

Computational fishing and structural analysis of MIPS protein from two important plant groups

Anjan Hazra

Post Graduate Dept. Of Botany, Barasat Govt. College, Kolkata, India

Email address: hazranjan93@gmail.com

Keywords: Myo inositol 1-phosphate synthase; computational protein fishing; physicochemical properties; homology modelling; Ramachandran Plot; Phylogeny

ABSTRACT. Myo Inositol 1-Phosphate Synthase (MIPS), which catalyzes the first step of inositol metabolism, has been reported from a diverse range of organism like bacteria to human including different groups of plants and animals. The present work is carried out to explore and analyze structural forms of the respective MIPS proteins from complete sequenced genome or proteome available on database of one representative from two important plant groups viz. bryophyte (*Physcomitrella patens*) and pteridophyte (*Selaginella moellendorffii*). Previously reported characteristic MIPS sequences was used to identify its homolog ones from those members under study. The explored sequences compared with a number of MIPS varieties from other plant members to study the conserveness or evolution of the protein/enzyme. ProtParam tool provided necessary theoretical physicochemical data of the predicted proteins, the three-dimensional structures were predicted through homology modelling with identified amino acid data. Structural evaluation and stereochemical analyses were performed using ProSA-web displaying Z-scores and Molprobity visualising Ramachandran plot.

1. INTRODUCTION

A rapid and fruitful development of structural and functional genomics and proteomics during the period of twentieth century's revolution in molecular and computational biology unfolded the chapter of proper storage, retrieval and analysis of genomic proteomic or transcriptomic data for all time. And there are a several options to get rid of a particular gene or protein or else like of interest which further can be analysed for different purpose. In very recent time computational determination of structural and functional aspects of a protein/gene derive an equal importance as like wet lab experimental outputs. For example, a number of computational tools are developed and evolved on regular basis to predict the secondary or quaternary structure of a protein or an enzyme and its validation that can be completed within a very short period of time.

Myo inositol 1-phosphate synthase (MIPS) is the key functional enzyme of myo-Inositol biosynthesis pathway. It belongs to NAD-5 super family and act as an NAD⁺ binding oxidoreductase (Majumder et al 1997) and found to be present throughout a evolutionarily wide range of living organisms (Majumder et al, 2003). The distribution and evolution of MIPS genes from a probable ancestor to cyanobacteria, eubacteria and archaea, and ultimately to higher eukaryotes such as plants and animals, indicate that MIPS started evolving from different sources (Majumder et al., 2003) and that MIPS enzyme activity is distributed in evolutionary diverse phyla, from eubacteria, archaebacteria, cyanobacteria, algae, fungi to higher plants and animals. In plants the enzyme has been reported from many organisms belonging to algae, fungi, gymnosperm, angiosperm etc. Although a few previous work has been carried out from bryophyte (Chhetri et al 2006) and pteridophyte members (Chhetri et al 2005,2006,2008, Jha et al 2013) but still those does not reveal any sequence or structural data from any member of the mentioned groups and that made interested in the current investigation through In Silico method. A large number of computational tools are available for making predictions regarding the identification and structure prediction of proteins. When a whole genome data are reported from a bryophyte (*Physcomitrella patens*) and

pteridophyte (*Selaginella moellendorffii*) those were used to match and find out the MIPS sequence followed by the bioinformatics methodology known as Protein fishing previously followed by many researchers and in recent Basu et al (2015).

The enzyme MIPS probably exists in homotetrameric form in native state as evidenced by crystallography (Norman et al, 2002; Stieglitz et al 2004 etc). It involves in internal oxidoreduction and cyclization & needs NAD⁺ as a cofactor which is mostly reconverted. MIPS convert glucose 6-phosphate to L-myo-inositol 1-phosphate in an NAD⁺ dependent manner. The NAD⁺ binding in eukaryotic MIPS involves a Rossman fold characterized by a GXGGXXG motif typical of an oxidoreductase (Kleiger and Eisenberg 2002).

Crystal structure analysis of MIPS from *Saccharomyces cerevisiae* revealed that each monomer of the homo-tetrameric MIPS has three functionally important structural domains namely the NAD binding Rossman fold, the catalytic binding site and the core domain. Those studies also exemplified induced fit model for binding of the substrate with the catalytic domain of the enzyme. (Stein & Geiger, 2002).

2. METHODOLOGY

2.1 Subject Sequences: Two different MIPS variety sequences, one from the lower plant group green algae (*Chlamydomonas reinhardtii*, Accession No. EDP05930) and another from halophytic monocotyledon (*Porteresia coarctata*, Accession No. AAP74579) were downloaded from NCBI Genpept database (www.ncbi.nlm.nih.gov) and used for sequence homology searching in target plants proteome.

