

©**0**⊗≡

DOI: 10.5114/bta.2016.60780

Myo-inositol 1-phosphate synthase – the chosen path of evolution

ANJAN HAZRA*, PARAMITA NANDY (DATTA)

Barasat Government College, Kolkata-700124, India

Abstract

Myo-inositol is a cyclohexanehexol containing cyclic polyol that has an important role in both abiotic and biotic stress responses, and it is specifically found to be accumulated in halophytes under salt stress where it acts as an osmolyte. Biosynthesis of *myo*-inositol is catalyzed by *myo*-inositol 1-phosphate synthase (EC 5.5.1.4, MIPS). The enzyme has been reported from a wide range of organisms belonging to prokaryotes and eukaryotes. In the current investigation, the naturally most fit isoform (in terms of selection pressure) or sequence variety of MIPS, known so far, was identified from the highest evolved plant group angiosperm. Subsequently, homologues of this MIPS variety were analysed from each of the taxonomic groups of the plant kingdom. Two common domains in MIPS nucleotide sequences and six conserved domains in the amino acid sequences were isolated, of which two amino acid domains were found to be unique for plants. According to the phylogenetic tree analysis based on MIPS amino acid sequences, MIPS proteins under current study are found to be clustered in branches in a way that confirms a common plant taxonomical lineage. Molecular clock analysis confirmed a much higher relative time of divergence from the prokaryotic cyanobacteria to eukaryotes, than the divergence within the eukaryotic community. As revealed by our study, MIPS started evolving from the lowermost plant group and with some modification through time it attained its highest adapted state in angiosperm via all intermediate plant groups. Interestingly, *Porteresia* MIPS is reflected as an isolated entity from other angiospermic members.

Key words: *myo*-inositol 1-phosphate synthase, evolution, natural selection, conserveness, phylogenetic tree, time of divergence

Introduction

The prosperity of genomics and its allied tools along with computational biology in recent times paves the way to an understanding of the evolution of proteins or genes. Phylogenetic perspectives can be focused on proteins across divergent prokaryotic and eukaryotic taxa marked via both nucleotide/amino acid sequences and functional or catalytic domains in their structural organisation.

The free inositol molecule and its derivatives are recognised as being important in the biological kingdom, and have acquired diverse functions over the course of evolution. So far, inositol and its derivatives have been shown to be involved in growth regulation, membrane biogenesis, hormone regulation, signal transduction, pathogen resistance and stress adaptation in higher plants (Loewus and Murthy, 2000; Stevenson et al., 2000; Michell, 2008). The *myo*-inositol is a cyclohexanehexol that has an important role in both abiotic and biotic stress responses (Shinozaki et al., 2006) specifically found to be accumulated in halophytes under salt stress (Sengupta et al., 2008) and acting as an osmolyte. It has been observed that *myo*-inositol derivatives regulate salt stress response by serving as compatible solutes and signalling molecules (Kido et al., 2013). According to the biosynthetic pathway of *myo*-inositol, the first key step is formation of *myo*-inositol 1-phosphate from glucose 6-phosphate (Loewus and Kelly, 1962; Majumder et al., 1997) catalysed by *myo*-inositol 1-phosphate synthase (EC 5.5.1.4, MIPS) – Fig. 1, which is our current focus of interest.

The enzyme has been reported from a wide range of organisms belonging to prokaryotes and eukaryotes (Majumder et al., 1997), it accumulates mostly under diffe-

^{*} Corresponding author: Barasat Government College, Kolkata-700124, India; e-mail: hazranjan93@gmail.com



myo-inositol 1-phosphate synthase (MIPS)

Fig. 1. A diagram of myo-inositol synthesis from glucose 6-phosphate

rent abiotic stress conditions such as drought, salinity etc. and acts as an osmoregulator in mammals too (Agranogff and Fischer, 2001). Expression studies of multiple genes encoding MIPS have revealed the possibility of specialised roles for individual enzyme isoforms. In higher organisms, the enzyme has two forms: cytosolic and organellar. MIPS activity has been localised in cytosolic and plastidic compartments in peas (Pisum sativum) (Imhoff and Bourdu, 1973), Vigna radiata and Euglena gracilis (Adhikari et al., 1987), Citrus paradisi (Abu-abied and Holland, 1994), Arabidopsis thaliana (Johnson and Sussex, 1995), Mesembryanthemum crystallinum(Ishitani et al., 1996), Oryza sativa(Ray Chaudhuri and Majumder, 1996; Hait et al., 2002), Zea mays and Hordeum vulgare (Keller et al., 1998), Brassica napus (Larson and Raboy, 1999), Glycine max (Iqbal et al., 2002), Sesamum indicum (Chun et al., 2003) and Phaseolus vulgaris (Johnson and Wang, 1996). Cytosolic MIPS has also been isolated and characterised from Perilla frutescens and A. thaliana, indicating its presence during normal seed growth (Chung et al., 1999; Jin et al., 2000; Baud et al., 2004). There are also a number of reports on MIPS from bryophyte (Chhetri et al., 2006) and pteridophyte members (Chhetri et al., 2005, 2006a, 2006b; Basak et al., 2013). Recent advances in structural and functional genomics and proteomics have made it easy to track the evolution of this essential enzyme.

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 humans, indicate that MIPS started evolving from different sources (Majumder et al., 2003) and that MIPS enzyme activity is distributed in evolutionarily diverse phyla, from eubacteria, archaebacteria, cyanobacteria, algae, fungi to higher plants and animals.

The first MIPS protein reported from an archaebacterial source was from *Archaeoglobus fulgidus* (Chen et al., 2000), followed by other phylogenetically diverse organisms, such as red algae, bryophytes, gymnosperm and fungi. Phylogenetic analysis of fungal MIPS has shown a high degree of homology with the *Saccharomyces MIPS* gene. In addition, all the fungal MIPS proteins have an extended N-terminal end of about 28 amino acid residues, which is absent in other MIPS proteins. Analyses of the total protein sequences (Bachhawat and Mande, 2000; Majumder et al., 2003) indicate that the eukaryotic MIPS family is homogenous, but not similar to the prokaryotic *MIPS* genes. This could be explained by a monophyletic origin of the eukaryotic MIPS genes.

The enzyme *myo*-inositol 1-phosphate synthase probably exists in native state in a homotetrameric form as evidenced by crystallography (Norman et al., 2002; Stieglitz et al., 2004). It is involved in internal oxidoreduction and cyclisation and needs NAD+ as a cofactor (mostly in reconverted form). MIPS converts glucose 6-phosphate to *myo*-inositol 1-phosphate in an NAD+ dependent manner. The NAD+ binding site in MIPS involves a Rossman fold characterised by a GXGGXXG motif typical for oxidoreductases (Kleizer and Eisenberg, 2002). Two distinct types of MIPS structures are reported through crystallography, one for prokaryotic and the other for eukaryotic enzymes. Crystal structure analysis of MIPS from the eukaryotic member *Saccharomyces cerevisiae* has revealed that each monomer of the homotetrameric MIPS has three functionally important structural domains; namely, the NAD binding Rossman fold, the catalytic binding site and the core domain. This study also exemplifies a case of induced fit model for binding of the substrate with the catalytic domain of the enzyme (Stein and Geiger, 2002).

In this investigation a collection of MIPS sequences (both amino acid and nucleotide) deposited in the NCBI (National Centre for Biotechnology Information), EBI (European Bioinformatics Institute) and Phytozome v10 database (up to March 2015) were used for a set of evolutionary analyses. Although having a significantly varying level of sequence availability the selected sequences were picked only after systematic consideration. The sequences, which contain the common domain that is shared by the higher vascular characteristic plants and members of each other group up to the lowermost unicellular organism of the plant kingdom, were selected for further analysis.

The genomic structure and organization of MIPS has been determined in several organisms, including yeast, *Arabidopsis thaliana, Citrus paradisi, Nicotiana tabacum,* and *Zea mays.* All these sequences have shown regions of high conservation at the nucleotide level (Hegeman et al., 2001). The amino acid sequences of selected *MIPS* genes from eubacteria, archaea, parasites, fungi, plant and animal sources reveal four stretches of eukaryotic MIPS sequences as "highly conserved domains" (Majumder et al., 2003). These four domains are involved in MIPS protein binding and are essential for MIPS functions, such as cofactor NAD+ binding and reaction catalysis (Majumder et al., 1997).

In our current investigation, sequence homology is considered for species selection to find the varieties of MIPS with stretches or portions conserved throughout the plant kingdom from the lowermost algae to the most highly evolved angiosperm. The other possible varieties of MIPS in different algal and fungal groups were discarded, since they did not continue to evolve to the most advanced type. The nucleotide sequences of MIPS from some well described salt tolerant angiosperms like *Porteresia*, *Avicennia* etc. targeted in the present study as the most effective and evolved sequences of MIPS, together with the evolutionary line in each lower group, have been searched and those that share the maximum sequence similarity were selected as representative species of these particular groups.

In addition to changes due to random, non-adaptive evolutionary forces, for any protein evolving through divergent lineages, diversity might arise due to adaptive changes in its gene sequences to suit the environment of the organism; however, a core functional structure must remain conserved through all such changes, to preserve the original function of the protein (Monizngo, 1996; Clothia, 1997). As an ancient MIPS, proteins/genes constitute an ideal model for studying evolution throughout the biological kingdom for its blanket distribution and prolonged presence. So, the reflection of the sequence divergence in respective protein structures was subsequently studied along with generation of a phylogenetic relationship among selected taxa.

