Austin Journal of Proteomics, Bioinformatics & Genomics

Open Access

Austin Publishing Group

Research Article

In-Silico Characterisation of a Hypothetical Protein (LA_1016) of Leptospira Interrogans Serovar Lai Strain 56601

Samaha TH*

Regional Medical Research Centre (ICMR), Port Blair, Andaman and Nicobar Islands, India

***Corresponding author:** Samaha TH, Regional Medical Research Centre (Indian Council of Medical Research), Port Blair, Andaman and Nicobar Islands, India

Received: October 04, 2017; **Accepted:** October 23, 2017; **Published:** October 30, 2017

Abstract

Major fractions of the proteome of Leptospira are hypothetical proteins, necessitates their identification and characterisation to disclose their role in the pathogenesis. The presented *in-silico* characterisation approach analyzed the Leptospiral protein (LA_1016); in terms of physicochemical parameters, family relationships, motif and fold details, subcellular localisation nature and interaction networks. And annotated the hypothetical protein as an extracellular protein, containing a coiled-coil motif and an immunoglobulin-like fold. At the same time no domain, family or functional evidence could be identified with confidence. Virulent factor analysis, toxin prediction, antigen prediction and secondary structure identification were also carried out. And the protein was predicted as exotoxin and as an antigen. The revealed secondary structure proportions were; 21.19%, 10.80%, 28.12%, and 39.89% respectively for alpha helix, beta turn, extended strand and random coil. In addition, the epitope analysis explored the potential linear B-cell epitopes of the protein. The possibility of these antigenic epitopes as potential candidates for the peptide driven diagnostic and vaccine strategies can be considered. Besides, the revealed extracellular localisation nature points to the importance of the protein as a valuable diagnostic target with the potential of early diagnosis of Leptospirosis. Thus, the study gives the possibility of LA_1016 as a potential diagnostic target and vaccine candidate having implication in the pathogenesis of Leptospirosis.

Keywords: Leptospira; *In-silico* characterisation; LA_1016; Hypothetical proteins; Coiled-coil motif; B-cell epitopes

Introduction

Leptospires are corkscrew-shaped bacteria responsible for an emerging zoonotic infectious disease Leptospirosis [1]. Only a few portions of Leptospiral proteins are reviewed, besides the data regarding their full characterisation and their role in the biology of Leptospira is lacking. In the case of Leptospira interrogans serovar Lai, only approximately 11.5% of total proteins are reviewed. Currently, most of the total proteins are assigned as hypothetical proteins, which are proteins whose existence has been predicted through conceptual translation from the open reading frames but there is no experimental data validating their existence [2]. The hypothetical proteins constitute major fractions of the proteome of prokaryotes [3]. Thus, identification and characterisation of these proteins may lead to the invention of new functions, interacting cascades and their contribution to the proteome of an organism [4].

In-silico characterisation of hypothetical proteins is an indispensable tool which may help in the identification of new proteins and also for designing and focusing experiments on unknown proteins. There are so many publicly available prediction servers aimed for the *in-silico* prediction of protein characters such as physicochemical parameters, subcellular localisation, protein family, domain and fold identification, toxic property prediction, antigen detection, and homology modelling. Among the *in-silico* characterisation approaches, protein similarity searches are one of the first steps for the identification of unknown proteins. Can be employed through sequence alignment and homology detection by servers such as BLASTp and protBLAST [5]. Which enable the comparison of the query sequence with the database through sequence alignment and identify library sequences that match the query sequence?

Protein sequence information and physicochemical properties can be identified using Prot Param server. Proteins secreted through the classical protein secretory pathway are depend on a signal peptide, which is a stretch of 16-30 amino acid sequences at the amino terminus of the protein [6]. The signal peptide sequence and cleavage site can be identified through SignalP software which is based on a combination of several artificial neural networks. Some proteins are secreted through a signal peptide-independent, non-classical secretory pathway. However, their mechanism of secretion is unknown; the SecretomeP software enables the prediction of these classes of secreted proteins [7]. TMHMM, HMMTOP, and CCTOP are topology prediction tools enable the identification of the transmembrane helices presence within the protein sequences, which is a characteristic of membrane spanning proteins [8-10]. Protein subcellular localisation prediction is one strategy to infer the functions of unknown proteins. Prediction tools such as Cello, Psortb, PSL pred, SoSuiGramN, and MetaLocGramN are specialized for the prediction of protein subcellular localisation from amino acid sequences.

Citation: Samaha TH. *In-Silico* Characterisation of a Hypothetical Protein (LA_1016) of Leptospira Interrogans Serovar Lai Strain 56601. Austin J Proteomics Bioinform & Genomics. 2017; 4(2): 1022.

Table 1: Tools used for the *in-silico* characterisation of hypothetic protein LA_1016.

