

## Screening of New Potential Drug Target for *Salmonella typhi* Using Biostatistics Analysis

Rajat Kumar Saraf<sup>1</sup>, Nikita Bora<sup>1</sup>, Prashant Tiwari<sup>2</sup>, Prashant Ankur Jain<sup>1</sup>

<sup>1</sup>Jacob School of Biotechnology and Computational Biology, SHIATS, Allahabad, India.

<sup>2</sup>Department of Bioinformatics, CytoGene Research & Development, Lucknow, India.

### Abstract

The sequenced genome of Human and various pathogens has provided access to huge amount of data related to their genome as well as their proteome. This immense assembly of information proves to be very useful in the identification of the novel targets in pathogens. Such drug targets are then used to control the dreadful actions of various fatal diseases. In this study, the disease chosen was typhoid. It was found that typhoid is mainly caused due to the bacteria *Salmonella typhi*. The whole proteome of *Salmonella typhi* was retrieved from both UniProt and NCBI. The filtered essential proteins of pathogen, non homologous to human, represents potential drug targets. The complete proteome set of *Salmonella typhi* contains 19732 proteins (Uniprot). 6292 proteins were retrieved as non-redundant by CD-HIT program at 60% (identical) threshold. MSA analysis was performed to get the conserved region within these proteins. Sub-cellular localization was predicted using PA-SUB server to locate the outer membrane proteins which could be probable vaccine candidates. These represent a vast number of potential therapeutic drug targets because of their involvement in major biological processes in cell.

**Keywords:** Target Identification, *Salmonella typhi*, drugs, statistical analysis, *Homo sapiens*.

### 1. Introduction

*Salmonella enterica* serovar typhi is a human-specific gram-negative pathogen causing enteric typhoid fever, a severe infection of the reticuloendothelial system [1],[2],[3]. It has two strains CT18 (multiple drug resistant) [4] and Ty with a complete proteome of 4718

proteins. The early administration of antibiotic treatment has proven to be highly effective in eliminating infections, but indiscriminate use of antibiotics has led to the emergence of multidrug-resistant strains of *S. enterica* serovar Typhi [5]. Chloramphenicol was the drug for the treatment of this infection till plasmid mediated chloramphenicol resistance was encountered [6]. Following this ciprofloxacin became the mainstay of treatment being a safer and more effective drug than Chloramphenicol but after clinical resistance to treatment with ciprofloxacin in the patients suffering from enteric fever, the choice left now is an expensive drug like ceftriaxone or cefixime [6]. Resistance against ceftriaxone have been reported to CDC (Centre for Drug Control) [7] mild to moderate side effects have been shown for ceftriaxone. The novel targets identified by us using subtractive genomics will help enable understanding the biology of the pathogen to provide a more cost effective medication.

Pathogenesis of *S. Typhi* is incompletely understood, and treatment failure is not uncommon in the era of multidrug resistance [8]. The *Salmonella* genome contains clusters of virulence associated genes called pathogenicity islands (PAIs). Of 17 PAIs identified so far [9], functions of only SPI 1, 2, and 7 are partially known. Functional characterization of other PAIs will help to identify new drug/vaccine targets.

COBRA integrates and represents our current knowledge of network components and interactions and has been applied to several prokaryotic and eukaryotic organisms, including *Escherichia coli*, *Helicobacter pylori* [10], *Haemophilus Influenzae* [11], *Saccharomyces cerevisiae* [12] and *Geobacter sulfurreducens* [13]. By systematizing and providing context for transcriptomic, proteomic, and metabolomic

data, these models allow for the imposition of constraints, which together define the possible phenotypic behavior allowed by these in silico organisms. The use of constrained based modeling has significant potential to identify new antibacterial targets [14]. Modeling allows simulation of single gene knockouts and, more importantly, the effect of combinatorial deletions, which might allow the rational development of combination drug therapy. The use of such a modeling approach has been demonstrated for H. Influenza [11] and for Mycobacterium tuberculosis [15]. As the number of pathogens with such models grows [16], [17],[18],[19],[20],[21],[15] one might expect that more basic principles of bacterial metabolism needed for pathogenesis can be derived. The 4,857-kilobase (kb) chromosome of *Salmonella enteric serovar typhimurium* LT2 (*S. Typhimurium* LT2) accounts for 4,489 coding sequences (CDS/ORFs) including 39 pseudogenes and a 94 kb plasmid encoding 108 ORFs [22].

