J Appl Biotechnol Rep. 2022 September;9(3):707-718

Journal of
Applied Biotechnology Reports

doi 10.30491/JABR.2021.302174.1437

Original Article

# Homology Modelling and Docking Studies of Arsenate Reductase of *Bacillus megaterium*

**Mst. Nusrat Jahan Arbi [1], Naima Thahsin [1], Azmery Nurjahan [1], Sabrina Sharmin [2], Md. Zahidul Islam [1*], Umme Salma Zohora [1], Mastura Khatun Ruma [1], Ripa Moni [1], Mohammad Shahedur Rahman [1]**

[1] Department of Biotechnology and Genetic Engineering, Jahangirnagar University, Savar, Dhaka-1342, Bangladesh
[2] Department of Pharmacy, Brac University, 66 Mohakhali, Dhaka 1212, Bangladesh

**Corresponding Author:** Md. Zahidul Islam, PhD, Associate Professor, Department of Biotechnology and Genetic Engineering, Jahangirnagar University, Savar, Dhaka-1342, Bangladesh. Tel: +88-01787-086965, E-mail: zahidul@juniv.edu

**Received** August 29, 2021; **Accepted** December 6, 2021; **Online Published** September 10, 2022

**Abstract**

**Introduction:** *Bacillus megaterium* is a ubiquitous bacterial strain that produces the enzyme arsenate reductase that catalyzes the reduction of less toxic arsenate (V) to more toxic arsenite (III). Due to the functional significance of this enzyme, the present study was carried out to construct and validate the Three Dimensional (3D) structure of arsenate reductase of *B. megaterium* and study its interaction with arsenate.

**Materials and Methods:** The 3D model was generated by MODELLER using the known crystal structure of the enzyme. The superimposition of the model with the template structures was done by PyMOL. PATCHDOCK was used to perform molecular docking of the enzyme with arsenate ion and Fire Dock was used to refining the docked complexes. The highest geometric score containing docked complex was visualized and the intra-molecular interaction within it was evaluated.

**Results:** The evaluation of the 3D computed model showed good qualities including fine stereochemical properties, satisfactory compatibility between the structure and its amino acid sequence, acceptable residues error value, etc. The model as well as its phylogenetic relatives (*Bacillus* and *Staphylococcus*) showed the same active site motif which is CTGNSCRS. The crucial amino acids involved in binding with the arsenate ion ($AsO_4^{3-}$) were Cys10, Thr11, Gly12, Asn13, Ser14, Cys15, His43, and Asp106. Among these, the first six amino acids fell in the conservative motif (CTGNSCRS).

**Conclusions:** Studying this interaction can be helpful for more research to inhibit the binding or any other approach that will stop the inimical conversion of less toxic arsenate to more toxic arsenite.

**Keywords:** *Bacillus megaterium*, Arsenate Reductase, Arsenate, Homology Modeling, Molecular Docking
**Citation:** Arbi MNJ, Thahsin N, Nurjahan A, Sharmin S, Islam MZ, Zohora US, et al Homology Modelling and Docking Studies of Arsenate Reductase of *Bacillus megaterium*. J Appl Biotechnol Rep. 2022;9(3):707-18. doi:10.30491/JABR.2021.302174.1437

## Introduction

Many natural phenomena and industrial growth have led to the heavy metal contamination, a real threat to the all living beings and environment. Predominantly, the geogenic origin and also anthropogenic sources cause the metal contamination in water and soil. Some metals are important and play a great role in human body. However, some are toxic even at low concentration and can lead to chronic effects. Among the heavy metals like chromium, cadmium, copper, mercury, lead etc., arsenic is one of the major causes of many health and environmental issues. It can cause cytotoxicity, carcinogenicity and mutagenicity in nature at low concentrations.[1] Arsenic is a metalloid and its two toxic ubiquitous forms include trivalent (arsenite) and pentavalent (arsenate) arsenic.[2] Arsenic can also be available in an inorganic form which is also toxic. Arsine gas also has toxicity and may cause hemolysis. However, in the biological system the conversion of arsenate to arsenite happens and both of their chronic exposure bring detrimental effects.[3] The *ars* operon in the system contains *ArsC* gene that encodes arsenate reductase enzyme. This enzyme reduces arsenate to arsenite.[4] For this conversion, the enzyme binds to the arsenate ion available in the circumstances.[5] The other one that also encodes this enzyme is known as *Arr* (arsenate respiratory reductase). This gene is mainly available in anaerobic bacteria which respires utilizing the arsenate as the electron acceptor. This enzyme binds to the arsenate ion ($AsO_4^{3-}$) and reduces it to arsenite ($AsO_3^{3-}$).[6] Arsenite ($AsO_2$) which is more toxic than arsenate, disrupts sulfhydryl groups of proteins, dithiols and interferes with enzyme function whereas arsenate ($AsVO_4$) acts as a phosphate analog and can interfere with phosphate uptake and transport.[7-9] Millions of people are exposed to arsenic containing water in a daily basis due to its environmental extensiveness and are colossally affected because of higher arsenic concentrations in natural aquifers.[10] People of South Asia subcontinent particularly Bangladesh, Myanmar, India, Nepal and Pakistan are mostly affected.[11] The problem caused by arsenic has become so detrimental that a vast study on this, is a very crucial demand of time. Many physical,

chemical and biological methods have been used to decrease the concentration of arsenic in contaminated water as its full eradication is quite impossible.[12] However, physical and chemical treatments are very costly and they associate a plethora of fatalistic impacts on the environment. Therefore, bioremediation can be a great way to deal with it.[13] The aim of bioremediation in this case is to convert the more toxic arsenite to less toxic arsenate. Many species of bacteria can transform arsenite to arsenate in anoxic conditions by using arsenite as an electron donor.[11] But the matter of concern is that arsenic is mostly found as an arsenate form in the surface water and is up taken by most of the organism. Into the system, it is converted to arsenite and this more toxic form is then released to the environment. Arsenate reductase is the enzyme that is responsible for the less toxic arsenate (V) conversion to more toxic arsenite (III) form in the nature.[14,15] To counter this detrimental conversion process, this enzyme should be focused on. The presence of this enzyme in different bacteria like *Escherichia coli, Staphylococcus aureus, Chrysiogenes arsenates, Bacillus subtillis, B. megaterium* etc. are confirmed. In *B. subtilis,* the arsenate reductase is coupled with the thioredoxin and thioredoxin reductase to catalyze the reduction of arsenate to arsenite.[16,17] The enzyme of *B. subtilis* has a conserved motif CTGNSCR (residues 10–16) which performs as an arsenate binding loop.[16] Among these residues, Cys10 is a key residue which performs the initial nucleophilic attack on arsenate.[16,17] Though the crystallographic structure of arsenate reductase of *B. subtilis* is available, but the structure of the enzyme of *B. megaterium*, another arsenic reactive bacteria, has not been deposited yet in the Protein Data Bank (PDB). In the absence of experimental structures, it is necessary to use the computational methods to predict the 3D protein structure of arsenate reductase enzyme. Through a computational study of this enzyme and its binding to the ligand arsenate ion, a depth insight into its structure, characteristics and function can be theorized. In recent days, many researchers have used computational tools to model and analyze the arsenate reductase enzyme of different bacterial species.[18,19] In this research, amino acid sequence of arsenate reductase from *B. megaterium* was used to execute *in silico* study. This bacteria is a gram positive, rod shaped and endospore forming bacteria.[20] Unlike many other bacteria, its genome is fully coded and *ArsC* gene is present in this bacteria which expresses the arsenate reductase protein.[21,22] The completed genomes allow access to whole genome DNA arrays and high throughput proteomics.[23] It is an important enzyme of functional significance as it catalyzes the conversion of arsenate to arsenite in *B. megaterium*.[24] This study was carried out in order to use the computational methods to predict the 3D structure of arsenate reductase (ArsC) of *B. megaterium* and its docking with the arsenate ion to get a comprehension of the structure and its function. To reverse it or to know the clear mechanism it is a prerequisite to know the structure and characteristics of the enzyme that is responsible for this conversion.

