## **Review** Article

# The Role of Lectins and HD-ZIP Transcription Factors in Isoprenoid Based Plant Stress Responses\*\*

SANGITA KUMARI<sup>1</sup>, SMRITI SHRIDHAR<sup>1</sup>, DALJIT SINGH<sup>1</sup>, PIYUSH PRIYA<sup>1</sup>, ROHIT FARMER<sup>1</sup>, JASREET HUNDAL<sup>1</sup>, PRIYANKA SHARMA<sup>1</sup>, KRUTIKA BAVISHI<sup>1</sup>, KATHRIN SCHRICK<sup>2</sup> and GITANJALI YADAV<sup>1</sup>\*

<sup>1</sup>Computational Biology Laboratory, National Institute of Plant Genome Research, New Delhi 110 067, India <sup>2</sup>Division of Biology, Kansas State University, Manhattan, Kansas, USA

(Received 26 July 2011; Revised 15 March 2012; Accepted 29 March 2012)

It was over half a century ago when the overwhelming array of chemicals found in plants was postulated to be more than just by-products of primary metabolism. Ever since, extensive research has been conducted on plant secondary metabolites which are now known to be the end points of sophisticated survival mechanisms that plants have developed as a response to various kinds of stresses. Stress, defined by its negative effect on the growth and development of an individual, can be internal (metabolic or genetic), external (biotic or abiotic), permanent or acute. To cope, organisms must develop tolerance, resistance or avoidance mechanisms. Isoprenoids, often released as volatiles from plants, constitute the most diverse groups of natural products and play an essential part in plant defense systems, both directly (as emitted volatiles) and indirectly (the principle of inviting friends to feast on foes). Research over the last decade has resulted in a significant improvement in our understanding of the isoprenoid biosynthesis but there remains much to learn about the complex regulatory network controlling the various steps of these pathways and their dynamic co-ordination. Here we identify novel plant proteins and provide a putative role for them in isoprenoid based stress responses, along with insights into future perspectives for research.

#### Key Words : Isoprenoid Biosynthetic Pathway; Systems Biology; Network Analysis; Structural Modelling; Lectins; HD-ZIP Transcription Factors

### Introduction

For thousands of years natural products have been harvested for their medicinal properties in an effort that is now called Bioprospecting. The astonishing chemical diversity of biologically active substances in various organisms reflects an equally staggering diversity in function. Isoprenoids represent one of the most varied class of natural compounds in all organisms but their largest structural diversity occurs in plants where their roles range from primary metabolism with functions in photosynthesis, respiration, regulation of growth and development to secondary metabolism such as role in ecological interaction of plants with the environment [1-3]. The classification of isoprenoids is based on the number of five carbon units they have. The important classes of isoprenoids are hemiterpenes ( $C_5$ ), monoterpenes ( $C_{10}$ ), sesquiterpenes ( $C_{15}$ ), diterpenes ( $C_{20}$ ), triterpene ( $C_{30}$ ) and tetraterpenes ( $C_{40}$ ). Isoprene, the simplest isoprenoids having a single  $C_5$  unit is a volatile product significantly released by some plants. The emission of isoprene by plants is an important component of atmospheric biochemistry. Isoprene-emitting plants are better able to resist thermal and oxidative stress as compared to non-emitting plants

<sup>\*</sup>Author for Correspondence: E-mail: gy@nipgr.ac.in; Tel: +91-11-26735103; Mob. 9911126926

<sup>\*\*</sup>Based on the talk given by Dr. Gitanjali Yadav for INSA Medal for Young Scientists in 2011

[4]. Monoterpenes are the major constituent of essential oil responsible for plant originated flavors and fragrances. In addition to being component of essential oils, many sesquiterpenes are reported to act as phytoalexin, antibiotic and feeding deterrent that may help plants in defense and communication. The diterpenes are pharmacologically most important class of isoprenoids that include anticancerous taxol and forskolin (glaucoma treatment). The triterpenes and tetraterpenes include some vital primary metabolites such as brassinosteroids, phytosterols and carotenoids that are essential and usually evident [3].

Despite the diversity in function and structure,

all isoprenoids derive from the universal  $C_5$  precursors, isopentenyl pyrophosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP). The fundamental pathway of terpenoid biosynthesis has three distinct stages. The first stage in the pathway results in formation of the basic  $C_5$  isoprene units. This stage proceeds via two alternative, spatially separated pathways, the classical cytosolic mevalonic acid (MVA) pathway and the recently discovered, mevalonate independent, plastidic, methyl-erythritol phosphate (MEP) pathway as shown in Fig. 1. Interestingly, it was through bioinformatics and computational genomics efforts that the enzymes of the MEP pathway were first elucidated [5]. This

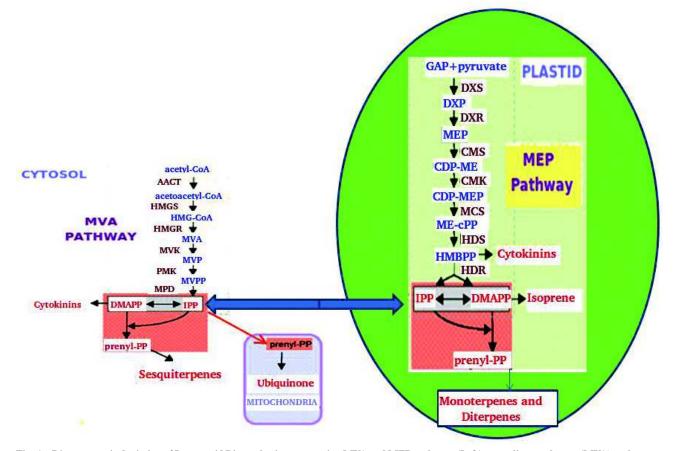


Fig. 1: Diagrammatic depiction of Isoprenoid Biosynthesis representing MVA and MEP pathway. (Left) cytosolic mevalonate (MVA) pathway: AACT, Acetoacetyl-CoA thiolase; HMGS, 3-hydroxy-3-methylglutaryl coenzyme A synthase; HMG-CoA, 3-hydroxy-3-methylglutaryl coenzyme A; HMGR, 3-hydroxy-3-methylglutaryl coenzyme A reductase; MVA, mevalonate; MVK, mevalonate kinase; MVP, mevalonate phosphate; PMK, phosphomevalonate kinase; MVPP, mevalonate diphosphate; MPD, mevalonate diphosphate decarboxylase; DMAPP, dimethylallyl diphosphate; IPP, isopentenyl diphosphate; Prenyl-PP, prenyl diphosphate; (Right) The plastidic MEP pathway: GAP, glyceraldehyde 3-phosphate: DXS, 1-deoxy-D-xylulose-5-phosphate synthase; DXP, 1-deoxy-D-xylulose-5phosphate; DXR, 1-deoxy-D-xylulose-5-phosphate reductoisomerase; MEP, methyl-erythritol phosphate; CMS, 2C-methyl-D-erythritol 4-phosphate cytidylyltransferase; CDP-ME, 4-Diphosphocytidyl-2-C-methyl-D-erythritol; CMK, 4-(cytidine 5'-diphospho)-2-C-methyld-erythritol kinase; CDP-MEP, 4-diphosphocytidyl-2-C-methyl-D-erythritol 2-phosphate; MCS, 2-C-methyl-D-erythritol 2,4cyclodiphosphate synthase; ME-CPP, methyl- erythritol 2,4-cyclodiphosphate; HDS, hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase; HMBPP, (E)-4-hydroxy-3-methyl-but-2-enyl diphosphate; HDR, 4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase. pathway is also a strong candidate for the discovery of novel antimicrobial and fungicidal drug candidates since it is absent in humans but essential for viability in several major pathogens, like the malarial parasite Plasmodium falciparum [6]. One of the major breakthroughs in the immune response to microbial pathogens came when an intermediate of the MEP pathway, namely (E)-4-hydroxy-3-methyl-but-2-enyl diphosphate (HMB-PP), was identified as the most potent activator known till date of the gammadelta T-cells [7]. The second isoprenoid biosynthetic stage is the condensation of two or three C<sub>5</sub> units to form  $C_{10}C_{15}$  or  $C_{20}$  prenyl diphosphates (PDPs), catalyzed by short chain prenyl transferases. PDPs such as geranyl diphosphate (GDP), farnesyl diphosphate (FDP) and geranylgeranyl diphosphate (GGDP) are the immediate precursors to mono-, sesqui- and diterpenes respectively. The second stage involves condensation of two or three  $C_5$  units to form  $C_{10}$ C<sub>15</sub> or C<sub>20</sub> prenyl diphosphates (PDPs), catalyzed by short chain prenyl transferases. In the final stage of terpene biosynthesis, these PDPs are converted to end products by a large family of unique enzymes called the terpene synthases (TPSs). In the final stage of terpene biosynthesis, these PDPs are converted to end products by a large family of unique enzymes called the terpene synthases (TPSs). One of the most outstanding properties of these enzymes is their ability to make multiple products from a single substrate.

Most of the work in characterization of the MEP pathway has been done with Escherichia coli proteins including the crystal structure of many of its component enzymes. By contrast only limited knowledge is available about the catalytic properties of plant enzymes. Several experiments involving terpenoid production have suggested that a significant coordination exists between both the MEP and the carotenoid pathways through the control of expression of key genes [8]. Although enzymes of isoprenoid biosynthesis have been elucidated and extensively studied through enzymology and crystallography [5] and considerable knowledge has been gained over the last two decades, there are several questions on physiological stress responses in plants involving individual proteins and genes that remain unanswered, thereby making the exploration of this acclimatization process and its regulation a complex and potentially rewarding field of study. Such a study is particularly relevant in an era when the complete sequences of several different organisms are now becoming available for the first time. Recent advances in genomics make it possible to identify a large number of genes as candidates for involvement in particular processes. For example, a computational genome wide analysis of functional linkages led to the identification of 27 novel pathways in bacteria [9]. A large number of genes encoding enzymes that synthesize volatile compounds have been reported. This well studied area is ideal for applying systems biology and for investigation of evolution not only of the enzymes individually but also of the pathways as a whole. In this report, we attempt to integrate recently gained knowledge from our ongoing work, with previously unknown aspects in this area, to place newly identified proteins on the isoprenoid based stress response map.

### Methods

### **Data Compilation**

This step involved a comprehensive large scale compilation of data, including genes, proteins, regulators, as well as interactions involved in the various events during terpene metabolism. Sequence and Structural data was compiled for both systems under focus, namely the MVA and the MEP pathways. Protein/Genetic Interaction data and mRNA expression values in addition to regulatory data were collected from Yeast (Saccharomyces cerevisae) (http://www.ncbi.nlm.nih.gov/genome/15), Arabidopsis (Arabidopsis thaliana) (http://www. ncbi.nlm.nih.gov/genome/4) and Rice (Oryza sativa) (http://www.ncbi.nlm.nih.gov/genome/10). The most comprehensive interaction data available to date is for S. cerevisiae at the BIOGRID website [10]. We also used STRING [11], and PAIR databases [12] for this analysis. Expression profile analyses were carried using PRIMe out [13], MPSS [14]. **GENEVESTIGATOR** [15] and VIRTUAL PLANT [16]. Protein Evolutionary rates and abundance measures for both the systems were extracted and collected from literature through PUBMED. Gene Ontology Data and Sub cellular localization data was extracted and compiled for the two systems from TAIR [17] and public databases. Homologues of various genes and enzymes involved in the MVA and MEP pathways were identified by a comparative analysis of available genomic and proteomic data. Specific sequence motifs and catalytic residues were identified for functional annotation of uncharacterized proteins. Genomic information was compared with 3D structural data wherever possible to enhance the predictive accuracy.

### **Computational Analyses**

Pairwise sequence alignment were carried out using BLAST [18], and the PHYLIP suite of programs [19], EMBOSS [20] and CLUSTALW [21] were used for phylogenetic analyses. Network analysis was carried out in Cytoscape [22], Pajek [23], R packages, in-

house FORTRAN and perl programs and UNIX scripting. For structural and cavity architectural analyses, we have made use of Discovery Studio version 3.0, VOIDOO [24], SCWRL [25], R packages, VMD [26] and PyMOL [27].

#### Results

#### **Pathway Evolution**

An attempt to understand the evolution of the two pathways revealed that the mevalonate pathway (MVA) enzymes all have archaebacterial ancestors, and show very frequent horizontal gene transfer. This pathway is widespread in archaea. In contrast, the MEP or non-mevalonate pathway is completely absent in archaea and the pathway enzymes show varied ancestries. For example, the enzyme 1-deoxy-D-xylulose-5-phosphate synthase (DXPS) appears to

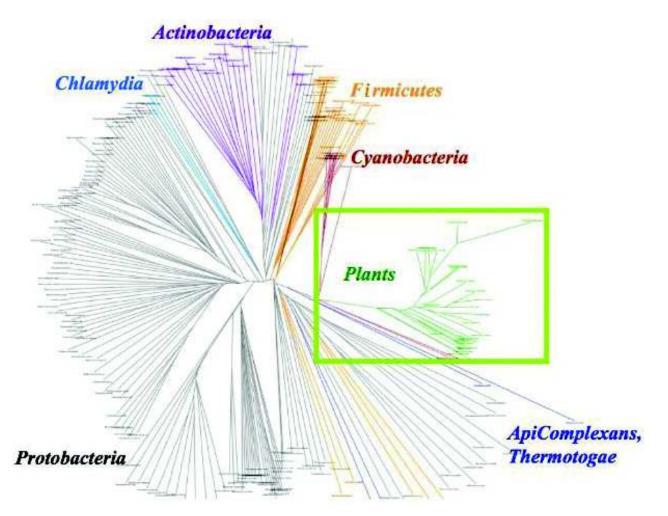


Fig. 2: Phylogenetic tree for the (a) 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR)

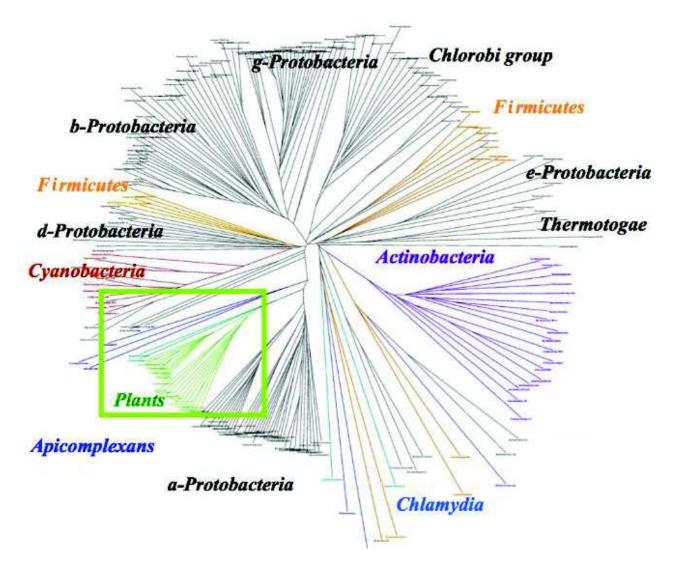


Fig. 2: Phylogenetic tree for the (b) 1-deoxy-D-xylulose-5-phosphate synthase (DXPS) enzymes of the Non-mevalonate pathway showing their varied ancestral decent. Green branches represent plants, other represent lower taxa such as proteobacteria (black), Chlamydia (cyan), Actinobacteria (purple), Cyanobacteria (red), and Firmicutes (orange). Note that plant DXR forms a sub-cluster within the cyanobacterial branch in this lineage, while the DXPS enzymes in plants appears to have evolved from a-proteobacteria (Gram negative organisms having different shapes)

have evolved from a-proteobacteria while 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR) from cyanobacteria as shown in Fig. 2 (a & b). 2-Cmethylerythritol-4-phosphate cytidyltransferase (MCT) and 4-(cytidine-5'-diphospho)-2-Cmethylerythritol kinase (CMK) have chlamydian ancestors, whereas 2-C-methylerythritol-2,4cyclodiphosphate synthase (MECPS) has reveals and aquifex ancestry. Several complexities are associated with these two pathways, which make them very difficult to analyse, the most important being the nomadic behaviour of the enzymes involved. The major controversy regarding the MEP pathway is its much debated cyanobacterial origin. The alternative hypothesis advocates the post-plastidial origin of this pathway via horizontal gene transfer. The last comprehensive study for these pathways was performed over a decade earlier and it included just about 14 organisms [28]. The current analysis, for the first time, includes information from more than 300 taxa, enabling much better sampling of sequence space and refinement of evolutionary inferences, than previously reported.

### Network Analysis and Interpretations

Using an integrated systems approach that enables assessment of multiple datasets together, rather than one source of data at a time, the information compiled for the isoprenoid biosynthetic pathways was superimposed in order to reduce data dimensionality, enabling a systems-level analysis of the various aspects and processes involved in plant stress response. The network analysis is based upon graph theory and our superimposition included data from sequence annotation, functional characterization, gene architecture, protein-protein interaction, genetic interactions, mRNA expression, regulatory interactions of transcription factors, gene ontology information and sub-cellular localization data. In this

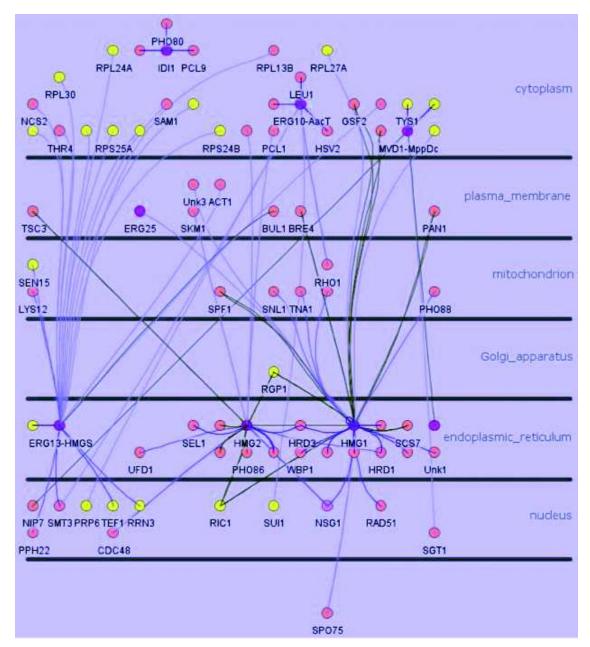


Fig. 3: Network Analysis of the mevalonic acid pathway showing integration of data from various sources as described in text. This model network of interactions was used to identify putative transcription factors which may be involved in the regulation of the isoprenoid pathways. Each circle in this figure represents a gene and each edge represents an interaction. Both physical (blue edges) and genetic (mint edges) interactions have been depicted. Yellow circles denote transcription factors while magenta circles represent pathway enzymes. Sub-cellular locations are also depicted, wherever known.

manner, a systems level understanding of the enzymes involved in the pathways was facilitated as an isoprenoid biosynthetic network shown in Fig. 3. Orthologs of the widely characterized yeast (S. cerevisiae) MVA genes and their reported interactors were searched in the available plant genomic data. In case of A. thaliana, this search resulted in a set of putatively interacting 'Interlog pairs'. Preliminary verification of predicted Interlog pairs was done by sub-cellular location data as well as mRNA expression levels, wherever available. In case of S. cerevisiae, 156 interactions have been reported for the eight genes of MVA pathway. For each of these 156, a study of the predicted A. thaliana interlogs was carried out to identify novel molecular players in the isoprenoid network that may have a role to play in regulation or cross-talk between the two pathways.

