

Structure prediction and evolution of a halo-acid dehalogenase of *Burkholderia mallei*

Alok R Rai^{1*}, Raghvendra Pratap Singh², Alok Kumar Srivastava² & Ramesh Chandra Dubey³

¹Department of Microbiology, Seth Kesarimal Porwal College, Kamptee Maharashtra 441002, India; ²National Bureau of Agriculturally Important Microorganisms (ICAR), Kushmaur, Kaithauli, Mau Nath Bhanjan, Uttar Pradesh-275101, India; ³Department of Botany and Microbiology, Gurukul Kangri University, Haridwar, Uttrakhand-249404, India; Alok R Rai – Email: raialok@sify.com; Phone: +91-9423407292; *Corresponding author

Received November 03, 2012; Accepted November 05, 2012; Published November 13, 2012

Abstract:

Environmental pollutants containing halogenated organic compounds e.g. haloacid, can cause a plethora of health problems. The structural and functional analyses of the gene responsible of their degradation are an important aspect for environmental studies and are important to human well-being. It has been shown that some haloacids are toxic and mutagenic. Microorganisms capable of degrading these haloacids can be found in the natural environment. One of these, a soil-borne *Burkholderia mallei* possesses the ability to grow on monobromoacetate (MBA). This bacterium produces a haloacid dehalogenase that allows the cell to grow on MBA, a highly toxic and mutagenic environmental pollutant. For the structural and functional analysis, a 346 amino acid encoding protein sequence of haloacid dehalogenase is retrieved from NCBI data base. Primary and secondary structure analysis suggested that the high percentage of helices in the structure makes the protein more flexible for folding, which might increase protein interactions. The consensus protein sub-cellular localization predictions suggest that dehalogenase protein is a periplasmic protein 3D2GO server, suggesting that it is mainly employed in metabolic process followed by hydrolase activity and catalytic activity. The tertiary structure of protein was predicted by homology modeling. The result suggests that the protein is an unstable protein which is also an important characteristic of active enzyme enabling them to bind various cofactors and substrate for proper functioning. Validation of 3D structure was done using Ramachandran plot ProSA-web and RMSD score. This predicted information will help in better understanding of mechanism underlying haloacid dehalogenase encoding protein and its evolutionary relationship.

Keywords: *Burkholderia mallei*, haloacid dehalogenase, homology modeling.

Background:

Intensification of agriculture and manufacturing industries has resulted in increased release of a wide range of dangerous and harmful byproducts, xenobiotic compounds and toxic waste to the environment named pollutants, a type of diversified problem and a big challenge of living being. In which halogenated organic compounds are one of the largest groups of environmental pollutants, and accordingly dehalogenases that catalyze the degradation of these compounds attract a great deal of attention from the viewpoint of environmental technology. Beside it 2-Haloacid dehalogenases are also useful for the production of optically

active 2-hydroxyal kanoic acids and 2-haloalkanoic acids, which are used as chiral synthons in chemical industry [1, 2]. Relationship between structures and functions of haloacid dehalogenases is interesting not only from the view point of basic enzymology, but also from the point of view of the molecular design of dehalogenases showing more efficient catalysis. In the biosphere *Burkholderia* sp. is one of versatile, gram-negative bacteria. They have been isolated in various niches including soil, water, plants, the rhizosphere and animals including humans. *Burkholderia cepacia* strains have been shown to be plant and/or animal pathogens, and yet they have also been found to possess antibacterial and antifungal

activity. Moreover, they have also been shown to be involved in the bioremediation of many types of environmental pollutants [3-5] due to the presence of the enzyme haloacid dehalogenase. Haloacids are metabolic products of naturally occurring compounds and are also disinfection by-products of sewage and water. It has been shown that some haloacids are toxic and mutagenic. Microorganisms capable of degrading these haloacids can be found in the natural environment. One of these, a soil-borne *Burkholderia mallei* posses the ability to grow on monobromoacetate (MBA). This bacterium produces a haloacid dehalogenase that allows the cell to grow on MBA, a highly toxic and mutagenic environmental pollutant.