## Materials and Methods

### *Retrieval and Analysis of Query Sequence*

The protein sequence of arsenate reductase of *B. megaterium* was retrieved in FASTA format from the NCBI database (National Centre for Biotechnology Information), an online data retrieval system (http://www.ncbi.nlm.nih.gov/). To appraise the physicochemical properties of the query sequence such as molecular weight, extinction coefficient, isoelectrical point, atomic composition, instability index aliphatic index, grand average of hydrophobicity (GRAVY) etc., the Expasy's ProtParam server was comprehensively used.[25-29] Pfam and CATH were used to identify the families of the protein sequence.[30,31] To predict the domains, various online tools such as CDD, CDART and SMART, were used.[32,33] A multitude of multiple sequence alignment of full-length proteins is reserved in CDD which was used to identify the conserved domains. In protein sequence, motifs are regarded as manifestation of protein families. Moreover, these play a key role in enzymes as they are associated with catalytic function. Motifs of the query sequence were identified using Motif Finder (https://www.genome.jp/tools/motif/MOTIF.html).

### *Multiple Sequence Alignment and Phylogenetic Analysis*

Multiple Sequence Alignment (MSA) of arsenate reductase enzyme was performed among the sequences from different strains of *B. megaterium* and also among the homologs from different species to detect conserved or functionally important regions. The sequences of the enzyme from different strains of *B. megaterium* was collected from NCBI database. On the other hand, the homologous sequences of the target enzyme were searched against UniProtKB/Swiss-Prot with the help of Basic Local Alignment Search Tool (BLAST).[34] Among the sequences from the BLAST search, those with higher sequence identity and query coverage were selected for sequence alignment. The multiple sequence alignment was executed by Molecular Evolutionary Genetics Analysis (MEGA) version 6.0 which accomplished the alignment using the neighbor-joining method in the program Multiple Sequence Comparison by Log-Expectation (MUSCLE).[35] Then, the MEGA was used to construct phylogenetic tree in the aim of investigating the evolutionary relationship of the query sequence with its homologous sequences. In MEGA, the important parameters which were set to build the phylogenetic tree were neighbor-joining method, 1000 bootstrap replications and the Poisson model.

### *Secondary Structure Prediction*

To identify the features of proteins and to recognize the folds, secondary structure prediction is a crucial step. It also

acts as an intermediate step for a 3D protein structure prediction. The PSIPRED tool was used to predict the secondary structure of the query sequence. This tool uses position specific scoring matrix produced by PSI-BLAST based on neural network methods for secondary structure prediction.[36]

### *Homology Modeling for 3D Structure Generation*

Homology modeling is to construct an atomic-resolution model of the "target" protein using its amino acid sequence and an experimental three-dimensional structure of a related homologous protein which is called template. In the present study, homology modeling was implemented with the help of the MODELLER version 9.19 to get the 3D structure of arsenate reductase of *B. megaterium*. At first, the query sequence was searched for similar sequence identity using BLAST against PDB to identify the corresponding templates.[34] The maximum percentage of sequence identity along with query coverage and E-value was considered to select the homologous templates from the BLAST search result. The structures of the selected templates were retrieved from the PDB (https://www.rcsb.org/). Then, the alignment file of the query sequence and the python script files containing necessary instructions to run MODELLER were generated. Multiple sequence alignment was then performed between the template sequences and the query sequence and on the basis of the alignment; the 3D model of the query sequence was built. The MODELLER was instructed to build five probable models. The best model among them was selected based on its DOPE score. Energy minimization was carried out for the best model by Swiss-PdbViewer.[37]

### *Validation of the Predicted 3D Model*

The 3D model of arsenate reductase constructed by MODELLER was verified using several *in silico* tools such as PROCHECK, Verify 3D, ERRAT and PyMOL. PROCHECK was used to generate and evaluate the Ramachandran plot for the model of target enzyme.[38] This plot allowed visualizing the distribution of all the residues of the enzyme into four particular regions which were the most favored regions, additional allowed regions, generously allowed regions and disallowed regions. The distribution of each residue was based on its backbone dihedral angles ψ against φ. Verify 3D helped to determine the compatibility of the 3D structure of arsenate reductase with its own amino acid sequence (1D) by comparing the 3D-1D profile of the query enzyme to good structures.[39] The ERRAT analyzed the statistics of non-bonded interactions between different atom types and calculated the overall quality factor of the protein model.[40] Then, the 3D structure of arsenate reductase was compared with the structure of one template protein by superimposition of both the structures. The template with the highest sequence similarity was chosen for this purpose. The superimposition of the structures and the calculation of RMSD (root mean square deviation) between them were executed by PyMOL (https://pymol.org/2/).

### *Molecular Docking*

The method of molecular docking was implemented to predict the preferred orientation of binding of arsenate ion to arsenate reductase enzyme to form a stable complex structure. To perform docking, the 2D structure of arsenate ion was retrieved in SDF format from PubChem (https://pubchem.ncbi.nlm.nih.gov/). The molecule of arsenate ion was then optimized using Gaussian software (www.gaussian.com) applying the universal force field (UFF) of molecular mechanics method and converted to PDB format using the same software. Docking was performed using PATCHDOCK server, where the optimized arsenate molecule and the constructed 3D model of arsenate reductase enzyme were provided as ligand and receptor, respectively.[41,42] One hundred models of docked complex were generated by the server, of which the first ten models acquiring higher scores were used to refine by FireDock.[43] Among these ten models, the highest geometric score containing docked complex was selected to analyze by Discovery Studio to identify the key residues of the enzyme involved in the interaction with the arsenate ion and the types of bonds formed between them.

## Results and Discussion

### *Sequence Analysis*

The FASTA sequence of the arsenate reductase of *B. megaterium* retrieved from the NCBI database had the accession number of KFM97827 which was comprised of 140 amino acids. Various physicochemical properties of this sequence were obtained from ProtParam server which is presented in Table 1. These properties can be very useful while analyzing the protein in laboratory. For example, the molecular weight (15687.54 Da) and the isoelectric point (pI) (5.51) of the protein were theoretically calculated with the help of this server which can be beneficial to know while performing gel electrophoresis for the protein. The extinction

**Table 1.** Physicochemical Properties of Arsenate Reductase of *B. megaterium* Computed using ProtParam Server

| Physicochemical Parameters | Values |
|---|---|
| Number of Amino Acid | 140 |
| Molecular Weight | 15687.54 |
| Theoretical pI | 5.51 |
| Formula | $C_{684}H_{1058}N_{198}O_{213}O_7$ |
| Total Numbers of Atoms | 2160 |
| Extinction Coefficient | 25230 $M^{-1}$ $cm^{-1}$ |
| Absorbance | 1.608 L $M^{-1}$ $cm^{-1}$ |
| Estimated Half Life | 30 hours (mammalian reticulocytes, *in vitro*)<br>>20 hours (yeast, *in vivo*)<br>>10 hours (*Escherichia coli*, *in vivo*) |
| Instability Index | 29.81 |
| Aliphatic Index | 65.57 |
| GRAVY | -0.648 |

**Table 2**. Homologous Sequences of Arsenate Reductase obtained from NCBI through BLAST Search

| Homologous Sequences from Different Species | Query Coverage | Percentage Identity | E-value |
|---|---|---|---|
| *Bacillus licheniformis*_Arsenate reductase | 100% | 84.29% | 6e-93 |
| *Bacillus subtilis*_Arsenate reductase | 100% | 80.71% | 2e-89 |
| *Geobacillus thermodenitrificans*_Arsenate reductase | 99% | 82.01% | 1e-88 |
| *Bacillus halodurans*_Arsenate reductase | 100% | 77.86% | 6e-84 |
| *Bacillus clausii*_Arsenate reductase | 100% | 75.71% | 1e-82 |
| *Oceanobacillus iheyensis*_Arsenate reductase | 100% | 72.86% | 1e-80 |
| *Bacillus cereus*_Arsenate reductase | 99% | 69.06% | 7e-74 |
| *Bacillus cytotoxicus*_Arsenate reductase | 99% | 69.06% | 4e-73 |
| *Brevibacillus brevis*_Arsenate reductase | 99% | 69.78% | 3e-72 |
| *Bacillus anthracis*_Arsenate reductase | 99% | 68.35% | 2e-71 |
| *Bacillus thuringiensis*_Arsenate reductase | 99% | 67.63% | 8e-71 |
| *Staphylococcus aureus*_Arsenate reductase | 96% | 61.48% | 1e-62 |

