Structure prediction and analysis of MxaF from obligate, facultative and restricted facultative methylobacterium

Raghvendra Pratap Singh1*, Ram Nageena Singh1, Manish K Srivastava1, Alok Kumar Srivastava1, Sudheer Kumar1, Ramesh Chandra Dubey2 & Arun Kumar Sharma1

1Microbial Genomics Lab, National Bureau of Agriculturally Important Microorganisms, Kusmaur, Maunath Bhanjan, Uttar Pradesh-275101, India; 2Department of Botany and Microbiology, Gurukul Kangri University, Haridwar, Uttrakhand-249404, India; Raghvendra Pratap Singh – Email: raghvendra_nbaim@rediffmail.com; Phone: +91-9455431180, Fax: 0547-253038; *Corresponding author

Received September 24, 2012; Accepted October 15, 2012; Published October 31, 2012

Abstract:

Methylobacteria are ubiquitous in the biosphere which are capable of growing on C1 compounds such as formate, formaldehyde, methanol and methylamine as well as on a wide range of multi-carbon growth substrates such as C2, C3 and C4 compounds due to the methylotrophic enzymes methanol dehydrogenase (MDH). MDH is performing these functions with the help of a key protein mxaF. Unfortunately, detailed structural analysis and homology modeling of mxaF is remains undefined. Hence, the objective of this research is the characterization and three dimensional modeling of mxaF protein from three different methylotrophs by using I-TASSER server. The predicted model were further optimize and validate by Profile 3D, Errat, Verify3D and PROCHECK server. Predicted and best evaluated models have been successfully deposited to PMDB database with PMDB ID PM0077505, PM0077506 and PM0077507. Active site identification revealed 11, 13 and 14 putative functional site residues in respected models. It may play a major role during protein-protein, and protein-cofactor interactions. This study can provide us an ab-initio and detail information to understand the structure, mechanism of action and regulation of mxaF protein.

Keywords: Methylobacteria, mxaF protein, homology modeling, functional site

Background:

Methylotrophic bacteria are a diverse group of organisms that possess a great number of specialized enzymes that enable them to grow on reduced carbon substrates without carbon–carbon bonds and use these as energy as well as a carbon source. They play an important role in biochemical cycling and possess a potential for use in bioremediation [1]. The members of Methylobacterium have a property to oxidize methanol due to the presence of methanol dehydrogenase (MDH), a pyrroloquinoline quinone (PQQ)-linked protein with an α2β2 tetramer structure [2], known as Methylotrophy, a capacity to aerobically utilize single carbon (C1) compounds as a sole source of carbon and energy by a bacterial metabolic pathway used to assimilate carbon or retrieve energy [3]. Until recently, the ability to oxidize methanol by gram-negative bacteria has been attributed almost exclusively to the MDH enzyme encoded by mxaF1. The study of amino acid sequences of mxaF protein show that several key amino acids that are required for the MDH enzyme activity are located in the deduced MxaF peptide [4]. The mxaF genes are well conserved among different classes of proteobacteria (alpha, beta, and gamma) in terms of both gene clustering and protein sequence identity [3, 5], suggesting a monophyletic origin for the mxaF (mox) encoded methanol oxidation machinery. Based on its
conservation, mxaF has served as a genetic marker for environmental detection of Methylotrophy [6]. Protein structure prediction is one of most intensely studied subjects in modern computational biology. It could be achieved at both the secondary structure level and the three-dimensional level. These are intimately related to each other. The accuracy of secondary structure prediction has increased gradually over the years as we gain a better understanding of the principles of sequence-structure relationships and the effect of evolution on the proteomes of organisms [7]. But the prediction of the three-dimensional structure of a protein from its amino acid sequence is a challenge that has fascinated researchers in different disciplines for many years. Because Proteins play a key role in almost all biological process like maintaining the structural integrity of the cell, transport and storage of small molecules, catalysis, regulation, signaling and the immune system. Their 3-D structure and functional properties depend intricately upon their structure. As a result there has been much effort, both experimental and computational, in determining protein structures.

