



# In silico characterization of *Leptospira interrogans* DNA ligase A and delineation of its antimicrobial stretches

Prasanta Kumar Koustasa Mishra<sup>1</sup> · Ramadevi Nimmanapalli<sup>1</sup>

Received: 24 July 2019 / Accepted: 11 September 2019 / Published online: 23 November 2019  
© Università degli studi di Milano 2019

## Abstract

**Purpose** In the present study, an attempt is being made to characterize the DNA ligase A (LigA) of *Leptospira interrogans* by computational methods.

**Methods** Several prediction servers (SwissProt, MoBye, TMHMM, PSIPRED, SignalP, etc.) were used to predict and interpret the physico-chemical parameters associated with LigA. A three-dimensional (3D) structure of the protein was created by homology-based modeling (I-TASSER). DNA-binding regions (PATCHDOCK) and interactome (STRINGS) of the protein were also predicted. A phylogenetic tree was constructed by MEGA version X. Finally, amino acid residues with antimicrobial activity were determined from the LigA sequence by AntiBP server.

**Results** Domains responsible for oligonucleotide binding (OB), BRCT (BRCA1 carboxy-terminal), and motifs like helix hairpin helix (HhH) were found to be present in the protein designating the super family it belongs to. Moreover, consensus residues, i.e., -KX/IDG-responsible for adenylation, are also found to be conserved within the amino acid sequence. In silico mutational analysis suggested that replacing any of the charged residues in the consensus (K or D) can lead to catalytic instability of the enzyme. Further, the protein was scanned for antimicrobial peptide (AMPs). Ten different stretches were found to have a potential bactericidal effect with significant scores.

**Conclusions** LigA of *Leptospira interrogans* is an acidic protein rich in alpha helixes which also contain 10 potential antimicrobial peptides in its amino acid sequence.

**Keywords** DNA ligase · Structural co-ordinates · In silico · Antimicrobial peptides · *Leptospira*

## Introduction

DNA ligases (E.C 6.5.1.1) play a very crucial role in the energy-dependent sealing of the phosphodiester bond of DNA/RNA molecules (Tomkinson et al. 2006). Starting from DNA replication to repair processes, it participates in nick sealing of the interrupted backbone of nucleic acid chain (Shuman 2009). The ligases can be classified broadly into two groups such as NAD<sup>+</sup>-dependent ligases (bacteria) and ATP-dependent ligases (eukaryotes and viruses) (Wilkinson et al. 2001). LigA is a family of NAD<sup>+</sup>-dependent ligases

found in bacteria (Pergolizzi et al. 2016). Generally, the ligation by DNA ligases requires either blunt-ended fragments or sticky-ended complementary overhangs (Bauer et al. 2017). The enzyme possesses typical residues where adenylation occurs. The activated enzyme subsequently initiates the ligation process (Gajiwala and Pinko 2014).

Structurally, the enzyme can be divided into three main domains as an oligonucleotide-binding domain (OB), the NTPase domain, and the BRCT domain. Apart from these, it also contains the helix-hairpin-helix domain and the tetra cysteine Zn-binding domain (Pergolizzi et al. 2016). The enzyme binds to DNA in the form of a C-shaped clamp (Nandakumar et al. 2007). The OB, NTPase, Zn finger, and HhH domains make contact with DNA while BRCT does not. A 19-nucleotide region near to the breaking point in the DNA backbone serves as the target site for the enzyme's OB domain during binding. The beta barrel-rich OB domain adopts a conformation with a concave surface within such that it can make contact with the minor groove of the DNA (Wang et al.

✉ Prasanta Kumar Koustasa Mishra  
prasantmodel@gmail.com

<sup>1</sup> Department of Veterinary Physiology and Biochemistry, Faculty of Veterinary and Animal Sciences, Rajiv Gandhi South Campus, Banaras Hindu University, Barkachha, Mirzapur, Uttarpradesh 231001, India

2009). An absolute absence of NAD<sup>+</sup>-dependent ligase among eukaryotes makes it a suitable target for drug discovery against microbes. The hydrophobic residues present in the DNA-binding tunnel of *E. coli* have been targeted by an analog-like 2-methylthio ATP which potentially inhibits the adenylation of ligase A (Miesel et al. 2007). Another group of chemical, i.e., pyridopyrimidine, was found to be able to block the active site of LigA of *Enterococcus* species (Swift and Amaro 2009). Much newer therapeutics can be designed by knowing the structural and chemical aspects of the enzyme ligase A. The structural and functional details of many bacterial LigA are unavailable. The LigA of *Leptospira interrogans* is one of them that warrant further scientific investigations.

*Leptospira* is a gram-positive (+) spirochete which causes leptospirosis otherwise known as Weil's disease (Inada et al. 1916). The genus *Leptospira* can be broadly divided into two species as *L. interrogans* (pathogenic) and *L. biflexa* (saprophytic) (Johnson 1984). The organism is responsible for more than 50,000 cases of infection in developing countries. Upon infection, the mortality rate may be up to 7–10% (Rao et al. 2003; Yaakob et al. 2015; Costa et al. 2015). *Leptospira* can infect both humans and animals. The resulting outcome may be exhibited in the form of renal failure, abortion, or in severe cases, can lead to the death of the infected animals. Being a highly pathogenic organism of zoonotic importance, it needs special attention from human and animal health perspective. Considering the hazardous effect of *Leptospira* infection, novel strategies like mechanism-based inhibition can be of great use in controlling the leptospirosis. Again, this approach of control can be highly benefited by a detailed understanding of the characteristics of different enzymes playing vital role(s) in survival as well as escape of these bacteria from the host-mediated response. Study on the structural and functional aspects of LigA of *L. interrogans* can help in addressing the disease, leptospirosis. In the present study, in silico approaches have been made to characterize the ligase A protein of *Leptospira interrogans*.

## Materials and methods

### Retrieval of protein sequence

The amino acid sequence of *Leptospira interrogans* LigA was retrieved from UniProt Knowledgebase (UniProtKB) (<https://www.uniprot.org/>) server against the UniProt accession number Q72MA6. The sequence was downloaded in FASTA format and subjected to protein BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>) analysis.

**Table 1** Physical and chemical properties of the protein determined by Proto param server

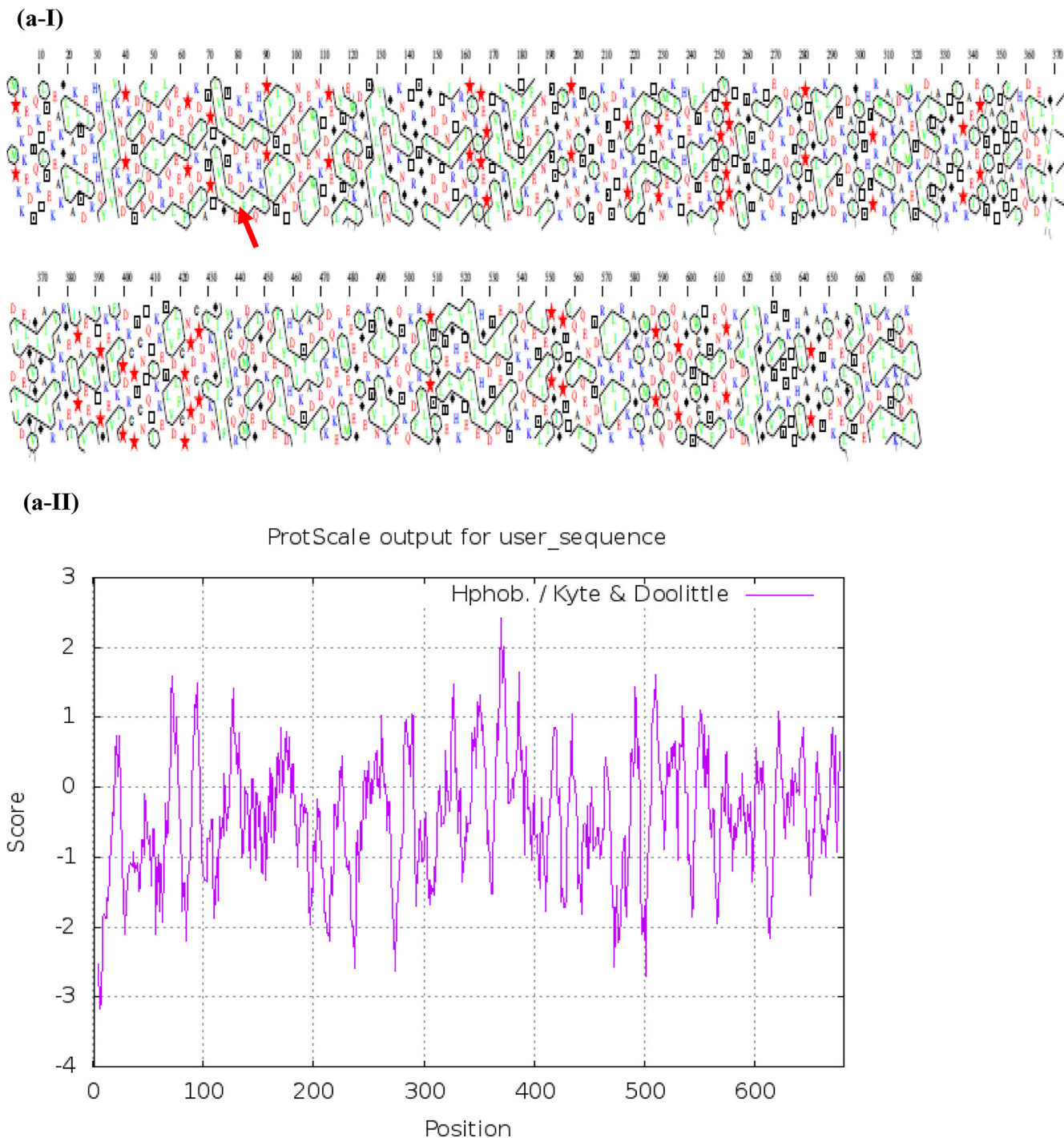
Properties	Values
Number of amino acids	681
Molecular weight	76,913.14 Da
Theoretical pI	6.44
Extinction coefficients	53,665
Estimated half-life	> 10 h ( <i>Escherichia coli</i> , in vivo)
Instability index (II)	36.19
Aliphatic index	89.19
Grand average of hydropathicity (GRAVY)	− 0.427

### Assessment of physico-chemical properties

The protein was analyzed by ProtParam (<https://web.expasy.org/protparam/>) server for various inherent characteristics like molecular weight, isoelectric pH(pI), half-life, stability, and hydropathy index. The hydrophobic patches of LigA were derived from the hydrophobic cluster analysis application (HCA 1.0.2) of the Mobyle portal (Alland et al. 2005). Further, hydrophobic residues were assessed by the ProtoScale server (<https://web.expasy.org/protscale/>) with Kyte and Doolittle algorithm. The window size was kept 15 and the scale was not normalized. The result was presented as a graph where y-axis represented a window range and x-axis contains amino acid residues. The aggregate-forming tendency of the protein in terms of its amino acid composition was determined by the Aggrescan3D server (Zambrano et al. 2015). Signal sequence and transmembrane domains were searched by submitting the sequence in SignalP (Armenteros et al. 2019), TMHMM server (Sonnhammer et al. 1998), OCTOPUS (<http://octopus.cbr.su.se/index.php?about=OCTOPUS>) (Viklund et al. 2008), and PSIPRED Workbench (<http://bioinf.cs.ucl.ac.uk/psipred/>) (Buchan and Jones 2019). Metal-binding regions were predicted by Metsite (Protein-metal ion contact prediction) of the PSIPRED Workbench server. The query was made against magnesium (Mg) binding by selecting a false positive rate of 1%. Identification of domains for the protein was done by SMART (<http://smart.embl-heidelberg.de/>) (Schultz et al. 1998), motifFinder (<https://www.genome.jp/tools/motif/>), and pfam (<https://pfam.xfam.org/>) (El-Gebali et al. 2018). The migration of protein as per their pI and charge was assessed by a virtual 2D predictor JVirGel v2.0 (Hiller et al. 2006).