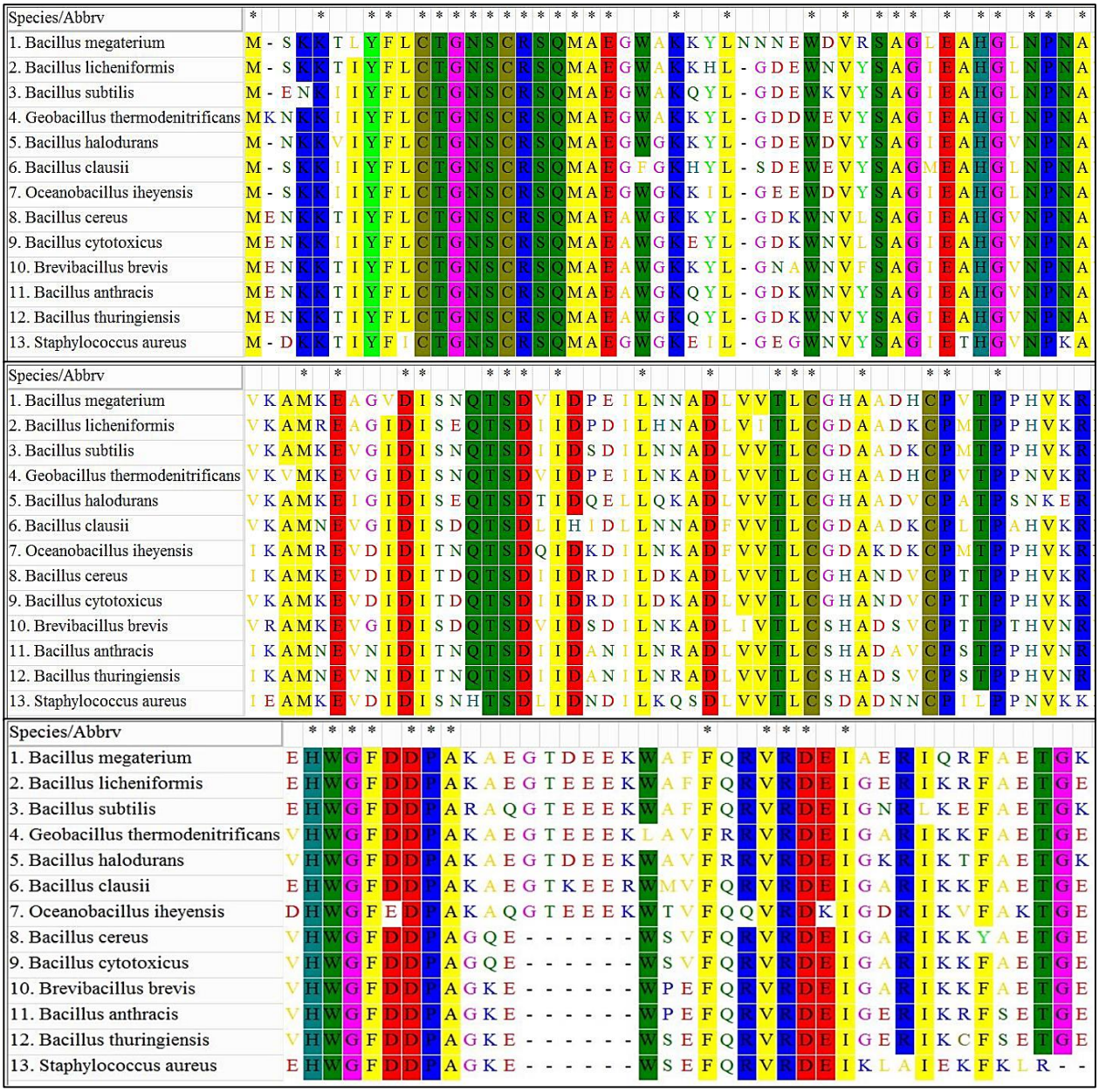


**Figure 1.** Multiple Sequence Alignment of Arsenate Reductase of *B. megaterium* with Homologous Sequences. Here the asterisk at the top of the sequences implies identical amino acids within the sequences. The regions which are conserved at 90% level or above are highlighted by colored blocks.

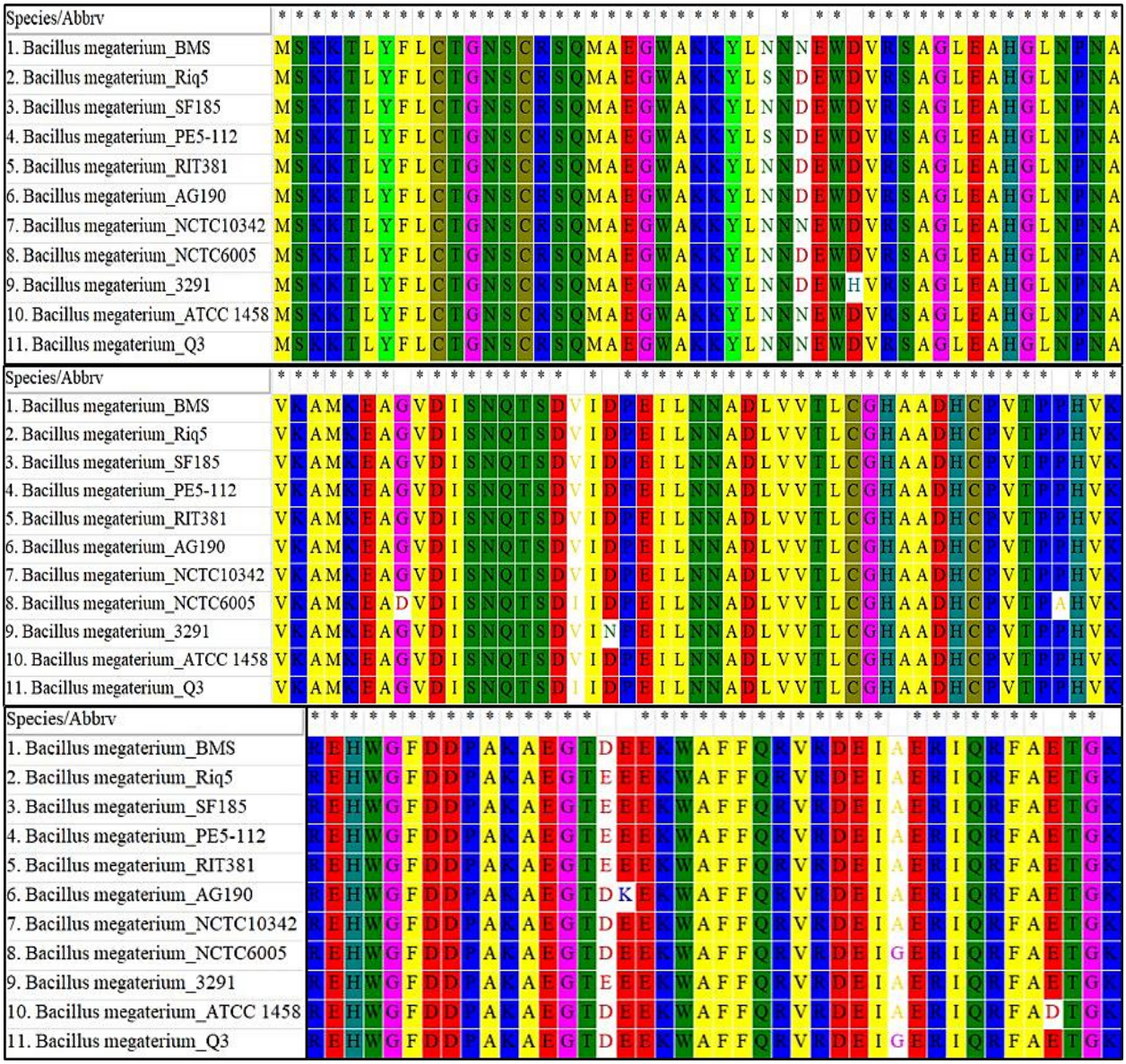


**Figure 2.** Multiple Sequence Alignment of Arsenate Reductase Enzyme of Various Strains of *B. megaterium*. Here the asterisk at the top of the sequences implies identical amino acids within the sequences. The regions which are conserved at 90% level or above are highlighted by colored blocks.

coefficient of the protein was estimated to be 25230 $M^{-1}$ $cm^{-1}$ at 280 nm wavelengths in water. The estimation of extinction coefficient helps to calculate the absorbance or optical density of a protein at a certain wavelength which was found to be 1.608 L $M^{-1}$ $cm^{-1}$ for the target protein. It would actually be useful when analyzing the protein with spectrophotometer. The half-life of the protein was predicted to be 30 h in mammalian reticulocytes (*in vitro*) which postulates that half of the amount of the protein would disappear within this time after its synthesis in the cell. On the other hand, in yeast (*in vivo*) and in *Escherichia coli* (*in vivo*), the half-life of the protein was assessed to be >20 h and >10 h, respectively. The enzyme scored an instability index of less than 40, which manifested its higher stability in the test tube. Another important feature obtained from ProtParam was the grand average of hydropathicity (GRAVY), the positive and negative value of which indicates the hydrophobic and hydrophilic nature of various proteins, respectively.[26] Since the target protein had a negative GRAVY score (-0.648), it pointed toward the hydrophilic nature of the protein.