2.2 Protein fishing by local alignment: Those two MIPS amino acid sequences were served as query for BLASTp against whole Proteome of each of our two target plant *i.e.* *Physcomitrella patens* and *Selaginella moellendorffii* using the California Department of Energy, JGI's database Phytozome v10 (<http://phytozome.jgi.doe.gov/pz/portal.html#!search?show=BLAST>). Working comparison matrix was BLOSUM62. The proteins with top hit on BLASTp or with significant similarity were fished out and reported as probable MIPS variety sequences of those target organisms.

2.3 Protein's physicochemical properties: The theoretical physicochemical properties were computed using the Expasy's Prot-Param (Gasteiger *et al.*, 2005) (<http://us.expasy.org/tools/protparam.html>) prediction server (Table 1). The output parameters were theoretical isoelectric point (pI), molecular weight, total number of positive and negative residues, extinction coefficient (Gill and Von Hippel, 1989), instability index (Guruprasad *et al.*, 1990), aliphatic index (Ikai *et al.*, 1980) and grand average hydrophathy (GRAVY) (Kyte and Doolittle, 1982).

2.4 Multiple sequence alignment: To confirm the presence of previously reported conserved regions of the protein a multiple sequence alignment among the diverse variety of MIPS protein from different group of plants was laid down with the help of online server (<http://multalin.toulouse.inra.fr/multalin>).

2.4 Phylogenetic analysis of sequences: Phylogenetic tree was generated by the UPGMA method using MEGA 6 software (Fig. 4) followed by validation with 100 replicates of Bootstrapping.

2.5 Quaternary structure prediction: The explored sequence's three dimensional structure prediction was performed using Modeller 9.12. In brief, this method aligns a query sequence to one or more template structures as determined by crystallization/X-ray diffraction, or NMR spectrometry. Suitable templates (PDB id 1LA2) were selected through pdbBLAST at NCBI. E-value and sequence similarity (Table 2) was carefully considered in this respect. As the protein exist in homotetrameric form in its native state so the top hit of homotetrameric structure report was

selected for modelling template. The NAD⁺ Ligand in the template structure have been transferred to the model because the ligand are annotated as biologically relevant to this particular protein or enzyme. Besides, to establish the model it should be in contact with Ligand and there should have no clash between the ligand and protein when interacting residues with the ligand are conserved between the template-target. Molecule visualisation was done with Chimera 1.10 (Pettersen et al 2004)

2.6 Evaluation of theoretical model: To check the stereochemical stability the psi and phi dihedral angles of each amino acid residues in the protein structures were determined with Ramachandran plots (Ramachandran *et al.*, 1963) analysis from Molprobit (Chen et al 2010) server (<http://molprobit.biochem.duke.edu/>). The predicted *in silico* models were also evaluated using the ProSA web server, ProSA calculates the overall quality score. The output provides a z-score for the model that indicates the overall model quality; this value was determined from the plot during prediction.

2.7 Submission to PMDB database: The models generated were successfully submitted in Protein model database (PMDb) (<https://bioinformatics.cineca.it/PMDB/>).

3. RESULT AND DISCUSSION:

Both the two way of BLASTp *C. Reinhardtii* and *P. coarctata* (Fig. 1) against the *Selaginella moellendorffii* proteome showed a single hit of resultant protein sequence. It was **Gene 270269** which encode 511 residue long protein that found to be 72% identical having 89% positive similarity with myo Inositol 1-Phosphate synthase of *C. reinhardtii*. On the other hand this was 67.1% identical with 79% positive similarity in terms of *P. coarctata* belonged variety of mIPS.

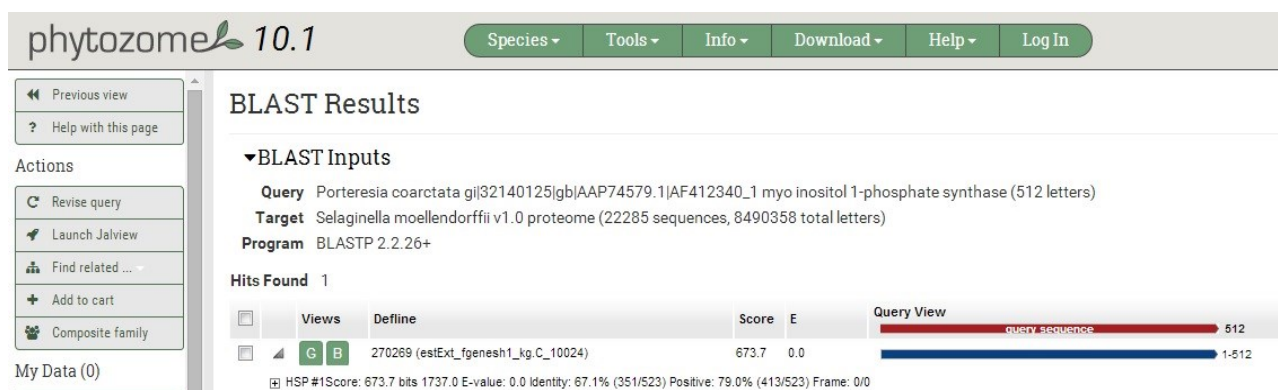


Fig 1: Snapshot of BLASTp result against *Selaginella moellendorffii* proteome

Likewise when the same methodology was followed in case of *Physcomitrella patens*, interestingly we found two genes (Fig. 2) from two different chromosome actually showed almost equal strength of similarity. One is **Phpat.004G003200** which have 66.7% identity & 78% positive similarity with both *C. reinhardtii* and *P. Coarctata* MIPS. Another is **Phpat.003G121000** that have 66.8% identity & 77.3% positive similarity with both *C. reinhardtii* and *P. coarctata*.