### Derivation of secondary structure of the protein

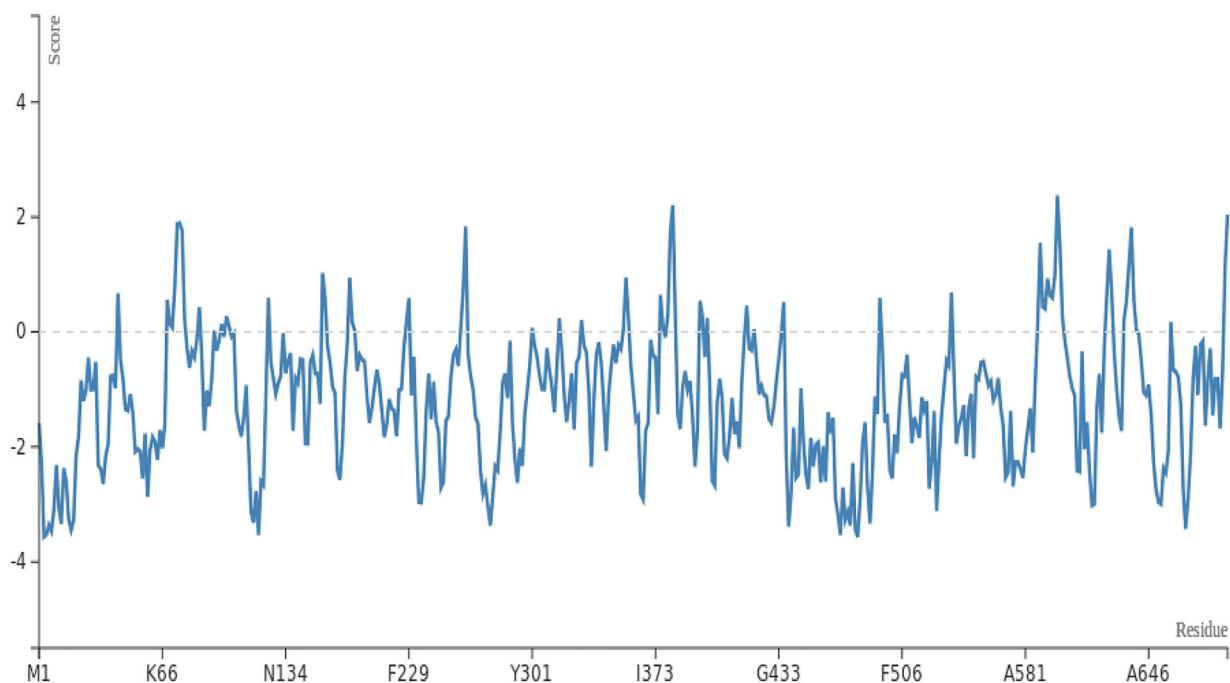
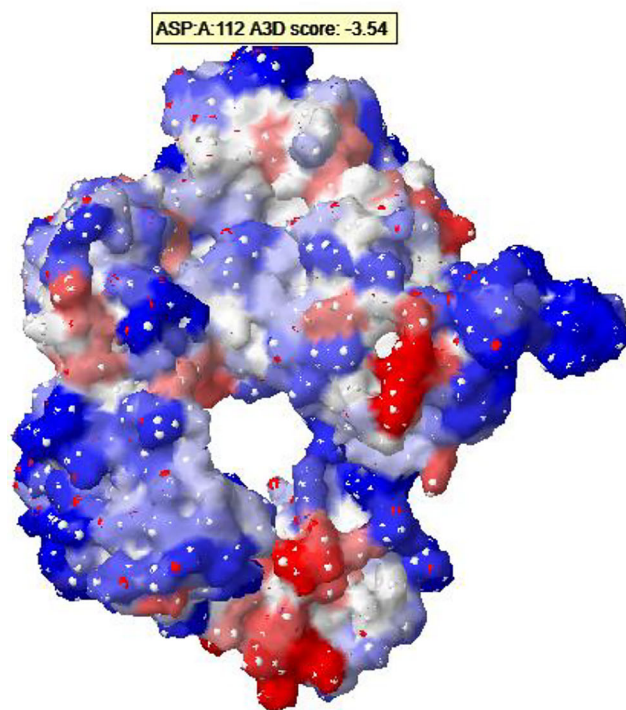
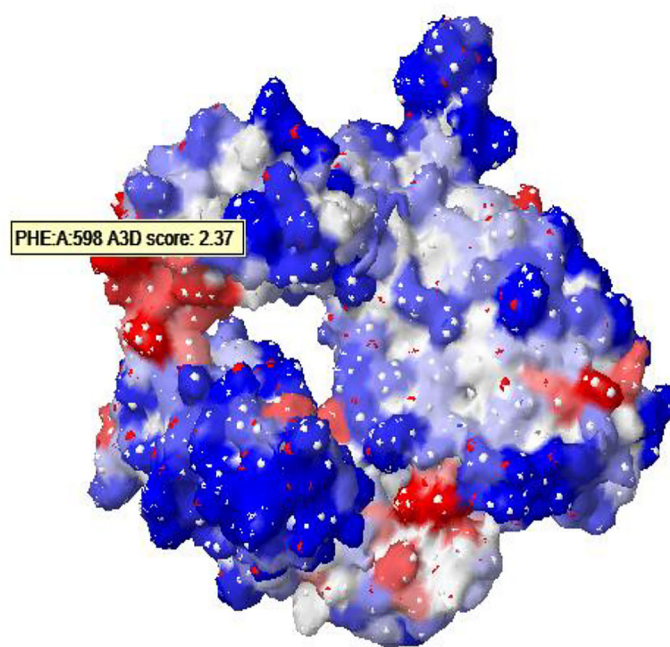
SOPMA ([https://npsa-prabi.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=NPSA/npsa\\_sopma.html](https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=NPSA/npsa_sopma.html)) and PSIPRED Workbench (<http://bioinf.cs.ucl.ac.uk/psipred/>) servers were used for



**Fig. 1** Characterization of Leptospiral ligase A by in silico analysis: The hydrophobic regions in the protein are indicated by as patches (a-I) and the hydrophobicity residues are plotted graphically (a-II) where  $-3$  represents the most hydrophilic residues, while  $3$  represents the most hydrophobic ones. Amino acids imparting solubility and aggregating formation ability to LigA are presented with their relative scores in the figure (a-III), where positive and negative values indicate aggregate and soluble tendencies of the residues. (a-iv) and (a-v) The position of most soluble and aggregate-forming residues respectively. There is no

transmembrane stretch present in the protein as predicted by TMHMM (b-1) and OCTOPUS servers (b-II). The protein was found to be non-membrane bound which confirms its cytosolic location (b-III, b-IV). This is further supported by the absence of any signal sequences as predicted by SignalP 5.0 server (c-I). The possible magnesium-binding regions are highlighted in red (c-II). (d-I) and (d-II) The domains present in the protein. Migration of the protein in a 2D gel electrophoresis is presented (e) which supports the physical characters predicted by the Swiss-Prot module



**(a-III)****(a-IV)****(a-V)****Fig. 1** (continued)

prediction of total secondary structure content in the protein. Default parameters were opted for the analysis. Protein folding is a process of the adoption of random conformations by several permutations (Honig 1999). Ramachandran's plot is an effective way to analyze the

dihedral angles present in the protein and thus is predictive for the restrictive conformations that a protein can and cannot adopt while folding into higher order structures (Zhou et al. 2011). The Ramchandran's plot was predicted by SAVESv5.0 (<http://servicesn.mbi.ucla.edu/SAVES/>) (Wodak et al. 2012).



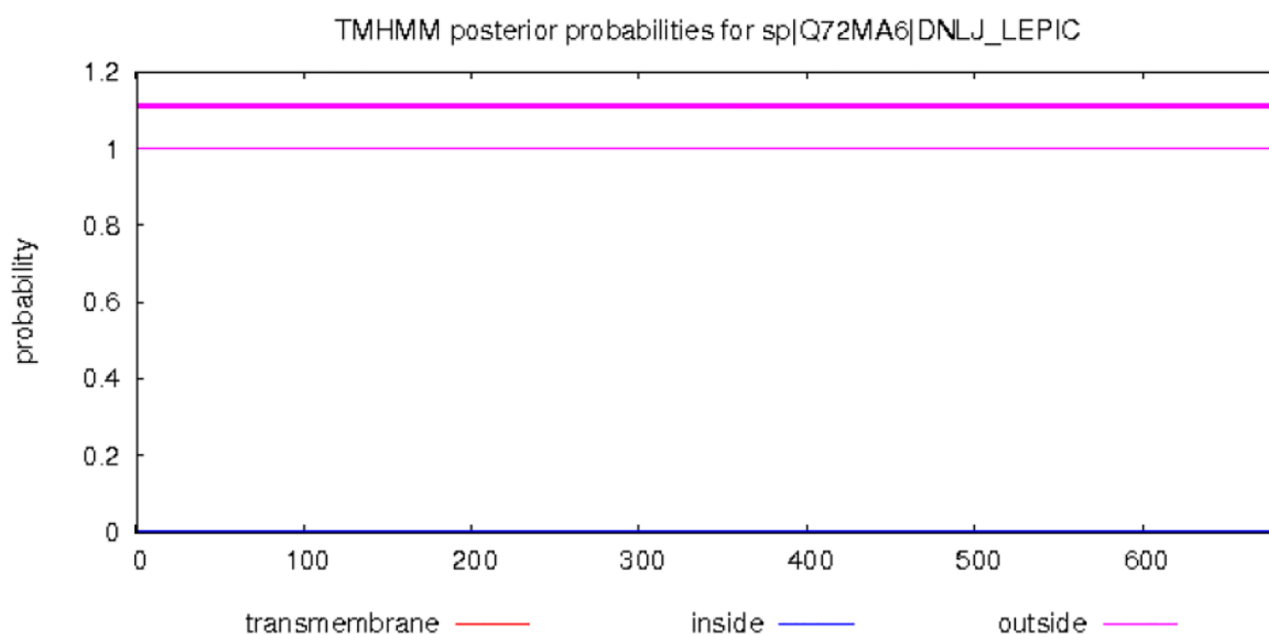
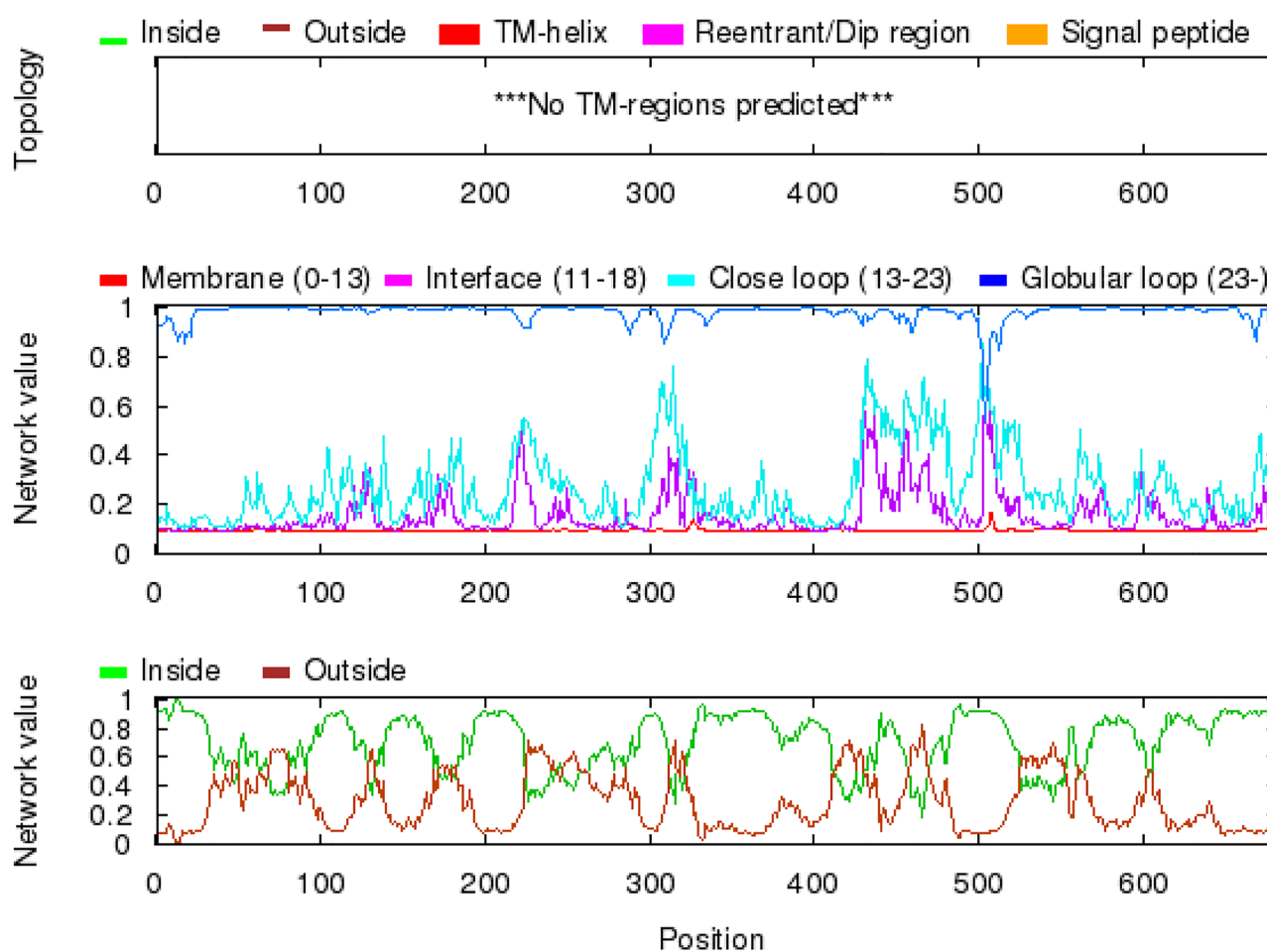
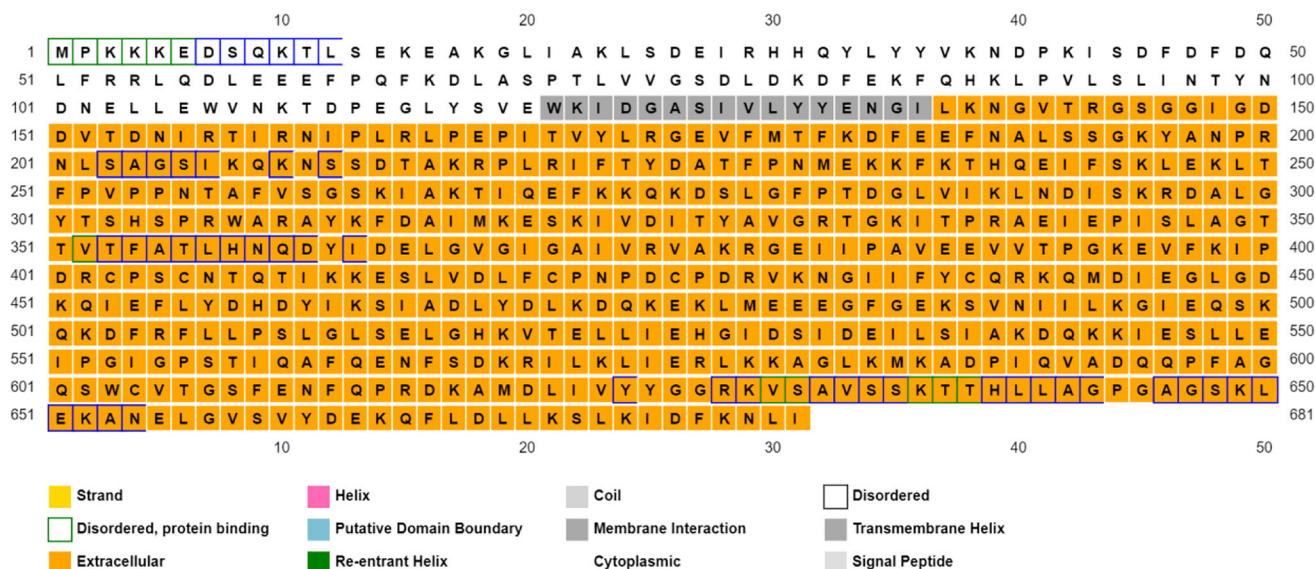
**(b-I)****(b-II)**