Another unique aspect of our reconstruction are the reactions involved in free radical quenching that are generally used by *Salmonella* to counteract the host cell response to infection [23],[24]. In addition to the classically known mechanisms mediated by superoxide dismutases, catalases and the enzymes that make and recycle glutathione and thioredoxin, our reconstruction includes mechanisms for reactive nitrogen species (RNS) and reactive oxygen species (ROS) resistance and removal. Some of those include: a soluble flavohemoglobin (Hmp) with a NO dioxygenase activity that is induced solely by RNS; and an alkyl hydroperoxide reductase (AhpC) belonging to the peroxiredoxin family that is involved in dissipating both ROS and RNS [24],[25],[26],[27]. The inclusion of reactions for ROS (superoxide, peroxide) and RNS (nitrate, nitrite, nitric oxide, and nitrous oxide) species make this model a suitable platform to analyze their role in infection and pathogenesis.

## Material and Method

### 1. Retrieval of proteomes of host (*Homo sapiens*) and pathogen (*Salmonella typhi*)

The whole proteome of *Salmonella typhi* was retrieved from both UniProt and NCBI. UniProt sequence retrieved were 19732 whereas for *Homo sapiens* whole proteome was 70,000.

## 2. Essential proteins identification for *S.typhi*

The complete proteome set of *Salmonella typhi* contains 19732 proteins (Uniprot). 6292 proteins were retrieved as non-redundant by CD-HIT program at 60 % (identical) threshold. MSA analysis was performed to get the conserved region within these proteins after removing the gaps from the result of MSA we considered this as a single protein. BlastP was performed taking MSA result as query sequence against the non redundant protein to get the template or good probabilistic target.

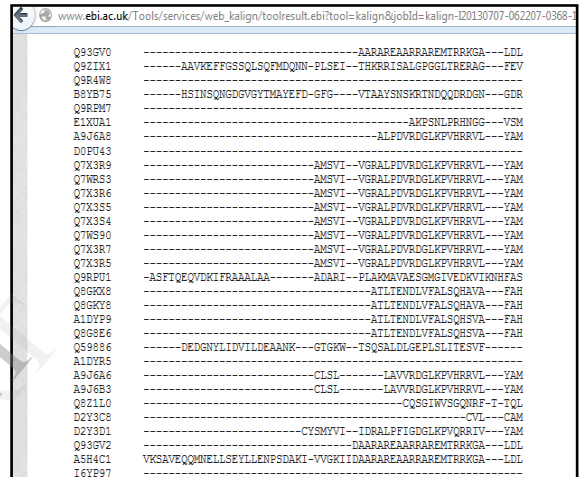


Fig 1: Multiple Sequence Alignment of full non redundant sequence of *salmonella typhi*.

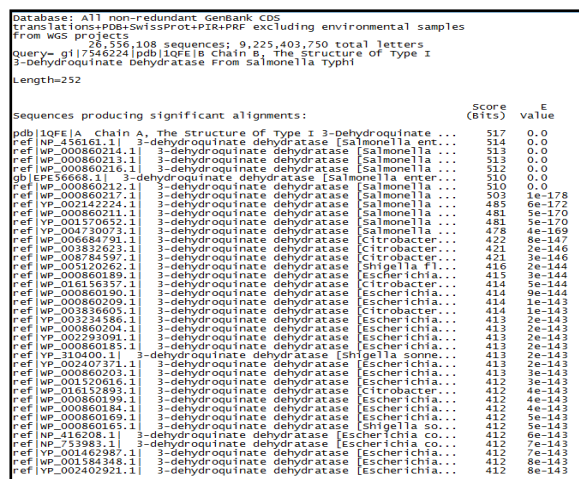


Fig 2: BLAST output of MSA result

## Metabolic Pathway Analysis

To understand the interactions of target protein with other molecules could be analyzed by looking into the metabolic pathways in which the disease specific target proteins are involved. Thus the significance of the target novel protein for the pathogen is explained by the metabolic pathway analysis. Metabolic pathway analysis of the essential proteins of *Salmonella typhi* was done by KEGG Automatic Annotation Server (KASS). The comparative analysis of metabolic pathways of the host and the pathogen revealed 3 pathways that were found to be distinctive in *Salmonella typhi*. 7 proteins were not participating in any of the metabolic pathways.

