ORIGINAL ARTICLE **Open Access**

ISSN 2278-1404

International Journal of Fundamental & Applied Sciences

Comparative modeling and *in silico* characterization of FtsH from *Mycobacterium tuberculosis*

Ashwini K and Vemula Vani

Department of Microbiology, M.S. Ramaiah college of Arts, Science and Commerce, Bangalore- 54, Karnataka, INDIA

Manuscript received 30th Nov, 2015, revised 01st Jan, 2016, accepted 3rd Jan, 2016

Abstract

Background: Tuberculosis (TB) is a major global health problem causing over one million deaths per year. The causative agent responsible for causing TB is *Mycobacterium tuberculosis.* In host-parasite diseases like tuberculosis, proteins as drug target are first preference. Multidrug-Resistant Tuberculosis (MDR TB) occurs when Mycobacterium *tuberculosis* strain is resistant to isoniazid and rifampin, two of the most powerful first-line drugs. FtsH is a membrane bound ATP dependent Zinc-metalloprotease which proteolytically regulates the level of specific membrane and cytoplasmic proteins that participate in diverse cellular function. FtsH is essential membrane – bound protease that degrades integral membrane proteins as well as cytoplasmic proteins. **Aim:** The objectives of the present study are to determine the three- dimensional (3D) structure of topological domain of FtsH from *Mycobacterium tuberculosis* using comparative modeling and its *in silico* characterization. **Methodology:** The sequence for FtsH was retrieved from UNIPROT database and sequence analysis was carried out using BLAST and FUGUE for the selection of template. The crystal structure of FtsH from *Thermatogamaritima* was selected as a template. The protein modeling was carried out using ModWeb and Swissmodeller. The obtained 3D model of the FtsH was visualized and analyzed using Chimera. This modeled protein structure was refined by loop modeling. Later, the quality of the protein structure was verified by its energy and stereochemical properties. Further, the *in sillico* characterization of the FtsH was carried out. **Result:** The 3D structure of FtsH, obtained from this study can be used in developing novel inhibitors using the methods of rational drug designing.

Keywords*:* FtsH, Comparative modeling, tuberculosis, *in silico* characterization, 3D structure

1. Introduction

Tuberculosis, a multisystemic disease with myriad presentations and manifestations, is a most common cause of infectious disease related mortality worldwide. As the conventional treatment is lengthy andcomplex, there is pressing need for new drugs, preferably with noble modes of action to avert the problem of cross-resistance[1]. Several new targets have been proposed including proteins which play very important role in the pathogenicity. FtsH is one such protein which we have selected for our study. FtsH from Mycobaterium tuberculosis is a membrane bound ATP – dependent zinc metalloprotease which proteolytically regulates the level of specific membrane and cytoplasmic proteins that participate in diverse cellular function[2]. As of now no crystal structure available for FtsH. Combined with evidence that FtsH is essential and involved in adaptation to survive *in vivo,* indicates that FtsH may be a suitable target for drug development.

The experimental methods to determine the protein 3D structure like X-ray crystallography, nuclearmagnetic resonance are technically demanding, time consuming and may not keep with which new protein sequences are being discovered by genomics research. Although a large number of genes being discovered, the number of protein structures being solved by experimental methods is limited. Alternative strategies for structure prediction and modeling of proteins are computational methods.

The major computational methods for predicting the structure of proteins are *ab initio* methods and comparative modeling. Comparative protein structure modeling remains the most accurate prediction method[3]. Comparative Modeling is also known as homology

****Corresponding author***
Full Address :
Department of Microbiology, M.S. Ramaiah college of Arts, Science and Commerce, Bangalore- 54, Karnataka, INDIA
Phone no. +91 9632119023
E-mail: vemula.vani@gmail. com

modeling of proteins is an alternative method to determine the 3-Dimensional structure of protein.With the progression of structural genomics projects, comparative modelling remains an increasingly important method of choice. It helps to bridge the gap between the available sequence and structure information by providing reliable and accurate protein models. Comparative modeling is a technique for predicting or generating detailed 3D structures of proteins based on coordinates of known homologues.

The main steps to create a comparative model are as follows: 1) Identification of structural homologues.2) Selection of structural homologues used as templates for modeling. 3) Alignment of templates withthe protein sequence to be modeled. 4) Model building. 5) Evaluation and refinement of the model.

