated as a black deposit. The predicted tellurite resistance genes are located on the chromosome of this strain at loci ESA_01775-01804 (Kucerova et al. 2010). These genes are homologs of well characterised genes carried on plasmids in other *Enterobacteriaceae* (Gilmour et al. 2004). This warrants further investigation as the mechanisms of tolerance and relevance of genomic location are still poorly understood (Chasteen et al. 2009). *C. sakazakii* BAA-894 carries a large plasmid (pESA3, 131 kb) which encodes 127 genes which have not been described in any detail in the literature (Kucerova et al. 2010). Our study has shown that more than one fourth of these genes encoded proteins performed heavy metal binding interactions. All metal binding sequences were accessed from UniprotKB/Swissprot database (Bairoch et al. 2004; Magrane 2011) and indicated that they are all specific to the protein family. Specific heavy metal binding interaction with protein sequences is helpful for protein functional annotation and structural study (George et al. 2005; Livingston and Barton 1996; Ouzounis et al. 1998; Pupko et al. 2002). The validation result shows metal ion binding interaction with similar protein crystal structure of identified plasmid pESA3 amino acid sequences. Most toxic heavy metals were recruited as cadmium, mercury, lead and arsenic, all of which the World Health Organization list chemicals of major public concern (Brathwaite, Rabone, 1985). Most toxic heavy metal corresponding proteins phylogeny generated robust trees aid interpreting biological data with reference to heavy metal tolerance. Furthermore, an attempt was made to explain to understand the pattern for binding with most toxic heavy metals. Very recently, it was hypothesized that member of the *Enterobacteriaceae* family shows similar pattern for arsenic reduction (Chaturvedi N, Pandey PN., 2014) and is more common

in this family compared to other members of the gammaproteobacteria. In addition, this study revealed the extent of heavy metal binding proteins on the larger plasmid of *C. sakazakii*, which could be further used to address public health concerns for exposure to heavy metals.

Conclusion

The genomic analysis of the large plasmid in *C. sakazakii* has revealed possibilities for further environmental care and bioremediation strategy with respect to heavy metal toxicity. The organism was shown to be tolerant of copper and silver and can reduce tellurite (IV) to the elemental form. Moreover the detailed metal binding affinities of identified proteins will support future studies. The ability to predict metal-binding proteins can accelerate the development of more efficient bioremediation, biosorption, bioaccumulation and many other environmental protection strategies.

Acknowledgements

The authors are thankful for the support of the DBT-BIF facility in Center of Bioinformatics, University of Allahabad, India. NC acknowledges a PhD fellowship from UGC, New Delhi, India.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, (1990) Basic local alignment search tool. J Mol Biol 215: 403–410.

Bairoch A, Boeckmann B, Ferro S, Gasteiger E, (2004) Swiss-Prot: juggling between evolution and stability. Brief Bioinform 5: 39-55.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Rapp BA, Wheeler DL, (2000) GenBank Nucleic Acids Res 28:15-18.

Bernstein FC, Koetzle TF, Williams GJB, Meyer Jr EF, Brice MD, Rogers JR, Kennard O, Shimanouchi T, Tasumi M, (1977) The Protein Data Bank: a computer-based archival file for macromolecular structures. J Mol Biol 112:535-542.

Brathwaite RL, Rabone SDC (1985). "Heavy Metal Sulphide Deposits and Geochemical Surveys for Heavy Metals in New Zealand". Journal of the Royal Society of New Zealand 15 (4): 363–370

Chasteen TG, Fuentes DE, Tantalean JC, Vasquez CC, (2009). Tellurite: history, oxidative stress, and molecular mechanisms of resistance. FEMS Microbiol Rev 33: 820–832.

De Castro E, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, Hulo N (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res 34: W362-5.