Due to the high impact of mxaF protein, it is necessary to extract structural information from sequences, has become increasingly important for bioinformatics research. So a detailed structural analysis of mxaF, a large subunit of methanol dehydrogenase protein has yet to be reported. In this study it was achieved by homology modeling. A comparative and detailed structural analysis of methanol dehydrogenase protein was assessed among facultative, restricted facultative and obligate methylotroph strains. The 3-D structure of methanol dehydrogenase was developed for three methylotrophic strains follow Serine pathway for methylorophic metabolism and was compared.

**Methodology:**
The amino acid sequence of MxaF protein of three different types of methylobacteria was retrieved from the NCBI database (CAI30806, CAA69318 and AAR88789). The interrogatory sequences from mxaF of methylotrophs were skim to find out the related protein structure to be used as a template by the BLAST program against Protein Data Bank database.

**Primary structural analysis**
Expsys’s ProtParam server has been applied for study of physicochemical characterization like theoretical isoelectric point (pI), molecular weight, molecular formula, total number of positive and negative residues, instability index, extinction coefficient, aliphatic index and grand average hydropathy (GRAVY) bonding pattern is predicted by using the tool CYS_REC. The predicted results were shown in Table 1 (see supplementary material).

**Secondary structural analysis**
For the enumeration of the secondary structural features of MxaF protein sequences, PSI>PRED view, a new highly accurate secondary structure prediction method was employed. PSI>PRED incorporates two feed-forward neural networks which perform an analysis on output obtained from PSI-BLAST (position specific iterated BLAST) [17]. Results are shown in Table 2 (see supplementary material).

**Discussion:**
The primary structure of MxaF protein of Methylosinus trichosporium, Hyphomicrobium zavarzini and Methylobacterium podarium was speculated and compared Table 3 (see supplementary material). The computed isoelectric point (pI) for the computed pI value of all three strains are 5.13, 5.78 and 5.46 and is less than 7 (pI<7), reveals the acidic nature of protein. Total predicted negative residues of mxaF protein are more in comparison to positive residues. These results also supported the acidic nature of target proteins of above Methylotrophic strains. Among those Methylotrophic strains, the extinction coefficient of H. zavarzini is high, that indicates the presence of high concentration of Cys, Trp and Tyr which helps in the quantitative study of protein-protein and protein-legend interactions in solution. The sulphide (S-S) bonding pattern has been shown that the M. tricosporium and H. zavarzini have two cystein at position 145, 174 but M. podarium have three cystein at position 3, 126, 145.

These results unveil that the mxaF protein residue of these methylotrophs has high enthalpy at folded state because disulphide bond increase the enthalpy of the folded state by stabilizing local interaction. The thermal stability of mxaF protein was determined by Instability Index. The predicted Instability index of mxaF of target strains were 16.74, 22.65 and 27.18 reveals that mxaF protein is thermostable because Instability index of a protein smaller than 40 make it stable while more than 40 make it unstable. The mxaF protein of above strain are stable at wide range of temperature due to the Higher aliphatic index (51.73, 51.73 57.62) since higher AI indicates increased stability while lower AI indicate increased flexibility in the protein structure. GRAVY (Grand average hydropathy) value of mxaF protein is -0.812, -0.970 and -0.728 expresses the hydrophilicity of protein of target strains due to their lower value because lower value indicates possible better interaction with water. The GRAVY value of H. zavarzini was lower (-0.970) in comparison to other strains, indicates the better solubility of mxaF protein. Most algorithms for protein structure prediction currently in use are based on machine learning techniques in which PSIPRED view has been shown to be capable of achieving an average Q3 score of 76.5%, a highest level of...
accuracy published for any method to date [25]. The secondary structure of mxaF proteins of target methylotrophs were predicted and analyzed by PSIPRED view and were shown in Table 2 (see supplementary material). The results were expressed that all the residues lies under the strands and coil and have only little differences among them. There is no alpha helix found in the predicted structure. It reveals the unfavorable structural property of protein in non-polar solvent. These results also can help in experimental verification of a predicted folding motif because it may be gained by measurements of protein secondary structural elements of which the motif is composed [26].

