

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

Raghvendra Pratap Singh<sup>1</sup>\*, Ram Nageena Singh<sup>1</sup>, Manish K Srivastava<sup>1</sup>, Alok Kumar Srivastava<sup>1</sup>, Sudheer Kumar<sup>1</sup>, Ramesh Chandra Dubey<sup>2</sup> & Arun Kumar Sharma<sup>1</sup>

<sup>1</sup>Microbial Genomics Lab, National Bureau of Agriculturally Important Microorganisms, Kusmaur, Maunath Bhanjan, Uttar Pradesh-275101, India; <sup>2</sup>Department of Botany and Microbiology, Gurukul Kangri University, Haridwar, Uttarakhand-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, Verify3-D 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 biogeochemical 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  $\alpha_2\beta_2$  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 mxa (mox) encoded methanol oxidation machinery. Based on its

conservation, *mx*aF has served as a genetic marker for environmental detection of Methylophily [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 *mx*aF protein, it is necessary to extract structural information from sequences, has become increasingly important for bioinformatics research. So a detailed structural analysis of *mx*aF, 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 methylophily strains. The 3-D structure of methanol dehydrogenase was developed for three methylophily strains follow Serine pathway for methylophily 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) [8]. The interrogatory sequences from *mx*aF of methylophily were skim to find out the related protein structure to be used as a template by the BLAST program [9] against Protein Data Bank database.

## Primary structural analysis

ExPasy's ProtParam server [10], has been applied for study of physiochemical characterization like theoretical isoelectric point (pI), molecular weight, molecular formula, total number of positive and negative residues, instability index [11], extinction coefficient [12], aliphatic index [13] and grand average hydropathy (GRAVY) [14]. The sulphide (S-S) bond pattern is predicted by using the tool CYS\_REC [15]. 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, PSIPRED view [16], a new highly accurate secondary structure prediction method was employed. PSIPRED 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).

## Homology modeling, Structure refinement and identification of functional site

3-D model of *MxaF* protein sequence of three different methylobacterium was generated by I-TASSER, a web based server. Further the models were evaluated by VARIFY 3D [18], Profile 3D [19] and Errat [20], to check the correctness of the overall fold/structure, errors over localized regions and stereochemical parameters such as bond lengths and angles. Visualization and protein contact map of target proteins were carried out by Accelry's Discovery Studio software. Structural validation of target proteins model were done by PROCHECK which determine stereochemical aspects along with main chain and side chain parameters with comprehensive analysis. The shows that various The *MxaF* residues falling under allowed, favoured and in disallowed regions was predicted by Ramachandran plot perform by PROCHECK [21]. Structure based protein functional site of *mx*aF of three methylophily were predicted by Q-site Finder [22].

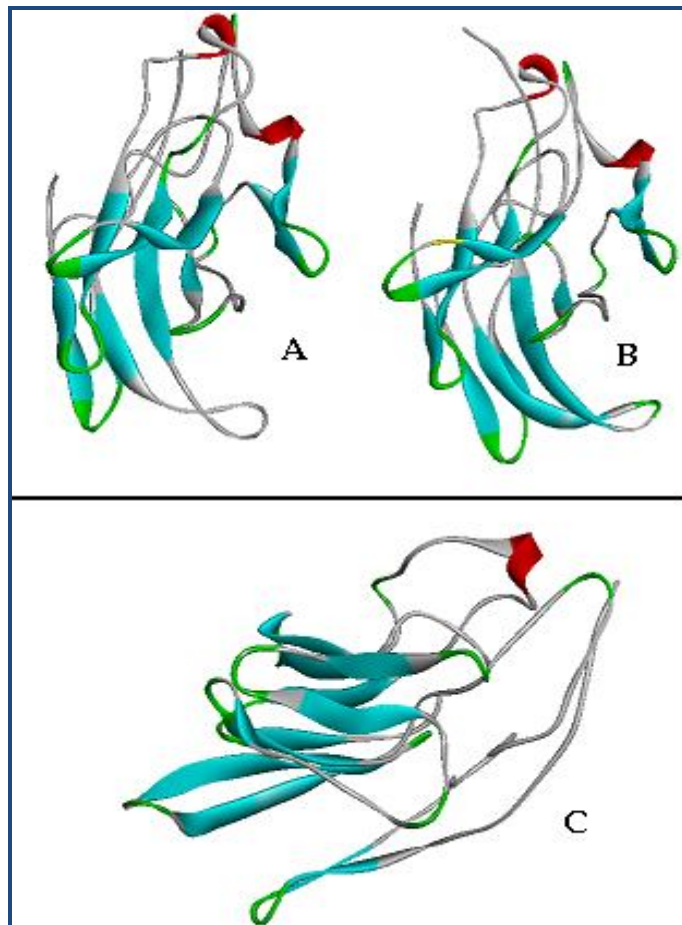
## Discussion:

The primary structure of *MxaF* protein of *Methylosinus trichosporium*, *Hyphomicrobium zavarzinii* 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 [23]. Total predicted negative residues of *mx*aF protein are more in comparison to positive residues. These results also supported the acidic nature of target proteins of above Methylophily strains. Among those Methylophily strains, the extinction coefficient of *H. zavarzinii* 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-ligand interactions in solution. The sulphide (S-S) bonding pattern has been shown that the *M. trichosporium* and *H. zavarzinii* have two cystein at position 145, 174 but *M. podarium* have three cystein at position 3, 146, 175.

These results unveil that the *mx*aF protein residue of these methylophily has high enthalpy at folded state because disulphide bond increase the enthalpy of the folded state by stabilizing local interaction [24]. The thermal stability of *mx*aF protein was determined by Instability Index. The predicted Instability index of *mx*aF of target strains were 16.74, 22.65 and 27.18 reveals that *mx*aF protein is thermostable because Instability index of a protein smaller than 40 make it stable while more than 40 make it unstable [11]. The *mx*aF 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 indicate increased stability while lower AI indicate increased flexibility in the protein structure [23]. GRAVY (Grand average hydropathy) value of *mx*aF 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. zavarzinii* was lower (-0.970) in comparison to other strains, indicates the better solubility of *mx*aF protein.

Most algorithms for protein secondary 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 *mx*aF 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 unfavored 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].



**Figure 1:** (A) *Hyphomicrobium zavarzinii*; (B) *Methylobacterium podarium*; (C) *Methylosinus trichosporium*

Three dimensional structure of *mx*aF 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 *mx*aF 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 *mx*aF 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 *mx*aF were further assessed and confirmed by VARIFY 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 *mx*aF 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 *mx*aF 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 *mx*aF 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 *mx*aF. 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 *mx*aF 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 *mx*aF 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 *mx*aF in reaching conclusions about the biochemistry and biological function of the modeled protein. Structure prediction and functional analysis of *mx*aF 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 Model.* 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 Mc *et al. Appl Environ Microbiol.* 2008 **74**: 1305  
[19] Singh RP *et al. J Applied Sciences in Environmental Sanitation,* 2011 **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 Evol.* 1997 **1**: S163 [PMID: 9071025]  
[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]

Edited by P Kanguane

Citation: Singh *et al.* Bioinformation 8(21): 1042-1046 (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 credited



## Supplementary material:

**Table 1:** Physico-chemical characterization three Methylo-trophic strain

Organism	Accession number	M.F.	M.W.	p.I.	-R	+R	EC	II	AI	GRAVY
<i>Methylosinus trichosporium</i>	CAI30806.1	C <sub>929</sub> H <sub>1375</sub> N <sub>249</sub> O <sub>281</sub> S <sub>8</sub>	20.78	5.13	28	19	33920	16.74	57.62	-0.812
<i>Hyphomicrobium zavarzinii</i>	CAA69318.1	C <sub>913</sub> H <sub>1359</sub> N <sub>253</sub> O <sub>276</sub> S <sub>8</sub>	20.55	5.78	26	21	55920	22.65	51.73	-0.970
<i>Methylobacterium podarium</i>	AAR88789.1	C <sub>924</sub> H <sub>1397</sub> N <sub>253</sub> O <sub>280</sub> S <sub>12</sub>	20.92	5.46	27	21	46410	27.18	56.20	-0.728

**Table 2:** Secondary Structure Analysis of three Methylo-trophic strain

Organisms	Accession no.	Alpha Helix	Strand	Random coil
<i>Methylosinus trichosporium</i>	CAI30806.1	00	55	126
<i>Hyphomicrobium zavarzinii</i>	CAA69318.1	00	52	132
<i>Methylobacterium podarium</i>	AAR88789.1	00	57	122

**Table 3:** 3-D Model Analysis of Methylo-trophic strain

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

**Table 4:** putative functional site residues of Methylo-trophic strain

Name	Putative functional site residues
<i>Methylosinus trichosporium</i>	GLY92, ILE93, SER106, ALA107, ASN108, LYS109, VAL114, LYS118, GLN119, VAL120, LEU122.
<i>Hyphomicrobium zavarzinii</i>	ILE94, SER107, ALA108, ASN109, LYS110, ASP113, VAL115, ASN116, PHE118, LYS1219, SER120, VAL121, LEU123
<i>Methylobacterium podarium</i>	ASN91, GLY92, ILE93, SER106, ALA107, ASP108, LYS109, VAL114, PHE117, LYS118, THR119, VAL120, ASP121, LEU122