The family and domain of the enzyme arsenate reductase were determined by Pfam, CATH, CDD, CDART, and SMART. It was found that this enzyme belongs to the Low Molecular Weight Phosphatase (LMWP) family which is included in the Protein Tyrosine Phosphatase (PTP) super family. The LMWP is a group of enzymes that have a highly conserved active site motif which is CXGNXCRS (here, X = any amino acid). This motif was found to be present in the sequence of the query enzyme within 10-17 amino acids (contain active site motif CTGNSCRS). The CDD server

also provided insights into protein's function describing its role in arsenic detoxification through reduction of arsenate to arsenite by coupling with thioredoxin (Trx)/thioredoxin reductase (TrxR) or glutathione (GSH)/glutaredoxin (Grx). Studying the CATH imparted that the query enzyme belongs to the alpha-beta class which forms a three layer alpha-beta-alpha sandwich architecture incorporating Rossmann fold topology.

### *Multiple Sequence Alignment and Phylogenetic Analysis*

In order to demonstrate the evolutionary relationship of arsenate reductase of *B. megaterium* with its homologs, it was aimed to perform multiple sequence alignment and construct phylogenetic tree. In this purpose, the homologous sequences of the target enzyme were identified through BLAST search. The result of BLAST revealed that the target sequence had very high similarity with eight species of *Bacillus* and also with some other closely related groups which were from different genus (*Geobacillus, Oceanobacillus*), but belonged to the same family (Bacillaceae). The target sequence also had similarity with some species of *Staphylococcus*. A total of 12 species were selected from BLAST for alignment, all of which had ≥96% query coverage and >61% sequence identity regarding the target sequence (Table 2). The multiple sequence alignment exhibited that large number of residues of the enzyme of the selected species were highly conserved (Figure 1). Moreover, multiple sequence alignment was also executed among different strains of *B. megaterium* which showed that most of the residues were identical (Figure 2). Both alignments revealed that the active site motif (CTGNSCRS) of *B. megaterium* was completely conserved among all the strains and also among all the selected species. The phylogenetic tree made the relationship of the target enzyme with its homologs more clear (Figure 3). The pattern of branching in the phylogenetic tree postulated that the arsenate reductase of *B. megaterium* was more closely related to *B. licheniformis* compared to others, since both of them diverged from the most recent common ancestor. This clade then converged with *B. subtilis* which made it the next closely related species to *B. megaterium*. In the phylogenetic tree, the *S. aureus* was represented as the out group.

### *Secondary Structure Prediction*

According to the prediction of PSIPRED, the structure of arsenate reductase was composed of α-helices, β-strands and coil structures (Figure 4). It was found that most of the residues of the enzyme were polar and small non-polar amino acids. Besides, there were also some hydrophobic and aromatic amino acids which were comparatively less in number.

### *Modeling of 3D Structure*

The 3D model of arsenic reductase of *B. megaterium* was obtained from the homology modeling program MODELLER. To construct this model, the selected homologous templates identified from BLAST search were the crystal structure of *B. subtilis* arsenate reductase (PDB ID-1JL3), solution structure of Trx-arsenate reductase complex from *B. subtilis* (PDB ID-2IPA), and X-ray structure of arsenate reductase from *S. aureus* (PDB ID-1JF8) (Table 3).

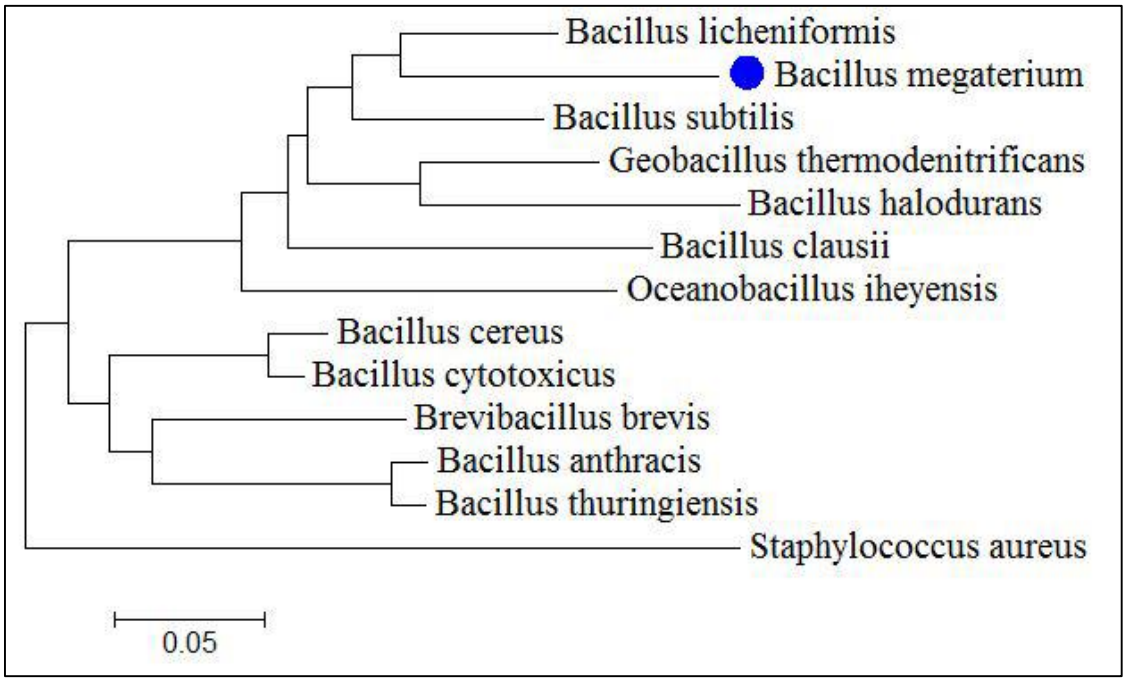


**Figure 3.** Phylogenetic Relationship of Arsenate Reductase of *B. megaterium* with Sequences from Different Species. The position of the target sequence is highlighted by a blue circle.