Fig 2: Snapshot of BLASTp result against *Physcomitrella patens* proteome

Now the explored amino acid sequences can be considered for further analysis after they named MIPSsm and MIPSpp respectively. Computational tools provide researchers to understand physico-chemical and structural properties of proteins. The theoretical calculation of different useful parameters (Table 1) like mol. Wt., extinction coefficient, isoelectric point etc reveal an overview on the primary structure and stability of the proteins.

Table 1: Physico-chemical properties of MIPSsm and MIPSpp

Subject Protein	Length	negatively charged residues (Asp + Glu)	positively charged residues (Arg + Lys)	M. wt.	Extinction coefficients	Theoretical Isoelectric point	Aliphatic index	Instability index:	Grand average of hydropathicity
MIPSpp	511	64	51	56776.0	51590	5.41	88.88	32.55	-0.209
MIPSsm	511	63	55	56767.9	53205	5.73	87.32	38.79	-0.229

MIPSsm have a few more positively charged residues (Arg+Lys) and a greater instability index than the MIPSpp. Isoelectric point (pI) is the pH value at which the molecule carries no charges or the negative and positive charges are equal. Proteins are found to be stable and compact at their pI. The computed pI value ($pI < 7$) will be useful for developing buffer systems for purification by isoelectric focusing method.

Conserved domains: The Multiple Sequence Alignment of protein is a powerful tool for grouping proteins into families and allows subsequent analysis of evolutionary issues (Balasubramanian *et al.*, 2012). The comparison of homologous protein sequences is the most effective means of identifying common active sites or binding domains. The multiple sequences alignment of amino acids with other group of plant members MIPS variety sequence (Fig. 3) obtained from database revealed some distinctly conserved domains that supports the earlier observation highlighted by Majumder *et al.*, 2003 which are essential for functional property of the protein. One of them is the most important Rossmann fold sequence GXGGXXG type that functions as a NAD⁺ binding domain. The domain lies in between 75 to 85 no. residue in the MSA image (Fig. 3).