Materials and methods

Species selection

A stress (salt) tolerant enzyme is supposed to show its highest level of activity when it is expressed in a salt tolerant species. Porteresia coarctata, Avicennia marina etc. are some of the halophilic angiosperms where MIPS proteins play a significant role in withstanding salinity. In the present study, P. coarctata has been considered as the reference where the expressed variety of myoinositol 1-phosphate synthase is sufficient for the species to thrive in the highly saline environment of the Sundarban delta (Nandy Datta et al., 2009). Moreover, the representative sequence of pino1 (MIPS coding gene present in *P. Coarctata*) has been found to be similar to many other angiosperm members as well. Thereafter, its homologue sequences with maximum identity and respective source organisms from each group of plant kingdom were picked for further analysis, e.g. rate and pattern estimation of sequence divergence through the evolutionary line. The Physcomitrella patens & Selaginella moellendorffii proteome report in the database (Phytozome v10) was used to identify the respective MIPS protein sequences, since no sequence was available in the database representing the bryophyte and pteridophyte groups (Hazra, 2015). In the case of gymnosperm, a single partial amino acid sequence of the protein was available and this could be used in sequence alignment only.



Fig. 2. Graphical summary of BLAST MIPS sequence alignment for selected species (Porteresia as query)



Fig. 3. Graphical summary of BLAST MIPS sequence alignment for selected species (Chlamydomonas as query)



Fig. 4. BLAST result showing yeast MIPS sharing two different common domains with Porteresia and Chlamydomonas

Database

All available reported complete cDNA or mRNA sequences of MIPS coding gene ino1, which is the key responsible for MIPS expression, as well as the reported amino acid sequences of *myo*-inositol 1-phosphate synthase were retrieved and then curated after screening from the databases of the National Centre for Biotechnology Information, USA (www.ncbi.nlm.nih.gov) and the European Bioinformatics Institute, Cambridgeshire, UK (www. ebi.ac.uk), updated March 2015 (Supplementary Table 1).

Nucleotide sequence homology

The subject species were selected on the basis of the percentage of sequence homology by BLAST® (Altschul et al., 1990) result (blastn, www.blast.ncbi.nlm.nih.gov/Blast.cgi).

Multiple sequence alignment

Amino acid and nucleotide sequences from selected organisms were further analysed by a number of multiple sequence alignments (MULTALIN/ Corpet, 1988) using an online server (www.multalin.toulouse.inra.fr/ multalin).

Quaternary structure prediction

Three-dimensional structure predictions of the respective MIPS proteins from different individuals were performed using comparative modelling in Modeller 9.12 (Sali et al., 2006). In brief, this method aligns a query sequence to one or more templates with known structures as determined by crystallization/X-ray diffraction, or NMR spectrometry. Suitable templates (PDB id 1U11 and 1LA2) were selected through pdbBLAST at NCBI. As the protein exists in homotetrameric form in its native state (Stein and Geiger, 2002), the top hit on the homotetrameric structure report was selected as a working template. The ligand modelling command was used to integrate an appropriate ligand with the protein. Molecular visualisation was conducted with Chimera 1.10 (Pettersen et al. 2004).

Evaluation of a theoretical model

To check the stereochemical stability of psi (ψ) and phi (ϕ) dihedral angles of each amino acid residues in the protein structures, Ramachandran plot (Ramachandran et al., 1963) analysis from the Molprobity server (www.molprobity.biochem.duke.edu), (Chen et al., 2010) was used.

Molecular clock

The relative rate and time of divergence were determined by the molecular clock analysis method using a Molecular Evolution Genetic Analysis (MEGA6, Tamura et al., 2013).

Results and discussion

The isoform of MIPS naturally chosen by selection pressure might be represented by one belonging to *Porteresia coarctata*, because in this organism the MIPS protein/enzyme function supports to the highest degree plant's ability to survive in stress (high saline) conditions (Majee et al., 2004). So, utilizing this MIPS mRNA sequence as a query in the NCBI-BLAST programme against the non-redundant database enabled us to identify its near relative, i.e. the most adapted varieties from each plant group. The nucleotide BLAST report revealed a significant high similarity and identity in the MIPS nucleotide coding sequences among all angiosperm with only a few sequence variations and that are taxa specific. Moreover, significant similarity was observed in some members of algae, fungi, bryophyte, pteridophyte etc. A number of the closest homologue sequences from angiosperm to algae were selected on the basis that they were members of the chosen MIPS evolution path towards *P. Coarctata* MIPS isoform that evolved under extreme environmental conditions.

On the basis of the next BLAST report focusing on those 16 selected MIPS sequences belonging to different species, as the subjects (Supplementary Table 1), two domains (Fig. 2 and Fig. 3) were identified to be very common or similar upstream and downstream of the MIPS nucleotide coding sequences, respectively. These two domains were further analysed by multiple sequence alignment (described later in this section) for a precise observation at each site.

The fungi were characterised by a slightly lower level of MIPS sequence similarity when compared with angiosperm than other members; however, they showed a high sequence matches among themselves (Fig. 2 and Fig. 3), which is in congruence with a previous report (Majumder et al., 2003). This was probably due to the fact that fungal MIPS has high conserveness within the group and are less evolved in respect to the adaptability against salt stress. Interestingly, one member of the fungal group, the unicellular eukaryotic model organism yeast (Saccharomyces cerevisiae) had a unique sequence composition for MIPS that shared two separate common domains with Chlamydomonas and Porteresia (Fig. 4). It is not among organisms chosen for our analyses, but still it is being considered here because of its routine use in evolutionary analyses. Moreover, the only high resolution eukaryotic MIPS crystal structure reported till date is for Saccharomyces. Hence, it was distinctly separated from other members of its family and this was also reflected in the subsequent phylogenetic tree based on the MIPS protein sequence.

Meanwhile, the BLAST result unveiled one more important point. Although *Chlamydomonas* belong in plant kingdom to a primitive lower algae group, MIPS DNA sequence showed almost complete identity with

A. Hazra, P. Nandy (Datta)



Fig. 5. Multiple sequence alignment of upstream MIPS nucleotide common domains for selected species



Fig. 6. Multiple sequence alignment of downstream MIPS nucleotide common domain for selected species

Porteresia coarctata MIPS variety. Once the naturally most adapted variety (*P. coarctata* MIPS) showed such high similarity with so called "primitive member" it was subsequently regarded as the evolutionary chosen variety of MIPS from evolutionary inferior group. Therefore in the next BLAST analysis *Chlamydomonas* was put as a query and all previously selected species as subjects (Fig. 3).

The Multalin reports precisely revealed the conserved sequence motifs in common MIPS nucleotide domains of the selected species (Fig. 5 and Fig. 6). Those domains probably appeared first in the primitive algal organisms like *Chlamydomonas*, and became naturally selected through the course of evolution. As such most of its existing characters (nucleotides) remained unchanged till the highest evolved organism. When the next