No.	Server Name	Reference	Purpose
1	BLASTp	Camacho et al. 2009	Similarity search
2	protBLAST	Altschul et al. 1999	
3	Muscle	Edgar. 2004	Multiple sequence alignment
4	BoxShade Server	http://sourceforge.net/projects/boxshade/	Alignment visualization
5	ProtParam	Gasteiger et al. 2005	Physicochemical characterisation
6	SignalP	Petersen et al. 2011	Signal peptide prediction
7	SecretomeP	Bendtsen et al. 2004	Non classical secretion prediction
8	Psortb	Yu et al. 2010	Subcellular localisation prediction
9	PSL pred	Bhasin et al. 2005	
10	Cello	Yu et al. 2006	
11	SOSUIGramN	Imai et al. 2008	
12	MetaLocGramN	Magnus et al. 2012	
13	TMHMM	Möller et al. 2001	Topology prediction
14	HMMTOP	Tusnády & Simon 2001	
15	CCTOP	Dobson et al. 2015	
16	Motif	Kanehisa et al. 2002	Motif discovery
17	Pfam	Finn et al. 2015	Family relationship identification
18	SuperFamily	Gough et al. 2001	Superfamily search
19	COILS	Lupas et al. 1991	Coiled-coil motif identification
20	PFP-FunDSeqE	Shen & Chou 2009	Fold recognition
21	InterPro	Hunter et al. 2009	Functional classification
22	STRING	Szklarczyk et al. 2015	Interaction network analysis
23	BTX pred	Saha & Raghava 2007	Toxin prediction
24	VICM pred	Saha & Raghava 2006	Virulent protein identification
25	MP3	Gupta et al. 2014	
26	BC pred	Chen et al. 2007	B-cell epitope prediction
27	IEDB	Parker et al. 2007	
28	VaxiJen	Doytchinova & Flower 2007	Vaccine candidate prediction
29	PSIPRED	McGuffin et al. 2000	Secondary structure prediction
30	SOPMA	Geourjon & Deléage 1995	

Table 2: Homologous proteins of LA_1016.

No.	Protein Name	Accession No. (NCBI)	Identity	E –Value
1	*Leptospira interrogans* serovar Copenhageni strain L1-130 (LIC_12641)	AAS71201	99%	0
2	*Leptospira interrogans* serovar Canicola strain LT1962 (LEP1GSC148_4102)	EMF71238	99%	0
3	*Leptospira interrogans* serovar Icterohaemorrhagiae strain Verdun (LEP1GSC116_1222)	EMO06089	99%	0
4	*Leptospira interrogans* serovar Copenhageni strain LT2050 (LEP1GSC150_2081)	EMG23115	98%	0
5	*Leptospira interrogans* serovar Manilae (LIMLP_13220)	AKP26801	97%	0

Motif and domain identification of unknown proteins can infer about the functions of proteins. Protein family relationships can be employed through Pfam database search, which is a collection of protein families [11]. SUPERFAMILY annotations are based on hidden Markov model domain search at the superfamily level [12]. Coiled-coils are structural motif having biological implications. Expasys' COILS server calculates the probability of coiled-coil conformation within a protein sequence. The fold recognition software's uses amino acid sequences to identify proteins belong to a fold; evolutionary distant but structurally similar proteins [13]. Prediction server like PFP-FunD SeqE identifies the fold pattern of proteins, which are the pattern of polypeptide chains in space [14]. Inter Pro protein functional characterisation depends on domain and repeats based family classification and sequence analysis [15]. And STRING database allows identification of functional and physical interactions of proteins from the predicted and known interactions.

Table 3: The subcellular localisation analysis.

No.	Analysis	Result
1	SignalP 4.0	Signal peptide probability: 0.985, SP cleavage site probability: 0.784, between position 21 and 22
2	SecretomeP 2.0	Positive SecP score (0.629)
3	TMHMM V.2.0	No transmembrane helices present
4	HMMTOP	No transmembrane helices present
5	CCTOP	Not a transmembrane protein
6	PSORT V.3.0	Unknown localisation
7	CELLO V.2.5	Outer membrane/extracellular localisation
8	SOSUIGramN	Extra-cellular localisation
9	PSL pred	Extra-cellular localisation
10	MetaLocGramN 0.99	Extra-cellular protein with probability 58.4%

Prediction server aiming identification of virulent proteins and toxic property of hypothetical proteins are valuable tools. Toxin prediction servers like BTX pred predicts the possibility of a toxic peptide from the amino acid or dipeptide composition. VICM pred and MP3 servers are designed for the functional characterisation of the proteins as virulent molecules. Reverse vaccinology is an important approach in vaccine research, depends on the prediction of protective antigens with the benefit of whole genome sequence data. VaxiJen is such an alignment independent bioinformatics prediction tool, prediction of a proteins' potential as vaccine candidate is based on the physicochemical properties [16].

Materials and Methods

Sequence, similarity identification

The sequence information the hypothetical protein (LA_1016) identified through MALDI-MS analysis (previous data) was collected from the NCBI database [17]. The sequence data was submitted to several prediction servers for the *in-silico* characterisation (Table 1). BLASTp and protBLAST servers were utilized for the sequence alignment and similarity searches [18,19]. Multiple sequence alignment of the protein with the homologous proteins were performed with the MUSCLE and visualised using the BoxShade server [20].