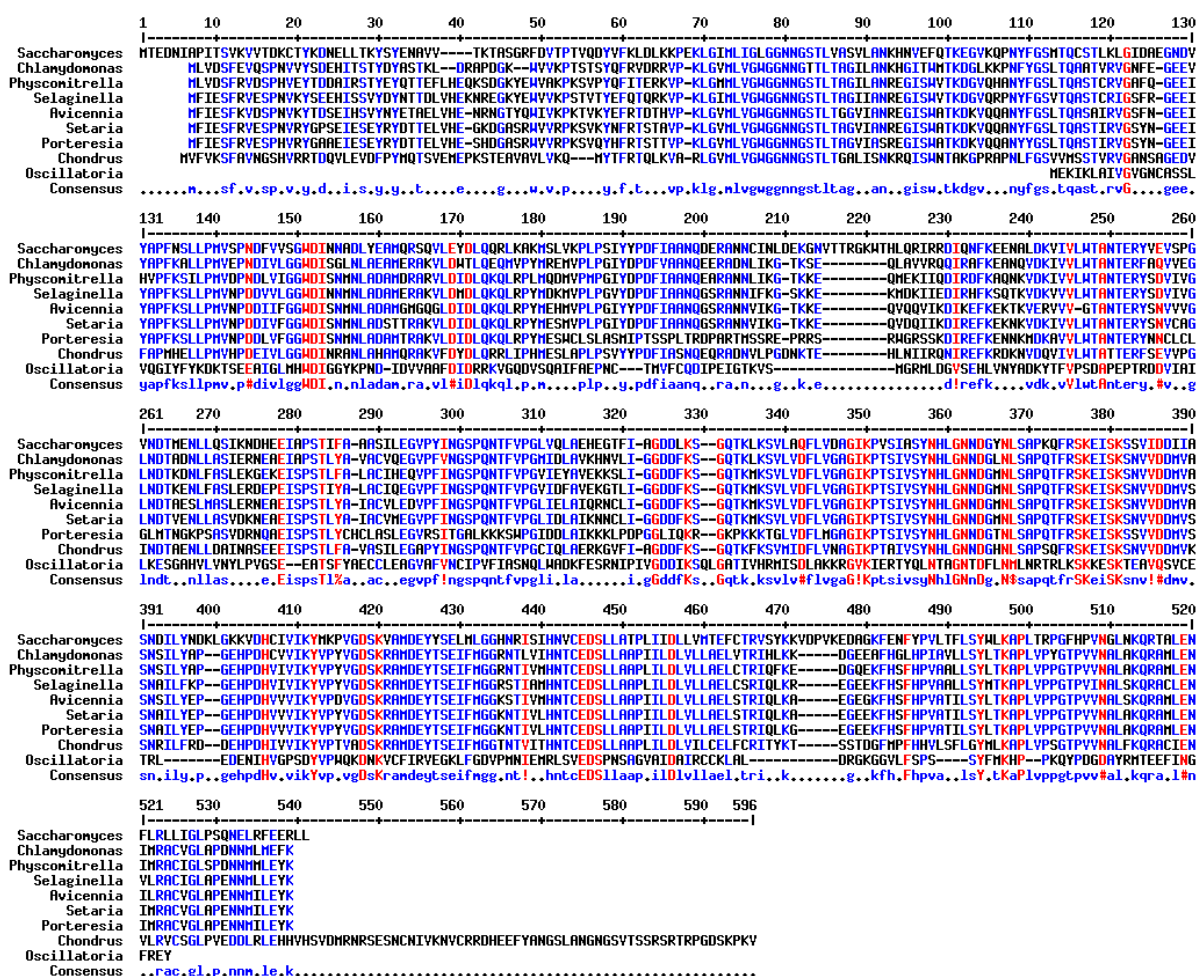


Fig. 3: Multiple sequence alignment of MIPS proteins among different plant, yeast and the subject sequences from *Selaginella* & *Physcomitrella* under study

Phylogenetic analysis: Various sequence information of gene or protein offers the possibility to investigate the gene function *in vivo* with parallel phylogenetic analyses of multigene families to address their evolution within and across species (Vincentz *et al.*, 2003). Within families, the protein structure and catalytic residues that determine the substrate specificity are generally conserved.

After confirmation of the explored proteins existence in NAD-5 super family with the help of CD search tool at NCBI, phylogenetic analysis of MIPS proteins among selected organisms (Fig. 4) also provided powerful information about ancestry of the protein. According to result the cyanobacteria *Oscillatoria* and Yeast forms an out-group and the members under current study are found to be in the same branch which suggests a common ancestor that conforms to the common plant taxonomical lineage.

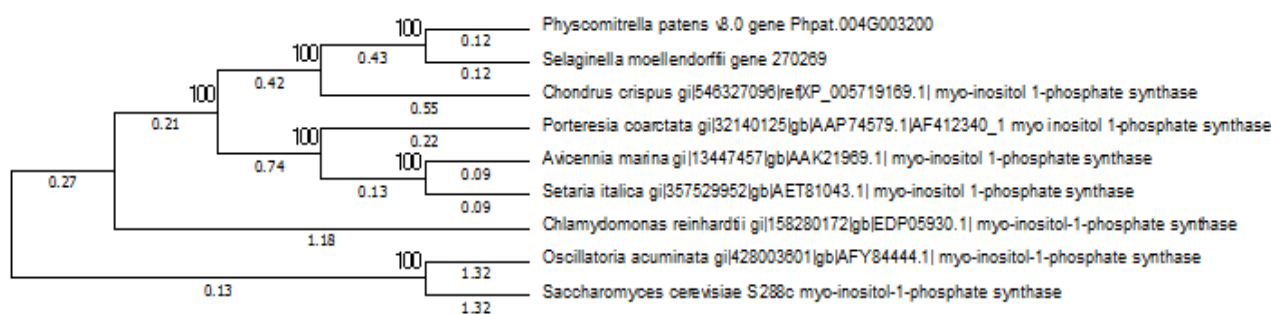


Fig. 4: Evolutionary relationships among taxa in respect to MIPS

The evolutionary history of the MIPS protein among different plants and yeast and cyanobacteria was inferred using the UPGMA method [Sneath and Sokal 1973]. The optimal tree with the sum of branch length = 7.32697491 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches [Felsenstein, 1985]. The evolutionary distances were computed using the Poisson correction method [Zuckerandl and Pauling, 1965] and are in the units of the number of amino acid substitutions per site. The analysis involved 9 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 364 positions in the final dataset. [Data adopted from: MEGA6 software]

Homology modelling: Protein 3D structure is important in understanding protein interactions, function and their localization [Jones 2001]. Quaternary structure plays an important role in defining protein function by facilitating allostereism and cooperativity in the regulation of ligand binding (Matthew *et al.*, 2001). Homology modeling is the most common structure prediction method. In this study, the quaternary structures of *Physcomitrella* and *Selaginella* MIPS proteins were predicted by considering the biological assembly that provided the functional state of the protein. The NAD⁺ ligands also incorporated in time of model generation (Fig. 5) to fix the functionality of the protein with its structure. NAD⁺ characteristically bound to the GWGGNNG domain which exists in a alpha-helix of each monomer.

