bioRxiv preprint doi: https://doi.org/10.1101/521518; this version posted January 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Identification of potential new vaccine candidates in *Salmonella typhi* using reverse vaccinology and subtractive genomics-based approach

Sumit Mukherjee [a] *, Kaustav Gangopadhay [b], Sunanda Biswas Mukherjee [c] *

[a] Department of Physical Sciences, Indian Institute of Science Education and Research- Kolkata, Mohanpur - 741246, Nadia, West Bengal, India

[b] Department of Biological Sciences, Indian Institute of Science Education and Research- Kolkata, Mohanpur - 741246, Nadia, West Bengal, India

[c] Department of Bioinformatics, Maulana Abul Kalam Azad University of Technology, BF-142, Sector 1, Salt Lake, Kolkata - 700064, West Bengal, India

***Corresponding author**

E-mail: mukherjee.sumit89@gmail.com

E-mail: somabiswas1991@gmail.com

bioRxiv preprint doi: https://doi.org/10.1101/521518; this version posted January 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ABSTRACT:

Salmonella enterica serovar typhi is the causative agent of typhoid fever in human. The available vaccine lacks the effectiveness and further overuse of the antibiotics throughout the past decades lead to the emergence of antibiotic-resistant strains. To reduce the spread of antibiotic resistance, it is essential to develop new vaccines. In this study using an extensive bioinformatic analysis based on reverse vaccinology and subtractive genomics-based approach, we searched the genome of *S typhi*. We found that two outermost membrane proteins LptD and LptE, which are responsible for the final transport of the LPS across the membrane having most of the essential signature for being vaccine target. Both of these proteins are found highly conserved and possess shared surface epitopes among several species of *Salmonella* pathogens. This indicates that these two proteins could be the potential target for vaccine development, which could be effective for a broad range of *Salmonella* pathogens.

Keywords: Reverse vaccinology; Subtractive genomics; LPS transport; Salmonella typhi.

bioRxiv preprint doi: https://doi.org/10.1101/521518; this version posted January 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1. Introduction:

Salmonella enterica serovar typhi (*S typhi*) is a gram-negative human-restricted pathogenic bacterium that is the causative agent of an acute febrile illness known as typhoid fever. The World Health Organization (WHO) recognizes infection by *S typhi* as a persistent threat in developing and underdeveloped countries. The CDC reported that, globally, 22 million people affected by *S typhi* annually, resulted in approximately 600,000 deaths [1]. Mechanism of the pathogenesis of *S typhi* is still not completely understood. Consequent use of broad-spectrum antibiotics revealed the emergence of multi-drug resistant strains which is a major public health concern [2]. Vaccination is the most effective way, which can help to reduce the spread of antibiotic resistance. The resistance of the vaccine is very rare because pathogen populations create less variation for vaccine resistance than they do for drug resistance. The main underlying principles behind it that vaccines tend to work prophylactically by inducing immune responses against multiple pathogen proteins (antigens), and various epitopes on each antigen while drugs tend to work therapeutically by interfering with a specific step in a particular metabolic pathway [3]. At present, mainly two vaccines are available against typhoid fever which includes inactivated whole cell vaccines and live attenuated vaccines [4][5]. All of these vaccines have less than 80% efficacy [6][7]. Therefore, the development of new vaccines is necessary. Identification of potential vaccine target is the first step towards the development of a new vaccine. The advent of high–throughput sequencing technology provided a vast amount of genome sequences data of pathogens, which can be mined to explore the new potential vaccine targets. Reverse vaccinology is an *in-silico* approach that intended to detect potential vaccine protein targets from the genome of the pathogenic organisms [8][9]. Reverse vaccinology was first applied to pathogen *Neisseria meningitides* serogroup B (MenB) to identify vaccine candidates [10]. The vaccine deriving from the mining of the genome of *Neisseria meningitides* serogroup B was entered into the phase I clinical testing in 2002 after successful preclinical studies [11]. In November 2012, the first vaccine named Bexsero against meningococcal

bioRxiv preprint doi: https://doi.org/10.1101/521518; this version posted January 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

serogroup B was granted for marketing by European Medicinal Agency and a few months later in January 2013, it was also approved by the European Commission [12]. In the meantime, reverse vaccinology based approach has been applied to mining genome of pathogens such as *Bacillus anthracis* [13], *Streptococcus pneumonia [14]*, *Mycobacterium tuberculosis [15]*, *Cryptosporidium hominis [16]*, *Helicobacter pyroli* [17] and *Treponema pallidum [18]*.