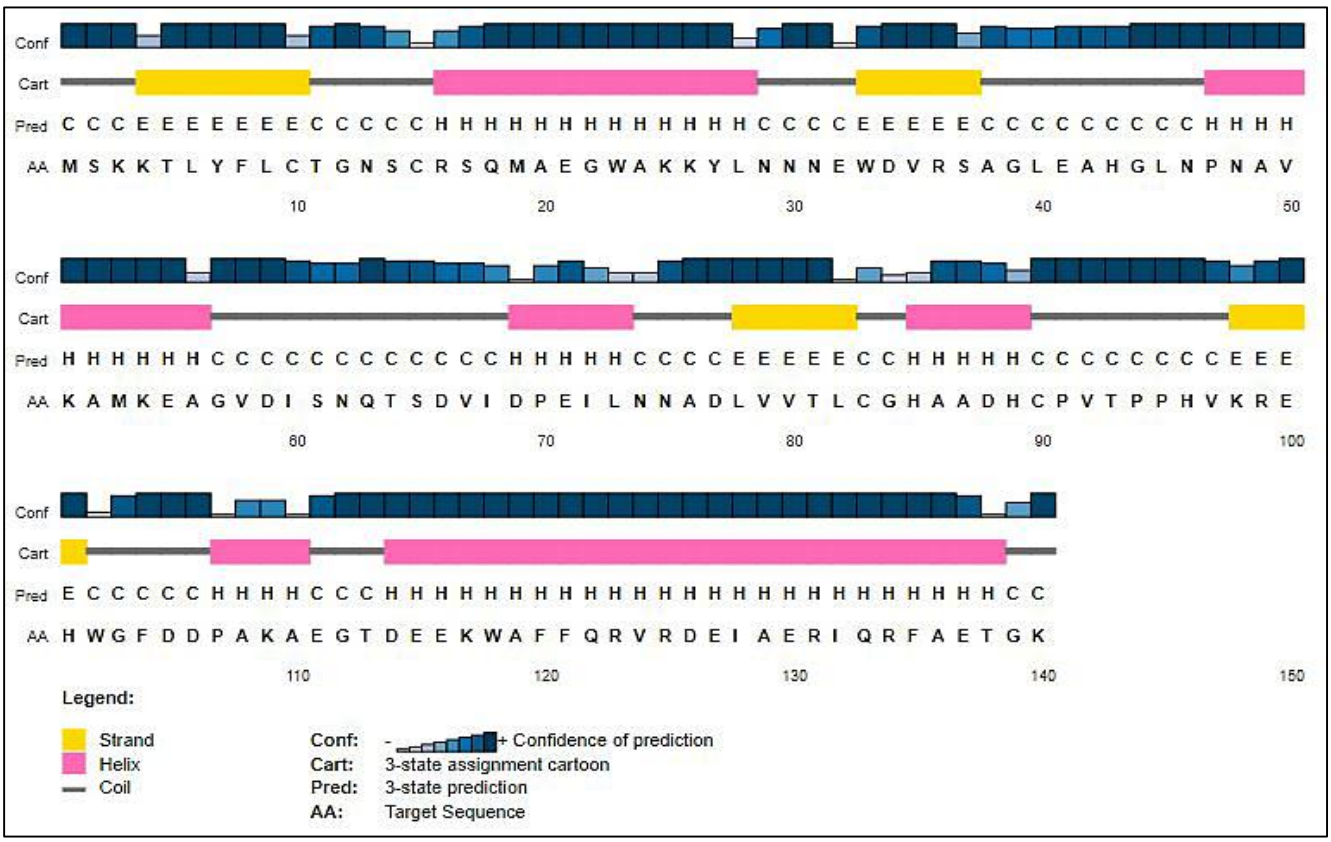


**Figure 4.** Secondary Structure Prediction of the Arsenate Reductase of *B. megaterium*. Here the purple and yellow colored regions show the α-helix and β-strand forming residues, respectively and regions with coil structure are represented by grey colored line.

Based on the multiple sequence alignment between these templates and the query enzyme, five models of the enzyme were generated by MODELLER. These five models were primarily evaluated by DOPE score and among them; model-2 was chosen to be the best model on the basis of the lowest value of the DOPE score (Table 4). The energy of this protein model was found to be -2092.378 KJ/mol from the calculation performed by Swiss-PdbViewer. To make the structure stable and free from steric clashes, refinement of the model was done by minimizing the energy to -5975.973 KJ/mol. When visualized by PyMOL, the structure of the modeled enzyme was found to be made of five α-helices, which were labeled as α1, α2, α3, α4 and α5, and four β-sheets, labeled as β1, β2, β3, and β4 (Figure 5). The β-sheets are parallel with each other and are flanked by the α-helices on both sides. Among the α-helices, α5 is the longest which ends with C-terminal region of the protein. On the other hand, the N-terminal region starts with the β1 strand. The α-helices and the β-sheets are linked together by coil structures.

**Table 3.** Selected Templates for the Homology Modeling of Arsenate Reductase

| Description | PDB ID | Query Coverage | Identity | E-value |
|---|---|---|---|---|
| Crystal structure of *B. subtilis* (Arsc) (Chain A) | 1JL3 | 100% | 81% | 3e-85 |
| Solution structure of Trx-Arsc complex of *B. subtilis* (Chain B) | 2IPA | 100% | 79% | 4e-82 |
| X-ray structure of arsenate reductase from *Staphylococcus aureus* (Chain A) | 1JF8 | 96% | 60% | 4e-56 |

**Table 4.** DOPE Score of Five Generated Models by MODELLER

| Models of Arsenate Reductase | DOPE Score |
|---|---|
| Arsenate reductase_model.01 | -14719.20020 |
| Arsenate reductase_model.02 | -14888.80371 |
| Arsenate reductase_model.03 | -14833.30664 |
| Arsenate reductase_model.04 | -14529.88281 |
| Arsenate reductase_model.05 | -14477.31348 |

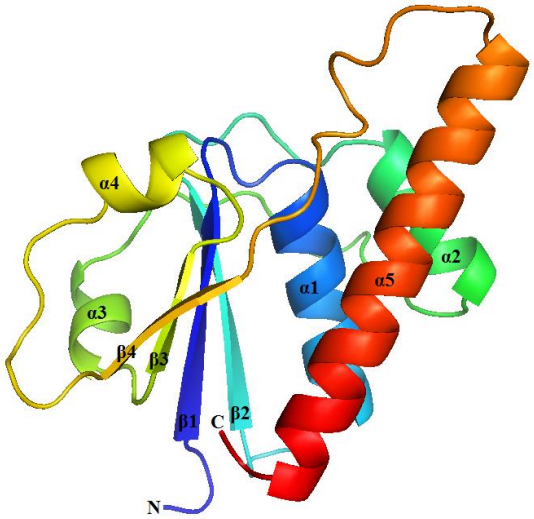


**Figure 5.** The 3D Structure of Arsenate Reductase of *B. megaterium* Representing by Cartoon Model and Rainbow Color; the Five Alpha Helices are Labeled as α1, α2, α3, α4 and α5 and the Four Beta Sheets as β1, β2, β3 and β4, Respectively.

### *Evaluation of the Model*

The stereochemical quality of the arsenate reductase model was assessed by analyzing the Ramachandran plot of the protein obtained from PROCHECK (Figure 6). According to the Ramachandran plot, 95.9% residues of the model was present in the most favored regions, while 3.3% and 0.8% residues were resided in the additional allowed regions and generously allowed regions, respectively (Table 5). There was no amino acid present in the disallowed regions. Thus, the analysis by PROCHECK revealed that the phi/psi angles of all the residues of the model were within the acceptable range, ensuring good stereochemical quality of the model.

**Table 5.** Distribution of Residues of Arsenate Reductase Model in Ramachandran Plot

| Regions of Ramachandran plot | No. of Residues | Percentage of Residues |
|---|---|---|
| Most favored regions | 118 | 95.9% |
| Additional allowed regions | 4 | 3.3% |
| Generously allowed regions | 1 | 0.8% |
| Disallowed regions | 0 | 0.0% |

