International Journal of Bioassays
ISSN: 2278-778X
www.ijbio.com

Research Article

PROTEIN FUNCTION PREDICTION FOR GPI 12P AND PIG-L WITH RETENTION OF ENDOPLASMIC RETICULUM LOCALIZED SIGNALING SEQUENCE

Godavarthi Rajan Bhagyasri*and KB Chandrasekhar
Department of Chemistry, Jawaharlal Nehru Technological University, Anantapur-515002, Andhrapradesh, India.

Received for publication: July 11, 2013; **Revised:** July 27, 2013; **Accepted:** August 14, 2013

Abstract: GPI anchor biosynthesis process involves a complex of 6-7 proteins and at least 5 enzymes with Gpi12p/PIG-L proteins exhibit catalysis pathway by a specific functions are validated in archea, bacteria and eukaryote, based on domain as well as motif prediction and identification of signaling peptide retention. However, the presence of signaling peptides are responsible for targeting proteins to the ER for subsequent transport through the secretory pathway involves transfer of N-acetyl glucosamine (GlcNAc) from UDP-GlcNAc to phosphatidylinositol (PI). GPI anchor biosynthesis can affect the function and even the viability of cells.

Keywords: Gpi Anchor, Gpi 12p, Pig-L, Signaling Sequence, Motif Prediction

INTRODUCTION

Glycosylphosphatidylinositol (GPI) anchors are structurally complex glycophospholipids that include amide, glycosidic, phosphodiester and hydroxyester linkages between their various components. Which can be attached to the C-terminal end of secretory proteins after they have been translocated into the ER during posttranslational modification and acts as a membrane anchor for a significant proportion of eukaryotic cell surface glycoproteins [1]. Experimental determination of GPI-anchored proteins was carried out by means of phospholipase C or D solubilization. Mostly, these anchored proteins are found in mammals [2] yeast [3] and plants [4] but are especially abundant in protozoan parasites [5]. The core structure and the biosynthetic steps of GPI are basically conserved in various organisms [6] Most of the GPI-anchored proteins are translocated to the plasma membrane. Although, there are some evidences of proteins residing in different compartments such as the ER or the Golgi apparatus [7]. The common core structure of GPI consists of inositol phospholipid, GlcN, three mannoses and an ethanolamine phosphate. All GPI-anchors have similar chemical structures, with slight distinction among different species. The core of the anchor molecule comprises a sugar moiety and a phosphatidylinositol molecule, linked to two long-chain fatty acids. The sugar moiety is composed of a glucosamine, three mannose residues and one phosphoethanolamine that can form an amide bond with the C-terminal residue of a polypeptide.

The GPI biosynthetic pathway is structurally, topologically, and spatially complex. While many aspects of the biosynthesis of GPI-protein complexes remain to be determined, it is apparent that both the biosynthesis of GPI precursors and post-translational modification of proteins with GPI take place in the endoplasmic reticulum. The process starts on the cytoplasmic side of this membrane and is completed on the lumenal side, so the intermediate glycophospholipid must be flipped across the membrane. In mammalian cells, the lipid precursor is a conventional phosphatidylinositol molecule, which is first attached to an *N*-acetylglucosamine residue [8]. This is de-acetylated before a saturated fatty acid is attached to the inositol residue, and this is followed by a sequence of reactions in which further carbohydrate moieties and phosphorylethanolamine are added.

General features of GPI-anchored proteins are summarized in the following. Upon synthesis and upon recognition of a N-terminal signal peptide, proteins are targeted to the ER, where the C-terminal portion of the protein interacts with the transamidase complex by means of hydrophobic residues. This complex is responsible for the removal of the C-terminal domain - known as propeptide and for the binding to a free GPI-anchor inserted into the internal leaflet of the ER membrane. Unfortunately, no consensus sequence can be found to describe the localization of the ω-site. Nevertheless, the C-terminal portion of the non-cleaved proteins can be roughly separated into different portions [9].

MATERIAL AND METHODS

Sequences retrieval:

This work proposed to study the GPI12-an essential gene involved in the second step of GPI biosynthesis. Identify putative Gpi12p/ PIG-L sequences from the available databases for Archea, Bacteria and Eukaryotes using BLASTp with the yeast and mammalian protein sequences as query.

***Corresponding Author:**
Godavarthi Rajan Bhagyasri,
Department of Chemistry,
Jawaharlal Nehru Technological University,
Anantapur-515002, Andhrapradesh, India.