In this present study, we have developed reverse vaccinology and subtractive genomics based computational strategy and screened the genome of *S typhi* and detected three highly potential vaccine candidates those contain most of the signatures of the ideal vaccine. All of the three proteins detected using this approach are essential for bacterial survival, extracellular membrane proteins and involved in the secretory pathway, containing the virulence determining factors and do not have any human homologous. Antigenic epitopes are present in those three proteins that have an affinity to bind B-cells and T-cells. Protein-protein interaction analysis reveals that those proteins actively interact with each other and all of them are involved in the LPS transport. Therefore, our findings put forward a set of new potential vaccine candidates of *S typhi*. This work can act as a benchmark and provide valuable input for further experimental and clinical testing.

bioRxiv preprint doi: https://doi.org/10.1101/521518; this version posted January 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2. Materials and Methods:

2.1. Dataset collection:

We retrieved the protein coding sequences of four strains (NC_003198.1, NC_016832.1, NC_021176.1, and NZ_CP012091.1) of *Salmonella enterica serovar typhi* from RefSeq database. We have performed the pan-genome analysis and those genes which are present in all of the four strains of *S typhi* were considered for our studies.

2.2. Selection of essential proteins:

Essential proteins of pathogens are crucial for major functions for their survival within the host. Database of Essential Genes (DEG) [19] consists of the experimentally validated data for the essential genes within a genome. To identify the essential genes we performed BLASTP [20] searches of the protein sequences of all the strains of *S typhi* against DEG databases. The parameters for considering the essential genes were set as follow: E-value ≤ 10-5, similarity ≥ 50 %, and the sequence coverage ≥ 30 %. Protein sequences of *S typhi* those have significant sequence similarity with the DEG protein sequences were retrieved.

2.3. Search for virulence factors in *S typhi*:

Virulence factors are molecules synthesized by pathogens that contribute to the pathogenicity to infect the host. We predicted the pathogenic ability of the *S typhi* essential proteins using MP3 server (http://metagenomics.iiserb.ac.in/mp3/index.php) [21] (Gupta et al., 2014). This web-server predicts pathogenic and virulent proteins using an integrated SVM-HMM approach with high accuracy.

2.4. Secretory pathway analysis:

Bacterial pathogens exploit diverse methods to invade mammalian hosts. One of the important strategies for many pathogens is the secretion of proteins across phospholipid membranes. Secreted proteins can take part in crucial roles in promoting bacterial virulence, by facilitating attachment to host cells and disrupting their functions [22]. SignalP web-server [23] was used to predict the signal

bioRxiv preprint doi: https://doi.org/10.1101/521518; this version posted January 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

peptide triggered protein secretion and SecretomeP 2.0 web-server [24] was used for predictions of non-classical i.e. not signal peptide triggered protein secretion.

2.5. Prediction of protein subcellular localization:

Subcellular localization of *S typhi* proteins was predicted using two different web servers: CELLO v.2.5 (http://cello.life.nctu.edu.tw/) [25] and SubCellProt (http://www.databases.niper.ac.in/SubCellProt) [26]. CELLO predicts protein subcellular localization using two-level support vector machine (SVM). While, SubCellProt is based on two machine learning approaches, k Nearest Neighbor (k-NN) and Probabilistic Neural Network (PNN). When at least two of these three approaches (k-NN, PNN, and SVM) predict the same localization of a protein we consider that as its subcellular localization.

2.6. Screening of non-homologous proteins:

To identify non-homologous proteins of *S typhi*, we used a systematic approach based on homology search. BLASTP [20] followed by TBLASTN filtering approach ($E<10^{-5}$ and use of low-complexity filters) was used against all the human proteins to detect the *S typhi* proteins which have no homologs with humans.