Structure elucidation is an expensive and time consuming process and also requires extensive expertise. Currently used techniques to reveal 3D structures are X-ray crystallography and Nuclear Magnetic Resonance (NMR) Imaging. Due to the techniques being expensive and time consuming, there has been an increasing gap of information between DNA/protein sequence information and structure information. Computational methods and molecular dynamic simulations are good alternatives to overcome these problems in protein structure prediction [6]. Comparative modeling is a computational technique for 3D structure prediction of proteins using known structures as templates [7]. Hence, the adopted approach enables to formulate a structure from an amino acid sequence in less time. Our current study describes the 3D model of the haloacid dehalogenase protein obtained through homology modelling. In addition, primary and secondary structure analysis, sub-cellular localization prediction was also performed and is explained.

Methodology:

Sequence retrieval and Physico-chemical Analysis

The amino acid sequence of haloacid dehalogenase protein was retrieved from the NCBI database (<http://www.ncbi.nlm.nih.gov/>) imported to pBlast search against PDB (Protein data bank) for sequence identity cut off for homology modeling. Physicochemical characterization, theoretical isoelectric point (pI), total number of positive and negative residues, extinction coefficient [8], and instability index [9], half life time, aliphatic index [10] and grand average hydropathy (GRAVY) [11] were computed using ExPASy's Prot-Param server [12]. The sulphide (S-S) bond pattern is predicted by using the tool CYS_REC (<http://linux1.softberry.com/berry.phtml?topic>).

Secondary Structure Prediction and Homology Modeling

Secondary structural features of considered protein sequences was predicted by employing Accelry's Discovery Studio visualize 2.5, a new highly accurate secondary structure prediction method. The 3-dimensional structural prediction was carried out by iterative threading assembly refinement (I-TASSER) server (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>). I-TASSER is a platform for automated protein structure and function prediction based on the sequence to structure to function paradigm. Visualization and structure Decipherization of target proteins were carried out by studio visualize 2.5 software.

Model evaluation functional Site Prediction

An assessment of how well the final model (or ensemble of models) compares with the experimental data, is an essential

part of the structure validation process. It was done by PROCHECK; verify 3-D and Prosa, (<http://www.came.sbg.ac.at/typo3/index.php?>) the best validation and evaluation server. Functional site prediction was carried out by employing the Q-site Finder and accurate protein function was assessed by 3D2GO (<http://www.sbg.bio.ic.ac.uk/phyre/pfd/index.html>)

Phylogenetic Analysis

Phylogenetics of haloacid dehydrogenase of *Burkholderia mallei* was performed in this study to assess the evolutionary relatedness among the different species as well as genus. The haloacid dehydrogenase protein sequences were retrieving from NCBI database. Alignment and evolutionary analysis through phylogram was performed by MEGA version 4.0 [13]

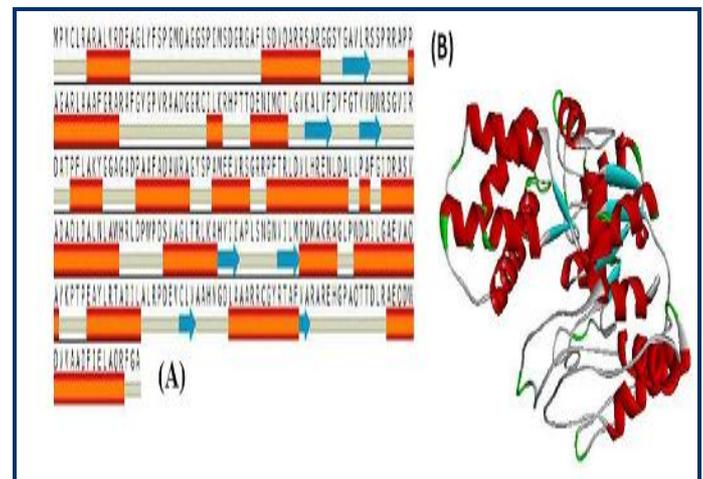


Figure 1: (A) Secondary and (B) Tertiary structure of Haloacid dehalogenase protein (PMDB Identifier code PM0078397).