Identifying conserve Domain of GPI anchoring formation proteins:

Domains are retrieved for Eukaryote , archea and bacteria of each Individual species by using online database, Batch CD-search tool[10][11][12] classified the domains into three categories mainly specific, Super family and Multi domain, based on sequence and structural similarity clusters. Conserve domains are predicted for a query sequence through FASTA sequence search, it provides domain output with brief description retrieved from different online database.

Motif prediction:

Scanned the sequences against the PROSITE-collection of motif database of protein families and domains. It is based on the observation that, there is huge number of different proteins; most of them are grouped on the basis of similarities in their sequences into a limited number of families. Proteins or protein domains belonging to a particular family generally derived as a functional attributes.

It is apparent of studying protein sequence families, that some regions have been determined as better conserved patterns. These regions are generally important for the function of a protein. By analyzing the constant and variable properties of such groups of similar sequences, it is possible to identify a protein family or domain, which distinguishes its members from all other unrelated proteins. Prosite scan [13][14][15][16] works based on fingerprints by the police for identification purposes. A fingerprint is generally sufficient to identify a given individual. Similarly, a protein signature can be used to assign a newly sequenced protein to a specific family of proteins and thus to formulate hypotheses about its function.

PROSITE works on patterns and profiles specific for more than thousand protein families or domains. Each of these signatures comes with documentation providing background information on the structure and function of these proteins.

Validating ER peptide:

Signal peptides (SPs) are responsible for targeting proteins to the ER for subsequent transport through the secretory pathway [17][18]. SPs generally consist of three regions: a positively charged n-region, a hydrophobic h-region, and a polar c-region leading up to the signal peptidase cleavage site. The most well conserved motif of SPs is the presence of a small and neutral amino acid at positions ÿ3 and ÿ1 relative to the cleavage site [19][20]. In mitochondrial targeting peptides (mTPs), Arg, Ala and Ser are over-represented while negatively charged amino acid residues (Asp and Glu) are rare. Only weak consensus sequences have been found, the most prominent being a conserved Arg in position ÿ2 or ÿ3 relative to the mitochondrial processing peptidase (MPP) cleavage site. Furthermore, mTPs are believed to form an amphiphilic a- helix that is of importance for import of the nascent protein into the mitochondrion [21][22][23]. Some matrix proteins are cleaved a second time by the mitochondrial intermediate peptidase (MIP), which removes an additional eight to nine residues from the mature protein [24][25]. A subset of the mitochondrial proteins are first imported into the matrix, where there the N-terminal cTP.

Subcellular localization was identifying through SPs (SignalP) [26] in a protein sequence. TargetP predict one of four different localizations (chloroplast, mitochondrion, ER/golgi/secreted, and other') to a query sequence, and also predict a potential cleavage site for presequence removal. TargetP is built from two layers of neural networks, where the first layer contains one dedicated network for each type of presequence (cTP, mTP, SP), and the second is an integrating network that outputs that the actual prediction (cTP, mTP, SP, other). A non-plant version of TargetP that distinguishes only between mTPs, SPs and other has also been constructed. TargetP is able to discriminate between cTPs, mTPs, and SPs with sensitivities and specificities higher than what has been obtained with other available subcellular localization predictors, and has a relatively well working cleavage site prediction for all involved target sequences.

RESULTS AND DISCUSSION

Conserved domain declaration:

Domains can be thought of as distinct functional and/or structural units of a protein. These two classifications coincide rather often, as a matter of fact, and what is found as an independently folding unit of a polypeptide chain also carries specific function. Domains are often identified as recurring (sequence or structure) units, which may exist in various forms. In molecular function domains may have been utilized as building blocks, and may have been recombined in different arrangements to modulate protein function. We define conserved domains as recurring units in molecular function, the extents of which can be determined by sequence analysis.

The biosynthesis of GPI anchor in all eukaryotes begins with the transfer of N-acetylglucosamine (GlcNAc) from UDP-GlcNAc to phosphatidylinositol (PI). This step is catalyzed by a complex of 6-7 proteins and a complex of at least 5 enzymes (Table 1) in GPI anchoring synthesis. In the second step of GPI biosynthesis, GlcNAc-PI is de-*N*-acetylated to yield glucosamine-PI (GlcN-PI). This step is catalyzed by the product of GPI12/*PIG-L* gene.