Table 2: Summarized values after PDB-BLAST for template selection

Model Protein	PMDB id	Template	Sequence similarity with template	Sequence identity with template	E value
MIPSpp	PM0080022	1LA2	99	51	1e-168
MIPSsm	PM0080023	1LA2	99	53	3e-174

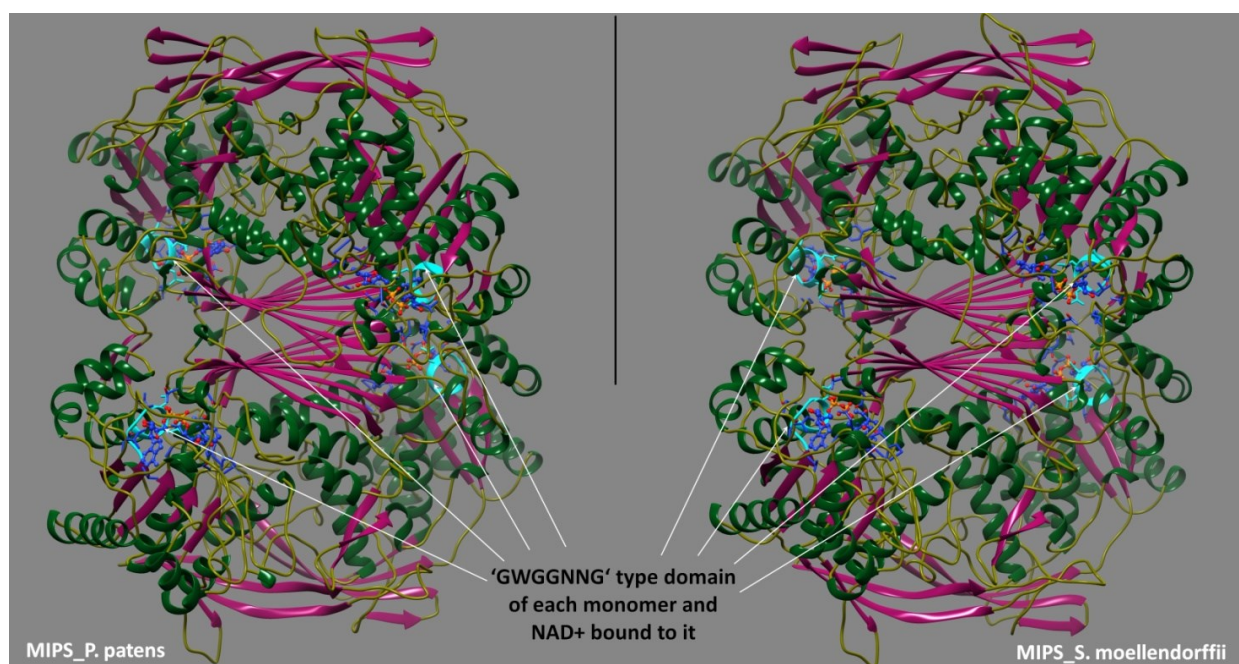


Fig. 5: The quaternary structures of MIPSsm and MIPSpp generated by Modeller and visualized by Chimera (For comparison both two models are attached in a single background)

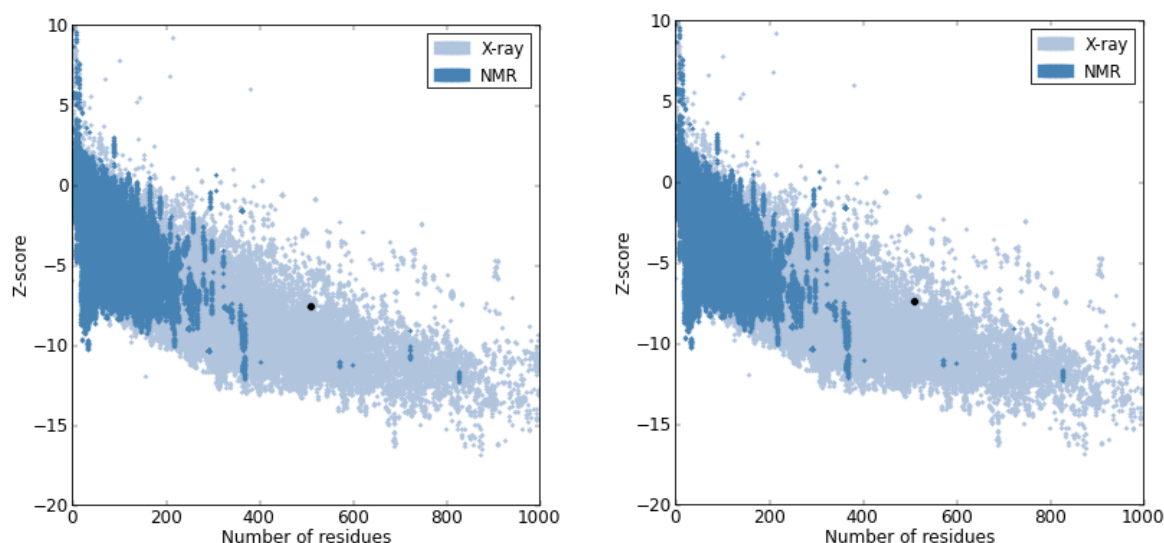


Fig 6: Screenshot of ProSA-web z-score plot (Left- MIPSpp, Right- MIPSsm)

The Z-score value, a measure of model quality that predicts the total energy of the structure (Wiederstein and Sippl, 2007), was determined for the MIPSsm & MIPSpp with the help of PROSA server. Z-score also measures the deviation of total energy of the structure with respect to an energy distribution derived from random conformations. The calculated scores i.e. both -7.66 for MIPSpp and -7.4 for MIPSsm are well within the range of scores (Fig. 6) typically found for proteins of similar size.

