

## Homology Modeling and Functional Analysis of LPG2 Protein of *Leishmania* Strains

Ganesh Chandra Sahoo<sup>1\*</sup>, Manas Ranjan Dikhit<sup>1</sup>, Mukta Rani<sup>1</sup>, Pradeep Das<sup>1</sup>

<sup>1</sup>Rajendra Memorial Research Institute of Medical Sciences,  
Agam Kuan, Patna, India- 80007

\*Corresponding author: Ganesh Chandra Sahoo, Rajendra Memorial Research Institute of  
Medical Sciences, Patna, Bihar, INDIA, E-mail: ganeshiitkgp@gmail.com

Received November 27, 2008; Accepted January 06, 2009; Published January 10, 2009

**Citation:** Ganesh CS, Manas RD, Mukta R, Pradeep D (2009) Homology Modeling and Functional Analysis of LPG2 Protein of *Leishmania* Strains. J Proteomics Bioinform 0: 032-050. doi:10.4172/jpb.1000059

**Copyright:** © 2009 Ganesh CS, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### Abstract

As drug resistance problem persists in case of Leishmaniasis, modeling and analysis of different essential proteins of *Leishmania* strains will help us further to discover novel lead compounds. Lipophosphoglycan 2 (LPG2) protein is required for the development of *Leishmania* throughout their life cycle, including for virulence to the mammalian host. LPG2 participates in a specialized virulence pathway, which may offer an attractive target for chemotherapy. Homology models of LPG2 of five *Leishmania* species have been constructed using the X-ray structures of different transporter proteins as templates, by comparative protein modeling principles. The resulting model has the correct stereochemistry as gauged from the Ramachandran plot and good three-dimensional (3-D) structure compatibility as assessed by the Procheck and Profiles-3D scores. Functional assignment of LPG2 protein of *Leishmania* strains by SVM revealed that along with transporters activity it also performs several novel functions e.g. iron-binding, sodium-binding, copper binding. It also belongs to protein of major facilitator family (MFS) and type II (general) secretory pathway (IISP) family. Important functional motifs have been identified in LPG2 protein of different *Leishmania* strains using different programs. Potential Ligand Binding Sites (LBSs) in LPG2 protein of these strains have been identified using Pocket Finder program. On the basis of structure of ligand binding sites, particular LPG2 inhibitors can be designed. The similarity in the molecular structure, function and differences in LBSs of LPG2 of *L. donovani*, *L. major*, *L. infantum*, *L. braziliensis* and *L. mexicana* provide evidences for selective and specific LPG2 inhibitors.

**Keywords:** LPG2 protein; Leishmaniasis; Comparative (homology) modeling; Phyre (Protein Homology/analogy Recognition Engine); SVM (Support Vector Machine); Ligand binding Sites (LBSs)

### Background

*Leishmaniasis* is identified by clinical syndromes caused by obligate intracellular protozoa of the genus *Leishmania* and transmitted from one host to another by the bite of blood sucking sand fly vectors. Visceral leishmaniasis also known as Kala-Azar (KA) is caused by *Leishmania donovani* and is fatal if it remains untreated (Bhattacharyya et al., 2002). It is typically a vector-borne zoonosis, with rodents as common reservoir hosts and humans as second-

ary hosts. Visceral leishmaniasis (VL), the most severe form (which is usually fatal if patients are untreated), which is due to *Leishmania donovani*, is common in less developed countries (Paris et al., 2004). *Leishmania* is endemic in large parts of the world with 600,000 new clinical cases reported annually and possibly more unreported (Vergnes et al., 2007).

A short list of drugs includes SAG (Sodium Antimony Gluconate), amphotericin-B, pentamidine, and the oral drug

miltefosine, which is in phase IV clinical trial in India (Bihar, Patna). Already a decrease in efficacy has been noted against this novel molecule (Croft et al., 2006). A comparative analysis of a genetically related pair of Sb (V)-sensitive and -resistant *Leishmania donovani* strains isolated from kala-azar patients revealed that the resistant isolate exhibited cross-resistance to other unrelated *Leishmania* drugs including miltefosine and amphotericin-B (Vergnes et al., 2007).

Lipophosphoglycan (LPG) is the major cell surface molecule of promastigotes of all *Leishmania* species. It is comprised of three domains i.e. a conserved GPI anchor linked to a repeating phosphorylated disaccharide (P2; PO<sub>4</sub>-6-Gal (β1-4) Man (α1- ) backbone variously substituted with galactose, glucose and arabinose residues in *L. major* and capped with a neutral oligosaccharide (Ng et al., 1994). The main surface glycoconjugate on promastigotes, lipophosphoglycan (LPG), is crucial for parasite survival (Winberg et al., 2007).

LPG2 encodes a 37 KDa protein of 341 amino acids, containing up to 10 transmembrane domains (Descoteaux et al., 1995; Ma et al., 2004). LPG2 is a member of a growing family of genes implicated in nucleotide-sugar transport. The family is large, covers several nucleotide-sugar specificities and is evolutionarily diverse including *Leishmania*, yeast, *C. elegans*, plants and humans (Ma et al., 2004). Because of its hydrophobicity, subcellular location, and similarity to other proteins implicated in transmembrane transport, LPG2 protein is golgi GDP-mannose transporter required for addition of disaccharide-phosphate units on lipophosphoglycan and related glycoconjugates (Descoteaux et al., 1995; Ma et al., 2004). The amino acid sequence of SQV-7 protein of *C. elegans* and *Leishmania donovani* protein, LPG2 are similar to each other (67 (20%) are identical), which is required for transport of GDP-mannose across membranes (Descoteaux et al., 1995; Ma et al., 2004). Such transporters are required to bring nucleotide sugars from the cytosol, where they are synthesized, into the endoplasmic reticulum and Golgi apparatus, where they are used as sugar-donor substrates by glycosyltransferases (Abeijon et al., 1997). The hydropathy plots of SQV-7 and LPG2 are highly similar (Kyte and Doolittle, 1982). Human cells have no detectable GDPmannose transport activity, yet there are at least two human proteins similar to LPG2; thus, it is likely that LPG2, SQV-7, and the two human proteins are members of a family of transporters that have a variety of nucleotide-sugar specificities (Ma et al., 2004).

For virulence and transmission, the protozoan parasite *Leishmania* assembles a complex glycolipid on the cell sur-

face, the lipophosphoglycan (LPG). Functional complementation identified the gene LPG2, which encodes an integral golgi membrane protein implicated in intracellular compartmentalization of LPG biosynthesis. Ipg2- mutants lack only characteristic disaccharide-phosphate repeats, normally present on both LPG and other surface or secreted molecules considered critical for infectivity. In contrast, a related yeast gene, VAN2/VRG4, is essential and required for general golgi function. These results suggest that LPG2 participates in a specialized virulence pathway, which may offer an attractive target for chemotherapy (Descoteaux et al., 1995).

Epitope tagging experiments localized the LPG2 protein to the parasite's golgi apparatus, with the C-terminus located on the luminal side (Descoteaux et al., 1995). Transient transfection of LPG2 expression constructs, suggested that LPG2 acts autonomously as the GDP-Man transporter. It is reported that LPG2 occurs in a hexameric complex in *Leishmania* and also showed that GDP-Man, GDP-Ara, and GDP-Fuc can be transported by this NST. These findings have important implications to the structure and function of the NST family in both *Leishmania* and other eukaryotes (Hong et al., 2000).

Entry of *Leishmania* into visceral organs can cause damage to visceral organs and neurodegenerative symptoms are likely to occur. Hence there is requirement of study on structural and functional characteristics of different proteins of *Leishmania* strains to target the protein to find a suitable anti-leishmanial drug. X-ray crystallographic structure is not available for this important protein of *Leishmania* species. Modeling of the LPG2 protein, assigning function to this protein, identifying different ligand binding sites will give us useful information regarding LPG2 protein.

## Methodology

### Structural modeling

The sequence of LPG2 protein (341 amino acids) of *Leishmania donovani* was downloaded for structural modeling from NCBI. Multiple alignments of the related sequences were performed using the online available ClustalW program accessible through the European Bioinformatics Institute (Thompson et al., 1994; <http://www.ebi.ac.uk/Tools/clustalw2/index.html>). No X-ray crystallographic or NMR structure of this protein of any *Leishmania* species has yet been determined. Tertiary structures of LPG2 protein of different *Leishmania* strains were modeled on the basis of different template structures from PHYRE, I-Tasser. Structure validation was performed using ANOLEA, Profiles-

3D, WHATIF, and Model-3D, molecular modeling tool (Profiles-3D) of discovery studio.

### Transmembrane region prediction

Different servers i.e. TMHMM, SOSUI, HMMTOP and TMpred servers were accessed to validate the TM region of LPG2 protein (Krogh et al., 2001; Hirokawa et al., 1998; Tusnády and Simon, 1998; Hofmann and Stoffel, 1993). TMHMM, a new membrane protein topology prediction method, is based on a hidden Markov model.

### Ligand binding site prediction

Pocket-Finder is a pocket detection algorithm based on Ligsite written by Hendlich et al. (1997). Pocket-Finder works by scanning a probe radius  $1.6\text{\AA}$  along all gridlines of grid resolution  $0.9\text{\AA}$  surrounding the protein. The probe also scans cubic diagonals. Grid points are defined to be part of a site when the probe is within range of protein atoms followed by free space followed by protein atoms. Grid points are only retained if they are defined to be part of a site at least five times (Hendlich et al., 1997).

### Protein function assignment of LPG2 protein by SVM

To know novel functions of LPG2 protein of different *Leishmania* strains were searched at BIDD (Cai et al., 2003; <http://jing.cz3.nus.edu.sg/cgi-bin/svmprot.cgi>). The web-based software, SVMProt, support vector machine (SVM) classifies a protein into functional families from its

primary sequence based on physico-chemical properties of amino acids (Cai et al., 2003). Novel protein function assignment of different proteins of SARS virus and *Japanese encephalitis* virus has already been reported (Cai et al., 2005; Sahoo et al., 2008).

