In silico Analysis for Functional Predicition of *Salmonella typhi* Gene in Human Infection through Threading Model

Surina Bhadu[1] and Vishal Agrawal[2*]

[1]*Department of Chemistry and Biochemistry, COBS & H, CCS HAU, Hisar 125004, Haryana, India*
[2]*Department of Biochemistry, Panjab University, Chandigarh, India*

*Corresponding author

A B S T R A C T

Salmonella typhi is the Gram negative bacteria that have the ability to infect humans via transferring an effector protein into host cells through type III secretion system. This effector protein starts interfering various host mechanisms which helps in bacterial survival. Antibiotics activate the immune system against infection but sometimes bacteria develop resistance by modulating its mechanism. The present study was carried out to explore an uncharacterized gene which may be involved in infection enhancement. YejF gene was selected from transcriptomic data analysis and its function prediction was done by *in silico* studies with the help of similarity in structure of YejF model to their homologs. This study revealed that YejF protein is a part of transporter family which is enrolled in probably transferring the effector protein into human cells.

Introduction

Salmonella typhi is a Gram negative facultative anaerobic bacteria that causes enteric systemic diseases (typhoid and paratyphoid fever) in humans (Crump *et al.,* 2015; Crump and Heyderman 2014; Crump and Mintz 2010). *S. typhi* infections are most commonly observed in underdeveloped and developing countries. *S. typhi* has flagella all over the body which help in attaching to the host cell so that it has sufficient of time to transfer the protein other molecule that has the

ability to cause infection (Kawamoto *et al.,* 2013).

S. typhi gets entry into humans with food and water. In developing countries, (slum areas) food and water are not disinfected the microorganisms. *S. typhi* causes typhoid fever only in humans, immunodeficient individuals take more time to fight against infection and in such cases bacteria gets sufficient time to persist growing deeper and deeper in the tissues (Crump, Luby and Mintz 2004; Okoro *et al.,* 2012; Klemm *et al.,* 2016).

Bacteria are more advanced than the technologies as they become multidrug-resistant by modulating the infection pathways. *S. typhi* is resistant to all the drugs present in the market like Nalidixic acid (Cooke and Wain 2004; Spellberg *et al.,* 2008; Koirala *et al.,* 2012). The reason for this is that there is no animal model available to understand the mechanism of *S. typhi* infection because of which no new drug is developed that which may overcome the resistance problem.

To develop new drug which may combat the bacterial resistance is the main concern to protect humans from typhoid fever (Harish and Menezes 2011; Menezes *et al.,* 2012). The present *in silico* studies were conducted to assign the putative role of uncharacterized *Salmonella typhi* gene which are enrolled in the establishing the infection in humans. Faucher *et al.,* 2006 transcriptomics study says that there are several genes which are involved in establishing the infection as there levels change over 24 hours of infection.

Out of all these genes there are few whose functions are not characterized. *yejF* gene was selected for further analysis because its mRNA levels increased significantly after 24 hours of infection and this gene may have an important role in establishing the infection.

Materials and Methods

Various bioinformatics softwares were used to predict the function of *yejF*.

The YejF sequence was retrieved from the online server UniProt. UniProt is a server that provides the gene sequence and the general information of a particular protein on the basis of protein sequence as well as functional annotation with the help of protein sequence repository (Consortium, 2010). The URL address of UniProt is www.uniprot.org

Sequence information

The sequence information was collected from a database Ecocyc which deals with different metabolic pathways, transporters and regulators in many organisms (Keseler *et al.,* 2013). The URL address for the Ecocyc database is www.ecocyc.org.

The number of different domains, specific regions were characterized with the help of a protein family database i.e. Pfam which deals on the basis of Hidden Markov Models (Finn *et al.,* 2014). For more specific results, this database utilizes both manually and automatically curated sequences. Pfam can be accessed by the URL address pfam.sanger.ac.uk/

Based on the information that during evolution change occurs in the gene functions, the conserved domain sequence and structure from their homologs were retrieved. To explore the YejF homologs, BLAST (Basic Local Alignment Search Tool) was run which compared the sequence of nucleotide and protein in various organisms (Zhang *et al.,* 2000).

BLAST provides the results on the basis of sequence similarity as well as sequence coverage. The BLAST is a heuristic program. The URL address of BLAST is www.ncbi.nlm.nih.gov/BLAST/

Structure information

Based on the information of YejF sequence a YejF model was generated. There were two options for model generation- (i) homology model and (ii) threading model. Homology model is generated if homologs have high sequence similarity which can be based on BLAST results. The threading model was a better option for YejF which was done with LOMETS (Local Meta-Threading Server).