This analysis revealed the involvement and existence of novel genes, proteins and putative transcription factors. For example a total of twelve transcription factors were identified among the hydroxy-methyl glutaryl CoA synthase (HMGS) interlogs in Arabidopsis, of which two belonged to the type-III HD-ZIP family and contain a unique domain hitherto functionally uncharacterized in plants. Since HMGS is one of the key enzymes in the pathway, we studied this domain further in order to understand its possible implications, and the details are presented in the next section.

In addition to the interaction data, the isoprenoid network described above contained expression data also. A preliminary scan of the Arabidopsis genes was carried out to identify gene families enriched for induction during isoprenoid biosynthetic pathway expression. Among these, it was interesting to note that genes of enzymes in both the mevalonate and non-mevalonate pathways were found to co-express with several members of the lectin superfamily. In order to better understand these lectins and their roles in stress, we studied this superfamily in greater detail, as explained in the penultimate section.

### Interaction Data and the Start Domain

The candidate plant domain identified from

interaction data was classified as the steroidogenic acute regulatory protein (StAR)-related lipid transfer (START) domain, a protein domain spanning ~210 residues that has been known to bind and transport lipid molecules to the inner mitochondrial membrane in mammals [29]. This domain functions in a variety of distinct mammalian physiological processes, such as lipid transfer between intracellular compartments, lipid metabolism and modulation of signaling events. Mutation or misexpression of START proteins is linked to pathological processes, including genetic disorders, autoimmune disease and cancer [30]. However, the precise function of this domain in plants has remained obscure despite the recognition that it is highly amplified in plant genomes [31]. Sequence data from recently completed Plant Genome Projects showed that Vitis vinifera has 26 such domains, while Populus and Physcomitrella patens subsp. Patens have eight and 20 domains respectively. Arabidopsis has 35 homologs of this domain in its genome and several of these domains are present in homeodomain transcription factors (HD-START), suggesting a role in regulation of key cellular processes [32]. Fig. 4 shows an alignment of these domains with the closest structural counterparts depicting the secondary structural segments as well as the consensus regions. The identification of two of these domains in the transcription factors associated with the isoprenoid network of Arabidopsis implies their role in recognition of terpene based signals, possibly terpene based activation and downstream regulation of stress responses in plants. The domain may be hypothesized as a ligand-binding module that regulates transcription factor activity, since a few HD-START transcription factors have been reported to play key roles in plant cell development of disease resistance [33]. Other domains that are frequently found to be associated with this domain are bZip 1, MEKHLA, PH, DUF1336 and RPN2 (26S proteasome regulatory complex component) [31].

The mammalian START domains have a conserved 'helix-grip' fold that forms an inner tunnel wide enough to accommodate hydrophobic lipids such as cholesterol or ceramides [34]. Although the identity of the lipids that bind each START domain is known for only a few members of the family, it is