### ELM (eukaryotic linear motif) server

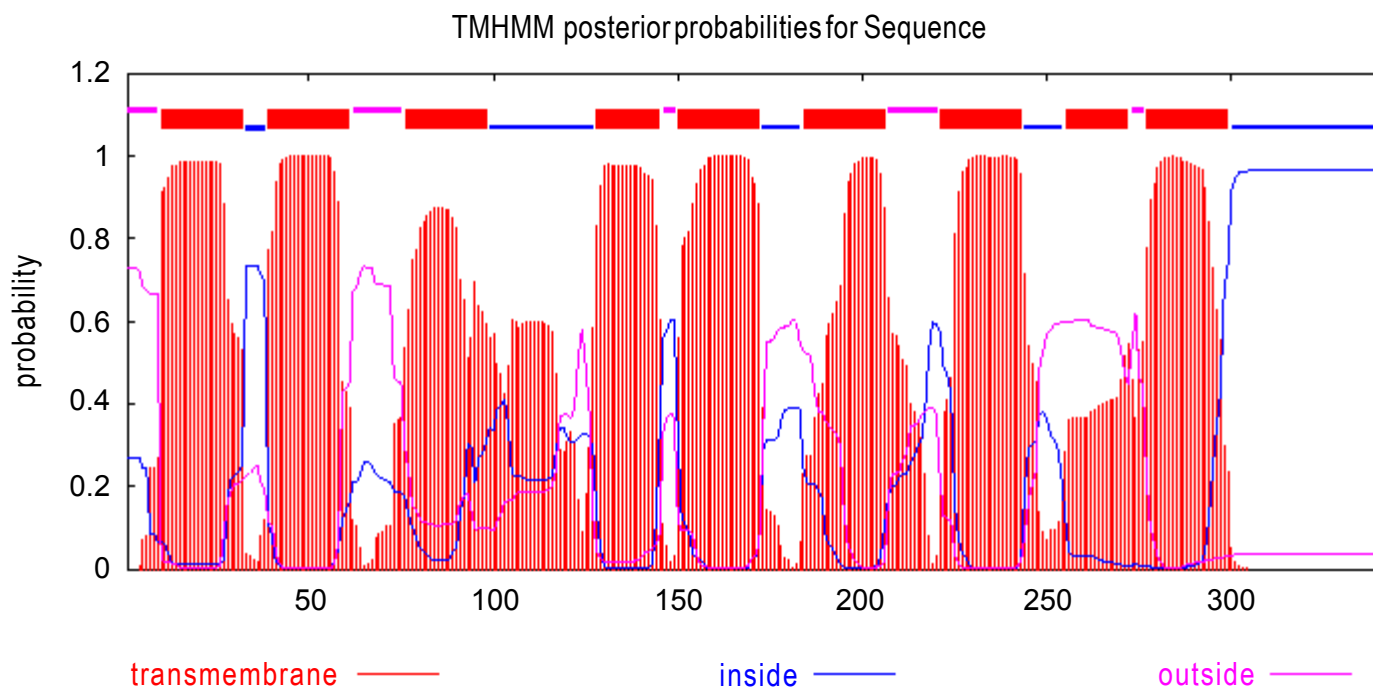
Functional sites in eukaryotic proteins which fit to the description “linear motif” are currently specified as patterns using regular expression rules. ELM server provides core functionality including filtering by cell compartment, phylogeny, globular domain clash (using the SMART/ Pfam databases) and structure (Puntervoll et al., 2003). Individual functions assigned to different sequence segments combine to create a complex function for the whole protein.

### Predict protein server

Predict Protein provides PROSITE sequence motifs, low-complexity regions (SEG), nuclear localization signals, regions lacking regular structure (NORS) and predictions of secondary structure, solvent accessibility, globular regions, transmembrane helices, coiled-coil regions, structural switch regions, disulfide-bonds, sub-cellular localization, and functional annotations (Puntervoll et al., 2003; Rost et al., 2004; Bairoch et al., 1997; Ceroni et al., 2004).

## Result and Discussion

Structure, function and ligand binding site analysis of LPG2



**Figure 1:** Involvement of different amino acids in formation of nine trans-membrane regions of LPG2 protein of different *Leishmania* strains.

SeqA Name	Len(aa)	SeqB Name	Len(aa)	Score
1 LPG2_Ldonovani_promastigote_Ld	341	2 LPG2_Lmaxicana_MNYC/BZ/62/M379	341	94
1 LPG2_Ldonovani_promastigote_Ld	341	3 LPG2_Lmajor_Friedlin	341	94
1 LPG2_Ldonovani_promastigote_Ld	341	4 LPG2_Linfantum	341	99
1 LPG2_Ldonovani_promastigote_Ld	341	5 Lbraziliensis_MHOM/BR/75/M2904	341	78
2 LPG2_Lmaxicana_MNYC/BZ/62/M379	341	3 LPG2_Lmajor_Friedlin	341	89
2 LPG2_Lmaxicana_MNYC/BZ/62/M379	341	4 LPG2_Linfantum	341	94
2 LPG2_Lmaxicana_MNYC/BZ/62/M379	341	5 Lbraziliensis_MHOM/BR/75/M2904	341	78
3 LPG2_Lmajor_Friedlin	341	4 LPG2_Linfantum	341	93
3 LPG2_Lmajor_Friedlin	341	5 Lbraziliensis_MHOM/BR/75/M2904	341	78
4 LPG2_Linfantum	341	5 Lbraziliensis_MHOM/BR/75/M2904	341	78

**Table 1:** Multiple alignment scores of LPG2 protein of different *Leishmania* strains.

protein will lead to identification of novel targets for design of suitable lead compounds inhibiting the specific functions of *L. donovani*, *L. major* and *L. infantum*.

### Transmembrane region prediction

Different servers have been accessed for accurate prediction analysis of transmembrane region e.g. TMHMM, HMMTOP, SOSUI and TMPred. Trans-membrane prediction analysis found that LPG2 protein of *L. infantum* and *L. donovani* are having same number of transmembrane regions (ten) and involvement of particular amino acids in the TM regions is similar. The LPG2 sequences of *L. infantum* and *L. donovani* are very similar to each other except at two positions 220 and 221 where threonine (T) is replaced by isoleucine (I) and methionine (M) respectively in *L. infantum*. Even if the number of transmembrane region are same (ten) in *L. mexicana*, *L. donovani* and *L. infantum*, it is found that 3<sup>rd</sup> and 4<sup>th</sup> transmembrane regions are different (3<sup>rd</sup> TM region comprises of aa 72-94 and 4<sup>th</sup> TM comprises 98-120 amino acids) in *L. mexicana* whereas in *L. infantum* and *L. donovani* it is coded by 77-99 and 101-123 amino acids. LPG2 of *L. braziliensis* and *L. major* both comprise of nine TM regions but the transmembrane regions are quite different, only in few cases these are similar. Involvement of different amino acids in formation of nine different trans-membrane regions of different *Leishmania* strains is shown in (Figure1).

Multiple alignment of amino acid sequences of LPG2 protein of different *Leishmania* strains shows that they are very close to each other ranging from 78-99% (Table 1). Multiple alignment of LPG2 protein shows that *Ldv* and *L. infantum* are having 99% identity (Figure 2). From phylogram, LPG2 protein of *L. braziliensis* is found to be far from other *Leishmania* strains (Figure 3). Many amino acid changes are found to be present towards the carboxy terminus in the

annotated amino acid sequences.

### Structure analysis of LPG2 protein

Five different models of LPG2 protein of different *Leishmania* strains were screened for profiles-3d score of DS (Accelrys). The best model of different *Leishmania* strains were screened by profiles-3d score and the best was selected for further analysis. About 14 - 27 helices have been predicted for LPG2 protein of different models of various *Leishmania* strains. The best model of LPG2 protein of *L. infantum* contains 22 helices, whereas 19 helices are there in case of *L. donovani* and *L. major*. The models for LPG2 protein of different *Leishmania* strains are shown in fig 4. The profiles-3D scores of best predicted models of LPG2 protein of various *Leishmania* strains (Table 2), shows that highest score (111.29) has been found in case of LPG2 protein model of *L. major* and lowest score (91.23) has been found in case of *L. braziliensis*. Invalid regions have been detected in different models. Highest numbers of invalid regions have been found in one model of *L. mexicana* whereas lowest number of invalid regions is found in other models of *L. mexicana*, *L. major* and *L. donovani*. On further side chain and loop modeling refinements of the predicted models of LPG2 protein, the profiles-3D score is found to be decreased. In Ramachandran plots (Procheck), 88-91% residues belong to core region, 6-8 % residue in allowed region, 1-2.5% in generously allowed regions and 1-3% in disallowed regions (Table 3). From Ramachandran plot, it is known that maximum residues in LPG2 protein are responsible for formation of helices. Also transmembrane region prediction analysis detected nine or ten transmembrane regions to be present in LPG2 protein of all these strains. Hence there is involvement of helices in formation of LPG2 protein.