The LOMETS generates the ten best models from the nine different top most threading servers (Wu and Zhang, 2007). The YejF model and the homology proteins can be visualized using the offline ICM Browser. This software provides the function of superimposition from which structure similarity in homologous proteins with the YejF model can be easily analysed. ICM Browser can be downloaded from URL address www.molsoft.com

Results and Discussion

529 amino acids long sequence of YejF protein is retrieve from the UniProt. UniProt also gave the information that YejF protein has 58.72 KD molecular weight.

```
MTSPLLAIEN LSVGFRQQQH VRPVVNAISL
QVNAGETLAL VGESGSGKSV TALSILRLLP 60
TPPAVYLSGD IRFHGESLLH ASEQTLRGVR
GNKIAMIFQE PMVSLNPLHT LEKQLYEVLS 120
LHRGMRREAA RAEMIGCLDR VGIRQASQRL
RDYPHQLSGG ERQRVMIAMA LLTRPELLIA 180
DEPTTALDVS VQAQILSLLR ELQRELNMGL
LFITHNLSIV KKLADSVAVM QHGKCVENQR 240
ADTLLSAPTH PYTQKLLNSE PTGDPVPLPA
GQAPLLEVDK LRVAFPIRKG ILKRVVDHNV 300
VVNNISFTLH PGETLGLVGE SGSGKSTTGL
ALLRLIRSEG RIVFDGQSLD TLNRRQLLPV 360
RHRIQVVFQD PNSSLNPRLN VLQIIEEGLR
VHQPTLSGAQ REQQVKAVMM EVGLDPETRH 420
RYPAEFSGGQ RQRIAVARAL ILKPSLIILD
EPTSSLDKTV QAQILALLKS LQQKHRLAYI 480
FISHDLHVVR ALCHQVIVLR QGEVVEQGQC
ERVFTAPQQA YTRQLLALS          529
```

Sequence information

Ecocyc predicted that YejF is most probably a part of transporter family which is involved in an ATP Binding Cassette (ABC transporter). YejF can also be a part of the operon YejABEF.

Pfam reported that YejF protein is constitute of two different domains have ABC transporter role (Figure 1). Most probably they able to transfer a dipeptide or oligopeptide which may act as a effector protein in the humans to begin the infection.

BLAST results showed that several homologs are present to YejF. Top 5 ranked homologs are shown in Table 1.

Structure information

As the BLAST result showed that there is less sequence similarity between the YejF and its homologs, so the low sequence similarity homologs generates homology model of poor quality. To overcome this problem threading model would be a better option because it does not require high sequence similarity.

LOMETS server built the best threading model by using high statistical calculation from the nine best threading model servers. The top three ranked threading model listed by LOMETS are shown in Table 2. 3bk_7 generated by the FFAS-3D server was selected as the threading model out of the ten because it showed high sequence coverage and high alignment amongst all.

ICM Browser

High quality visualization of the model showed that YejF consists of two different domains as predicted by Pfam and both the domains are approximately of equal length. N-terminal of YejF has methionine at 1st position and the C-terminal has serine at 529th position. Superimposition of the YejF model with the protein having similar structure with the help of PDB database.

Table.1 Top 3 BLAST hits that have significant sequence alignment

BLAST Hits	Score	Total Score	Query coverage	Identity	PDB accession
Crystal structure of NBD of a dipeptide ABC transporter [thermoanaerobacter tengcongensis MB4]	145	275	96%	37%	4FWI_B
Crystal structure of methionine Importer Metni [Escherichia coli K-12]	125	250	98%	32%	3DHW_C
Inward Facing Conformation of the Metni Methionine Abc transporter Cy5 Native Crystal Form [Escherichia coli K-12]	124	246	98%	31%	3TUI_C

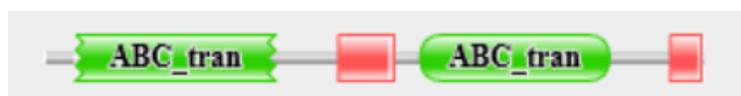**Table.2** Top 3 threading model generated for YejF by LOMETS software

Rank	Template	Align_length	Coverage	Seq_id	Confidence Score	Program
1	3bk7_A	476	0.899	0.22	High	FFAS-3D
2	3ozxA	465	0.879	0.18	High	PRC
3	4fwiB0	277	0.523	0.35	High	PGenTHREADER

Table.3 Threading model with the resolved structural proteins

Sr No.	PDB Id_Chain	Colour Code	RMSD (Å)	Residues aligned	Total residue	Percent-identity (%)	Organism
1	3bk7-A	Blue	1.3	473	593	21	*Pyrococcus abyssi*
2	1yqt-A	Yellow	1.5	460	515	21	*Pyrococcus furiosus*
3	3ozx-A	Green	1.7	456	514	17	*Sulfolobus solfataricus*
4	3ozx-B	Orange	1.8	455	515	17	*Sulfolobus solfataricus*
5	3j16-B	White	2.8	478	608	18	*Saccharomyces cerevisiae*
6	3j15-B	Pink	2.9	475	593	21	*Saccharomyces cerevisiae*
7	4crm-P	Cyan Blue	2.8	477	608	18	*Saccharomyces cerevisiae*

Best known structure protein having the highest structure similarity with YejF model was selected. Different colour codes were given to all the protein for better analyzing the structure during superimposition.