Physicochemical characterisation

Physical and chemical parameters such as molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index, and grand average of hydropathicity predictions were performed using Expasys' ProtParam tool [21].

Subcellular localisation analysis

Different prediction tools such as SignalP [22], SecretomeP [23], Psortb [24], PSL pred [25], Cello [26], SOSUIGramN [27], MetaLocGramN [28], TMHMM [8], HMMTOP [9] and CCTOP [10] were used for the topology prediction.

Motif, domain, fold, coil, family and superfamily identification

Protein motif search was carried out using Motif (Genome Net) server [29]. Pfam and SuperFamily database searches were done to assign the proteins' evolutionary relationships. Protein folding pattern recognition were carried out using PFP-FunD SeqE server (14). For the detection of coiled-coil conformation within the protein, the COILS server was employed [30]. Protein sequence analysis and classification serverInterPro was employed for the functional analysis of the protein [15]. And STRING 10.0 search was carried out for the identification of possible functional interaction network of the protein [31].

Virulent factor identification and vaccine candidate prediction

BTX pred toxin prediction, based on amino acid or dipeptide composition through a support vector machine (SVM) module was used for the prediction of the possibility of protein as a toxin [32]. For virulent factor analysis and annotation, VICM pred and MP3 (integrated SVM-HHM approach) servers were employed [33,34]. Antigen prediction was employed through VaxiJen server [16]. And BC pred and IEDB Parker hydrophilicity predictions were used for the identification of antigenic B-cell epitopes of the protein [35,36]. In addition a BLAST analysis against the human genome was performed for the detection of the autoimmune nature of the protein.

Secondary structure analysis

PSIPRED and SOPMA servers were used for the prediction of the proteins' secondary structure [37,38].

Results

Sequence and similarity information

The proteins' BLAST searches revealed the similarity with few proteins belong to the other serovars of pathogenic Leptospira interrogans, which was as shown in Table 2. And Figure 1 depicts the multiple sequence alignment of the protein with the homologous proteins. At the same time no similarity was detected against the non-pathogenic Leptospira biflexa proteome.

Physicochemical features

The protein consist of 722 amino acids, among the most abundant were Leu and Lys (69) followed by Glu and Ser (64), Asn (53), Ile (48), Gly and Val (40), Asp, Phe and Thr (36), Ala and Arg (31), Pro (30), Tyr (25), Gln (24), His and Met (10), Trp (5) and Cys (1). The calculated molecular weight was 82145.1Da and theoretical pI was 7.17. Total number of positively charged residues (Arg+Lys) and the total number of negatively charged residues (Asp+Glu) was found