The objectives of this present study are to obtain 3D structure of the topological domain of FtsH using comparative mdelling and its *in silico* characterization. This medelled structure can be used as potential drug target for developing inhibitors.

2. Materials and Methods

2.1 Retrieval of FtsH sequence from Uniprot database

In our study the sequence details of the protein FtsH was retrieved from UniProt database. The UniProt Knowledgebase (UniProtKB) is the central hub for the collection of functional information on proteins, with accurate, consistent and rich annotation [4].

2.2 Identification of Template

In order to search for the template for model building, the FtsH sequence was submitted to PDB BLAST server. To confirm the results obtained from PDB-BLAST, the sequence was also submitted to FUGUE, which is used for recognizing distant homologues by sequence-structure comparison.BLAST is a heuristic that finds short matches between two sequences and attempts to start alignments from these ‘hotspots’. In addition to performing alignments, BLAST (http://www/ncbi.nih.gov/BLAST/) provides statistical information to help decipher the biological significance of the alignment; this is the ‘expect’ value, or false positive rate [5-7].

2.3 Model building

For building the model we have used two different software, MODWEB and SWISS- MODEL. MODWEB is a queryable database of annotated protein structure models. The models are derived by ModPipe, an automated modeling pipeline relying on the programs PSI-BLAST and MODELLER. The database also includes the fold assignments and alignments on which the models were based. MODWEB[8] contains theoretically calculated models, which may contain significant errors, not experimentally determined structures. Thus, special care is taken to assess the quality of the models.

SWISS-MODEL is a structural bioinformatics web-server dedicated to homology modelling of protein 3D structures. Comparative modelling methods make use of experimental protein structures ("templates") to build models for evolutionary related proteins ("targets"). SWISS-MODEL consists of three tightly integrated components: (1) The SWISS-MODEL pipeline - a suite of software tools and databases for automated protein structure modelling. (2) The SWISS-MODEL Workspace - a web-based graphical user workbench. (3) The SWISS-MODEL Repository - a continuously updated database of homology models for a set of model organism proteomes of high biomedical interest.

2.4 Secondary structure prediction

SOPMA was used for secondary structure prediction of topological structure of FtsH. SOPMA (Self-Optimized Prediction Method with Alignment) is an improvement of SOPM method. SOPMA correctly predicts 69.5% of amino acids for a three-state description of the secondary structure (alpha-helix, beta-sheet and coil) in a whole database containing 126 chains of non-homologous (less than 25% identity) proteins[9].

2.5 Loop modeling

Modeling of the errored loops in FtsH structure was carried out using Swiss-PDB Veiwer. Swiss -pdb veiwer is an application that provides a user friendly interface allowing analyzing several proteins at the same time. SPDBV was used to remodel the regions which showed instability in the verify 3D graph. The regions were selected using control panel and suitable loop was selected from the database which assures the stability of the selected loop including the overall structure [10,11].

2.6 Visualization of the model

Chimera 1.5.3 was used for visualizing the model. Chemera is highly extensible program for interactive visualization and analysis of the molecular structures and related data, including density map, supramolecular asemblies , sequence alignments, docking results, trajectories, and conformational ensemles. High quality images and animations can be generated.

2.7. Evaluation of the model

The evaluation of the obtained model for FtsH was carried out using verify 3D program. The three dimensional (3D) profile of a protein structure is a table computed from the atomic coordinates of the structure that can be used to score the compatibility of the 3D structure model with any amino acid sequence. The stereo -chemical quality of the FtsH structure was analyzed by Ramchandran plot using the software RAMPAGE. RAMPAGE[12] is an offshot of RAPPER which generates a Ramchandran plot [13-15] .

2.8 Computation of physical and chemical properties of FtsH

The physical and chemical properties of FtsH was computed by protparam. It is a tool which allows the computation of various physical and chemical parameters for a given protein sequence. The computed parameters include the molecular weight, theoretical pI , amino acid composition, atomic composition, extension coefficient, estimated half life, instability index, aliphatic index, and grand average of hydropathicity.