Chaturvedi N, Pandey PN., (2014), Phylogenetic analysis of gammaproteobacterial arsenate reductase proteins specific to Enterobacteriaceae family, signifying arsenic toxicity, Interdiscip Sci. 2014 Mar;6(1):57-62

Falquet L, Pagni M, Bucher P, Nicolas, Hulo, Sigrist, CJA, Hofmann K, Bairoch A, (2002) The PROSITE database, its status in 2002. Nucleic Acids Res 30: 235–238.

Forsythe, S.J., Dickins, B., & Jolley, K.A. (2014) Cronobacter, the emergent bacterial pathogen Enterobacter sakazakii comes of age; MLST and whole genome sequence analysis. BMC Genomics 15:1121

George RA, Spriggs, RV, Bartlett GJ, Gutteridge A, MacArthur MW, Porter CT, Al-Lazikani B, Thornton JM, Swindells MB, (2005) Effective function annotation through catalytic residue conservation. Proc Natl Acad Sci USA 102: 12299–12304.

Gilmour MW, Thompson NR, Sanders M, Parkhill J, Taylor D (2004) The complete nucleotide sequence of the resistance plasmid R478: defining the backbone components of incompatibility group H conjugative plasmids through comparative genomics. Plasmid 52: 182-202.

Golovin A, Henrick K (2008) MSDmotif: exploring protein sites and motifs. BMC Bioinformatics 9: 312.

Golovin A, Henrick K (2009) Chemical substructure search in SQL. J Chem Inf Model 49: 22-7.

Golovin A, Dimitropoulos D, Oldfield T, Rachedi A, Henrick K (2005) MSDsite: A database search and retrieval system for the analysis and viewing of bound ligands and active sites. Proteins 58: 190-9.

Iversen C, Forsythe S (2003) Risk profile of Enterobacter sakazakii, an emergent pathogen associated with infant milk formula. Trends in Food Science and Technol 14: 443-454.

Joseph S, Forsythe SJ (2011) Predominance of Cronobacter sakazakii ST4 with neonatal infections. Emerg Infect Dis 17: 1713-1715.

Joseph S, Sonbol H, Hariri S, Desai P, McClelland M, Forsythe SJ (2012) Diversity of the Cronobacter genus as revealed by multi locus sequence typing. J Clin Microbiol 50: 3031-3039.

Kucerova E, Clifton SW, Xia XQ, Long F, Porwollik S, et al (2010) Genome sequence of Cronobacter sakazakii BAA-894 and comparative genomic hybridization analysis with other Cronobacter species. PLoS ONE 5, e9556

Kucerova E, Joseph S, Forsythe S (2011) Cronobacter: diversity and ubiquity. Quality Assurance and Safety of Foods and Crops 3: 104-122.

Livingstone CD, Barton GJ (1996) Identification of functional residues and secondary structure from protein multiple sequence alignment. Meth Enzymol 266: 497–512.

Magrane M, the UniProt consortium (2011) UniProt Knowledgebase: a hub of integrated protein data, Database, bar009.

Mergea M (2007) Plasmids pMOL28 and pMOL30 of Cupriavidus metallidurans are specialized in the maximal viable response to heavy metals. J Bacteriol 189: 7417-7425.

Nicolas H, Sigrist CJA, Le Saux V, Langendijk-Genevaux S, Bordoli L, Gattiker A, De Castro E, Bucher P, Bairoch A (2004) Recent improvements to the PROSITE database. Nucleic Acids Res. 32(Database issue), D134–D137.

Nies DH (1999) Microbial heavy-metal resistance. Appl Microbiol Biotechnol 51: 730-735.
Osaili T, Forsythe S (2009) Desiccation resistance and persistence of Cronobacter species in infant formula. Intl J Food Microbiol 136: 214-220.

Ouzounis C, Perez-Irratxeta C, Sander C, Valencia A (1998) Are binding residues conserved? Pac Symp Biocomput 401–412.

Pilon-Smits E (2005) Phytoremediation. Ann Rev Plant Biol 56: 15–39.

Pupko T, Bell RE, Mayrose I, Glaser F, Ben-Tal N (2002) Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. Bioinformatics 18: S71–S77.