Gpi12p/ PIG-L proteins are ER resident membrane proteins which catalyze the removal of the acetyl moiety from GlcNAc-PI to yield glucosamine-PI (GlcN-PI), a crucial intermediate in GPI biosynthesis. However, due to the difficulty in expression and purification of these proteins, the biochemical characterization of pure proteins or their active domains(Table 1 & 2) are predicted through bioinformatics approach, have been successful shorten the domain identities with functional description.

Motif pattern identification from predicted domains:

Conserved domains contain conserved sequence patterns or motifs, which allow for their detection in polypeptide sequences. The distinction between domains and motifs is not sharp, however, especially in the case of short repetitive units. Functional motifs are also present outside the scope of structurally conserved domains. Found that enzymes involved in GPI anchoring synthesis has unique active site H - Histidine and metallic binding site C – Cysteine are observed clearly [Table 3]. Functional studies identified that there is no active and binding site found in Gpi12p/PIG-L though, proteins suggested a role for metal binding in the activity of the enzyme by amalgamate with other enzymes [Table 1]. O-GlcNAc (O-linked N-acetylglucosamine) is a form of protein glycosylation site found exclusively in the nucleus and cytoplasm of Organism cells. Some glycosylation site motifs are showing same in archea, bacteria and eukaryote enzymes, involve in GPI anchoring synthesis, which are highlighted in green color font in [Table 3] as well as considering with [Table 4], one motif has similar in archea and eukaryote. However, it tells about the functional relationship between the organisms implicitly with glycosylation site. Many proteins are modified at their serine and threonine hydroxyl groups by the attachment of O-GlcNAc motifs [Table 3 & 4] are predicted through PROSITESCAN online database. Proteins that regulate trafficking into and out of the nuclear pore are extensively O-GlcNAcylated. Phosphorylated O-GlcNAc proteins form reversible multimeric complexes with other proteins and these associations are often regulated by phosphorylation. O-GlcNAc proteins may play a key role in pathogenesis of tumors and various cancer cells. O-GlcNAc residues regulate the assembly of the preinitiation complex and are therefore important in transcriptional initiation. There is slight common among archeal putative PIG-L proteins, bacteria and the eukaryotic based on N-glycosylation sites, although have unique Phosphorylation site features [Table 3 & 4] within the selected organisms.

Retention of ER signaling peptide:

In general, the GPI anchor is synthesized in the endoplasmic reticulum (ER). Once the synthesis is complete, the whole anchor is transferred on to the carboxyl terminal of proteins that contain the GPI-attachment signal sequence. The GPI-anchored proteins are then transported to cell surface. The core GPI structure is conserved in most eukaryotic systems studied till date and is composed of $NH_2CH_2CH_2PO_4H$-6Man 1-2Man 1-6Man 1-4GlcN 1-6D-myo-inositol-1-HPO_4-lipid EtN-P-Man3GlcN-PI-lipid), where the lipid can be diacylglycerol, alkylacylglycerol or ceramide. Further modifications of the GPI structure occurs however in a species-specific and tissue-specific manner. Thus, for a specificity of 0.95, the corresponding sensitivities were found to be 0.63-0.96 non-plant predictor, as in Table 5, With similar resulting performances except that the limiting (in terms of size) categories in general scored lower sensitivities, while the largest categories of non-plant proteins, mTP and ``other'', respectively, scored better sensitivities. Reliability class (RC), according to the difference, _, between highest and second-highest network output score. If _ > 0.8, then RC. 1; if 0.6 < _ < 0.8, then RC. 2, etc. RC is useful in indication of the level of certainty in the prediction for a sequence. Overall, 99% of the sequences with RC. 1, and 93/95% of the sequences with RC. 2 were correctly predicted (Table 5). While the specificity within an RC is an expression of the reliability of a specific prediction, increased by only considering predictions at a particular RC as better. Based on the Loc –Localization as shown in Table 5. Property indicates that the GPI-attachment signal sequence conserved in most eukaryotic systems only.