Fig. 1 (continued)

## (b-III)



## (b-IV)

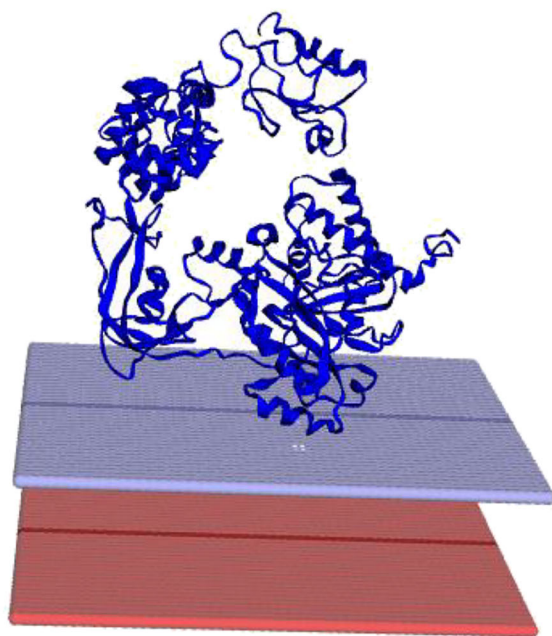


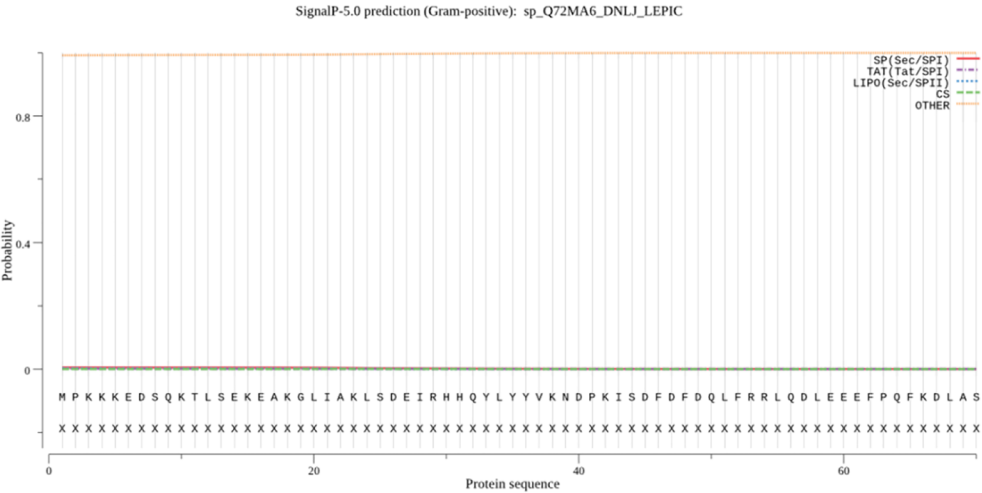
Fig. 1 (continued)

### Prediction of three-dimensional (3D) conformation and validation of the predicted structure

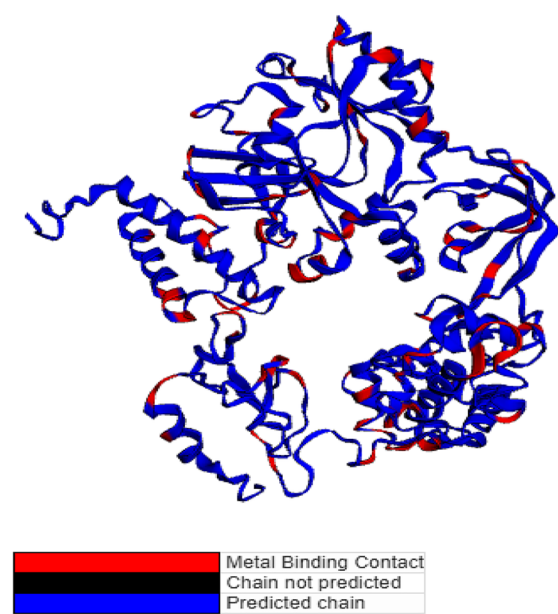
It was ascertained that the X-ray diffraction-derived crystal structure of the candidate protein of *Leptospira* was unavailable in the PDB (<http://www.rcsb.org/>) structure database. Though 15 different structures are present in the SWISS-MODEL (<https://swissmodel.expasy.org/>) server, all of them

are in silico predicted based on homology models. So, another attempt was made to predict the structure of *L. interrogans* LigA by a different prediction module. The structure was predicted by submitting the protein sequence in I-TASSER (<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>) (Yang et al. 2015) server for protein homology modeling. I-TASSER uses LOMET as the multiple threading gateway. Out of the various templates used during threading, the structure of *E. coli* DNA ligase (PDB ID-2owoA) was intensively

(c-I)



(c-II)



(d-I)

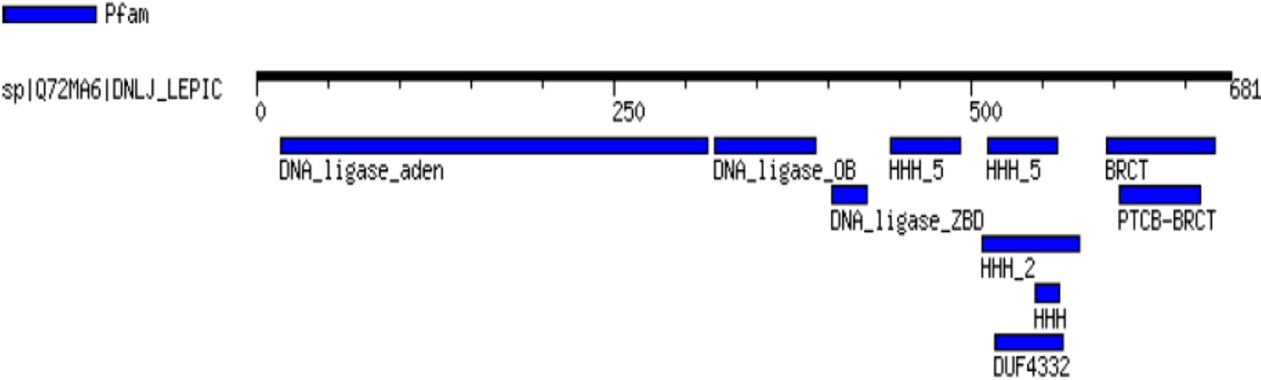
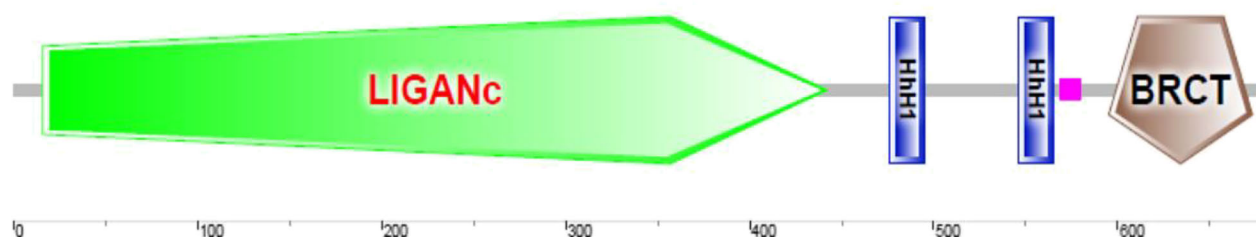


Fig. 1 (continued)



## (d-II)



## (e)

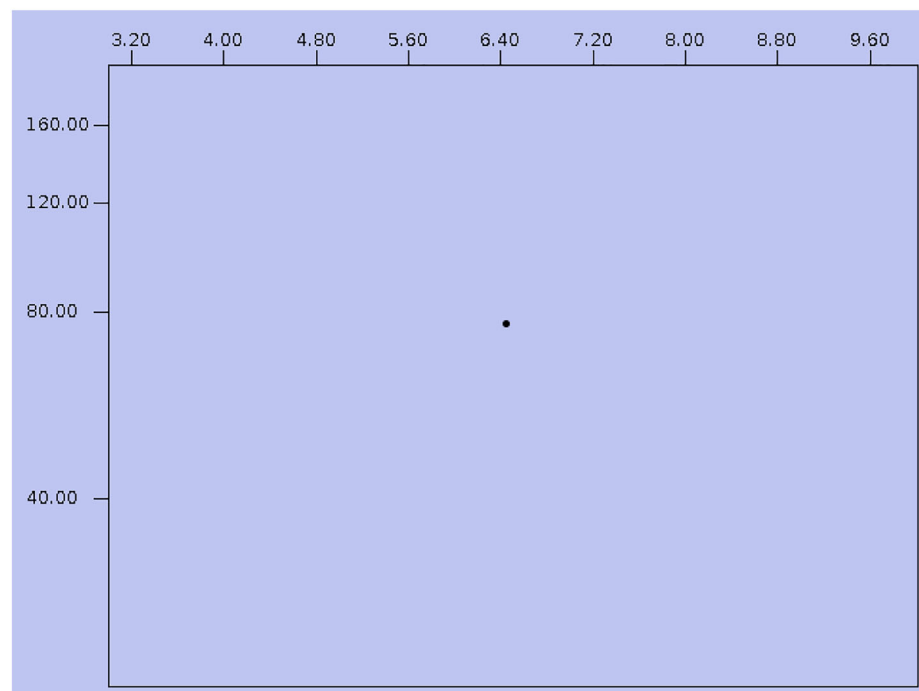


Fig. 1 (continued)

used. The predicted structure was obtained in the form of a protein database (pdb) compatible file. The predicted structure was finetuned by Galaxy refine tool of GalaxyWEB server