1EM2_A   PDBID	α <b>l</b> 2000000000000000000000000000000000000	ηι	β2 <b>TTT</b>	β3	a2 0000000	α3 2222	β4	->TTT-	β5	_	β6	÷ —	β7
2Z9Z_A/PDBID	al 000000000000000000000000000000000000	<b>Γ</b> β1β2		β4 β5	- 200	α2 ηΙ 000000 000		β6		β7	β8	>	TT
2Z9Z_A PDBID	FVQKVEEMVQNHMTYSLQDVGGDA	NWQLVVEEGEMKVY	RREVEENG	IVIDPLKATHA	VKGVTGHE	VCNYFWNVDV	RNDWETTIE	NFHVVET	LADNAII	IYQTHKRVWPASQRDVLY.		.LSVIRKI	PALTENDP
gi 332657038 gi 332659099	MLKLAERMVMSFCSGVGASTAHAW MLKLAERMVMSFCTGVGASTAHAW	TTMSTIGSD.DVRV TTLSTIGSD.DVRV	/MTRKSMDDPGRP /MTRKSMDDPGRP	PGIVLSAATSF PGIVLSAATSF	WIPVAPKR	VFDFLRDENS VFDFLRDENS	SRKEWDILSN SRSEWDILSN	GGMVQEM	IAHIANGE IAHIANGF	IEPGNCVSLLRVNSGNS	 	SQSNMLIL GQSNMLIL	QESCIDAS QESCIDAS
ai 1222422025	MT VI A FDMUTCFCACUCACTAUTS	ITTICCTCAE DUDI	IMTDVCUDDDCDD	DOTUTOAATCE	WIDDVDVDV	IPDET DDEMC	DNEWDTICN	COUVOEM	AUTANCE	DTCNCUCTIDUNCANC		COCMMITI	OFCCTDDT
gi 332656526	MLKLALWIGTONFCGGVCASSLQKW MLKLAQRMTFNFCSGISAPSVHNW ILKLAQRMKLNFYSGITASCIHKW IVKLAQRMKLNFYSGITASCIHKW	ISKLTVGNVDEDVRI	MIRKSVNNPGEP	PGIVLSAATSV	WLPAAPQRI	LYDFLGNERI	ARCEWDILSN	GGPMREM	IAHIAKGP IAHITKGÇ	DQGVSLLRSNAMNA	 	NQSSMLIL NQSSMLIL	QEISIDAA QETCIDAS
gi 332008796 gi 332659674	ILKLAQRMKLNFYSGITASCIHKW	EKLLAENVGQDTR I	LTRKSLEP	SGIVLSAATSL	WLPVTQQRI	LFEFLCDGKC	CRNQWDILSN	GASMENT	LLVPKGQ	QEGSCVSLLRAAGNDQ		NESSMLIL	QETWNDVS
gi 332198190	VLKMAQRMTQSFYRAIAASSYHQW	TKIT. TKTGQDMRV	SSRKNLHDPGEP	TGVIVCASSSL	WLPVNQAI	LFDFFRDEAF	RHEWDILIN	GAHVQSI	ANLSKG	DRGNSVAIQTVKS	 	REKSIWVL	QDSSTNSY
gi 332191531 gi 332197328	VLKMAQRMTQSFYRAIAASSYHQW IMRLAHRMVSNFCLSVGTSNNTRS MMRLAQRMISNYCLSVSRSNNTRS	TVVSGLDEF.GIRV	TSHKSRHEP	NGMVLCAATSF	WLP ISPON	VFNFLKDERI	RPQWDVLSN	GNSVQEV	AHITNGS	SNPGNCISVLRGFNASS		SQNNMLIL S NNMLIL	QESCIDSS
ori   3831459	MLKIAERIARTEFA, GMTNATGS	STIFSGVEGE DIRV	/MTMKSVNDPGKP	PGMITCAATSE	WIPAPPNTY	VEDELREATH	IRHNWDVLCN	GEMMHKI	AETTNGI	DKRNCASLIRHGHTSK		S. KMMIV	OETSTDPT
gi 332193611 gi 332005030	VMSLGERMLKNFAWIMKMSDKLDI VMHLGERMLRNFAWMMKMVNKLDF	.PQQSGANNS.GVR1 SPOSETNNS.GIR1	ISVRTN.TEAGQP IGVRIN.NEAGOP	PGLIVCAGSSL PGLIVCAGSSL	SLPLPPLQ	VYDFLRNLEV VYDFLKNLEV	/RHQWDVHCQ /RHOWDVLCH	GNPVTEA	ARFVTGE	DQKNNVTFLQPSSVG		.EYKLMIL .NTKLMIL	QDGFIDAL ODSFKDAL
gi 332640397	VMKLGERMVKNFNEMLTMSGKIDF LMKLSQRMVKTFCLNIINSHGQ	PQQSKN.GVRV	/SIRMN.IEAGQP	PGIVVSASSSL	AIPLTPLQ	VFAFLQNLDI	RQQWDILSY	GTVVNEI	ARIVTGS	SETNCVTILRVHPTHEENNI	DKMVVQDS	CKDDMLML	<b>QDCYMDAL</b>
gi 332658537 gi 332008060	IMRLSQRLVKTFCUNISTAYGQS	TALSETTKDIVF	RITTRKMCEPGQP	TGVVLCAVSTT	WLPFSHHQU	VFDLIRDQHH	IQSLLEVLFN	GNSPHEV	AHIANGS	HEGNSISLERINVESN	S S	WHNVELML	QESCIDNS
gi 332195204 gi 332641765	IMRLSQRLVKTFCVNISTAYGQS DYRKLWDKTVVEHEQLQVDSN.TC EYRKQWDKTVVEHEQLQVDSN.SC EFRSKWDDMLLYSSTLERCKD.TC	IEIGRTIKKFPLLT	CS.REYVLA	. WRLWQGKEKK	FYCFTKECI	DHNMVPQQRK	YVRVSYFRS	GWRIRQV	PGR	NACEIKMFHQENAG		. LNVEMAK	LAFSKGIW
gi 332196157	EFRSKWDDMLLYSSTLERCKD.TG	STMVVQWVRKFPFFC	CSDREYIIG	.RRIWDAGRV.	FYCITKGV	QYPSVPRQNK	(PRRVDLYYS	SWCIRAV	ESKRGDO	SEMTSCEVLLFHHEDMG	 	. IPWEIAK	LGVRQGMW
gi 332009079 gi 332643192	EFRPTWDTMLSNSTTVEECPS.TO	TMIVRWIRKFPFF(	SDREYVIG	. RRIWNCGNS.	YYCVTKGV:	SVPSIPPNNK	QKRVDLFYS	SWCIRPV	ESRRDDO	WTSACEVLLFHHEDMG		. IPREIAK	LGVKRGMW
gi 31376367	EFRPKWDFMLANFKTLEEDTQ.TC	TMIVQWRKKFPFFC	SDREYIIG	.RRIWESGKK.	YYCVTKGVI	PYPALPKRDK	PRRVELYFS	SWVIRAV	ESRKGDO	QQTACEVSLVHYEDMG	 	. IPKDVAK	LGVRHGMW
gi 330253013 gi 222424068	SPHATCNPPCCYGTTLPTDPSCDI NNTFFYGATLOKDPSLTI	PCSWTTTDPSTFL1	IRGKTYLDD	. OKKVKAKGT.	LMEMVA	AADWLKSDKF GADWISSDKF	REDDLGSRPG REDDLGGRIG	GIVQKYA GLVOEYA	AKGG	PEFFFIVNIQ	· · · · · · · · · ·	.VPGSTTY .VPGSAMY	SLVLYYMM SLALYYML
gi 332658730 gi 332007885	AHNIDLSCFSGNLKRNENENA	RNCWRISDGNNFK	RGKNFGQE	. KRKIPAGKH.	IMDLVA	AVDWFKDSKF	RIDHVARRKG	CAAQVAA	EKG	LFSMVVNVQ		.VPGSTHY	SMVFYFVM
gi 227202528	EFRSKWDDMLLYSSTLERCKD.TC EFRFWDTMLSNSTTVEECPS.TC EFRFKWDTMLAYFKTLEEDPK.TC EFRFKWDFMLANFKTLEEDTQ.TC SPHATCNPPCCYGTILPTDPSCDI NTFFYGATLQKDPSLTI AHNIDLSCFSGNLKRNENEN ALNIDLSCFSGNLKRNENEN SIDPSQFQGSLFKGNGCKT TPDLASSID	SNCWRISDGNNFK	RGKTYLKD	.NAKVMGGQP.	LATLIS	SVDWFKDIKE	VDNIALHPK	CLIQSEP	GKK	LPFILVINLQ	 	.VPGSIHY .VPAKPNY	CLVLYYAA
-11000050010	TIDINOBLEC OTRIDODINO.			1 CHARGE VICE V	VODIQIUM	LA TAPPPOOL	TODIOUD	ODDALTI	UDMIGDO	VOLUTONDEL CADOQUECO		DIDDIC	QLWDHODN OLWNHODN
gi 332009984	ILDLASALEGSIKQAGEADF TLDLTSSLEVG.PSPENASGNSFS TLDLASALEIGSAGTKASTDQSGN TLDLASALDVGSRTAGDSCGSKG LLNLSSFMVNVFCSGVCGITGQRN LLNLSSFMVNVFCSGVCGITGQRN	SSSSSRCILTIAFQF	PFENNLQENV.	AGMACQYVRSV	ISSVQRVA	MAISPSGISP	SLGSKLSPG	SPEAVIL	AQWISQS	SYSHHLGSELLTIDSLGSD.	 	DSVLK	LLWDHQDA
gi 222424857 gi 31711762	TLDLASALEIGSAGTKASTDOSGN TLDLASALDVGSRTAGDSCGSRGN	ISTCARSVMTIAFEE	GIESHMQE. HV	A SMARQY VRGI A SMARQY VRSV	ISSVQRVA	LALSPSHISS	SQVGLRTPLG	TPEAQTL SPEAHTL	ARWICQS	YRGYMGVELL. KSNSDGN		ESILK	NLWHHTDA
gi 332003746	LLNLSSFMVNVFCSGVCGITGQRW	INRLNTVGVSANNIF	RMFTQESRGMSGI	PCVLVSATGLA	RMHTKPEVI	MFGLINGAEK	QEIWSYLES	AKDMKEL	IRIGRHE	NSWNEVSVFSIEWKGS		KEWYLI	QETYYDES
gi 332659870 1EM2_A PDBID	EVAGRASESAUEREIIRUGREAIA	AVVDUI LAUEENWAP	EANNEIGDIVII	TEVPENGATET	TUTE TECEN	AFTAIOFAIT	JOPERMVLWN	NIVIA	COLLORY	EDNILISIDVSAGAAG		GVVSPR	DEVNVRRI
gi 57222140	VRFASESCIQELLSSSEAGRFGNA	SNGWKMVRHDSNG	/EISKRDSG	. SLHAFRSRRI	LTSVSPEQI	FITVANAIDA	AAKQWEGDLV	EATHIRD	IDENLSV	IRLRFGENSKPLFRNR		.EFIIYER	RETMQDGT
1EM2_A/PDBID	TT TT	TT		pri		000000	<u>α4</u> 000000000	000000	000				
	010	011					012		~~~~				
2Z9Z_A PDBID	β9 ★→ TT →	β11	.00.		η2 	e	β12	→	. 00000	α4 <u>000000000000000000000000000000000000</u>	Lee.		
2Z9Z_A   PDBID 2Z9Z_A   PDBID gi 1332657038	β9 TT ETWIVCAFSVDHDSAPLNNRCVR. GSVVIYAPVDIVAMNVVISCC	β11	PPEGNQEISRD		η2		β12	VNPGGW.	00000 APASVLR	α4 000000000000000000000000000000000000	KTAGKPI	LF	
2Z9Z_A PDBID 2Z9Z_A PDBID gi 332657038 gi 332659099	β9 TT β10 ETWIVCAFSVDHDSAPLNNCVR. GSYVIXAPVDIVAMNVVLSGG. .GSYVIXAPVDIIAMNVVLSGG.	β11 AKINVAMICQTLVS DPDYVALLPSGFAI DPDYVALLPSGFAI	PPEGNQEISRD LPDGSVGGGDG LPDGSARGGGGSA	ANA SAGAGVEG	η2 .NQHQEMVS SGEGNNLEV		β12 ILCKITYVA SLLTVAFQI SLLTVAFQI	VNPGGW. VNPGGW. LVDSVP. LVDSVP.	00000 APASVLR .TAKLSL .TAKLSL	α4 οοοοοοοοοοοοοοοοοοοοοοοοοοοοοοοοοοοο	KTAGKPI KTAGKPI VSCDVGG LACDGA.	LF GA	
2Z9Z_A PDBID 2Z9Z_A PDBID gi 332657038 gi 322659099 gi 222423025 gi 332646640	B <sup>9</sup> TT ETWIVCAFSVDHDGAPLNNRCVR. GSVUIXAPVDIVAMNVVLSGG. GSVVIXAPVDIVAMNVVLSGG. ASFVIXAPVDIVAMNVVLSGG. GAVVVXAPVDIPAMOAVMNG.	β11 AKINVAMICQTLVS DPDYVALLPSGFAI DPDYVALLPSGFAI DPDYVALLPSGFAI DSAYVALLPSGFAI	PPEGNQEISRD LPDGSVGGGDG LPDGSARGGGGSA LPDGNANSGAP LPDGOAGTORC.	NASAGAGVEG	η2 NQHQEMVS GGEGNNLEV		β12 ILCKITYVA SLLTVAFQI SLLTVAFQI SLLTVAFQI SLLTVAFQI	NVNPGGW LVDSVP. LVDSVP. LVDSVP. LVDSVP.	00000 APASVLR .TAKLSL .TAKLSL .TAKLSL .TAKLSL	α4 ουοοοοοοοοοοοοοοοοοοοοοοοοοοοοοοο ανακεγρκΕικεττουοοοοοοοοοοοοοο gsWatvnslikctverikaa gsWatvnslikctverikaa gsWatvnnlisctverikas esWetvnnlisctvokikaa	KTAGKPI SCDVGG LACDGA. MSCETA.	LF GA	
2Z9Z_A   PDBID 2Z9Z_A   PDBID gi   332657038 gi   332659099 gi   222423025 gi   33264640 gi   332656526	β9         TT           ETWIVCAFSVDHDSAPLNRCVR. GSVVIYAPVDIVAMVVLSGG.         GSVVIYAPVDIVAMVVLSGG.           ASFVIYAPVDIVAMVVLSGG.         GSVVVYAPVDIPAMQAVMNGG.           GAVVVYAPVDIPAMUAVMNGG.         GALVVYAPVDIPAMUAVMNGG.	β11 AKINVAMICQTLVS DPDYVALLPSGFAI DPDYVALLPSGFAI DPDYVALLPSGFAI DSAYVALLPSGFAI DSSYVALLPSGFAV	22. PPEGNQEISRD. LPDGSVGGDG. LPDGSARGGGGSP LPDGNANSGAP. LPNGQAGTQRC. LPDGGIDGGGS.	ANA SAGAGVEG	η2 NQHQEMVS GGEGNNLEV AAEERNSIG GDGDQRPVG	VTITGSCGG NGGCMEEGG	β12 ILCKITYVA SLLTVAFQI SLLTVAFQI SLLTVAFQI SLLTVAFQI SLLTVAFQI	VNPGGW LVDSVP. LVDSVP. LVDSVP. LVDSVP. LVNSLP. LVNNLP.	QQQQQ APASVLR .TAKLSL .TAKLSL .TAKLSL .TAKLSL .TAKLTV .TAKLTV	a4 avAkrey kFlkretsyve gsVatvnslikctverikaa gsVatvnslikctverikaa gsVatvnslikctverikaa gsVatvnslikctverikaa esVetvnnisctvokikaa	KTAGKPI VSCDVGG LACDGA. MSCETA. LHCDST.	LF	
2Z9Z_A   PDBID gi   332657038 gi   332657038 gi   322657039 gi   222423025 gi   332645640 gi   332655526 gi   332658576 gi   332659674	B <sup>9</sup> TT ETWIVCAFSVDHDAPLNRCVR. GSVVIYAPVDIVAMNVVLSGG. GSVVIYAPVDIVAMNVVLSGG. GSVVIYAPVDIVAMNIVLNGG. GALVVYAPVDIPAMQAVMNGG. GALVVYAPVDIPAMHVVMNGG. GALVVYAPVDIPSMNIVMSG.	β11 AKINVAMICQTLVS DPDYVALLPSGFAI DPDYVALLPSGFAI DPDYVALLPSGFAI DSSYVALLPSGFAI DSSYVALLPSGFAV DSAYVALLPSGFSI NSDSVKFLPSGFSI		ANA SAGAGVEG	η2 NQHQEMVS GGEGNNLEV AAEERNSIG GDGDQRPVG FDTDGGLVN		β12 ILCKITYVAFQI SLLTVAFQI SLLTVAFQI SLLTVAFQI SLLTVAFQI CLLTVGFQI	VNPGGW LVDSVP. LVDSVP. LVDSVP. LVNSLP. LVNSLP. LVNSLP. LVNSLP. LVSLP.	00000 APASVLR TAKLSL TAKLSL TAKLSL TAKLSL TAKLTV TAKLTV TAKLNV TAKLNV	a4 avgreedoorecoore gsvatvnslikctverikas gsvatvnslikctverikas gsvatvnnilsctverikas esvetvnnilsctvekikas esvetvnnilsctvekikas gsvatvnkilsctvekikas gsvetvnnilsctvekikas	KTAGKPI VSCDVGG LACDGA. MSCETA. LHCDST. LCCES. LRIPA.	LF	
2292_A   PDBID gi   332657038 gi   332657038 gi   322659099 gi   222423025 gi   332646640 gi   332058526 gi   332058796 gi   332659674 gi   332659674	β9 TT ETWIVCAFSVDHDAPLNRCVR. GSYVIYAPVDIVAMNVVLSGG. GSYVIYAPVDIVAMNVVLSGG. ASFVIYAPVDIVAMNVVLSGG. GALVVYAPVDIPAMQAVMNGG. GALVVYAPVDIPAMNYVMSG. GALVVYAPVDIPSMNTVMSG. CALVVYAPVDIPSMTUVMGG. ELVVVAPVDIPSMTUVMSG. CANVYAPVETNIELUVKGG.	β11 AKINVAMICOTIVS DPDYVALLPSGFAI DPDYVALLPSGFAI DPDYVALLPSGFAI DSAYVALLPSGFAI DSAYVALLPSGFAI DSSYVALPSGFSI DSSYVALPSGFSI DPSN-QILPSGFSI DPSN-QILPSGFSI	DO PPEGNQELSRD LPDGSVGGGGS. LPDGSARGGGGSP LPDGANASGAP LPNGQAGTQRC LPDGGIDGGGS LPDGSSSSDQ VPDGVNGSYHRG. IPDGVESRPLVII	NA SAGAGVEG	n2 NQHQEMVS GGEGNNLEV AAEERNSIG GDGDQRPVG FDTDGGLVN	UTTGSCGG VTTTGSCGG GGDGG NGGCMEEGG QESKG NTGGG QDDRNSQGG	β12 SLLTVAFQI SLLTVAFQI SLLTVAFQI SLLTVAFQI SLLTVAFQI CLLTVGFQI CLLTFGLQI SLLTLALQT	VNPGGW LVDSVP. LVDSVP. LVDSVP. LVDSVP. LVNSLP. LVNSLP. LVNSLP. LVNSLP.	20200 APASVLR .TAKLSL .TAKLSL .TAKLSL .TAKLSL .TAKLTV .TAKLTV .TAKLNV .TAALIQ .AAKLNW	a4 AVMKREYEKELKRETSYVOE GSVATVNSLIKCTVERIKAA GSVATVNSLIKCTVERIKAA GSVATVNSLIKCTVERIKAS ESVETVNNLISCTVQKIKAA ESVETVNLISCTVQKIRAA ESVETVNLISCTVQKIRAA ESVETVNLISCTVQKIRAA ESVETVNLISCTVAKIRAA	LOO KTAGKPI AVSCDVGG ALACDGA. MSCETA. ALHCDST. ALQCES. ALRIPA. ALDLQT. ALDLQT.	LF	
2292_A   PDBID 2292_A   PDBID gi   322657038 gi   322655039 gi   222423025 gi   332646640 gi   332646640 gi   332645626 gi   332659674 gi   332198190 gi   3321981531 gi   332197328	B <sup>9</sup> TT ETWIVCAFSVDHD <sup>C</sup> APLNNCUR, GSYVIYAPVDIVAMNVVLSGG, GSYVIYAPVDIVAMNVVLSGG, GSYVIYAPVDIPAMNVVLSGG, GAVVYAPVDIPAMNVVMGG, GALVVYAPVDIPAMNYMKGG, GAVVYAPVDIPSMNYMKG, SAALVIYPVDIPSLVKKGE, SAALVIYPVDIPSLVKKGC, GAVVYAPVDINTCUVLAGH, SAALVIYPVDIPALNIAMSCG, GAFVVYSPVDIAALNIAMSCG,	β11 AKINVAMICOTIVS DPDYVALLPSGFAI DPDYVALLPSGFAI DSAYVALLPSGFAI DSAYVALPSGFAI DSAYVALPSGFSI DSSYVALPSGFSI NSDSVKFLPSGFSI DTSYIPILPSGFAI DTSYIPILPSGFAI DTSYIPILSSGFTI	22. PPECNQEISRD. LPDGSARGGGCG. LPDGSARGGGSP LPNGQAGTQRC. LPDGSIDGGGS. LPDGSSSSDQ. VPDGVNGSYHRG. IPOGVESRPLVII SPDGSSKGG SPDGNGSNEQGG	ANA SAGAGVEG I I I I I I I I I I I I I I I I I I I	n2 NGHQEMVS GGEGNNLEV AAEERNSIG GDGDQRPVG FDTDGGLVN ST	VITTGSCGG VITTGSCGG NGGCMEEGG NGGCMEEGG QCDRNSQGG QDDRNSQGG SSGRASASG	β12 ILCKITYVAFQI SLLTVAFQI SLLTVAFQI SLLTVAFQI SLLTVAFQI CLLTFGLQI SLLTVAFQI SLLTVAFQI SLLTVAFQI SLLTVAFQI SLLTVAFQI SLITVGFQI	VNPGGW. LVDSVP. LVDSVP. LVDSVP. LVNSLP. LVNSLP. LVNSLP. LVSIP. LVGINP. LVGINP. VSGLQ. VVSNLP.	QQQQQ APASVLR .TAKLSL .TAKLSL .TAKLSL .TAKLSL .TAKLTV .TAKLTV .TAKLTV .TAKLNV .TAALIQ .PAKLNM .TAKLNM	a4 accession of the second se	KTAGKPI VSCDVGG LACDGA. LACDGA. LLHCDST. LLHCDST. LLRIPA. LLDLQT. LLCPSTA LLSGPTAS	LF. GA.	
2292_A   PDBID 2292_A   PDBID gi   332657038 gi   332655099 gi   222423025 gi   332645640 gi   3326556526 gi   332655674 gi   332198190 gi   332198190 gi   332197328 gi   3831459 ri   332193130	B T T T T T T T T T T T T T	β11 AKINVAMICOTLVS DPDYVALLPSGFAI DPDYVALLPSGFAI DPDYVALLPSGFAI DSSYVALLPSGFAI DSSYVALLPSGFAI DSSYVALLPSGFSI DSSYVALLPSGFSI DSSYVFILPSGFSI DTSYIPILPSGFAI DPSYIPLSGFAI DPSYIPILPSGFAI DPSYIPILPSGFAI	PPEGNQEISRD. LPDGSNGGGDG. LPDGSNGGGGSA LPDGSNGGGGSA LPDGGIDGGGS. LPDGGIDGGGS. VPDGVNGSYHRG SPDGNGSNSLQGC SPDGNGSNSLQGC SPDGNGSSESU	NASAGAGVEG(	η2 NOHQEMVS GGEGNNLEV AAEERNSIG GDGDQRPVG FDTDGGLVN 	VTTTSSGSGG VTTTGSCGG GGDGG NGGCMEEGG GG QDRNSQGG NTGGG QDDRNSQGG GG SSGRASASG GG	β12 ILCKITYVAF SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I CLITVGF0I SLITVAF0I SLITVGF0I SLITVGF0I SLITVGF0I SLITVGF0I SLITVGF0I SLITVGF0I	VNPGGW LVDSVP. LVDSVP. LVDSVP. LVNSLP. LVNSLP. LVNSLP. LVSGINP. LVSGLQ. VVSGLQ. VVSNLP. LVESGP. VVSQLQ.	APASVLR .TAKLSL .TAKLSL .TAKLSL .TAKLSL .TAKLTV .TAKLTV .TAKLNV .TAALIQ .PAKLNM .PAKLNM .EARLSV .TDINLK	a4 avjakretyk F Lkretsyve gsvatvnslikctverika gsvatvnslikctverika gsvatvnslikctverika gsvatvnslikctverika gsvetvnslistvov kika gsvetvnslistvov kika gsvetvnslistvikika gsvetvnslia gsvetvnsliktvikika gsvetvnsliktvikika gsvetvnsliverika gsvetvnslistvikika gsvetvnsliverika gsvetvnsliverika gsvetvnsliverika gsvetvnsliverika gsvetvnsliverika gsvetvnsliverika gsvetvnsliverika gsvetvnsliverika gsvetvnsliverika gsvetvnsliktverika gsvetvnsliktverika gsvetvnsliverika gsvetvnsliktverika gsvetvnsliverika gsvetvnsliverika gsvetvnsliverika gsvetvnsliverika gsvetvnsliverika gsvetvnsliverika gsvetvnsliktverika gsvetvnsliverika gsvetvn	KTAGKPI VSCDVGG LACDGA. MSCETA. LLACDST. LLQCES. LLDLQT. LLDLQT. LLCPSTA LSGPTAS FPCQTA. UNCE	LF	
2292_A PDBID 2292_A PDBID gi  332657038 gi  332657038 gi  322645640 gi  332645640 gi  3322645640 gi  332088796 gi  332198190 gi  332197328 gi  332197328 gi  332197328 gi  332193511 gi  332193511 gi  332105330	B T T T T T T T T T T T T T	β11 AKINVAMIC QTLVS DPDYVALLPSGFAI DPDYVALLPSGFAI DPDYVALLPSGFAI DSSYVALPSGFAI DSSYVALPSGFAI DSSVKFLPSGFSI DSSVKFLPSGFSI NSDSVKFLPSGFSI DPSNIQLPSGFSI DPSNIQLPSGFSI DPSYIPLLSSGFTI DPSYIPLLPSGFAI DPSTIPILPSGFFI DPSTIPILPSGFFI	PPEGNQEISRD. LPDGSNGGGDG. LPDGSNGGGGS LPDGSNAGGGS LPDGGIDGGGS. LPDGGIDGGGS. VPDGVNGSYHRG SPDGVNGSYHRG SPDGVNGSNSEQGG PPDGTGRPGKEC SRDSHPSSEV.	NA SAGAGVEG	η2 NOHQEMVS GGEGNNLEV AAEERNSIG GDGDQRPVG FDTDGGLVN ST	N TTSSCSCGG VTTTCSCGG GGDGG NGGCMEEGG G. GG GGDENSCGG GDDENSQGG GDDENSQGG GSGRASASG GSGRASASG G. GGSY HGGSY	β12 ILCKITYVAF SLITVAF0T SLITVAF0T SLITVAF0T SLITVAF0T CLLTVGF0T CLLTVGF0T SLITVGF0T SLITVGF0T SLITVGF0T SLITVGF0T SLITAF0T	VNPGGW LVDSVP. LVDSVP. LVDSVP. LVNSLP. LVNSLP. LVNSLP. LVSGIQ. VVSGLQ. VVSGLQ. VVSGLY. LVESGP. LVESGP.	APASVLR .TAKLSL .TAKLSL .TAKLSL .TAKLTV .TAKLTV .TAKLTV .TAALIQ .PAKLNM .PAKLNM .EARLSV YTDLNLK SPDTNLE	α4 ΔΟΟΟ Ο Ο Ο Ο Ο Ο Ο Ο Ο Ο Ο Ο Ο Ο Ο Ο Ο	KTAGKPI KTAGKPI MSCCTA. LACDGA. MSCETA. LLRCDST. LDLQT. LDLQT. LDLQT. LLDLPT. LLDLFTAS KILSGPTAS FFPCQTA. 1LNCE	LF. SA. TTA.	
2292_A PDBID gi  332657038 gi  332657038 gi  322659099 gi  222423025 gi  332645640 gi  3322645640 gi  332208796 gi  332298190 gi  332191531 gi  332191531 gi  332191532 gi  332193511 gi  3322055030 gi  3322055300 gi  332645377 gi  332645377 gi  332645357 gi  332645537 gi  332645547 gi  332655557 gi  3326555557 gi  3326555557 gi  3326555557 gi  3326555557 gi  33555557 gi  335555557 gi  3355555557 gi  33555555555555557 gi  3355555555555555555555555555555555555		β11 AKINVAMICQTLVS DPDYVALLPSGFAI DPDYVALLPSGFAI DPDYVALPSGFAI DSSYVALPSGFAI DSSYVALPSGFSI DSSYVALPSGFSI DSSYVALPSGFSI DSSYVALPSGFSI DPSNIQIDSGFSI DPSNIQIDSGFSI DPSNIDPSGFI DPSFIPILPSGFAI DPSFIPILPSGFVI DPSFIPILPSGFVI DFSFIPILPSGFVI	PPECNQEISRD. LPDCSNCGGCGC. LPDCSNCGGCS LPDCSNCGGCS LPDCGIDGGCS LPDCGIDGGSS LPDCSSSSDQ. VPDGVNCSYHRG. SPDCSNCSSSS SPDCGSNSSSEV SRDCRPSECA. SBDCRSSTVE. SDCRRSTVE.	NASAGAGVEG	η2 NOHQEMVS GGEGNNLEV AAEERNSIG GDGDQRPVG GDGDQRPVG TDTDGGLVN ST	TTSSCSCGG VTTTCSCGG GCGCGCG NGGCMEEGG GQESKG GNTGGG GDDRNSQGG GDDRNSQGG GSGRASASG GSGRASASG GGGSY DGGSY DGGSY	β12 ILCKITYVAF SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I CLLTVGF0I SLITVGF0I SLITVGF0I SLITVGF0I SLITVGF0I TLTVAF0I TLTVAF0I TLTVAF0I TLTVAF0I	VNPGGW. LVDSVP. LVDSVP. LVDSVP. LVNSLP. LVNSLP. LVNSLP. LVSLP. VSGLQ. VSGLQ. VSSCPSY LVSCPSY LVSCPSY LVSCPSY LVSCFSY	Q0000 APASVLR .TAKLSL .TAKLSL .TAKLSL .TAKLTV .TAKLTV .TAKLTV .TAKLNV .TAKLNM .EARLSV .TAKLNM .EARLSV .TAKLNK SPDTNLE RSREVNLE .TERLDL	a4 accorded accorded	KTAGKPI KTAGKPI MSCCTA. LLACDGA. MSCETA. LLHCDST. LLDLQT LLDLQT LLDLQT LLSGPTAS FFPCQTA. ILNCPEC. LLNCPEC.	LF. SA. TTA.	
2292_A PDBID gi  332657038 gi  332659099 gi  222423025 gi  332656526 gi  332656526 gi  33208796 gi  332191531 gi  332191531 gi  332191531 gi  332191531 gi  332191531 gi  332191531 gi  33219531 gi  3322059300 gi  332658577 gi  332008036	β9         β10           ETWIVCAFSVDHDAPUNMVVLSGG.         GSVVIYAPVDIVMNVVLSGG.           GSVVIYAPVDIVMNVVLSGG.         GSVVIYAPVDIPMNVVLSGG.           GSVVIYAPVDIPMNIVLMGG.         GALVVYAPVDIPMNIVLGG.           GALVVYAPVDIPMNIVLGG.         GALVVYAPVDIPMNIVLGG.           GALVVYAPVDIPMNIVLGG.         GALVVYAPVDIPMNIVLGG.           GARVYAPVDIPMNITUNGG.         GARVYAPVDIPMNITUNGG.           GARVYAPVDIPMUTIFMUTAGC.         GARVYAPVDISGG.           GGMVYAPVDIPMNITALASGG.         GGMVYAPVDISGG.           GGMVYAPVDIPMITGLVXACH.         SALUSSTVDPUACAALSGCV.           GGMVYAPMNIMITAYSAISGCV.         GGMIYAPMNIMITAYSAISGCV.           GGLIVYSTVDPVAVQLAMNATMHPAVSGEV.         GSLIVYSTVDPVAVQLAMNCE.	β11 AKINVAMICQTLVS DPDYVALLPSGFAI DPDYVALLPSGFAI DPDYVALLPSGFAI DSSYVALLPSGFAI DSSYVALLPSGFAI DSSYVALLPSGFAI DSSYVALLPSGFSI DSSNUALPSGFSI DPSNIQ1LPSGFSI DPSNIQ1LPSGFSI DPSNIPLPSGFFI DPSVIPLPSGFFI DPSVIPLPSGFVI DPSVIPLPSGFVI DPSVIPLPSGFVI DPSEIPLLPSGFVI	PPECNQEISRD. LPDCSNCGGCBG. LPDCSNCGGCBG LPDCSNCGGCS LPDCGNCGCC LPNCQAGTORC. LPDCSSSSDQ. POCSVCSSYHGG SPDCGTCSSSEV SRDCRPSECA. SSDCRSTVE. SSDCRSTVE. VPVNPSGISVMS	NASAGAGVEG G	η2 NOHQEMVS GEEGNNLEV AAEERNSIG GDGDQRPVG DTDGGCLVN ST	N TTSSGSCGG VTTTGSCGG NGGCMEEGG ODDNSQGG SSGRASASG SSGRASASG DGSN DGSN CGSS CSGS SGGS SGGS SGGS SGGS SGGS S	β12 ILCKITYVA SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVGF0I SLITVGF0I TLITVAF0I TLITVAF0I TLITVAF0I TLITVAF0I CLITVGF0I CLITVGF0I	WNPGGW. UVDSVP. UVDSVP. UVDSVP. UVNSLP. UVNSLP. UVNSLP. UVSLP. UVSGLQ. WVSGLQ. WVSGLQ. VVSGPSY. UVSGPSY. UVSGPSY. UVSGPSY. USGSVT. USSVP.	ACCOLO TAKISL TAKISL TAKISL TAKISL TAKIV TAKINV TAKINV PAKINM PAKINM EARLSV YTDLNIK SPDTNLE RSREVNE TERIDL TAKINM	a4 avgreaded by bar	KTAGKPI VSCDVGG MSCETA. MSCETA. LLHCDST. LLCES. LLCIPA LLCIPA LLCIPA LLCPTAS. FPCQTA. LLSPTAS. LLCE LLCCPEC LLSNTITP	LF. SA TTA VIASSADV	SN.