Sigrist CJA, Cerutti L, Hulo N, Gattiker A, Falquet L, Pagni M, Bairoch A, Bucher P (2002) PROSITE: a documented database using patterns and profiles as motif descriptors. Brief Bioinform 3: 265-274.

Sigrist CJA, Cerutti L, de Castro E, Langendijk-Genevaux PS, Bulliard V, Bairoch A, Hulo N (2010) PROSITE, a protein domain database for functional characterization and annotation. Nucleic Acids Res 38(Database issue) 161-166.

Silver S (1996) Bacterial resistance to toxic metal ions-a review. Gene 179: 9-19.

Silver S, Phung LT (1996) Bacterial heavy metal resistance: new surprises. Ann Rev Microbiol 50: 753-89.

Tamura K., Peterson D., Peterson N., Stecher G., Nei M., and Kumar S. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution

Thilakaraj R, Raghunathan K, Anishetty S, Pennathur G (2007) In silico identification of putative metal binding motifs. Bioinformatics 23: 267–271

Valls M, de Lorenzo V (2002) Exploiting the genetic and biochemical capacities of bacteria for the remediation of heavy metal pollution. FEMS Microbiol Rev 26: 327–338.

Wackett LP, Dodge AG, Ellis LBM (2004) Microbial genomics and the periodic table. Appl Environ Microbiol 70:647-655

White C, Sharman AK, Gadd GM (1998) An integrated microbial process for the bioremediation of soil contaminated with toxic metals. Nat Biotechnol 16: 572–575.

Zhang W, Colman RW (2000) Conserved amino acids in metal-binding motifs of PDE3A are involved in substrate and inhibitor binding. Blood 95: 3380-6.

Table
Click here to download Table: Table.doc

Table 1 Gene-locus-tag with Protein-ID and predicted interaction with corresponding heavy metals of pESA3 of C. sakazakii BAA-894.

Serial number	Gene Locus-Tag	Matched Prosite-ID	Domain name	Putative binding heavy metals	Number of Prosite Scan Hits against UniprotKb/ TrEMBL
1	ESA_pESA3p05431	PS51186	Gcn5-related N acetyltransferase (GNAT) domain	Ni^{2+} Zn^{2+} AsO_3^{3-}	682
2	ESA_pESA3p05442	PS50109	Histidine kinase domain	Hg^{2+} Mn^{2+}	493
3	ESA_pESA3p05446	PS51462	Nudix hydrolase domain signatures	Mn^{+} Gd^{3+} Zn^{2+}	960
4	ESA_pESA3p05448	PS50977	TetR-type HTH domain signature	Ni^{2+}	343
5	ESA_pESA3p05450	PS50850	Major facilitator superfamily (MFS)	Hg^{2+}	1001
6	ESA_pESA3p05454	PS50977	TetR-type HTH domain signature	Ni^{2+}	343
7	ESA_pESA3p05462	PS50883	EAL domain	AsO_3^{3-} Cu^{2+} Hg^{2+} Mn^{2+}	36
8	ESA_pESA3p05463	PS50937	MerR-type HTH domain signature	Zn^{2+}	75
9	ESA_pESA3p05464	PS50111	Bacterial chemotaxis sensory transducers signature	Pb^{2+}	51
10	ESA_pESA3p05466	PS51186	Gcn5Gcn5-related N acetyltransferase (GNAT) domain	Ni^{2+} AsO_3^{3-} Zn^{2+}	682
11	ESA_pESA3p05470	PS50987	ArsR-type HTH domain signature	Zn^{2+}	41
12	ESA_pESA3p05472	PS51098	PTS EIIB domain profiles and cysteine phosphorylation site signature	Zn^{2+}	108
13	ESA_pESA3p05474	PS50949	GntR-type HTH domain	Cd^{2+} Zn^{2+}	
14	ESA_pESA3p05485	PS51353	Arsenate reductase arsC family	AsO_3^{3-}	59
15	ESA_pESA3p05487	PS50987	ArsR-type HTH domain signature	Zn^{2+}	41
16	ESA_pESA3p05490	PS50949	GntR-type HTH domain	Cd^{2+} Zn^{2+}	217
17	ESA_pESA3p05497	PS00871	Chaperonins clpA/B signatures	Pt^{2+}	72
18	ESA_pESA3p05498	PS00142	Neutral zinc metallopeptidases, zinc binding region signature	Zn^{2+} Co^{3+} Cd^{2+} Cu^{2+}	1008