Three dimensional structure of mxaF protein of three methylotrophic strains were predicted and compared. The comparative protein structure analysis for methylotrophs is still untouched and unavailable. The tertiary structure prediction was performed by I-TASSER server by using the best align template (4aahA). The template was selected to analyze 3-D structure because a high level of sequence identity should guarantee a more accurate alignment between the target sequence and template structure [27]. Out of five generated similar models of the target sequence, the best one have been chosen to employing the criteria of good alignment with template, C-Score, TM score and RMSD values Table 3 (see supplementary material). The developed 3-D model of mxaF protein of methylotrophs was deposited to the PMDB database and their PMDB accession number is given in Table 3. The Predicted models were visualized through Accelry's Discovery Studio visualize 2.5 (Figure 1). The generated contact map of mxaF protein of methylotrophs explains the reduced representation of the target structure that helps in to the superimposition and similarity with other protein. The quality of predicted structures of mxaF were further assessed and confirmed by VERIFY 3D [28] Profile 3D [19] and Errat [20]. The scores (from -1 to +1) were added and plotted for individual residues. The residues falling in the area where the orange line crosses 0.0 have low prediction accuracy and less stable conformation whereas, most of the residues fall above 0.15-0.4 so we can say that the model is of good quality. The stereochemical quality and accuracy of the predicted model of mxaF were evaluated after the refinement process using Ramachandran Map calculation with the PROCHECK program [21]. The Ramachandran plot has been shown a tight clustering of phi ~ -50 and psi ~ -50. In the plot analysis, the residues were classified according to its regions in the quadrangle. The red regions in the graph point out the most allowed regions whereas the yellow regions represent allowed regions. Glycine is restrained by triangles and other residues are represented by squares.

The analysis report of Ramachandran plot concluding phi and psi angles to contribute in conformation of amino acids excluding glycine and proline. PROCHECK analysis of mxaF protein reveals in Ramachandran plot concluding phi and psi angles to contribute in conformation of amino acids excluding glycine and proline with 85.3%, 83.2%, 85.7% residue in most favoured region, 11.3%, 15.4%, 12.3% (16 amino acid) in additional allowed region, 2.7%, 1.3%, 1.9% generously allowed region and 0.7%, 0.0%, 0.0% residue in disallowed region in M. trichosporium, H. zavarzinii and M. podarium respectively. Q-Site Finder server was employed for the prediction of functional sites in the modeled mxaF proteins. Server were detected the 11, 13 and 14 putative functional site residues with significant matches in the modeled protein of H. zarzvinni, M. podarium and M. trichosporium respectively. The putative residues are given in Table 4 (see supplementary material), which could be important for protein interactions and/ or activity of mxaF. The predicted 3-D structures of the target methylotrophs have been shown good stereochemistry among the strains, indicating reasonable good quality. The Overall 3-D structure of mxaF is well conserved among the methylotrophs who differ at the level of their nutrition.

**Conclusion:**
Precise evaluation and modeling of proteins is a major goals and key aspect of computational Biology. The methylotrophs play a vital role in biogeochemical cycling and have potentiality for use in bioremediation due to the mxaF a major sub-unite of MDH protein. So the structural exploration and 3-D model was generated for the first time of three different methylotrophs which varying at nutrition level. It offers an alternative way to obtain structural information well before the structure of the new protein is determined by X-ray crystallography or NMR. Physicochemical and functional studies performed for characterization of mxaF in reaching conclusions about the biochemistry and biological function of the modeled protein. Structure prediction and functional analysis of mxaF will give an insight to the location of these proteins along with site of utilization of methanol. The present study would aid in
detailed molecular mechanism of biostimulation of contaminant biodegradation and biotransformation through mxaF protein of methylotrophs.