2292_A   PDBID 2292_A   PDBID gi   332657038 gi   332659099 gi   222423025 gi   332645640 gi   332658674 gi   33208796 gi   332191531 gi   332191531 gi   332191531 gi   332191531 gi   332191531 gi   332191531 gi   3322085030 gi   332643397 gi   332643397 gi   332643397 gi   332643397 gi   33268376 gi   332195204 gi   3321641765 gi   3321645767 gi   3321645767 gi   3321645767 gi   3321645767 gi   3321645767 gi   3321645767 gi   3321658577 gi   3321658577 gi   3321658577 gi   3321658577 gi   3321645767 gi   3321645767 gi   3321645767 gi   3321645767 gi   3321645767 gi   3321645767 gi   3321645767 gi   3321658577 gi   332168577 gi   337177 gi   337177 gi	β9         β10           ETWIVCAFSVDHDAPLNRCVR.         GSVVIYAPVDIVAMVVLSGG.           GSVVIYAPVDIVAMVVLSGG.         GSVVIYAPVDIPAMQAVMNGG.           GSVVIYAPVDIPAMQAVMNGG.         GALVVYAPVDIPAMQAVMNGG.           GALVVYAPVDIPAMQAVMNGG.         GALVVYAPVDIPAMQAVMNGG.           GALVVYAPVDIPAMDAVMNTULNGG.         GALVVYAPVDIPAMDAVMNGG.           GALVYAPVDIPAMDILVAGG.         GAVVYAPVDIPAMDNTWSGG.           GARVYYAPVDIPAMDILVAGG.         GANVYAPVDIPAMDILVAGG.           GARVYAPVTNIT         GANVYAPVDITS           GARVYAPVTNIT         GANVAPVDINATANSGE.           GGWVYAPVDIPAMDIT         GSALVYSPVDIPAVGANCE.           GGMVYAPMDMATMHPAVSGEV         GSLVYSTVDVDVGLANCE.           GGMVYAPMDMATMHPAVSGAVG.         GSLLYYSTVDVGLANCE.           SSVUCKMENALCKVILATSHRTQGP         SVUCKMENALCKVILATSHRTQGP	β11 AKINVAMICCTLVS DPDYVALLPSGFAI DPDYVALLPSGFAI DPDYVALLPSGFAI DSAYVALLPSGFAI DSSYVALLPSGFSI DSSYVALLPSGFSI DSSYVALLPSGFSI DPSNIQ1LPSGFSI DPSNIQ1LPSGFSI DPSYIPLSGFFI DPSYIPLPSGFFI DPSYIPLPSGFFI DPSTIPILPSGFFI DPSTIPILPSGFFI DPSTIPILPSGFVI DPSTIPLLPSGFVI DSSNIPILPGFSV DSSNIPILPGFSV DSSNIPILPSGFVI ILSAVSTMKKIPSE	PPECNQEISRD. LPDCSNCGGCG LPDCSNCGGCS LPDCSNCGGCS LPDCQACGC. LPDGSDCGCS LPDGSDCGCS LPDGSDSSSDQ. VPCVNCSNCS SPDCTGRCSSSEV. SDGRRSSEVE. SDGRRSSEVE. SDGRRSSTVE. VPVNPPGCISVNS LETGTDNVTGSME LESGTDDITNSSG	NASAGAGVEG( , , , , , , , , , , , , , , , , , , ,	η2 NOHOEMVS GGEGNNLEV AAEERNSIG DDCDQPVG PDTDGGLVN ST ST L L M	N TTSSGSCGG VTTTGSCGG NGGCMEEGG QDDRNSQGG SSGRASASG C DGRS SGRASASG C DGGS SGRASASG SSGRAS SSGRASASG SSGRASASG SSGRASASG SSGRASASG SSGRASASG SSGRASG SSG SSG	β12 ILCK TVVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I TLITAF0I TLITVAF0I TLITVAF0I TLITVAF0I TLITVAF0I CLITVGI0V KKMRKPSTK KLIRKPSKK	WVNPGGW. UVDSVP. UVDSVP. UVDSVP. UVNSLP. UVNSLP. UVNSLP. UVNSLP. UVSLP. UVSSLP. UVSSLP. UVSSKAN. UVSSKAN. USSSSKAN. USSSKAN. USSSKA	ADASULR TAKLSL TAKLSL TAKLSL TAKLSL TAKLTV TAKLTV TAKLTV TAKLTV TAKLNM PAKLNM PAKLNM PAKLNM EARLSV YTDLNLK SPDTNLE RSREVNE TERLDL TAKLDL LVG	a4 avia key ke lkretsyve gsvatvnslikctverikaa gsvatvnslikctverikaa gsvatvnslikctverikaa gsvatvnslikctverikaa e svetvnni sctvokikaa e svetvnni sctvokikaa e svetvnni sctvokikaa gsvetvnni sctvokikaa e svetvnni sctvokikaa gsvetvnni sctvokikaa e svetvnni sctvokikaa gsvetvnni sctvokikaa gsvetvnni sctvokikaa svetvni sctvokikaa svetvni sctvokikaa svetvni sctvokikaa svetvni sctvokikaa svetvni sctvokikaa svetvni sctvokikaa svetvni sctvokikaa svetvni sctvokikaa svetvi sctvokikaa svetvi sctvokikaa svetvi sctvokikaa stvo	KTAGKPI VSCDVGG LLACDGA. MSCETA. LLCDST. LLCDST. LLCDST. LLCPSTA LLSQTDS. LLCPSTA LLSGTTAS' FPCQTA. LLCPEC. LLCPEC. LLSNTITP' LLTKLNK. FLSKMRK	LF. SA 	SN. NQEVS. PQTTI. SQNAGI.
2292_A PDBID 2292_A PDBID gi  322657038 gi  322657038 gi  322652039 gi  222423025 gi  332645640 gi  332645640 gi  33208796 gi  332198190 gi  332197328 gi  332197328 gi  332197328 gi  332197328 gi  332197328 gi  3321932030 gi  3322640397 gi  3320080760 gi  3320840507 gi  3320840567 gi  3326417655 gi  332096157 gi  332090870		β11 AKINVAMICCTLVS DPDYVALLPSGFAI DPDYVALLPSGFAI DPDYVALLPSGFAI DSAYVALLPSGFAI DSAYVALLPSGFAI DSAYVALLPSGFSI DSSYVALLPSGFSI DFSNIQILPSGFSI DFSNIQILPSGFSI DPSTIPILPSGFII DPSTIPILPSGFII DPTIPILPSGFII DFSTIPILPSGFII DFSTIPILPSGFII DFSTIPILPSGFII DFSTIPILPSGFII DSSNIPILPSGFSI DSSNIPILPSGFSI DSSNIPILPSGFSI DSSNIPILPSGFSI DSSNIPILPSGFSI DSSNIPILPSGFSI DSSNIPILPSGFSI DSSNIPILPSGFSI DSSNIPILPSGFSI DSSNIPILPSGFSI DSSNIPILPSGFSI DSSNIPILPSGFSI DSSNIPILPSGFSI DSSNIPILPSGFSI DSSNIPILPLSFSI DSSNIPILPSGFSI DSSNIPILPSGFSI DSSNIPILPSGFSI DSSNIPILPSGFSI DSSNIPILPSGFSI DSSNIPILPSGFSI DSSNIPILPSGFSI DSSNIPILPSGFSI DSSNIPILPSGFSI DSSNIPILPSGFSI DSSNIPILPSGFSI DSSNIPILPSGFSI DSSNIPILPSGFSI DSSNIPILPSGFSI DSSNIPILPSGFSI DSSNIPILPSGFSI DSSNIPILPSGFSI DSSNIPILPSGFSI DSSNIPILSGFSI DSSNIPILSGFSI DSSNIPILPSGFSI DSSNIPILSG	PPEGNQEISRD. LPDGSVGGGDG. LPDGSNGGGGS LPDGSNGGGGS LPDGSNGGGS LPDGSNGGGS LPDGSVGGGS LPDGSVGSVHRG VPDGVNGSVHRG SPDGTGSVHRG SPDGTGRGSSE SCRSSPEGEA SDGRRSTVE. VPVNPSGVEGSS VPVNPSGISVNS LETGTDNVTGSMB LESGTDITINSSC VSABEFNMERGSISVNS VSABEFNMERGSISVNS	NASAGAGVEG( , , , , , , , , , , , , , , , , , , ,	η2 NOHQEMVS GGEGNNLEV AAEERNSIG BOCDQRPVG FDTDGGLVN ST ST LL LL	N TITSSGSCGG VITTGSCGG NGGCMEEGG ODDRNSQGG SSGRASASG 	β12 ILCKTYVA SLITVAF0I SLLTVAF0I SLLTVAF0I SLLTVAF0I SLLTVAF0I SLITVGF0I SLITVGF0I SLITVGF0I SLITVGF0I SLITVGF0I SLITVGF0I CLLTVGF0I CLLTVGF0I CLLTVAF0I CLLTVAF0I CLLTVAF0I CLLTVGI0V KKMRKPSTK KKMRKPSKK . KMPKIL	VNPGGW. UVDSVP. UVDSVP. UVDSVP. UVDSVP. UVNSLP. UVNSLP. UVNSLP. UVSGLQ. VVSGLQ. VVSGLQ. VVSGLQ. UVSGPSY. UVSGKANI USSPSY. USSGSVT.	ADAGUA TAKLSL .TAKLSL .TAKLSL .TAKLSL .TAKLTV .TAKLTV .TAKLTV .TAKLTV .TAKLTV .TAKLNM .PAKLNM .PAKLNM .PAKLNM .PAKLNM .TAKLNL .TAKLNL .TAKLLL	a4 accession of the second se	LLC KTAGKPI VSCDVGG LACDGA. MSCBTA. LHCDST. LLCST. LQCES. LDLQTA. LDLQTA. LDLQTA. LDLQTA. LLCSCTA. LLCSCTA. LNCPEC. LNCPEC. LNCPEC. LNCPEC. LNCPEC. KRM. K	LF JA TTA VIIASSADV. GGSLNQT TTPDSSOT	SN. NQEVS. TQTTTI. SQNAGI. Teo
2292_A PDBID 2292_A PDBID gi  32655038 gi  322655039 gi  222423025 gi  332655626 gi  332655626 gi  332208796 gi  332198190 gi  33219310 gi  33219310 gi  33219320 gi  332640397 gi  33206610 gi  3322195204 gi  332219517 gi  332208179 gi  33219517 gi  33219517 gi  33219517 gi  33219517 gi  33219517 gi  33219517 gi  33219517 gi  332196175 gi  33208079 gi  33208079 gi  33208079 gi  33208079 gi  332081987 gi  33208319 gi	β9         β10           ETWIVCAFSVDHDSAPLNNRCVR. GSYVIYAPVDIVAMNVVLSGG.         GAPLNNRCVR. GSYVIYAPVDIAMNVVLSGG.           GSYVIYAPVDIYAMNVVLGG.         GAVVYAPVDIYAMNTVNGG.           GALVYAPVDIPAMHVVMRG.         GAVVYAPVDIPAMHVVMRG.           GALVYAPVDIPAMHVVLOVGG.         GALVYAPVDIPAMITVNGG.           GALVYAPVDIPAMITVNGG.         GAVVYAPVDIPAMITVNGG.           GALVYAPVDIPAMITNNGG.         GAVVYAPVDIPAMITMGV.           GALVYAPVDINTNGG.         GGAVYAPVDINTMGV.           SALVIYPVDALANIAMSG.         GGAVYAPVDINTMGV.           SALVIYPVDALANIAMSG.         GGMVYAPVDINTMGC.           GGMVYAPVDINTMGV.         GGMVYAPVDINTMGV.           GGMVYAPVDINTMACAAISGDI         GGMVYAPMDINTACAAISGDI           GGMVYAPMNLMATAHFAVSGEV.         GSLIYYSTVDPVO VOLAMICE.           SVCKMENALCKYLATSHRPGPF         GAVKKLEPCIRAYCARAAGAG.           GAVKKNEPCIRAXCARAAGAG.         GAVKKNEPCIRAXCARAAGAG.	β11 AKINVAMICOTLVS DPDYVALLPSGFAI DPDYVALLPSGFAI DPDYVALLPSGFAI DSSYVALLPSGFAI DSSYVALLPSGFAI DSSYVALLPSGFAI DSSYVALLPSGFAI DSSYVALLPSGFAI DSSYVALLPSGFAI DPSNIDLPSGFAI DPSNIDLSGFAI DPSTIPILPSGFAI DPSTIPILPSGFAI DPSTIPILPSGFAI DPSTIPILPSGFVI DPSTIPILPSGFVI DPSTIPILPSGFVI DPSTIPILPSGFVI DSSNIPILPVGFS ILSAVIAMANINTK LSSPSAFMAGINTK	PEGNQEISRD. LPDGSNGGGDG. LPDGSNGGGGSA LPDGSNGGGGSA LPDGSNGGGSA LPDGSSSSDQ. UPDGVNGSYHRG SPDGNGSVHRG SPDGNGSNSUGG SPDGNGSNSUGG SPDGNGSSSU SRDSHPSSEV SRDSHPSSEV SRDSHPSGVEGSS VPVNPSGVEGSS VPVNPSGVEGSS VPVNPSGVEGSS ITGDHLISLSNG ITDHLISLSNG	NASAGAGVEG	η2 NOHQEMVS GGEONNLEV AAEERNSIG GDGDQRPVG FDTDGGLVN ST ST L M		β12 ILCKITYVA SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVGF0I SLITVGF0I SLITVGF0I SLITVGF0I SLITVGF0I SLITVAF0I TLIVAF0I TLIVAF0I CLITVAF0I CLITVAF0I SLITVAF0I RAANKKLL RAANKKLL RAANKKLL RAANKKLL	VVNPCGW/ VVDSVP. VDSVP. VDSVP. VDSVP. VVSSLP. VVSSLP. VVSSLP. VVSSLQ. VVSSLQ. VVSSLQ. VVSSLQ. VSSLAN. UVSSLAN. LASNVP. LASNVP. LASNVP. LASNVP. LASNVF. LASN	ACOULD A PASVLR .TAKLSL .TAKLSL .TAKLSL .TAKLSL .TAKLTV .TAKLTV .TAKLNV .TAKLNV .TAKLNM .PAKLNM .PAKLNM .EARLSV YTDLNLK SPDTNLE .TERLDL .TAKPLD .TAKPD .TAKP	CAL CONTRACTIVE INFORMATION AVAKRETYKE LKRFTSYVOE GSVATVNSLIKCTVERIKAA GSVATVNSLIKCTVERIKAA GSVATVNSLIKCTVERIKAA GSVATVNSLIKCTVERIKAA GSVETVNLISCTVQKIKAA ESVETVNLISCTVQKIKAA GTVKSVETIMAHTIVKIKSS ESVETVNLINTTVHQIKTA SSVATTVNILSSTVQRIKAA MSVTVNILSSTVQRIKAA MSVTVNILSSTVQRIKAA SSTVTINNHLCATVNQISS STVTINNHLCATVNQISS STVTINNHLCATVNQISS GALCLSRCHSALGAKVALAY GALCLSRCHSALGAKVALAY GALCLSRCHSALGAKVALAY GALCLSRCHSALGAKVALAY	LOO KTAGKPI VSCDVGG LACDGA. MSCETA. LLHCDST. LQCES. LLQCES. LLQCES. LLQCES. LLQCETA. LLGCTA. LLGCTA. LLSGPTAS. FPCQTA. LLSCFTAS. LNCPEC. LLSCFTAS. SKRM RKRM.C.	LF SA UIASSADV. GAQLSQT TRSESQT	SNVS NOEVS COTTI SONAGI TSS
2292_A PDBID 2292_A PDBID gi  332657038 gi  332657039 gi  222423025 gi  332645640 gi  3322645640 gi  332208796 gi  332198190 gi  332198190 gi  332193126 gi  33219310 gi  3322635337 gi  332058537 gi  332058537 gi  33205854 gi  332195204 gi  332195204 gi  3322196157 gi  3322196157 gi  3322196157 gi  3322196157 gi  3320979 gi  332643192 gi  332643192 gi  33263457 gi  332643192 gi  33263457 gi  332643192 gi  33263457 gi  332643192 gi  33263457 gi  33263457 gi  33263457 gi  332643192 gi  33263457 gi  332557 gi  33557 gi  335577 gi  335577 gi  335577 gi  335577 gi  335577 gi  3355777 gi  3355777 gi  33557777 gi  335577777777777777777777777777777	P     TT     FIT     FIT	β11 A K I NVAMIC C T LVS D D YVALLP SGFAI D D YVALLP SGFAI D D YVALLP SGFAI D D YVALLP SGFAI D S YVALLP SGFSI D S YVALLP SGFSI D S YVALLP SGFAI D S YVALLP SGFAI S S SAVAALA STTK S S S S S S S S S S S S S S S S S S S	PEGNQEISRD. LPDCSNGGGDG. LPDCSNGGGGS LPDCSNGGGGS LPDCGNCGGS LPDCGIDCGGS VPDCVNGSYHRG SPDG. SPDG. SPDG. SPDG. SSDGRSS SDGRPS SDGRSTVE. SDGRST	NASACAGVEG	η2 NOHQEMVS GGEONNLEV AAEERNSIG GDGDQRPVG FDTDGGLVN ST ST LL M		β12 ILCKITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITSF0M TLITVAF0I TLITVAF0I TLITVAF0I TLITVAF0I TLITVAF0I TLITVAF0I TLITVAF0I TLITVAF0I TLITVAF0I TLITVAF0I TLITVAF0I TLITVAF0I TLITVAF0I TLITVAF0I TLITVAF0I CLITVGI0V CLITVGI0V SUBSTANISKIL RAANISKIL DAGSODWIKM DHIRNDWIKM	VVNPGGWA VVNPGGWA VDSVP. VDSVP. VVSVP. VVSVP. VVSVP. VVSVP. VVSVP. VVSVP. VVSVP. VVSVP. VVSVP. VVSVP. VVSVP. VVSVP. VVSVP. VSVP VSVP	ACCOUNT A PASVLR .TAKLSL .TAKLSL .TAKLSL .TAKLSL .TAKLTV .TAKLTV .TAKLNV .TAKLSL	a4 avg cooccoccoccoccoccoccoccoccoccoccoccocco	LAD KTAGKPI KTAGKA MSCETA. LACDGA. LHCDST. LQCES. LLRIPA. LQCESTA. LDLOT. LQLEDC. LNCESTA LSCETAS. HKCE LNCEC. LNCEC	LF SA TTA VIASSADV. GGTSLNQT TTRSESQT TTRSESQT THLIDASK	SN NOEVS. TOTTII. SONAGI. TSS.
2292_A PDBID 2292_A PDBID gi  332657038 gi  332657038 gi  322645640 gi  332655029 gi  332645640 gi  3322008796 gi  332198190 gi  332197328 gi  332197328 gi  332197328 gi  332195201 gi  3322050300 gi  3322050300 gi  3322050300 gi  3322050300 gi  3322195204 gi  332195204 gi  332195204 gi  332195204 gi  332045130 gi  332045130 gi  332045130 gi  3320645120 gi  33263501 gi  33263501 gi  32224240666	β9         β10           ETWIVCAFSVDHDSAPLNRCVR.         GSYVIYAPVDIVAMVVLSGG.           GSYVIYAPVDIAMVVLSGG.         GSYVIYAPVDIAMVVLSGG.           GSYVIYAPVDIAMVVLSGG.         GSYVIYAPVDIAMVVLSGG.           GSYVIYAPVDIAMVVLSGG.         GAVVYAPVDIAMVVLSGG.           GAVVYAPVDIAMVVLSGG.         GAVVYAPVDIAMVVLSGG.           GAVVYAPVDIAMVVLSGG.         GAVVYAPVDIAMVVLSGG.           GAVVYAPVDINTOLILLAGA.         GAVVYAPVDINTOLILAGA.           GAVVYAPVDINTOLILAGAS.         GAKVYAPVDINTOLILAGAS.           GAVVYAPVDINTOLILAGAS.         GGMVYAPMUNNIAASSAISGQV.           GGMVYAPMUNNIAAYSAISGQV.         GGMVYAPMUNNIACASAISGQV.           GGMVYAPMUNNIAASSAISGQV.         GGMVYAPMUNNIACASAISGQV.           GSLIYYSTVDVQVQULAMNCE.         SVCKMENALCKYLATSHRTQCP.           GAVKKLEPGLRAYCTHELSBGGT.         GAVKKLEPGLRAYCSARKSDGS.           GAVKKLNSGLRAYCSARKSPGS.         GAVKKLNSGLRAYCSARKSPGS.           GAVKKLNSGLRAYCSARKSDS.         STPIEEHPLLVSFVNGDAYNS	β11 AKINVAMICCTLVS DPDYVALLPSGFAI DPDYVALLPSGFAI DPDYVALLPSGFAI DPDYVALLPSGFAI DSSVVALPSGFAI DSSVVALPSGFAI DSSVVALPSGFSI DSSVVFLPSGFSI NSDSVKFLPSGFSI NSDSVKFLPSGFSI DSSVVFLPSGFSI DPSNIDLPSGFSI DPSTIPLPSGFAI DPSTIPLPSGFSI DPSTIPLPSGFSI DFSTIPLPSGFVI DFSTIPLPSGFVI DFSTIPLPSGFVI DFSTIPLPSGFVI DSSNTPILPUSGSVI LSSAVSLMKKIPSE LSSAVAALTTK LSSSAMASITTK LSSSAMASITTK LSSSAMASITTK LSSSAMASITTK RFKLIPYLSKGSWI	PECNQEISRD. PPECNQEISRD. PPOSARGGGSA LPDCSARGGGSA LPDCGIDGGGS. LPDCGIDGGGS. VPDGVNGSYHRG SPOGUNGSNSEQGC PPOGTGRPGKEC SRDSRPSSEV. SRDGRPSEGEA. SSDCRSTVE. SDCRRSTVE. SSDCRSTVE. LESGTDDITNSSE USADEMERGSI ITODHLISLSNG LNMDLVETSGAEL LNMDLVETSGAEL NMDLVESS.SKL	NASAGAGVEG	η2 NOHQEMVS GGEGNNLEV AAEERNSIG GDGDQRPVG DTDGGLVN ST ST		β12 ILCKITYVA SLITVAFOI SLITVAFOI SLITVAFOI SLITVAFOI SLITVAFOI SLITVAFOI SLITVAFOI SLITVGFOI SLITVGFOI SLITVGFOI LITVAFOI TLITVAFOI TLITVAFOI TLITVAFOI TLITVAFOI CLITVAFOI KLIRNPSKK KLIRNPSKK KLIRNPSKK KLIRNPSKK LANDPSTK KLIRNPSKK LANDPSTK KLINGPSTK KLINGPSTK KLINGPSTK KLINGPSTK KLINGPSTK LGUDGSST	WWNPGGW UVDSVP. UVDSVP. UVDSVP. UVDSVP. UVDSVP. UVDSVP. UVNLP. UVNLP. UVSLP. UVSCPV. UVSCPV. UVSCPV. UVSCPV. USSCANN UVSGVA. USSCANNP. ULASGNV. USSCANNP. US	ACCOUNT TAKLSL TAKLSL TAKLSL TAKLSL TAKLYV TAKLYV TAKLYV TAKLYV TAKLNV TAKLNV TAKLNN TAKLN	a4 accorded a construction of the second and and a construction o	LACDGA. KKTAGKPI LACDGA. MSCETA. LLACDGA. LLQCBS. LLQCBS. LLQCBS. LLQCBS. LLQCBS. LLQCBS. LLQCBS. LLQCSS. LNCPSC. LLNCPSC. LLNCPSC. LLNCPSC. LSNTIP. LSNTIP. LSNTIP. LSNTIP. KKRQLEG 	LF SA TTA VIASSADV. GGTSLNQT TTRSESQT FNHLDASK. NYLDVSK	SN. NQEVS. TQTTII. SQNAGI. ISS. ISS. ISSIIP.
2292_A PDBID gi  332657038 gi  332657038 gi  322655039 gi  222423025 gi  332655626 gi  3322645640 gi  332208796 gi  332198190 gi  332199130 gi  332199130 gi  33219531 gi  33219531 gi  33219531 gi  33219537 gi  332058337 gi  332058337 gi  332058237 gi  332058237 gi  332195204 gi  332195204 gi  332195204 gi  332195204 gi  33219517 gi  33209079 gi  332263817 gi  332263837 gi  3322424068 gi  3322638730 gi  33200785 gi  33200785 gi  3322638370 gi  33268730 gi  33200785 gi  332268730 gi  33200785 gi  33200785 gi  33200785 gi  33200785 gi  33200785 gi  33220785 gi  33200785 gi  33200785 gi  332200785 gi  33220785 gi  33220785 gi  33200785 gi  33220785 gi  33200785 gi  33220785 gi  33200785 gi  33200785 gi  33200785 gi  33200785 gi  33200785 gi  3320785 gi  33200785 gi  33200785 gi  3320785 gi  33200785 gi  33200785 gi  33200785 gi  3320785 gi  33200785 gi  33200785 gi  33200785 gi  3320785 gi  33200785 gi  3320785 gi  33200785 gi  3320785 gi  3320785 gi  3320785 gi  33200	B     T     T     T      ETWIVCAFSVDHDSAPLNRCVR.     GSYVIYAPVDIVAMNVVLSGG.     GSYVIYAPVDIVAMNVVLSGG.     GSYVIYAPVDIVAMNVVLGG.     GAVVYAPVDIPAMNVVLGG.     GAVVYAPVDIPAMHVVMGG.     GAVVYAPVDIPAMHVVMGG.     GAVVYAPVDIPAMHVVMGG.     GAVVYAPVDIPAMHVVMGG.     GAVVYAPVDINTILVKGG.     GGAVYAPVDIPAMHVVMGG.     GGAVYAPVDIPAMLVAAGA.     GGAVYAPVDIPAMLVAAGA.     GGAVYAPVDIMTICUVLAGA.     GGAVYAPVDIMTICUVLAGA.     GGAVYAPVDIPAMLVAAGA.     GGAVYAPVDIPAMLVAAGA.     GGAVYAPVDIMTICUVAGA.     GGAVYAPMDIMTACA.     GGAVYAPVDIMTICOVACA.     GGAVYAPVDIMTICOVACA.     GGAVKAPPOLNACA.     GGAVYAPVDIMTICOVACA.     GGAVYAPVDIMTICOVACA.     GGAVYAPVDIMTICOVACA.     GGAVYAPVDIMTICOVACA.     GGAVXAPMDIMTACA.     GGAVKANPOLNACA.     GGAVKANPOLNACA.     GAVKKLNSCLRAVCSARKSDS.     STPIEEHPLLSSFVNGDAYNNSI     KELVPGS.LPGRFVNGDAFNNSI     KELVPGS.LPGRFVMGDAFNNSI     KELVPGS.LPGRFVNGGAPFNSI     KELVPGS.LPGRFV	β11 AKINVAMICQTLVS DPDYVALLPSGFAI DPDYVALLPSGFAI DPDYVALLPSGFAI DPDYVALLPSGFAI DSSYVALLPSGFAI DSSYVALLPSGFSI NSDSVKFLPSGFSI DSSYVALLPSGFSI DSSYVALLPSGFSI DSSTUPILPSGFSI DPSTIPILPSGFMI DPSTIPILPSGFMI DPSTIPILPSGFMI DFSTIPILPSGFMI DFSTIPILPSGFWI DFSTIPILPSGFWI DFSTIPILPSGFWI DFSTIPILPSGSTI ILSAVIKKEVPSE ILSAVSIMKKIPSE ILSAVSIMKKIPSE ILSAVSIMKKIPSE ILSAVSIMKITSE ISSAQMASITTK ISSRAQMASIT ISSRAQMASIT ISSRAQMASIT ISSRAQMASIT ISSRAQMASIT ISSRAQMASIT ISSRAQMASIT ISSRAQMASIT ISSRAQMASIT ISSRAQMASIT ISSRAQMASIT ISSRAQMASIT ISSRAQMASIT ISSRAQMASIT ISSRAQMASIT ISSRAQMASIT ISSRAQMASIT ISSRAQMASIT ISSRAQMASIT ISSRAQMASI ISSRAQMASIT ISSRAQMASI I	PECNQEISRD. PPECNQEISRD. PPOSARGGGSA UPDCSARGGGSA UPDCGIDGGGS. UPDCGIDGGGS. UPDCSYSSDQ. VPDGVNGSYHRG. IPDGVSRPLVI PPOGTNGSNSEQGC PPOGTGRPGKEC SRDSRPSCEA. SBDCRSTVE. SBDCRSTVE. UPVNPSCESSEV. VSAEEFMNERGSI ITODHLISLSNGP UMDSAES.SRE VKQSVGKKACLUC VRQSVGSTPCLLG	ANASAGAGVEG	η2 NOHQEMVS GGEGNNLEV AAEERNSIG GDGDQRPVG DTDGGLVN ST ST		β12 ILCKITYVA SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITISF0M TLITVAF0I TLITVAF0I TLITVAF0I TLITVAF0I TLITVAF0I KLIAKPSIK KKIAKPSIK KKIKKKL DQFGVDWKW LGVDIGSST IDVDIGSST	WWNPGGW WUNSVP. UVDSVP. UVDSVP. UVDSVP. UVDSVP. UVDSVP. UVNLP. UVNLP. UVSLP. UVSCPS. UVSCPS. UVSCPS. UVSCPS. USCSKAN. USCSKAN. USCSKAN. USCSKAN. USCSKAN. USCSKAN. USCSKAN. USCSKAN. USCSKAN. USCSKAN. USCSKAN. MASSVC.	ACCOUNT A PASVLR .TAKLSL .TAKLSL .TAKLSL .TAKLTV .TAKLTV .TAKLTV .TAKLTV .TAKLTV .TAKLNV .TAKLNV .TAKLNV .TAKLNV .TAKLNV .TAKLNV .TAKLNV .TAKLNV .TAKLNV .TAKLVV	a4 accord and a second and a second and a second accord a	LACDGA. KTAGKPI LACDGA. MSCETA. LHCDST. LACDST. LACDST. LACDA. LACPSA. LACPSA. LACPSA. LACPSA. LSOFTAS. FPCQTA. LNCPSTA. LNCPSC. LN	LF SA TTA VIASSADV. GAQLSQT: TTRSESQT TTRSESQT THLDASK. SHIELSS. SHUELSS.	SN SNGEVS. JOITTI SQNAGI ISS AISIIP SVKER AIVPNLDS