Discussion:

The present study focused on structural analysis of an important biodegrading protein of *Burkholderia mallei* named haloacid dehalogenase protein. ProtParam was used to analyze different physicochemical properties from the amino acid sequence. The haloacid dehalogenase protein contains 346 amino acids, with a molecular weight of 37141.3 Daltons. Total number of negatively charged and positively charged residues is 37 and 43 respectively, which show the basic nature of protein. The predicted molecular formula of haloacid dehalogenase is $C_{1643}H_{2597}N_{495}O_{469}S_{11}$ contain 2652 bond in which 1994 single bond, 392 double bond, 152 Partial Double and 120 aromatic bond are present. The isoelectric point of haloacid dehalogenase is 9.29 which is above 7 indicates that a positive charge protein corresponds to having more positive charge residues, and an instability index of 46.83 suggest an unstable protein. The negative GRAVY index of -0.075 is indicative of a hydrophilic and soluble protein. The protein sequence was found to be rich in the amino acid leucine, suggesting a preference for α -helices in 3D structure. The sulphide (S-S) bond pattern is predicted by employing the tool CYS_REC (<http://linux1.softberry.com/berry.phtml?topic>) has shown that haloacid dehalogenase protein has 4 cysteines in positions 4, 94, 288 and 303, but have no S-S bonding. This result shows that the target protein residue have low enthalpy at folded state because disulphide bond increase the enthalpy of the folded state by stabilizing local interactions and lack of

enhancer against proteases because disulphide bonds maintaining the integrity of a protein structure against local unfolding events.