	1	10	20	30	40	50	60	70	80	90	100	110	120	130
Oscillatoria Pacilluc	1								-			Consei	rved N	lotif 1
Sphaerulina	MAPHAN	DSEAYSAGTP	LTR	GGSST	DFT	VDATNVQY	TGQDITS	KYRYRTTEV		KR-YDG-	KYVAKPKET	VYDFKTQRNY	G-KYGYML	GLGGNNGT
Fusarium	MODUOO	TTCCTNCOO	TNGK	655H1	DTTEN	VNSPNVIT	IDHEIRS	KTITRIIKY		DYDHNG-	KTYHIPNEI	LTUFKYURKI	P-KYGNNL	UCUCCHNCT
Colletotrichum	MAPHAE	/NSGSAHGNG	PSPHPNGN	GAAATTATTTI	NOOHRELET	VNSPNVVY		KYTYPTTKU-			KYVATPTET	VYFEKTORNU	P-RVGHHIN	GUGGXNGT
Aspergillus	МАРНА-	rsdavangav	NGS	ARTTT	API FT	VNSPNVFY	TONETKS	RYAYHTTNT		TRTAFG-	KMVATPKVT	NYOFKYDRKY	G-KYGHMI A	GHGGNNGS
Dacryopinax	MAPHA-	<i>TGPAPELGPY</i>	TAPOPPYH	PTAER	RDTYYL	VDSDRTTY	TEEHITE	KFTYRGAOY		TRE-GG-	DIKYTPTKE	EFEFRTKRKY	P-KLGINJ	GLGGNNGS
Chondrus					MYFYKSFA	VNGSHVRR	TDQYLEV	DFPYMQTSV		EMEPKST	EAVAYLYKQ	MYTFRTQLKY	A-RLGYML	GHGGNNGS
Chlanydononas					MLYDSFE	VQSPNVVY	SDEHITS	TYDYASTKL		DRAPDG	KHYYKPTST	SYQFRYDRRY	/P-KLGYML	GHGGNNGT
Setaria					MFIESFR	VESPNVRY	GPSEIES	EYRYDTTELY		EGKDGAS	RHYYRPKSY	KYNFRTSTAY	/P-KLGYHL	GHGGNNGS
Glycine					MFIENFK	VECPNVKY	TETEIQS	VYNYETTELVI		ENRNGTY	QHIYKPKSY	KYEFKINIH	P-KLGYML	GAGGNNGS
Hvicennia					METECED	YUSPNYKT	IUSEIHS	CYDYDTTCI U	1	ENKNGIT	UNTARAK	OVUEDTETTU	P-KLGYNLA	CUCCNNCC
Oct peococcus					MI VETET	VICODTODETC	GHHETES	EIKTUIIELYI	DAASTTANCI S	ESHDUHS	DEGI VDEED	UTHERISIIY TETHDTRVTU	P-KLGYNLY	CUCCNNCS
Saccharomuces					MTERNTAP	TTSVKVVTDK	TYKNNFLLT	KYSYENAVV-	MIII TIIII TIIIIITII TIIIIITTIIIITII TIIII TIIIIITII TIIII TIIIITTIII TIIII TIIII TIIIIITTIIIITII TIIIIITII TIIIIITII TIIIIITII TIIIIITII TIIIIITII TIIIIITII TIIIIITII TIIIIITII TIIIIITIIITII TIIIIITII TIIIIITIIIITII TIIIIITII TIIIITIIIITII TIIIIITII TIIIIITIIITII TIIIIITII TIIIIITIIITII TIIIIITII TIIIITIIITII TIIIIITII TIIIITIIITII TIIIITIIITII TIIIITIIITII TIIIITIIITII TIIIIITII TIIIITIIIITII TIIIITIIITIIIITII TIIIITIIITIIIITII TIIIIITII TIIIIITIIITI	TKTASG	RENVTRTVO	NYVEKI DI KK	PEKI GTMI	GL GGNNGS
Pinus					meonin	1011111010	Inditett	RIGIENI		minud		DITINEDEN	a Encorne	deddinido
Consensus						vv		····y·····	•		P	.y.f	. k.g.ml.	g.ggnng.
	131	140	150	160	170	180	190	200	210	220	230	240	250	260
	I	+	+	+	+	+	+	+	+	+	+	+	+	
Oscillatoria				MEKIK	LAIYGYGNC	ASSLYQGIYFY	KDKTSEEAJ	GLMHHDIGGYN	KPND-IDYYAA	FDIDRRKYG	QDYSQAIFA	EPNCTMY	FCQDIPEI	iTK¥S
Bacillus		ATVOTCUL EU	'GTI GAVAT	TTISGLFAYNO	NLAPLRGYI	SSEREFETLQL	TPI DOT	OFCCURTOVE	SI TEAVKPHOT	TOPDTI OUT	and a second sec			
	20.00	TITUTUYLEY					The bag	IL DOMDIGNO	JETEHTKKNUT	THELTCAKT	EMNLNDY	PYHQAPLA	ANYNDFYKG\	YRLK
Sphaerulina	TYTAGE		TREGPRAA	NYYGSYYMSSI	YKLGTDAET	GADYNIPLQSL	LPMYHPNDL	AIGGNDISGM	DLASAMDRAQY	LEPTLKQLV	emnlndv Kkemagmkp	PYHQAPLA LPSIYYPDF1	inyndfykg\ Caanqgdrac	YRLK
Sphaerulina Fusarium	TVTAGI TVTAGI		TREGPRAA	NYYGSYYMSS1 Nyygsyymgs1	YKLGTDAET	GADYNIPLQSL HQEINIPFHNI	LPMYHPNDL	AIGGHDISGMU	DLASAMDRAQY	LEPTLKQLY	emnlndy Kkemagmkp Kkemaemyp	PYHQAPLA LPSIYYPDFI LPSIYYPDFI	invndfykgi Caanqgdrad Caanqedrad	YRLK
Sphaerulina Fusarium Metarhizium	TVTAGI TVTAGI TVTAGI	LANRRNL THE	TREGPRAA TKEGPRES TKEGPREA	NYYGSYYMSS1 Nyygsyymgs1 Nyygsyymgs1	YKLGTDAET LKLGHDAKT MKLGSDAKT	GADYNIPLQSL HQEINIPFHNI NADINIPFHNY	LPMYHPNDL LPMYHPNDL LPMYHPNDL	AIGGADISCH VIGGADISCH VIGGADISCH VIGGADISCH	DLASAMDRAQY NLAQAMDRAQY NLAQAMDRAQY	IQEPILQKI LEPTLKQLY LEPTLKAQY LEPTLKSLY	emnlndy Kkemagnkp Kkemaenyp Kkemnenyp	PYHQAPLA LPSIYYPDFI LPSIYYPDFI LPSIYYPDFI	nyndfykg\ Caanqgdrac Caanqedrac Caanqedrac	YRLK NYIP NYLE
Sphaerulina Fusarium Metarhizium Colletotrichum Popopoilluo	TVTAGI TVTAGI TVTAGI TVTAGI TVTAGI	LANRRNL THE LANRRQLAND LANRRNL YHE LANRRNL YHE	TREGPRAA TKEGPRES TKEGPREA TREGPRAA	NYYGSYYMSSI NYYGSYYMGSI NYYGSYYMASI NYYGSYIMGSI NYYGSI MASI	YKLGTDAET LKLGHDAKT MKLGSDAKT YKLGTDADT	GADYNIPLQSL HQEINIPFHNI NADINIPFHNY ARDYNIPFHDL GEEINTPEHDH	LPMYHPNDL LPMYHPNDL LPMYHPNDL VPMYHPDDL LPMTHPNDL	AIGGADISGA VIGGADISCA VIGGADISKA AIGGADISCA VIGGADISCA	DLASAMDRAQY NLAQAMDRAQY NLAQAMDRAQY SLADAMDRAQY	LEPTLKQLY LEPTLKQLY LEPTLKAQY LEPTLKSLY LEPTLKAQL	emnlndy Kkemagmkp Kkemaemyp Kkemnemyp Kkemadmyp Rkemadmyp	PYHQAPLA LPSIYYPDFI LPSIYYPDFI LPSIYYPDFI LPSIYYPDFI	inyndfykgi Caanqgdral Caanqedral Caanqedral Caanqedral	/YRLK DNYIP DNYLE DNYLE DNLIP
Sphaerulina Fusarium Metarhizium Colletotrichum Aspergillus Dacruoninay	TYTAGI TYTAGI TYTAGI TYTAGI TYTAGI TYTAGI	LANRRNL THE LANRRQLAND LANRRNL VHE LANRRGLAND LANRRGLAND LANRRRL VHE	TREGPRAA TKEGPRES TKEGPREA TREGPRAA TREGERAA	NYYGSYYMSSI NYYGSYYMGSI NYYGSYYMASI NYYGSYIMGSI NYYGSLYMSSI NYYGSYI PASI	TYKLGTDAET TLKLGHDAKT TKLGSDAKT TYKLGTDADT TYKLGTDSKT TYRLGTDSKT	GADYNIPLQSL HQEINIPFHNI NADINIPFHNY ARDYNIPFHDL GEEINIPFHDM GKDEHL PYCDL	LPMYHPNDL LPMYHPNDL LPMYHPNDL VPMYHPDDL LPMIHPNDL LPMIHPNDL	AIGGADISGM YIGGADISGM YIGGADISKM AIGGADISGM YIGGADISGM	DLASAMDRAQY NLAQAMDRAQY NLAQAMDRAQY SLADAMDRAQY NLADAMDRAQY NLADAMDRAQY DI DEAMPRAEY	LUEPILUKI LEPTLKQLY LEPTLKAQY LEPTLKSLY LEPTLKAQL LEPTLKQLY	emnlndy Kkemagmkp Kkemaemyp Kkemaemyp Kkemaemyp Rkemaemkp VPI Mrgtkp	PVHQAPLA LPSIYYPDFI LPSIYYPDFI LPSIYYPDFI LPSIYYPDFI LPSIYYPDFI LPSIYYPDFI	ANYNDFYKG\ CAANQGDRAC CAANQEDRAC CAANQEDRAC CAANQEDRAC CAANQEDRAC	YRLK DNYIP DNYLE DNYLE DNLIP DNYIE
Sphaerulina Fusarium Metarhizium Colletotrichum Aspergillus Dacryopinax Chondrus	TYTAGI TYTAGI TYTAGI TYTAGI TYTAGI TFTATI TI TGAI	LANRRNL THE LANRRQLAND LANRRNL VHE LANRRGLAND LANRRRLVHE LANRRRLVHE LANKNHYTHY	TREGPRAA TKEGPRES TKEGPREA TREGPRAA TREGERAA TKEGIQYP ITAKGPRAP	NYYGSVYHSST NYYGSVYHGST NYYGSVYHAST NYYGSVIHGST NYYGSLYHSST DYYGSVLRAST NI FGSVYHSST	TYKLGTDAET TLKLGHDAKT TKLGSDAKT TYKLGTDADT TYKLGTDSKT TYRLGTDPGT TYRLGANS-A	GADYNIPLQSL HQEINIPFHNI NADINIPFHNY ARDYNIPFHDL GEEINIPFHDM GKDFHLPYCDL GEDYFAPMHFI	LPHYHPNDL LPHYHPNDL LPHYHPNDL LPHYHPDDL VPHYHPDDL LPHYHPNDL LPHYHPNDJ LPHYHPNDJ	AIGGADISGM VIGGADISGM VIGGADISKM AIGGADISGM VIGGADISGM TIGGADISSL VIGGADISSL	DLASANDRAQY NLAQANDRAQY NLAQANDRAQY SLADANDRAQY NLADANDRAQY PLDEANRRAEY NLAHANDRAY	IQEPILQKI LEPTLKQLY LEPTLKAQY LEPTLKSLY LEPTLKQL LEPTLKQLY LPYDLQRQL FNYNLQRRI	emnlndv Kkenagmkp Kkenaemvp Kkenadmvp Rkenaemkp Vplmrgikp Tphmesi ap	PYHQAPLA LPSIYYPDFI LPSIYYPDFI LPSIYYPDFI LPSIYYPDFI LPSIYYPDFI LPSYYPPFI LPSYYPDFI	invndfykgi Caanqgdrac Caanqedrac Caanqedrac Caanqedrac Caanqedrac Caanqedrac Caanqedrac	/YRLK DNVIP DNVLE DNVLE DNLIP DNVIE DNVIP
