

# Characterizing hp's from leptospira interrogans



**ASSIGN  
BUSTER**

## Introduction:

Genome sequencing projects and genetic engineering has revealed many aspects of complex cellular environment containing large number of proteins. Despite sequences of most of organisms are available and proteins coded are studied experimentally, there are some proteins whose functions are unknown, need to be characterised(5). Such proteins are known as Hypothetical proteins (HP) sequences of which are known but there is no evidence of experimental study(6). There is extensive need to study and classify these hypothetical proteins which can open new way to design drug molecules against infectious organisms. Functional annotation of HP's involved in infection, drug resistance, and essential biosynthetic pathways is important for development of the potent antibacterials against infectious agents. Improved understanding of these proteins may make them potential targets of antimicrobial drugs [26]. *Leptospira interrogans* is gram negative spirochete, having an internal flagella is pathogenic which causes Leptospirosis (1)(2), other serovars (strains) are distinguished on the basis of cell surface antigens. These are infectious to animals, but through animal urine can be spread to human(3). *Leptospira* enters in body via broken skin, mucosa and spreads in body, if immune system fails to stop the growth of bacteria it cause severe hepatic and renal dysfunctions(4). This present study highlights their silicostudies to characterize HP's from *Leptospira interrogans*.

## Methods:

### Sequence Retrieval:

<https://assignbuster.com/characterizing-hps-from-leptospira-interrogans/>

KEGG (Kyoto Encyclopedia of Gene and Genomes) is a large collection of databases having entries of genes, proteins, pathways in metabolism and diseases, drug and ligands of organism (7). We have selected the Sequences of 12 hypothetical proteins of *Leptospira interrogans* randomly from KEGG database ([www. genome. jpg/kegg](http://www.genome.jp/kegg)).

Pfam:

Pfam is curated Protein families database, it uses jackhmmer programme (HMMR3). To give profile HMM (Hidden Markov Model) with PSI-BLAST, which were searched against UniProt(9). However, to include protein in a family its domain and sequence bit scores must be equal or above the Gathering Thresholds (GA). Pfam gives Pfam A families which are manually curated and Pfam B families generated automatically(8).

Batch CD search:

Hypothetical Protein sequences were searched for conserved domains at batch CD search, which gives results by using MSA and 3D structures for homologous domains available on Pfam and SMART(9)(10).

ExpASy-ProtParam tool:

ProtParam tool ([www. expasy. org/tools/protparam. html](http://www.expasy.org/tools/protparam.html)) was used to estimate physicochemical parameters of hypothetical proteins(11). Query protein can be submitted in form of SWISS/TrEMBL ID or protein sequence. Server provides directly calculated values of pI/MW (Isoelectric point, Molecular Weight), Percentage of each amino acid, Extinction coefficient

<https://assignbuster.com/characterizing-hps-from-leptospira-interrogans/>

(EC), Instability Index (II)(12), Aliphatic Index (AI) and GRAVY (Grand Average of Hydrophobicity).

SOSUI server:

Amphiphilicity index and Hydropathy index of query protein sequences were calculated by SOSUI server which categorises protein into cytoplasmic or transmembrane nature(13).

Protein-Protein Interaction network:

Protein in the cell environment interacts with other proteins, in silico these interactions were studied by STRING v9. 1 (Search Tool for Retrieval of Interacting Genes). STRING is a large repository of protein-protein interactions involving functional interactions, stable complexes, and regulatory interactions among proteins(14, 15). Figure 1. Shows resulting protein-protein interaction network of selected hypothetical proteins, for better understanding interaction networks should be seen on server site.

Disulfide-Bonding in Protein

Disulfide bonds among cysteine residues in protein plays an important role in folding it into functional and stable conformation. DISULFIND server utilizes SVM binary server to predict bonding state of cysteins, these cysteins are paired by Recursive Neural Network to show disulfide bridges(16).