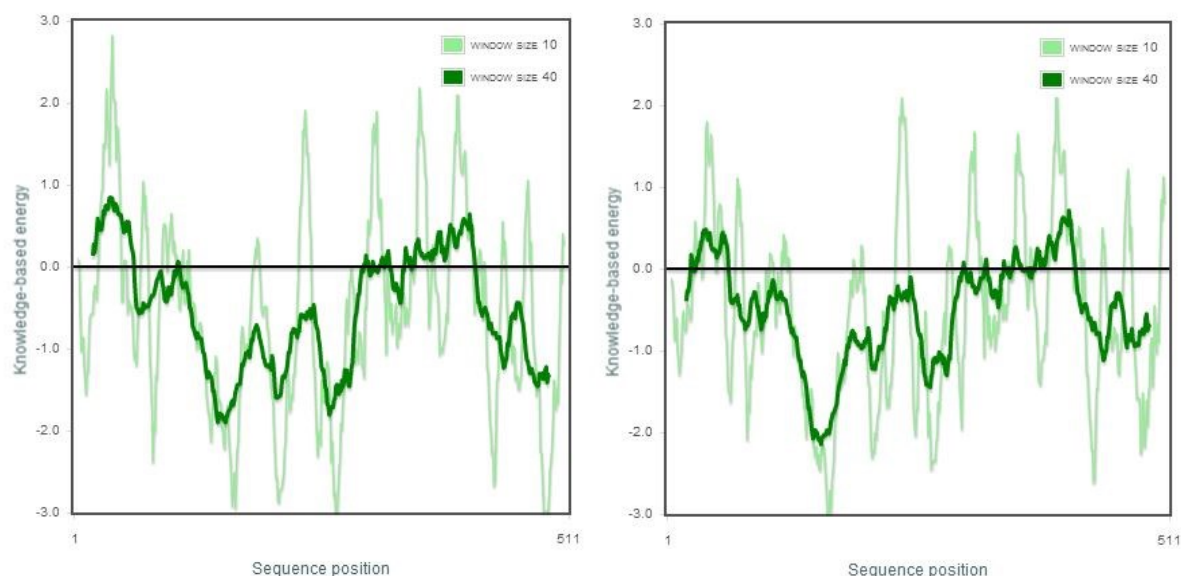
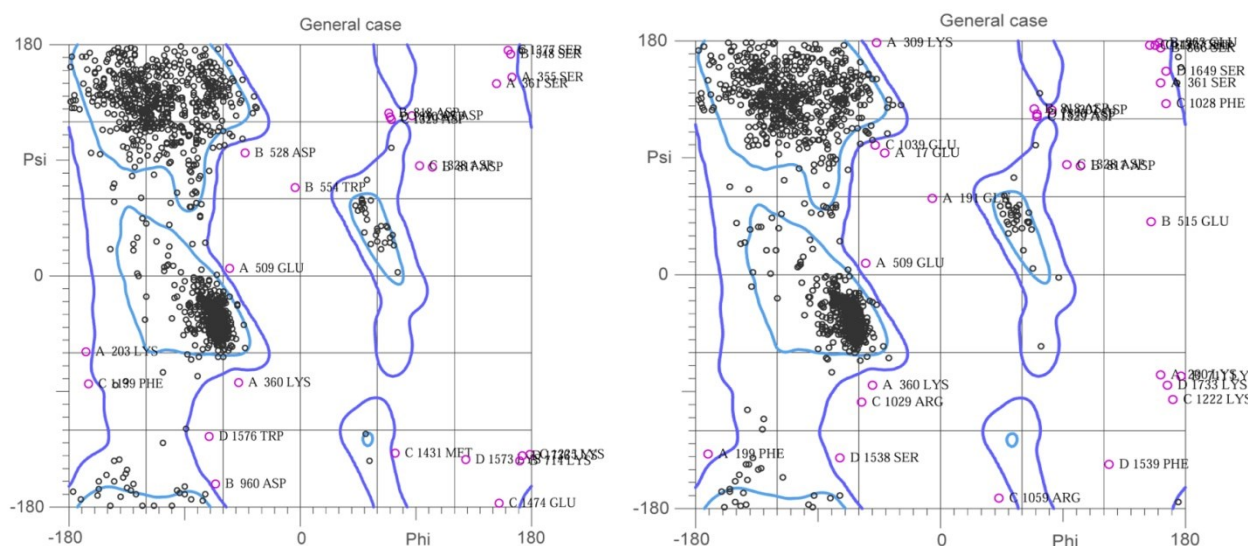