The models of different *Leishmania* strains were based

CLUSTAL 2.0.8 multiple sequence alignment

```

LPG2_Ldonovani_promastigote_Ld      MNHTRSVMEAVLAVITYSFCVSMILVNKLIMNTYDMNFFPGILVLQTGG 50
LPG2_Linfantum                      MNHTRSVMEAVLAVITYSFCVSMILVNKLIMNTYDMNFFPGILVLQTGG 50
LPG2_Lmaxicana_MNYC/BZ/62/M379     MNHTRSVMEAVLAVITYSFCVSMILVNKLIMNTYDMNFFPGILVLQTGG 50
LPG2_Lmajor_Friedlin               MNHTRAVMEAILAVITYSFCVSMILVNKLIMNTYDMNFFPGILVLQTGG 50
Lbraziliensis_MHOM/BR/75/M2904     MNHSHIVMEAAALSVMAYSFCSVSMILVNKLIMNTYDMNFFPGILLQTGG 50
***:: **** *:***:*****:*****:*****:*****

LPG2_Ldonovani_promastigote_Ld      ALVIVALAKAARFIEYPAFSDVAKKWLPLTLLFVAMLFTSMKSLGTMSV 100
LPG2_Linfantum                      ALVIVALAKAARFIEYPAFSDVAKKWLPLTLLFVAMLFTSMKSLGTMSV 100
LPG2_Lmaxicana_MNYC/BZ/62/M379     AVVIVALAKAARFIDYPAFSDVAKQWLPLTLLFVAMLFTSMKSLGTMSV 100
LPG2_Lmajor_Friedlin               ALVIVALAKAARFVEYPAFSDVAKKWLPLTLLFVAMLFTSMKSLATMSV 100
Lbraziliensis_MHOM/BR/75/M2904     ALMIVTLAKATRFIDYSAFSLDVAKQWLPLTLLFVSMSTSMKSLGMSV 100
*::**::****:***:*.***:****:*****:*** *****:***

LPG2_Ldonovani_promastigote_Ld      AAQTIKKNLAVVLIALGDKFLYGKAQTPMVYFSAFALMILGSLGAKGDKW 150
LPG2_Linfantum                      AAQTIKKNLAVVLIALGDKFLYGKAQTPMVYFSAFALMILGSLGAKGDKW 150
LPG2_Lmaxicana_MNYC/BZ/62/M379     AAQTIKKNLAVVLIALGDRFLYGKAQTPMVYFSAFALMILGSLGAKGDRW 150
LPG2_Lmajor_Friedlin               AAQTIKKNLAVVILIALGDKFLYGKAQTPMVYFSAFALMILGSLGAKGDKW 150
Lbraziliensis_MHOM/BR/75/M2904     AAQTIKKNLAVVFIALGDKFLYGKAQTPSVYVSAFALMFFGSYLGAAGDKW 150
*****:*****:****:*****:***** ** *****:*** *****:

LPG2_Ldonovani_promastigote_Ld      VTAWGLVWTFNLNIVSTVSYTLYMKAVLGSVSNISIGRYGPFVFNLLSLPF 200
LPG2_Linfantum                      VTAWGLVWTFNLNIVSTVSYTLYMKAVLGSVSNISIGRYGPFVFNLLSLPF 200
LPG2_Lmaxicana_MNYC/BZ/62/M379     VTAWGLVWTFNLNIVSTVSYTLYMKAVLGSVSNISIGRYGPFVFNLLSLPF 200
LPG2_Lmajor_Friedlin               VTAWGLIWTLLNIVSTVSYTLYMKAVLGSVSNISIGRYGPFVFNLLSLPF 200
Lbraziliensis_MHOM/BR/75/M2904     VTPWGLFWTFNLNIAATVLYTLYMKTMLGSLVSNISIGRYGPFVFNLLSLPF 200
**_***_*:***:*** ** *****:*****:*****:*****

LPG2_Ldonovani_promastigote_Ld      FLIMGVGDIMPFSAAIGDITTFGKLVLTFSVLVSSVMTFSVFWCMSITSP 250
LPG2_Linfantum                      FLIMGVGDIMPFSAAIGDITMFGKLVLTFSVLVSSVMTFSVFWCMSITSP 250
LPG2_Lmaxicana_MNYC/BZ/62/M379     FLIMGVGDIMPFSAAIGETITTFGRLLVLAFLVSVSSVMTFSVFWCMSITSP 250
LPG2_Lmajor_Friedlin               FLIMGVGEIMPFSAAIGETITTLGKLVLTFSVLVSSVMTFSVFWCMSITSP 250
Lbraziliensis_MHOM/BR/75/M2904     FLVMGVGEIMPFAATAISETTSLGKLVLFVSVLVSSVMTFSVFWCMSITSP 250
**::****:***:***:***:*** ** *****:*****:*****:*****

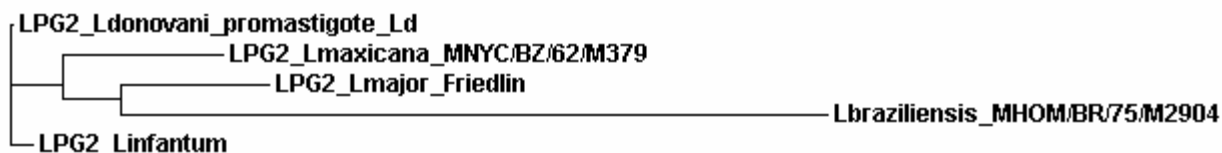
LPG2_Ldonovani_promastigote_Ld      FLIMGVGDIMPFSAAIGDITTFGKLVLTFSVLVSSVMTFSVFWCMSITSP 250
LPG2_Linfantum                      FLIMGVGDIMPFSAAIGDITMFGKLVLTFSVLVSSVMTFSVFWCMSITSP 250
LPG2_Lmaxicana_MNYC/BZ/62/M379     FLIMGVGDIMPFSAAIGETITTFGRLLVLAFLVSVSSVMTFSVFWCMSITSP 250
LPG2_Lmajor_Friedlin               FLIMGVGEIMPFSAAIGETITTLGKLVLTFSVLVSSVMTFSVFWCMSITSP 250
Lbraziliensis_MHOM/BR/75/M2904     FLVMGVGEIMPFAATAISETTSLGKLVLFVSVLVSSVMTFSVFWCMSITSP 250
**::****:***:***:***:*** ** *****:*****:*****:*****

LPG2_Ldonovani_promastigote_Ld      TTMSVVGSLNKIPLTFLGMLVFHQFPTATGYLGIMIALSAGFLYTHLNIR 300
LPG2_Linfantum                      TTMSVVGSLNKIPLTFLGMLVFHQFPTATGYLGIMIALSAGFLYTHLNIR 300
LPG2_Lmaxicana_MNYC/BZ/62/M379     TTMSVVGSLNKIPLTFLGMLVFHQFPTATGYLGILIALSAGFLYTHLNIR 300
LPG2_Lmajor_Friedlin               TTMSVVGSLNKIPLTFLGMLAFHQFPTATGYLGIMVALSAGFLYTHLNIR 300
Lbraziliensis_MHOM/BR/75/M2904     TTMSVIGSLNKIPLTFLGMLVFHQFPTATGYVGIVVALAAGFLYTHLNIT 300
*****:*****:*****:*****:*****:*****:*****:*****

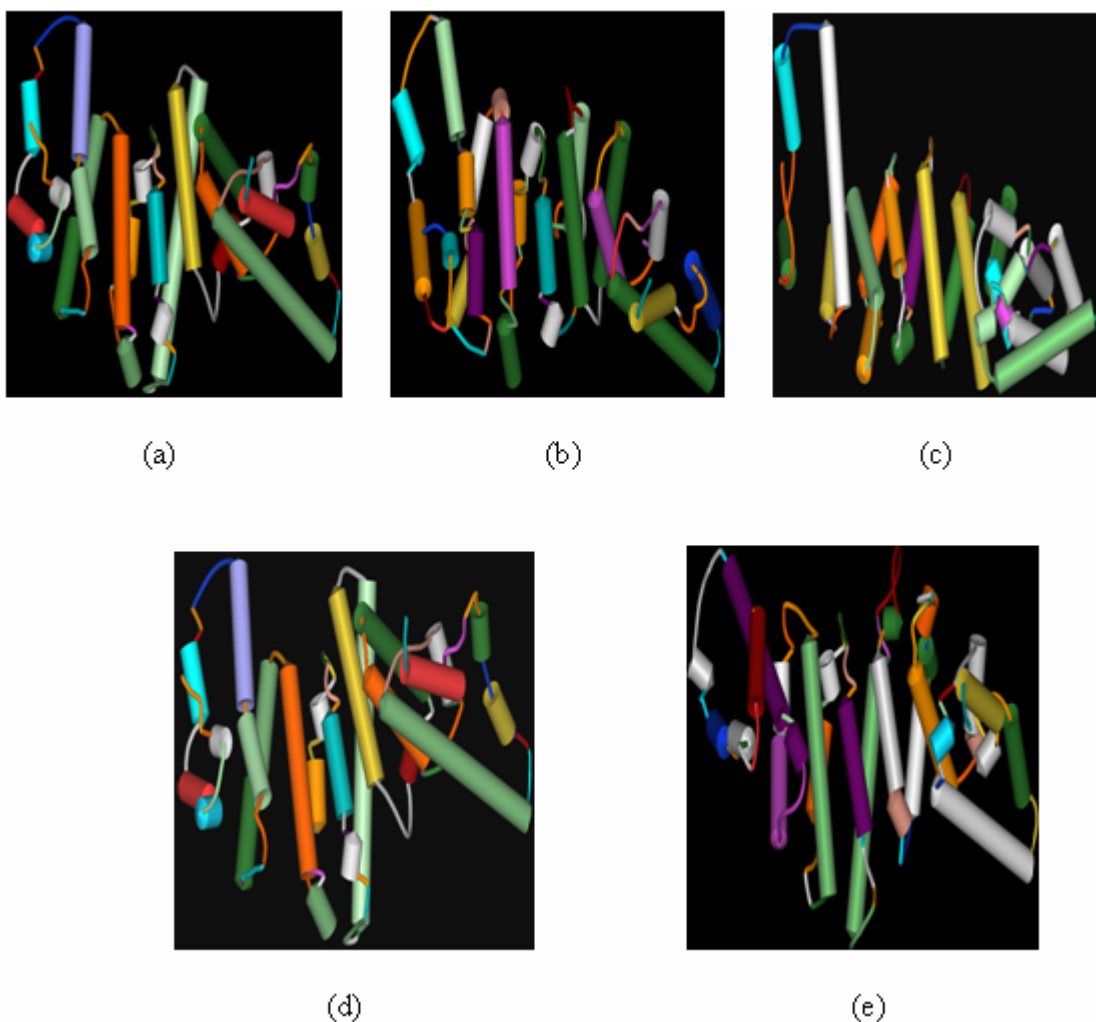
LPG2_Ldonovani_promastigote_Ld      ANRAKASSDTEHQMQQAGKTTAESIVLVRADENSNDTSKSE 341
LPG2_Linfantum                      ANRAKASSDTEHQMQQAGKTTAESIVLVRADENSNDTSKSE 341
LPG2_Lmaxicana_MNYC/BZ/62/M379     ANHAKASSDTEHQMQQAGKPSAESIVLVRADENSNDTSKSE 341
LPG2_Lmajor_Friedlin               ANRAKASSDLEHQMQQTGKTTTESIVLVRADENSNDTSKSE 341
Lbraziliensis_MHOM/BR/75/M2904     ANRAKASATDAENHRQQTMMSSPFAVLAPPEEKNIDTVKSV 341
**::***:*** ** *****:*****:*****:*****

```

Figure 2: Multiple alignment of amino acid sequences of LPG2 protein of different *Leishmania* strains.



**Figure 3:** Phylogram showing phylogenetic relationship of LPG2 protein of different *Leishmania* strains (*L. donovani*, *L. major*, *L. infantum*, *L. mexicana* and *L. braziliensis*).