2.7. Prediction of potential antigenic epitopes:

The potentiality of the peptides to activate the human immune system are identified in order to generate memory cells against *S typhi*. To understand the immunomodulatory effect of these peptides, we have analyzed different types of epitopes. LBtope [27] was used to predict the linear B-cell epitopes. We used the LBtope model [27] with default threshold of 60% probability cut-off. Conformational B-cell epitopes were predicted using CBTOPE [28]. The major role of the MHC is to bind to antigenic regions (or peptides) and present these peptides on the cell surface, where appropriate T-cells recognize these peptides. Thus, it is important to predict MHC binders, as these binders have the ability to activate the T-cells of the immune system. nHLAPred [29] was used to predict MHC class-I binding peptide. In addition to the prediction of MHC binders which are

bioRxiv preprint doi: https://doi.org/10.1101/521518; this version posted January 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

potential T-cell epitopes, we also predicted CTL epitopes using a direct method, CTLPred [30]. CTLPred [30] is a direct method that predicts T-cell epitopes (CTL) from a primary sequence of the antigen instead of using the intermediate step where MHC Class I binders are predicted.

2.8.Molecular weight calculation:

Protein molecular weight calculated using ExPasypI/Mw tool [31].

2.9. Protein interaction network analysis:

A protein-protein interaction network of the selected proteins was constructed using the STRING database [32]. In the present study, all interactors with low confidence scores i.e., less than 0.400 were screened in order to avoid false positives and false negatives. Only medium and high confidence interactions with not more than 50 interactors were included in the interaction network. The protein networks constructed for selected proteins were then analyzed using Cytoscape 2.8.1 [33]. In order to evaluate the importance of the query node, each node in the protein network was deleted one by one and the change in clustering coefficient was analyzed [34][35]. Pathogen-specific essential proteins important in the metabolic network were regarded as potential targets.

bioRxiv preprint doi: https://doi.org/10.1101/521518; this version posted January 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3. Results:

Reverse vaccinology is a promising approach for genome-wide identification of potential vaccine candidates. Systematically genome-wide screening is crucial for the identification of vaccine candidates from pathogens. We developed a systematic extensive computational screening approach based on reverse vaccinology and subtractive genomics for the identification of potential vaccine candidates from *S typhi*. Most of the important characteristic features that required for being potential vaccine candidates were considered for our analysis. The detailed computational pipelines are described in the methods section. Figure 1 represents the flow diagram of our detailed screening approaches and computational pipelines. For the initial screening step, we have selected only the essential genes of *S typhi*. Essential genes are involved in basic metabolic pathways and are crucial for the survival of any organisms. Therefore, in microbial pathogens, those genes constitute attractive targets for antimicrobial drug and vaccine design [36][37][38][39]. After the first step of screening, we have found total 358 genes, which are essential for *S typhi*. In the second step from the essential genes datasets, we have screened those genes, which comprise potential virulence factors. Virulence factors facilitate the pathogens to invade the host and evade host defenses. Genes having virulence factor induces pathogenicity within hosts and therefore, those genes can be a potential target for vaccines [40]. We have found only 20 proteins from essential genes datasets having significant pathogenic abilities. In the next step, we have searched the secreted proteins. Secreted proteins have vital roles in promoting bacterial virulence, from enhancing attachment to eukaryotic cells to scavenging resources in an environmental niche, to directly intoxicating target cells and disrupting their functions [22]. Therefore, protein secretion systems have an impact on bacterial pathogenesis and secretory proteins are ideal for vaccine targets. From the set of 20 protein sequences, we have predicted that 11 proteins have a role in the secretory pathway. In the next screening steps, we have select the extracellular membrane proteins. Membrane proteins play a key role in various bacterial physiology and pathogenesis and are the most important targets for vaccine development. We have

bioRxiv preprint doi: https://doi.org/10.1101/521518; this version posted January 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