(<http://galaxy.seoklab.org/>) (Shin et al. 2014) to generate a 3D conformation with a higher propensity for the original conformation. SAVES v5.0 applets like Verify 3D,

**Table 2** Amino acid residues with magnesium-binding propensity

Property	Amino acids with their positions in the protein
Residues capable of $Mg^{++}$ binding	ASP101 ASP112 ASP151 ASP154 ASP214 ASP26 ASP315 ASP364 ASP428 ASP444 ASP45 ASP458 ASP47 ASP473 ASP49 ASP49 ASP503 ASP530 ASP533 ASP588 ASP662 ASP668 ASP67 ASP78 GLU106 GLU120 GLU133 GLU168 GLU233 GLU241 GLU271 GLU320 GLU365 GLU388 GLU395 GLU414 GLU446 GLU480 GLU481 GLU482 GLU486 GLU515 GLU522 GLU526 GLU534 GLU546 GLU550 GLU564 GLU59 GLU610 GLU663 GLU663 GLY115 GLY135 GLY140 GLY176 GLY205 GLY262 GLY300 GLY331 GLY367 GLY369 GLY371 GLY393 GLY447 GLY483 GLY485 GLY528 GLY553 GLY600 GLY607 GLY627 GLY645 GLY76 SER118 SER192 SER203 SER206 SER212 SER213 SER305 SER346 THR111 SER415 SER44 SER499 SER514 SER531 SER547 SER557 SER602 SER608 SER648 SER70 SER77 THR142 THR153 THR158 THR215 THR228 THR333 THR337 THR606

**Table 3** Different types of motifs present in LigA

Pfam	Residues positions	Description
DNA_lig_aden	17-316	NAD-dependent DNA ligase adenylation domain
DNA_lig_OB	321-391	NAD-dependent DNA ligase OB-fold domain
HHH_5	443-492	Helix-hairpin-helix domain
	511-560	
BRCT	594-670	BRCA1 C-Terminus (BRCT) domain
PTCB-BRCT	604-660	twin BRCT domain
DNA_lig_ZBD	402-427	NAD-dependent DNA ligase C4 zinc finger domain
HHH_2	508-575	Helix-hairpin-helix motif
HHH	545-561	Helix-hairpin-helix motif
DUF4332	516-564	Domain of unknown function (DUF4332)

PROCHECK, Prove, and ERRAT were used to check the accuracy of refined structure (<http://servicesn.mbi.ucla.edu/SAVES/>). Additionally, the structure was evaluated by assessing the secondary structural content by subjecting the modeled structure for an in silico circular dichroism spectroscopy analysis on the PDB2CD server (<http://pdb2cd.cryst.bbk.ac.uk/>) (Drew et al. 2018). The scan was done from a wavelength of 175 to 260 nm and plotted as a graph against the ellipticity or circular dichroism.

Structural homology further justified the predicted protein structure. For that, the crystal structure of DNA ligases from *Thermus filiformis* and *Enterococcus faecalis* was obtained from PDB with PDB ID “1V9P” and “4EFB” respectively. Both the structures were overlapped to check the propinquity by deconSTRUCT (<https://bio.tools/deconstruct>) (Zhang et al. 2010), TM-Align (Zhang and Skolnick 2005), and SuperPose version1.0 (<http://wishart.biology.ualberta.ca/SuperPose/>) (Maiti et al. 2004) servers. In the latter one, the two structures against each other option were considered for the analysis. The Gaussian width  $\delta$  was kept 0.3, while secondary structural elements (SSEs) were restrained to a minimum of three for matching. Additionally, structural co-ordinates of the adenylation domain of *Haemophilus influenzae* were derived from PDB with PDB ID “3UQ8” and superposed with the LigA structure.

## Phylogenetic analysis

Amino acid sequences of DNA ligase A/1 of eleven different species from different phyla were obtained by accessing UniProt KB. MEGA version X (Kumar et al. 2018) standalone software was used for the construction of the phylogenetic tree. The retrieved sequences were subjected for multiple alignment by ClustalW with a gap opening penalty of 10 and an extension penalty score of 0.10 and 0.20 for pairwise and multiple alignment respectively. The aligned file was saved as a mega-supporting format and later used for construction of a phylogenetic tree by Jones-Taylor-Thornton (JTT)

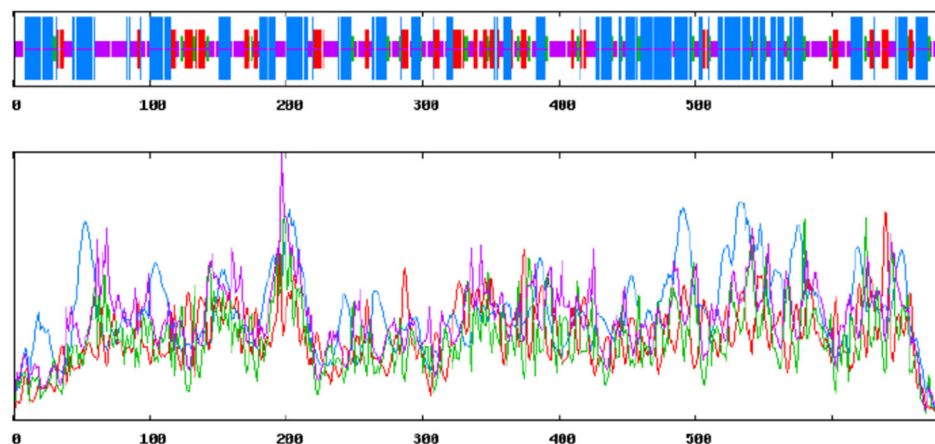
model (Jones et al. 1992) with maximum likelihood statistical method. Initially, the tree(s) were obtained automatically (for the heuristic search) by applying Neighbour-Join and BioNJ algorithms to a matrix of pairwise distance estimated using a JTT model and then selecting the topology with superior log likelihood value. The modeled 3-D structure was subjected for query in ConSurf server (<http://consurf.tau.ac.il/>) (Landau et al. 2005) to identify the evolutionary conserved residues/patches in the protein. This prediction used HMMER algorithm with *E* value cutoff of 0.0001 and multiple sequence alignment was performed by Bayesian calculation method. UniRef90 database was searched for homologous sequences.

## Identification of the DNA-binding region

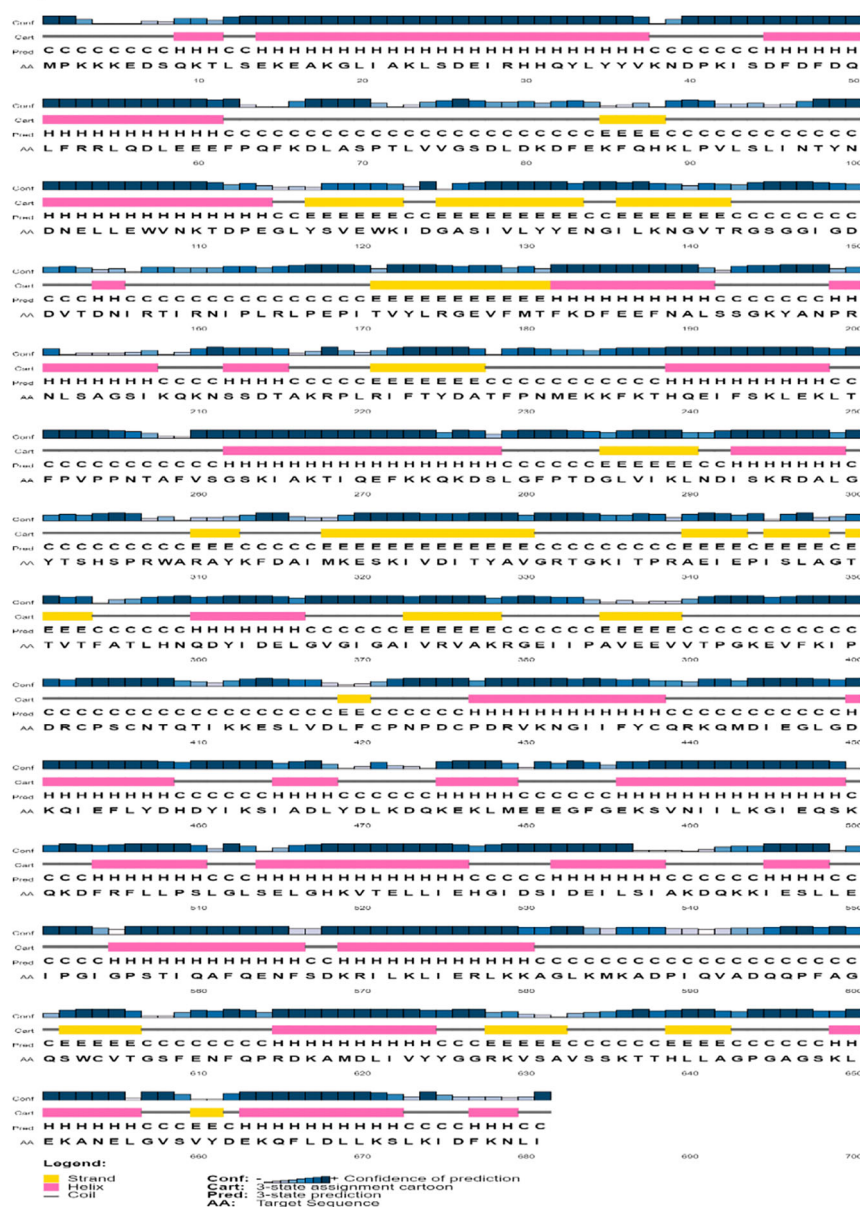
Two approaches were made to identify DNA-binding regions in the protein sequence. Amino acid residues possessing DNA-binding affinity were predicted from the protein sequence by MetaDBSite (<http://projects.biotec.tu-dresden.de/metadbsite/>) (Si et al. 2011) portal and DNA BIND PROT server (Haliloglu et al. 1997). The search parameters in the latter server were fixed as conservation threshold at 5, fast threshold percentage at 0.1 with exact residue neighbourhood method. The Gaussian network model was used in fast1 mode. In the second approach, docking of the LigA with a  $\beta$ -DNA dodecamer was performed to check the atomic orientation of the moieties during protein-DNA interaction. The structure of a twelve-nucleotide long  $\beta$ -DNA (CGCGAATTCGCG) was retrieved from PDB against the ID-1BNA. The predicted LigA structure and retrieved DNA structure were docked by online-based HDock (<http://hdock.phys.hust.edu.cn/>) (Yan et al. 2017) and PATCHDOCK servers (<https://bioinfo3d.cs.tau.ac.il/PatchDock/>) (Schneidman-Duhovny et al. 2005).

**Fig. 2** Assessment of secondary structure and prediction of protein conformation of ligA. The secondary structural content is shown in the (a-I) and (a-II). In (a-I), the blue color regions correspond to helix and sheets are represented by pink color. Helix (H) is represented in pink color and extended strand (E) in yellow in (a-II). Sterically hindrance in dihedral angles and conformational permitted regions are presented in the form of Ramachandran's plot (b). The predicted three-dimensional structure of protein is presented before (c) and after (d) chain/fold refinement. The circular dichroism of the protein as a function of wave length is presented graphically in e-I, while the secondary structural content in terms of  $\alpha$ -helix and  $\beta$ -sheet is presented in e-II. The later x-axis represents helix content and y-axis represents sheets. The star mark indicates the LigA. The structural co-ordinates of LigA from *L. interorgans* and *E. faecalis* were overlapped and analyzed for similarities. The output of analyzed results are presented in f-I where the blue color chain corresponds to the protein backbone of LigA of *Leptospira* and red chain refers to *Enterococcus*. The most overlapped regions between both the proteins are highlighted residue-wise in f-II, where black color indicates most similar and green/yellow color represents the distant amino acids. The distance between the aligned sequences is given amino acid-wise in f-III, where the red color states a distance below  $3\text{\AA}$ . (g) and (h) show the superposed structure of *Thermus filiformis* (red chain) and adenylation domain of *H. Influenza* (red chain) respectively with LigA (blue chain) of *Leptospira*