Table.1: Displaying types of enzymes with their conserved domain function involved in formation of GPI anchoring protein and below there is a functional description of related enzymes with respective of domain

Type	rganism	Domains		
		Specific	**Super family**	**Multi**
Enzymes	Archea	SIRT5_Af1_CobB; M20_ArgE_DapE_like6; M20_ArgE_DapE_like1; HDAC_classII; HDAC_classII_APAH; M20_ArgE_DapE_like2; Hist_deacetyl; HDAC_AcuC_like ; HDAC; SIR2 family; SIR2_Af2	SIR2; Zinc_peptidase_like; Arginase_HDAC	NAD-dependent deacetylase; Deacetylases; diaminopimelate aminotransferase; succinyl-diaminopimelate desuccinylase; acetylornithine deacetylase; AcuC
	Bacteria	HDAC; HDAC_classII; HDAC_classIV; HDAC_classII_APAH; HDAC_AcuC_like; SIRT7; SIR2-like; SIR2_Af2; SIRT5_Af1_CobB; SIR2H; Hist_deacetyl; M20_ArgE_DapE_like5	Arginase_HDAC; SIR2; Zinc_peptidase_like	AcuC; NAD-dependent deacetylase; ArgE;
	Eukaryota	SIR2; HDAC_classIV; HDAC3;HDAC8, HDAC2; HDAC1; HDAC_classI; HDAC_Hos1; M20_ArgE-related; HDAC10_HDAC6-dom1	SIR2; Arginase_HDAC; Zinc_peptidase_like	SIR2; AcuC; Peptidase_M20

Enzyme Function Specification:

Archea, Bacteria, Eukaryote Specific Domain:

SIRT5_Af1_CobB: Eukaryotic, archaeal and prokaryotic group (class3) which includes human sirtuin SIRT5, Archaeoglobus fulgidus Sir2-Af1, and E. coli CobB; and are members of the SIR2 family of proteins.

HDAC_AcuC_like: Class I histone deacetylase AcuC (Acetoin utilization protein)-like enzymes; AcuC (Acetoin utilization protein) is a class I deacetylase found only in bacteria and is involved in post-translational control of the acetyl-coenzyme A synthetase (AcsA). Deacetylase AcuC works in coordination with deacetylase SrtN (class III), possibly to maintain AcsA in active (deacetylated) form and let the cell grow under low concentration of acetate.

***HDAC*:** Histone deacetylase (HDAC) also contain bacterial acetylpolyamine amidohydrolase (APAH) as well as other classes I, II, IV and related proteins; The HDAC/HDAC-like family includes Zn-dependent histone deacetylase classes I, II and IV (class III HDACs, also called sirtuins, are NAD-dependent and structurally unrelated, and therefore not part of this family).

***M20_ArgE-related*:** M20 Peptidases with simlarity to acetylornithine deacetylases; Peptidase M20 family, acetylornithine deacetylase (ArgE, Acetylornithinase, AO, N2-acetyl-L-ornithine amidohydrolase, EC 3.5.1.16)-related subfamily. Proteins in this subfamily have not yet been characterized, but have been predicted to have deacetylase activity. ArgE catalyzes a broad range of substrates, including N-acetylornithine, alpha-N-acetylmethionine and alpha-N-formylmethionine, while DapE catalyzes the hydrolysis of N-succinyl-L,L-diaminopimelate (L,L-SDAP) to L,L-diaminopimelate and succinate..

***Sir2-like*:** Prokaryotic group of uncharacterized Sir2-like proteins which lack certain key catalytic residues and conserved zinc binding cysteines; and are members of the SIR2 superfamily of proteins, silent information regulator 2 (Sir2) enzymes which catalyze NAD+-dependent protein/histone deacetylation.

SIRT7, and several bacterial homologs; and are members of the SIR2 family of proteins, silent information regulator 2 (Sir2) enzymes which catalyze NAD+-dependent protein/histone deacetylation. Sir2 proteins have been shown to

***HDAC10_HDAC6-dom1*:** Histone deacetylase 6, domain 1 and histone deacetylase 10.

Super family domain:

SIR2: superfamily of proteins includes silent information regulator 2 (Sir2) enzymes which catalyze NAD+-dependent protein/histone deacetylation, where the acetyl group from the lysine epsilon-amino group is transferred to the ADP-ribose moiety of NAD+, producing nicotinamide and the novel metabolite O-acetyl-ADP-ribose. Sir2 proteins, also known as sirtuins, are found in all eukaryotes and many archaea and prokaryotes and have been shown to regulate gene silencing, DNA repair, metabolic enzymes, and life span.

Arginase_HDAC: Arginase-like and histone-like hydrolases; Arginase-like/histone-like hydrolase superfamily includes metal-dependent enzymes that belong to Arginase-like amidino hydrolase family and histone/histone-like deacetylase class I, II, IV family, respectively. These enzymes catalyze hydrolysis of amide bond. Arginases are known to be involved in

control of cellular levels of arginine and ornithine, in histidine and arginine degradation and in clavulanic acid biosynthesis.