analyzed the previously screened 11 proteins and predicted that 4 proteins are membrane proteins. Pathogen proteins those are not present in the humans are significant for vaccine targets because targeting inhibitors against these non-homologous sequences would have minimal side effects in humans [17][41]. Therefore, we have screened the protein sequences, which are non-homologous to human. From previously screened 5 protein datasets we have found 3 proteins which do not have sequence similarity with humans. Proteins having molecular weight <110 kDa are considered to be more effective targets because proteins having lower molecular weight can easily be purified and efficiently subjected to vaccine development [42]. We have found all of these three proteins have a molecular weight less than <100 kDa. Interacting partners of a protein at the cellular level can reveal many important signaling pathway-stimulating factors and some crucial cross-talks between potential therapeutic targets. Our protein interaction analysis revealed all these three proteins interact with each other and involved in bacterial LPS transport. The topological analysis of the constructed network indicate that all these three proteins interacts with other essential proteins involved in many important pathways for pathogen survival. In order to evaluate the importance of the each of these selected proteins, each protein in the network was deleted one by one and the change in clustering coefficient was analyzed. We found that all of these three proteins are crucial and deleting one node change the clustering coefficient. Next, we analyzed these three proteins to find the potential epitopes. We predicted that LptD and LptE both of these proteins contain multiple epitopes, and each of these protein contains one such epitopes each which can recognize B-cell, T-cell and can bind the MHC class-I. The epitopes are highlighted in the structure of LptDE complex and it has been observed that the epitopes in LptD and LptE are present in close proximity (Figure 2). These epitopes have high VaxiJen score (provided in Table 1) indicate that LptD and LptE could be the potential vaccine target.

bioRxiv preprint doi: https://doi.org/10.1101/521518; this version posted January 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4. Discussion:

Lipopolysaccharides (LPS) are unique glycolipids that contribute to permeability barrier properties of the outer membrane of Gram-negative bacteria, thus helping it to survive in the harsh environment and evading attacks from the host immune system [43]. LPS also forms a permeation barrier that inhibits the hydrophobic antibiotics from entering the bacteria which promotes the antibiotic resistance [43][44]. LPS biogenesis is a complex process which are responsible for the transport of LPS across the cellular envelope and the assembly of LPS at the cell surface [45]. Transport of LPS to the outer membrane needs different transporter proteins named Lipopolysaccharide Transporter (Lpt) proteins. The Lpt system contains seven components, the soluble cytoplasmic protein LptB, inner membrane LptF, and LptG, periplasmic LptC anchored to the Inner Membrane by a transmembrane helix, soluble periplasmic LptA, periplasmic LptE anchored to the Outer Membrane by a lipid anchor attached to Cys19 and Outer membrane localized LptD (Figure 3). All of the seven Lpt proteins are essential for the LPS transport to the outer membrane [46]. Two proteins among the group form a plug and barrel conformation in which LptE is located inside the barrel of LptD, which is important for the transport of LPS across the membrane. The primary function of LptDE complex is to place the LPS in the Outer Membrane. According to the current model of LptDE function, LPS arrives from the Inner membrane to the periplasm at the periplasmic N-terminal domain of LptD, causing a conformational change of LptDE [47]. This enables LPS molecules to enter the interior of the barrel. LPS can then move through the lumen of LptD and hence selectively passing through the lateral opening of LptD into the extracellular leaflet of Outer Membrane [47]. Recent studies demonstrated that LptD have the strong ability to generate the immune responses against Vibrio infection and showed a significant reduction of bacterial growth and LPS level, and increased susceptibility to antibiotics which supports that LptD could be the promising target for the development of effective vaccines [48]. Our finding highlighted that LptD, and LptE both are the outer membrane protein responsible for final transport of LPS to the outer leaflet, was widely

bioRxiv preprint doi: https://doi.org/10.1101/521518; this version posted January 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

distributed and highly conserved, and possesses shared surface epitopes among the several major Salmonella pathogens. Both of these extracellular membrane proteins are essential and involved in the secretory pathway, containing the virulence factors and do not have any human homologous, and possess antigenic epitopes. Therefore, both of these proteins have most of the important signatures of being potential vaccines. Thus, LptD and LptE could be the prospective vaccine antigen and therapeutic target for the prevention and control of *Salmonella* infection. Further experimental studies based on our findings would facilitate the development of new vaccines against *S typhi* in near future. Moreover, the reverse vaccinology and subtractive genomics based computational approaches used in this study can also be used for screening the genomes of different pathogens to detect the potential vaccine targets.

Conflict of interest

None

Acknowledgments:

SM and KG are thankful to the IISER Kolkata for financial support.