Sphaerulina Fusarium Metarhizium Colletotrichum Aspergillus Dacryopinax Chondrus Chlamydomonas	TYTAGI TYTAGI TYTAGI TYTAGI TYTAGI TFTATI TLTGAL TLTAGI	LANRRNL THE LANRRQL AHD LANRRQL AHD LANRRQL AHD LANRRGL AHD LANRRRL YHE LANKNHYTHY LSNKRQISHN LANKHGITHM	TREGPRAA TKEGPRES TKEGPREA TREGPRAA TREGERAA TKEGIQYP ITAKGPRAP (TKDGLKKP	NYYGSVYHSST NYYGSVYHGST NYYGSVYHGST NYYGSVYHGST NYYGSLYHSST DYYGSVLRAST NLFGSVYHSST NFYGSLT0AAT	VKLGTDAET ILKLGHDAKT IKLGSDAKT VKLGTDADT VKLGTDSKT VRLGTDPGT VRVGANS-A VRVGNF-E	GADYNIPLQSL HQEINIPFHNI NADINIPFHNY ARDYNIPFHDL GEEINIPFHDH GKDFHLPYCDL GEDYFAPMHEL GEEYYAPFKAL	LPHYHPNDL LPHYHPNDL LPHYHPNDL VPHYHPDDL LPHYHPNDL LPHYHPNDJ LPHYHPDEJ LPHYHPDEJ	AIGGADISGM YIGGADISGM YIGGADISGM AIGGADISGM YIGGADISGM YIGGADISSH YLGGADISGL	DLASAMDRAQY NLAQAMDRAQY NLAQAMDRAQY SLADAMDRAQY NLADAMDRAQY PLDEAMRRAQY NLADAMDRAQY NLAHAMQRAKY NLAEAMERAKY	IQEPILQKI LEPTLKQLY LEPTLKAQY LEPTLKAQY LEPTLKAQL LEPTLKQLY LPYDLQRQL FDYDLQRRL LDHTLQEQM	ehnlndv Kkemaghkp Kkemaemvp Kkemadmvp Rkemaemkp Vplmrgikp Iphmeslap Vpymremvp	PYHQAPLA LPSIYYPDFI LPSIYYPDFI LPSIYYPDFI LPSIYYPDFI LPSYYYPDFI LPSYYYPDFI LPSYYPDFI LPSYYPDFI LPSYYPDFI	inyndfykgi Caanqgdrac Caanqedrac Caanqedrac Caanqedrac Caanqedrac Caanqerac Caanqerac Caanqerac	/YRLK DNVIP DNVLE DNLIP DNVIE DNVIE DNVIP DNVIP
Sphaerulina Fusarium Metarhizium Colletotrichum Aspergillus Dacryopinax Chondrus Chlanydononas Setaria	TYTAGI TYTAGI TYTAGI TYTAGI TYTAGI TYTAGI TFTATI TLTGAL TLTAGI TLTAGI	LANRRNL THE LANRRQL AHD LANRRQL AHD LANRRGL AHD LANRRRL YHE LANKRNLYTHY ISNKRQISHN LANKHGITHM LANKHGITHM	TREGPRAA TKEGPRES TKEGPREA TREGPRAA TREGERAA TREGERAA TKEGIQYP ITAKGPRAP ITKDGLKKP	NYYGSYVHSST Nyygsyvhgst Nyygsyvhgst Nyygsivhgst Nyygslvhsst Dyvgsvlrast Nlfgsvvhsst Nfygsltqaat Nyfgsltqast	IVKLGTDAET ILKLGHDAKT INKLGSDAKT IVKLGTDADT IVRLGTDSKT IVRLGTDPGT IVRVGANS-A IVRVGNFE IIRVGSYN	GADYNIPLQSL HQEINIPFHNI NADINIPFHNY ARDYNIPFHDL GEEINIPFHDM GKDFHLPYCDL GEDYFAPMHEL GEEYYAPFKAL GEEIYAPFKSL	LPHYHPNDL LPHYHPNDL LPHYHPNDL VPHYHPDDL LPHYHPNDL LPHYHPNDJ LPHYHPDEJ LPHYPPDDJ LPHYPDDJ	AIGGHDISCHU VIGGHDISCHU VIGGHDISCHU AIGGHDISCH AIGGHDISCH TIGGHDISCH VLGGHDISCLU VLGGHDISGLU VFGGHDISCH	DLESANDRAQY NLAQANDRAQY NLAQANDRAQY SLADANDRAQY NLADANDRAQY PLDEAMRRAEY NLAHAMQRAKY NLAHAMQRAKY NLAEANERAKY NLADSTTRAKY	IQEPILQKI LEPTLKQLY LEPTLKAQY LEPTLKAQY LEPTLKAQL LEPTLKQLY LPYDLQRQL FDYDLQRRL LDHTLQEQM LDIDLQKQL	ehnlndv Kkemaghkp Kkemaemvp Kkemadmvp Rkemaemkp Vplmrgikp IPHMeslap Vpymremvp Rpymesmvp	PYHQAPLA LPSIYYPDFI LPSIYYPDFI LPSIYYPDFI LPSIYYPDFI LPSYYPDFI LPSYYPDFI LPSYYPDFI LPGIYDPDFY LPGIYDPDFI	Invndfykg\ Caanqedrad Caanqedrad Caanqedrad Caanqedrad Caanqedrad Caanqegrad Caanqeqrad Caanqegrad Caanqegrad	/YRLK MVIP DNVLE DNVIE DNVIE DNVIP DNVIP DNLIK
Sphaerulina Fusarium Metarhizium Colletotrichum Aspergillus Dacryopinax Chondrus Cholanydomonas Setaria Glycine	TYTAGI TYTAGI TYTAGI TYTAGI TYTAGI TFTATI TLTGAL TLTAGI TLTAGY TLTAGY	IANRRNL THE LANRRQLAHD LANRRQLAHD LANRRQLAHD LANRRGLAHD LANRRRLYHE LANKNHYTHY ISNKRQISHN LANKHGITHM LANREGISHA	TREGPRAA DTKEGPRES TKEGPREA DTREGPRAA TREGERAA TKEGIQYP ITAKGPRAP (TKDGLKKP DTKDKIQQA DTKDKIQQA	NYYGSYVHSST Nyygsyvhgst Nyygsyvhgst Nyygsyvhgst Nyygsvyhgst Dyygsvyrsst Nifgsvyhsst Nifgsltqaat Nyfgsltqaat Nyfgsltqasf Nyfgsltqasf	IVKLGTDAET ILKLGHDAKT INKLGSDAKT IVKLGTDADT IVRLGTDSKT IVRVGANS-A IVRVGNFEI IIRVGSYN IIRVGSFQ	GADYNIPLQSL HQEINIPFHNI NADINIPFHNU GREINIPFHDL GEEINIPFHDM GKDFHLPVCDL GEDYFAPMHEL GEEYYAPFKSL GEEIYAPFKSL GEEIYAPFKSL	LPHYHPNDL LPHYHPNDL LPHYHPNDL LPHYHPDDL LPHYHPNDL LPHYHPNDI LPHYHPDEJ LPHYPPDDJ LPHYNPDDJ LPHYNPDDJ	AIGGHDISGMI VIGGHDISGMI VIGGHDISKMI AIGGHDISKMI AIGGHDISGMI VIGGHDISSHI VIGGHDISGLI VFGGHDISNMI VFGGHDISNMI	DLESANDRAQY NLAQANDRAQY NLAQANDRAQY SLADANDRAQY NLADANDRAQY PLDEAMRRAEY NLAHAMQRAKY NLAEANERAKY NLAEANERAKY NLADANARAKY	IQEPILQKI LEPTLKAQY LEPTLKAQY LEPTLKSLY LEPTLKAQL LEPTLKQLY LPYDLQRQL FDYDLQRQL LDHTLQEQM LDIDLQKQL FDIDLQKQU	ennlndy Kkenagmkp Kkenaemvp Kkenadmvp Rkenadmvp Rkenadmvp Rkenadk Rkenadk VPLMRGIKp IPHMESLAP VPYMREMVP RPYMESMLP	PYHQAPLA LPSIYYPDFI LPSIYYPDFI LPSIYYPDFI LPSIYYPDFI LPSIYYPDFI LPSVYYPSFI LPSVYYPDFI LPGIYDPDFI LPGIYDPDFI LPGIYDPDFI	invndfykgv (Aanvgedrai (Aanvgedrai (Aanvgedrai (Aanvgedrai (Aanvgedrai (Aanvgerai (Aanvgerai (Aanvgerai (Aanvgerai)	YYRLK NYYIP NYLE NYIE NYIP NYIP NYIP NYLR NYIK
Sphaerulina Fusarium Metarhizium Colletotrichum Aspergillus Dacryopinax Chondrus Chlanydononas Setaria Glycine Avicennia	TYTAGI TYTAGI TYTAGI TYTAGI TYTAGI TFTATI TLTGAL TLTGGU TLTAGY TLTGGY TLTGGY	IANRRNLTHE LANRRQLAND LANRRQLAND LANRRQLAND LANRRQLAND LANRRQLYNE LANKNQTSHN LANKHQITHM LANKHQITHM LANREGISHA LANREGISHA	TREGPRAA TKEGPRES TKEGPREA TREGPRAA TREGERAA TKEGIQYP TKEGIQYP TKKOKYQQA TKDKIQQA TKDKYQQA	NYYGSYVHSST Nyygsyvhgst Nyygsyvhgst Nyygsvlhgst Nyygsvlrgst Dyvgsvlrgst Nlfgsvvhsst Nffgsltqast Nyfgsltqast Nyfgsltqast Nyfgsltqasf	IVKLGTDAET ILKLGHDAKT IVKLGSDAKT IVKLGTDSKT IVRLGTDSKT IVRVGANS-A IVRVGNFE IRVGSYN IIRVGSFQ IIRVGSFN	GADYNIPLQSL HQEINIPFHNI NADINIPFHNU ARDYNIPFHDL GEEINIPFHDH GKDFHLPYCDL GEDYFAPMHEL GEEYYAPFKSL GEEIYAPFKSL GEEIYAPFKSL	LPAYHPNDL LPAYHPNDL LPAYHPNDL LPAYHPNDL LPAYHPNDL LPAYHPNDJ LPAYHPNDJ LPAYNPDDJ LPAYNPDDJ LPAYNPDDJ	AIGGHDISCHI VIGGHDISCHI AIGGHDISCHI AIGGHDISCHI VIGGHDISCHI VLGGHDISCHI VLGGHDISCHI VFGGHDISCHI VFGGHDISCHI IFGGHDISNHI	DLASANDRAQY NLAQANDRAQY SLADANDRAQY SLADANDRAQY PLDEAMRRAQY NLABANDRAQY NLABANGRAKY NLABANGRAKY NLADSTTRAKY NLADANGRAGY NLADANGNGQGQ	LUEPTLLKQLY LEPTLKQQY LEPTLKAQY LEPTLKAQY LEPTLKAQL LEPTLKQLY LPYDLQRQL LDYDLQRQL LDUDLQRQL LDIDLQKQL LDIDLQKQL	ennlndy Kkenagnkp Kkenaenvp Kkenaenvp Kkenaenvp Rkenaenkp VPLNRGIKP IPHNESLAP RPYNESNVP RPYNESNVP RPYNESNLP RPYNESNLP	PVHQAPLF LPSIYYPDFI LPSIYYPDFI LPSIYYPDFI LPSIYYPDFI LPSYYPDFI LPSYYPDFI LPSYYPDFI LPGIYDPDFI LPGIYDPDFI LPGIYDPDFI	Inyndfykg\ (Aanvgodra) (Aanvgedra) (Aanvgedra) (Aanvgedra) (Aanvgedra) (Aanvgera) (Aanvgera) (Aanvgera) (Aanvgera) (Aanvgera)	YYRLK INYIP INYIP INYIP INYIP INYIP INYIK INYIK INYIK