***Zinc_peptidase_like*:** Zinc peptidases M18, M20, M28, and M42; Zinc peptidases play vital roles in metabolic and signaling pathways throughout all kingdoms of life. This family corresponds to several clans in the MEROPS database, including the MH clan, which contains 4 families (M18, M20, M28, M42). The peptidase M20 family includes carboxypeptidases such as the glutamate carboxypeptidase from Pseudomonas, the thermostable carboxypeptidase Ss1 of broad specificity from archaea and yeast Gly-X carboxypeptidase.

Multi domain:

AcuC: Deacetylases, including yeast histone deacetylase and acetoin utilization protein [Chromatin structure and dynamics / Secondary metabolites biosynthesis, transport, and catabolism]

ArgE: *Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase and related deacylases [Amino acid transport and metabolism] SIR2: NAD-dependent protein deacetylases, SIR2 family [Transcription]*

***Peptidase_M20*:** Peptidase family M20/M25/M40; this family includes a range of zinc metallopeptidases belonging to several families in the peptidase classification. Family M20 is Glutamate carboxypeptidases. Peptidase family M25 contains X-His dipeptidases.

Table.2: Mentioning the types of conserved domain involved in formation of GPI anchoring protein and below there is a functional description of each domains

Protein	Organism	Domains	
		Specific	Super family
GPI 12p / PIG - L	Archea	COG2120; PIG-L	PIG-L
	Bacteria	COG2120; PIG-L	PIG-L
	Eukaryote	COG2120; PIG-L	PIG-L

Domain Function Specification :

Archea, Bacteria and Eukaryote:

***COG2120*:** Uncharacterized proteins, LmbE homologs [Function unknown].

PIG-L: GlcNAc-PI de-N-acetylase; Members of this family are related to PIG-L an N-acetyl glucosaminyl phosphatidylinositol de-N-acetylase (EC: 3.5.1.89) that catalyzes the second step in GPI biosynthesis.

Table 3: Showing functional patterns of enzymes involved in GPI anchoring synthesis.

Organism	Motifs identified for Enzymes involved in GPI anchor synthesis				
	Active site	Metal Binding site	Phosphorylation site		N-glycosylation sites
			Protein Kinase C	Casein kinase II	
Archea	H	C	Phospho serine S Phospho Threonine T	Phospho serine S Phospho Threonine T	NATL, NLSS, NNSF , NRSS, NGSA, NTSA, NISS, NSSD, NKSE, NASG, NVTQ, NTTL, NGTQ, NGTE, NASR, NVSE, NATK, NKSY, NGTK, NTTD, NVTN, NDTR, NGTS, NGTE, NETY, NNSA , NATV, NGSI NNTT, NTTG, NGTE, NGTI, NGTV, NFTA, NASS, NDSS, NISG, NRTI, NWTT, NLTQ, NGTQ, NSTI, NSTG, NHTL, NYTV, NKST, NITH, NKSS, NMSR, NVTH, NGTE, NISI, NDSE, NRSN, NYSI, NLTD, NASA, NITD, NITG, NFTT, NLTG, NGTA, NWSQ, NISL, NLSR, NWST, NETG, NATR, NRTC, NNSH, NVTM, NITQ,NATG, NLSA, NSSQ, NNSI, NSSA, NVSN, NLTL, NGTD, NDSY
Bacteria	H	C	Phospho serine S Phospho Threonine T	Phospho serine S Phospho Threonine T	
Eukaryote	H	C	Phospho serine S Phospho Threonine T	Phospho serine S Phospho Threonine T	NCSN, NSSW, NCTT, NASS, NLSR, NLSI, NYSR, NWSG, NMSE, NLTV, NMIN, NLTI, NMTN, NNSN, NITI, NYSE, NKTL, NKTG, NLTG, NLSV, NQSE, NVSR, NMSL, NVSA, NRSW, NLSS, NMTN, NSSN, NVSY, NDSE, NHTE, NLST, NLSM, NLSQ, NLSE, NSSR, NGSG, NLSI, NSTE, NGTY, NGTM, NYTE, NMSN, NTTE, NRSF, NTSE, NRSY, NLTL, NYSD, NSSE, NLTV, NSTK, NATG, NLTE, NISA, NLSK, NVSV, NASD

Table.4: Showing functional patterns for Gpi12p/PIG-L proteins.