3. Results and discussion

3.1 Sequence analysis

The sequence of the topological domain of the protein, FtsH from *Mycobacterium tuberculosis* was retrieved from Uniprot KB in the FASTA format. Along with this, the annotated information about the same such as sequence length (760 AA), subcellular locations (cytoplasmic membrane), sequence similarity (belongs to AAA family), active site (460) was also retrieved from the Uniprot KB.

3.2 Identification of the template

The sequence which is obtained from the Uniprot KB was used as the input in BLAST-P server to find out the suitable template for the model building. In order to confirm the results the query sequence was again submitted to the FUGUE, a fold recognition program.

It was seen that the FtsH from *Thermatoga maritima* had 67% identity among the homologus sequences resulted from the BLAST-P server and the same protein showed the maximum Z score, 42.6 from FUGUE server. It shows the strongest match with the topological domain of FtsH from *Mycobacterium tuberculosis*.

3.3 The template for FtsH from *T.maritima* : 2CE7 A

2CE7 A, is the PDB code for the crystal structure of FtsH from *T.maritima*. It is a cell division protein obtained from *Thermotoga maritima,* which is a hyperthermophilic organism. The *Thermotoga maritima* is non-spore forming, rod shaped bacterium, it grows optimally at the temperature of 80^{0}c. *T.maritima* metabolizes many simple and complex carbohydrates. The length of this protein is 1860725 bp (1.8kb). This template shows 66% identity with the query sequence The structure has been solved and refined at 2.44Å resolution in a complex[16]. From the extracellular domain of FtsH of *M. tuberculosis*, the region from 199 – 448 was selected to build the model.

3.4 Model building

The refined sequence-sequence alignment obtained by BLAST-P was used to construct 3D model of topological domain of FtsH using MODWEB version r166 (Figure I).

3.5 Loop Modelling

The regions showing troughs in the Verify 3D graph of Modelled FtsH were considered for loop modelling with the help of Swiss PDB Viewer (SPDBV) (Figure II). The regions selected for loop modeling are listed below (Table I).

Table I: Regions selected for loop modelling

Loop No	Amino acid residues
1	247-250
2	241-246
3	235-240
4	230-234
5	232-236

While loop modeling, for each loop region anchor residues were carefully selected and the loop database of SPDBV was scanned. Of the loops obtained from the database, one was selected on its stereo-chemical compatibility (no bad phi/psi angle) and its side chains interaction with the rest of the structure (favorable interaction). Loops selected were added to the model one at a time and all the selected loop regions were remodeled. After remodeling the loop regions, the model was subjected to energy minimization using Swiss PDB Viewer.

After energy minimization, the model was again checked for its stereo-chemical quality and Verify 3D graph. In this model, few residues were appeared in disallowed region of Ramachandran plot. In Verify 3D graph also troughs were found, which indicates that the remodeled loops are energetically not stable.

From this model, again the errored loops were selected and remodeled using SPDBV. This was continued till we got the model, which satisfied the criteria of Ramachandran plot, Verify 3D graph and energy.

For the finally obtained FtsH model, the Verify 3D graph is given in Figure III. The regions of the troughs in the graph (Figure II) were found to be improved. The number of residues found in the different regions of Ramachandran plot (Figure IV) are as follows:

- Number of residues in favored region (~98.0% expected): 243 (98.0%)
- Number of residues in allowed region (~2.0% expected): 2 (0.2%)
- Number of residues in outlier region: 3 (1.2%)

The above mentioned results indicate that the FtsH

Score	Expect	Method	Identities	Positives	Gaps
353 bits (905)	3e-119()	Compositional matrix adjust.	167/250(67%)	204/250(81%)	0/250(0%)