**Figure 4:** Ribbon representations of the modeled LPG2 protein images of all different *Leishmania* strains using Discovery Studio 2.0 (Accelrys) software (a) *L. donovani*, (b) *L. major*, (c) *L. infantum* (d) *L. mexicana* and (e) *L. braziliensis*.

on template PDB coordinates of 2i68, 1ee4, 1pw4, 1ej1, 3b5d and 1xm9. The PDB '2i68' codes for transmembrane domain of the multi drug resistance antiporter from *E. coli* Emr E. This protein has antiporter activity and belongs to a family of plasma membrane proteins and proteins integral to membrane. The PDB '1pw4' codes for protein with transporter activity and belongs to MFS general substrate transporter fold and is involved in glycerol metabolic process and glycerol 3-phosphate transport. It is a protein of plasma

membrane and protein integral to membrane. The PDB '1xm9' codes for structure of armadillo repeat domain of plakophilin 1, belongs to all alpha proteins with alpha alpha superhelix fold which has ARM repeat forming plakophilin 1 domain. This protein is involved in molecular functions like signal transduction, cell adhesion and protein binding. The PDB '3b5d' codes for x-ray crystallographic structure of Emr E multidrug transporter in complex with TPP, involved in antiporter like transporter function. It is a plasma

Model Features Strain Names	Number and percentage of alpha helices	Number and percentage of 3 <sub>10</sub> (3 <sub>10</sub> ) helices	Number of chains	Profile 3-D scores
<i>L. major</i>	<b>17 / 68.9%</b> 7(min)-27(max) residues take part in formation of helices	<b>2 / 1.8%</b> 3 residues	<b>1</b>	<b>111.29/2 (Model 5)</b>
<i>L. mexicana</i>	<b>17 / 73%</b> 5(min)-28(max) residues take part in formation of helices	<b>1 / 0.9%</b> 3 residues	<b>1</b>	<b>101.8/2 (Model 5)</b>
<i>L. infantum</i>	<b>21 / 70.4%</b> <b>4(min)- 23(max)</b> residues take part in formation of helices	<b>1 / 0.9 %</b> 4 residues	<b>1</b>	<b>108.48/8 (Model 4)</b>
<i>L. braziliensis</i>	<b>19 / 69.2%</b> <b>4(min)- 24(max)</b> residues take part in formation of helices	<b>2 / 1.5%</b> 3 residues	<b>1</b>	<b>91.23/13 (Model 3)</b>
<i>L. donovani</i>	<b>24 / 68.3 %</b> <b>5(min)- 16(max)</b> residues take part in formation of helices	<b>3 / 2.3%</b> 3-4 residues	<b>1</b>	<b>101.8/2 (Model 5)</b>

**Table 2:** Promotif search result summary and Profiles-3D scores of modeled structures of LPG2 proteins of different *Leishmania* strains.

Residues	Number of Amino acids involved					Percentage of amino acids involved				
	Lbrzl	Lmjr	Linf	Lmx	Ldv	Lbrz	Lmjr	Linf	Lmx	Ldv
Residues in most favoured regions [A, B, L]	270	272	271	276	263	88.2	88.6	88.6	91.1	86.8
Residues in additional allowed regions[a,b,l,p]	28	25	20	19	28	9.2	8.1	6.5	6.3	9.2
Residues in generously allowed regions [~a,~b,~l,~p]	3	7	5	4	6	1	2.3	1.6	1.3	2
Residues in disallowed regions	5	3	10	4	6	1.6	1	3.3	1.3	2
Number of non-glycine and non-proline residues	306	307	306	303	303	100% for all strains				
Number of end-residues (excl. Gly and Pro)	2	2	2	2	2					
Number of glycine residues (shown as triangles)	21	22	23	25	25					
Number of proline residues	12	10	10	11	11					
Total number of residues	341	341	341	341	341					

[Abbreviations used in the Table:

Lbrzl → *Leishmania braziliensis*; Lmjr → *Leishmania major*;

Linf → *Leishmania infantum*; Lmx → *Leishmania mexicana*

Ldv → *Leishmania donovani*]

**Table 3:** Referring to Ramachandran Plots of LPG2 protein of five different strains of *Leishmania*.

membrane component. Few drugs have been targeted to this protein like dapson and tamoxifen. The PDB coordinates of '1ee4' code for crystal structure of yeast karyopherin (importin) in a complex with a c-myc NLS peptide. This karyophilin alpha protein have armadillo domain with ARM repeat. This protein is classified as transporter protein which is a transmembrane protein with transporter activity involved in import into nucleus, protein targeting to membrane and intracellular transport. This is a leucine rich alpha helical protein. The PDB coordinates of '1ej1' code for crystal structure of mRNA 5'cap binding protein (eIF 4E) bound to 7-methyl-GDP. This protein is having RNA binding and translation factor activity and belongs to class of alpha and beta proteins. From all these template analyses, it is found that except '1ej1', all other templates are transporter-like protein. LPG2 protein also belongs to this group of protein.

It is found that the best model of LPG2 protein of all these strains consisted of only one chain (Table 2). In all the models 68-73 percent is helical. The best model of LPG2 protein of *L. donovani* is having highest number (24) of helices where minimum five and maximum sixteen residues take part in formation of a helix. In *L. donovani* three  $3_{10}$  helices are present. In *L. mexicana* 28 residues has been found to be forming a helix and helices in this strain accounts for 73% of total residues.

**Functional assignment of LPG2 protein by SVM**

From the comparative analysis of LPG2 protein functional assignment of five *Leishmania* strains shows that it belongs to transporter group of proteins (Table 4). LPG2 protein also belonged to trans-membrane region protein. LPG2 protein of *L. mexicana* strain M379 belonging to copper

Functions from NCBI	Friedlin_ <i>L. major</i> _CAJ08033	<i>L. donovani</i> _ Sudanese1S_ AAC46914	MNYC/BZ/62/M379_ Belize country <i>L. mexicana</i>	MHOM/BR/75/M2904_ <i>L. braziliensis</i>	<i>L. infantum</i>
Nucleotide-sugar transporter	TC 2.A. Electrochemical Potential-driven transporters - Porters (uniporters, symporters, antiporters)	TC 2.A. Electrochemical Potential-driven transporters - Porters (uniporters, symporters, antiporters)	TC 2.A. Electrochemical Potential-driven transporters - Porters (uniporters, symporters, antiporters) (99.2%)	TC 2.A. Electrochemical Potential-driven transporters - Porters (uniporters, symporters, antiporters)	TC 2.A. Electrochemical Potential-driven transporters - Porters (uniporters, symporters, antiporters)
	Transmembrane	Transmembrane	Transmembrane	Transmembrane	Transmembrane
1.Posttranslational modification 2.Protein turnover			Sodium-binding (78.4 %)	TC 9.B. Incompletely Characterized Transport Systems - Putative uncharacterized transport proteins (73.8%)	
		Iron-binding (58.6%)	Iron-binding (73.8%)		Iron-binding
			Metal-binding	G Protein Coupled Receptors (58.6%)	
			TC 2.A.1 Major facilitator family (MFS) (58.6%)	TC 2.A.1 Major facilitator family (MFS) (58.6)	
chaperones / Intracellular trafficking and secretion			TC 3.A.5 Type II (general) secretory pathway (IISP) family		
			Copper binding (58.6%)		
			Magnesium binding (58.6%)		

**Table 4:** Comparative analysis of functional assignment of LPG2 protein in various *Leishmania* strains by SVMProt.



Results	<i>L. donovani</i>	<i>L. mexicana</i>	<i>L. major</i>	<i>L. infantum</i>	<i>L. braziliensis</i>
<b>ASN_GLYCOSYLATION</b> (N-glycosylation site)  N[ <sup>^</sup> P][ST][ <sup>^</sup> P]	<b>Pattern:</b> 2 → NHTR 335→ NDTs	<b>Pattern:</b> 2 → NHTR	<b>Pattern:</b> 2 → NHTR	<b>Pattern:</b> 2 → NHTR 335→ NDTs	<b>Pattern:</b> 2 → NHTR
<b>PKC_PHOSPHO_SITE</b> (Proteinkinase C) [ST].[RK]	<b>Pattern:</b> 91 → SMK 337 → TSK	<b>Pattern:</b> 91 → SMK 337 → TSK	<b>Pattern:</b> 91→SMK	<b>Pattern:</b> 91→SMK 337→TSK	<b>Pattern:</b> 91→SMK
<b>CK2_PHOSPHO_SITE</b> (Casein kinase II phosphorylation site) [ST].[2][DE]	<b>Pattern:</b> 6 → SVME 308 → SDTE 320 → TTAE 338 → SKSE	<b>Pattern:</b> 6 → SVME 338 → SKSE	<b>Pattern:</b> SDLE → 308	<b>Pattern:</b> 6→ SVME 308→SDTE 320→TTAE 328→SKSE	<b>Pattern:</b> TDAE → 308
<b>MYRISTYL</b> (N-myristoylation site) G[ <sup>^</sup> EDRKHPFYW].[2][STAGCN][ <sup>^</sup> P]	<b>Pattern:</b> 140 →GSSLGA 155 →GLVWTF 178 →GSVSNS 283 →GIMIAL	<b>Pattern:</b> 140→GSSLGA 155→GLVWTF 178→GSVSNS 283→GILIAL 336→GTSKSE	<b>Pattern:</b> 178→GSVSNS	<b>Pattern:</b> 140→GSSLGA 155→GLVWTF 283→GIMIAL	<b>Pattern:</b> 178→GSVSNS
<b>DISULFIND</b>	20-244 (length-224aa)	20-244 (length-224aa)	20-244 (length-224aa)	20-244 (length-224aa)	20-244 (length-224aa)
<b>Predicted secondary structure</b>	H→55.72% E →12.02% L → 32.26%	H→56.01% E→10.85% L→33.14%			
<b>Globularity</b>	<b>nexp</b> = 91 (number of predicted exposed residues)  <b>nfit</b> = 140 (number of expected exposed residues)  <b>diff</b> = - 49.00 (difference nexp-nfit) So , protein not be globular	<b>nexp</b> = 87 (number of predicted exposed residues)  <b>nfit</b> = 140 (number of expected exposed residues)  <b>diff</b> = - 53.00 (difference nexp-nfit) So , protein not be globular			

**Table 5:** It shows comparative analysis of different motifs of LPG2 protein of five *Leishmania* strains. The motifs were predicted by Predict Protein server.

binding (58.6%), magnesium binding (58.6%), metal-binding, iron-binding (73.8%) and sodium-binding (78.4%) protein function families. It also belongs to type II (general) secretory pathway (IISP) family and major facilitator family (MFS) (58.6%). LPG2 protein of *L. braziliensis* also belongs to incompletely characterized transport systems - putative uncharacterized transport proteins (73.8%), G protein coupled receptors (58.6%) and major facilitator family (MFS) (58.6%). From NCBI, it is also known that this protein is a nucleotide sugar transporter protein which is also likely to be post-translationally modified and belongs to chaperone or intracellular trafficking group of proteins.