Reference:
[1] Altschul, SF et al. Nucleic Acids Res. 1997 25: 3389
[2] Anthony C et al. Biochem J. 1994 304: 665
[3] Chistoserdova L et al. J Bacteriol. 2007 189: 4020
[4] Chistoserdova L et al. J Bacteriol. 2003 185: 2980
[5] De Marco P et al. FEMS Microbiol Lett. 2004 234: 75
[6] Eisenberg D et al. Methods in Enzymol. 1997 277: 396
[7] Gill SC & Von Hippel PH, Anal Biochem. 1989 182: 319
[8] Gundampati RK et al. J Mol Biol. 2012 18: 653 [PMID: 21562828]
[9] Guruprasad K et al. Prot Eng. 1990 4: 155
[10] http://blast.ncbi.nlm.nih.gov/Blast.cgi
[11] http://www.ncbi.nlm.nih.gov/entrez/
[12] Ikai AJ, J Biochem. 1980 88: 1895
[13] Jae Hoon Jeong et al. Mol Cells. 2002 13: 369
[14] John Pelton T & Larry Mclean R, Analytical Biochemistry. 2000 277: 167
[15] Kyte J & Doolittle RF, J Mol Biol. 1982 157: 105
[16] Laskowski RA et al. J App Cryst. 1993 26: 283
[17] Liam J et al. Bioinformatics. 1999 16: 404
[18] Donald M et al. Appl Environ Microbiol. 2008 74: 1305
[19] Singh RP et al. J Applied Sciences in Environmental Sanitation, 2001 6: 357
[20] Rost BJ, Struct Biol. 2001 134: 204
[21] Sekhar PN et al. In silico biology, 2006 6: 0041
[22] Suyama M et al. J Mol Biol. 1997 277: 396
[23] Wedemeyer WJ, Biochemistry. 2000 39: 7032
[24] http://us.expasy.org/tools/protparam.html
[25] http://linux1.softberry.com/berry.phtml?topic.
[26] http://globin.bio.warwick.ac.uk/psipred.
[27] http://toolkit.tuebingen.mpg.de/modeller/verify3d/
[28] Laurie A & Jackson R, Bioinformatics. 2005 21: 1908 [PMID: 15701681]

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

Edited by P Kanguane
Citation: Singh et al. Bioinformation 8(21): 1042-1046 (2012)
### Table 1: Physico-chemical characterization three Methylotrophic strain

| Organism                     | Accession number | M.F.   | M.W.     | p.I. | -R | +R | EC  | II  | AI  | GRAVY |
|------------------------------|------------------|--------|----------|------|----|----|-----|-----|-----|--------|
| Methylosinus trichosporium  | CAI30806.1       |        | 20.78    | 5.13 | 28 | 19 | 33920| 16.74| 57.62| -0.812 |
| Hyphomicrobium zavarzinii   | CA69318.1        |        | 20.55    | 5.78 | 26 | 21 | 55920| 22.65| 51.73| -0.970 |
| Methylobacterium podarium   | AAR88789.1       |        | 20.92    | 5.46 | 27 | 21 | 46410| 27.18| 56.20| -0.728 |

### Table 2: Secondary Structure Analysis of three Methylotrophic strain

| Organisms                     | Accession no. | Alpha Helix | Strand | Random coil |
|------------------------------|---------------|-------------|--------|-------------|
| Methylosinus trichosporium   | CAI30806.1    | 00          | 55     | 126         |
| Hyphomicrobium zavarzinii    | CA69318.1     | 00          | 52     | 132         |
| Methylobacterium podarium     | AAR88789.1    | 00          | 57     | 122         |

### Table 3: 3-D Model Analysis of Methylotrophic strain

| Name                         | PMDB Accession No. | C-score | Exp.TM-Score | Exp. RMSD | No.of decoys | Cluster density |
|------------------------------|--------------------|---------|--------------|-----------|--------------|----------------|
| Methylosinus trichosporium   | PM0077505          | 1.75    | 0.96+0.05    | 1.8+1.5   | 12600        | 1.2500         |
| Hyphomicrobium zavarzinii    | PM0077506          | 1.75    | 0.96+0.05    | 1.8+1.5   | 12600        | 1.2500         |
| Methylobacterium podarium     | PM0077507          | 1.76    | 0.96+0.05    | 1.8+1.5   | 12600        | 1.2500         |

### Table 4: putative functional site residues of Methylotrophic strain

| Name                         | Putative functional site residues                  |
|------------------------------|----------------------------------------------------|
| Methylosinus trichosporium   | GLY92, ILE93, SER106, A LA 107, A SN 108, LYS109, VAL114, LYS118, GLN119, VAL120, LEU122 |
| Hyphomicrobium zavarzinii    | ILE94, SER107, A LA 108, A SN 109, LYS110, A SP113, VAL115, A SN 116, PHE118, LYS1219, SER120, VAL121, LEU123 |
| Methylobacterium podarium     | A SN 91, GLY92, ILE93, SER106, A LA 107, A SP108, LYS109, VAL114, PHE117, LYS118, TH R119, VAL120, A SP121, LEU122 |