Fig.1 Domains predicted by Pfam. YejF is consists of two similar domains with ABC transporter family

Fig.2 Superimposition of YejF with other known proteins. YejF is very well superimposed with all other ABC transporter proteins

All the six selected proteins which had similar structure may or may not have similar sequence (Table 3). All the six selected proteins were ABC transporter proteins and were very well superimposed on the YejF model (Figure 2). This kind of superimposition showed that YejF is also an ABC transporter family protein.

Deaths due to *Salmonella typhi* infection is increasing continuously in the developing countries. *Salmonella typhi* modulates its infection pathways and become resistant to all the developed antibiotics presents till date. Therefore, to develop new drugs to inhibit *S. typhi* infection is the major concern in most affected countries.

The present study revealed that YejF protein is involved in the causing infection in humans. YejF is a part of a operon and itself a ABC transporter family protein that has the ability to transport oligopeptide into the humans via type III secretion system. If such a drug is synthesized which is oligopeptide in nature and has the ability to interfere with the operon YejABEF so that YejF is not able to transfer virulence protein in humans, it will be really a break through to prevent *Salmonella typhi* infection in coming years.

References

Consortium TU 2010. The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Research;* 38: 142-148.

Cooke FJ and Wain J. 2004. The emergence of antibiotic resistance in typhoid fever. *Travel Medicine and Infectious Disease;* 2: 67-74.

Crump JA and Heyderman RS. 2014. Invasive Salmonella infections in Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene*; 108: 673-5.

Crump JA and Mintz ED. 2010. Global trends in typhoid and paratyphoid fever. *Clinical* Infectious *Diseases;* 50: 241-246.

Crump JA, Luby SP and Mintz ED. 2004. The global burden of typhoid fever. *Bull World Health Organ;* 82: 346-53.

Crump JA, Sjolund-Karlsson M and Gordon MA. 2015. Epidemiology, clinical presentation, laboratory diagnosis, antimicrobial resistance, and antimicrobial management of invasive *Salmonella* infections. *Clinical Microbiology Review;* 28: 901-937.

Faucher SP, Porwollik S, Dozois, CM, McClelland M and Daigle F. 2006. Transcriptome of Salmonella enterica serovar Typhi within macrophages revealed through the selective capture of transcribed sequences. *Proceedings of the National Academy of Sciences of the United States of America;* 103: 1906-1911.

Finn RD, Bateman A, Clements J, Coggill P, Eberhard RY, Eddy SR, Heger A, Hetherington K, Holm L and Mistry J. 2014. Pfam: the protein families database. *Nucleic acids research;* 42: 222-230.

Harish BN and Menezes GA. 2011. Antimicrobial resistance in typhoidal salmonellae. *Indian Journal Medical Microbiology;* 29: 223-229.

Kawamoto A, Morimoto YV, Miyata T, Minamino T, Hughes KT, Kato T and Namba K. 2013. Common and distinct structural features of Salmonella injectisome and flagellar basal body. Scientific Reports 3: 3369.

Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, Bonavides-Martínez C, Fulcher C, Huerta AM, Kothari A and Krummenacker M. 2013. EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Research;* 41: 605-612.

Klemm EJ, Gkrania-Klotsas E, Hadfield J. Emergence of hostadapted *Salmonella enteritidis* through rapid evolution in an immunocompromised host 2016. *Nature Microbiology;* 1, DOI: 10.1038/nmicrobiol.2015.23.

Koirala KD, Thanh DP and Thapa SD. 2012. Highly resistant Salmonella enterica serovar Typhi with a novel gyrA mutation raises questions about the long-term efficacy of older fluoroquinolones for treating typhoid fever. *Antimicrobial Agents;* 56: 2761-2762.

Menezes GA, Harish BN and Khan MA. 2012. Antimicrobial resistance trends in blood culture positive *Salmonella typhi* isolates from Pondicherry, India, 2005-2009. *Clinical Microbiology Infection;* 18: 239-245.

Okoro CK, Kingsley RA and Quail MA. 2012. High-resolution single nucleotide polymorphism analysis distinguishes recrudescence and reinfection in recurrent invasive nontyphoidal *Salmonella typhi*murium disease. *Clinical Infection Diseases;* 54: 955-963.

Spellberg B, Guidos R and Gilbert D. 2008. The epidemic of antibiotic resistant infections: a call to action for the medical community from the Infectious Diseases Society of America. *Clinical Infection Diseases;* 46: 155-164.

Wu S and Zhang Y. 2007. LOMETS: A local meta-threading-server for protein structure prediction. *Nucleic Acids Research;* 35: 3375-3382.

Zhang Z, Schwartz S, Wagner L and Miller W. 2000. A Greedy Algorithm for Aligning DNA Sequences. *Journal of Computational Biology;* 7: 203-214.