Features:

```
Query 1  LYGPPGTGKTLLARAVAGEAGVPFFTISGSDFVEMFVGVGASRVRDLFEQAKQNSPCIIF 60
      L GPPGTGKTLLARAVAGEA VPFF ISGSDFVE+FVGVGA+RVRDLF QAK ++PCI+F
Sbjct 54  LVGPPGTGKTLLARAVAGEANVPFFHISGSDFVELFVGVGAARVRDLFAQAKAHAPCIVF 113

Query 61  VDEIDAVGRQRGAGLGGGHDEREQTLNQLLVEMDGFGDRAGVILIAATNRPDILDPALLR 120
      +DEIDAVGR RGAGLGGGHDEREQTLNQLLVEMDGF + G+I++AATNRPDILDPALLR
Sbjct 114 IDEIDAVGRHRGAGLGGGHDEREQTLNQLLVEMDGFDSKEGIIVMAATNRPDILDPALLR 173

Query 121 PGRFDRQIPVSNPDLAGRRAVLRVHSKGKPMAADADLDGLAKRTVGMTGADLANVINEAA 180
      PGRFD++I V PD+ GR+ +L +H++ KP+A D +L+ +AKRT G GADL N++NEAA
Sbjct 174 PGRFDKKIVVDPPDMLGRKKILEIHTRNKPLAEDVNLEIIAKRTPGFVGADLENLVNEAA 233

Query 181 LLTARENGTVITGPALEEAVDRVIGGPRRKGRIISEQEKKITAYHEGGHTLAAWAMPDIE 240
      LL ARE  IT  EEA+DRVI GP RK +IS EK+I AYHE GH + + +P+ E
Sbjct 234 LLAAREGRDKITMKDFEEAIDRVIAGPARKSLLISPAEKRIIAYHEAGHAVVSTVVPNGE 293

Query 241 PIYKVTILAR 250
      P+++++I+ R
Sbjct 294 PVHRISIIPR 303
```

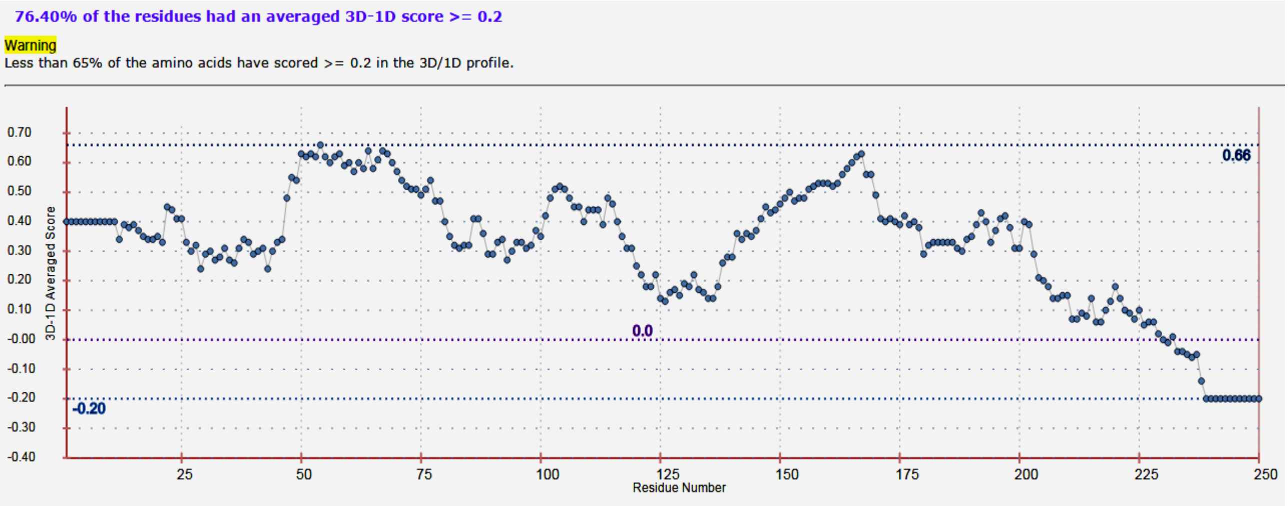
Region of the target sequence covered by the model : 84- 275

Template used : 2CE7 A

DOPE score(*) : - 0.03

Prosa 2003 Zscore(*) : - 43.41

Figure I: Alignment of template sequence (2ce7 A) with query sequence (FtsH) obtained from BLAST