Organism	Motifs identified for Gpi12p/ PIG-L		
	Phosphorylation site		
	Protein Kinase C	Casein Kinase II	N-glycosylation sites
Archea	Phospho serine S	Phospho serine S	NRSD, NFSN, NDTA, NDSI, NTTK, NDTH, NVST, NLTF, NGSF, NSSS, NLST, NRSY, NASL, NASI, NTSS
	Phospho Threonine T	Phospho Threonine T	
Bacteria	Phospho serine S	Phospho serine S	NLSY, NESV, NASQ, NISE, NESI, NYSE, NGTD, NLSL, NAST, NSSD, NSST, NVSA, NETK, NLSF, NISM, NFTV, NSSM, NLTR, NASK, NATH, NTSM, NASN, NRSD, NFSS, NFTD, NITD, NNSI, NESH, NITN, NATL, NVTK, NASL, NETA, NRTV, NISI, NFSQ
	Phospho Threonine T	Phospho Threonine T	
Eukaryote	Phospho serine S	Phospho serine S	NGTD, NATD, NVTL, NASV, NHTA, NVTV, NHSS, NSSN, NISV, NATH, NWSV, NQSD, NLTV, NFTK, NHTN, NDTA, NHSA, NQSD, NVTI, NLSS, NLSL, NATN, NLSD, NDTS, NLSD, NSSG, NLTI, NQTL, NDTD, NSSL, NNTS, NTSD, NLTR, NYTK, NNTR, NDTA, NRTW, NQSQ, NLTV, NFTK, NHTN, NVSE, NQTQ, NETS, NVTI, NSTF, NIST, NFTA, NYSG, NASN, NSTL, NKSS, NRTT, NSTQ, NMSS, NFTQ, NNTN, NKTT, NNTS, NTSI, NGTM, NFSK, NASL, NESA
	Phospho Threonine T	Phospho Threonine T	

Table.5: Table explaining about ER localization signal sequence retention percentage in archea, bacteria and eukaryote species.

Organism	Name	Len	mTP	SP	other	Loc	RC	TPlen
Archea								
	Methanosarcina barkeri	302	0.079	0.865	0.085	S	2	25
Bacteria								
	marine metagenome	228	0.067	0.742	0.193	S	3	15
	Arthrobacter aurescens	339	0.432	0.479	0.214	S	5	116
Eukaryote								
	Homo sapiens	252	0.024	0.983	0.030	S	1	17
	Pan troglodytes	252	0.018	0.984	0.038	S	1	17
	Nomascus leucogenys	252	0.026	0.981	0.033	S	1	15
	Callithrix jacchus	249	0.025	0.967	0.061	S	1	21
	Ailuropoda - melanoleuca	252	0.016	0.984	0.057	S	1	17
	Canis familiaris	362	0.799	0.008	0.379	M	3	26
	Mus musculus	252	0.113	0.931	0.018	S	1	24
	Rattus norvegicus	252	0.095	0.936	0.022	S	1	24
	Bos taurus	253	0.011	0.989	0.041	S	1	24
	Gallus gallus	268	0.324	0.932	0.004	S	2	25
	Xenopus tropicalis	259	0.070	0.942	0.034	S	1	22
	Populus trichocarpa	265	0.061	0.945	0.019	S	1	24
	Podospora anserinas	293	0.103	0.978	0.006	S	1	36
	Schizosaccharomyces japonicus	253	0.035	0.963	0.051	S	1	18
	Cryptococcus gattii	292	0.075	0.910	0.058	S	1	38

As in the table 5, all eukaryote species having signaling peptide (SP) in their sequence and also exhibiting mitochondrial targeting peptides (mTP) which are essential in GPI anchoring process. Length of the sequence (Len) and Locus (Loc) means the sequence having signal or mitochondrial region.

CONCLUSION

In the current study, we identified that there is vast difference in protein functions among the selected organism based up on their motif pattern prediction within the same domain. Although, eukaryote showing slight similarity with archea and bacteria through motif analysis. Even though domains are not related with

other organisms. According to the GPI-attachment signal sequence, TargetP explained that signal sequence conserved in most eukaryotic systems only. Within the eukaryote organism, Sacchromysis cervasia resembling similar functions with Homo sapiens.