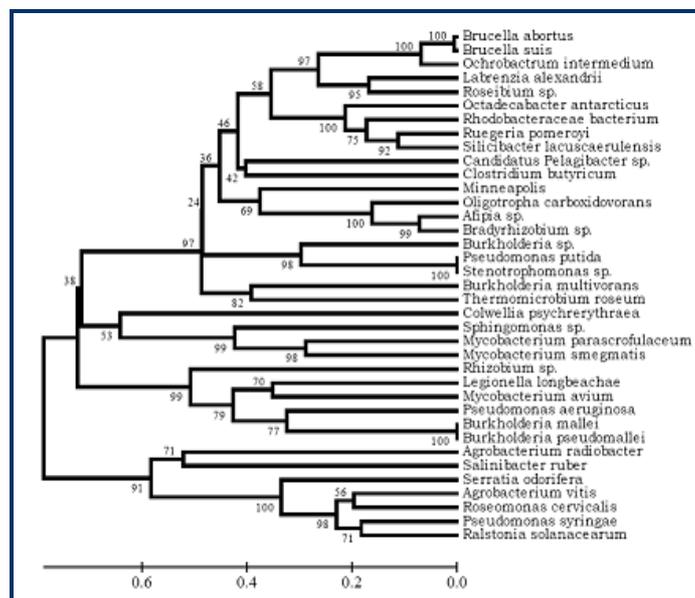


Figure 2: Phylogenetic tree of Haloacid dehalogenase

Secondary structure analysis was performed using Studio Visualizer and the protein was predicted to contain 149 Coil, 170 helix and 27 Beta sheet that consistent with protParam results (**Figure 1A**). The high percentage of helices in the structure makes the protein more flexible for folding, which might increase protein interactions. Homology modeling has been carried out by the I-TASSER server. BLASTp search against the PDB database identified 2no4B is a best template as it shows maximum identity to query sequence of the domain. Ten models were obtained in which the best model was selected based on discrete optimized protein energy (DOPE). Model visualization and contact map development was done by Studio visualizer. The predicted 3D structure of as shown in (**Figure 1B**) has been evaluated by several structure assessment methods including RMSD value, PROCHEK, verify 3-D and Prosa server. The RMSD value were found to be 14.7+/-3.6, indicates the degree to which 3D structure was calculated by superimposing template and query structure. Finally Ramachandran plots were obtained for the quality assessment of generated model. in case of haloacid dehalogenase PROCHECK displayed 83.5% of residues in the most favorable regions, with 11.0% additional allowed region and generously allowed and disallowed regions respectively (Data not shown) The comparable Ramachandran plot characteristics, RMSD values and verify 3-d score confirm the quality of the homology model of haloacid dehalogenase. The predicted 3D structure of

haloacid dehalogenase were submitted to the protein Model Database (PMBD) and assigned the PMBD ID PM0078397.

The consensus protein sub-cellular localization predictions suggest that dehalogenase protein is a periplasmic protein. The function of modeled haloacid dehalogenase protein of *Burkholderia mallei* was determined by 3D2GO server, suggesting that it is mainly employed in metabolic process followed by hydrolase activity and catalytic activity (Score 0.97, 0.96 and 0.80). Q-Site Finder server has been predicted the 25 functional sites including LEU10, VAL11, GLU14, ALA15, GLY16, LEU17, PHE19, MET31, SER32, AP33, GLY34, ARG35, PHE38, TRY53, ALA55, VAL56, LEU57, ARG58, TRP265, PRO217, ASP218, ARG312, ALA313, ASP336 and ILE 338. This result provides the abyssal information of the function of the haloacid dehalogenase protein. We have retrieved the protein sequence to display the greater evolutionary depth and relatedness of haloacid dehalogenase protein among the species as well as genus. Regarding this 37 different protein sequence was downloaded from NCBI and aligned through ClustalW program. Aligned file is then imported to MEGA version 4.0 to develop the phylogenetic tree. The generated dendrogram (**Figure 2**) has been separated in three major clusters shown that haloacid dehydrogenase is a very diverse protein among the bacteria. Protein interactions have also been predicted on the basis of the comparison of evolutionary histories, or phylogenetic trees, under the premise that interacting proteins are subject to similar evolutionary pressures resulting in similar topologies for the corresponding trees [14].

Acknowledgment:

We would like to express special gratitude to Dr. S.S Dhondge , Principal S.K Porwal College, Kamptee, for his inspiration and constant support.

References:

- [1] Motosugi K *et al. Biotech Bioeng.* 1984 **26**: 805
- [2] Hasan AKMQ *et al. Journal of Fermentation and Bioengineering.* 1991 **72**: 481
- [3] Chiarini L *et al. Burkholderia cepacia Trends Microbiology.* 2006 **14**: 277
- [4] Ramette A *et al. Appl Environ Microbiology.* 2005 **71**: 1193
- [5] Tsang JSH, *Adv Appl Microbiology.* 2004 **54**: 71 [PMID: 951043]
- [6] Liu HL & Hus JP, *Proteomics.* 2005 **5**: 2056 [PMID: 15846841]
- [7] Sali A & Blundell TL, *J Mol Bio.* 1993 **5**: 779 [PMID: 8254673]
- [8] Gill SC *et al. Anal Biochem.* 1989 **182**: 319
- [9] Guruprasad K *et al. Prot Eng.* 1990 **4**: 155
- [10] Ikai AJ, *J Biochem.* 1980 **88**: 1895
- [11] Kyte J *et al. J Mol Biol.* 1982 **157**: 105
- [12] Gasteiger E *et al. Humana Press.* 2005 **571**: 607
- [13] Tamura K *et al. Molecular Biology and Evolution.* 2007 **24**: 1596
- [14] Pazos F *et al. J Mol Biol.* 2005 **4**: 1002

Edited by P Kanguane

Citation: Rai *et al.* Bioinformation 8(22): 1111-1113 (2012)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credit