## Protein functional family prediction

Protein functional family prediction provides important information regarding structure, activity and metabolic roles. The essential proteins identified in the pathogen comprises of 10(target) putative uncharacterized proteins. Protein family classification allows probable function assignment for the uncharacterized protein. These 10 putative uncharacterized proteins were characterized and classified by SVM-Prot web server. Out of these 10 protein most of the proteins showing the more than 1 type family consider as primary and secondary function based on their R-value and P-value (%). These transmembrane and transporter proteins predicted by SVM-Prot web server may be considered as effective drug targets.

## Sub-cellular localization

Sub-cellular localization was predicted using Proteome Analyst Specialized Subcellular Localization Server v2.5 (PA-SUB) to locate the outer membrane proteins which could be probable vaccine candidates. 2 outer membrane proteins were identified (3-dehydroquinate dehydratase and dihydrofolate reductase). Membrane proteins have vital role in cellular communications, signal transduction, transport of ions, metabolites and other molecules. These represent a vast number of potential therapeutic drug targets because of their involvement in major biological processes in cell.

```

SeqID: 224H:A|PDBID|CHAIN|SEQUENCE
Analysis Report:
  CDSW- unknown [No details]
  CytoSW- unknown [No details]
  EC5W- unknown [No details]
  ModSW- unknown [No internal helices found]
  Motif- unknown [No motifs found]
  OMotif- unknown [No motifs found]
  PPSW- unknown [No details]
  Profile- unknown [No matches to profiles found]
  SCL-BLAST- unknown OuterMembrane [matched 730159: Copper homeostasis protein cutF precursor]
  SCL-BLAST- unknown OuterMembrane [No matches against database]
  Signal- unknown [No signal peptide detected]
Localization Scores:
  OuterMembrane 9.92
  Periplasmic 0.03
  Extracellular 0.03
  CytoplasmMembrane 0.01
  Cytoplasmic 0.01
Final Prediction:
  OuterMembrane 9.92

SeqID: 224H:A|PDBID|CHAIN|SEQUENCE
Analysis Report:
  CDSW- unknown [No details]
  CytoSW- unknown [No details]
  EC5W- unknown [No details]
  ModSW- unknown [No internal helices found]
  Motif- unknown [No motifs found]
  OMotif- unknown [No motifs found]
  PPSW- unknown [No details]
  Profile- unknown [No matches to profiles found]
  SCL-BLAST- unknown OuterMembrane [matched 730159: Copper homeostasis protein cutF precursor]
  SCL-BLAST- unknown OuterMembrane [No matches against database]
  Signal- unknown [No signal peptide detected]
Localization Scores:
  OuterMembrane 9.92
  Periplasmic 0.03
  Extracellular 0.03
  CytoplasmMembrane 0.01
  Cytoplasmic 0.01
Final Prediction:
  OuterMembrane 9.92
  
```

Fig 3: psort result for Lipoprotein

Bioinformatics & Drug Design Group [BIDD]  
**SVMProt: Protein Functional Family Prediction**  
 Protein functional families currently covered by SVMProt and statistics of data sets - prediction results.

Welcome to SVMProt network service

Query=Sequence  
 Length=466 amino acids

Classification Done.

