International Biodeterioration & Biodegradation 122 (2017) 82–91

Contents lists available at ScienceDirect

International Biodeterioration & Biodegradation

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

In silico approach for bioremediation of arsenic by structure prediction and docking studies of arsenite oxidase from *Pseudomonas stutzeri TS44*

Mymoona Akhter [a, b, *], Munazzah Tasleem [a], Mohammad Mumtaz Alam [a, b], Shakir Ali [a, c, *]

[a] Bioinformatics Infrastructure Facility, Jamia Hamdard, Hamdard Nagar, New Delhi 110062, India
[b] Drug Design and Medicinal Chemistry Lab, School of Pharmaceutical Education and Research, Jamia Hamdard, Hamdard Nagar, New Delhi 110062, India
[c] Department of Biochemistry, School of Chemical and Life Sciences, Jamia Hamdard, New Delhi 110062, India

ARTICLE INFO

Article history:
Received 31 October 2016
Received in revised form
21 March 2017
Accepted 24 April 2017

Keywords:
Homology modeling
3D structure
Arsenite-oxidase
P. stutzeri TS44
Intramolecular interaction
Bioremediation
Arsenite
Arsenate
Arg421

ABSTRACT

Pseudomonas stutzeri TS44 is a moderately halotolerant, arsenite-oxidizing bacterium, containing genes for arsenite oxidation, arsenic resistance, and ectoine/hydroxyectoine biosynthesis. This paper reports *in silico* studies to understand bioremediation to eliminate toxic metal arsenic in water, air and soil by arsenite oxidase (AO), the bacterial enzyme from *P. stutzeri* TS44 that can be used for a low cost and eco-friendly removal of arsenite from the environment. To understand the activity of AO in elimination of arsenite, sequence analysis was carried out and homologs, orthologs, domains, family, and conserved residues were identified followed by model generation using various homology modeling tools. The generated models were validated for the best quality protein structure and the best model was used for further optimization using energy minimization approach. Molecular docking studies were performed to study the binding interaction of AO with arsenite. The study predicts and validates the 3D structure of *P. stutzeri* TS44 arsenite oxidase and reports four active site residues (His197, Glu205, Arg421, and His425) from a close structural homolog of AO from *A. faecalis* (PDB ID: 1G8K_A). The molecular docking studies suggested the formation of a stable complex and *in silico* site-directed mutagenesis revealed the importance of Arg421, which resulted in a decrease in stability of the complex when mutated. The study implicates *P. stutzeri* TS44 arsenite oxidase as a non virulent protein for low cost and eco-friendly bioremediation of arsenite.

1. Introduction

Industrial growth has led to the increased discharge of industrial waste to the environment leading to heavy metal contamination, a major threat to human health and environment. Heavy metal contamination in soil and water mainly occurs from various anthropogenic sources. Some heavy metal ions are essential but most of them are toxic even at low concentration. Heavy metal treatment releases non-biodegradable compounds and produces sludge, thus it cannot be completely degraded or modified (Nies, 1999). Heavy metals like arsenic, chromium, cadmium, copper, mercury and lead etc, are cytotoxic, carcinogenic and mutagenic in nature even at low concentration levels (Salem and Farag, 2000). Arsenic is a metalloid, widely disseminated in the environment with an abundance of 0.0001% and exists specifically in two toxic forms including trivalent and pentavalent arsenic (Takei and Tsujimoto, 1998). Arsenite is discharged into the environment by various natural activities (Bahar et al., 2012; Ruchita Dixit et al., 2015). Arsenite (AsO_2^-) is capable of binding to sulfhydryl groups of proteins and dithiols, thereby inducing toxicity. The pentavalent form, arsenate ($As^VO_4^-$) is, however, less toxic, and a structural analog of phosphate that produces arsenylated derivatives to inhibit phosphorylation process. Arsenic concentration in water bodies ranges from 5 to 5000 μg L^{-1}. Permissible limit of arsenic in drinking water is 10 μg L^{-1}.

Due to the high magnitude solubility of arsenic in water, it is removed using conventional treatment methods. An alternative to

* Corresponding authors. Bioinformatics Infrastructure Facility, Jamia Hamdard (Hamdard University), Hamdard Nagar, New Delhi 110062, India.
E-mail addresses: jhuniv.btisnet@nic.in (M. Akhter), jhuniv.btisnet@nic.in (S. Ali).

http://dx.doi.org/10.1016/j.ibiod.2017.04.021
0964-8305/

conventional treatment of water is the biological treatment based on biological oxidation that transforms arsenite to arsenate with a cost effective and eco-friendly approach (Ahluwalia, 2007; Anderson et al., 1992). Microorganisms have developed high resistance to arsenic compounds, as seen in autotrophic bacteria, where As(III) act as the electron donor and oxygen as an electron acceptor during arsenite oxidation, while in heterotrophic bacteria, detoxification mechanism is activated by arsenite oxidase enzyme (Pepi et al., 2007; Tiancai Feng et al., 2014). Although bioremediation is a promising approach for removal and recovery of heavy metals from polluted soil and water, microorganisms are reported to develop resistance to heavy metals - they adopt various detoxifying mechanisms such as bioaccumulation, biosorption, biotransformation and biomineralization (Gadd, 2000; Lim et al., 2003; Lin, 2005). Therefore, the aim of this study was to identify arsenic-resistant bacteria from arsenic-contaminated soil and analyze the enzyme involved in bioremediation of arsenic using *in silico* approach. Arsenic-resistant *Pseudomonas* spp., which is found in polluted soil and water, is reported to possess remarkable adaptability in diverse environmental conditions (Pepi et al., 2007). *Pseudomonas stutzeri* is a Gram-negative, motile, rod-shaped bacterium which possesses metabolic diversity, and is widely distributed in the natural environment (Lalucat et al., 2006; Zhuangpeng Huang and Yongyou, 2016). To date five genome sequences for *P. stutzeri* have been reported, whereas one of the member, *P. stutzeri TS44* was isolated from highly arsenic-contaminated soil (Cai et al., 2009). The strain *P. stutzeri TS44*, which displays an arsenite-oxidizing rate of 59.1 μMh^{-1}, can be considered as a potential candidate for bioremediation of arsenic (Cai L et al., 2009). Eectoine/hydroxyectoine biosynthetic genes (ectABCD-ask) have been identified in *P. stutzeri* which helps in its survival under high-osmolarity conditions (Li et al., 2012).