2292_A PDBID 2292_A PDBID gi  322657038 gi  322657038 gi  322645640 gi  322645640 gi  322645640 gi  3322645640 gi  332208796 gi  332198190 gi  332199120 gi  332197328 gi  33219320876 gi  3321932080 gi  3322643397 gi  3322643397 gi  3322643192 gi  3322641765 gi  3322641765 gi  3322641765 gi  332641765 gi  332647765 gi  332647765 gi  332657301 gi  332657305 gi  322007858 gi  322702528	P      T	β11 AKINVAMICQTLVS DPDYVALLPSGFAI DPDYVALLPSGFAI DPDYVALLPSGFAI DPDYVALLPSGFAI DSSYVALLPSGFAI DSSYVALLPSGFAI DSSYVALLPSGFAI DSSYVALLPSGFAI DSSYVALLPSGFSI DPSNIDILPSGFAI DPSTIPLPSGFAI DPSTIPLPSGFAI DPSTIPLPSGFAI DFSTIPLPSGFAI DFSTIPLPSGFAI DFSTIPLPSGFAI DFSTIPLPSGFAI SSNIPLPVSS DSSNIPLPVSS DSSNIPLPVSS DSSNIPLPSS LSPSAIMAINTK KLSSRAFMAQINTK LSSRAQMASITTK LSSRAQMASITTK LSSRAQMASITTK SSNIPLPSSS DSSNIPLSSS DSSNIPLSSS DSSNIPLSSS DSSNIPLSSS DSSNIPLSS D	PPECNQEISRD. LPDGSNGGGDG. LPDGSNGGGGS LPDGNANSGAP. LPDGSNGGGS LPDGGIDGGSS. LPDGGIDGGSS. LPDGGIDGGSS. SPOGUSSYHGG SPDGTGRPGKEGS SRDGRPSRECA. SSDGRSTVE. SSDGRSTVE. SSDGRSTVE. LEIGIDNVIGSME LSGIDIIINSSC VSAEFYNERGSI LSSGIDIIINSSC VSAEFYNERGSS. SSDGRSSSEV. VSSVGKKACLUC VKSSVGKKACLUC VKSSVGKKACLUC VKSSVGKKACLUC	ANASAGAGVEG	η2 NOHQEMVS GGEGNNLEV AAEERNSIG DDCDQPVG PDTDGGLVN ST ST		β12 ILCKTYVA SL1TVAF0I SL1TVAF0I SL1TVAF0I SL1TVAF0I SL1TVAF0I SL1TVAF0I SL1TVAF0I SL1TVAF0I SL1TVAF0I SL1TVGF0I SL1TVGF0I TL1TVAF0I TL1TVAF0I TL1TVAF0I TL1TVAF0I TL1TVAF0I CL1TVAF0I CL1TVAF0I DL1TVAF0I SL1TVAF0I TL1TVAF0I TL1TVAF0I TL1TVAF0I S	WWPGGW UVDSVP. UVDSVP. UVDSVP. UVDSVP. UVDSVP. UVDSVP. UVDSVP. UVSSVP.	ACCOUNT A PASVLR .TAKLSL .TAKLSL .TAKLSL .TAKLSL .TAKLSL .TAKLSL .TAKLSL .TAKLNY .TAKLNY .TAKLNY .TAKLNY .TAKLNY .TAKLNY .TAKLNY .TAKLNY .TAKLNY .TAKLNY .TAKLNY .TAKLNY .TAKLNY .TAKLNY .TAKLY .TAKY .TAKLY .TAKY	44 2000 200 200 200 200 200 AVÄKREYFKË LKRFTSYVOE GSVATVNS LIKCTVERIKAA GSVATVNS LIKCTVERIKAA GSVATVNS LIKCTVERIKAA ESVETVNNLISCTVOKIKAA ESVETVNNLISCTVOKIKAA ESVETVNNLISCTVOKIKAA ESVETVNNLISCTVOKIKAA ESVETVNNLISCTVORIKAA GSVATINTIVSIV SVJATINTIVSAVQRIKAA SVATIENLISTVORIKAA KSVDTVSALISSTIQRIKAA KSVDTVSALISTTORIKAA KSVDTVSALISTTORIKAA GALCLSROFSALGAKVALAY G.LLTKAVIEGVARFAMG G.LLTKAVIEGVARFAMG G.LLTKAVIEGVARFAMG G.LLTKAVIEGVARFAMG G.LLTKAVIEGVARFAMG G.LLTKAVIEGVARFAMG G.LLTKAVIEGVARFAMG G.LLTKAVIEGVARFAMG G.LLTKAVIEGVARFAMG G.LLTKAVIEGVARFAMG G.LLTKAVIEGVARFAMG G.LLTKAVIEGVARFAMG G.LLTKAVIEGVARFAMG G.LLTKAVIEGVARFAMG G.LLTKAVIEGVARFAMG G.LLTKAVIEGVARFAMG G.LLTKAVIEGVARFAMG MUMIEMAFLIQANTEELEPE SLUVEMAFLUQANTEELEPE	LACDA KTAGKPI UNSCOVGG LACDGA MSCCTA LACDST LLACDST LLACDST LLACDST LLACDST LLACDST LLACDST LLCST LNCPSTA LNCPSTA LNCPSTA LSSNTIF FPCOTA LLSNTIF LSNTIF LSNTIF LSNTIF LSNTIF LSNTIF LSNTIF LSNTIF LSNTIF LSNTIF LLST LSSNTIF LSN	LF SA TTA VIASSADV. GTSINQT GGAQLSQT TRSESQT TRSESQT TRLDSSK. SHILDSS SHILDSS SHIELDSS	SN NQEVS TOTTTI SQNAGI TSS AISIIP SVKER AIVPNLDS VVSFEV
2292_A PDBID 2292_A PDBID gi  322657038 gi  322657038 gi  322655039 gi  222423025 gi  332645640 gi  3322659074 gi  332198190 gi  332193120 gi  332193120 gi  332193120 gi  33219321 gi  33219321 gi  33219321 gi  332219328 gi  3322643392 gi  332643192 gi  332658730 gi  332658730 gi  332658730 gi  3322539118 gi  3326353111 gi  3326353111 gi  3326353111 gi  3326353111 gi  3326353111 gi  3326353111 gi  3326353111 gi  332658730 gi  3326353111 gi  3326353111 gi  3326353111 gi  3326353111 gi  3326353111 gi  3326353111 gi  332658730 gi  3326353111 gi  33263531111 gi  3326353111 gi  3326353111 g	P     FT      FT	β11 AKINVAMICOTLVS DPDYVALLPSGFAI DPDYVALLPSGFAI DPDYVALLPSGFAI DPDYVALLPSGFAI DSSYVALLPSGFAI DSSYVALLPSGFAI DSSYVALLPSGFSI DSSYVALLPSGFSI DSSYVALLPSGFAI DSSYVALLPSGFAI DSSYVALLPSGFAI DPSYIPILPSGFAI DPSYIPILPSGFAI DPSYIPILPSGFAI DPSYIPILPSGFAI DPSYIPILPSGFAI DPSYIPILPSGFAI DPSYIPILPSGFAI DPSYIPILPSGFAI DPSYIPILPSGFAI DPSYIPILPSGFAI DPSYIPILPSGFAI DPSYIPILPSGFAI DPSYIPILPSGFAI DPSYIPILPSGFAI DPSYIPILPSGFAI DPSYIPILPSGFAI DPSYIPILPSGFAI DPSYIPILPSGFAI DPSYIPILPSGFAI RESUPILPSGFAI SENTON	PEGNQEISRD. PPEGNQEISRD. LPDCSARGGGBA LPDCSARGGGBA LPDCSARGGGBA LPDCSARGGGBA LPDCSARSSAD. LPDCSSSSD VPDCVNGSYHRG SPDCNGSNSEQGC FPDCTGKPGGKBC SRDSHPSSEV. SRDSHPSSEV. SRDSHPSSEV. SDCRTSTVE. VPVNPPCISS SDCRTSTVE. VPVNPPCISS SDCRTSTVE. VPVNPPCISS SDCRTSTVE. VPVNPPCISS SDCRTSTVE. VPVNPPCISS SDCRTSTVE. VPVNPPCISS SDCRTSTVE. VPVNPPCISS SDCRTSTVE. VPVNPSDCVCSS SDCRTSTVE. SDCRTSTVE. VPVNPSDCVCSS SDCRTSTVE. SDC	NASAGAGVEG	η2 NOHOEMVS GGEGNNLEV AAEERNSIG DDCDQRPVG PDTDGGLVN ST ST		β12 ILCK[TV[VA] SL1TVAF0] SL1TVAF0] SL1TVAF0] SL1TVAF0] SL1TVAF0] SL1TVAF0] SL1TVAF0] SL1TVAF0] TL1TAF0] TL1TAF0] TL1TAF0] TL1TAF0] TL1TVAF0] TL1TVAF0] TL1TVAF0] TL1TVAF0] TL1TVAF0] TL1TVAF0] DL1VAF0] DL1VAF0] IL1	WWPGGW UVDSVP. UVDSVP. UVDSVP. UVDSVP. UVDSVP. UVNLP. UVNLP. UVNLP. UVNSLP. UVSCP. UVS	ACCOUNT A PASVLR .TAKLSL .TAKLSL .TAKLSL .TAKLSL .TAKLTV .TAKLTV .TAKLTV .TAKLTV .TAKLNY .TAKLNY .TAKLNY .TAKLNY .TAKLNY .TAKLNY .TAKLNY .TAKLNY .TAKLNY .TAKLNY .TAKLNY .TAKLNY .TAKLNY .TAKLNY .TAKLNY .TAKLY .TAK	CAL CONCOURSE OF A VARANCE STATUS A VARANCE YE KE LKRFT SYVGE G SVATVNS LIKCTVER IKAA G SVATVNS LIKCTVER IKAA G SVATVNS LIKCTVER IKAA G SVATVNS LIKCTVER IKAA S SVATVNS LISCTVQK IKAA E SVETVNNL ISCTVQK IKAA G TVKSVET IMAHTIVK IKAA G TVKSVET IMAHTIVK IKAA G TVKSVET IMAHTIVK IKAA G TVKSVET IMAHTIVK IKAA S SVATUNI LISCTVQK IKAA S SVATUNI LISCTVQK IKAA VSATTVNT LISSTVQR IKAA G LISC STVSVIN ISTVER IKAI G ALCLSRCH SALGAKVALAY G ALCLSRCH SALGAKVALAY S IIGKALLAGAGQR LARR. G I LIKAVI E QVARTE LELPE S LIVEMAF LIQANTE ELPE S LIVEMAF LIVQANTA ELPE S S LIVEMAF LVQANTA ELPE S S LIVEMAF LVQANTA ELPE S S S LIVKMKVUDNE SN.	LLACDA KTAGKPI KTAGKPI LACDGA. MSCETA. LLACDST. LLACDST. LLACDST. LLACDST. LLACPEC. LLACPEC. LLSCPIC. LLSCPIC. LLSCPIC. LLSCPIC. KKROLEG. KKROLEG. LLLGTCR. KKROLEG. LLLGTCR. KLLGAVR: YLLCAFT. HCLAFT.	LF DV. IASSADV. GTSLNQT TRSSQT TRSSQT TRSSQT SHIELSS. SHIELSS. SHIELSS. SHIELSS. SHIELSS. SHIELSS.	SN. NQEVS. OTTII. SNAGI. TSS. TSS. TSS. TSS. TSS. TSS. TVPNLDS TVPNLDS TVPNLDS TVPNLDS TVPNLDS TVPNLDS TVPNLDS TVPNLDS
2292_A PDBID 2292_A PDBID gi  322657038 gi  322657038 gi  322659039 gi  222423025 gi  332658526 gi  332208796 gi  332208796 gi  332198190 gi  33219310 gi  33219310 gi  33219320 gi  332640397 gi  332640397 gi  332643195 gi  332208796 gi  3322641765 gi  3322081765 gi  332208176 gi  332208177 gi  332643192 gi  332658730 gi  332658730 gi  33205918 gi  3322658730 gi  3322658730 gi  3322658730 gi  3322658730 gi  3322658730 gi  3322658730 gi  3322658730 gi  3322658730 gi  33225918 gi  3325	β9         β10           ETWIVCAFSVDHDSAPLNNCVER         GSYVIYAPVDIXANVVLSGG.           GSYVIYAPVDIXANVVLSGG.         GSYVIYAPVDIXANVVLSGG.           GSYVIYAPVDIXANVVLSGG.         GSYVIYAPVDIXANVVLSGG.           GSYVIYAPVDIXANVVLSGG.         GSYVIYAPVDIXANVVLSGG.           GGLVYAPVDIXANVVLSGG.         GGLVYAPVDIXANVS           GGLVYAPVDIXANVVLSGG.         GGLVYAPVDIXANVS           GGLVYAPVDIXANVDIXANSGC.         GGNVYAPVDIXANIANSGC.           GGNVYAPVDIXANVDIXANSGC.         GGNVYAPVDIXANASCA           GGNVYAPNDIXANASCA         GGNVYAPNDIXANSSCA           GGNVYAPNDIXANASCA         GGNVYAPNDIXANSSCA           GGNVYAPNDIXANASCA         GSUSSTVDPAVOUSANSSCA           GSUSUSSTVDPAVOUSANSS.         SVUCKMENALCKYLATSHRPGPG           GAVKKINSCLRAYCSARKSDSS.         STPIEEHPLLSFVNGDDAYRNSI           KELVPGS.LLORFV GDDEFFNSI         KELVPGS.LLORFV GDDEFFNSI           DRPVNKISLGKAVGSS.SLORFV GSDSYRDA         LCCSLKTN.ASPVFTPANOGLI           LLCCSLKTN.ASPVFTPANOGLI         LLCCSLKTN.ASPVFTPANOGLI	β11 A K I NVAMIC C T LVS D D DYVALLP SGFAI D D DYVALLP SGFAI D DYVALLP SGFAI D DYVALLP SGFAI D SYVALLP SGFAI D SYVALLP SGFAI D SYVALLP SGFAI D SYVALLP SGFAI D SYVALLP SGFAI D SYVALLP SGFAI D P SY IP ILP SGFAI D P SY IP ILP SGFAI D P ST IP ILP SGFAI D SYVALLP SGFAI D SYNAL SGFAI SGF	PEGNQEISRD. PPEGNQEISRD. LPDCSARGGGSA LPDCSARGGGSA LPDCGIDGGGS. LPDCGIDGGGS. VPDCVNGSYHRG. SPDCNCSYHRG. SPDCNCSYHRGSE SPDCNCSYHRGSE SPDCNCSYCS RDSHPSSEV. SCORRSIVE. SCORRSIVE. LESCTDDITNSSC UPVNPSCISVN VFVNPSCISVN VFVNPSCISVN VFVNPSCISVN USAEEFNNERGSI ITDDHLISLSNC VKSSVGKACLIC VKSVGKACLIC VKSVGKACLIC UNSVGSTPCLLC VKSVGKACLIC UNSVGSTPCLLC VKSVGKACLIC LDKTLDDS.GRA	NASAGAGVEG A A A A TSSGEGV. TTTISG. AEVIG. SPVTE. DEDRSR. SQLEIN. SQLEIN. SQLEIN. SQLEVC. KAVDCN. KAVDCN. KAVDCN. KAVDCN. KAVC.	η2 NOHQEMVS GGEONNLEV AAEERNSIG GDGDQRPVG FDTDGGLVN ST ST		β12 ILCKITYVA SLITVAFOI SLITVAFOI SLITVAFOI SLITVAFOI SLITVAFOI SLITVAFOI SLITVAFOI SLITVAFOI SLITAGFOI SLITAGFOI SLITAGFOI SLITAGFOI TLITVAFOI SIGNO	WNPGGW. UVDSVP. UVDSVP. UVDSVP. UVNSLP. UVNSLP. UVNSLP. UVNSLP. UVSSVP	ACCOUNT APASVLR TAKLSL TAKLSL TAKLSL TAKLSL TAKLYV TAKLYV TAKLYV TAKLYV TAKLYV TAKLYV TAKLNM PAKLNM PAKLNM PAKLNM PAKLNM PAKLNM TAKLNV TAKLYV TAKL	a4 accorded and a second and a second and a second accorded and a second accorded and a second accorded and a second accorded ac	LLACIGA. INSCETA. LLACIGA. INSCETA. LLACIGS. LLACIGS. LLICESTA. LLICESTA. LLICESTA. LLICESTA. LLICESTA. LLICESTA. LLICESTA. LLICESTA. SCRM. S	LF SA UV.IASSADV. GGSISLQT. TRSESQT TRSESQT TRSESQT SHIELSS. SHVELSS. JSWSFV. JWNSFV. VNNSFV.	SNSN. NOEVS TOTTII. SONAGI TSS AISIIP. SVKER. AIVPNLES AIVPNLES AIVPNLES AIVPNLES AVSFEV
2292_A PDBTD 2292_A PDBTD gi  322657038 gi  322657039 gi  222423025 gi  332655029 gi  332656526 gi  3322635674 di  332198190 gi  332193100 gi  33219310 gi  33219320 gi  33219320 gi  33226337 gi  33205030 gi  332643192 gi  3322643192 gi  332263730 gi  332658730 gi  33203984 gi  332039184 gi  33203918 gi  33200	P9         D10           ETWIVCAFSVDHDSAPLNNRCVR. GSYVIYAPVDIVAMNVVLSGG.         GAPLNNRCVR. GSYVIYAPVDIVAMNVVLSGG.           GSYVIYAPVDIVAMNVVLSGG.         GAVVYAPVDIPAMNVVLSGG.           GALVYAPVDIPAMNVVLSGG.         GALVYAPVDIPAMNVVLSGG.           GALVYAPVDIPAMNVVLSGG.         GALVYAPVDIPAMNVVLSGG.           GALVYAPVDIPAMNIVLSG.         GALVYAPVDIPAMNIVKSG.           GALVYAPVDIPALNIANSSC.         GALVYAPVDIPALNIANSSC.           SALVIYPVDIALALIANSSC.         GAKVYAPVDINTUNGG.           GGMVYAPVDIALALANIANSSC.         GAVYAPVDINTUNGCALALASISGUV           GGMVYAPMDLMIACAANISGC.         GAVKINSULANDSI           SALVIYPVDVQUAALANIANSSC.         GGWNYAPMDLMIACAANISGC.           SALVIYSTVDVQUQUAANSS.         GSULYSTVDPVQUAANSSC.           SVVCKMENALCKXILATSHRTQGP         GAVKKINSGLRAYCSARKSGSS.           GAVKKINSGLRAYCSARKSGSS.         STPIEEHPLLSSTWODDAYNSS           KELVPGS.LLORFVQGDDEFNSS         KELVPGS.LLORFVQGDDEFNSS           KELVPGS.LLORFVQGDDEFNSS         KELVPGS.LLORFVQGDDEFNSS           KELVPGS.LLORFVQGDDEFNSS         KELVPGS.LLORFVPGDDEFNSS           KELVPGS.LLORFVPGDDEFNSS         KELVPGS.LLORFVPGDDEFNSS           KELVPGS.LLORFVPGDDEFNSS         KELVPGS.LLORFVPGDDEFNSS           KELVPGS.LLORFVPGDDEFNSS         KELVPGS.LORFVPGDDEFNSS           KELVPGS.LLORFVPGDDEFNSS         KELVPGS.L	β11 A K I NVAMIC C T LVS D D YVALLP SCFAI D D YVALLP SCFAI D D YVALLP SCFAI D D YVALLP SCFAI D S YVALLP SCFAI S S YVALLP SCFAI S S S S S S S S S S S S S S S S S S S	PEGNQEISRD. PPEGNQEISRD. LPDCSARGGGSA LPDCGANGGGGSA LPDCGIDGGGS. LPDCGIDGGGS. VPDCVNGSYHRG SPDGNGSNSQQC PDGTGRSSQQC SPDGNGSNSQQC SRDSHPSSSEV. SDGRPGEGA. SSDGRSTVE. LESGTDDITNSSG VPVNPSGISVNG VPVNPSGISVNG VSAEEFNERGSI ITDDHLISLSNG VKSSVGKACLIC VKSVGKACLIC VKSVGKACLIC VKSVGKACLIC VKSVGKACLIC VKSVGKACLIC VKSVGKACLIC VKSVGKACLIC VKSVGKACLIC	NASACAGVEG	η2 NOHQEMVS GGEGNNLEV AAEERNSIG GDGDQRPVG FDTDGGLVN ST ST L L		β12 ILCKITYVA SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITSF0M TLITVAF0I SCIT	WWPGGW UVDSVP. UVDSVP. UVDSVP. UVDSVP. UVDSVP. UVNSLP. UVNSLP. UVSSLP. UVGINP. UVSSLP. UVSSV UVS	ACCOUNT A PASVLR .TAKLSL .TAKLSL .TAKLSL .TAKLYV	a4 avia krevy kF Lkret syvee g svatvns Likctve rikaa g svatvns Lisctvok ikaa g svetvnn Lisctvok rikaa g svetvnn Lisctvok rikaa g svetvnn Likattik rikaa g svetvn Likattik rikaa g svetvn Likattik rikaa g svetve rikaa	LLQ KTAGKPI KTAGKPI KVSCDVGG LACDGA. MSCETA. LLRCBT. LLQCBTA. LLQCBTA. LLQCBTA. LLQCBTA. LLQCBTA. LLQCBTA. LLQCBTA. LLQCBTA. LLQCBTA. KRM. KRGLEG. KRM. KRGLEG. KRM. KRGLEG. KRM. KRGLEG. KRGLEG. KRGLEG. KRGLEG. KRGLEG. KRGLEG. KRGLEG. KLGTVR. HCLAFM. HCLGFV. HCLGFV. HCLGFV.	LF SA DV: TTA VIASSADV. GGTSLNQT TTRSESQT TTRSESQT TTRSESQT THLDDSK. NHLDDSK SHVELSS. SHVELSS. SHVELSS. SHVELSS. SHVELSS. VSWFV- VNWSFV VNWSFV TNWSFV TNWSFV TNWSFV	SN NQEVS. TQTTII. SQNAGI. TSS. AISIIP. SVKER. AIVPNLES IVPNLDS VVSFEV.
2292_A PDBID 2292_A PDBID gi  332657038 gi  332657039 gi  222423025 gi  332658059 gi  332658059 gi  3322658526 gi  332198190 gi  332198190 gi  33219310 gi  33219320 gi  33219320 gi  33219320 gi  3322005030 gi  3322005030 gi  3322058537 gi  332058517 gi  332007885 gi  222424068 gi  332007885 gi  222424068 gi  332007885 gi  222424068 gi  332007885730 gi  332007895 gi  222424057 gi  33200784 gi  33200784 gi  33200784	P     TT     ETWIVCAFSVDHDSAPINRCVR.     GSYVIYAPVDIVAMNVVLSGG.     GSYVIYAPVDIVAMNVVLSGG.     GSYVIYAPVDIVAMNVVLSGG.     GSYVIYAPVDIVAMNVVLSGG.     GSYVIYAPVDIVAMNTVLGG.     GAVVYAPVDIPAMHVVMGG.     GAVVYAPVDIPAMHVVMGG.     GAVVYAPVDIPAMHVVMGG.     GAVVYAPVDIPAMHVVMGG.     GAVVYAPVDIPAMHVVMGG.     GAVVYAPVDIPAMHVVMGG.     GAVVYAPVDIPAMHVVMGG.     GAVVYAPVDIPAMHVVMGG.     GGMVYAPVDIPAMHVVMGG.     GGMVYAPVDISTOULAGA.     SALUIYTPVDLALNIAMSGE.     GAVVYAPVDISTOULAGA.     GGMVYAPMDIMICACASISCOV     GGMVYAPMDIMICALASISCOV     GGMVYAPMDIMICACASISCOV     GGMVXASIS     F F F GINACAL     GAVKKLINSGLANCS      F F F MNCACI      GAVKKLINSGLANCS      F F F MNCACI            LCCSLKTN. ASPVF F F ANQACI	β11 A K INVAMIC C T LVS D D DYVALLP SGFAI D DYVALLP SGFAI D DYVALLP SGFAI D DYVALLP SGFAI D DYVALLP SGFAI D SYVALLP SGFAI D SYVALLP SGFSI D SSVVALLP SGFSI N SDSVKFLP SGFSI N SDSVKFLP SGFSI D SNYALLP SGFSI D SNYALP SGFSI N SGFSI SGFS	22. PPEGNQEISRD. LPDGSNGGGDG. LPDGSNGGGGS LPDGCNANSGAP. LPDGSNGGGSS LPDGUNGSSDQ. UPDGSSSSDQ. VPDGVUSSPLVII SPDG.SSSE SPDGTGRPGKEG SRDGRPBEGEA. SSDGRSTVE. VPVNPSGEEA. SSDGRSTVE. LESGTDDITNSS VPVNPSGISVNE USADEFMERGSI ITODHLISLSNG VQSVGKKACLVC VGSVGKKACLVC VGSVGKKACLVC VGSVGKKACLVC VGSVGKKACLVC VGSVGGTPCLLC VKAVGTKACLLDE UKGSVGKKACLVC VKSVGSTPCLLC VKAVGTKACLVC VKSVGSTPCLLC VKAVGTKACLVC VKSVGSTPCLLC VKAVGTKACLVC VKSVGSTPCLLC VKSVGSTPCLLC VKSVGSS, GRRF LDKIFDDN.GRKT LPKIFDDN.GRKT IPCESQE	NASAGAGVEG	η2 NOHQEMVS GGEGNNLEV AAEERNSIG GDGDQRPVG DTDGGLVN ST ST		β12 ILCKITYVA SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVGF0I SLITVGF0I SLITISF0M TLITVAF0I SCITTA	WWNPGGW. WVNSUP. UVDSVP. UVDSVP. UVDSVP. UVDSVP. UVDSVP. UVDSVP. UVDSVP. UVDSVP. UVDSVP. UVDSVP. UVDSVP. UVSSVP. UV	ACCOUNT A PASVLR . TAKLSL . TAKLSL . TAKLSL . TAKLYV . TAKLYV . TAKLYV . TAKLYV . TAKLYV . TAKLNY . TAKLNY . TAKLNY . TAKLNY . TAKLNY . TAKLNY . TAKLYN . TAKLY	a4 avg cooccoccoccoccoccoccoccoccoccoccoccocco	LQC KTAGKPI INSCETA LACDGA. LACDGA. LACDGA. LACDA. LACDA. LACDA. LACDA. LACDA. LACDA. LACDA. LACDA. LACDA. LACDA. LACDA. LACDA. LACDA. LACDA. LACDA. KRC. LACDA. KRM. KRQLEG KRC. LLCTCR. LLCTCR. LLCTCR. KRC. KRC. KRC. LLCTCR. KRC	LF SA TTA VIASSADV. GGTSLNQT TTRSESQT TTRSESQT TTRSESQT TTRSESQT SHVELSS. SHVELSS. SHVELSS. SHVELSS. SHVELSS. TTRSESQT	SN. NQEVS. TQTTII. SQNAGI. ISS. NISIIP. VKER. IVPNLES VKER. IVPNLES VSFEV.
2292_A PDBID 2292_A PDBID gi  322657038 gi  322659039 gi  222423025 gi  332655069 gi  332655626 gi  332208796 gi  332198190 gi  332193100 gi  33219310 gi  33219310 gi  33219310 gi  33205030 gi  332643192 gi  3320643192 gi  3322643192 gi  3322643192 gi  3322643192 gi  3322643192 gi  3322643192 gi  3322643192 gi  3322643192 gi  3322643192 gi  33225013 gi  332643192 gi  332253113 gi  33225318 gi  33223318 gi  33223318 gi  33223318 gi  33223318 gi  33223318 gi  33223318 gi  33223318 gi  33223318 gi  33223318 gi  33203985 gi  227202528 gi  332039184 gi  332039184	TT ETWIVCAFSVDHDGAPLNNRCVR. GSYVIYAPVDIVAMNVVLSGG.I. ASFVIYAPVDIVAMNVVLSGG.I. ASFVIYAPVDIVAMNVVLSGG.I. GALVVYAPVDIPAMNVVLSGG.I. GALVVYAPVDIPAMNTVLNGG.I. GALVVYAPVDIPAMNTWNGG.I. GALVVYAPVDIPSMNTVMSG.I. GAWVYAPVDTPSMNTVMSG.I. GAWVYAPVDTSLUVKGG.I. SALVIYAPVDTSLUVKGG.I. GAWVYAPVDTNITULVLAGH.I SALVIYAPVDTSLUVKGG.I. GGWAYAPVDTNITULVLAGH.I SALVIYAPVDTSLUVGG GGMUYAPMNLNTAYSAISGCV. GGWAYAPMDLNTACAAISGDII GGWIYAPMDNATMHFAVSGEV. GGWAYAPMDLNTACAAISGDII GGWIYAPMDATMHFAVSGEV. GSLUVYSTVDVVGULGOAMNGE.I. SVVCKMENALCKYIATSHRPOCP SVVCKMENALCKYIATSHRPOCP GAVKKINSCLRAVGAKKSOS. STPIEEHPLLVSFVNGDAYNSI KTPLEEHPLLSSFVNGDAYNSI KELVPGS.LLQRFVDGDEFNNSI KELVPGS.LLQRFVDGDEFNNSI KELVPGS.LLQRFVDGDEFNNSI KELVPGS.LLQRFVDGDEFNNSI KELVPGS.LPQRFVDGDEFNNSI KELVPGS.LPQRFVDGDEFNNSI KELVPGS.LPQRFVDGDEFNNSI	β11 AKINVAMICCTIVS DPDYVALLPSGFAI DPDYVALLPSGFAI DPDYVALLPSGFAI DPDYVALLPSGFAI DSSVVALPSGFAI DSSVVALPSGFAI DSSVVALPSGFAI DSSVVALPSGFAI DSSVVALPSGFAI DSSVVFLPSGFSI NDSVKFLPSGFSI DSSVVFLPSGFSI DSSVFFLPSGFSI DSSVFFLPSGFSI DSSVFFLPSGFSI DSSVFFLPSGFSI DSSVFFLPSGFSI DSSVFFLPSGFSI DSSVFFLPSGS ILSAVILNKEVPSE ILSAVILNKEVPSE ILSAVSLMAHINTK KLSRSAMASITTK KLSRSAMASITTK KLSRSAMASITTK KLSRSAMASITTK LSSRSAMASITTK RFKLIPHISKGSWI	PPEGNQEISRD. PPEGNQEISRD. PPOSARGGG5A LPDCSARGGG5A LPDCGIDGGGS. LPDCGIDGGGS. VPDGVNGSYHRG SPDGUVSRPLVII SPDG.SSEQ PPOGTGRPGKEC SRDSRPSSEV. SRDGRPSEGEA. SDCRRSTVE. VPVNPSCISVNG SRDGRPSEGEA. SDCRRSTVE. LESGTDDITNSGAEL LMMDLVETSGAEL LMMDLVETSGAEL LMMDLVETSGAEL LMMDLVESS.SRL VQSVGKKACLUC VRASVGKACLUC VRASVGKA	NASAGAGVEG	η2 NOHQEMVS GGEGNNLEV AAEERNSIG GDGDQRPVG DTDGGLVN ST ST		β12 ILCKITYVA SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVAF0I SLITVGF0I SLITISF0M TLITVAF0I CIITVAF0I CIITVAF0I CIITVAF0I CIITVAF0I CIITVAF0I CIITVAF0I CIITVAF0I CIITVAF0I CIITVAF0I CIITVAF0I TLITVAF0I TLITVAF0I CIITV	WWNPGGW. WVNPGW. WVDSVP. UVDSVP. UVDSVP. UVDSVP. UVDSVP. UVDSVP. UVDSVP. UVDSVP. UVDSVP. UVSCP. UVSCPV. UVSCPV. UVSCPV. UVSCPV. UVSCPV. USSVV. USSVV. USSVV. USSVV. USSVV. USSVV. USSVV. USSVV. USSVV. USSVV. MARGVN. MARGVV. MARGVV. MARGVV. MARGVV. MARGVV. MARGVV. MARGVV. MARGVV. MARGVV. MARGVV. MARGVC. MARG	ACCOUNT A PASVLR . TAKLSL . TAKLSL . TAKLYL . TAKLYV . TAKLY	a4 ava karves k k k k k k k k k k k k k k k k k k k	LACDGA. IXTAGKPII IXVSCDVGG LACDGA. IMSCETA. LLACDA. LLACDA. LLACDA. LLOCSI. LLOCSI. LLOCSI. LLOCSI. LNCPSIA. LLNCPSC. LLNCPSC. LLNCPSC. LLNCPSC. LLNCPSC. LLNCPSC. LLNCPSC. LLNCPSC. LLNCPSC. LLNCPSC. LLNCPSC. LLNCPSC. LLNCPSC. LLNCPSC. LLNCPSC. ILNCPSC. LLN	LF SA DV. TTA VIASSADV. GGTSLNQT TTRSESQT TTRSE	SN. NQEVS. TQTTII. SQNAGI. SS. NISIIP. VKER. AIVPNLES AIVPNLES AVSFEV.