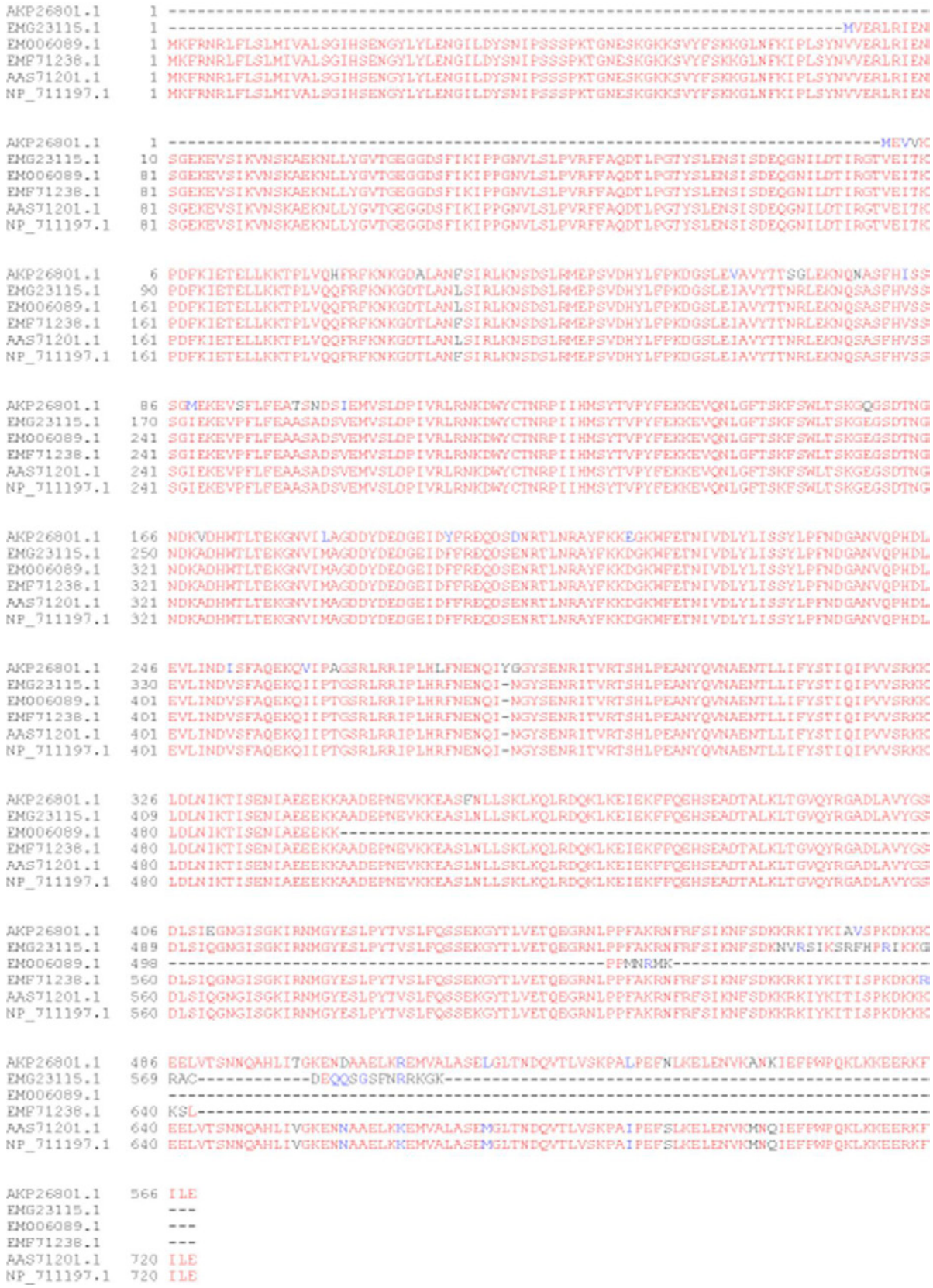

Figure 1: Multiple sequence alignment. MSA of the protein LA_1016 (NP_711197) with homologous proteins of Leptospira.

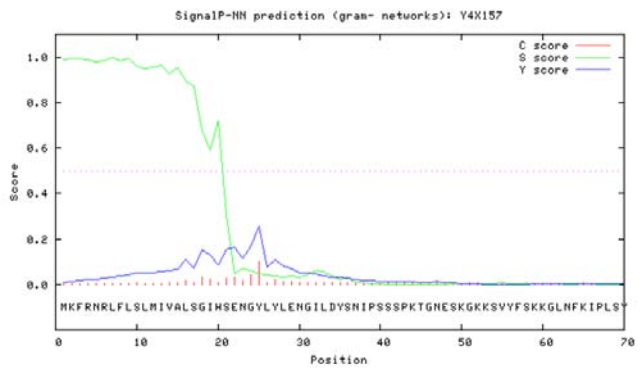

Figure 2: Signal peptide prediction (SignalP). Signal peptide probability: 0.985, Max cleavage site probability: 0.784 between pos. 21 and 22. Red, green and blue lines represents the predicted C score (cleavage site value), S score (signal peptide value), and Y score (combination of C- and S-scores) respectively.

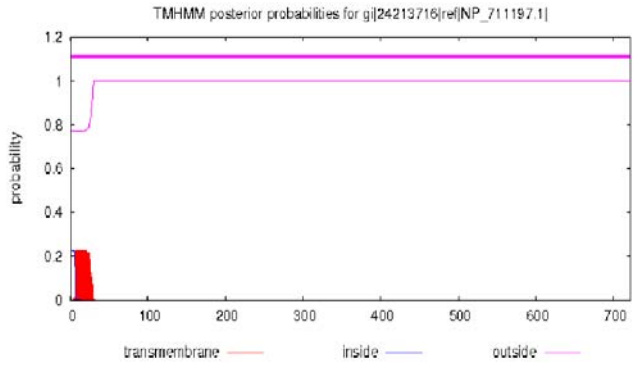

Figure 3: TMHMM transmembrane helix prediction. SEQUENCE Length: 722, Number of predicted TMHs: 0, Exp. number of AAs in TMHs: 4.70655, Exp. number of first 60 AAs: 4.70589, Total probability of N-in: 0.225, TMHMM2.0 sequence outside 1-722.

to be 100. The computed instability index, aliphatic index and grand average of hydropathicity (GRAVY) were 28.69, 83.56, and -0.572 respectively. Extinction coefficient predicted at 280nm was 64750. Protein half -life computed was found to be 30 hours in mammalian reticulocytes (*in vitro*), > 20 hours and >10 hours in yeast and E. coli (*in vivo*) respectively. And the molecular formula of protein was identified as C3674H5808N986O1125S11.

Subcellular localisation nature

Subcellular localisation analysis results presented in Table 3, displayed the proteins' extracellular localisation nature. The graphical representation of SignalP and TMHMM prediction results were displayed in Figure 2 & 3.