Based on Support Vector Machine classification, your protein may belong to the following families:

Protein Family Name	R.Value	P.Value(%)
Zinc-binding	5.9	99.0
All lipid-binding proteins	2.5	90.3
EC 4.1.-.- Lyases - Carbon-Carbon Lyases	1.6	76.2
Metal-binding	1.1	62.2
Calcium-binding	1.0	58.6
Outer membrane	1.0	58.6
DNA repair	1.0	58.6
Copper-binding	1.0	58.6
Magnesium-binding	1.0	58.6

Citation: C.Z. Cai, L.Y. Han, Z.L. Ji, X. Chen, Y.Z. Chen. (2003) SVM-Prot: Web-Based Support Vector Machine Software for Functional Classification of a Protein from its Primary Sequence. *Nucleic Acids Res.* 31: 3692-3697

Fig 13: protein family prediction of Lipoprotein

## Result and Discussion

Considering the above properties, it was found that out of the 10 target proteins, 3 proteins satisfied most of the properties. These 3 target proteins were Type 1 fimbrial protein, Dihydrofolate reductase and 3-Dehydroquinate dehydratase which could be the potential targets.

Various parameters like melting point, water solubility, Log P, Log S, Pka for the known drug compounds were analysed from DrugBank. Both experimental and predicted properties of the different parameters were studied.

**Table 1: ANOVA Analysis**

Summary of analysis	Result
Total no. of proteins (uniprot)	19732
Total no. of non-redundant sequence by using CD-Hit at 60% identical	6292
Essentials proteins (DEG)	353
Total protein chosen as target (all different)	10
Total no. of protein participating in metabolic pathways	3
Pathways unique to <i>Salmonella typhi</i>	3
Outer membrane essential protein of <i>Salmonella typhi</i>	1
ANOVA analysis for known drug molecules (experimental)	Ho1: .91, Ho2: .81
ANOVA analysis for known drug molecules (predicted)	Ho1: 2.80, Ho2: 32.66
ANOVA analysis for new drug target molecules (predicted)	Ho1: .60 Ho2: 66
Probabilistic result for the single protein toward single properties	.20

Table 1: For the new potential drug molecules, one hypothesis i.e., Ho1 was accepted and Ho2 was rejected.

The ligands present in the 3 target proteins were analyzed and based on it template ligands were formed and properties analyzed through OSIRIS. The different properties which were similar to DrugBank properties were taken like ClogP, solubility, Molecular weight, Druglikeness and Drug Score. These properties were then compared among the different template ligands through ANOVA. The drug score combines druglikeness, cLogP, logS, molecular weight and toxicity risks in one handy value than may be used to judge the compound's overall potential to qualify for a drug.

To validate the calculated properties of the ligands for the 3 drug target proteins 3-Dehydroquinase dehydratase, Type 1 fimbrial protein, Dihydrofolate

reductase to that of known drug compounds taken from Drug Bank (Amoxicillin, Azithromycin, Chloramphenicol, Ceftriaxone, Nalidixic acid, ciprofloxacin, fluoroquinolone), an ANOVA hypothesis was set.

## Conclusion

The large scale genome sequencing projects have increased the availability of completely sequenced genomic and proteomic data in public domain [28]. Screening and analysis of these large biological sequence data provide new opportunities to understand and combat both infectious and genetic diseases in humans. There is a budding need for new drugs and vaccines to treat and prevent emerging and neglected infectious diseases. Subtractive genomics is a powerful tool for exploring new therapeutic targets. The current study based on subtractive proteomics approach helped in the identification and characterization of the potential essential proteins that could be targets for efficient drug designing against *Salmonella typhi*. Screening these potential targets against drug bank might be useful in the discovery of potential therapeutic compounds against *Salmonella typhi*.

The bacteria *Salmonella typhi* was studied and the proteome of this bacterium was analyzed. Statistical and probabilistic calculation was performed to validate the result obtaining from proteomic analysis of *Salmonella typhi*. The 3 target proteins were found to be efficient towards the bacteria *Salmonella typhi* and the properties of these proteins were found to be statistically similar to that of the known drug compounds towards this bacterium.

## Acknowledgment

We are very thankful to CytoGene Research & Development for providing the computational facility laboratory.

## References

1. P Everest et al., (2001). The molecular mechanisms of severe typhoid fever. Trends in Microbiology, 316-320.
2. Galan JE., (1996). Molecular genetics bases of *Salmonella* entry into host cells. Mol Microbiol, 263-271.