List of abbreviations:

LPS: Lipopolysaccharide

S typhi: *Salmonella enterica serovar typhi*

BLAST: Basic Local Alignment Search Tool

bioRxiv preprint doi: https://doi.org/10.1101/521518; this version posted January 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References:

[1] J.A. Crump, E.D. Mintz, Global trends in typhoid and paratyphoid Fever., Clin. Infect. Dis. (2010). doi:10.1086/649541.

[2] F.J. Cooke, J. Wain, The emergence of antibiotic resistance in typhoid fever, Travel Med. Infect. Dis. (2004). doi:10.1016/j.tmaid.2004.04.005.

[3] D.A. Kennedy, A.F. Read, Why does drug resistance readily evolve but vaccine resistance does not?, Proc. R. Soc. B Biol. Sci. (2017). doi:10.1098/rspb.2016.2562.

[4] C.A. Guzman, S. Borsutzky, M. Griot-Wenk, I.C. Metcalfe, J. Pearman, A. Collioud, D. Favre, G. Dietrich, Vaccines against typhoid fever, Vaccine. (2006). doi:10.1016/j.vaccine.2005.07.111.

[5] K.A. Date, A. Bentsi-Enchill, F. Marks, K. Fox, Typhoid fever vaccination strategies, Vaccine. (2015). doi:10.1016/j.vaccine.2015.04.028.

[6] A. Fraser, M. Paul, E. Goldberg, C.J. Acosta, L. Leibovici, Typhoid fever vaccines: Systematic review and meta-analysis of randomised controlled trials, Vaccine. (2007). doi:10.1016/j.vaccine.2007.08.027.

[7] B.E. Mahon, A.E. Newton, E.D. Mintz, Effectiveness of typhoid vaccination in US travelers, Vaccine. (2014). doi:10.1016/j.vaccine.2014.04.055.

[8] R. Rappuoli, Reverse vaccinology, Curr Opin Microbiol. (2000). doi:10.1016/S1369-5274(00)00119-3.

[9] R. Rappuoli, A. Covacci, Reverse Vaccinology and Genomics, Science (80-.). (2003). doi:10.1126/science.1092329.

[10] M. Pizza, V. Scarlato, V. Masignani, M.M. Giuliani, B. Aricò, M. Comanducci, G.T. Jennings, L. Baldi, E. Bartolini, B. Capecchi, C.L. Galeotti, E. Luzzi, R. Manetti, E. Marchetti, M. Mora, S. Nuti, G. Ratti, L. Santini, S. Savino, M. Scarselli, E. Storni, P. Zuo, M. Broeker, E. Hundt, B. Knapp, E. Blair, T. Mason, H. Tettelin, D.W. Hood, A.C. Jeffries, N.J. Saunders, D.M. Granoff, J.C. Venter, E.R. Moxon, G. Grandi, R. Rappuoli, Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing, Science (80-.). (2000). doi:10.1126/science.287.5459.1816.

[11] M.M. Giuliani, J. Adu-Bobie, M. Comanducci, B. Aricò, S. Savino, L. Santini, B. Brunelli, S. Bambini, A. Biolchi, B. Capecchi, E. Cartocci, L. Ciucchi, F. Di Marcello, F. Ferlicca, B. Galli, E. Luzzi, V. Masignani, D. Serruto, D. Veggi, M. Contorni, M. Morandi, A. Bartalesi, V. Cinotti, D. Mannucci, F. Titta, E. Ovidi, J.A. Welsch, D. Granoff, R. Rappuoli, M. Pizza, A universal vaccine for serogroup B meningococcus., Proc. Natl. Acad. Sci. U. S. A. (2006). doi:10.1073/pnas.0603940103.

[12] I. Delany, R. Rappuoli, K.L. Seib, Vaccines, reverse vaccinology, and bacterial pathogenesis, Cold Spring Harb. Perspect. Med. (2013). doi:10.1101/cshperspect.a012476.

[13] N. Ariel, A. Zvi, H. Grosfeld, O. Gat, Y. Inbar, B. Velan, S. Cohen, A. Shafferman, Search for potential vaccine candidate open reading frames in the Bacillus anthracis virulence plasmid pXO1: In silico and in vitro screening, Infect. Immun. (2002). doi:10.1128/IAI.70.12.6817-6827.2002.