Fig. 4: Alignment of the 35 Arabidopsis START domain proteins with their mammalian counterparts, depicting the placement of secondary structural elements and the high conservation in sequence. This conservation in sequence enabled reliable structural models to be built using the mammalian structures as templates for plant sequences

believed that identification of the ligand can help in assigning distinct physiological and pathological roles to different START-domain-containing proteins. For this, it would be essential to study the inner tunnels of the respective domains, which aids, or severely restricts the correct substrate by virtue of its shape, size and chemical properties. Accordingly, structural modelling of the ligand binding tunnel/pocket was carried out for a few of the Arabidopsis START domains. This involved mutational analysis of the domains with vertebrate START proteins serving as a template to probe the function of the plant START domain as a regulatory molecule. Detailed structural analysis of full models as well as chimeras generated from in silico mutations of residues known to be lining the active site cavities (as per crystal structure data) revealed that in contrast to the large cholesterol binding tunnels of mammalian domains, most plants domains may have small and/or medium sized cavities (Fig. 5). A detailed study of steric constraints in the plant protein pockets revealed that plant START domains may be a novel class of previously unknown lipid translocators with regulatory functions (GY unpublished data). This prediction requires experimental verification that is being presently addressed and has shown highly promising results (KS, personal communication to GY). The next step is to identify the possible terpene based ligands that 679