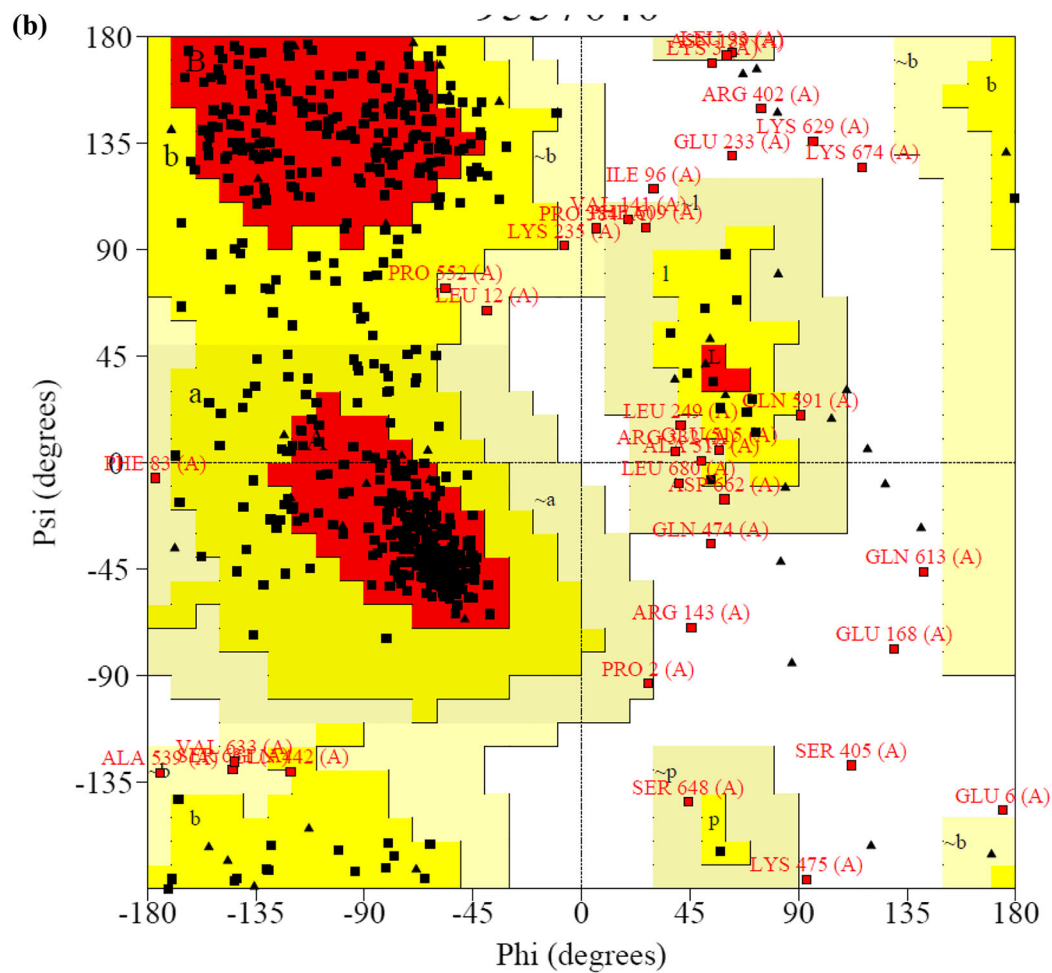
(a-I)



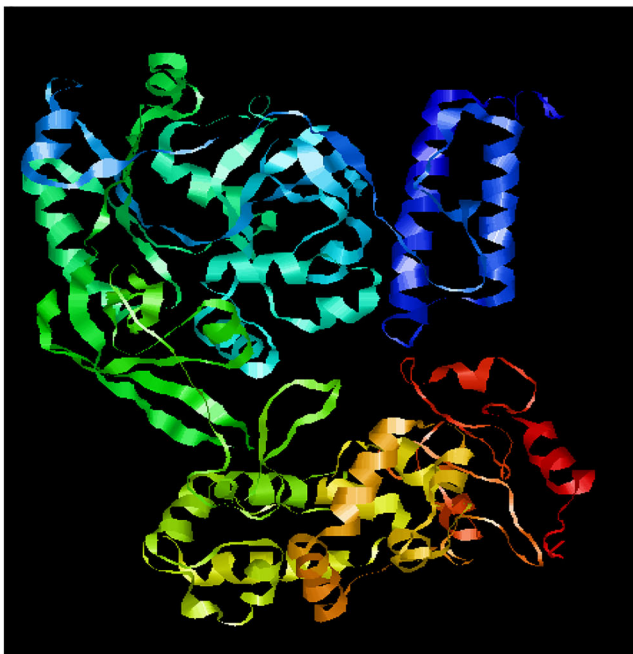
(a-II)







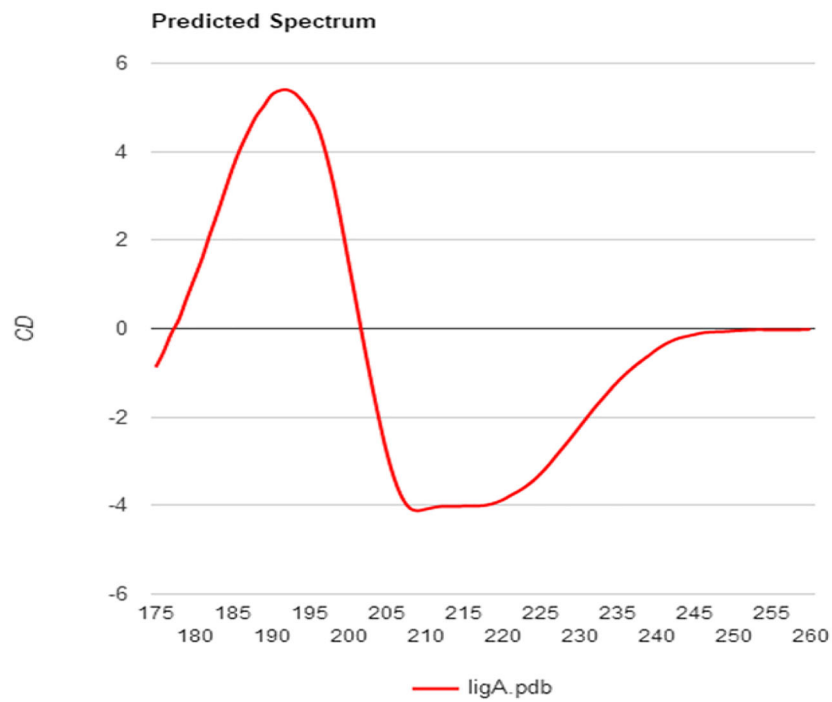
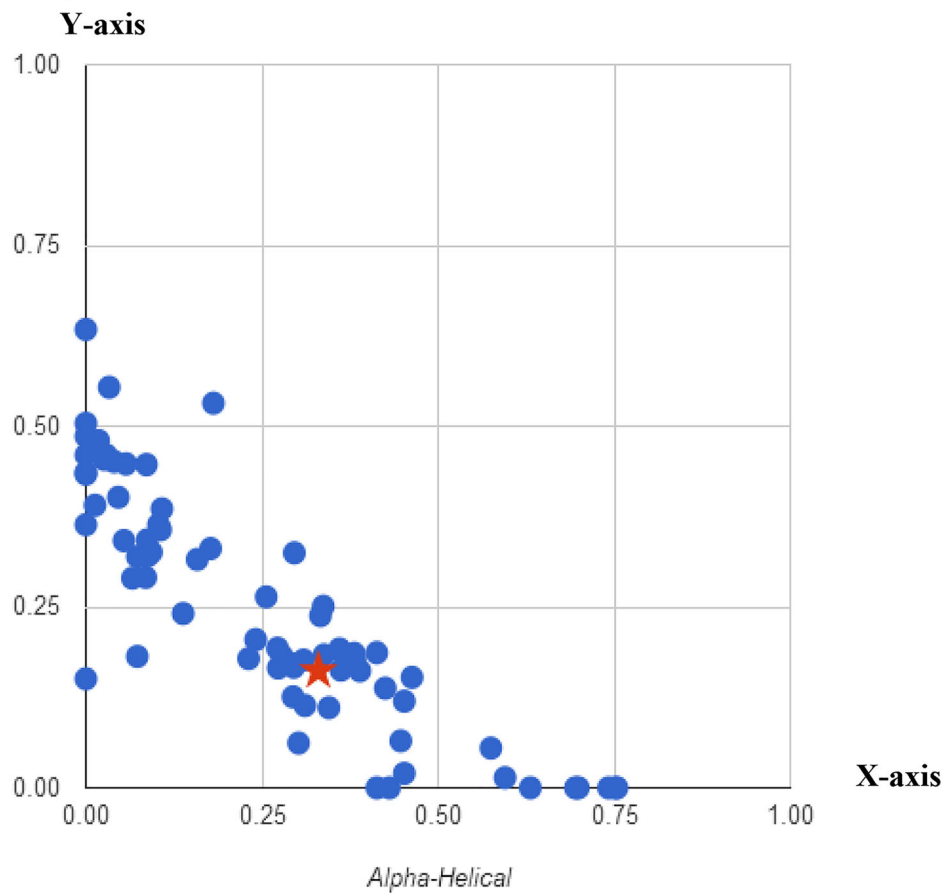
(2c)



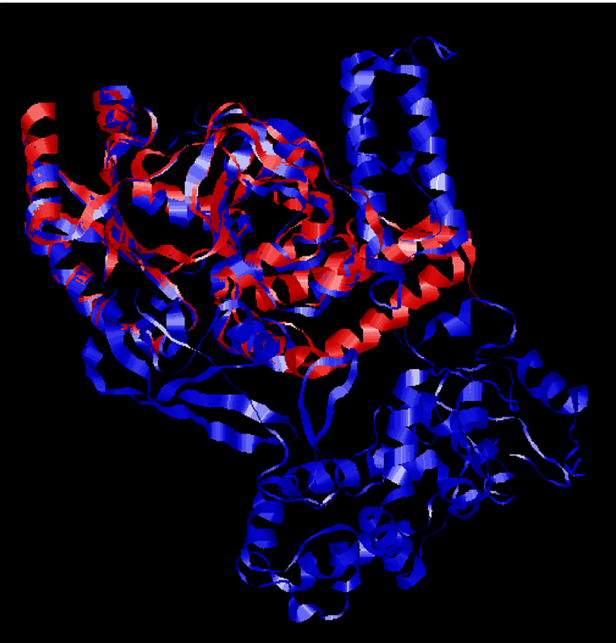
(2d)



Fig. 2 (continued)

**(2e-I)****(2e-II)****Fig. 2** (continued)

(2f-I)



(2f-II)

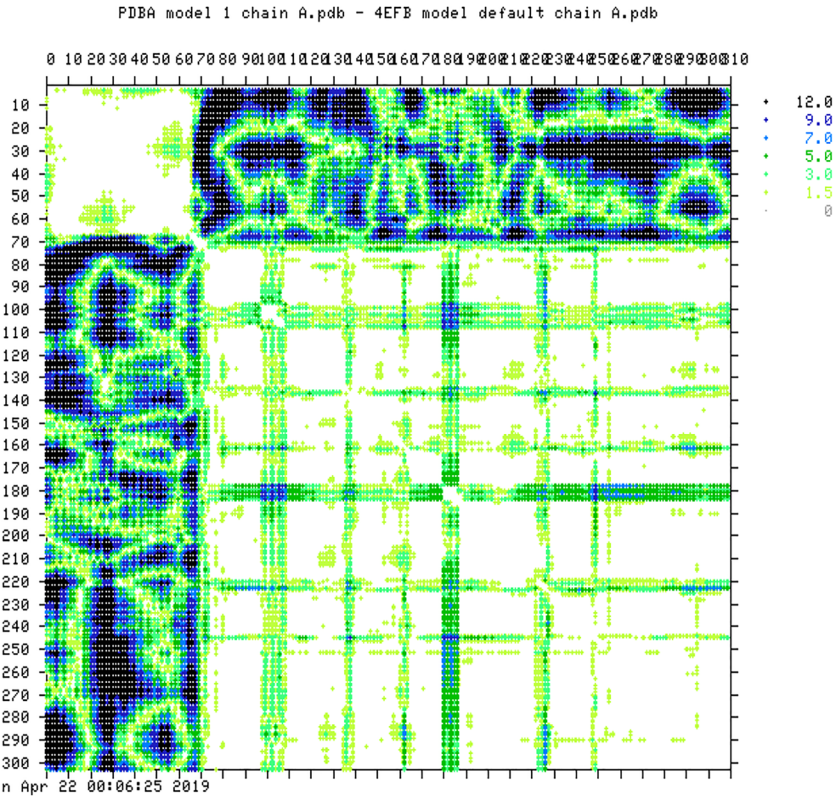
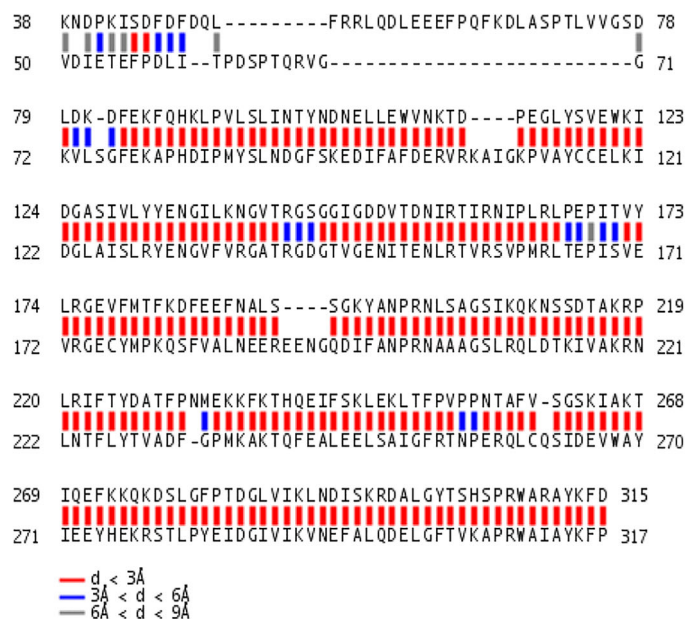