Protein Structure Prediction:

Protein structure prediction server (PS) <sup>2</sup> (17) requires query sequence in fasta format to generate 3D structure by comparative modelling(18). Server <https://assignbuster.com/characterizing-hps-from-leptospira-interrogans/>

utilizes consensus strategy to find template using PSI-BLAST and IMPALA.

Query sequence and template aligned by T-coffee, PSI-BLAST, and IMPALA [13]. 3D structures are predicted from template using MODELLER and visualised by CHIME, Raster3D. Resulting 3D structural model of selected hypothetical proteins are shown in Figure 2.

Ligand Binding site Prediction:

Q-site finder(19) server was used for binding site prediction in selected proteins. Server uses energy based methods to find clefts on protein surface for ligands(20). These hot spots for ligand binding have predicted after ranking their physicochemical properties as hydrophobicity, desolvation, electrostatic & van der waal potentials.

Discussion:

ProtParam tool computes different physicochemical parameters depending on the queries submitted to the databases. Isoelectric focusing separates proteins according to pI where pH gradients are developed(21). Predicted pI via server may not be adequate because in case of high number of basic amino acids and lower buffer capacity. By using pH gradients and calculated pI, proteins can be separated experimentally. MW of proteins along with pI is used for the 2D gel electrophoresis. EC shows a light absorbed by a protein relative to their composition at a specific wavelength. EC given (Table 1) are calculated with reference to Tryptophan, Cysteine, Tyrosine (11). Instability index (II) refers to the stability of the protein in test tube(22). Among studied proteins giÇ€24214908, giÇ€24215664, giÇ€24216444, giÇ€24213620, giÇ€24213945 were found to be unstable, and rest are stable (proteins with II

<https://assignbuster.com/characterizing-hps-from-leptospira-interrogans/>

above 40 are unstable). Aliphatic amino acid constitutes the aliphatic index (a relative volume of aliphatic side chains). Increased AI results into a hydrophobic interactions and thus gives thermostatic stability to protein, predicted AI and II shows inverse relation for stability except these two proteins gi|24215664 and gi|24215909. GRAVY(23) values are a ratio of all hydropathy values of amino acids to the number of residues in sequence. Smaller the GRAVY(23) more hydrophilic is protein, gi|24214908 and gi|24213945 proteins found the most hydrophilic. In case of 3D structure hydrophilic domains tends to be on exterior surface, while hydrophobic domains avoids external environment and forms internal core of the protein. Search of family for hypothetical proteins based on conserved domains having consensus sequence in their structure is given in Table 3.

Hypothetical protein gi|24214908 found to be a member of GH18\_CFLE\_spore\_hydrolase, Cortical fragment Lytic Enzyme bearing a catalytic domain from glycosyl hydrolase, an enzyme used in breaking a spore peptidoglycans so as to activate it for germination when favourable conditions are available. Hypothetical protein gi|24215649 from PDZ\_serine\_protease involved in protein reassembly and work as a heat shock protein. Protein gi|24215664 belongs to Leucine-rich Repeats (LRR), ribonuclease inhibitor like family. LRR are motifs having role in protein interactions in complex networks. S-adenosylmethionine decarboxylase (AdoMetDC) enzyme for biosynthesis of spermine and spermidine by decarboxylation of SAM belongs to Ado\_Met\_dc family (gi|24217373). Pilz domain in gi|24213620 is found in bacterial cellulose synthase and other proteins that forms biofilm around a bacterium and involve in effluxing drug(24). Hypothetical protein gi|294827583 (FecR superfamily) is involved