3. Jones BD et al., (1996). Salmonellosis: host immune responses and bacterial virulence determinants. *Annu Rev Immunol*, 533-561.
4. J Parkhill et al., (2001). Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature*, 848-852.
5. B Rowe et al., (1997). Multidrug-resistant *Salmonella typhi*: a worldwide epidemic. *Clinical Infectious Diseases*, 106-109.
6. A Kapil et al., (1994). *S. typhi* with transferable chloramphenicol resistance isolated in Chandigarh during 1983-87. *Indian J Pathol Microbiol*, 179-183.
7. E. Steinburg et al., (1999). Antimicrobial Resistance of *Salmonella typhi* in the United States: the National Antimicrobial Monitoring System (NARMS).
8. Crump JA et al., (2010). Global trends in typhoid and paratyphoid Fever. *Clin Infect Dis*, 241-246
9. Vernikos GS et al., (2006). Interpolated variable order motifs for identification of horizontally acquired DNA: Revisiting the *Salmonella* pathogenicity islands. *Bioinformatics*, 2196-2203.
10. Schilling CH et al., (2002). Genome-scale metabolic model of *Helicobacter pylori* 26695. *J Bacteriol*, 4582-4593.
11. Raghunathan A et al., (2004). In Silico Metabolic Model and Protein Expression of *Haemophilus influenzae* Strain Rd KW20 in Rich Medium. *OMICS*, 25-41.
12. Forster J et al., (2003). Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res*, 244-253.
13. Mahadevan R et al., (2006). Characterization of metabolism in the Fe(III)-reducing organism *Geobacter sulfurreducens* by constraint-based modeling. *Appl Environ Microbiol*, 1558-1568.
14. Trawick JD et al., (2006). Use of constraint-based modeling for the prediction and validation of antimicrobial targets. *Biochem Pharmacol*, 1026-1035.
15. Jamshidi N et al., (2007). Investigating the metabolic capabilities of *Mycobacterium tuberculosis* H37Rv using the in silico strain iNJ661 and proposing alternative drug targets. *BMC Syst Biol*, 1-26.
16. Thiele I et al., (2005). Expanded metabolic reconstruction of *Helicobacter pylori* (iT341 GSM/GPR): an in silico genome-scale characterization of single- and double deletion mutants. *J Bacteriol*, 5818-5830.
17. Heinemann M et al., (2005). In silico genomescale reconstruction and validation of the *Staphylococcus aureus* metabolic network. *Biotechnol Bioeng*, 850-864.
18. Oberhardt MA et al., (2008). Genome-scale metabolic network analysis of the opportunistic pathogen *Pseudomonas aeruginosa* PAO1. *J Bacteriol*, 2790-2803.
19. Becker SA et al., (2005). Genome-scale reconstruction of the metabolic network in *Staphylococcus aureus* N315: an initial draft to the two-dimensional annotation. *BMC Microbiol*, 5-8.
20. Baart GJ et al., (2007). Modeling *Neisseria meningitidis* metabolism: from genome to metabolic fluxes. *Genome Biol*, 8-136.
21. Chavali AK., (2008). Systems analysis of metabolism in the pathogenic trypanosomatid *Leishmania major*. *Mol Syst Biol*, 4-177.
22. McClelland M et al., (2001). Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature*, 852-856.
23. Shiloh MU et al., (2000). Reactive nitrogen intermediates and the pathogenesis of *Salmonella* and mycobacteria. *Curr Opin Microbiol*, 35-42.
24. Janssen R et al., (2003). Responses to reactive oxygen intermediates and virulence of *Salmonella typhimurium*. *Microbes Infect*, 527-534.
25. Mills PC et al., (2005). Detoxification of nitric oxide by the flavorubredoxin of *Salmonella enteric* serovar Typhimurium. *Biochem Soc Trans*, 198-199.
26. De Groote MA et al., (1995). Genetic and redox determinants of nitric oxide cytotoxicity in a *Salmonella typhimurium* model. *Proc Natl Acad Sci USA*, 6399-6403.
27. Poole LB., (2005). Bacterial defenses against oxidants: mechanistic features of cysteine-based peroxidases and their flavoprotein reductases. *Arch Biochem Biophys*, 240-254.
28. Bhawna Rathi et al., (2009). Genome subtraction for novel target definition in *Salmonella typhi*. *Biomedical Informatics*, 143-150.