				Fe^{3+} Hg^{2+} Mn^{2+} Ni^{2+} Pb^{2+}	
19	ESA_pESA3p05503	PS51257	Prokaryotic membrane lipoprotein lipid attachment site	Cd^{2+} Fe^{3+} Zn^{2+}	1000
20	ESA_pESA3p05508	PS51352	Thioredoxin family active site signature and domain	Zn^{2+} Cd^{2+} Cu^{2+} Ni^{2+} AsO_3^{3-}	1045
21	ESA_pESA3p05516	PS50893	ATP-binding cassette, ABC transporter-type, signature	Zn^{2+} Hg^{2+} Cd^{2+} Mn^{2+}	1298
22	ESA_pESA3p05518	PS50983	Iron siderophore/cobalamin periplasmic-binding domain	Fe^{3+} Cd^{2+} Zn^{2+}	62
23	ESA_pESA3p05519	PS50850	Major facilitator superfamily (MFS)	Hg^{2+}	1001
24	ESA_pESA3p05527	PS50111	Bacterial chemotaxis sensory transducers signature	Pb^{2+}	51
25	ESA_pESA3p05530	PS50404	Soluble glutathione S-transferase N- and C-terminal domain	Fe^{3+} Cd^{2+} Cu^{2+} Zn^{2+}	350
26	ESA_pESA3p05532	PS50111	Bacterial chemotaxis sensory transducers signature	Pb^{2+}	51
27	ESA_pESA3p05536	PS00154	P-type ATPases phosphorylation site	Al^{3+}	621
28	ESA_pESA3p05537	PS51184	JmjN and JmjC domains	Fe^{3+} Ni^{2+} Zn^{2+} Mn^{2+} Hg^{2+}	294
29	ESA_pESA3p05541	PS00909	Mandelate racemase / muconate lactonizing enzyme family signatures	Mn^{2+} Ba^{2+}	64
30	ESA_pESA3p05544	PS50943	Cro/C1-type HTH domain profile	Hg^{2+} Zn^{2+} Cd^{2+}	247
31	ESA_pESA3p05545	PS51384	Ferredoxin reductase-type FAD-binding domain	Zn^{2+}	658
32	ESA_pESA3p05546	PS50850	Major facilitator superfamily (MFS)	Hg^{2+}	1001
33	ESA_pESA3p05551	PS01156	TonB-dependent	Fe^{3+}	76

			receptor proteins signatures		

Table 2 Plasmid pESA3 protein sequences showing similarity with protein crystal structure and binding with potential metal(s) ion and percentage of identity and query coverage. PDB identifiers corresponding for every protein were highlighted in bold.