in Iron transport system in bacterial membranes, Fe<sup>3+</sup> (insoluble) loaded on citrate carrier is sensed by FecR protein found in periplasmic space in bacterial membrane(25). Protein sites are predicted as cytoplasmic, host associated, extracellular, cytoplasmic membrane proteins. SOSOI server predictions (Table 6) shows that positively charged amino acids are more at the end of trans membrane region. Protein-protein interaction study has shown some hypothetical proteins are involved in essential cellular process such as transport across membrane, biosynthesis of molecules, translational regulation. Hypothetical protein gi|24214908 (Figure 1) interacts with SUA5 protein which is known as one of translational regulator from YrdC/SUA5 family. Search for gi|24215909 shown to be involved in chloride transport with chloride channel protein (EriC gene). Protein gi|24217373 found to be interacted with S-layer like protein (slpM) which forms layer around bacteria to attach other surfaces and protect it from environment. Additionally it involve in cell devising processes and transport across membrane. Protein gi|294827687 had shown interaction with proteins for bleomycin resistance, chorismate synthase (Trp biosynthesis) and Mammalian Cell entry (MCE) like proteins. Figure 2 shows 3D structures of proteins gi|24214908, gi|24213620, gi|24214753, gi|24213945 predicted from amino acid sequence on PS<sup>2</sup> server by using templates 1vf8A, 3bo5A, 1f9zA, and c2efsA respectively.

#### Conclusion:

Development of potential bioinformatics tools and databases has opened new platform for in-silico study. Currently it is very needful to annotate and characterize hypothetical proteins in *Leptospira interrogans* serovar. These <https://assignbuster.com/characterizing-hps-from-leptospira-interrogans/>

hypothetical proteins may have an imperative role in producing many virulence factors and cause serious infection or disease. We have analyzed 12 hypothetical proteins from KEGG database and categorized its physicochemical properties and recognized domains and families using various bioinformatics tools and databases. The structures were modeled and their ligand binding sites were identified. Physicochemical predictions made for hypothetical proteins, which can be used to find therapeutic agents against infections caused by *Leptospira interrogans*. Some of hypothetical proteins serves as channel proteins, ribosomal proteins or are involved in cell cycle process. Families which were identified for these hypothetical proteins are involved in normal cellular processes and the resistance against drugs. Ligand binding hotspots were found with Q-sitefinder which shown amino acids involved in interaction with ligands. It will help in study of molecular docking for development of potent and effective target against Leptospira infection.

#### Acknowledgement:

This study was supported by NIPER Guwahati academic staff. We are very grateful for their excellent support in every manner.

#### References:

1. Chou L-F, Chen Y-T, Lu C-W, Ko Y-C, Tang C-Y, Pan M-J, et al. Sequence of *Leptospira santarosai* serovar Shermani genome and prediction of virulence-associated genes. *Gene* [Internet]. 2012; 511: 364-70. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23041083>



2. Langston CE, Heuter KJ. Leptospirosis. A re-emerging zoonotic disease. *Vet. Clin. North Am. Small Anim. Pract.* [Internet]. 2003; 33: 791-807. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0195561603000263>
3. Kohn B, Steinicke K, Arndt G, Gruber AD, Guerra B, Jansen A, et al. Pulmonary abnormalities in dogs with leptospirosis. *J. Vet. Intern. Med. Am. Coll. Vet. Intern. Med.* [Internet]. 2010; 24: 1277-82. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20738768>
4. Picardeau M, Brenot A, Saint Girons I. First evidence for gene replacement in *Leptospira* spp. Inactivation of *L. biflexa* flaB results in non-motile mutants deficient in endoflagella. *Mol. Microbiol.* [Internet]. 2001; 40: 189-99. Available from: <https://www.scopus.com/inward/record.url?eid=2-s2.0-0035050686&nerID=40&md5=ea6dce51e08375c70cdc92eb578e74b1>
5. Adinarayana KPS, Sravani TS, Hareesh C. A database of six eukaryotic hypothetical genes and proteins. *Bioinformation.* 2011; 6: 128-30.
6. Hsieh W-J, Pan M-J. Identification *Leptospira santarosai* serovar shermani specific sequences by suppression subtractive hybridization. *FEMS Microbiol. Lett.* [Internet]. 2004; 235: 117-24. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15158270>
7. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* [Internet]. 2004; 32: D277-D280. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/14681412>
8. <http://pfam.sanger.ac.uk/>.

9. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, et al. The Pfam protein families database. *Nucleic Acids Res.* [Internet]. 2012 Jan [cited 2013 Sep 20]; 40(Database issue): D290–301. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3245129&tool=pmcentrez&rendertype=abstract>
10. Letunic I, Doerks T, Bork P. SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res.* [Internet]. 2011; 40: D302–5. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22053084>
11. Wilkins MR, Gasteiger E, Bairoch a, Sanchez JC, Williams KL, Appel RD, et al. Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol.* [Internet]. 1999 Jan; 112: 531–52. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10027275>
12. Mohan R, Venugopal S. Computational structural and functional analysis of hypothetical proteins of *Staphylococcus aureus*. *Bioinformatics* [Internet]. 2012 Jan; 8(15): 722–8. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3449381&tool=pmcentrez&rendertype=abstract>
13. Mitaku S, Hirokawa T, Tsuji T. aid in the characterization of amino acid preference at membrane - water interfaces. 2002; 18(4): 608–16.
14. Lewis ACF, Saeed R, Deane CM. Predicting protein-protein interactions in the context of protein evolution. *Mol. Biosyst.* [Internet]. 2010; 6: 55–64. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20024067>
15. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. STRING v9. 1: protein-protein interaction networks, with

- increased coverage and integration. *Nucleic Acids Res.* [Internet]. 2013 Jan [cited 2013 Sep 17]; 41(Database issue): D808-15. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3531103&tool=pmcentrez&rendertype=abstract>
16. Vullo A, Marta VS. Disulfide Connectivity Prediction using Recursive Neural Networks and Evolutionary Information. : 1-12.
  17. <http://ps2.life.nctu.edu.tw>.
  18. Chen C-C, Hwang J-K, Yang J-M. (PS)2: protein structure prediction server. *Nucleic Acids Res.* [Internet]. 2006 Jul 1 [cited 2013 Sep 24]; 34(Web Server issue): W152-7. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1538880&tool=pmcentrez&rendertype=abstract>
  19. <http://www.bioinformatics.leeds.ac.uk/qsitefinder>.
  20. Burgoyne NJ, Jackson RM. Predicting protein interaction sites: binding hot-spots in protein-protein and protein-ligand interfaces. *Bioinformatics* [Internet]. 2006 Jun 1 [cited 2013 Sep 23]; 22(11): 1335-42. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16522669>
  21. Bjellqvist B, Hughes GJ, Pasquali C, Paquet N, Ravier F, Sanchez JC, et al. The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis* [Internet]. 1993; 14: 1023-31. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/8125050>
  22. Guruprasad K, Reddy BVB, Pandit MW. Correlation between stability of a protein and its dipeptide composition•: a novel approach

for predicting in vivo stability of a protein from its primary sequence.

1990; 4(2): 155-61.

23. Kyte J, Doolittle RF, Diego S, Jolla L. A Simple Method for Displaying the Hydrophobic Character of a Protein. 1982; 105-32.
24. Amikam D, Galperin MY. PilZ domain is part of the bacterial c-di-GMP binding protein. *Bioinformatics* [Internet]. 2006 Jan 1 [cited 2013 Sep 24]; 22(1): 3-6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16249258>
25. Van Hove B, Staudenmaier H, Braun V. Novel two-component transmembrane transcription control: regulation of iron dicitrate transport in *Escherichia coli* K-12. *J. Bacteriol.* [Internet]. 1990 Dec; 172(12): 6749-58. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=210789&tool=pmcentrez&rendertype=abstract>
26. Galperin MY, Koonin EV: Searching for drug targets in microbial genomes. *Curr Opin Biotechnol* 1999, 10(6): 571-578.