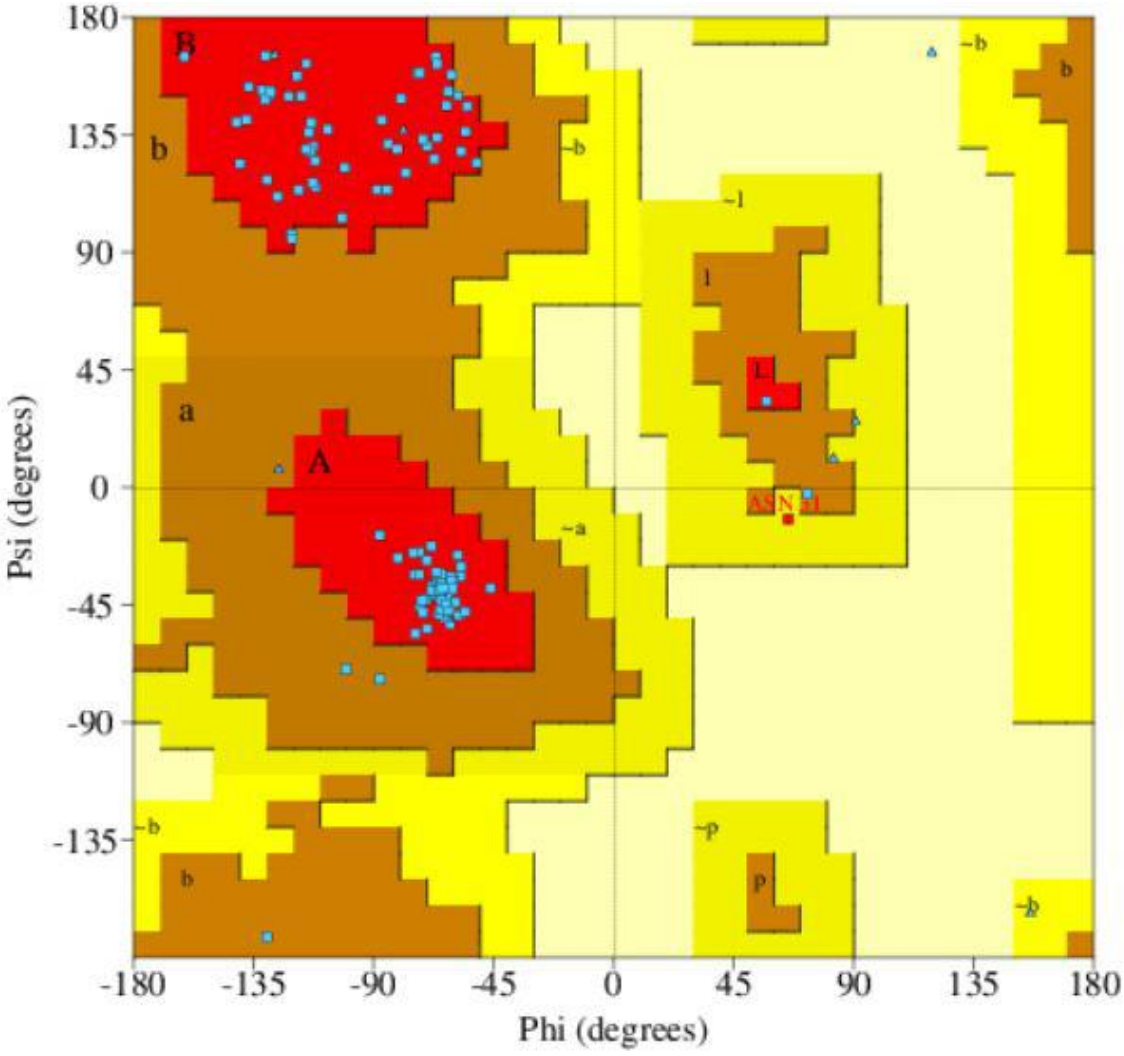


**Figure 6.** Ramachandran Plot of Arsenate Reductase Model Showing the Distribution of Amino Acids in Most Favored Regions, Additional Allowed Regions, Generously Allowed Regions and Disallowed Regions.

According to the calculation done by Verify 3D, 100% of the residues of arsenate reductase model had the 3D-1D average score above the threshold level which was 0.2 (Figure 7). It confirmed very good compatibility of the protein's 3D structure with its amino acid sequence. The ERRAT plot (Figure 8) for the arsenate reductase model showed that the error value of all the residues were below the 99% confidence limit and most of them also had less than 95% error value, except residues 57-59. The ERRAT server also provided the overall quality factor of the protein model. It expressed the percentage of protein responsible for residual error value to fall below the 95% rejection limit. Usually, protein structures with high resolutions get quality factor around 95 or higher. Since the quality factor of arsenate reductase was found to be 97.7273, it indicated a very good quality of the model.

The superimposition of arsenate reductase model and the template (1JL3) structure exhibited very high similarity between these two structures (Figure 9). The root mean square deviation (RMSD) between the target and the template was 0.382 which was lower than 1. This low deviation meant that the constructed protein model was well grounded.

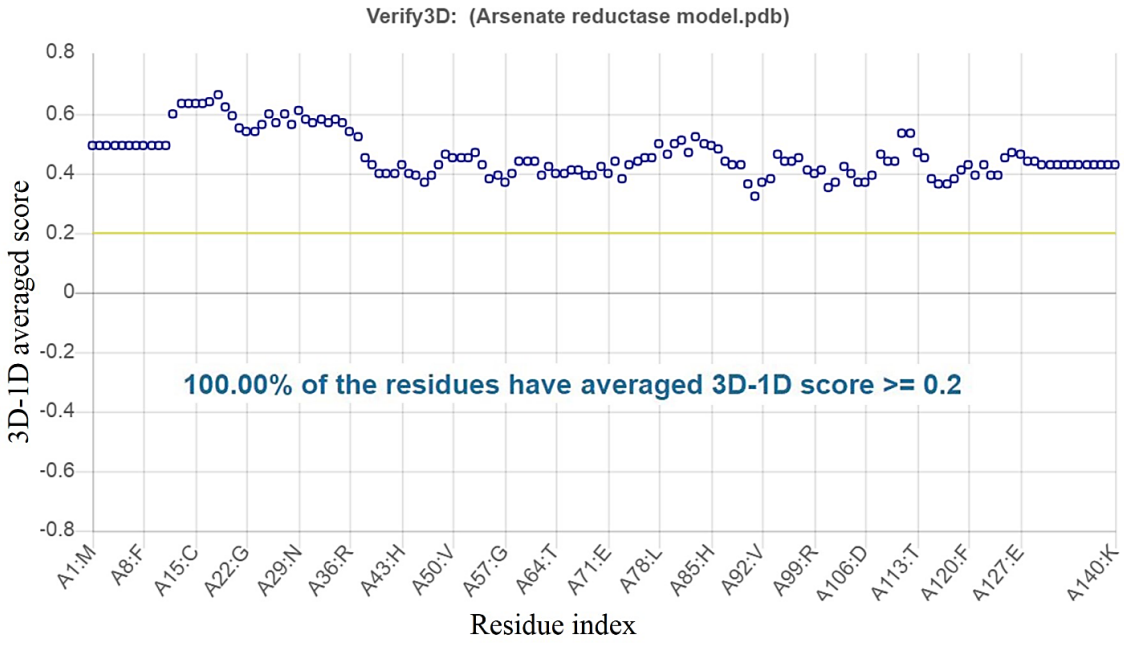


**Figure 7.** 3D-1D Profile of Arsenate Reductase Model Obtained from Verify 3D.

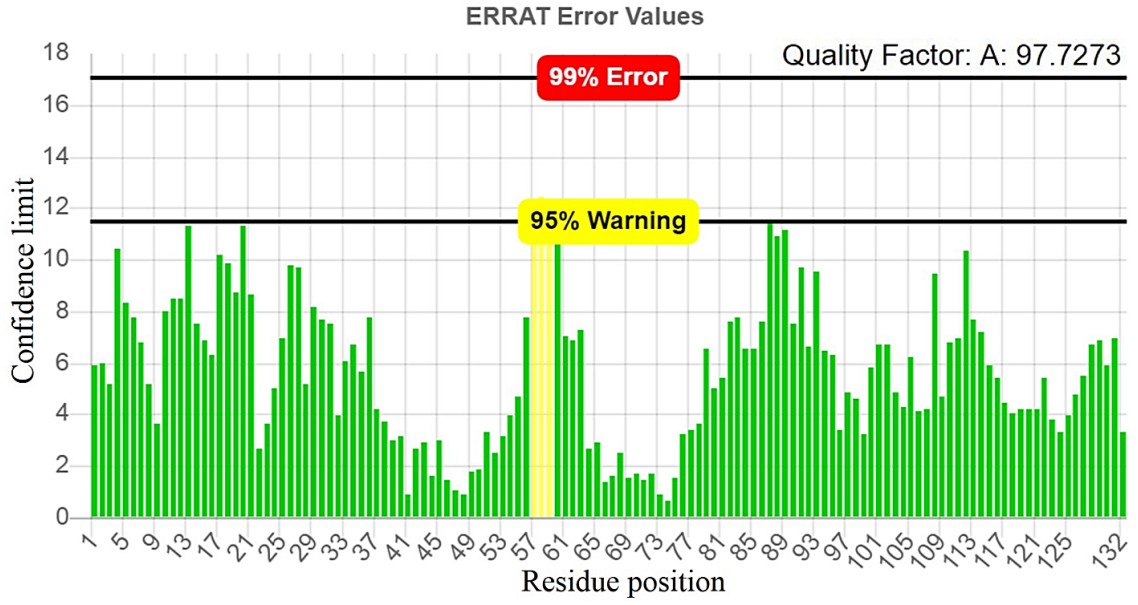


**Figure 8.** Plot Showing Error Values and Quality Factor of Arsenate Reductase Model Obtained from ERRAT.

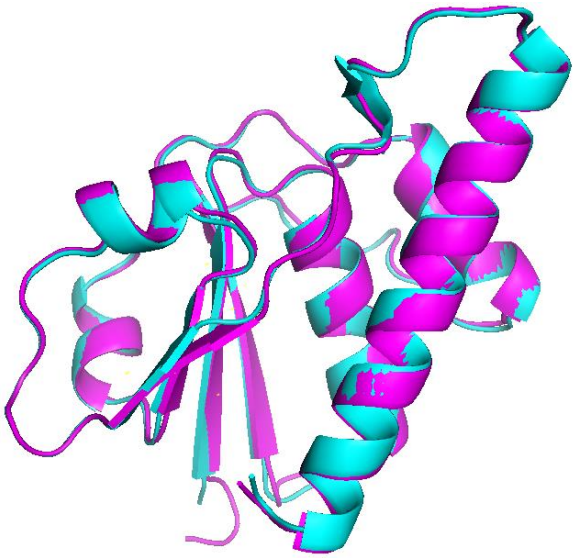

**Figure 9.** Superimposition of Arsenate Reductase Model (Purple Color) with the Structure of Template Protein (1JL3) (Cyan Color).