Serial No.	Gene-locus-tag	(PDB-ID) Similarity with Protein Crystal Structure	PDB % Identity	Query Coverage	Metals ion
1	ESA_pESA3p05431	**(3RL5)** Rat metallophosphodiesterase MPPED2 H67R Mutant	34	33	Ca^{2+}
2	ESA_pESA3p05442	**(3A0R)** Crystal structure of histidine kinase ThkA (TM1359) in complex with response regulator protein TrrA (TM1360)	26	52	Hg^{2+}
3	ESA_pESA3p05446	**(2QJT)** Crystal structure of a bifunctional NMN adenylyltransferase/ADP ribose pyrophosphatase complexed with AMP and MN ion from *Francisella tularensis*	40	45	Mn^{2+}
4	ESA_pESA3p05448	**(3B81)** Crystal structure of predicted DNA-binding transcriptional regulator of TetR/AcrR family	30	60	Na^{+}
5	ESA_pESA3p05450	**(2VLI)** Structure of *Deinococcus radiodurans* tunicamycin resistance protein	33	72	Cd^{2+}
6	ESA_pESA3p05454	**(2GUH)** Crystal Structure Of The Putative Tetr-Family Transcriptional Regulator	33	36	Mg^{2+}
7	ESA_pESA3p05462	**(3KZP)** Crystal structure of putative diguanylate cyclase/phosphodiesterase from *Listeria monocytigenes*	26	53	AsO_3^{3-}
8	ESA_pESA3p05463	**(1EXI)** Crystal Structure of Transcription Activator Bmrr, from *B. subtilis*, bound to 21 Base Pair Bmr Operator And Tpsb	24	28	Zn^{2+}, Sb^{3+}
9	ESA_pESA3p05464	**(2CH7)** Crystal Structure Of The Cytoplasmic Domain Of A Bacterial Chemoreceptor From *Thermotoga maritima*	31	35	Pb^{2+}
10	ESA_pESA3p05466	**(2Q0Y)** Crystal structure of GCN5-related N-acetyltransferase (YP_295895.1) from *Ralstonia eutropha* JMP134 at 1.80 A resolution	26	75	Zn^{2+}
11	ESA_pESA3p05470	**(2JSC)** NMR structure of the cadmium metal-sensor CMTR from *Mycobacterium tuberculosis*	36	35	Cd^{2+}
12	ESA_pESA3p05472	**(3BP8)** Crystal structure of Mlc/EIIB complex	39	31	Zn^{2+}
13	ESA_pESA3p05474	**(2OOI)** The crystal structure of gene product SA0254 from *Staphylococcus aureus* subsp. *aureus* N315	25	45	Zn^{2+}
14	ESA_pESA3p05485	**(1I9D)** Arsenate reductase from *E. coli*	75	99	AsO_3^{3-}

15	ESA_pESA3p05487	**(2JSC)** NMR structure of the cadmium metal-sensor CMTR from *Mycobacterium tuberculosis*	33	74	Cd^{2+}
16	ESA_pESA3p05490	**(2ZC0)** Crystal structure of an archaeal alanine:glyoxylate aminotransferase	28	66	Zn^{2+}
17	ESA_pESA3p05497	**(1QVR)** Crystal structure analysis of ClpB.	45	73	Pt^{2+}
18	ESA_pESA3p05498	**(4HMY)** Structural Basis For Recruitment And Activation Of The Ap-1 Clathrin Adaptor Complex By Arf1	29	54	Mg^{2+},
19	ESA_pESA3p05503	**(3LW5)** Phers From Staphylococcus Haemolyticus- Rational Protein Engineering And Inhibitor Studies	47	35	Fe^{3+}
20	ESA_pESA3p05508	**(2FWE)** Crystal structure of the C-terminal domain of the electron transfer catalyst DsbD (oxidized form)	36	11	Ni^{2+}
21	ESA_pESA3p05516	**(3C41)** ABC protein ArtP in complex with AMP-PNP/Mg^{2+}, **(3G5U)** Structure of P-glycoprotein Reveals a Molecular Basis for Poly-Specific Drug Binding	28 32	91 83	Mg^{2+} , Hg^{+2}
22	ESA_pESA3p05518	**(2R7A)** Crystal structure of a periplasmic heme binding protein from *Shigella dysenteriae*	27	46	Fe^{3+}
23	ESA_pESA3p05519	**(1UP8)** Recombinant vanadium-dependent bromoperoxidase from red algae C*orallina pilulifera*	32	17	Ca^{2+}
24	ESA_pESA3p05527	**(2CH7)** Crystal structure of the cytoplasmic domain of a bacterial chemoreceptor from *Thermotoga maritima*	27	46	Pb^{2+}
25	ESA_pESA3p05530	(**4IEL**) Crystal Structure Of A Glutathione S-Transferase Family Protein From Burkholderia Ambifaria, Target Efi-507141, With Bound Glutathione	30	78	Mg^{2+}
26	ESA_pESA3p05532	**(2CH7)** Crystal structure of the cytoplasmic domain of a bacterial chemoreceptor from *Thermotoga maritima*	36	81	Pb^{2+}
27	ESA_pESA3p05533	**(3B8E)** Crystal structure of the sodium-potassium pump	79	28	Mg^{2+}
28	ESA_pESA3p05537	**(2XDV)** Crystal structure of the catalytic domain of flj14393	62	26	Cd^{2+}
29	ESA_pESA3p05541	**(1WZM)** *Thermoactinomyces vulgaris* R-47 alpha-amylase II (TVA II) mutatnt R469K	38	32	Ca^{2+}
30	ESA_pESA3p05544	**(3GN5)** Structure of the *E. coli* protein MqsA (YgiT/b3021)	38	50	Zn^{2+}
31	ESA_pESA3p05545	**(2GPJ)** Crystal structure of a	29	94	Ca^{2+}