Sphaerulina Fusarium Metarhizium Colletotrichum Aspergillus Dacryopinax Chondrus Chlamydononas Setaria Glycine Avicennia Oryza	TVTAGI TVTAGI TVTAGI TVTAGI TVTAGI TTTAGI TLTGAI TLTGAI TLTAGY TLTGGY TLTGGY TLTAGY	IANRRNLTHE LANRRQLAHD LANRRQLAHD LANRRQLAHD LANRRRLAHE LANKHATTH ISNKRQISHN LANKHGITHA IANREGISHA IANREGISHA IANREGISHA	TREGPRAA TKEGPREA TREGPREA TREGERAA TREGERAA TREGEQYP TREGERAA TKEGIQYP TRAKGPRAP TKBKIQQA TKDKIQQA TKDKYQQA TKDKYQQA	NYYGSVVHSST NYYGSVVHGST NYYGSVVHGST NYYGSVVHGST DYYGSVVHSST DYVGSVLHSST DYYGSLTQAAT NYFGSLTQAAT NYFGSLTQASF NYFGSLTQASF NYFGSLTQASF NYFGSLTQASF	VKLGTDAET ILKLGHDAKT VKLGTDADT VKLGTDSTT VRLGTDSTT VRVGNFE IRVGSYN IIRVGSFQ IIRVGSFN IIRVGSFN	GADVNIPLQSL HQEINIPFHNI NADINIPFHNY ARDVNIPFHDL GEEINIPFHOW GKDFHLPVCDL GEEVYAPFKAL GEEIYAPFKSL GEEIYAPFKSL GEEIYAPFKSL GEEIYAPFKSL	LPHYHPNDL LPHYHPNDL LPHYHPNDL VPHYHPNDL LPHYHPNDJ LPHYHPNDJ LPHYHPNDJ LPHYHPDDJ LPHYNPDDJ LPHYNPDDJ LPHYNPDDJ LPHYNPDDJ	AIGGADISGAN YIGGADISGAN YIGGADISGAN AIGGADISGAN YIGGADISGAN YIGGADISGAN YIGGADISGAN YFGGADISAN YFGGADISAN YFGGADISAN YFGGADISAN YFGGADISAN	DLASANDRAQY NLAQANDRAQY NLAQANDRAQY SLADANDRAQY NLADANDRAQY NLADANDRAQY NLADANDRAQY NLADANGRAKY NLADANGRAKY NLADANGRAGQG NLADANGRAGY NLADANGRAQG	IQEPILUKU LEPTLKQLY LEPTLKAQY LEPTLKAQY LEPTLKQL LEPTLKQL LPYDLQRQL FDYDLQRQL LDHTLQEQM LDIDLQKQL LDIDLQKQL LDIDLQKQL	EMNLADY Kkenaghkpi Kkenaghvp Kkenadhvp Rkenadhvp Rkenaghkp Ryplagikp Iphmeslap Vpynrehvp Rpyneshlp Rpyneshlp Rpyneshlp Rpyneshlp	PVHQAPLF LPSIYYPDFI LPSIYYPDFI LPSIYYPDFI LPSIYYPDFI LPSYYYPSFI LPSYYYPSFI LPSYYPDFI LPGIYDPDFV LPGIYDPDFI LPGIYDPDFI LPGIYPDFI SLASHIPTSS	Invindfykgi Caanqedrai Caanqedrai Caanqedrai Caanqedrai Caanqedrai Caanqerai Caanqeerai Caanqeerai Caanqeerai Caanqeerai Caanqeerai Caanqeerai	YYRLK INVIE INVIE INVIE INVIE INVIP INVIK INVIK INVIK INVIK INVIK INSR
Sphaerulina Fusariun Metarhiziun Colletotrichum Aspergillus Dacryopinax Chondrus Chlanydononas Setaria Glycine Avicennia Oryza Ostreococcus	TVTAGI TVTAGI TVTAGI TVTAGI TVTAGI TTAGI TLTGAL TLTGAL TLTAGY TLTGGY TLTGGY TLTAGY TLTAGY	IGNRRNLTHE LANRRQLAND ANRRNLYNE LANRRGLAND ANRRRLYNE CANKNHVTHY ISNKRQISNN IANKHGITNM IANREGISNA CANREGISNA CANREGISNA CANREGISNA CASREGISNA	TREGPRAN TKEGPRES TREGPREA TREGERAN TREGERAN TREGERAN TREGERAN TREGERAN TREGEQUP TROKIQQA TKDKYQQA TKDKYQQA TKDKYQQA TKCGAQEP	NYYGSVVHSST NYYGSVVHGST NYYGSVVHGST NYYGSVLHGST NYYGSLHSST DYGSLVHSST DYGSLVHSST NYFGSLTQAST NYFGSLTQAST NYFGSLTQAST NYFGSLTQASS NYFGSLTQASS NYFGSLTQASS NYFGSLTQASS NYYGSLTQAST	VKLGTDAET IKLGHDAKT IKKLGSDAKT VKLGTDADT VKLGTDSKT VRVGNS-A VRVGNS-A VRVGNSN IRVGSFN IIRVGSFN IIRVGSFN IIRVGSYN IRVGSYN IRVGSYN	GADYNIPLQSL HQEINIPFHNI ARDINIPFHNY ARDYNIPFHDI GEEINIPFHDI GKOFHLPYCDL GEOYFAPHHEL GEEIYAPFKSL GEEIYAPFKSL GEEIYAPFKSL GEEIYAPFKSL GEEIYAPFKSL	LPHYHPNDL LPHYHPNDL LPHYHPNDL LPHYHPNDL LPHYHPNDL LPHYHPNDJ LPHYHPDDJ LPHYPPDDJ LPHYNPDDJ LPHYNPDDJ LPHYNPDDL LPHYNPDDL LPHYNPDDL LPHYNPDDL LPHYNPDDL	AT GOHDISGNI YIGGHDISGNI AIGGHDISGNI AIGGHDISGNI YIGGHDISGNI YIGGHDISGNI YIGGHDISGNI YFGGHDISNNI YFGGHDISNNI YFGGHDISNNI YFGGHDISNNI YFGGHDISNNI YFGGHDISNNI YIGGHDINANI	DLLASHIDARQQ VILAQANDRQQ VILAQANDRQQ VILAQANDRQQ VILADANDRQQ VILADANDRQQ VILADANDRQQ VILADANDRQQ VILADANDRQQ VILADANTRAKY VILADANTRAKY VILADANTRAKY VILADANTRAKY VILADANTRAKY	LUEPTLLKQLY LEPTLKQLY LEPTLKAQY LEPTLKAQLY LEPTLKQLU LPYDLQRQL LPYDLQRQL LDYDLQRQL LDTDLQKQL LDTDLQKQL LDTDLQKQL LDTDLQRQL LDTDLQRQL	EMNLNDY Kkenaghkpi Kkenaghvp Kkenadhvp Rkenaghkp Iphneslap Vpyhrenvp Rpyheshup Rpyheshup Rpyheshup Rpyheshup Rpyheshuc Rpyheskuc	PYNQAPLL LPSIYYPDFI LPSIYYPDFI LPSIYYPDFI LPSIYYPDFI LPSYYPDFI LPSYYPDFI LPSYYPDFI LPGIYDPDFV LPGIYDPDFV LPGIYDDFFI LPGIYPDFI SLASMIPTSS RPSAFDPKFV	Invindfykgi Caanqedrai Caanqedrai Caanqedrai Caanqedrai Caanqebrai Caanqebrai Caanqebrai Caanqebrai Caanqebrai Caanqebrai Spl Trdppan Yadnqlaba	YYRLK NWYLE NWYLE NWYLE NWYIP NUYIK NUYIK NUYIK TYSSR INYIK TYSSR INYIK
Sphaerulina Fusarium Metarhizium Colletotrichum Aspergillus Dacryopinax Chlanydomonas Chlanydomonas Setaria Glycine Avicennia Ostreococcus Saccharonycess	TYTAGI TYTAGI TYTAGI TYTAGI TYTAGI TFTATI TLTGAL TLTAGY TLTAGY TLTAGY TLTAGY TLTAGY TLTAGY	IGNRRNLTHE LANRRQLAHD LANRRNLYHE LANRRQLAHD LANRRQLAHD LANKNAYTHY SSNKRQISHA LANKEQISHA LANREGISHA LANREGISHA LANREGISHA LANREGISHA LANKERYTHR LANKERYTHR	TREGPRAN TKEGPRES TREGPRAN TREGPRAN TREGERAN TKEGIQYP ITAKGRAPA ITKDKLQA ITKDKLQA ITKDKYQQA TKCKYQQA TKCKYQQA TKCKYQQA TKEGAQEP	NYYGSVVHSSI NYYGSVVHGSI NYYGSVIHGSI NYYGSVIHGSI NYYGSVIHGSI NYFGSLVHSSI NIFGSVVHSSI NYFGSLTQASI NYFGSLTQASI NYFGSLTQASI NYFGSLTQASI NYYGSLTQASI NYFGSHTQCSI	VKLGTDAET ILKLGHDAKT VKLGTDADT VKLGTDADT VRLGTDSKT VRLGTDSKT VRVGANS-A VRVGNFE IRVGSYN IRVGSFQ IRVGSFN IRVGSFN CRVGHFE LKLGIDA-E	GADYNIPLQSI HQEINIPFHNU ARDYNIPFHNU GEEINIPFHDU GEOVFAPHHEL GEEVYAPFKSL GEEIYAPFKSL GEEIYAPFKSL GEEIYAPFKSL GEEIYAPFKSL GEEYYPFKSL GADVYAPFNSL	LPHYHPNDL LPHYHPNDL LPHYHPNDL LPHYHPDDL LPHYHPDDJ LPHYHPNDJ LPHYPPNDJ LPHYPPDDJ LPHYNPDDJ LPHYNPDDJ LPHYNPDDJ MPLADPNEY LPHYSPNDF	AT GOHDISGNI VIGGHDISKMI ATGGHDISKMI ATGGHDISGMI VIGGHDISGMI VIGGHDISGMI VIGGHDISGMI VFGGHDISNMI VFGGHDISNMI VFGGHDISNMI VFGGHDISNMI VFGGHDISNMI VFGGHDISNMI	DLLASINDRAQV NLAQANDRAQV SLADANDRAQV SLADANDRAQV NLADANDRAQV PLDEAMRRAEV NLAHANQRAV NLAHANQRAV NLABSTTRAKV NLADANGRAQG NLADANGRAQG NLADANTRAKV NLADANGRAQG JLYEAMQRSQV	LQEPTLLKQLY LEPTLLKQLY LEPTLLKAQY LEPTLKAQY LEPTLKQLY LPYDLQRQL LPYDLQRQL LDYDLQRQL LDUDLQKQL LDIDLQKQL LDIDLQKQL LDIDLQKQL LDIDLQRQL	ethilndy kkenaghkp kkenaghvp kkenadhvp kkenadhvp kkenadhvp vplmrgikp iphneslap vpymrehvp rpymeshlp rpymeshlp rpymeshlp rpymeshlp Rpymeshcl rpymegikp kaknslvkp	PVHQAPLE LPSIYYPDFI LPSIYYPDFI LPSIYYPDFI LPSIYYPDFI LPSYYPDFI LPSYYPDFI LPSYYPDFI LPGIYDPDFI LPGIYDPDFI LPGIYDPDFI LPGIYDPDFI LPGIYPDFFI LPSIYYPDFI	Invindfykgi (Aanuqedrai (Aanuqedrai (Aanuqedrai (Aanuqedrai (Aanuqedrai (Aanuqerai (Aanuqerai (Aanuqerai (Aanuqerai (Aanuqerai (Aanuqerai (Aanuqerai (Aanuqerai (Aanuqerai) (Aanuqerai) (Aanuqderai)	YYRLK NWVIP NWVLE NWVIP NWVIP NWVIP NWLIK NWVIK NWVIK NYIK NYIE NYIE NYIE NYIE