[14] T.M. Wizemann, J.H. Heinrichs, J.E. Adamou, A.L. Erwin, C. Kunsch, G.H. Choi, S.C. Barash, C.A. Rosen, H.R. Masure, E. Tuomanen, A. Gayle, Y.A. Brewah, W. Walsh, P. Barren, R. Lathigra, M. Hanson, S. Langermann, S. Johnson, S. Koenig, Use of a whole genome approach to identify vaccine molecules affording protection against Streptococcus pneumoniae infection, Infect. Immun. (2001). doi:10.1128/IAI.69.3.1593-1598.2001.

[15] J.C. Betts, Transcriptomics and proteomics: Tools for the identification of novel drug targets and

bioRxiv preprint doi: https://doi.org/10.1101/521518; this version posted January 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

vaccine candidates for tuberculosis, in: IUBMB Life, 2002. doi:10.1080/15216540212651.

[16] P.A. Manque, F. Tenjo, U. Woehlbier, A.M. Lara, M.G. Serrano, P. Xu, J.M. Alves, R.B. Smeltz, D.H. Conrad, G.A. Buck, Identification and immunological characterization of three potential vaccinogens against Cryptosporidium species, Clin. Vaccine Immunol. (2011). doi:10.1128/CVI.05197-11.

[17] A. Naz, F.M. Awan, A. Obaid, S.A. Muhammad, R.Z. Paracha, J. Ahmad, A. Ali, Identification of putative vaccine candidates against Helicobacter pylori exploiting exogenome and secretome: A reverse vaccinology based approach, Infect. Genet. Evol. (2015). doi:10.1016/j.meegid.2015.03.027.

[18] A.K. Jaiswal, S. Tiwari, S.B. Jamal, D. Barh, V. Azevedo, S.C. Soares, An in silico identification of common putative vaccine candidates against treponema pallidum: A reverse vaccinology and subtractive genomics based approach, Int. J. Mol. Sci. (2017). doi:10.3390/ijms18020402.

[19] R. Zhang, DEG: a database of essential genes, Nucleic Acids Res. (2004). doi:10.1093/nar/gkh024.

[20] S.F. Altschul, BLAST Algorithm, in: Encycl. Life Sci., 2005. doi:10.1038/npg.els.0005253.

[21] A. Gupta, R. Kapil, D.B. Dhakan, V.K. Sharma, MP3: A software tool for the prediction of pathogenic proteins in genomic and metagenomic data, PLoS One. (2014). doi:10.1371/journal.pone.0093907.

[22] E.R. Green, J. Mecsas, Bacterial Secretion Systems – An overview, Am. Soc. Microbiol. (2015). doi:10.1128/microbiolspec.VMBF-0012-2015.Bacterial.

[23] T.N. Petersen, S. Brunak, G. Von Heijne, H. Nielsen, SignalP 4.0: Discriminating signal peptides from transmembrane regions, Nat. Methods. (2011). doi:10.1038/nmeth.1701.

[24] J.D. Bendtsen, L. Kiemer, A. Fausbøll, S. Brunak, Non-classical protein secretion in bacteria, BMC Microbiol. (2005). doi:10.1186/1471-2180-5-58.

[25] C.-S. Yu, Y.-C. Chen, C.-H. Lu, J.-K. Hwang, Prediction of protein subcellular localization, Proteins Struct. Funct. Bioinforma. (2006). doi:10.1002/prot.21018.

[26] P. Garg, V. Sharma, P. Chaudhari, N. Roy, SubCellProt: Predicting protein subcellular localization using machine learning approaches, In Silico Biol. (2009). doi:10.3233/ISB-2009-0384.

[27] H. Singh, H.R. Ansari, G.P.S. Raghava, Improved Method for Linear B-Cell Epitope Prediction Using Antigen's Primary Sequence, PLoS One. (2013). doi:10.1371/journal.pone.0062216.

[28] H.R. Ansari, G.P. Raghava, Identification of conformational B-cell Epitopes in an antigen from its primary sequence, Immunome Res. (2010). doi:10.1186/1745-7580-6-6.