		siderophore-interacting protein (sputcn32_0076) from *shewanella putrefaciens* cn-32 at 2.20 a resolution			
32	ESA_pESA3p05546	**(2H39)** Crystal structure of an ADP-Glucose Phosphorylase from *Arabidopsis thaliana* with bound ADP-Glucose	27	17	Zn^{2+}
33	ESA_pESA3p05551	**(3FHH)** Crystal structure of the heme/hemoglobin outer membrane transporter ShuA from *Shigella dysenteriae*	27	59	Pb^{2+}

Figure
Click here to download Figure: Figure.doc

Figure. 1 Frequency histogram of specific metal binding residues percentage for entire 33 sequences with predicted interaction with corresponding heavy metals, encoded on pESA3 of *C. sakazakii* BAA-894.

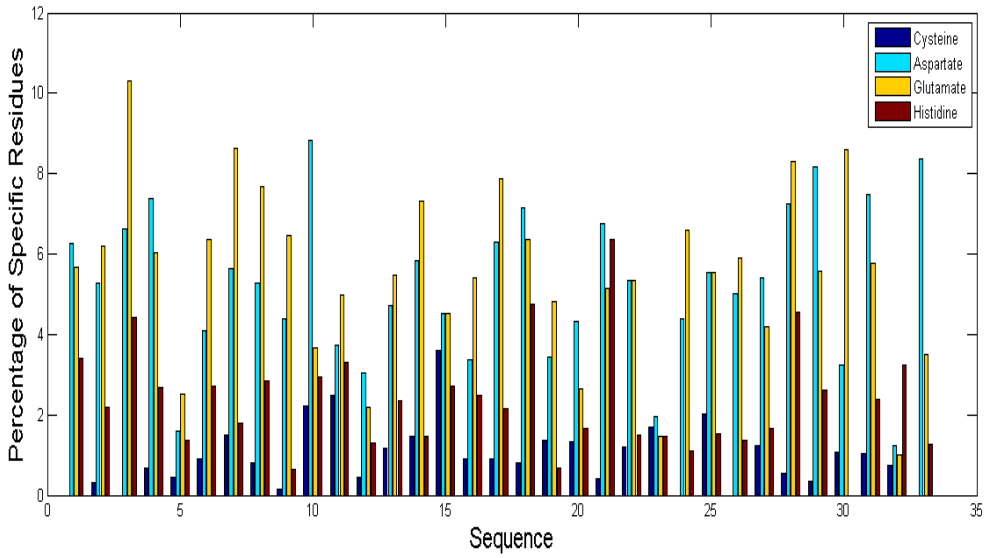

Fig 2: Fig: 2. Representation of evolutionary tree of amino acid sequences encoded on the plasmid pESA3 which may interact with mercury, cadmium, lead and arsenic. All nodes are named as gene locus tag along with corresponding metals. Red circles (A, B and C) indicate most related clades corresponding to the respective metal, showing significant similarity for lead as well as cadmium and arsenic.