Fig. 7A. Multiple sequence alignment of MIPS complete amino acid sequences for selected species – alignment of N terminal portion of the MIPS sequence indicating the strictly conserved GXGGNNG motif

BLAST analysis was performed with *Chlamydomonas* MIPS set as a query and all remaining MIPS isoforms as subjects, then they also showed high similarity. Fungal members, as stated earlier, conserved some specific domains confined to themselves only.

The multiple sequence alignment of amino acids among the members selected for the whole study (Fig. 7) revealed some distinctly conserved domains (GWGGNNG, LWTANTERY, NGSPQNTFVPGL, SYNHLGNNDG) that support earlier observations highlighted in a previous report (Majumder et al. 2003). The authors state that a clear difference between prokaryotic and eukaryotic MIPS protein sequences was observed when compared among each other and that amino acid sequence alignment of the prokaryotic MIPS failed to show any striking similarity across different prokaryotic taxa, as seen in a comparison between eubacterial and archaebacterial MIPS sequences. It is noticeable that the selected set of subject species in the present study is almost completely different from the earlier one, because here attention was only paid to the plant members, whereas the previous one was based on an all living kingdoms. Nevertheless, the similar conservativeness detected in our analysis corroborates the species selection and methodology of the present study.

Moreover, we report here the existence of two additional domains (FRSKEISKSNVVDDMV, CEDSLLAS PLIIDL) that were found conserved amongst the selected species (Fig. 7). The later one was present in the prokaryotic system (*Bacillus, Oscillatoria*) as well, except for the initial 'CE' residues. The uniformity of these domains in some places was slightly deviated between the members of fungal group (Table 1).

It is now clear that the MIPS enzyme is so important in the biological system that it still continues to conserve and maintain its essential catalytic domains from the primitive time. Due to selection pressure, however, slight to major modifications occur in other portions of the sequence.

The study of molecular evolution includes the pattern and rate of substitution of nucleotide or amino acid residues at different sites. *In silico* tools enable an estimation of the rate and divergence of particular sequences through the sites. Logically, the substitution rate should be considerably lower in core functional domains of proteins than in other parts. If mutation in those domains still happen, they should have lesser effect on proteins' structure. This is so because an enzyme is to maintain its minimum functional property in any circumstances and therefore its core structure also. Accor-

101



Fig. 7B. Multiple sequence alignment of MIPS complete amino acid sequences for selected species – remaining portion up to C terminal of the same showing other five common/conserved domains

dingly, in the subsequent part of our investigation light has been shed first on the 3D structures of the MIPS proteins to predict their structural stability and then to perform the molecular clock analyses.

Protein 3D structure is important in understanding protein interactions, their functions and cellular localisation (Jones, 2001). Quaternary structure plays an essential role in defining protein functioning by facilitating allosterism and cooperativity in the regulation of ligand binding (Matthew et al., 2001). Homology modelling is the most common structure prediction method. In this regard, a template structure is needed. A template stands for the 3D structure of the same or related protein from another organism already visualised through crystallography. In our investigation most suitable templates for structural prediction of the MIPS sequences were identified using pdbBLAST. Finally, two prokaryotic MIPS of *Mycobacterium tuberculosis* (pdb id-1GR0) and Archaeoglobus fulgidus (pdb id-1U1I) were used to predict the MIPS 3D structures belonging to Oscillatoria and Bacillus. The only eukaryotic MIPS template reported for Saccharomyces cerevisiae (pdb id-1LA2) was employed for structural prediction of other MIPS proteins (Table 2). In this study, the quaternary structures of MIPS proteins of the individuals under investigation were predicted by considering the biological assembly that provided the functional state of the protein. As mentioned before, in order to function the protein interacts with an NAD+ molecule at the N terminal GXG GNNG motif (Kleizer and Eisenberg, 2002). It was also necessary to check whether the MIPS protein structures were stable with NAD+ substrate molecule or not. For this reason, the NAD+ ligand was also incorporated during MIPS model generation (Supplementary Fig. 1A-P).

Sr. no.	Conserved domain sequences	Features	Comments			
1	GWGGNNG	Rossman fold sequence for NAD binding	strongly conserved among all memberspreceded by 'MLV'			
2	LWTANTERY	eukaryotic conserved domain	• 'L' is strictly conserved from algae to angiosperm			
3	NGSPQNTFVPGL	eukaryotic conserved domain	 'PFI' Preceding the domain <i>Porteresia</i> is somehow unique in this domain the terminal 'L' is not conserved and distinctly different 			
4	SYNHLGNNDG	eukaryotic conserved domain	 preceded by 'AGIKPTISIV' in which 'P' is deviated in fungi as 'L' 'V' is present in <i>Chlamydomonas</i> and all higher plants and <i>Saccharomyces</i> closely followed by 'NLSS' where- terminal residue is 'S' in fungi and 'A' in all others 			
5	FRSKEISKSNVVDDMV	reported conserved domain	• Ostreococcus had only a little difference in some sites			
6	CEDSLLASPLIIDL	reported conserved domain	 except for 'CE', it is almost conserved in prokaryotes also the 8th residue is 'S/I' in fungi and 'A/L' in all others 			

Table 1. Features of the conserved domains of MIPS proteins identified from amino acid MULTALIN



Fig. 8. A model of MIPS monomer from *Oscillatoria acuminata*: A– 3D structural view of the protein, B – Ramachandran plot of the respective protein model

Basically, it was transferred to the model from the template structure where it is annotated as biologically relevant to this particular protein or enzyme. According to our results, NAD+ characteristically bound to the GWGG NNG domain which existed in an alpha-helix form in each analysed monomer (Fig. 8).