Motif, domain, fold, coil, family and superfamily and functional details

Motif search by Motif server (genome net) identified a galactokinase against NCBI cdd, but no homologs in Prosite or Pfam databases. Domain search on Pfam was negative. Moreover, there was no significant hit against Superfamily search. Fold pattern recognition by PFP-FunDSeqE tool revealed the presence of an 'immunoglobulin-like' fold within the protein sequence. And COILS server revealed the

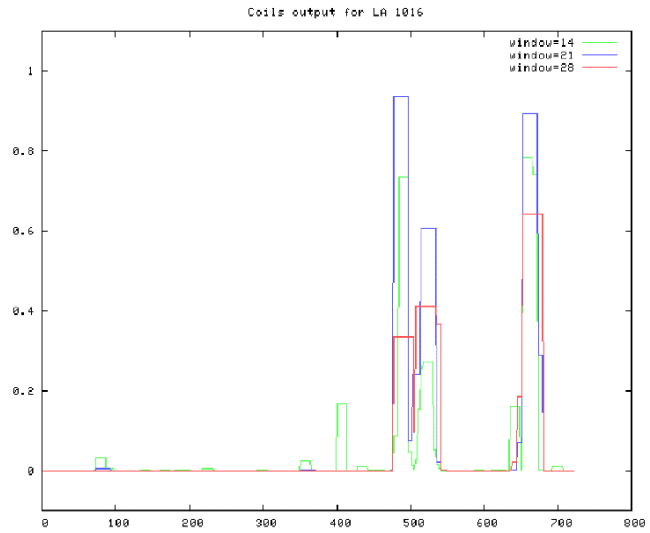

Figure 4: Coils prediction. Depicts the heptads corresponding to the residue windows 14 (green), 21(blue) and 28 (red).

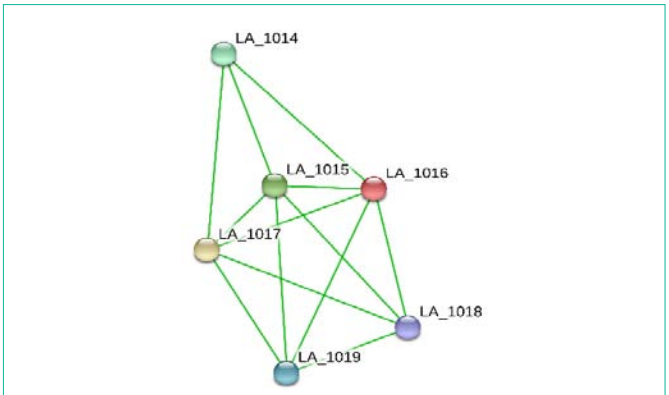

Figure 5: Protein-protein interaction network of LA_1016. Shows the predicted functional partners (LA_1017, LA_1015, LA_1014, LA_1019 and LA_1018).

presence of a coiled-coil motif (Figure 4). There was no functional relation information by InterPro functional analysis. The predicted protein interaction network through the STRING analysis as shown in Figure 5. And the identified functional partners with scores were; LA_1017 (0.859), LA_1015 (0.859), LA_1014 (0.469), LA_1019 (0.445), and LA_1018 (0.445).

Virulent factor identification and vaccine candidate prediction

The protein possibility as an exotoxin was predicted through a support vector machines (SVM) module based on amino acid sequences with 96% accuracy through BTX pred server. VICM pred analysis classified the protein as metabolism related molecule through a support vector machine based pipeline. AndVaxiJen prediction presented the protein as an antigen with a significant score 0.6185 (threshold 0.4). The explored potential linear B-cell epitopes of the protein were listed in Table 4. And Figure 6 illustrates the antigenic epitopes of the protein identified through the Parker Hydrophilicity method. Moreover, no similarity result was showed against human proteome BLAST.

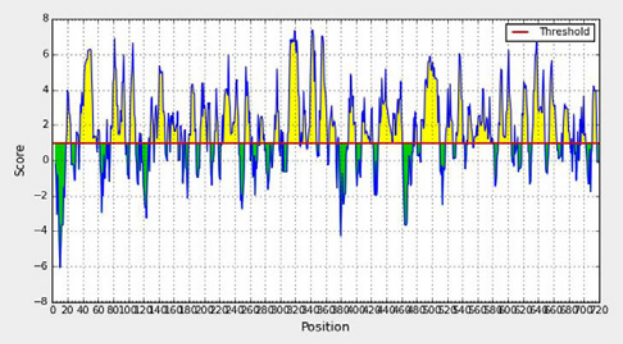

Figure 6: B-cell epitope prediction. The predicted linear B-cell epitopes of the protein.

Table 4: Predicted linear B-cell epitopes of the protein.

No.	Position	Epitope	Length	Score
1	623	KKRKIYKITISPKDKKKEEL	20	1
2	489	ENIAEEEKKAADEPNEVKKE	20	1
3	34	DYSNIPSSSPKTGNESKGKK	20	0.999
4	307	SWLTSKGEGSDTNGNDKADH	20	0.994
5	352	REQDSENRTLNRAYFKKDGK	20	0.988
6	101	GVTGEGGDSFIKIPPGNVLS	20	0.983
7	560	DLSIQGNGISGKIRNMGYES	20	0.959
8	156	VEITKPDFKIETELLKKTPL	20	0.924
9	699	VKMNQIEFPWPQKLKKEERK	20	0.911
10	598	ETQEGRNLPPFAKRNFRFSI	20	0.909
11	282	IIHMSYTVPYFEKKEVQNLG	20	0.865

Secondary structure

The SOPMA secondary structure prediction server analysis revealed the proportions of alpha helix, beta turn, extended strand and the random coil of protein as 21.19%, 10.80%, 28.12%, and 39.89% respectively. Figure 7 illustrates the secondary structure alignments of the protein, identified through the PSI PRED server.