Various post-translational modification sites for LPG2 protein of various *Leishmania* strains were identified using protein predict program. Two asparagine glycosylation sites (at 2<sup>nd</sup> amino acid [NHTR] and at 335<sup>th</sup> amino acid [NDTS]) have been found in *L. infantum* and *L. donovani* whereas only one glycosylation site at 2<sup>nd</sup> amino acid [NHTR] was found in other three *Leishmania* strains (Table 5). Two protein kinase C activation sites were predicted in three *Leishmania* strains and one PKC site was predicted in other two strains. Four casein kinase II phosphorylation (CKP) sites are present in two *Leishmania* strains whereas in other strains one or two CKPs are present. One to five numbers of N-myristoylation sites are present in different *Leishmania* strains. Only one disulfide bond between 20<sup>th</sup> and 244<sup>th</sup> amino acid residues was detected in this protein of different *Leishmania* strains (Table 5).

ELM server detected short functional sites in LPG2 protein. In *L. donovani*, *L. major* and *L. infantum*, phosphothreonine motif binding a subset of FHA domains having a preference for an acidic amino acid at the pT+3 position (Nucleus, Replication fork) (LIG\_FHA\_2) has been predicted at 318-324[GKTTAES] position. In *L. braziliensis* the predicted site and amino acids involved were different (Site 306-312 and sequence is SATDAEN) (Table 6). Since phosphothreonine motif of LPG2 protein of *L. braziliensis* is different, according to this motif, LPG2 protein of all other four *Leishmania* strains are homologous to each other which is also confirmed by Clustal W and phylogenetic analysis. From analogy point of view, all the five strains are equal as the phosphothreonine motif is found in all five *Leishmania* strains.

The MAP kinase (MAPK) cascades convey a signal in form of phosphorylation events. MAPKs are phosphorylated by MAP kinase kinases (MAPKKs), phosphorylate various targets, such as transcription factors and MAPK-activated protein kinases (MAPKAPKs), and are dephos-

phorylated and inactivated by several MAPK-phosphatases (MKPs) (Sturgill and Wu, 1991; Ahn et al., 1992; Nishida et al., 1993; Marshall et al., 1995). In LPG2 protein of three *Leishmania* strains, MAPK interacting molecules (e.g. MAPKKs, substrates, phosphatases) carrying docking motif helping to regulate specific interaction in the MAPK cascade, have been detected in 75- 82 residues (Table 6).

Amino acid sequence "GRLVLA" (223<sup>rd</sup> - 228<sup>th</sup>) of *L. mexicana*, was found to have an RXXL motif that binds to the Cdh1 and Cdc20 components of APC/C thereby targeting the protein for destruction in a cell cycle dependent manner (Table 6).

Cell cycle depends upon the well orchestrated activation and deactivation of cyclin-dependent kinases which phosphorylate a number of substrates required for entry into the next phase of the cell cycle (Takeda et al., 2001). Three substrate recognition sites have been identified in LPG2 protein of *Leishmania* strains that interact with cyclin and thereby increase phosphorylation by cyclin / cdk complexes which are required for cell cycle events. Predicted proteins also have the MOD\_CDK sites which are used by cyclin inhibitors (Table 6).

USP7 plays an important role in regulating cell proliferation and apoptosis through p53 and Mdm2 interactions (Saridakis et al., 2005). Motif containing amino acid sequence [AKASS-304<sup>th</sup> -308<sup>th</sup>] of LPG2 protein of four *Leishmania* strains is the USP7 NTD domain binding motif variant based on the MDM2 and P53 interactions, this site is absent in *Leishmania braziliensis* (Table 6).

WW domain is one of the domains mediating cellular processes which require physical interactions between proteins (Bedford et al., 1998). Two to four numbers of class IV WW domains interaction motifs (phosphorylation-dependent interaction motifs found in both nuclear and cytosolic proteins) are present in LPG2 protein of *Leishmania* strains (Table 6).

CK1 is a "phosphate-directed" protein kinase which is able to phosphorylate with high efficiency Ser/Thr residues specified by a prephosphorylated side chain (either pS or pT) at position n-3 (or less effectively n-4) (Meggio et al., 1979; Donella-Deana et al., 1985; Flotow et al., 1990). This observation led to the concept of "hierarchical phosphorylation" (Roach et al., 1991) and the term "primed phosphorylation" to indicate the ability to phosphorylate residue(s) specified by another phosphorylated residue at a predetermined (critical) position. This feature is shared by a small number of acidophilic Ser/Thr kinases, notably CK1 and

Identified functional /sequence domain	Amino acid sequence and number in the protein sequence of different strains of Leishmania					Comments about reported Domain and protein function
	<i>L. mexicana</i>	<i>L. donovani</i>	<i>L. major</i>	<i>L. infantum</i>	<i>L. braziliensis</i>	
<b>LIG_APCC_Dbox 1</b> (R.L.)	GRLVLA→223-228	-----	-----	-----	-----	An RXXL motif binds to either to Cdh1 / Cdc20 components of APC/C thereby targeting protein for destruction in a Cell cycle dependent manner. [Nucleus,cytosol]
<b>LIG_CYCLIN_1</b> ([RK].L.{0,1}[FYLVMP])	KNLAV→107-111 RFLY→119-122	KWLPL→76-80 KNLAV→107-111 RFLY→119-122	KWLPL→76-80 KNLAV→107-111 RFLY→119-122	KWLPL→76-80 KNLAV→107-111 RFLY→119-122	KNLAV→107-111 RFLY→119-122	Substrate recognition site, interacting with cyclin, thereby increasing phosphorylation by cyclin / cdk Complexes. Predicted protein also should have the MOD_CDK site. Also used by cyclin inhibitors [Nucleus,Cytosol]
<b>LIG_FHA_1</b> (..(T). [ILV].)	NHTRSVM→2-8 LGTMSVA→95-101 AQTPMVY→125-131  VMTFSVF→236-242 PTTMSV→250-256 TATGYLG→277-283	NHTRSVM→2-8 LGTMSVA→95-101 AQTPMVY→125-131  VLTFSVL→226-232 VMTFSVF→236-242 PTTMSV→250-256 TATGYLG→277-283	NHTRAVM→2-8 LATMSVA→95-101 AQTPMVY→125-131 GETTTLG →217-223 VLTFSVL→226-232 VMTFSVF→236-242 PTTMSV→250-256 TATGYLG→277-283 TTTESIV→320-326	NHTRSVM→2-8 LGTMSVA→95-101 AQTPMVY→125-131  VLTFSVL→226-232 VMTFSVF→236-242 PTTMSV→250-256 TATGYLG→277-283	KATRFID→59-65 AQTPSVY→125-131  SETTSLG→217-223  VMTFSVF→236-242 PTTMSVI→250-256 TATGYVG→277-283	Phosphothreonine motif binding a subset of FHA domains that have preference for an acidic amino acid at the Pt+3 position [Nucleus]
<b>LIG_FHA_2</b> (..(T). [DE].)	-----	GKTTAES →318-324	GKTTTES→318-324	GKTTAES→318-324	SATDAEN→306-312	Phosphothreonine motif binding a subset of FHA domains that have a preference for an acidic amino acid at the pT+3 positions. [Nucleus,Replication fork]
<b>LIG_MAPK_1</b> [[KR]{0,2}[KR].{0,2}[KR].{2,4}[ILVM].[ILVF]]	-----	KKWLPLTL→75-82	KKWLPLTL→75-82	KKWLPLTL→75-82	-----	Molecules interacting with MAPK (MAPKKs, substrates, phosphatases) carry docking motif that helps to regulate specific interaction in MAPK cascade. The classic motif approximates (R/K)xxx#x#, where # is a hydrophobic residue [Nucleus,Cytosol]