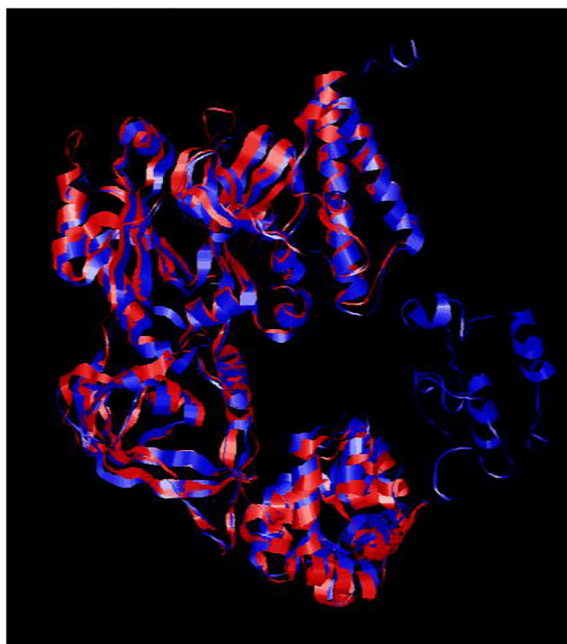
Fig. 2 (continued)



## (2f-III)



## (2g)



## (2h)



Fig. 2 (continued)

### Estimation of the effect of point mutation at the catalytic site

It is established that the enzyme LigA contains a residue of lysine-any amino acid–aspartic acid-glycine (-KXDG-) which is necessary for adenylation and deadenylation of the ligase. A sequence search revealed that the LigA homolog of

*Leptospira* does contain a similar sequence at position 122 in the form of “lysine-isoleucine-aspartic acid-glycine” (KIDG). A mutational analysis of these residues was done by converting K and D to G and G to aspartic acid. The possible effect was predicted by two servers Provean ([http://provean.jcvi.org/seq\\_submit.php](http://provean.jcvi.org/seq_submit.php)) (Choi and Chan 2019) and PredictSNP (<https://loschmidt.chemi.muni.cz/predictsnp1/>)

(Bendl et al. 2014). The latter is a secondary server that combinedly predicts the mutational effect from several primary servers available for forecasting of events that may result due to mutation. It takes inputs from PredictSNP, MAPP, PhD-SNP, PolyPhen-1, PolyPhen-2, SIFT, SNAP, nsSNPanalyzer, and PANTHER.

### Prediction of interactome and antimicrobial peptides (AMP)

String (<https://string-db.org/>) (Szkarczyk et al. 2018) gives a network of proteins that interacts physically or related functionally. As the database of the server did not contain *Leptospira* LigA, so *E. coli* protein was considered as a candidate for analysis considering its sequence similarity with the former.

The amino acid sequence of *Leptospira* LigA was searched for any possible antimicrobial peptide. AntiBP Server (<http://crdd.osdd.net/raghava/antibp/>) (Lata et al. 2007), which uses a support vector machine (SVM) based method for prediction of antimicrobial peptides with an overall accuracy of more than 90%, was used for the prediction of AMP. Amino terminus was used for prediction with a SVM threshold of 0.0. Additionally, the sequence was scanned in CAMP3 server (<http://www.camp3.bicnirrh.res.in/predict>) (Waghu et al. 2015) for AMPs. The query followed four different algorithms like SVM, artificial neural network, random forest, and discriminant analysis.

## Results

### Physico-chemical characterization of *Leptospira* LigA protein

The amino acid sequence of ligase A of *L. interrogans* was obtained from UniProt. It contains 681 amino acids in its skeleton with a molecular weight of 76.9 kDa. Protein blast search performed against *Leptospira* genus suggested conserved residues across various bacterial species. The physical and chemical properties of the protein were determined by Proto param server and presented in a tabular form (Table 1). The pI of the protein was found to be 6.44 which suggest proportionately more numbers of acidic residues in the structure. Though the gravity value of  $-0.427$  indicates non-polar nature of the protein with an instability index of 36.19, the protein was considered to be quite stable in vitro (a GRAVY value of  $< 40$  is indicative of a stable protein in vitro). The half-life of the protein in *E. coli* host was found to be more than 10 h. Over 30 hydrophobic patches can be seen in the protein (Fig. 1a-I), which is contributed by the abundant hydrophobic residues present in the protein (Fig. 1a-II). Aggrescan3D initially minimizes the 3D

structures energetically which are subjected to an intrinsic aggregation propensity scale for amino acids. The tendency to aggregate is calculated for spherical regions around the C $\alpha$  carbon of every residue. The score for LigA is presented in the form of a graph as shown in Fig. 1(a-III). Residues with most negative (i.e., highly soluble residue) and positive (i.e., highly aggregate forming residue) score are presented in Fig. 1(a-IV) and (a-V) respectively. The protein does not contain any transmembrane (Fig. 1b-I, II, III, and IV) or signal sequences (Fig. 1c-I). Probable metal (magnesium) binding regions present in the structure were predicted and shown in Fig. 1(c-II). The amino acid residues with magnesium-binding propensity are given in Table 2. One hundred one such residues were found in the LigA.

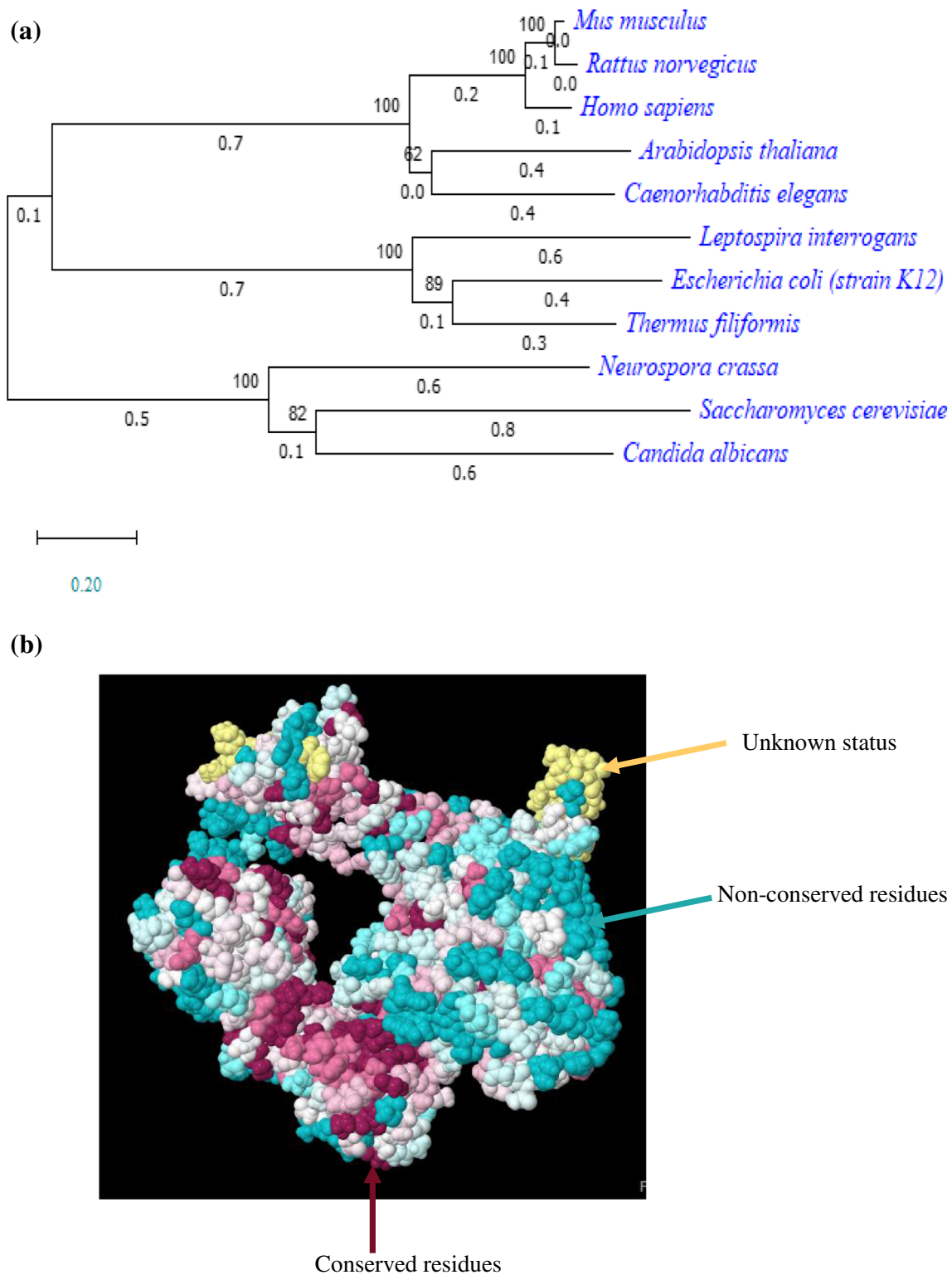
A total of nine (pfam and SMART server suggested 5 motifs, while motifFinder suggested 9 motifs) different types of motifs (Fig. 1d-I, II) are present in LigA (Table 3). Apart from the motif at position 516–564 with unknown function (DUF4332), most others have a significant role in DNA-binding and are well reported before (Lee et al. 2000). The protein in 2D gel electrophoresis migrated near to a relative molecular weight of 80 kDa and pH of near to 6.4 (Fig. 1e) which is as per the predicted protein characteristics.

### Evaluation of secondary structure and derivation of structural model

The secondary structural content was derived from the SOPMA (Fig 2 a-I) and Octopus (Fig 2 a-II) servers. It was found that the major portion of the protein will tend to adopt alpha helix (43.6%) followed by random coils (34.07%) and extended strand (14.83%) (Table 4). A very few residues may opt for beta turns (7.5%). Around 97.8% of residues are in favored to ordered secondary structure as indicated by Ramachandran's plot (Fig. 2b). The predicted 3D structure also suggested maximum alpha helix content both before (Fig. 2c) and after structural refinement (Fig. 2d). Validation of the protein conformation was done by multiple applets available in SAVES

**Table 4** Adoption of protein

Type of secondary structure	Percentage
Alpha helix (Hh)	43.61%
$3_{10}$ helix (Gg)	0.00%
Pi helix (Ii)	0.00%
Beta bridge (Bb)	0.00%
Extended strand (Ee)	14.83%
Beta turn (Tt)	7.49%
Bend region (Ss)	0.00%
Random coil (Cc)	34.07%



**Fig. 3** Phylogenetic analysis: A phylogenetic tree of LigA (compared from amino acid sequences of 11 different species) was constructed by MEGA X software. The output is presented in the form of a rooted dendrogram (a). The predicted three-dimensional structure of

Leptospiral LigA was subjected for analysis of the evolutionary conserved regions and the result is presented in (b), where the maroon-colored regions represent highly conserved regions and the yellow-colored regions are of unknown origin

v5.0 platform. An average score of 0.2 (Fig. 2d) indicated that most of the residues are well fitted in the predicted

structure. Above 96% of the residues in the predicted structure remain in allowed regions of the



**Table 5** Residues interacting with DNA

Property	Residues
Amino acid patch capable of binding to DNA	PRO384 SER127 LEU90 LYS276 VAL386 HIS358 ARG200 TYR99 ILE383 GLY194 LEU95 ALA377LYS313 ALA316 GLU103 ASN201 ARG379 LYS89 ILE148 ARG175 ASN189 TYR225 LYS208 PRO91GLN87 LEU93 GLU106 LYS378 ALA197 HIS304 ASN97 ASP226 THR302 PHE86 PRO306 ALA204ASN198 ASP315 GLY380 LYS110 GLY147 ARG310 THR283 LYS273 SER303 GLU381 SER305 ALA348TRP308 SER145 GLU177 GLU387 ALA385 PRO282 LYS122 ARG307 PHE314 ARG143 GLY205 MET318ILE96 TYR301 ALA126 LYS85 PHE185 SER94 GLY146 GLY176 THR356 TYR362 ILE317 TYR196 LYS195

Ramachandran's plot, among which 74.9% lies in the most favored regions. Moreover, a query was made against the modeled LigA structure for in silico derivation of circular dichroism of the protein which designated it as an alpha helix-rich protein (Fig. 2e–I, II). The final attempt to validate the 3D structure was by superposing the structural coordinates of a protein homolog from *Enterococci* (Fig. 2f–I, II, III) and *Thermus filiformis* (Fig. 2g) with the *Leptospira* LigA. The superposed structure of the *Enterococci* contained 249 aligned residues with a root mean square deviation (RMSD) value of 2.3 Å<sup>0</sup> and which is within the acceptable range for predicted overlap. The RMSD value for *T. filiformis* and *H. influenza* (Fig. 2h) adenylation domain overlap was found to be 1.70 and 1.76 respectively.