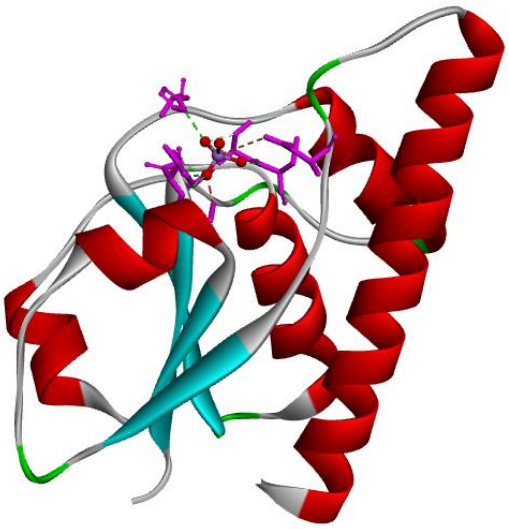

**Figure 10.** Docked Complex of Arsenate Reductase Enzyme with Arsenate Ion. The arsenate ion is represented by ball and stick model, the enzyme is represented by solid ribbon and the residues of the enzyme involved in interaction with ligand are represented by solid sticks.

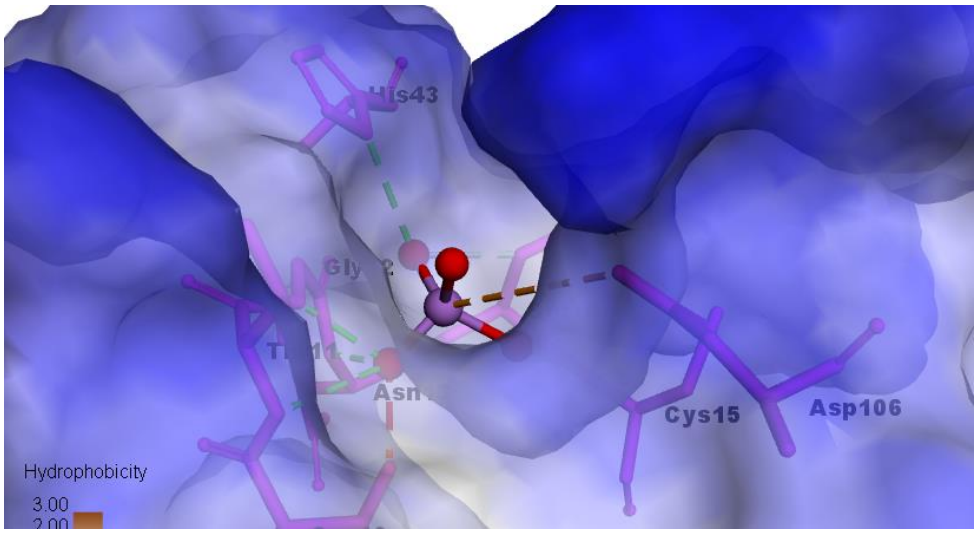


**Figure 11.** Binding Pocket of Arsenate Reductase Represented by Surface Model Based on Hydrophobicity.

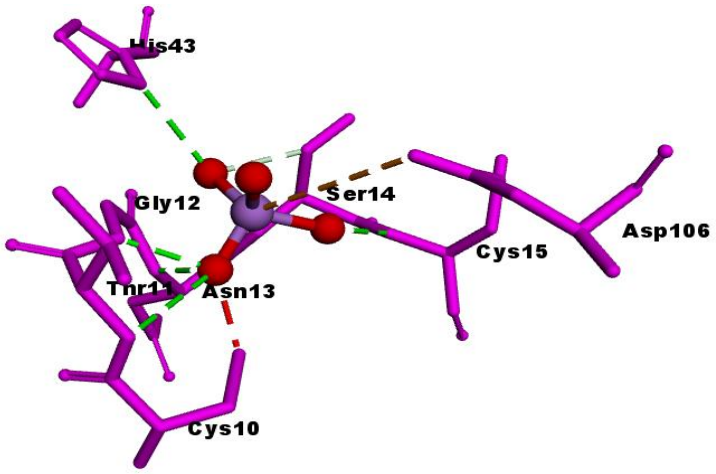


**Figure 12.** Interaction of Arsenate Ion with The Amino Acids of Arsenate Reductase.

### *Study of Intramolecular Interaction within the Docked Complex*

The docked complex of arsenate reductase enzyme with arsenate ion (Figure 10) was obtained from PATCHDOCK and FireDock server which helped to analyze the interaction between them. When visualized by Discovery Studio, the arsenate ion was found to be resided in a pocket of the enzyme (Figure 11). The key amino acids of the enzyme which built the binding pocket and were involved in the interaction with arsenate were Cys10, Thr11, Gly12, Asn13, Ser14, Cys15, His43, and Asp106 (Figure 12). Among these residues, Cys10, Thr11, Gly12, Asn13, Ser14, and Cys15 are part of the classical conserved motif (CTGNSCRS). These six residues and also His43 formed hydrogen bonds with the oxygen molecules of the arsenate ion. On the other hand, the Asp106 residue formed electrostatic bond with the arsenic molecule. The distances of hydrogen bonds between the interacting residues and the arsenate ion were within the range of 2.1-3.4 Å and the distance of the electrostatic bond was about 5.3 Å (Table 6).

**Table 6.** Bond Types and Bond Distances between Receptor and Ligand

| Atoms of Amino Acids of Receptor | Atoms of Ligand | Types of Bonds | Bond Distances |
|---|---|---|---|
| ASP106:Oδ2 | As | Electrostatic | 5.27718 |
| THR11:N | O | Hydrogen bond | 3.35485 |
| GLY12:N | O | Hydrogen bond | 2.30709 |
| ASN13:N | O | Hydrogen bond | 2.42332 |
| CYS15:N | O | Hydrogen bond | 2.50510 |
| HIS43:Nδ1 | O | Hydrogen bond | 2.83209 |
| SER14:Cβ | O | Hydrogen bond | 2.62944 |
| CYS10:Sγ | O | Hydrogen bond | 2.15936 |

## Conclusion

The present study predicts and validates the 3D structure of arsenate reductase enzyme of *B. megaterium*. The molecular docking study shows the formation of a stable complex of arsenate reductase with arsenate. In the docked complex, the binding site of the enzyme was found to be comprised of Cys10, Thr11, Gly12, Asn13, Ser14, Cys15, His43, and Asp106. The binding of this enzyme with arsenate ion led to the detrimental conversion of arsenate to arsenite. The computational 3D structure of this enzyme and its interaction with the ion can provide a comprehensive insight of this process. This understanding can be exploited in further studies to inhibit this interaction and subsequently can procure a potent means of bioremediation.

## Authors' Contributions

MNJA and NT contributed to the conceptualization, data curation, writing-original draft, and editing; AN, SS, MKR, and RM contributed to the data curation, formal analysis, writing-review, and editing; MZI performed conceptualization, formal analysis, writing-review, and editing; USZ and MSR performed conceptualization, formal analysis.

## Conflict of Interest Disclosures

The authors declare that they have no conflicts interest.

## Acknowledgment

This work was supported in part by a research grant 2017-2018 from Jahangirnagar University to Md. Zahidul Islam.