Fig 7. Screenshot of ProSA-web plot of residue scores (Left- MIPSpp, Right- MIPSsm)

The residue score plot (Fig. 7) shows local model quality by plotting energies as a function of amino acid sequence position i . In general, positive values correspond to problematic or erroneous parts of the input structure. A plot of single residue energies usually contains large fluctuations and is of limited value for model evaluation. Hence the plot is smoothed by calculating the average energy over each 40-residue fragment $s(i, i+39)$, which is then assigned to the 'central' residue of the fragment at position $i+19$ (thick line). A second line with a smaller window size of 10 residues is shown in the background of the plot (thin line) [adopted from ProSA web server]. The plots indicate that in both the cases the modelled structures are found to have good stability and overall quality.

Table 3: Stereo chemical properties of modelled proteins

Model Protein	PMDB id	residue in favoured region	Outlier	Z- score
MIPSpP	PM0080022	93% (1888/2031)	35/2031	-7.6
MIPSSm	PM0080023	92.4% (1878/2032)	46/2032	-7.44

**Fig 8:** Screenshot of Molprobrity Ramachandran Plot (Left- MIPSpP, Right- MIPSSm)

Per Residue geometry and the overall geometry of the protein structures was analyzed with Ramachandran plots (Ramachandran *et al.*, 1963). In a polypeptide, the main chain N-C α and C α -C bonds are relatively free to rotate. These rotations are represented by the torsion angles ϕ and ψ respectively. Fig. 8 shows the quantitative evaluation of *Physcomitrella* and *Selaginella* MIPS structures generated using Ramachandran plots provided by the Molprobrity web server. Both *Physcomitrella* and *Selaginella* MIPS had almost equal number of residues in favoured (>92%) regions. The prominence of residues in the right handed α -helix and β -sheet regions suggested that the structure was rigid with more right helices. *Physcomitrella* MIPS showed higher quality compared to the *Selaginella* MIPS protein in respect to having fewer residues in the outlier region and more residues in the allowed region (Table 3). *Selaginella* MIPS had 97.74% of residues in the allowed region. More residues were observed in the allowed region of *Physcomitrella* (98.28%). There were more outliers in the *Selaginella* MIPS (2.26%) than in *Physcomitrella* (1.72 %). These values were almost similar with the the template X ray crystallography structure present in Protein Data Bank. After validation 3D models were successfully submitted to PMDB with id **PM0080022** and **PM0080023**.

4. CONCLUSION

When the enzyme myo-inositol 1-Phosphate synthase is crucially important in diversification evolution and comparison study, the void of sequence data from the previously mentioned important group can at least be mortally fulfilled by the referred sequence of current study. Furthermore it is well known that due to low resolution and high R factors of many crystal structures, homology modelling is now regarded as an important technique for not only in obtaining the 3D-structure of a putative protein but also for refining the existing low accuracy experimental

structures. The above analysis through homology models may provide an insight to simulations and related computational studies such as drug docking calculations and protein-drug or enzyme-target interactions to obtain useful insights into structure–function relationships. The partial phylogenetic analysis including the reported data and its positive inference finally adds a strong support in current method of study.

Acknowledgement

The major credit goes to Dr. Paramita Nandy (Datta) of PG Dept. Of Botany, BGC for introducing the analytical works on MIPS, I am also grateful to Sayak Ganguly, Aditi Gangopadhyay and Hirak Chakraborty of DBT-BIF, Presidency University, Kolkata for instructing me in different bioinformatics tools.