<b>LIG_USP7_1</b> ([PA][^P][^FYWILS [^P])	AKASS→ 304-308	AKASS→ 304-308	ATMSV→96-100 AKASS→ 304-308	AKASS →304-308	-----	The USP7 NTD domain binding motif variant based on the MDM2 and P53 interactions. [Nucleus]
<b>LIG_WW_4</b> (... [ST]P.)	KAQTPM→124-129 SITSPT→246-251	KAQTPM→124-129 SITSPT→246-251	KAQTPM→124-129 SITSPT→246-251	KAQTPM→124-129 SITSPT→246-251	KAQTPS→124-129 KWVTPW→149-154 SITSPT→246-251 MNSSPF→319-324	Class IV WW domains interaction motif; phosphorylation-dependent interaction [Nucleus, Cytosol]
<b>MOD_CK1_1</b> ([PA][^P][^FYWILS [^P])	SVITYSF→ 13-19 SFCSVSM→18-24  SMKSLGT→91-97 SLGTMSV→94-100 STVSYTL→165-171  SVMTFSV→235-241 SITSPPT→246-252  SSDTGHQ→307-313 SAESIVL→321-327 SKGTSKS→334-340	SFCSVSM→18-24  SMKSLGT→91-97 SLGTMSV→94-100 STVSYTL→165-171  SVMTFSV→235-241 SITSPPT→246-252 SPTTMSV→249-255 SSDTEHQ→307-313  SNDTSKS→334-340	SFCSVSM→18-24  SMKSLAT→91-97 SLATMSV→94-100 STVSYTL→165-171  SVMTFSV→235-241 SITSPPT→246-252 SPTTMSV→249-255	SFCSVSM→18-24  SMKSLGT→91-97 SLGTMSV→94-100 STVSYTL→165-171  SVMTFSV→235-241 SITSPPT→246-252 SPTTMSV→249-255  SKGTSKS→334-340	SFCSVSM→18-24 SAFSLDV→67-73 SMLSTSM→86-92 SMKSLGS→91-97 SLGMSV→94-100  SETTSLG→217-223 SVMTFSV→235-241 SITSPPT→246-252 SPTTMSV→249-255	CK1 phosphorylation site [nucleus, cytosol]
<b>MOD_CK2_1</b> (... [ST]P.)	HTRSVME→ 3-9  KGTSKSE→335-341	HTRSVME→ 3-9 KASSDTE→305-311 AGKTTAE→317-323 NDTSKSE→335-341	KASSDTE→305-311 TGKTTTE→317-323	HTRSVME→ 3-9 KASSDTE→305-311 AGKTTAE→317-323 NDTSKSE→335-341	KSATDAE→305-311	CK2 phosphorylation site [nucleus, cytosol, protein kinase CK2 complex]
<b>MOD_GSK3_1</b> (S. ([ST])...)	MLFTSMKS→87-94 IVSTVSYT→163-170  VLGSVNS→176-183  LAFSVLVS→227-234 VLVSSVMT→231-238 MSITSPPT→245-258 TTMSVVG→251-258  DEISKGTS→331-338	MLFTSMKS→87-94 IVSTVSYT→163-170  VLGSVNS→176-183  LTFSVLVS→227-234 VLVSSVMT→231-238 MSITSPPT→245-258 TTMSVVG→251-258  AGKTTAES→317-324 DENSNDTS→331-338	MLFTSMKS→87-94 IVSTVSYT→163-170  VLGSVNS→176-183  LTFSVLVS→227-234 VLVSSVMT→231-238 MSITSPPT→245-258 TTMSVVG→251-258 MQQTGKTT→314-321 TGKTTTES→317-324 LVRTDENS→327-334 DENSNDTS→331-338	MLFTSMKS→87-94 IVSTVSYT→163-170  VLGSVNS→176-183  LTFSVLVS→227-234 VLVSSVMT→231-238 MSITSPPT→245-258 TTMSVVG→251-258 AGKTTAES→317-324 DENSNDTS→331-338	LFVSMST→83-90 MLSTSMKS→87-94 QTPSVVY→126-133 IAATVLYT→163-170 YMKTMLGS→172-179 MLGSVNS→176-183 TAISSTTS→214-221 LVFVLVS→227-234 VLVSSVMT→231-238 MSITSPPT→245-258 TTMSVVG→251-258 RQQTKMNS→314-321	GSK3 phosphorylation recognition site. [Nucleus, Cytosol]
<b>MOD_N-GLC_1</b> (. (N)[^P][ST].)	MNHTRS→1-6	MNHTRS→1-6 SNDTSK→334-339	MNHTRS→1-6 ENSSDN→332-337	MNHTRS→1-6 SNDTSK→334-339	MNHSHI→1-6 LNITAN→297-302 MNSSPF→319-324	Generic motif for N-glycosylation. Shakin-Eshleman et al. showed that Trp, Asp, and Glu are uncommon before the Ser/Thr position. Efficient glycosylation usually occurs when ~60 residues or more separate the glycosylation acceptor site from the C-terminus. [extracellular, endoplasmic reticulum, Golgi apparatus]

MOD_ProDKin_1 (... ([ST])P..)	KAQTPMV→124-130 SITSPTT→246-252	KAQTPMV→124-130 SITSPTT→246-252	KAQTPMV→124-130 SITSPTT→246-252	KAQTPMV→124-130 SITSPTT→246-252	KAQTPSV→124-130 KWVTPWG→149-155 SITSPTT→246-252 MNSSPFA→319-325	Proline-Directed Kinase (e.g. MAPK) phosphorylation site in higher ukaryotes. [Nucleus, Cytosol]
TRG_ENDOCYTIC_2 (Y.[LMVIF])	YPAF→66-69 YFSF→131-134 YGPV→187-190 YNNL→192-195 YLG1→281-284 YTHL→294-297	YPAF→66-69 YFSF→131-134 YGPV→187-190 YNNL→192-195 YLG1→281-284 YTHL→294-297	YPAF→66-69 YFSF→131-134 YGPV→187-190 YNNL→192-195 YLG1→281-284 YTHL→294-297	YPAF→66-69 YFSF→131-134 YGPV→187-190 YNNL→192-195 YLG1→281-284 YTHL→294-297	YSAF→66-69 YVSF→131-134 YGPV→187-190 YNNL→192-195 YVGI→281-284 YTHL→294-297	Tyrosine-based sorting signal responsible for the interaction with mu subunit of AP (Adaptor Protein) complex. [cytosol, plasma membrane, clathrin-coated endocytic vesicle]
TRG_NES_CRM1_1: ([DEQ].{0,1}[LIM].{2,3}[LIVMF].{2,3}[LMVF].[LMIV].{0,3}[DE][DE].{0,1}[LIM].{2,3}[LIVMF].{2,3}[LMVF].[LMIV].{0,3}[DEQ])	DMNFPFGILVLQ→36-47 QTILKNLAVVLIALGD→103-118	DMNFPFGILVLQ→36-47 QTILKNLAVVLIALGD→103-118	DMNFPFGILVLQ→36-47	DMNFPFGILVLQ→36-47 QTILKNLAVVLIALGD→103-118	DMNFPFGILLQ→36-47 QTIIKNLAVVLIALGD→103-118	Some proteins re-exported from the nucleus contain a Leucine-rich nuclear export signal (NES) binding to the CRM1 exportin protein. [Nucleus, Cytosol]
LIG_14-3-3_3 [RHK][STALV].[ST]. [PESRDIF]	-----	-----	KTTTES→ 319-324	KAQTPS→ 124-129	-----	Consensus derived from natural interactors which do not exactly match the mode1 and mode2 ligands. [Nucleus, Cytosol, Internal side of plasma membrane]
MOD_SUMO [VILMAFP]K.E]	-----	-----	FKSE →338-341	-----	-----	Motif recognised for modification by SUMO-1. [Nucleus, PML body]
LIG_TRAF2_1 [PSAT].[QE]E]	-----	-----	-----	-----	APEE→328-331 PEEE→329-332	Major TRAF2-binding consensus motif. Members of the tumor necrosis factor receptor (TNFR) superfamily initiate intracellular signaling by recruiting the C-domain of the TNFR-associated factors (TRAFs) through their cytoplasmic tails. [Cytosol]

**Table 6:** Sequence domains identified in LPG2 protein of different *Leishmania* strains from ELM server.

CK2, glycogen synthase kinase 3 (GSK3), and the Golgi apparatus casein kinase (G-CK) (Pinna and Ruzzene, 1996). Casein kinase I have a wide variety of substrates. CK1 phosphorylation motifs (7-11 numbers) (throughout the LPG2 sequence) have been identified in LPG2 protein of *Leishmania* strains. Reports on various CK1 inhibitors are known which can be

applied to LPG2 protein by docking to find novel drug candidates for leishmaniasis treatment (Table 6) (Rena et al., 2004). Protein kinase CK2 is a pleiotropic and ubiquitous serine or threonine kinase, which is highly conserved during evolution (Faust and Montenarh, 2000). Protein kinase CK2 can phosphorylate many protein substrates in addition to casein. CK2

phosphorylation motifs have been identified in LPG2 protein of *Leishmania* strains, one at the beginning of amino terminus and others towards carboxy terminus.

Glycogen synthase kinase 3 (GSK3) is a well conserved serine/threonine kinase that is implicated in different cellular processes controlling cell proliferation and programmed cell death (Frame and Cohen, 2001). In resting cells, GSK3 is a constitutively active kinase that phosphorylates a wide range of protein substrates to directly inhibit their biochemical activities, interfere with their sub-cellular localization, or promote their degradation (Ali et al., 2001). GSK3 phosphorylation recognition sites are found in LPG2 protein throughout the sequence of *Leishmania*. On comparison of GSK3 sites, amino acids involved in formation of this recognition site are same but in few cases it is different.

N-glycosylation motifs have been detected in different strains at the beginning of N-terminal and towards carboxy terminus. In LPG2 protein of *L. braziliensis*, three N-glycosylation motifs are present. Among the different types of glycosylation, the N-linked attachment of sugars to the polypeptide backbone is by far the most abundant modification (Vijay, 1998).

Two proline-directed kinase (e.g. MAPK) phosphorylation sites (124-130 and 246-252) are found in four *Leishmania* strains whereas in *L. braziliensis*, four PDK phosphorylation sites are present. One way that they achieve this is through direct interactions with substrate residues flanking the phosphorylation (P) site (Pinna and Ruzzene, 1996; Lu et al., 2002; Songyang et al., 1996). The preference for proline at the P+1 position may be linked to downstream signaling mechanisms mediated by Pin1 proline isomerization (Zhou et al., 1999).

Tyrosine-based sorting signal sequences have been found throughout the LPG2 protein sequence of *Leishmania* which is responsible for the interaction with mu subunit of AP (adaptor protein) complex. Multiple sorting steps within eukaryotic cells are mediated by tyrosine-based sorting motifs. These motifs are recognized by the medium-chain subunits of heterotetrameric adaptor complexes (Stephens and Banting, 1998).

Some proteins re-exported from the nucleus contain a leucine-rich nuclear export signal (NES) binding to the CRM1 exportin protein. CRM1 mediates the nuclear export of proteins exposing leucine-rich nuclear-export signals (NESs) (Kutay and Güttinger, 2005). Two such motifs [36<sup>th</sup> - 47<sup>th</sup> and 103<sup>rd</sup> - 118<sup>th</sup> amino acids] are detected in LPG2 protein of *Leishmania* strains except *L. major* where only one site [36<sup>th</sup> -

47<sup>th</sup> amino acids] is present. Hence the LPG2 protein manufactured in nucleus is transported to cytosol or Golgi due to presence of these motifs in LPG2 protein itself.

Interestingly, the sequences “KTTTES” and “KAQTPS” of LPG2 protein of *L. major* and *L. infantum* respectively matches a 14-3-3 interaction consensus site derived from natural interactors. Dimerization of 14-3-3 is essential for a good interaction (Cahill et al., 2001; Tzivion et al., 1998). P53 tumor suppressor protein, which plays a major role in maintaining genomic stability, has only one 14-3-3 binding site (Waterman et al., 1998). Hence LPG2 protein of *Leishmania* may have genomic stability role, which is required to be proved experimentally.