Discussion

The sequence information collected from the NCBI database revealed the protein as "uncharacterised" protein. Further sequence alignment and comparison using the BLASTp and protBLAST searches showed the similarity to only a few proteins belongs to the pathogenic candidate Leptospira interrogans. Furthermore, the absence of sequence similarity to the non-pathogenic candidate Leptospira biflexa can be implicated to the virulent nature of the protein. ProtParam physicochemical characterisation classified the protein as stable with the instability index 28.69 [39], which is an estimation of stability of protein in a test tube [4]. The identified high aliphatic index data gives an indication of proteins' stability over a wide temperature range, and extinction coefficient value correlates to the concentration of Cys, Trp, and Tyr. Moreover, the negative GRAVY (-0.572) indicates the hydrophilic and the soluble nature of the protein [40,41].

The proteins' signal peptide and cleavage site prediction (between 21and 22 positions) by SignalP analysis gives an indication of the secretory nature of the protein. Further TMHMM, HMMTOP, CCTOP results strengthen this data by avoiding the possibility of trans membrane helix presence within the protein sequence. The SignalP and SecretomeP results point to the secretion of protein though classical secretory pathways. Besides other subcellular localisation tools such as Cello, Psortb, SoSuiGramN, and MetaLocGramN were also explained the subcellular localisation of protein as extracellular. Which may give insights into the functional relations of the protein.

In case of this protein no domain features and family relationships could be predicted with confidence, which relates to the unknown nature of the protein. Concurrently the analysis showed the presence of an immunoglobulin-like fold and coiled-coil motif. Immunoglobulin-like (Ig-like) fold proteins have a role in immune functions. They can participate in interactions with other immunoglobulin-like domains via their beta sheets (a pair of Greek-key beta barrel) [42]. So protein fold recognition may allow the understanding of protein structure and function; based on the fact that proteins with different sequence information can adopt similar structures and similar functions [43]. Likewise, coiled-coil motifs are structural units formed from the supercoiling of alpha helices, with implications in biological functions such as regulation of gene expression [44]. Facilitates the protein-protein interactions by cementing the interfaces between the proteins. Its predictions are helpful for the identification of protein domain boundaries and potential protein interacting partners [45].

Protein-protein interaction identification helps to imply the functions of the proteins based on their interacting network [4]. The SRING analysis showed the interactions of this protein with other Leptospiral proteins such as LA_1017, LA_1015, LA_1014, LA_1019, and LA_1018. Among, LA_1017 is a TPR repeat containing protein. The TPR repeat motifs mediate the protein-protein interactions and modulate the assembly of multi protein complexes. Thus the interaction of the protein with these interacting partners and involvements in their functions are remained to be studied.

Toxin prediction analysis revealed the possibility of protein as an exotoxin. Exotoxins are soluble proteins secreted by the bacteria, diffuse through the organism to act at distant places from the site of infection. Exotoxins include neurotoxins, enterotoxins and cytotoxins and are involved in various biological functions [32]. As an exotoxin, this proteins' ability of binding to specific receptors on the cell's plasma membrane and pore formation remains to be studied, which may open up extra dimensions in leptospirosis research. Furthermore, the virulence factor analysis presented the protein as a metabolism related molecule. Protein possibility as an antigenic candidate from the result of VaxiJen and less chance of autoimmune induction gives an indication of this proteins' potential for the consideration as a vaccine candidate. In addition, the revealed B-cell epitopes of the protein are antigenic regions with the potential of peptide driven diagnostic and vaccine candidate efficiency. Even though *in vitro* confirmation of the annotated virulent characteristics are mandatory for the full characterisation of the protein.

The presented *in-silico* characterisation data gives an indication of the proteins' significance in the biology of Leptospira. The secondary structure information may give insights into the higher order structure and functional annotation of the protein. And the revealed extracellular nature points to the possibility of this protein