In this study, *arsenite oxidase* (AO) from *P. stutzeri TS44* was analyzed to identify its family, domain, motifs and evolutionary relationships. The structure of AO from *P. stutzeri TS44* has not been reported yet, and in its absence, homology models built from close structural homologs could be an ideal starting point for understanding binding interaction with As(III). The models were generated and validated to ensure good quality protein structure and structural integrity using various structure analysis tools. Models were further refined in terms of energy minimization to remove steric clashes. Based on the sequence and structural homology, the active site and binding pocket were identified, and the key residues involved in ligand binding were determined. The best model thus obtained was selected for docking studies to analyze various important and conserved active site residues and interaction patterns crucial to understand both the substrate's mode of action and the catalytic mechanism that contribute to binding with As(III). Mutational studies were carried out to understand the significance of the predicted active-site residues in the generated model. Mutants were determined for the arsenite oxidase and mutant three-dimensional models were generated and validated. The best generated model was subjected to docking with As(III) for interaction study. The study provides insights into the ligand binding process as well as structural features of the ligands responsible for biological.

2. Materials and methods

The computational framework used for sequence annotation given in Fig. 1, is divided into five phases namely, Phase I, II, III, IV, and V. The Phase I includes the sequence retrieval, characterization, sequence comparison, function prediction and analysis for the presence of virulence factor. Phase II comprises of reference template recognition for model generation and its evaluation using various bioinformatics tools. In phase III, docking studies were executed for understanding the interaction between arsenite and AO. Phase IV comprised of mutant residue identification, mutant model generation and docking studies of the mutant model and As(III). Phase V includes identification of closed interaction between the heavy metal ligand and the receptor (generated wild type and mutant model).

2.1. Sequence analysis

Protein sequences of AO from *P. stutzeri TS44* were downloaded in FASTA format using NCBI, the online data retrieval system - National Center for Biotechnology Information (Dooley, 2004). Expasy's ProtParam server (Gasteiger et al., 2003), was extensively used to measure physiochemical properties of AO protein sequences such as molecular weight, isoelectric point, extinction coefficient (Gill and von Hippel, 1989), aliphatic index (Ikai, 1980), instability index (Guruprasad et al., 1990) and grand average of hydropathicity (GRAVY) (Kyte and Doolittle, 1982).

Identification of virulence factor in protein is an important step for pointing out the virulent protein that could be a target for the development of antimicrobial drugs. Here, a non-virulent protein is required that can be used to oxidize arsenite to arsenate. Virulence in the protein sequences was predicted using two robust tools: Virulentpred (Garg and Gupta, 2008) and VICMpred (Saha and Raghava, 2006). These tools predict bacterial virulence factors based on support vector machine (SVM) methods with high accuracy, of which VICMpred is dedicated specifically for identification of functions of gram-negative bacterial proteins.

Sequence similarity search was performed for the identified protein sequences to determine homologous sequences, family, domains and thereby predict functions. Sequence comparison was performed using PSI-BLAST search against the non-redundant (nr) protein sequence database to identify the homologs (Friedberg et al., 2000). Sequences with low query coverage (<50%) and low sequence similarity (<65%) were excluded for further analysis. Protein sequences showing highest value for the respective parameters were considered as close homologs for deriving probable functions. To identify close structural homologs, PSI-BLAST search against Protein Data Bank (PDB) database was carried out.

Various bioinformatics tools were adopted for precise assignment of functions to the proteins. Families of the protein sequence was identified using Pfam (Finn et al., 2016) and CATH (Sillitoe et al., 2015). For predicting domains, CDD (Derbyshire et al., 2015), CDART (Geer et al., 2002) and SMART (Letunic et al., 2015) online tools were used. CDD has a collection of multiple sequence alignment of ancient domains and full-length proteins that are used to identify conserved domains. CDART searches for similar protein using protein domain profiles and group proteins similar to the query sequence and assigns a score based on their architecture; whereas, SMART is a Simple Modular Architecture Research Tool that annotates and explore protein domain architecture. The results obtained from the above tools were analyzed to derive consensus, which was considered together with the results computed using ProFunc tool, for functional assignment.

Motifs in protein sequence are considered as signature of protein families, assisting in the prediction of function. Motifs play an important role in enzymes as they are associated with catalytic functions. Identification of motifs in query sequence was executed by the InterProscan tool, which is composed of various highly developed methods for protein signature recognition including Pfam, PANTHER, Prosite, ProDom, PRINTS, SUPERFAMILY etc. Other tools used for motif searching are MEME that uses expectation maximizing technique to fit a two component finite mixture model to the set of sequences and MotifFinder.