76.40% of the residues had an averaged 3D-1D score >= 0.2

Warning

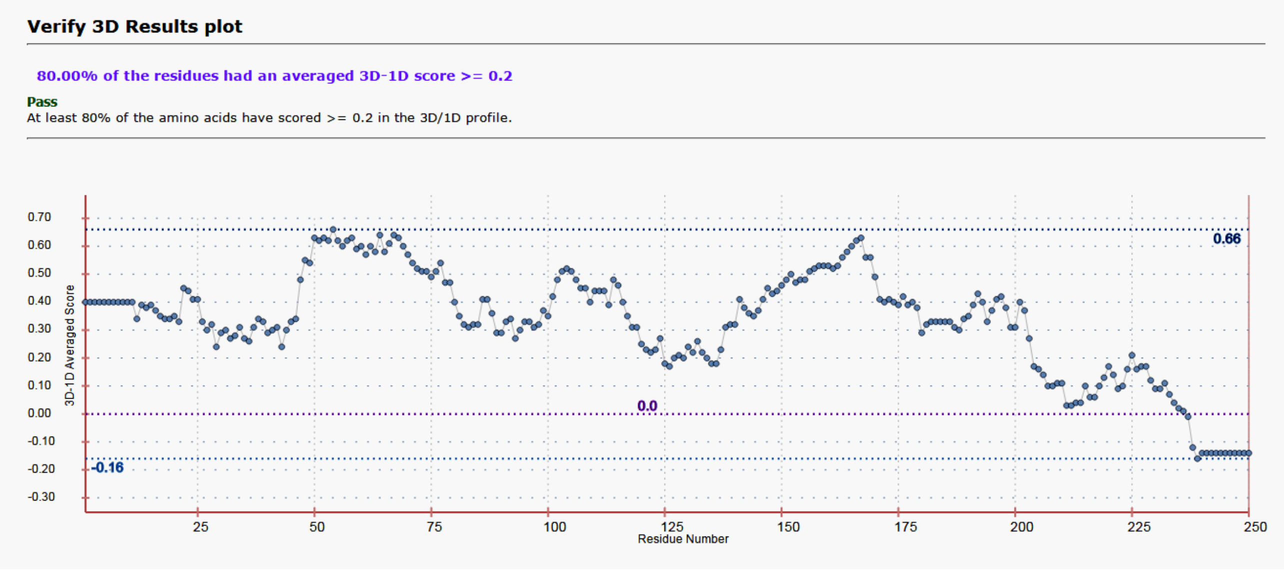
Less than 65% of the amino acids have scored >= 0.2 in the 3D/1D profile.

Figure II: Verify 3D Plot Before Loop Modelling

model (Figure V) is stereo-chemically satisfactory and the 3D structure obtained from this study can be used in developing novel inhibitors using the methods of rational drug designing.

3.6 Secondary structure prediction

The secondary structure of the topological domain of FtsH was predicted by the improved self-optimized prediction method (SOPMA) software (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html). The protein sequence of the topological domain of FtsH was input, and four conformational states, including helices, sheets, turns and coils, were analyzed. The parameters of similarity threshold and window width were set to 8 and 17, respectively, whilst the remaining parameters were left as default.



Verify 3D Results plot

80.00% of the residues had an averaged 3D-1D score >= 0.2

Pass
At least 80% of the amino acids have scored >= 0.2 in the 3D/1D profile.