## References

1. Salem HM, Eweida EA, Farag A. Heavy metals in drinking water and their environmental impact on human health. ICEHM2000, Cairo University, Egypt. 2000:542-56.
2. Oremland RS, Stolz JF. Arsenic, microbes and contaminated aquifers. Trends Microbiol. 2005;13(2):45-9. doi:10.1016/j.tim.2004.12.002
3. Baker BA, Cassano VA, Murray C. Arsenic exposure, assessment, toxicity, diagnosis, and management: guidance for occupational and environmental physicians. J Occup Environ Med. 2018;60(12):e634-9. doi:10.1097/JOM.0000000000001485
4. Dey S, Rosen BP. Dual mode of energy coupling by the oxyanion-translocating ArsB protein. J Bacteriol. 1995;177(2):385-9. doi:10.1128/jb.177.2.385-389.1995
5. Kaur S, Kamli MR, Ali A. Diversity of arsenate reductase genes (arsC genes) from arsenic-resistant environmental isolates of *E. coli*. Curr Microbiol. 2009;59(3):288-94. doi:10.1007/s00284-009-9432-9
6. Poddar N, Badilla C, Maghool S, Osborne TH, Santini JM, Maher MJ. Structural and functional investigation of the periplasmic arsenate-binding protein ArrX from *Chrysiogenes arsenatis*. Biochemistry. 2021;60(6):465-76. doi:10.1021/acs.biochem.0c00555
7. Nurchi VM, Buha Djordjevic A, Crisponi G, Alexander J, Bjшrklund G, Aaseth J. Arsenic toxicity: molecular targets and therapeutic agents. Biomolecules. 2020;10(2):235. doi:10.3390/biom10020235
8. Tawfik DS, Viola RE. Arsenate replacing phosphate: alternative life chemistries and ion promiscuity. Biochemistry. 2011;50(7):1128-34. doi:10.1021/bi200002a
9. Ahmann D, Roberts AL, Krumholz LR, Morel FM. Microbe grows by reducing arsenic. Nature. 1994;371:750. doi:10.1038/371750a0
10. Nordstrom DK. Worldwide occurrences of arsenic in ground water. Science. 2002;296(5576):2143-5. doi:10.1126/science.10723
11. Jain RM, Mody K, Mishra A, Jha B. Physicochemical characterization of biosurfactant and its potential to remove oil from soil and cotton cloth. Carbohydr Polym. 2012;89(4):1110-6. doi:10.1016/j.carbpol.2012.03.077
12. Marchiset-Ferlay N, Savanovitch C, Sauvant-Rochat MP. What is the best biomarker to assess arsenic exposure via drinking water?. Environ Int. 2012;39(1):150-71. doi:10.1016/j.envint.2011.07.015
13. Sharma P, Jha AB, Dubey RS, Pessarakli M. Reactive

oxygen species, oxidative damage, and antioxidative defense mechanism in plants under stressful conditions. J Bot. 2012;2012:217037. doi:10.1155/2012/217037
14. Rosen BP, Weigel U, Monticello RA, Edwards BP. Molecular analysis of an anion pump: purification of the ArsC protein. Arch Biochem Biophys. 1991;284(2):381-5. doi:10.1016/0003-9861(91)90312-7
15. Ji G, Garber EA, Armes LG, Chen CM, Fuchs JA, Silver S. Arsenate reductase of *Staphylococcus aureus* plasmid pI258. Biochemistry. 1994;33(23):7294-9. doi:10.1021/bi00189a034
16. Bennett MS, Guan Z, Laurberg M, Su XD. *Bacillus subtilis* arsenate reductase is structurally and functionally similar to low molecular weight protein tyrosine phosphatases. Proc Natl Acad Sci USA. 2001;98(24):13577-82. doi:10.1073/pnas.2413971
17. Li Y, Hu Y, Zhang X, Xu H, Lescop E, Xia B, et al. Conformational fluctuations coupled to the thiol-disulfide transfer between thioredoxin and arsenate reductase in *Bacillus subtilis*. J Biol Chem. 2007;282(15):11078-83. doi:10.1074/jbc.M700970200
18. Chauhan D, Srivastava PA, Agnihotri V, Yennamalli RM, Priyadarshini R. Structure and function prediction of arsenate reductase from Deinococcus indicus DR1. J Mol Model. 2019;25(1):15. doi:10.1007/s00894-018-3885-3
19. Rahman M, Hossain M, Saha SK, Rahman S, Sonne C, Kim KH. Homology Modeling and Probable Active Site Cavity Prediction of Uncharacterized Arsenate Reductase in Bacterial spp. Appl Biochem Biotechnol. 2021;193(1):1-18. doi:10.1007/s12010-020-03392-w
20. Kartinen Jr EO, Martin CJ. An overview of arsenic removal processes. Desalination. 1995;103(1-2):79-88. doi:10.1016/0011-9164(95)00089-5
21. Gladysheva TB, Oden KL, Rosen BP. Properties of the arsenate reductase of plasmid R773. Biochemistry. 1994;33(23):7288-93. doi:10.1021/bi00189a033
22. Nahrstedt H, Schruder C, Meinhardt F. Evidence for two recA genes mediating DNA repair in *Bacillus megaterium*. Microbiology. 2005;151(3):775-87. doi:10.1099/mic.0.27626-0
23. Bunk B, Kucklick M, Jonas R, Munch R, Schobert M, Jahn D, et al. MetaQuant: a tool for the automatic quantification of GC/MS-based metabolome data. Bioinformatics. 2006;22(23):2962-5. doi:10.1093/bioinformatics/btl526
24. Islam MZ, Arbi MN, Moni R, Zohora US, Nishat FZ, Khanam R, et al. Reduction of Arsenate in a new isolate of *Bacillus megaterium*. Jahangirnagar Univ J Biol Sci. 2019;8(2):47-59. doi:10.3329/jujbs.v8i2.49839
25. Gill SC, Von Hippel PH. Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem. 1989;182(2):319-26. doi:10.1016/0003-2697(89)90602-7
26. Guruprasad K, Reddy BB, Pandit MW. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting *in vivo* stability of a protein from its primary sequence. Protein Eng Des Sel. 1990;4(2):155-61. doi:10.1093/protein/4.2.155
27. Ikai A. Thermostability and aliphatic index of globular proteins. J Biochem. 1980;88(6):1895-8. doi:10.1093/oxfordjournals.jbchem.a133168
28. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol. 1982;157(1):105-32. doi:10.1016/0022-2836(82)90515-0
29. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 2003;31(13):3784-8. doi:10.1093/nar/gkg563
30. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, et al. InterPro in 2017—beyond protein family and domain annotations. Nucleic Acids Res. 2017;45(D1):D190-9. doi:10.1093/nar/gkw1107
31. Das S, Sillitoe I, Lee D, Lees JG, Dawson NL, Ward J, et al. CATH FunFHMMer web server: protein functional annotations using functional family assignments. Nucleic Acids Res. 2015;43(W1):W148-53. doi:10.1093/nar/gkv488
32. Geer LY, Domrachev M, Lipman DJ, Bryant SH. CDART: protein homology by domain architecture. Genome Res. 2002;12(10):1619-23. doi:10.1101/gr.278202
33. Letunic I, Copley RR, Schmidt S, Ciccarelli FD, Doerks T, Schultz J, et al. SMART 4.0: towards genomic data integration. Nucleic Acids Research. 2004;32(suppl_1):D142-4. doi:10.1093/nar/gkh088
34. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol biol. 1990;215(3):403-10. doi:10.1016/S0022-2836(05)80360-2
35. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013;30(12):2725-9. doi:10.1093/molbev/mst197
36. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J Mol biol. 1999;292(2):195-202. doi:10.1006/jmbi.1999.3091
37. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. Electrophoresis. 1997;18(15):2714-23. doi:10.1002/elps.1150181505
38. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr. 1993;26(2):283-91. doi:10.1107/S0021889892009944
39. Luthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. Nature. 1992;356(6364):83-5. doi:10.1038/356083a0
40. Bowie JU, Eisenberg D. An evolutionary approach to folding small alpha-helical proteins that uses sequence information and an empirical guiding fitness function. Proc Natl Acad Sci USA. 1994;91(10):4436-40. doi:10.1073/pnas.91.10.443
41. Duhovny D, Nussinov R, Wolfson HJ. Efficient unbound docking of rigid molecules. InInternational workshop on algorithms in bioinformatics, Springer, Berlin, Heidelberg. 2002:185-200. doi:10.1007/3-540-45784-4_14
42. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic Acids Res. 2005;33(suppl_2):W363-7. doi:10.1093/nar/gki481
43. Andrusier N, Nussinov R, Wolfson HJ. FireDock: fast interaction refinement in molecular docking. Proteins. 2007;69(1):139-59. doi:10.1002/prot.21495