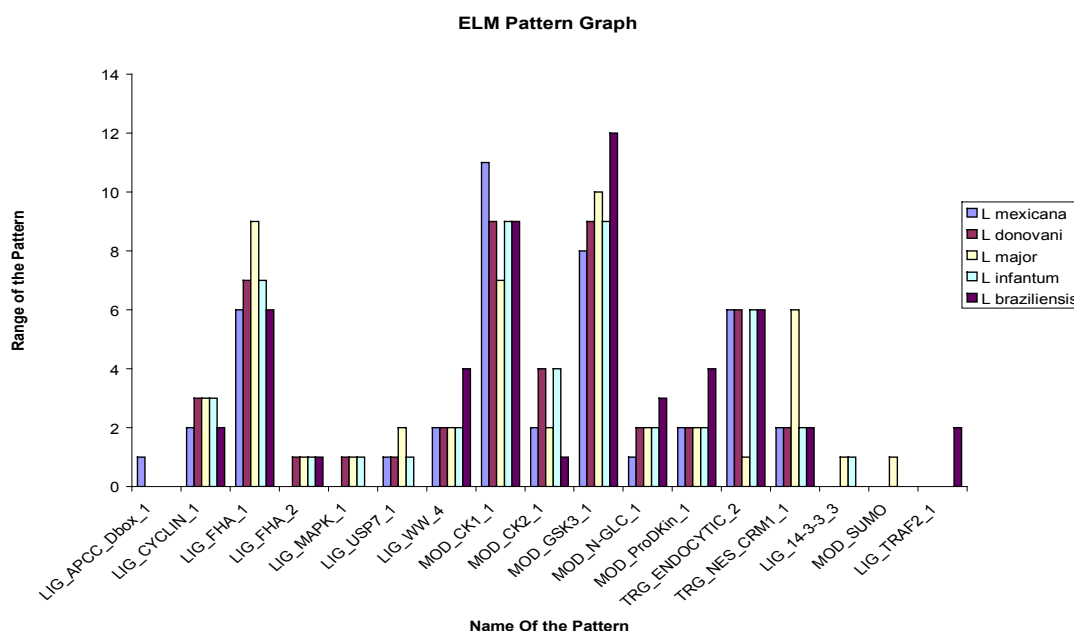
Only novel motif [FKSE] recognized for modification by SUMO-1 has been identified in LPG2 protein of *L. major*. The small, ubiquitin-related protein SUMO-1 is highly conserved from yeast to humans (Muller et al., 2001) and has been associated with subnuclear localization of many cellular proteins. Like ubiquitination, sumoylation leads to attachment of SUMO-1 to target proteins through the  $\gamma$ -NH<sub>2</sub> group of lysine residues, using a cascade of E1, E2, and E3 enzymes. SUMO-1 is an important determinant of protein localization, required for the speckled nuclear distribution of the proteins PML, TEL, and HIPK2.

Likewise major TRAF2-binding consensus motif found in members of the tumor necrosis factor receptor (TNFR) superfamily, which initiate intracellular signaling by recruiting the C-domain of the TNFR-associated factors (TRAFs) through their cytoplasmic tails, is also present in LPG2 protein of only one strain i.e. *L. braziliensis*.

In this analysis we have identified many motifs in LPG2 protein of *Leishmania*, which were previously reported in other proteins of different organisms, i.e. yeast to mammals. It is learned from this computational analysis that in LPG2 protein of *L. braziliensis* several distinct motifs are present which are not found in other *Leishmania* strains. The graphical view of all the motifs that present in LPG2 protein of all *Leishmania* Strains are shown in (Figure 5).

### Protein ligand binding site analysis

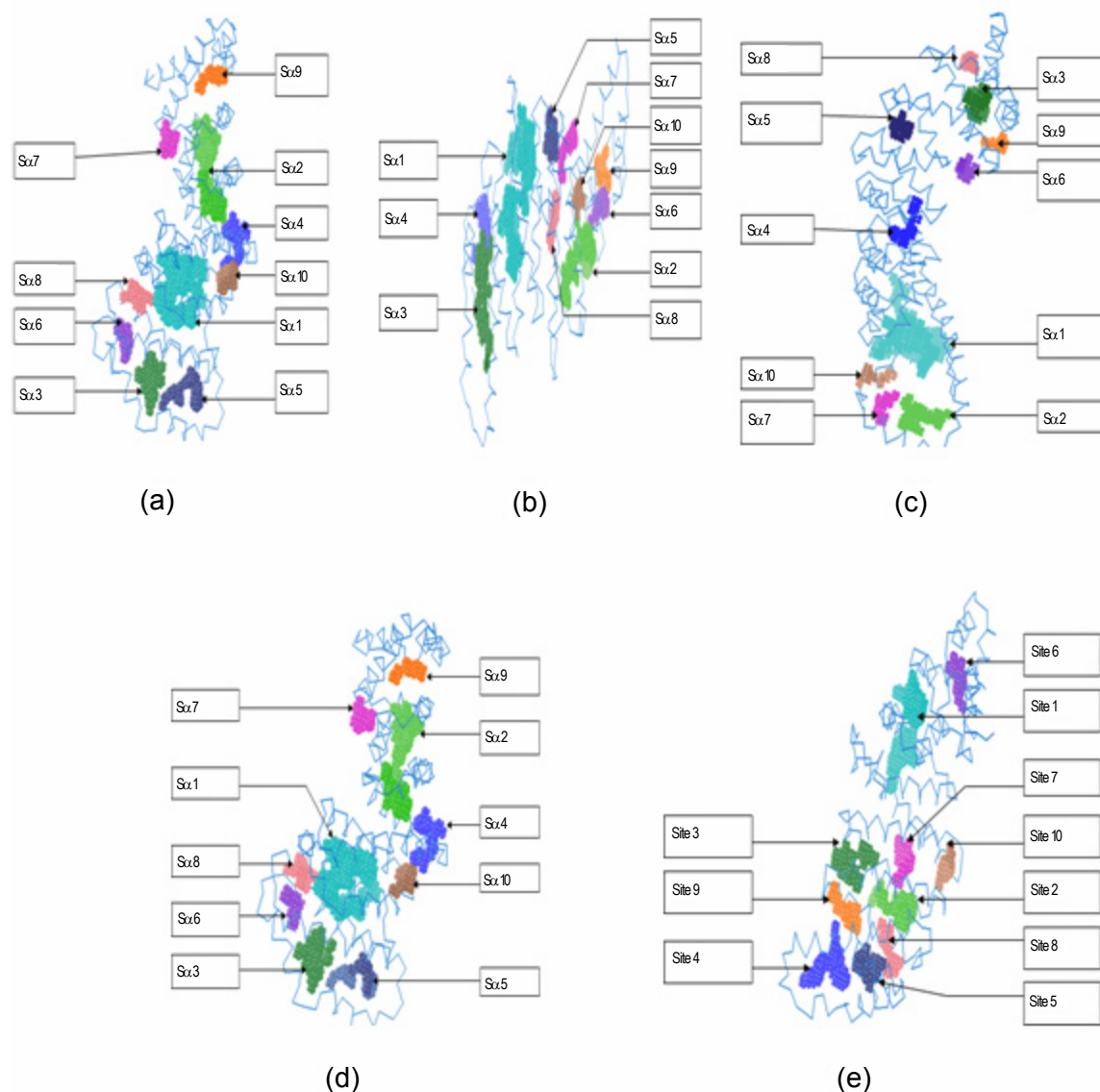
Potential ligand binding sites in LPG2 protein of all *Leishmania* strains (*L. donovvani*, *L. major*, *L. infantum*, *L. mexicana* and *L. braziliensis*) have been found by using pocket finder program (Table 7) (Ruppert et al., 1997). Different LBSs of LPG2 protein of *Leishmania* are shown in (Figure 6). First LBS of all the strains involve more than 35 amino acids. At most thirty eight amino acids (*L. mexicana*)



**Figure 5:** This graph shows involvement of different conserved motifs in LPG2 protein of different *Leishmania* strains from ELM Server. Different colors have been used to distinguish the presence of various motifs in different *Leishmania* strains.

Ligand Binding Site (LBSs)	Total number of amino acids in <i>L. mexicana</i>	Total number of amino acids in <i>L. donovani</i>	Total number of amino acids in <i>L. braziliensis</i>	Total number of amino acids in <i>L. major</i>	Total number of amino acids in <i>L. infantum</i>
Site 1	38	37	30	35	37
Site 2	36	36	17	24	9
Site 3	17	17	18	21	16
Site 4	12	12	16	13	11
Site 5	10	10	14	10	9
Site 6	12	11	14	10	7
Site 7	10	10	13	10	4
Site 8	9	9	10	14	7
Site 9	9	9	14	8	8
Site 10	10	10	13	7	7

**Table 7:** It shows various possible Ligand Binding Sites (LBSs) and total number of amino acids and their positions that involved in each binding site of *L. mexicana*, *L. donovani*, *L. braziliensis*, *L. major* and *L. infantum*.



**Figure 6:** Prediction of Ligand Binding Sites (LBSs) of LPG2 protein of five *Leishmania* strains by the pocket finder program (a) *L. donovani*, (b) *L. major*, (c) *L. infantum* (d) *L. mexicana* and (e) *L. braziliensis*

involved in formation of binding sites have been found and in few sites a single amino acid is changed or lacking in formation of ligand binding sites, have been detected. Lowest number (four) of amino acids involved in formation of LBSs has been found in *L. infantum* (7<sup>th</sup>). It is known from this analysis that some LBSs are similar in all the strains but other LBSs are specific to each strain. All LBSs of *L. donovani* and *L. mexicana* are very similar, except at 1<sup>st</sup> and 6<sup>th</sup> where one amino acid is absent in *L. donovani* (1<sup>st</sup> site: G-188<sup>th</sup> and 6<sup>th</sup> site: V-328<sup>th</sup>).

All these structural, functional and proteomic analysis about LPG2 protein of *Leishmania* species will lead to identification of novel lead compounds to eliminate *Leishmania* infection in Bihar state, India.

## Future Perspectives

As the structure of LPG2 protein is known from this study, novel lead compounds can be designed on the basis of ligand protein interaction (docking) scores of available antileishmanial drugs with LPG2 protein, defining the highest dockable compound and designing various analogues of the presently available drugs or defining a novel molecule on the basis of different binding sites of LPG2 protein of different *Leishmania* strains.