## Phylogeny

A phylogenetic tree (rooted) for LigA was constructed with 11 sequences (amino acid) from different species by MEGA X software (Fig. 3a). Ligases from *E. coli* and *Thermus filiformis* were found to be in the same clad and descendent from *Leptospira* where the latter served as internal node. The distribution of the gene is well conserved across different phyla and species. The amino acid residues which are highly conserved were predicted from Consurf server. The color coding in Fig. 3b indicates the evolutionary status of the residues.

## Delineation of DNA-binding residues and evaluation of the effect of mutation at adenylation site

As per the MetaDBSite result, out of the 681 residues, a large patch of 73 residues was found to be capable of interacting with DNA. The residues are enlisted in Table 5. Additionally, DNA BIND PROT server suggested 12 residues (from C-terminal of the protein) with potential DNA-binding affinity (Table 6). Further, docking analysis supported the residues involved in binding of the enzyme to DNA (Fig. 4–I, II).

The effect of mutation at the adenylation site was studied by substituting a neutral amino acid (glycine)

in the place of a charged amino acid, i.e., lysine<sub>122</sub>. Similarly, the residue important for deadenylation, i.e., aspartic acid<sub>124</sub>, was substituted with glycine. Additionally, the glycine<sub>125</sub> was exchanged with aspartic acid. All these three substitutions were found to be detrimental to the protein.

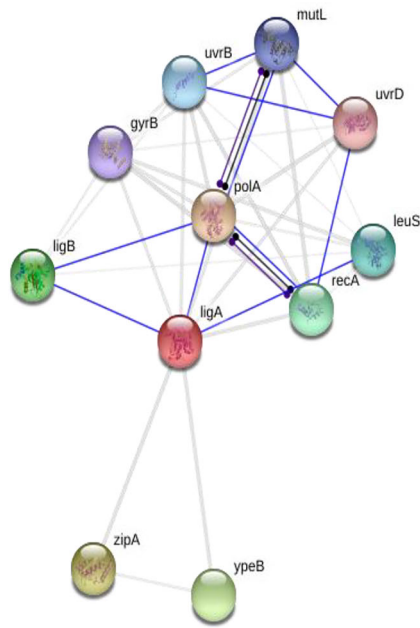
## Derivation of interactome and antimicrobial peptides of LigA

The iterated interactome of LigA suggests ten different proteins like mutL, zipA, uvr B/D, recA, and gyrB (Fig. 5a). The output showed 11 nodes with an average node degree of 6.36 and an average local clustering coefficient of 0.9. The model was suggestive of 3 different binding proteins to LigA as ligB, polA, and leuS. The co-expression pattern of the protein was also predicted from the string server and presented in Fig. 5b. The peptide stretches of the protein with a significant score (> 1 for antiBP server and 0.5 for CAMP server) for the antimicrobial effect are presented in Table 7. These sequences are the common stretches predicted from both the servers with significant score. A total of 10 peptides were found to be eligible for fulfilling the criteria.

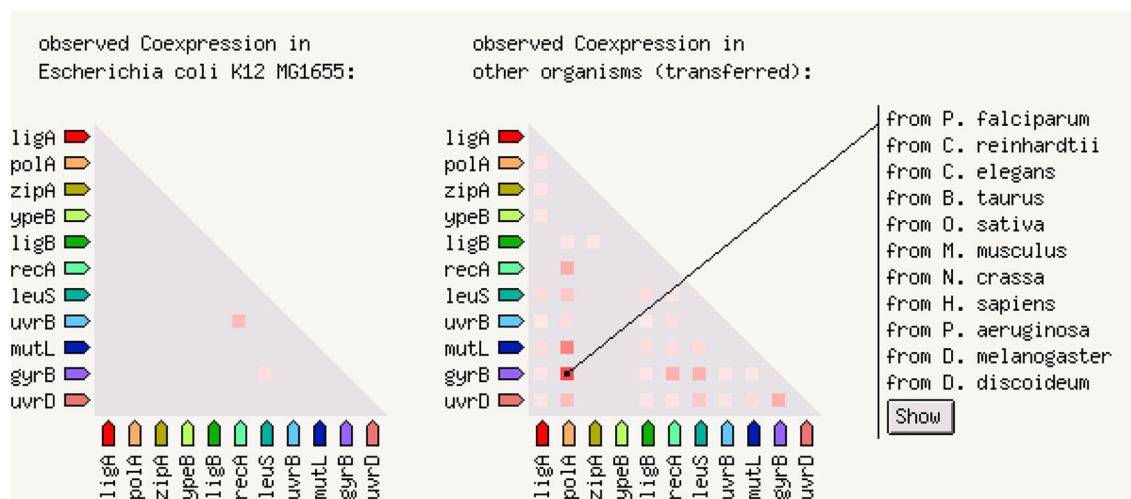
**Table 6** Residues with potential DNA-binding affinity

Position	Residues
440	Arg
502	Lys
499	Ser
510	Ser
521	Thr
439	Gln
437	Tyr
507	Leu
506	Phe
445	Ile
443	Met
438	Cys

(a)



(b)



**Fig. 4** DNA-binding regions were predicted by docking analysis. Four of the probable docking sites (pointed by colored arrows) are presented with bound oligonucleotide in I. The docking model with maximum stability and best fit is being represented in II. Some of the DNA-interacting

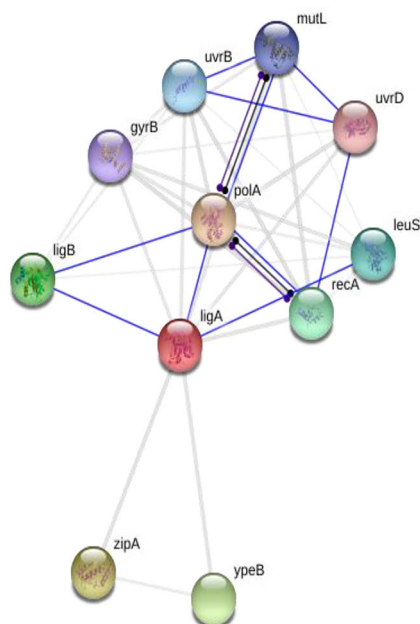
residues of LigA are presented with their positions in the latter figure. The residues like Gly147 and Glu103 (encircled by yellow line) are also part of the DNA-binding patch predicted by MetaDBSite

## Discussion

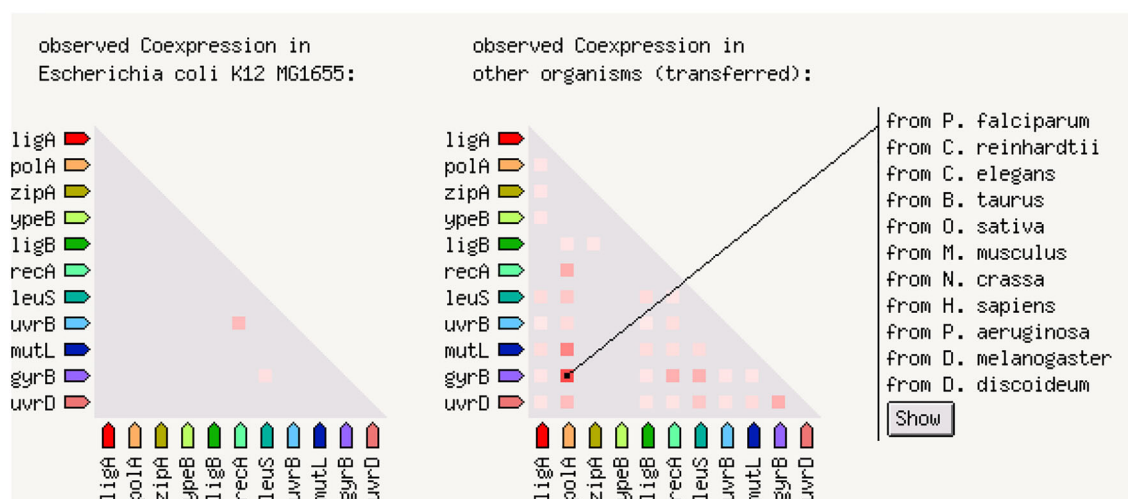
In the present study, DNA ligase A enzyme of *Leptospira interrogans* is being characterized. With the help of SwissProt's Protparam module, it was predicted that the protein bears a proportionately large amount of acidic amino acids (Asp+Glu) which confers the overall pI. This was further confirmed by virtual two-dimensional gel electrophoresis which separates the molecules based on their charge and isoelectric pH. The protein does not contain any sorting signals which confine it to the protoplasmic area of the bacteria.

Though it contained many hydrophobic patches but no trans-membrane sequences were found in the protein. A consensus adenylation site is present in the protein between the residue 122–125 as -KIDG-. The structure of DNA ligase A (*Leptospira*) by X-ray diffraction is not available until now (September 2019). So, a 3D structure of the protein was predicted by I-TASSER server. The structure was predicted with the help of iterative threading against the available templates which further got refined by Galaxy refine tool. The sterically favorable dihedral angle content in the protein was found to be more than 90% in the predicted structure. The structure is rich

(a)



(b)



**Fig. 5** Interactome analysis: The protein homolog from *E. coli* was analyzed by STRING server. The predicted interactome is represented in (a), where blue color lines from ligA indicate its direct interaction with polA, leuS, and ligB. The co-expression pattern of the proteins

iterated from the analysis is represented in (b), where the left and right panels show the expression in *E. coli* and in other organisms (enlisted at the side)

in alpha helices which was proved later by computational CD and prediction of secondary structure content. A structural comparison (Fig. 6) was made against one of the available homology model from SWISS-MODEL repository for LigA of *L. interrogans* serogroup Icterohaemorrhagiae serovar Lai (strain 56601) (<https://swissmodel.expasy.org/repository/uniprot/Q8EYU4>). It was found that the RMSD value was 5.42 between the predicted and available structure co-ordinates, stating there is a significant difference between them.