would fit the small plant START cavities, by employing various strategies like expression patterns, sub-cellular localization patterns, roles of other domains attached to these domains, and computational efforts like molecular docking approaches. Further progress requires molecular docking efforts for analysis of enzyme substrate complexes with the goal of explaining and ultimately predicting the stereo-specificity and substrate specificity of these enzymes.

### Expression Data and the Lectin Connection

Lectins have been known to be synthesized and accumulated after microbial attack and related situations and their roles in the natural defense mechanism of plants have been well documented [35]. Upon binding to carbohydrate or carbohydrate containing proteins, and after being ingested by herbivores, lectins bind to epithelial cell lining of the digestive tracts and interfere with nutrient absorption [36]. Our finding that the isoprenoid biosynthetic pathways are co-induced with specific lectins led us to further investigate this superfamily across known plant genomes. Due to the huge diversity in sequence and carbohydrate specificity of lectin domains across different organisms, the first obstacle in a comprehensive analysis was the lack of methods that could improve glycoprotein detection, and/or

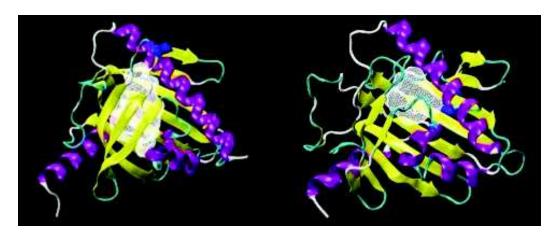
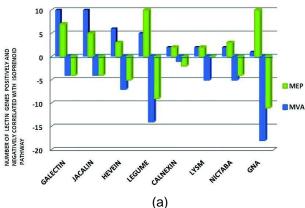


Fig. 5: Structural modelling of START domains from *A. thaliana*. The mammalian template domain (left) and plant model domain (right) are shown, depicting the conspicuous difference in cavity architecture, size and volume despite similarity in overall sequence as observed in multiple alignments in Fig. 4. Note that the mammalian cavity (white mesh; left) is much larger than the plant cavity (white mesh; right), which sterically cannot accommodate a cholesterol molecule (with an approx spatial volume of 700A<sup>3</sup>). We believe this is due to specific residues that block the START tunnel extension area in plants. All tunnels are depicted in mesh representation, domains helices are magenta while the beta sheet and turn regions are depicted in yellow and cyan respectively.

prediction of their sugar binding specificity. Our first effort therefore, was to develop an algorithm for identification, annotation and assignation of broad substrate specificity to plant lectin domains, using amino-acid information, and based upon a rigorous analysis of evolutionary relationships between several hundred characterized polypeptide sequences of lectin domains with known specificities, followed by optimization of all parameters for building accurate profile Hidden Markov Models (HMMs) [37]. A comparison of our results with those obtained from general protein annotation databases like PFam, Smart-db and Panther-db, revealed that our algorithm (PLecDom) performed better than any of these programs. Encouraged by the success of our approach, and its ease with handling large scale input, we applied the program to genomic, protein as well as EST data from over fifty plant genome sequencing projects currently underway [37]. In order to make the program accessible to researchers in the area, we converted the algorithm into an online automated server with a very simple user friendly interface, and help pages. This server is available freely to the scientific community at http://www.nipgr.res.in/ plecdom.html.

More than 7000 plant lectins were identified in this effort, and the specific domains identified in network analysis were analysed for their possible roles in the isoprenoid biosynthetic pathway regulation and/or induction, by comparing detailed expression profiles across tissues, developmental stages and stress conditions in both A. thaliana and O. sativa. The results are depicted in Fig. 6. As can be seen in Fig. 6a, the expression profiles of isoprenoid pathway enzymes are positively correlated with several lectin genes in the MVA and MEP pathways of Arabidopsis. In all, we have identified 38 lectins that are upregulated during MVA pathway induction in Arabidopsis, and 58 lectins whose expression is negatively correlated with MVA pathway genes. Further, we have identified 42 lectins whose expression profiles are upregulated during MEP pathway induction and 39 lectins that were downregulated with respect to the MEP pathway in Arabidopsis. Similarly, in case of the rice, expression profiles of 61 and 90 lectins were found to correlate







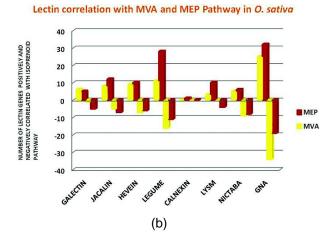


Fig. 6(a&b): Lectin correlation with the enzymes of isoprenoid biosynthetic pathway (MVA and MEP) in (a) A. thaliana and (b) O. sativa. In this figure each bar represents the number of lectin genes positively and negatively correlated with cytosolic Mevalonic Acid (MVA) and plastidic 2-C-Methyl-D-Erythritol 4-Phosphate (MEP) Pathway. Green bars represent plastidic MEP pathway whereas blue bars correspond to MVA pathway in A. thaliana. In case of O. sativa, red bars denote plastidic MEP pathway and yellow bars correspond to MVA pathway. In each case, positive Yaxis represents genes that are positively correlated whereas negative Y-axis represents the number of genes that are negatively correlated. Note that GNA-related and Legumes are the major correlates among all lectin families indicating the vital role of these two lectin superfamilies in isoprenoid based stress response. Lectins are classified by their sequence based families, namely, Galectin (galactose-binding lectin); Jacalin [lectin found in Artocarpus integrifolia (jackfruit) seed]; Hevein (N-acetyl-D-glucosamine specific lectin found in the latex of Hevea brasiliensis); Legume (sugar binding lectins found in plants of Leguminosae family); Calnexin (membrane bound lectin found in endoplasmic reticulum); LYSM (lectins having LysM domain first described in lysozyme); Nictaba [The tobacco (Nicotiana tabacum) agglutinin]; GNA (Galanthus nivalis agglutinin related lectin family)

Table 1a: List of accession numbers of lectin genes found to be correlated with isoprenoid biosynthetic pathway (MVA) in *A. thaliana.* The first column lists accession numbers of lectin genes found to be positively correlated with MVA pathway of *A. thaliana.* Second column shows, for each lectin gene, the name of the sequence based lectin family to which it belongs. The third and fourth column represents the accession numbers and families of lectin genes found to be negatively correlated with MVA pathway of *A. thaliana*, respectively. As mentioned already, lectins are classified by their sequence based family names that is Galectin (galactose-binding lectin); Jacalin (lectin found in *A. integrifolia* (jackfruit) seed); Hevein (N-acetyl-D-glucosamine specific lectin found in the latex of *Hevea brasiliensis*); Legume (sugar binding lectins found in plants of Leguminosae family); Calnexin (membrane bound lectin found in Endoplasmic Reticulum); LYSM (lectins having LysM domain first described in lysozyme); Nictaba (The tobacco (*Nicotiana tabacum*) agglutinin); GNA (*G. nivalis* agglutinin related lectin family).

Accession no. of lectins +vely correlated with MVA	Lectin name	Accession no. of lectins -vely correlated with MVA	Lectin name
AT1G02360	Hevein domain	AT1G05850	Hevein domain
AT1G05850	Hevein	AT1G08450	Calreticulin/Calnexin
AT1G09210	Calreticulin/Calnexin	AT1G11340	GNA
AT1G11280	GNA	AT1G11350	GNA
AT1G21880	LysM lectin	AT1G21880	LysM
AT1G27120	Galectin	AT1G34300	GNA
AT1G45130	Galectin	AT1G55000	LysM
AT1G52060	Jacalin	AT1G61420	GNA
AT1G52070	Jacalin	AT1G61460	GNA
AT1G53070	Legume	AT1G61610	GNA
AT1G56340	Calreticulin/Calnexin	AT1G63090	Nictaba
AT1G56680	Hevein	AT1G65390	Nictaba
AT1G74800	Galectin	AT1G65790	GNA
AT1G77630	LysM	AT1G65800	GNA
AT2G02230	Nictaba	AT1G67520	GNA
AT2G28470	Galectin	AT1G73040	Jacalin
AT2G33070	Jacalin	AT1G77630	LysM
AT2G39310	Jacalin	AT1G78820	GNA
AT2G43590	Hevein	AT1G78860	GNA
AT2G43610	Hevein	AT2G02360	Nictaba
AT3G13750	Galectin	AT2G19130	GNA
AT3G16390	Jacalin	AT2G23770	LysM
AT3G16430	Jacalin	AT2G33580	LysM
AT3G16440	Jacalin	AT2G37710	Legume
AT3G16450	Jacalin	AT2G43570	Hevein
AT3G16460	Jacalin	AT2G43620	Hevein

(Table 1a contd ...)

Accession no. of lectins +vely correlated with MVA	Lectin name	Accession no. of lectins -vely correlated with MVA	Lectin name
AT3G16470	Jacalin	AT2G43730	Jacalin
AT3G52840	Galectin	AT3G04720	Hevein
AT3G53000	Nictaba	AT3G08870	Legume
AT3G53380	Legume	AT3G12500	Hevein
AT3G54080	Legume	AT3G16390	Jacalin
AT4G01700	Hevein	AT3G16460	Jacalin
AT4G21060	Galectin	AT3G16530	Legume
AT4G36360	Galectin	AT3G52840	Galectin
AT5G01090	Legume	AT3G53000	Nictaba
AT5G20710	Galectin	AT3G53380	Legume
AT5G55830	Legume	AT3G53810	Legume
AT5G56870	Galectin	AT3G54420	Hevein
		AT3G59700	Legume
		AT4G01700	Hevein
		AT4G02410	Legume
		AT4G04960	Legume
		AT4G11900	GNA
		AT4G19840	Nictaba
		AT4G27300	GNA
		AT4G28350	Legume
		AT4G29050	Legume
		AT4G32300	GNA
		AT5G01540	Legume
		AT5G01550	Legume
		AT5G03350	Legume
		AT5G18470	GNA
		AT5G35370	GNA
		AT5G56870	Galectin
		AT5G60270	Legume
		AT5G60900	GNA
		AT5G62620	Galectin
		AT5G63800	Galectin

# (Continuation of Table 1a)

Table 1b: List of accession numbers of lectin genes found to be correlated with isoprenoid biosynthetic pathway (MEP) in *A. thaliana*. The first column lists accession numbers of lectin genes found to be positively correlated with MEP pathway of *A. thaliana*. Second column shows, for each lectin gene, the name of the sequence based lectin family to which it belongs. The third and fourth column represents the accession numbers and families of lectin genes found to be negatively correlated with MEP pathway of *A. thaliana*, respectively. As mentioned already, lectins are classified by their sequence based family names that is Galectin (galactose-binding lectin); Jacalin (lectin found in *A. integrifolia* (jackfruit seed); Hevein (N-acetyl-D-glucosamine specific lectin found in the latex of *H. brasiliensis*); Legume (sugar binding lectins found in plants of Leguminosae family); Calnexin (membrane bound lectin found in Endoplasmic Reticulum); LYSM (lectins having LysM domain first described in lysozyme); Nictaba ( The tobacco (*N. tabacum*) agglutinin); GNA (*G nivalis* agglutinin related lectin family)

Accession no. of lectins +vely correlated with MVA	Lectin name	Accession no. of lectins -vely correlated with MVA	Lectin name
AT1G05850	Hevein	AT1G11280	GNA
AT1G08450	Calreticulin/Calnexin	AT1G12710	Nictaba
AT1G09210	Calreticulin/Calnexin	AT1G34300	GNA
AT1G11300	GNA	AT1G56340	Calreticulin/Calnexin
AT1G11330	GNA	AT1G56680	Hevein
AT1G11350	GNA	AT1G61370	GNA
AT1G12710	Nictaba	AT1G61550	GNA
AT1G19715	Jacalin	AT1G67520	GNA
AT1G45130	Galectin	AT1G74800	Galectin
AT1G51940	LysM	AT1G78860	GNA
AT1G55000	LysM	AT2G02230	Nictaba
AT1G65790	GNA	AT2G26820	Nictaba
AT1G65800	GNA	AT2G33070	Jacalin
AT1G74800	Galectin	AT2G43580	Hevein
AT1G78820	GNA	AT2G43610	Hevein
AT1G78860	GNA	AT2G43690	Legume
AT2G02360	Nictaba	AT3G04720	Hevein
AT2G32810	Galectin	AT3G21380	Jacalin
AT2G37710	Legume	AT3G45420	Legume
AT2G39310	Jacalin	AT3G51710	GNA
AT2G43620	Hevein	AT3G53000	Nictaba
AT3G04720	Hevein	AT3G53380	Legume
AT3G08870	Legume	AT3G53810	Legume
AT3G13750	Galectin	AT3G54080	Legume
AT3G16390	Jacalin	AT3G54420	Hevein
AT3G16450	Jacalin	AT4G03230	GNA

(Table 1b contd ...)

Accession no. of lectins +vely correlated with MVA	Lectin name	Accession no. of lectins -vely correlated with MVA	Lectin name
AT3G16460	Jacalin	AT4G04960	Legume
AT3G16530	Legume	AT4G21060	Galectin
AT3G52840	Galectin	AT4G21390	GNA
AT4G02410	Legume	AT5G03700	GNA
AT4G02420	Legume	AT5G06740	Legume
AT4G11900	GNA	AT5G18470	GNA
AT4G19840	Nictaba	AT5G38540	Jacalin
AT4G27300	GNA	AT5G38550	Jacalin
AT4G29050	Legume	AT5G55830	Legume
AT5G01090	Legume	AT5G59260	Legume
AT5G01540	Legume	AT5G61790	Calreticulin/Calnexin
AT5G01550	Legume	AT5G62620	Galectin
AT5G03350	Legume	AT5G63810	Galectin
AT5G56870	Galectin		
AT5G60900	GNA		
AT5G63800	Galectin		

#### (Continuation of Table 1b)

positively with MVA and MEP pathway enzyme genes respectively, whereas 69 and 54 lectins were negatively correlated with pathway enzyme expression profiles. This is interesting in view of the fact that none of the pathway enzymes in either pathway has any direct interaction with lectins. Overall, the data shows that Arabidopsis has 68 genes from eight family of lectins upregulated with isoprenoid biosynthetic pathway during various expression studies, while in case of rice; 1-deoxy-Dxylulose 5-phosphate synthase (DXPS) enzyme of the MEP pathway has the maximum number of positively correlated lectins among all the enzymes of these pathways. This enzyme has also shown upregulation in various biotic stress conditions as suggested by MPSS data analysis (data not shown). In summary, the data suggests that it is the G. nivalis agglutinin (GNA) related lectins and Legumes that play substantial roles in isoprenoid related stress

responses. A list of the identified lectins given along with their family classification is provided in Table 1 (for *A. thaliana*) and Table 2 (for *O. sativa*). Efforts to further characterize these roles for each lectin identified are currently underway in our laboratory.