Figure III: Verify 3D plot after loop modeling

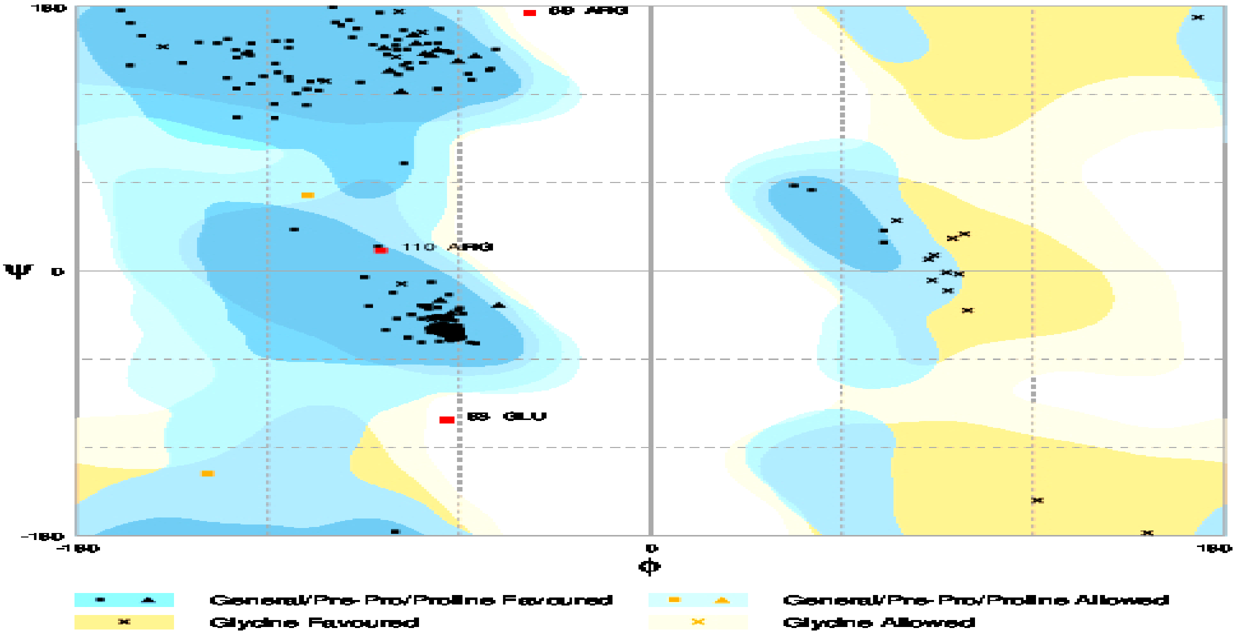


Figure IV: Ramachandran plot assessment after loop modeling

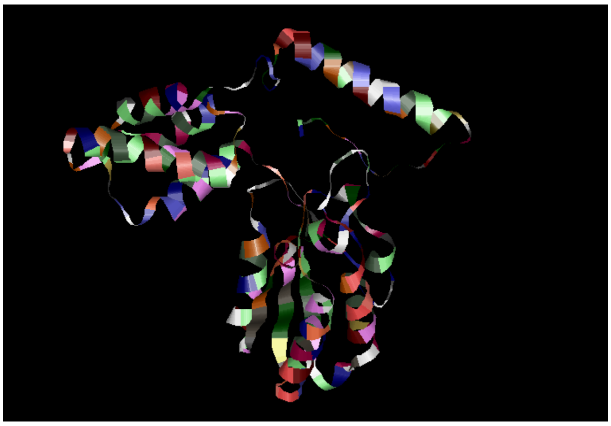


Figure V: The 3D structure of FtsH of *M. tuberculosis* after loop modeling as viewed in Chimera 1.5.3.

The predicted secondary structure results for the FtsH protein are shown in Figure VI. The results revealed that the proportion of random coils, β turns, α helices and extended strands (β folds) accounted for 32.00%,

SOPMA:

Alpha helix	(Hh)	:	80 is 32.00%
3_{10} helix	(Gg)	:	0 is 0.00%
Pi helix	(Ii)	:	0 is 0.00%
Beta bridge	(Bb)	:	0 is 0.00%
Extended strand	(Ee)	:	51 is 20.40%
Beta turn	(Tt)	:	31 is 12.40%
Bend region	(Ss)	:	0 is 0.00%
Random coil	(Cc)	:	88 is 35.20%
Ambiguous states	(?)	:	0 is 0.00%
Other states		:	0 is 0.00%

20.40%, 12.40% and 35.20% of the secondary structure, respectively.

Sequence length : 250

3.7 Analysis of Physical and chemical properties of the FtsH

The analysis of physical and chemical properties of the FtsH was carried out using ProtParam tool. The computed parameters including the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity are mentioned below:

Total number of negatively charged residues (Asp + Glu): 32

Total number of positively charged residues (Arg + Lys): 30

Number of amino acids: 250

Formula: $C_{1170}H_{1901}N_{345}O_{352}S_6$

Total number of atoms: 3774

Extinction coefficients: Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient 9970

Abs 0.1% (=1 g/l) 0.374, assuming all pairs of Cys residues form cystines

Ext. coefficient 9970

Abs 0.1% (=1 g/l) 0.374, assuming all Cys residues are reduced

Estimated half-life: The N-terminal of the sequence considered is L (Leu).

The estimated half-life is: 5.5 hours (mammalian reticulocytes, *in vitro*).

3 min (yeast, *in vivo*).