The straight-forward way to justify the predicted protein model as a simulator of the actual *in vivo* structural form is a computational estimation of its biophysical properties. In the present study per residue and overall geometries of the model protein structures were analysed with Ramachandran plots (Ramachandran et al., 1963). In a polypeptide, main chain N-C_a and C_a-C bonds

are relatively free to rotate. These rotations are represented by the torsion angles φ and ψ , respectively. Ramachandran plot analysis helps to visualise the dihedral angles φ vs. ψ of each amino acid residue from a quaternary protein structure. In the plot, residues lie in favoured (sterically allowed regions if strict or greater radii of the atoms are considered), allowed (sterically allowed if shorter radii of the atoms are considered) or outlier (sterically disallowed conformations) regions. A perfect model should have the least number of residues in outlier regions of the Ramachandran plot. Here, all the MIPS model structures in our experiments have their maximum (>90%) residues (Table 2, Supplementary Fig. 2)

No.	Organism	Template	Seq. identity (%)	Outlier
01 a	Oscillatoria acuminata (monomer)	1GR0	56.53	8/365 (0.2%)
01 b	Oscillatoria acuminata	1U1I	21.80	95/1458 (6.5%)
02	Bacillus cereus	1U1I	38.34	53/1564 (3.4%)
03	Aspergillus nidulans	1LA2	52.53	41/2125 (1.9%)
04	Metarhizium robertsii	1LA2	51.72	47/2137 (2.2%)
05	Fusarium oxysporum	1LA2	52.33	43/2153 (2%)
06	Sphaerulina musiva	1LA2	53.20	50/2121 (2.4%)
07	Colletotrichum higginsianum	1LA2	52.18	48/2209 (2.2%)
08	Ostreococcus tauri	1LA2	50.92	65/2127 (3.1%)
09	Chlamydomonas reinhardtii	1LA2	52.36	34/2015 (1.7%)
10	Chondrus crispus	1LA2	52.57	80/2260 (3.5%)
11	Physcomitrella patens	1LA2	51.90	35/2031 (1.7%)
12	Selaginella moellendorffii	1LA2	52.29	46/2032 (2.3%)
13	Glycine max	1LA2	52.20	41/2027 (2%)
14	Avicennia marina	1LA2	51.86	54/2023 (2.7%)
15	Porteresia coarctata	1LA2	44.10	62/2035 (3%)
16	Setaria italica	1LA2	51.85	42/2027 (2.1%)

Table 2. Summary of the overall quality analysis of predicted MIPS protein models belonging to selected members and the percentage of their identity with respective templates based on which the models are built

in favoured regions. Among prokaryotic members least amount of residues observed in outlier region was *Oscillatoria* MIPS monomer form (0.2%) and the maximum outlier was in *Oscillatoria* MIPS homotetramer form (6.5%). In eukaryotes the lowest percentage of outliers was present in *Chlamydomonas* and *Physcomitrella* (1.7%) and the highest was in *Chondrus* (3.5%). The prominence of residues in the αR and β regions suggested that the structure was rigid with more right helices.

Protein domains are compact and constitute more or less independent folding units of structure that recur in different molecular contexts; they are conserved, and are considered functional and evolutionary units of classification (Caetano-Anollés et al., 2009; Chothia and Gough, 2009). A number of popular protein domain classification schemes are available that are based on features like sequence and structure. For example, the Structural Classification of Proteins (SCOP) is one of the important taxonomical resources which groups domains with known 3D structures into families, super-families, and folds (Murzin et al., 1995).

Phylogenetic analysis of MIPS proteins among selected organisms (Fig. 9) provided useful information about the ancestry of the protein and closely related varieties. According to our results, the phylogenetic tree is rooted by cyanobacteria *Oscillatoria* MIPS protein and subsequently the *Saccharomyces* became the common ancestor for MIPS proteins of other, more complex eukaryotes including multicellular fungi. All fungal MIPS members belong to the same sister group. MIPS proteins from plant members analysed in our study were found to be clustered in the tree branches in a manner as they had previously been classified by plant scientists. So in other words, the MIPS sequence based tree strongly confirms the common plant taxonomical lineage.

The evolutionary history of the MIPS protein from prokaryotic cyanobacteria to highest evolved angiosperm through different plant groups and fungi was inferred using the UPGMA method (Sneath and Sokal, 1973). The optimal tree with the sum of branch length is 3.47994888 is shown in Fig. 10. The percentage of replicate trees in which the associated taxa were 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 (Zuckerkandl and Pauling, 1965) and are in



Fig. 9. Tree generated by UPGMA method showing evolutionary relationships of taxa inferred by MIPS amino acid sequence divergence (bootstrap values in percentages from 100 replicates and the branch lengths are indicated above and below the branches)



Fig. 10. Evolutionary relationships among selected taxa representing prokaryotes and different groups of eukaryotes (timetree) based on MIPS amino acid sequences, values next to the branches are relative times of divergence as revealed by RelTime method (Tamura et al., 2012) employing MEGA6

the units of the number of amino acid substitutions per site. The analysis involved 16 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 361 positions in the final dataset. Divergence times for all branching points in the topology were calculated with the RelTime method (Tamura et al., 2012) using the branch lengths contained in the inferred tree (Fig. 10) (information on statistical analysis were partially adapted from MEGA6 software).

The standard molecular clock describes a constant rate of molecular evolution and provides a powerful framework for evolutionary timescales. Proteins change by accumulating mutations while continuing to fold into stable structures, while preserving their functions. The interplay between stability and function is important in molecular evolution (Caetano-Anollés and Mittenthal, 2010). Experimental studies have shown that mutations in divergent lineages have only modest and additive effects on protein stability (Serrano et al., 1993) and that a substantial number of mutations promote mutational robustness and have no detectable effects on protein structure (Bloom et al., 2005). This suggests that much of protein evolution is neutral and provides strong support to the existence of a molecular clock. Molecular clock analysis of MIPS proteins revealed a much higher relative time of divergence from the prokaryotic cyanobacteria to higher eukaryote community, followed by the same trend, i.e a long time path of evolution also from Saccharomyces. Interestingly Porteresia MIPS is reflected as a highest evolved isolated entity from other angiospermic members probably due to its survival in extreme saline conditions with maximum necessity of this enzyme.

Conclusions

The availability of an enormous amount of sequence or structural data from a number of continuous and exhaustive wet lab experiments enabled the performance of an analytical study that allows us to draw evolutionary conclusions on MIPS proteins. It is clear that among different groups of plants the MIPS genes and proteins show complex evolutionary changes due to natural selection pressure through time. However, they were able to preserve some conservative core catalytic domains across all the varieties and present themselves as an example of a good fit for natural selection. It is well known that due to the low resolution and high R factors of many crystal structures, homology modelling is now regarded as an important technique not only for obtaining the 3D-structure of a putative proteins but also for refining the existing low accuracy experimental structures. The above analysis through homology models may provide an insight into 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.

Acknowledgements

The present study was not funded by any organisation. Infrastructure facility was partly supported by the PG Department of Botany, Barasat Govt. College and DBT Centre for Bioinformatics, Presidency University. The authors are grateful to these organisations.

We would like to express our sincere thanks to the editors and reviewers of this article before publication for their effort and critical evaluation.

References

- Abu-abied M., Holland D. (1994) *The gene cINO1 from Citrus* paradise is highly homologous to tur1 and Ino1 from the yeast and Spirodela encoding for myo inositol 1-phosphate synthase. Plant Physiol. 106: 1689.
- Adhikari J., Majumder A.L., Bhaduri T.J., Dasgupta S. (1987) Chloroplast as a local of L-myo-inositol-1-phosphate synthase. Plant Physiol. 85: 611-614.
- Agranoff B.W., Fisher SK. (2001) *Inositol, lithium and the brain.* Psychopharmacol. Bull. 35: 5-18
- Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. (1990) Basic local alignment search tool. J. Mol. Biol. 215: 403-410.
- Bachhawat N., Mande S. (2000) *Complex evolution of the inositol-1-phosphate synthase gene among archaea and eubacteria.* Trends Genet. 16: 111-113.
- Basak A., Jha T.B., Adhikari A. (2013) Biosynthesis of myoinositol in lycopods:characteristics of the pteridophytic l-

myoinositol-1-phosphate synthase and myoinositol-1-phosphate phosphatase from the strobili of Lycopodium clavatumand Selaginella monospora. Acta Physiol. Plant. 34: 1579-1582.