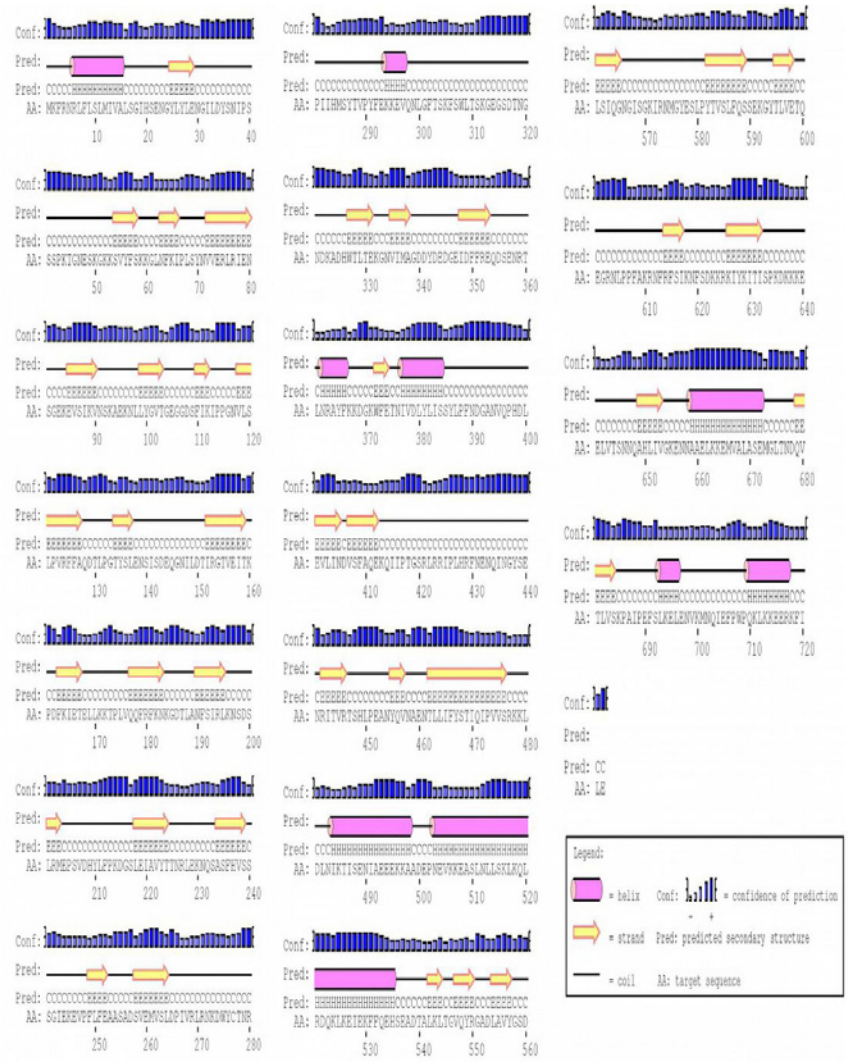

Figure 7: LA_1016 secondary structure. Shows the alignments of alpha helix, beta turn, extended strand and random coil (PSI PRED analysis).

as a valuable diagnostic target. As extracellular proteins are readily exposed to the host immune system, this proteins' involvement in the immuno-pathogenesis of Leptospirosis is to be discovered. Moreover, the development of diagnostic assays targeting these protein or antigenic peptides may lead to the early diagnosis of Leptospirosis.

Conclusion

Identification and characterisation of new virulent molecules will strengthen basic knowledge on Leptospirosis. The *in-silico* characterised hypothetical protein LA_1016 is an unknown protein with less similarity to existing characterised or annotated proteins. And the proteins' revealed characteristics such as extracellular nature, coiled-coil motif presence, immunoglobulin-like fold presence, toxic possibility and antigenic nature emphasise the significance of this protein. The possibility of this protein as a potential vaccine candidate and diagnostic target also could be considered. Further in-vitro confirmation of the annotated characteristics may lead to the identification of a novel protein, having an impact in the pathogenesis of Leptospirosis. So extended research has to be carried out to experimentally validate these possibilities and to find out the proteins' role in the Leptospiral biology.

Acknowledgements

The author gratefully thanks Dr. P. Vijayachari, (Director, Regional Medical Research Centre, ICMR, Port Blair) for giving support to conduct the study.

References

1. WHO. Human leptospirosis: guidance for diagnosis, surveillance and control. WHO Libr. 2003; 45: 1-109.
2. Ijaq J, Chandrasekharan M, Poddar R, Bethi N. Annotation and curation of uncharacterized proteins- challenges. Front Genet. 2015; 6: 1-7.
3. Shahbaaz M, Hassan MI, Ahmad F. Functional annotation of conserved hypothetical proteins from Haemophilus influenzae Rd KW20. PLoS One. 2013; 8: e84263.
4. Mohan R, Venugopal S. Computational structural and functional analysis of hypothetical proteins of Staphylococcus aureus. Bioinformation. 2012; 8: 722-728.
5. Remmert M, Biegert A, Hauser A, Söding J. HHblits: lightning-fast iterative

protein sequence searching by HMM-HMM alignment. Nat Methods. 2011; 9: 173-175.

6. Wang G, Chen H, Xia Y, Cui J, Gu Z, Song Y, et al. How are the non-classically secreted bacterial proteins released into the extracellular milieu? Current Microbiology. 2013; 67: 688-695.

7. Bendtsen JD, Kiemer L, Fausbøll A, Brunak S. Non-classical protein secretion in bacteria. BMC Microbiol. 2005; 5: 58.

8. Möller S, Croning MD, Apweiler R. Evaluation of methods for the prediction of membrane spanning regions. Bioinformatics. 2001; 17: 646-653.

9. Tusnády GE, Simon I. The HMMTOP transmembrane topology prediction server. Bioinformatics. 2001; 17: 849-850.

10. Dobson L, Reményi I, Tusnády GE. CCTOP: a Consensus Constrained TOPology prediction web server. Nucleic Acids Res. 2015; 43:W408-W412.

11. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2015; 44: D279-D285.

12. Gough J, Karplus K, Hughey R, Chothia C. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. J Mol Biol. 2001; 313: 903-919.