The classical domain structures of a typical DNA ligase are present in the *Leptospira* ligase A. The domain responsible for

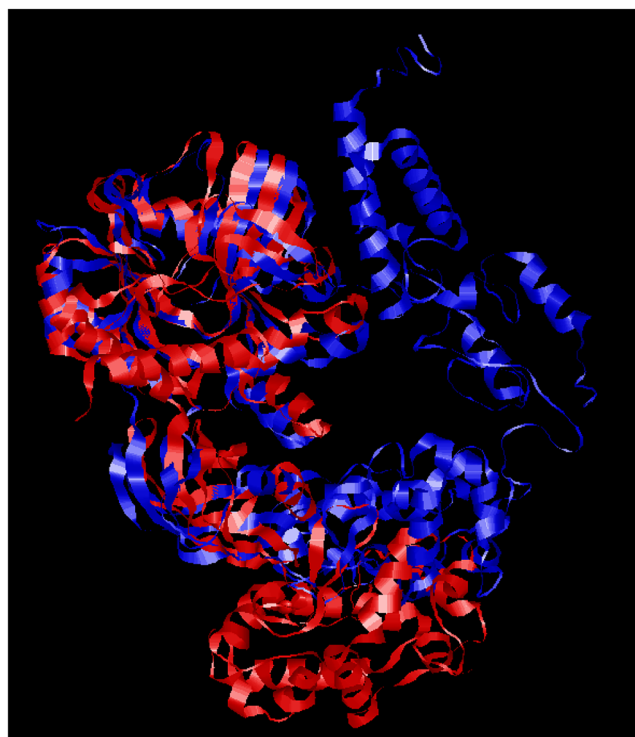
interaction with  $\text{NAD}^+$  spans the N-terminus, while BRCT domain is present at the C-terminus as reported earlier in other DNA ligases (Lee et al. 2000). A prediction with the help of N<sub>SITE</sub>P<sub>RED</sub> server (Chen et al. 2011) suggests that there are 22 residues (*Leptospira* LigA) that can interact with ATP and 39 with GTP. Furthermore, the phylogenetic analysis supported the fact that the amino acid sequence of LigA is conserved across many forms of life. Superposing of LigA of *L. interrogans* with *E. faecalis* ligase A indicated a very close fit between the aligned residues. Though it has been established that a central tunnel in the ligase accommodates

t-

**Table 7** Peptide stretches of the protein with a significant score for the antimicrobial effect

Residues	Location	Score (AntiBP)	Score (CAMP)
KLIERLKKAGLKMKA	573–587	1.903	0.9
SIDEILSIKADQK	531–543	1.47	0.85
ILKLIERLKKAGLKM	570–585	1.42	0.9
ANPRNLSAGSIKQKN	196–211	1.22	0.5
ILKGIEQSKQKDFRF	492–506	1.11	0.9
GSKIAKTIQEFKKQK	262–276	1.094	0.764
GIEQSKQKDFRLLP	495–509	1.079	0.766
SIDEILSIKADQKKI	531–545	1.072	0.669
QIEFLYDHDYIKSIA	455–466	1.038	0.783
VVGSDLDKDFEFKQH	74–88	1	0.564

he DNA during nick sealing (Timson and Wigley 1999), but docking with a dodecamer nucleotide gave some additional regions with potential DNA attachment. This could be due to the shorter size of the nucleotide fragment available for docking analysis. As the ligase needs a 19 region flanking sequence to the nick site for sealing the DNA backbone (Tomkinson et al. 2006), the structure taken for analysis might not be of sufficient length to span the entire enzyme. The interactome analysis of the protein suggested that it directly



**Fig. 6** A structural comparison was made by superimposing the 3D coordinates of the I-TASSER (blue) predicted and SWISS-MODEL repository-derived LigA (red) structure, with the help of the TM-align server

interacts with a less efficient counterpart, i.e., DNA ligase B. Apart from that, it may also interact with a tRNA ligase, responsible for charging of the activated leucine with the cognate tRNA<sup>leu</sup>. The protein was found to be co-expressed with various other proteins/enzymes involved in DNA lesion repair processes (Fig. 5b). Surprisingly, the ligase A was found to have 10 potential antimicrobial regions. It might happen so that during protein turnover, selective cleavage of the protein may result in many such peptides which can help the bacteria in survival from other bacteria and to propagate as in the case of gut microbiota. But this hypothesis needs further investigations. Blast search against CAMP<sub>R3</sub> database suggested that, among the predicted peptides, KLIERLKKAGLKMKA shares 57% identity with cathelin-related peptide of mouse which is known to be potent antimicrobial agent (Kościuczuk et al. 2012), GSKIAKTIQEFKKQK shares 56% identity with cationic peptide 3c of *Cupiennius salei* (Kuhn-Nentwig et al. 2012), and VVGSDLDKDFEFKQH shares 64% identity with human ISG20 (Jiang et al. 2008) but others are novel AMPs. Generating peptides (identified from sequences of LigA) and testing their efficacy against various microbial organisms can help in establishing the ligase's moonlighting action as an AMP.

## Conclusions

The ligase A protein of *leptospira interrogans* was characterized by several online-based servers. The protein was found to be rich in acidic and hydrophobic residues. It can be said that the protein will adopt an alpha helix-rich conformation during folding. Structural diversity can be seen from other homologs from different species but many signature sequences are seen being conserved in the protein backbone. Mutational analysis of a crucial site responsible for adenylation and deadenylation showed compromised protein function. Surprisingly, the protein contains different stretches which may act as antimicrobial. Further analysis by constructing peptides from these predicted sequences can provide us some new antimicrobial agents.

**Acknowledgments** The authors are highly thankful to the Dean, FVAS, RGSC-BHU, and Vice-Chancellor, Banaras Hindu University, for providing the required supports during the analysis.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Research involving human participants and/or animals** N/A

**Informed consent** N/A



## References

- Alland C, Moreews F, Boens D, Carpentier M, Chiusa S, Lonquety M, Renault N, Wong Y, Cantalloube H, Chomilier J, Hochez J (2005) RPBs: a web resource for structural bioinformatics. *Nucleic Acids Res* 33:W44–W49
- Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* 37:420
- Bauer RJ, Zhelkovsky A, Bilotti K, Crowell LE, Evans TC Jr, McReynolds LA, Lohman GJ (2017) Comparative analysis of the end-joining activity of several DNA ligases. *PLoS One* 12: e0190062
- Bendl J, Stourac J, Salanda O, Pavelka A, Wieben ED, Zendulka J, Brezovsky J, Damborsky J (2014) PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. *PLoS Comput Biol* 10:e1003440
- Buchan DW, Jones DT (2019) The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res* 47:W402–W407
- Chen K, Mizianty MJ, Kurgan L (2011) Prediction and analysis of nucleotide-binding residues using sequence and sequence-derived structural descriptors. *Bioinformatics* 28:331–341
- Choi Y, Chan AP (2019) PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* 31:2745–2747
- Costa F, Hagan JE, Calcagno J, Kane M, Torgerson P, Martinez-Silveira MS, Stein C, Abela-Ridder B, Ko AI (2015) Global morbidity and mortality of leptospirosis: a systematic review. *PLoS Negl Trop Dis* 9:e0003898
- Drew ED, Mavridis L, Janes RW (2018) PDB2CD: a web-based application for the generation of circular dichroism spectra from protein atomic coordinates. *Biophys J* 114:46a
- El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer EL (2018) The Pfam protein families database in 2019. *Nucleic Acids Res* 47:D427–D432
- Gajiwala KS, Pinko C (2014) Structural rearrangement accompanying NAD<sup>+</sup> synthesis within a bacterial DNA ligase crystal. *Structure* 12:1449–1459
- Haliloglu T, Bahar I, Erman B (1997) Gaussian dynamics of folded proteins. *Phys Rev Lett* 79:3090
- Hiller K, Grote A, Maneck M, Münch R, Jahn D (2006) JVirGel 2.0: computational prediction of proteomes separated via two-dimensional gel electrophoresis under consideration of membrane and secreted proteins. *Bioinformatics* 22:2441–2443
- Honig B (1999) Protein folding: from the levinthal paradox to structure prediction. *J Mol Biol* 293:283–293
- Inada R, Ido Y, Hoki R, Kaneko R, Ito H (1916) The etiology, mode of infection, and specific therapy of Weil's disease (*spirochaetosis icterohaemorrhagica*). *J Exp Med* 23:377
- Jiang D, Guo H, Xu C, Chang J, Gu B, Wang L, Block TM, Guo JT (2008) Identification of three interferon-inducible cellular enzymes that inhibit the replication of hepatitis C virus. *J Virol* 82:1665–1678
- Johnson RC (1984) *Leptospira*. Bergey's manual of systematic bacteriology 1, pp 62–67
- Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* 8: 275–282
- Kościuczek EM, Lisowski P, Jarczak J, Strzałkowska N, Jóźwik A, Horbańczuk J, Krzyżewski J, Zwierzchowski L, Bagnicka E (2012) Cathelicidins: family of antimicrobial peptides. A review. *Mol Biol Rep* 39:10957–10970
- Kuhn-Nentwig L, Fedorova IM, Lüscher BP, Kopp LS, Trachsel C, Schaller J, Vu XL, Seebeck T, Streitberger K, Nentwig W, Sigel E (2012) A venom-derived neurotoxin, CsTx-1, from the spider *Cupiennius salei* exhibits cytolytic activities. *J Biol Chem* 287: 25640–25649
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 35:1547–1549
- Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, Pupko T, Ben-Tal N (2005) ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res* 33: W299–W302
- Lata S, Sharma BK, Raghava GP (2007) Analysis and prediction of antibacterial peptides. *BMC bioinformatics* 8:263
- Lee JY, Chang C, Song HK, Moon J, Yang JK, Kim HK, Kwon ST, Suh SW (2000) Crystal structure of NAD<sup>+</sup>-dependent DNA ligase: modular architecture and functional implications. *EMBO J* 19:1119–1129
- Maiti R, Van Domselaar GH, Zhang H, Wishart DS (2004) SuperPose: a simple server for sophisticated structural superposition. *Nucleic Acids Res* 32:W590–W594
- Mavridis L, Janes RW (2016) PDB2CD: a web-based application for the generation of circular dichroism spectra from protein atomic coordinates. *Bioinformatics* 33:56–63
- Miesel L, Kravec C, Xin AT, McMonagle P, Ma S, Pichardo J, Feld B, Barrabee E, Palermo R (2007) A high-throughput assay for the adenylation reaction of bacterial DNA ligase. *Anal Biochem* 366: 9–17
- Nandakumar J, Nair PA, Shuman S (2007) Last stop on the road to repair: structure of *E. coli* DNA ligase bound to nicked DNA-adenylate. *Mol Cell* 26:257–271
- Pergolizzi G, Wagner GK, Bowater RP (2016) Biochemical and structural characterization of DNA ligases from bacteria and archaea. *Biosci Rep* 36:e00391
- Rao RS, Gupta N, Bhalla P, Agarwal SK (2003) Leptospirosis in India and the rest of the world. *Braz J Infect Dis* 7:178–193
- Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res* 33:W363–W367
- Schultz J, Milpetz F, Bork P, Ponting CP (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci* 95:5857–5864
- Shin WH, Lee GR, Heo L, Lee H, Seok C (2014) Prediction of protein structure and interaction by GALAXY protein modeling programs. *Bio Design* 2:1–1
- Shuman S (2009) DNA ligases: progress and prospects. *J Biol Chem* 284: 17365–17369
- Si J, Zhang Z, Lin B, Schroeder M, Huang B (2011) MetaDBSite: a meta approach to improve protein DNA-binding sites prediction. *BMC Syst Biol* 5:S7
- Sonnhammer EL, Von Heijne G, Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *In* *Ismb* 6:175–182
- Swift RV, Amaro RE (2009) Discovery and design of DNA and RNA ligase inhibitors in infectious microorganisms. *Expert Opin Drug Discovery* 4:1281–1294
- Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ (2018) STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 47:D607–D613
- Timson DJ, Wigley DB (1999) Functional domains of an NAD<sup>+</sup>-dependent DNA ligase. *J Mol Biol* 285:73–83
- Tomkinson AE, Vijayakumar S, Pascal JM, Ellenberger T (2006) DNA ligases: structure, reaction mechanism, and function. *Chem Rev* 106:687–699

- Viklund H, Bernsel A, Skwark M, Elofsson A (2008) SPOCTOPUS: a combined predictor of signal peptides and membrane protein topology. *Bioinformatics* 24:2928–2929
- Waghu FH, Barai RS, Gurung P, Idicula-Thomas S (2015) CAMPR3: a database on sequences, structures and signatures of antimicrobial peptides. *Nucleic Acids Res* 44:D1094–D1097
- Wang LK, Zhu H, Shuman S (2009) Structure-guided mutational analysis of the nucleotidyltransferase domain of *Escherichia coli* DNA ligase (LigA). *J Biol Chem* 284:8486–8494
- Wilkinson A, Day J, Bowater R (2001) Bacterial DNA ligases. *Mol Microbiol* 40:1241–1248
- Wodak SJ, Vagin AA, Richelle J, Das U, Pontius J, Berman HM (2012) Deviations from standard atomic volumes as a quality measure for protein crystal structures, *International Tables for Crystallography F*, pp 664–665
- Yaakob Y, Rodrigues KF, John DV (2015) Leptospirosis: recent incidents and available diagnostics—a review. *Med J Malaysia* 70:351
- Yan Y, Zhang D, Zhou P, Li B, Huang SY (2017) HDock: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res* 45:W365–W373
- Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y (2015) The I-TASSER suite: protein structure and function prediction. *Nat Methods* 12:7
- Zambrano R, Jamroz M, Szczasiuk A, Pujols J, Kmiecik S, Ventura S (2015) AGGRESCAN3D (A3D): server for prediction of aggregation properties of protein structures. *Nucleic Acids Res* 43:W306–W313
- Zhang Y, Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 33:2302–2309
- Zhang ZH, Bharatham K, Sherman WA, Mihalek I (2010) deconSTRUCT: general purpose protein database search on the substructure level. *Nucleic Acids Res* 38:W590–W594
- Zhou AQ, O’Hern CS, Regan L (2011) Revisiting the Ramachandran plot from a new angle. *Protein Sci* 20:1166–1171

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.