REFERENCES

1. Ferguson M. A, Williams A. F, Annu. Rev. Biochem, 1988, 57, 285±320
2. Kinoshita T, Inoue N. and Takeda J, Adv. Immunol, 1995, 60, 57±103
3. Caro L H. P, Tettelin H, Vossen J. H, Ram A. F. J, van den End, H, Klis F. M. Yeast13, 1997, 1477±1489.
4. Takos A. M., Dry I. B, Soole, K. L, FEBS Lett, 1997,405, 1±4

5. McConville M. J, and Ferguson M. A. J, Biochem. J.1993, 294, 305±324

6. Englund P. T, Annu. Rev. Biochem.1993, 62, 121±138

7. Chatterjee S, Mayor S: The GPI-anchor and protein sorting, Cell Mol Life Sci 2001, 58, 1969-1987.

8. Caras I.W, Weddell G.N, Etal, Signal for attachment of a phospholipid membrane anchor in decay accelerating factor, Science 238 (1987) 1280–1283.

9. Eisenhaber B, Bork P, Eisenhaber F: Prediction of potential GPI-modification sites in proprotein sequences. J Mol Biol, 1999, 292,741-758.

10. Marchler-Bauer A et al, CDD, a Conserved Domain Database for the functional annotation of proteins, 2011, Nucleic Acids Res.39 (D), 225-9.

11. Marchler-Bauer A et al, CDD, specific functional annotation with the Conserved Domain Database, Nucleic Acids Res, 2009, 37(D), 205-10.

12. Marchler-Bauer A, Bryant SH, CD-Search, protein domain annotations on the fly., Nucleic Acids Res, 2004, 32(W),327-331.

13. Sigrist CJA, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L, Xenarios I. New and continuing developments at PROSITE,Nucleic Acids Res. 2012.

14. Sigrist CJA, Cerutti L, Hulo N, Gattiker A, Falquet L, Pagni M, Bairoch A, Bucher P, PROSITE: a documented database using patterns and profiles as motif descriptors, Brief Bioinform,2002, 3,265-274.

15. De Castro E, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, Hulo N, ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins.Nucleic Acids Res, 2006, 1,34(W),362-5.

16. Sigrist C.J.A., De Castro E., Langendijk-Genevaux P.S., Le Saux V., Bairoch A., Hulo N, ProRule: a new database containing functional and structural information on PROSITE profiles, Bioinformatics, 2005 Nov 1, 21,4060-6.

17. Rapoport, T. A, Transport of proteins across the endoplasmic reticulum membrane, Science, 1992, 258, 931-936.

18. Von Heijne, G, The signal peptide. J. Membr Biol, 1990, 115, 195-201.

19. Von Heijne, G, Patterns of amino acids near signal sequence cleavage sites, Eur. J. Biochem, 1983,133,17-21.

20. Von Heijne, G, Signal sequences, the limits of variation. J. Mol. Biol, 1985, 184, 99-105.

21. Gavel Y, Nilsson L. & von Heijne G, Mitochondrial targeting sequences, Why non-amphiphilic peptides may still be amphiphilic, FEBS Letters, 1988, 235,173-177.

22. Roise, D, Recognition and binding of mitochondrial presequences during the import of proteins into mitochondria, J. Bioenerg. Biomembr, 1997, 29, 19-27.

23. Waltner M. & Weiner, H, Conversion of a nonprocessed mitochondrial precursor protein into one that is processed by the mitochondrial processing peptidase, J. Biol. Chem, 1996, 271, 21226-21230.

24. Kalousek, F., Hendrick, J. P. & Rosenberg, L. E, Two mitochondrial matrix proteases act sequentially Predicting Subcellular Localization of Proteins 1015 in the processing of mammalian matrix enzymes, Proc. Natl Acad. Sci. USA, 1988, 85, 7536-7540.

25. Isaya, G. & Kalousek, F, Mitochondrial intermediate peptidase. In Signal Peptidases von Heijne, G., ed.,1988, pp. 87-103, R.G. Landes Company, Austin.

26. Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G, Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites, Protein Eng, 1997, 10, 1-6.

27. Eisenhaber B, Maurer-Stroh S, Novatchkova M, Schneider G, Eisenhaber F, Enzymes and auxiliary factors for GPI lipid anchor biosynthesis and post-translational transfer to proteins, Bioassays, 2003, 25(4),367-85.

Source of support: Nil
Conflict of interest: None Declared