13. Exarchos TP, Papaloukas C, Lampros C, Fotiadis DI. Mining sequential patterns for protein fold recognition. J Biomed Inform. 2008; 41: 165-179.

14. Shen HB, Chou KC. Predicting protein fold pattern with functional domain and sequential evolution information. J Theor Biol. 2009; 256: 441-446.

15. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, et al. InterPro: The integrative protein signature database. Nucleic Acids Res. 2009; 37: D211-D215.

16. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics. 2007; 8: 4.

17. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2013; 44: D7-D19.

18. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009; 10: 421.

19. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990; 215: 403-410.

20. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004; 32:1792-1797.

21. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins M., Appel R., et al. Protein Identification and Analysis Tools on the ExPASy Server. In: The Proteomics Protocols Handbook. 2005; 571-607.

22. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011; 8: 785-786.

23. Bendtsen JD, Jensen LJ, Blom N, Von Heijne G, Brunak S. Feature-based prediction of non-classical and leaderless protein secretion. Protein Eng Des Sel. 2004; 17: 349-356.

24. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, et al. PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics. 2010; 26: 1608-1615.

25. Bhasin M, Garg A, Raghava GPS. PSLpred: Prediction of subcellular localization of bacterial proteins. Bioinformatics. 2005; 21: 2522-2524.

26. Yu C-S, Chen Y-C, Lu C-H, Hwang J-K. Prediction of protein subcellular localization. Proteins. 2006; 64: 643-651.

27. Imai K, Asakawa N, Tsuji T, Akazawa F, Ino A, Sonoyama M, et al. SOSUI-GramN: high performance prediction for sub-cellular localization of proteins in Gram-negative bacteria. Bioinformation. 2008; 2:417-421.

28. Magnus M, Pawlowski M, Bujnicki JM. MetaLocGramN: A meta-predictor of protein subcellular localization for Gram-negative bacteria. In: Biochimica et Biophysica Acta - Proteins and Proteomics. 2012; 1425-1433.

29. Kanehisa M, Goto S, Kawashima S, Nakaya A. The KEGG databases at GenomeNet. Nucleic Acids Res. 2002; 30:42-46.

30. Lupas a, Van Dyke M, Stock J. Predicting coiled coils from protein sequences. Science. 1991; 252:1162-1164.

31. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: Protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015; 43: D447-452.

32. Saha S, Raghava GPS. BTXpred: prediction of bacterial toxins. In Silico Biol. 2007; 7: 405-412.

33. Saha S, Raghava GPS. VICMpred: An SVM-based Method for the Prediction of Functional Proteins of Gram-negative Bacteria Using Amino Acid Patterns and Composition. Genomics Proteomics Bioinformatics. 2006; 4: 42-47.

34. Gupta A, Kapil R, Dhakan DB, Sharma VK. MP3: A software tool for the prediction of pathogenic proteins in genomic and metagenomic data. PLoS One. 2014; 9:e93907.

35. Chen J, Liu H, Yang J, Chou KC. Prediction of linear B-cell epitopes using amino acid pair antigenicity scale. Amino Acids. 2007; 33: 423-428.

36. Parker JMR, Guo D, Hodges RS. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. Biochemistry. 1986; 25: 5425-5432.

37. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. Bioinformatics. 2000; 16: 404-405.

38. Geourjon C, Deléage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci. 1995; 11: 681-684.

39. Guruprasad K, Reddy B V, 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. Protein Eng. 1990; 4: 155-161.

40. Sharma AR, Chakraborty C, Lee SS, Sharma G, Yoon JK, Doss CGP, et al. Computational biophysical, biochemical, and evolutionary signature of human R-spondin family proteins, the member of canonical Wnt/b-catenin signaling pathway. Biomed Res Int. 2014; 2014:1-22.

41. Sarwar MW, Saleem IB, Ali A, Abbas F. Insilico Characterization and Homology Modeling of Arabitol Dehydrogenase (ArDH) from Candida albican. Bioinformation. 2013; 9: 952-957.

42. P. Bork, L. Holm CS. The Immunoglobulin Fold: Structural Classification, Sequence Pattern and Common Core. J Mol Biol. 1994; 1: 309-320.

43. Hou J, Sims GE, Zhang C, Kim S-H. A global representation of the protein fold space. Proc Natl Acad Sci U S A. 2003; 100: 2386-2390.

44. Mason JM, Arndt KM. Coiled coil domains: Stability, specificity, and biological implications. ChemBioChem. 2004; 5: 170-176.

45. Walshaw J, Woolfson DN. Socket: a program for identifying and analysing coiled-coil motifs within protein structures. J Mol Biol. 2001; 307: 1427-1450.

Austin J Proteomics Bioinform & Genomics - Volume 4 Issue 2 - 2017
ISSN : 2471-0423 | www.austinpublishinggroup.com

Citation: Samaha TH. *In-Silico* Characterisation of a Hypothetical Protein (LA_1016) of Leptospira Interrogans Serovar Lai Strain 56601. Austin J Proteomics Bioinform & Genomics. 2017; 4(2): 1022.