Reference

- [1]. Basak A, Jha T. B. & Adhikari J., Biosynthesis of myo-inositol in lycopods: characteristics of the pteridophytic l-myo inositol 1-phosphate synthase and myoinositol- 1-phosphate phosphatase from the strobili of *Lycopodium clavatum* and *Selaginella monospora*, *Acta Physiol Plant* 34 (2012) 1579-1582
- [2]. Basu P., Ganguli S., Gupta S., Datta A., Exploring Computational Protein Fishing (CPF) to identify Argonaute Proteins from Sequenced Crop Genomes *International Letters of Natural Sciences Vol 33* (2015) pp 27-36
- [3]. Balasubramanian J, Shahul Hammed MK, Tamilselvan R and Vijayakumar N., Artificial neural network: A forecast in pharmaceutical science. *Nerve* (2012) 1:7-12
- [4]. Castrignanò T, D'Onorio P, Meo D, Cozzetto D, Talamo I G, and Tramontano A., The PMDB Protein Model Database Nucleic Acids Research Volume 34, Issue suppl (2005) Pp. D306-D309
- [5]. Chen et al., MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica D* 66 (2010) 12-21.
- [6]. Chhetri, D. R, Mukherjee A. K, Adhikari J., Partial purification and characterization of L-myo-inositol-1-phosphate synthase of pteridophytic origin. *Acta Physiol. Plant.*, 28 (2006) 101-107.
- [7]. Chhetri D. R, Choudhuri M, Mukherjee AK, Adhikari J., L-myo-inositol-1-phosphate synthase: partial purification and characterization from *Gleichenia glauca*. *Biol. Plant.* 49 (2005) 59-63.
- [8]. Chhetri D. R, Yonzon S, Mukherjee A. K, Adhikari J., L-myo-inositol-1-phosphate synthase from *Marchantia nepalensis*: partial purification and properties, *Gen. Appl. Plant physiology*, (2006) 32(3-4), 153-164
- [9]. Corpet F., 1988, Multiple sequence alignment with hierarchical clustering" *Nucl. Acids Res.*, 16 (22), 10881-10890
- [10]. Eswar N., Marti-Renom M. A., Webb B., Madhusudhan M. S., Eramian D., Shen M., Pieper U., Sali A., Comparative Protein Structure Modeling With MODELLER. Current Protocols in Bioinformatics, John Wiley & Sons, Inc., Supplement 15 (2006) 5.6.1-5.6.30
- [11]. Felsenstein J., Confidence limits on phylogenies: An approach using the bootstrap *Evolution* 39 (1985) 783-791.
- [12]. Gasteiger E., Protein Identification and Analysis Tools on the ExPASy Server. In: John M. Walker ed, The Proteomics Protocols Handbook, *Humana Press*, (2005) 571-607

-
- [13]. Gill SC., Von Hippel PH Extinction coefficient, *Anal Biochem* 182 (1989) 319- 328.
- [14]. Ikai A. J., Thermo stability and aliphatic index of globular proteins. *J Biochem* 88 (1980) 1895-1898. PMID:7462208
- [15]. Jones D. T., Protein structure prediction in genomics. *Brief Bioinform* 2(2) (2001) 111-125
- [16]. Kleiger G, and Eisenberg D., GXXXG and GXXXA Motifs Stabilize FAD and NAD (P)-binding Rossmann Folds Through C α -H $\cdots$ O Hydrogen Bonds and van der Waals Interactions *J. Mol. Biol.* 323 (2002) 69-76
- [17]. Kyte J, Doolittle RF, A simple method for displaying the hydropathic character of a protein, *J Mol Biol* 157 (1982) 105- 132
- [18]. Majumder A.L., Chatterjee A., Ghosh D.K., Majee M., Diversification and evolution of L-*myo*-inositol 1-phosphate synthase. *FEBS Lett.* 553 (2003) 3–10
- [19]. Majumder A.L., Johnson M.D., Henry S.A., 1L-*myo*-inositol 1-phosphate synthase. *Acta Biochim. Biophys.* 1348 (1997) 245–256.
- [20]. Matthew DG and Mark SH., Quaternary structure of rice non-symbiotic hemoglobin. *J Biol Chem* 276 (2001) 6834-6839
- [21]. Norman R.A., McAlister M.S.B., Murray Rust J., Movahedzadeh F., Stoker N.G., & McDonald N.Q., Crystal structure of inositol 1 phosphate synthase from *Mycobacterium tuberculosis*, a key enzyme in phosphatidyl inositol synthesis. *Structure* ,vol10 (2002) 393-402
- [22]. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE., UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem.* 25(13) (2004) 1605-12
- [23]. Ramachandran GN, Ramakrishnan C and Sasisekharan V Stereochemistry of polypeptide chain configurations. *J Mol Biol* 7 (1963) 95-99
- [24]. Banerjee R, Dhani R, Chhetri D, and Adhikari J, Occurrence of *myo*-inositol-1-phosphate phosphatase in pteridophytes: characteristics of the enzyme from the reproductive pinnules of *Dryopteris filix-mas* (L.) Schott, *Braz. J. Plant Physiol.*, 19(2) (2007) 109-117
- [25]. Sneath P.H.A, and Sokal R.R., *Numerical Taxonomy*. Freeman, San Francisco. (1973)
- [26]. Stein A.J, & Geiger J.H., The crystal structure and mechanism of L *myo* inositol 1 phosphate synthase. *Journal of Biological Chemistry* 277 (2002) 9484-9491
- [27]. Tamura K., Stecher G., Peterson D., Filipski A., and Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* 30 (2013) 2725-2729.
- [28]. Vincentz M, Bandeira-Kobarg C, Gauer L, Schlogl P and Leite A., Evolutionary pattern of angiosperm bZIP factors homologous to the maize Opaque2 regulatory protein. *J Mol Evol* 56 (2003) 105-116
- [29]. Wiederstein M. & Sippl M.J, ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research* 35 (2007) W407-W410
- [30]. Zuckerkandl E. and Pauling L, Evolutionary divergence and convergence in proteins. Edited in *Evolving Genes and Proteins* by V. Bryson and H.J. Vogel, pp *Academic Press* (1965) 97-166