- Baud S., Vaultier M.N., Rochat C. (2004) *Structure and expression profile of the sucrose synthase multigene family in Arabidopsis.* J. Exp. Bot. 55: 397-409.
- Bloom J.D., Silberg J.J., Wilke C.O., Drummond D.A., Adami C., Arnold F.H. (2005) *Thermodynamic prediction of protein neutrality*. Proc Natl Acad. Sci. 102(3): 606-611.
- Caetano-Anollés G., Mittenthal J.E. (2010) Exploring the interplay of stability and function in protein evolution. Bio-Essays 32(8): 655-658.
- Caetano-Anolle's G., Wang M., Caetano-Anolle's D., Mittenthal J.E. (2009) *The origin, evolution and structure of the protein world.* Biochem. J. 417(3): 621-637.
- Chen L., Zhou C., Yang H., Roberts M.F. (2000) *Inositol-1*phosphate synthase from Archaeoglobus fulgidus is a class II aldolase. Biochemistry 39: 12415-12423.
- Chen V.B., Arendall W.B., Headd J.J., Keedy D.A., Immormino R.M., Kapral G.J., Murray L.W., Richardson J.S., Richardson D.C. (2010) *MolProbity: all-atom structure validation* for macromolecular crystallography. Acta Crystallogr. D66: 12-21.
- Chhetri D. R, Yonzone S, Mukherjee A. K, Adhikari J. (2006) *L-myo-inositol-1-phosphate Synthase from Marchantia nepalensis: partial purification and properties.* Gen. Appl. Plant Physiol. 32(3-4): 153-164.
- Chhetri D.R., Choudhuri M., Mukherjee A.K., Adhikari J. (2005) *L-myo-inositol-1-phosphate synthase: partial purification and characterization from Gleichenia glauca.* Biol. Plant. 49: 59-63.
- Chhetri D.R., Mukherjee A.K., Adhikari J. (2006) *Partial purification and characterization of Lmyo-inositol-1-phosphate synthase of pteridophytic origin.* Acta Physiol. Plant. 28: 101-107.
- Chothia C., Gerstein M. (1997) *Protein evolution. How far can* sequences diverge? Nature 385: 579-581.
- Chothia C., Gough J. (2009) *Genomic and structural aspects* of protein evolution. Biochem. J. 419(1): 15-28.
- Chun J.A. (2003) Isolation and characterization of a myo-inositol 1-phosphate synthase cDNA from developing sesame (Sesamum indicum L.) seeds: functional and differential expression, and salt-induced transcription during germination. Planta 216: 874-880.
- Chung C.H., Kim J.L., Lee Y.C., Choi Y.L. (1999) *Cloning and characterization of a seedspecific fatty acid desaturase cDNA from Perilla frutescens.* Plant Cell Physiol. 40: 114-118.
- Corpet F. (1988) *Multiple sequence alignment with hierarchical clustering.* Nucl. Acids Res. 16 (22): 10881-10890.
- Eswar N., Marti-Renom M.A., Webb B., Madhusudhan M.S., Eramian D., Shen M., Pieper U., Sali A. (2006) *Comparative Protein Structure Modeling With MODELLER*. Current Protocols in Bioinformatics. John Wiley & Sons, Supplement 15: 5.6.1-5.6.30.

- Felsenstein J. (1985) *Confidence limits on phylogenies: An approach using the bootstrap.* Evolution 39: 783-791.
- Hait N.C., RayChaudhuri A., Das A., Bhattacharyya S., Majumder A.L. (2002) Processing and activation of chloroplast L-myo-inositol 1-phosphate synthase from Oryza sativa requires signals from both light and salt. Plant Sci. 162: 559-568.
- Hazra A. (2015) Computational Fishing and Structural Analysis of MIPS Protein from Two Important Plant Groups. Int. Lett. Natural Sci. 42: 18-27.
- Hegeman C.E., Good L.L., Grabau E.A. (2001) Expression of D-myo-inositol-3-phosphate synthase in soybean. Implication for Phytic acid biosynthesis. Plant Physiol. 125: 1941-1948.
- Imhoff V., Bourdu R. (1973) *Formation d'inositol par les chloro plasts isoles de pois.* Phytochemistry 12: 331-336.
- Iqbal M.J. (2002) A pyramid of loci for partial resistance to Fusarium salani f. sp. Glysines maintains Myo-inositol-1-phosphate synthase expression in soybean roots. Theor. Appl. Genet. 105: 1115-1123.
- Ishitani M., Majumder A.L., Borhouser A., Michalowski C.B., Jensen R.G., Bohnert H.J. (1996) Coordinate transcriptional induction of myo-inositol metabolism during environmental stress. Plant J. 9: 537-548.
- Jin U.H. (2000) *Characterization of a methionine-rich storage* protein cDNA from perilla (Perilla frutescens) seed. Plant Physiol. 27: 701-707.
- Johnson M.D., Sussex I.M. (1995) 1-L-myo-inositol 1-phosphate synthase from Arabidopsis thaliana. Plant Physiol. 107: 613-619.
- Johnson M.D., Wang X. (1996) Differentially expressed forms of 1-L-myo-inositolphosphate synthase (EC5.5.1.4) in Phaseolus vulgaris. J. Biol. Chem. 271: 17215-17218.
- Jones D.T. (2001) *Protein structure prediction in genomics.* Brief Bioinform. 2(2): 111-125.
- Keller R., Brearley C.A., Trethewey R.N. (1998) Reduced inositol content and altered morphology in transgenic potato plants inhibited for 1D-myo-inositol 3-phosphate synthase. Plant J. 16: 403-410.
- Kido E.A., Ferreira Neto J.R., Silva R.L., Belarmino L.C., Bezerra Neto J.P., Soares-Cavalcanti N.M. (2013) Expression dynamics and genome distribution of osmoprotectants in soybean: identifying important components to face abiotic stress. BMC Bioinformatics. 14: 1-11.
- Kleiger G., Eisenberg D. (2002) GXXXG and GXXXA Motifs Stabilize FAD and NAD (P)-binding Rossmann Folds Through C α-H... O Hydrogen Bonds and van der Waals Interactions. J. Mol. Biol. 323: 69-76.
- Larson S.R., Raboy V. (1999) *Linkage mapping of maize and barley myo-inositol 1-phosphate synthase DNA sequences: correspondence with low phytic acid mutation.* Theor. Appl. Genet. 99: 27-36.
- Loewus F.A., Kelly S. (1962) Conversion of glucose to inositol in parsley leaves. Biochem. Biophys. Res. Commun. 7: 204-208.
- Loewus F.A., Murthy P.N. (2000) Myo- inositol metabolism in plants. Plant Sci. 150: 1-19.

- Majee M., Maitra S., Dastidar K.G., Pattnaik S., Chatterjee A., Hait N.C., Das K.P., Majumder A.L. (2004) A novel salt-tolerant L-myo-inositol-1-phosphate synthase from Porteresia coarctata (Roxb.) Tateoka, a halophytic wild rice. Molecular cloning, bacterial overexpression, characterization, and functional introgression into tobacco-conferring salt tolerance phenotype. J. Biol. Chem. 279: 28539-28552.
- Majumder A.L., Chatterjee A., Ghosh D.K., Majee M. (2003) Diversification and evolution of Lmyo-inositol 1-phosphate synthase. FEBS Lett. 553: 3-10.
- Majumder A.L., Johnson M.D., Henry S.A. (1997) *1L-myo-inositol 1-phosphate synthase*. Acta Biochim. Biophys. 1348: 245-256.
- Margulis L. (1996) *Archaeal-eubacterial mergers in the origin* of *Eukarya: phylogenetic classification of life.* Proc. Natl. Acad. Sci. 93: 1071-1076.
- Matthew D.G., Mark S.H. (2001) Quaternary structure of rice non-symbiotic hemoglobin. J. Biol. Chem. 276: 6834-6839.
- Michell R.H. (2008) *Inositol derivatives: evolution and functions.* Nature reviews Molecular Cell Biology. 9: 151-161
- Monizngo A.F., Marcote E.M., Robertson J.D. (1996) Chitinases, chitosanases, and lysozymes can be divided into procaryotic and eucaryotic families sharing a conserved core. Nat. Struct. Biol. 3: 133-140.
- Murzin A.G., Brenner S.E., Hubbard T., Chothia C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. 247(4): 536-540.
- Nandy Datta P., Dasgupta N., Das S. (2009) Differential expression of physiological and biochemical characters of some Indian mangroves towards salt tolerance. Physiol. Mol. Biol. Plants. 15(2): 151-160.
- Norman R.A., McAlister M.S.B., Murray R.J., Movahedzadeh F., Stoker N.G., McDonald N.Q. (2002) Crystal structure of inositol 1 phosphate synthase from Mycobacterium tuberculosis, a key enzyme in phosphotidyl inositol synthesis. Structure 10: 393-402.
- Pettersen E.F., Goddard T.D., Huang C.C., Couch G.S., Greenblatt D.M., Meng E.C., Ferrin T.E. (2004) UCSF Chimera – a visualization system for exploratory research and analysis. J. Comput. Chem. 25(13): 1605-1612.
- Ramachandran G.N., Ramakrishnan C., Sasisek-haran V. (1963) Stereochemistry of polypeptide chain configurations. J. Mol. Biol. 7: 95-99.
- Ray Chaudhuri A., Majunder A.L. (1996) Salinity induced enhancement of L-myo-inositol 1- phosphate synthase in rice (Oryza sativa L.). Plant Cell Environ. 19: 1437-1442.
- Sengupta S., Patra B., Ray S., Majumder A.L. (2008) Inositol methyl transferase from a halophytic wild rice, Porteresia coarctata Roxb. (Tateoka): regulation of pinitol synthesis under abiotic stress. Plant Cell Environ. 31(10): 1442-1459.
- Stein A.J., Geiger J.H. (2002) The crystal structure and mechanism of L myo inositol 1 phosphate synthase. J. Biol. Chem. 277: 9484-9491.

- Stevenson J.M., Perera I.Y., Heilmann I., Persson S., Boss W.F. (2000) *Inositol signaling and Plant Growth.* Trends Plant Sci. 5: 252-258.
- Tamura K., Battistuzzi F.U., Billing-Ross P., Murillo O., Filipski A., Kumar S. (2012) *Estimating Divergence Times in Large Molecular Phylogenies.* Proc. Nat. Acad. Sci. 109: 19333-19338.
- Tamura K., Stecher G., Peterson D., Filipski A., Kumar S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol. Biol. Evol. 30: 2725-2729.
- Yamaguchi-Shinozaki K., Shinozaki K. (2006) Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. Ann. Rev. Plant Biol. 57: 781-803.
- Zuckerkandl E., Pauling L. (1965) Evolutionary divergence and convergence in proteins. [in:] Evolving Genes and Proteins. Ed. V. Bryson, H.J. Vogel. Academic Press, New York: 97-166.