### Conclusion

As mentioned earlier, the vast amount of publicly available data offers a unique opportunity to translate genomic information into useful new biological knowledge for understanding terpene biosynthetic diversity, its regulation, and the molecular basis of scent emission, using sequence and structural data, background biological knowledge and novel inference techniques, that could eventually provide experimentally testable predictions of mutations leading to the precise production of specified isoprenoid end products. The molecular mechanisms by which specific isoprenoids influence stress Table 2a: List of accession numbers of lectin genes found to be correlated with isoprenoid biosynthetic pathway (MVA) in *Oryza* sativa. The first column lists accession numbers of lectin genes found to be positively correlated with MVA pathway of *O. sativa*. Second column shows, for each lectin gene, the name of the sequence based lectin family to which it belongs. The third and fourth column represents the accession numbers and families of lectin genes found to be negatively correlated with MVA pathway of *O. sativa*, respectively. As mentioned already, lectins are classified by their sequence based family names that is Galectin (galactose-binding lectin); Jacalin (lectin found in *A. integrifolia* (jackfruit) seed); Hevein (N-acetyl-D-glucosamine specific lectin found in the latex of *H. brasiliensis*); Legume (sugar binding lectins found in plants of Leguminosae family); Calnexin (membrane bound lectin found in Endoplasmic Reticulum); LYSM (lectins having LysM domain first described in lysozyme); Nictaba (The tobacco (*N. tabacum*) agglutinin); GNA (*G. nivalis* agglutinin related lectin family)

Accession no. of lectins +vely correlated with MVA	Lectin name	Accession no. of lectins -vely correlated with MVA	Lectin name
LOC_Os01g18400	Hevein	LOC_Os01g06500	Nictaba
LOC_Os01g24710	Jacalin	LOC_Os01g18400	Hevein
LOC_Os01g25280	Jacalin	LOC_Os01g25280	Jacalin
LOC_Os01g48040	GNA	LOC_Os01g47900	GNA
LOC_Os01g57560	GNA	LOC_Os01g48040	GNA
LOC_Os01g65030	GNA	LOC_Os01g57560	GNA
LOC_Os01g66610	GNA	LOC_Os01g65030	GNA
LOC_Os01g66680	GNA	LOC_Os01g66610	GNA
LOC_Os01g72810	GNA	LOC_Os01g66680	GNA
LOC_Os02g12730	Galectin	LOC_Os02g19530	Legume
LOC_Os02g48200	Legume	LOC_Os02g42780	Legume
LOC_Os02g48210	Legume	LOC_Os02g48200	Legume
LOC_Os02g52850	GNA	LOC_Os02g48210	Legume
LOC_Os02g56750	Nictaba	LOC_Os02g52850	GNA
LOC_Os02g56820	Nictaba	LOC_Os02g56750	Nictaba
LOC_Os03g04060	Hevein	LOC_Os02g56820	Nictaba
LOC_Os03g06940	Galectin	LOC_Os03g02550	Nictaba
LOC_Os03g12150	GNA	LOC_Os03g04060	Hevein
LOC_Os03g60810	Legume	LOC_Os03g12150	GNA
LOC_Os04g01310	GNA	LOC_Os03g30890	GNA
LOC_Os04g03579	Legume	LOC_Os03g56160	Legume
LOC_Os04g09390	Hevein	LOC_Os03g60810	Legume
LOC_Os04g15580	GNA	LOC_Os04g01310	GNA
LOC_Os04g23760	GNA	LOC_Os04g03579	Legume
LOC_Os04g28780	GNA	LOC_Os04g15580	GNA
LOC_Os04g34390	GNA	LOC_Os04g23760	GNA
LOC_Os04g41620	Hevein	LOC_Os04g28780	GNA
LOC_Os04g41680	Hevein	LOC_Os04g34250	GNA
LOC_Os04g42740	GNA	LOC_Os04g34270	GNA

(Table 2a contd ...)

Accession no. of lectins +vely correlated with MVA	Lectin name	Accession no. of lectins -vely correlated with MVA	Lectin name
LOC_Os04g44900	Legume	LOC_Os04g34290	GNA
LOC_Os04g44910	Legume	LOC_Os04g34390	GNA
LOC_Os04g49480	Legume	LOC_Os04g41620	Hevein
LOC_Os04g53998	GNA	LOC_Os04g41680	Hevein
LOC_Os04g54180	GNA	LOC_Os04g42740	GNA
LOC_Os05g04690	Hevein	LOC_Os04g44900	Legume
LOC_Os06g17490	Legume	LOC_Os04g44910	Legume
LOC_Os06g22290	Legume	LOC_Os04g53998	GNA
LOC_Os06g37750	GNA	LOC_Os04g54130	GNA
LOC_Os06g40030	GNA	LOC_Os04g54180	GNA
LOC_Os06g51050	Hevein	LOC_Os05g04690	Hevein
LOC_Os06g51060	Hevein	LOC_Os05g07300	GNA
LOC_Os07g09670	Galectin	LOC_Os05g13770	Legume
LOC_Os07g36570	GNA	LOC_Os05g42210	GNA
LOC_Os08g03020	Legume	LOC_Os06g06960	GNA
LOC_Os08g03070	Legume	LOC_Os06g22290	Legume
LOC_Os09g27890	LysM	LOC_Os06g37750	GNA
LOC_Os09g28180	GNA	LOC_Os06g40030	GNA
LOC_Os09g37600	LysM	LOC_Os06g51050	Hevein
LOC_Os09g37834	GNA	LOC_Os07g03860	Legume
LOC_Os10g04270	Jacalin	LOC_Os07g03870	Legume
LOC_Os10g06680	GNA	LOC_Os07g03880	Legume
LOC_Os10g18400	Galectin	LOC_Os07g09670	Galectin
LOC_Os10g39680	Hevein	LOC_Os07g36570	GNA
LOC_Os11g03860	GNA	LOC_Os08g03020	Legume
LOC_Os11g05240	GNA	LOC_Os08g03070	Legume
LOC_Os11g39530	Jacalin	LOC_Os08g05480	Nictaba
LOC_Os12g03594	Nictaba	LOC_Os09g28180	GNA
LOC_Os12g03740	Nictaba	LOC_Os10g04270	Jacalin
LOC_Os12g12720	Jacalin	LOC_Os10g06680	GNA
LOC_Os12g30180	Nictaba	LOC_Os10g39680	Hevein
LOC_Os12g44320	GNA	LOC_Os11g03860	GNA
		LOC_Os11g05240	GNA
		LOC_Os11g39530	Jacalin
		LOC_Os12g03594	Nictaba
		LOC_Os12g03740	Nictaba
		LOC_Os12g12720	Jacalin
		LOC_Os12g14440	Jacalin

LOC\_Os12g30180

LOC\_Os12g44320

Nictaba

GNA

# (Continuation of Table 2a)

Table 2b: List of accession numbers of lectin genes found to be correlated with isoprenoid biosynthetic pathway (MEP) in *O. sativa.* The first column lists accession numbers of lectin genes found to be positively correlated with MEP pathway of *O. sativa.* Second column shows, for each lectin gene, the name of the sequence based lectin family to which it belongs. The third and fourth column represents the accession numbers and families of lectin genes found to be negatively correlated with MEP pathway of *O. sativa,* respectively. As mentioned already, lectins are classified by their sequence based family names that is Galectin (galactose-binding lectin); Jacalin (lectin found in *A. integrifolia* (jackfruit) seed); Hevein (N-acetyl-D-glucosamine specific lectin found in the latex of *H. brasiliensis*); Legume (sugar binding lectins found in plants of Leguminosae family); Calnexin (membrane bound lectin found in Endoplasmic Reticulum); LYSM (lectins having LysM domain first described in lysozyme); Nictaba (The tobacco (*N. tabacum*) agglutinin); GNA (*G. nivalis* agglutinin related lectin family)

Accession no. of lectins +vely correlated with MVA	Lectin name	Accession no. of lectins -vely correlated with MVA	Lectin name
LOC_Os01g06500	Nictaba	LOC_Os01g06240	GNA
LOC_Os01g18400	Hevein	LOC_Os01g18400	Hevein
LOC_Os01g24710	Jacalin	LOC_Os01g24710	Jacalin
LOC_Os01g25280	Jacalin	LOC_Os01g25280	Jacalin
LOC_Os01g36550	LysM	LOC_Os01g48040	GNA
LOC_Os01g47900	GNA	LOC_Os01g57560	GNA
LOC_Os01g48000	GNA	LOC_Os02g12730	Galectin
LOC_Os01g57480	GNA	LOC_Os02g39330	Hevein
LOC_Os01g57560	GNA	LOC_Os02g48200	Legume
LOC_Os01g65030	GNA	LOC_Os02g48210	Legume
LOC_Os01g66250	GNA	LOC_Os02g52850	GNA
LOC_Os01g66610	GNA	LOC_Os02g56750	Nictaba
LOC_Os01g66680	GNA	LOC_Os02g56820	Nictaba
LOC_Os01g72810	GNA	LOC_Os03g02550	Nictaba
LOC_Os02g12730	Galectin	LOC_Os03g06940	Galectin
LOC_Os02g19530	Legume	LOC_Os04g01310	GNA
LOC_Os02g39330	Hevein	LOC_Os04g03579	Legume
LOC_Os02g45320	Nictaba	LOC_Os04g09390	Hevein
LOC_Os02g48200	Legume	LOC_Os04g21130	Nictaba
LOC_Os02g48210	Legume	LOC_Os04g23760	GNA
LOC_Os02g52850	GNA	LOC_Os04g28780	GNA
LOC_Os02g53000	LysM	LOC_Os04g34390	GNA
LOC_Os03g02685	LysM	LOC_Os04g42740	GNA
LOC_Os03g04060	Hevein	LOC_Os04g44900	Legume
LOC_Os03g04110	LysM	LOC_Os04g44910	Legume
LOC_Os03g06940	Galectin	LOC_Os04g49480	Legume

(Table 2b contd ...)

Accession no. of lectins +vely correlated with MVA	Lectin name	Accession no. of lectins -vely correlated with MVA	Lectin name
LOC_Os03g12150	GNA	LOC_Os04g54180	GNA
LOC_Os03g30890	GNA	LOC_Os05g04690	Hevein
LOC_Os03g56160	Legume	LOC_Os06g17490	Legume
LOC_Os03g60810	Legume	LOC_Os06g22290	Legume
LOC_Os04g01320	GNA	LOC_Os06g37750	GNA
LOC_Os04g03579	Legume	LOC_Os06g51050	Hevein
LOC_Os04g09390	Hevein	LOC_Os07g09670	Galectin
LOC_Os04g15580	GNA	LOC_Os07g36570	GNA
LOC_Os04g22120	Legume	LOC_Os08g03020	Legume
LOC_Os04g23700	GNA	LOC_Os08g03070	Legume
LOC_Os04g34250	GNA	LOC_Os08g03160	Legume
LOC_Os04g34270	GNA	LOC_Os08g05480	Nictaba
LOC_Os04g34290	GNA	LOC_Os09g27890	LysM
LOC_Os04g41620	Hevein	LOC_Os09g28180	GNA
LOC_Os04g41680	Hevein	LOC_Os09g37600	LysM
LOC_Os04g53998	GNA	LOC_Os09g37780	GNA
LOC_Os04g54002	GNA	LOC_Os09g37834	GNA
LOC_Os05g04690	Hevein	LOC_Os10g04270	Jacalin
LOC_Os05g13770	Legume	LOC_Os10g06680	GNA
LOC_Os05g42210	GNA	LOC_Os10g38040	LysM
LOC_Os05g43170	CAL	LOC_Os10g39680	Hevein
LOC_Os05g43240	Jacalin	LOC_Os11g03860	GNA
LOC_Os06g06940	GNA	LOC_Os11g05240	GNA
LOC_Os06g06960	GNA	LOC_Os11g39530	Jacalin
LOC_Os06g17490	Legume	LOC_Os12g03594	Nictaba
LOC_Os06g22290	Legume	LOC_Os12g03740	Nictaba
LOC_Os06g37560	Galectin	LOC_Os12g30180	Nictaba
LOC_Os06g37750	GNA	LOC_Os12g44320	GNA
LOC_Os06g40030	GNA		
LOC_Os06g51050	Hevein		
LOC_Os06g51060	Hevein		

# (Continuation of Table 2b)

(Table 2b contd ...)

# (Continuation of Table 2b)

1	
LOC_Os06g51360	LysM
LOC_Os07g03790	Legume
LOC_Os07g03830	Legume
LOC_Os07g03840	Legume
LOC_Os07g03860	Legume
LOC_Os07g03870	Legume
LOC_Os07g03880	Legume
LOC_Os07g03900	Legume
LOC_Os07g03930	Legume
LOC_Os07g04040	Legume
LOC_Os07g18230	Legume
LOC_Os07g38250	Legume
LOC_Os07g38800	Legume
LOC_Os07g38810	Legume
LOC_Os08g02996	Legume
LOC_Os08g03020	Legume
LOC_Os08g03070	Legume
LOC_Os08g03160	Legume
LOC_Os08g05480	Nictaba
LOC_Os09g27890	LysM
LOC_Os09g37600	LysM
LOC_Os09g37834	GNA
LOC_Os09g37890	GNA
LOC_Os10g04270	Jacalin
LOC_Os10g39680	Hevein
LOC_Os11g10290	GNA
LOC_Os11g39420	Jacalin
LOC_Os11g39530	Jacalin
LOC_Os12g03594	Nictaba
LOC_Os12g14440	Jacalin
LOC_Os12g30180	Nictaba
LOC_Os12g34320	GNA
LOC_Os12g44320	GNA

responses in plants are not clearly understood. This work was carried out with the aim to integrate information at the level of genes, proteins, their interactions and expression profiles, enabling a reliable systems analysis. Insights from the network analysis have led to the identification of a new class of plant amplified transcription factors, which are hypothesized to be regulating at least partially, the biosynthetic crosstalk between the pathways. Detailed sequence and structural analysis of these transcription factors proteins supports the idea that they may be involved in the crosstalk between the MVA-MEP pathways. The most fascinating aspect of this endeavor has been the discovery of several new directions for potential research, including prediction of substrate specificity of plant lectin domains and START domains, and bioprospecting for new molecular players that may have strong agribiotechnological, pharmaceutical and economic applications.

#### Acknowledgements

The authors are thankful to the Director, NIPGR for encouragement and support. The authors (SK) and (PP) are recipients of Senior Research Fellowship from CSIR, India. The author (SS) was funded by grants to National Institute of Plant Genome Research under BTIS project, from the Department of Biotechnology, Government of India. One of the authors (KB) was a recipient of the Summer Research Fellowship Programme (SRFP) of the Indian Science Academies at the time of the work. This work was supported by grants to one of the authors (GY) under the Senior Innovative Young Biotechnologist Award (IYBA) by DBT, India.

#### References

- 1. Fraenkel GS Science 129 (1959) 1466
- Kirby J and Keasling J D Annu Rev Plant Biol 60 (2009) 335
- Buchanan B B, Gruissem W and Joneas R L (Eds) Biochemistry and Molecular Biology of Plants Rodney Croteau, Toni M Kutchan and Norman G Lewis (2000) pp 1250
- Sharkey TD, Wiberley AE and Donohue AR Ann Bot 101 (2008) 5
- 5. Cane DE Science **287** (2000) 818
- Jomaa H, Wiesner J, Sanderbrand S, Altincicek B, Weidemeyer C, Hintz M, Turbachova I, Eberl M, Zeidler J, Lichtenthaler H K, Soldati D and Beck E *Science* 285 (1999) 1573
- Eberl M, Hintz M, Reichenberg A, Kollas A K and Wiesner J and Jomaa H FEBS Lett 544 (2003) 4
- Laule O, Furholz A, Chang H S, Zhu T and Wang X, Heifetz P B, Gruissem W and Lange M *Proc Natl Acad Sci USA* 100 (2003) 6866
- 9. Date S V and Marcotte E M *Nat Biotechnol* **21** (2003) 1055
- Stark C, Breitkreutz B J, Chatr-aryamontri A, Boucher L, Oughtred R, Livstone M S, Nixon J, Auken K V, Wang X, Shi X, Reguly T, Rust J M, Winter A, Dolinski K and

Tyers M Nucleic Acids Res 39 (2011) D698

- Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen L J and Mering C V *Nucleic Acids Res* 39 (2011) D561
- Lin M, Shen X and Chen X Nucleic Acids Res 39 (2011) D1134
- Akiyama K, Chikayama E, Yuasa H, Shimada Y and Tohge T, Shinozaki K, Hirai M Y, Sakurai T, Kikuchi J and Saito K *In Silico Biol* 8 (2008) 339
- Nakano M, Nobuta K, Vemaraju K, Tej S S, Skogen J W and Meyers B C *Nucleic Acids Res* 34 (2006) D731
- Zimmermann P, Laule O, Schmitz J, Hruz T and Bleuler S and Gruissem W *Mol Plant* 1 (2008) 851
- Katari M S, Nowicki S D, Aceituno F F, Nero D, Kelfer J, Thompson P, Cabello J M, Davidson R S, Goldberg A P, Shasha D E, Coruzzi G M and Gutierrez R A *Plant Physiol* 152 (2010) 500
- Rhee S Y, Beavis W, Bernadine T Z, Chen G, Dixon D, Doyle A, Garcia-Hernandez M, Huala E, Lander G, Montoya M, Miller N, Mueller LA, Mundodi S, Reiser L, Tacklind J, Weems D C, Wu Y, Xu I, Yoo D, Yoon J and Zhang P *Nucleic Acids Res* **31** (2003) 224
- Altschul S F, Gish W, Miller W, Myers E W and Lipman D J J Mol Biol 215 (1990) 403

- 19. Felsenstein J Science 246 (1989) 941
- 20. Rice P, Longden I and Bleasby A *Trends Genet* **16** (2000) 276
- 21. Thompson J D, Higgins D G and Gibson T J *Nucleic Acids Res* 22 (1994) 4673
- 22. Lopes C T, Franz M, Kazi F, Donaldson S L, Morris Q and Bader G D *Bioinformatics* **26** (2010) 2347
- 23. Batagelj V and Mrvar A Connections 21 (1998) 47
- 24. Kleywegt G J and Jones T A Acta Crystallogr D Biol Crystallogr 50 (1994) 178
- 25. Canutescu AA, Shelenkov AA and Dunbrack R L J *Protein* Sci 12 (2003) 2001
- Humphrey W, Dalke A and Schulten K J Molec Graphics 14 (1996) 33
- 27. DeLano W L, The PyMOL Molecular Graphics System on World Wide Web http://www.pymol.org" *DeLano Scientific* (2002)
- Lange B M, Rujan T, Martin W and Croteau R *Proc Natl* Acad Sci USA 97 (2000) 13172

- 29. Ponting C P and Aravind L *Trends Biochem Sci* 24 (1999) 130
- 30. Alpy F and Tomasetto C J Cell Sci 118 (2005) 2791
- Schick K, Nguyen D, Karlowski W M and Mayer K F Genome Biol 5 (2004) R41
- 32. Ariel F D, Manavella P A, Dezar C A and Chan R L *Trends Plant Sci* **12** (2007) 419
- Fu D, Uauy C, Distelfeld A, Blechl A, Epstein L, Dubcovsky J, Fahima T, Sela H and Chen X Science 323 (2009) 1357
- Romanowski M J, Soccio R E, Breslow J L and Burley S K Proc Natl Acad Sci USA 99 (2002) 6949
- Loon L C V, Pierpoint W S, Boller T and Conejero V *Plant Molecular Biology Reporter* 12 (1994) 245
- Peumans W J and Damme E J V *Plant Physiol* **109** (1995) 347
- Shridhar S, Chattopadhyay D and Yadav G Nucleic Acids Res 37 (2009) W452.