## Acknowledgments

This study was supported by Indian Council of Medical Research (ICMR), Govt. of India. We are thankful to Dr.



Meera Singh of ICMR for helping us during establishment of our division. We acknowledge Dr. Sindhu Prava Rana for helping us in preparation of the manuscript.

## Reference

1. Abeijon C, Mandon EC, Hirschberg CB (1997) Transporters of nucleotide sugars, nucleotide sulfate and ATP in the Golgi apparatus. *Trends Biochem Sci* 22: 203-207. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
2. Ahn NG, Seger R, Krebs EG (1992) The mitogen-activated protein kinase activator. *Curr Opin Cell Biol* 4: pp992-999.» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
3. Ali A, Hoeflich KP, Woodgett JR (2001) Glycogen synthase kinase-3: properties, functions, and regulation. *ChemRev* 101: 2527-2540.» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
4. Bairoch A, Bucher P, Hofmann K (1997) The PROSITE database, its status in 1997. *Nucleic Acids Res* 25: 217-221. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
5. Bedford MT, Reed R, Leder P (1998) WW domain-mediated interactions reveal a spliceosome-associated protein that binds a third class of proline-rich motif: The proline glycine and methionine-rich motif. *Proc Natl Acad Sci USA* 95: 10602-10607. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
6. Bhattacharyya A, Mukherjee M, Duttagupta S (2002) Studies on Stibanate unresponsive isolates of *Leishmania donovani*. *J Biosci* 27: 503-508. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
7. Cahill CM, Tzivion G, Nasrin N, Ogg S, Dore J, et al. (2001) Phosphatidylinositol 3-kinase signaling inhibits DAF-16 DNA binding and function via 14-3-3-dependent and 14-3-3-independent pathways. *J Biol Chem* 276: 13402-13410. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
8. Cai CZ, Han LY, Chen X, Cao ZW, Chen YZ (2005) Prediction of Functional Class of the SARS Coronavirus Proteins by a Statistical Learning Method. *J Proteome Res* 4: 1855-1862. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
9. Cai CZ, Han LY, Ji ZL, Chen X, Chen YZ (2003) SVM-Prot: Web-Based Support Vector Machine Software for Functional Classification of a Protein from Its Primary Sequence. *Nucleic Acids Res* 31: 3692-3697. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
10. Ceroni A, Frasconi P, Passerini A, Vullo A (2004) Disulfide connectivity prediction using recursive neural networks and evolutionary information. *Bioinformatics* 20: 653-659.» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
11. Croft SL, Sundar S, Fairlamb AH (2006) Drug resistance in leishmaniasis. *Clinical Microbiol Rev* 19: 111-126. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
12. Descoteaux A, Luo Y, Turco SJ, Beverley SM (1995) A specialized pathway affecting virulence glycoconjugates of *Leishmania*. *Science* 269: 1869-1872.» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
13. Donella DA, Grankowski N, Kudlicki W, Szyszka R, Gasior E, et al. (1985) A type-1 casein kinase from yeast phosphorylates both serine and threonine residues of casein. Identification of the phosphorylation sites. *Biochim Biophys Acta* 829: 180-187. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
14. Faust M, Montenarh M (2000) Subcellular localization of protein kinase CK2. A key to its function. *Cell Tissue Res* 301: 329-40. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
15. Flotow H, Graves PR, Wang A, Fiol CJ, Roeske RW, et al. (1990) Phosphate groups as substrate determinants for casein kinase I action. *J Biol Chem* 265: 14264-14269. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
16. Frame S, Cohen P (2001) GSK3 takes centre stage more than 20 years after its discovery. *Biochem J* 359: 1-16. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
17. Hendlich M, Rippmann F, Barnickel G (1997) LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. *J Mol Graph Model* 15: 359-363. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
18. Hirokawa T, Boon CS, Mitaku S (1998) SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* 14: 378-379.» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
19. Hofmann K, Stoffel W (1993) TMbase - A database of membrane spanning proteins segments. *Biol Chem* 374: 166-170. » [Google Scholar](#)
20. Hong K, Ma D, Beverley SM, Turco SJ (2000) The *Leishmania* GDP-mannose transporter is an autonomous, multi-specific, hexameric complex of LPG2 subunits. *Biochemistry* 39: 2013-2022. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
21. Krogh A, Larsson B, Von HG, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305: 567-80. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
22. Kutay U, Güttinger S (2005) Leucine-rich nuclear-export signals: born to be weak. *Trends Cell Biol.* 15: 121-124.» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
23. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol*

- Biol 157: 105-132.» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
24. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 123: 2947-8.» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
25. Lu KP, Liou YC, Zhou XZ (2002) Pinning down proline-directed phosphorylation signaling. *Trends Cell Biol* 12: 164-172.» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
26. Ma DQ, Russell DG, Beverley SM, Turco SJ (2004) Reconstitution of GDP-Man Transport Activity with Purified. *J Biol Chem* 280: 2018-2035.» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
27. Marshall CJ (1995) Specificity of receptor tyrosine kinase signaling: transient versus sustained extracellular signal-regulated kinase activation. *Cell* 80: pp179-185.» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
28. Meggio F, Donella DA, Pinna LA (1979) Studies on the structural requirements of a microsomal cAMP-independent protein kinase. *FEBS Lett* 106: 76-80.» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
29. Muller S, Hoegge C, Pyrowolakis G, Jentsch S (2001) SUMO, ubiquitin's mysterious cousin. *Nat Rev Mol Cell Biol* 2: 202-10.» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
30. Ng K, Handman E, Bacic A (1994) Biosynthesis of lipophosphoglycan from *Leishmania major*: characterization of ( $\beta$ 1-3)-galactosyltransferase(s). *Glycobiology* 4: 845-853.» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
31. Nishida E, Gotoh Y (1993) The MAP kinase cascade is essential for diverse signal transduction pathways. *Trends BiochemSci* 18: pp128-131.» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
32. Paris C, Loiseau PM, Bories C, Bre'ard J (2004) Miltefosine Induces Apoptosis-Like Death in *Leishmania donovani* Promastigotes. *Antimicrob Agents Chemother* 48: pp852-859.» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
33. Pinna LA, Ruzzene M (1996) How do protein kinases recognize their substrates. *Biochim Biophys Acta* 1314: 191-225.» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
34. Puntervoll P, Linding R, Gemünd C, Chabanis DS, Mattingsdal M, et al. (2003) ELM server: a new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Res* 31: 3625-3630.» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
35. Rena G, Bain J, Elliott M, Cohen P (2004) D4476, a cell-permeant inhibitor of CK1, suppresses the site-specific phosphorylation and nuclear exclusion of FOXO1a. *EMBO reports* 5: 60-65.» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
36. Roach PJ (1991) Multisite and hierarchical protein phosphorylation. *J Biol Chem* 266: 14139-14142.» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
37. Rost B, Yachdav G, Liu J (2004) The PredictProtein Server. *Nucleic Acids Res* 32: W321-W326.» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
38. Ruppert J, Welch W, Jain AN (1997) Automatic identification and representation of protein binding sites for molecular docking. *Protein Sci* 6: 524-533.» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
39. Sahoo GC, Dikhit MR, Das P (2008) Functional assignment to JEV proteins using SVM. *Bioinformation* 3: 1-7.» [Pubmed](#) » [Google Scholar](#)
40. Saridakis V, Sheng Y, Sarkari F, Holowaty MN, Shire K, et al. (2005) Structure of the p53 Binding Domain of HAUSP/USP7 Bound to Epstein-Barr Nuclear Antigen 1 Implications for EBV-Mediated Immortalization. *Mol Cell* 18: 25-36.» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
41. Songyang Z, Lu KP, Kwon YT, Tsai LH, Filhol O, et al. (1996) A structural basis for substrate specificities of protein Ser/Thr kinases: Primary sequence preference of casein kinases I and II, NIMA, phosphorylase kinase, calmodulin-dependent kinase II, CDK5, and Erk1. *Mol Cell Biol* 16: 6486-6493.» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
42. Stephens DJ, Banting G (1998) Specificity of interaction between adaptor-complex medium chains and the tyrosine-based sorting motifs of TGN38 and Igp120. *Biochem J* 335: 567-572.» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
43. Sturgill TW, Wu J (1991) Recent progress in characterization of protein kinase cascades for phosphorylation of ribosomal protein S6. *Biochim Biophys Acta* 1092: pp350-357.» [Pubmed](#) » [Google Scholar](#)
44. Takeda DY, Wohlschlegel JA, Dutta A (2001) A Bipartite Substrate Recognition Motif for Cyclin-Dependent Kinases. *J Biol Chem* 276: 1993-1997.» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
45. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673-4680.» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
46. Tusnády GE, Simon I (1998) Principles Governing Amino Acid Composition of Integral Membrane Proteins: Application to Topology Prediction. *J Mol Biol* 283: 489-506.» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
47. Tzivion G, Luo Z, Avruch J (1998) A dimeric 14-3-3

- protein is an essential cofactor for Raf kinase activity. Nature 394: 88-92. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
48. Vergnes B, Gourbal B, Girard I, Sundar S, Drummelsmith J, et al. (2007) A Proteomics Screen Implicates HSP83 and a Small Kinetoplastid Calpain-related Protein in Drug Resistance in *Leishmania donovani* Clinical Field Isolates by Modulating Drug-induced Programmed Cell Death. Mol Cell Proteomics 6: 88-101. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
49. Vijay IK (1998) Developmental and Hormonal Regulation of Protein N Glycosylation in the Mammary Gland. J Mammary Gland Biol Neoplasia 3: 325-336. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
50. Waterman MJ, Stavridi ES, Waterman JL, Halazonetis TD (1998) ATM-dependent activation of p53 involves dephosphorylation and association with 14-3-3 proteins. Nat Genet 19: 175-178. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
51. Winberg ME, Rasmusson B, Sundqvista T (2007) Leishmania donovani: Inhibition of phagosomal maturation is rescued by nitric oxide in macrophages. Exp Parasitol 117: 165-170. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
52. Zhou XZ, Lu PJ, Wulf G, Lu KP (1999) Phosphorylation-dependent prolyl isomerization: A novel signaling regulatory mechanism. Cell Mol Life Sci 56: 788-806. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)