2 min (*Escherichia coli*, *in vivo*).

Instability index:

The instability index (II) is computed to be 28.82

This classifies the protein as stable.

Aliphatic index: 94.88

Grand average of hydropathicity (GRAVY): - 0.116

Top of Form

Bottom of Form

Atomic composition:

Carbon	C	1170
Hydrogen	H	1901
Nitrogen	N	345
Oxygen	O	352
Sulfur	S	6

4. Conclusion

Tuberculosis (TB) is among the top three leading causes of death by a single infectious agent worldwide. The FtsH is considered to be an attractive target because it is essential for the cell division. The protein structure model obtained from the present study can be used to understand the molecular interaction of this FtsH with other proteins. This modelled structure could also be used in rational drug designing to develop novel inhibitors which can be used for tuberculosis treatment.

Conflict of interest

The author's declares none.

References

1. Stephen H. Gillespie. Evolution of Drug Resistance in *Mycobacterium tuberculosis*: Clinical and Molecular Perspective. *Antimicrob. Agents Chemother.* (2002) **46** 2267-274

LYGPPGTGKTLLARAVAGEAGVPFFTISGSDFVEMFVGVGASRVRDLFEQAKQNSPCIIFVDEID
AVGRQ
eeccttcchhhhhhhhhhttttcceeeecccceeeeeeecccchhhhhhhhhhttcccceeeeehhhhhccc
RGAGLGGGHDEREQTLNQLLVEMDGFGDRAGVILIAATNRPDILDPALLRPGRFDRQIPVSNPD
LAGRRA
ttccccccccchhhhhhhhhhhhttcccttteeeeeeccccccccccthccttcccccccccccccccche
VLRVHSKGKPMAADADLDGLAKRTVGMTGADLANVINEAALLTARENGTVITGPALEEAVDR
VIGGPRRK
eeeectttccccchhhhhhhhhhhheecchhhhhhhhhhhhhhhhhcttttceeecccchhhhhhhhhtcccct
GRIISEQEKKITAYHEGGHTLAAWAMPDIEPIYKVTILAR
Teeeehttcceeeeectttcheeeecccccccchheeeeehh

Figure VI: Secondary structure prediction of topological domain of FtsH of *M. tuberculosis*

2. David M Roberts, Yoann Personne, Juliane Ollinger & Tanya Parish. Proteases in *Mycobacterium tuberculosis* pathogenesis: potential as drug targets. (2013) **8(5)** 621-631
3. de Bakker P.I.W., M.A. DePristo, R.P. Shetty, T.L. Blundell.. Discrete restraint-based protein modeling and the Cα-trace problem. *Protein Science*. (2003) **12** 2032-2046.
4. Cathy Wu CH1, Apweiler R, Bairoch A, Natale DA, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Mazumder R,O'Donovan C, Redaschi N, Suzek B. The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res*. (2006) **34(Database issue)** D187-91
5. Altschul,S.F., Gish,W., Miller,W., Myers, E.W. and Lipman,D.J.. Basic local alignment search tool. *J. Mol. Biol*. (1990) **215** 403–410.
6. Altschul, S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* (1997) **25** 3389–3402.
7. Scott McGinnis and Thomas L. Madden. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res*. (2004) W20-W25
8. Eshwar Narayanan, Urusulla ,Ben Webb; Pieper, et al., MODWEB (http://salilab.org/modweb) is a web server for automated comparative protein structure modeling that relies on MODPIPE for its functionality (2011).
9. Geourjon C and Deléage G .SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl Biosci*. (1995) **11(6)** 681-4.
10. Guex, N. and Peitsch, M.C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* (1997) **18** 2714-2723.
11. Peitsch, MC, Guex N. & Schwede T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis*. (2009) **30 Suppl 1:S** 162-73
12. Ramachandran, G.N., Sasisekharan, V. Conformation of polypeptides and proteins. *Adv. Protein Chem.* (1968) **23** 283-438.
13. Bowie JU, Lüthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science* (1991) **253 (5016)** 164-170.
14. David Eisenberg, Roland Lüthy, James U. Bowie.

<u>Questions raised</u>

Q1. Mention the table in text at appropriate place

Q2. Mention the figure 1 in text at appropriate place