[29] S. Lata, M. Bhasin, G.P.S. Raghava, Application of machine learning techniques in predicting MHC binders., Methods Mol. Biol. (2007). doi:10.1007/978-1-60327-118-9_14.

[30] M. Bhasin, G.P.S. Raghava, Prediction of CTL epitopes using QM, SVM and ANN techniques, Vaccine. (2004). doi:10.1016/j.vaccine.2004.02.005.

[31] Swiss Institute of Bioinformatics, ExPASy - Compute pI/Mw tool, ExPASy Bioinforma. Resour. Portal. (2015).

[32] C. von Mering, M. Huynen, D. Jaeggi, S. Schmidt, P. Bork, B. Snel, STRING: A database of predicted functional associations between proteins, Nucleic Acids Res. (2003). doi:10.1093/nar/gkg034.

[33] M.E. Smoot, K. Ono, J. Ruscheinski, P.L. Wang, T. Ideker, Cytoscape 2.8: New features for data integration and network visualization, Bioinformatics. (2011). doi:10.1093/bioinformatics/btq675.

[34] K. Raman, N. Chandra, Mycobacterium tuberculosis interactome analysis unravels potential pathways to drug resistance, BMC Microbiol. (2008). doi:10.1186/1471-2180-8-234.

bioRxiv preprint doi: https://doi.org/10.1101/521518; this version posted January 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[35] A. Jadhav, B. Shanmugham, A. Rajendiran, A. Pan, Unraveling novel broad-spectrum antibacterial targets in food and waterborne pathogens using comparative genomics and protein interaction network analysis, Infect. Genet. Evol. (2014). doi:10.1016/j.meegid.2014.08.007.

[36] M. Juhas, L. Eberl, J.I. Glass, Essence of life: Essential genes of minimal genomes, Trends Cell Biol. (2011). doi:10.1016/j.tcb.2011.07.005.

[37] M. Juhas, L. Eberl, G.M. Church, Essential genes as antimicrobial targets and cornerstones of synthetic biology, Trends Biotechnol. (2012). doi:10.1016/j.tibtech.2012.08.002.

[38] R.J. Haselbeck, D. Wall, B. Jiang, T. Ketela, J. Zyskind, H. Bussey, J.G. Foulkes, T. Roemer, Comprehensive essential gene identification as a platform for novel anti-infective drug discovery., Curr. Pharm. Des. (2002). doi:10.2174/1381612023394818.

[39] T. Roemer, B. Jiang, J. Davison, T. Ketela, K. Veillette, A. Breton, F. Tandia, A. Linteau, S. Sillaots, C. Marta, N. Martel, S. Veronneau, S. Lemieux, S. Kauffman, J. Becker, R. Storms, C. Boone, H. Bussey, Large-scale essential gene identification in Candida albicans and applications to antifungal drug discovery, Mol. Microbiol. (2003). doi:10.1046/j.1365-2958.2003.03697.x.

[40] J.A. Andersson, J. Sha, T.E. Erova, E.C. Fitts, D. Ponnusamy, E. V. Kozlova, M.L. Kirtley, A.K. Chopra, Identification of New Virulence Factors and Vaccine Candidates for Yersinia pestis, Front. Cell. Infect. Microbiol. (2017). doi:10.3389/fcimb.2017.00448.

[41] S. Mukherjee, A. Panda, T.C. Ghosh, Elucidating evolutionary features and functional implications of orphan genes in Leishmania major, Infect. Genet. Evol. 32 (2015) 330–337. doi: 10.1016/j.meegid.2015.03.031.

[42] D. Barh, N. Barve, K. Gupta, S. Chandra, N. Jain, S. Tiwari, N. Leon-Sicairos, A. Canizalez-Roman, A. Rodrigues dos Santos, S.S. Hassan, S. Almeida, R. Thiago Jucá Ramos, V. Augusto Carvalho de Abreu, A. Ribeiro Carneiro, S. de Castro Soares, T. Luiz de Paula Castro, A. Miyoshi, A. Silva, A. Kumar, A.N. Misra, K. Blum, E.R. Braverman, V. Azevedo, Exogenome and Secretome Derived Broad Spectrum Novel Drug and Vaccine Candidates in Vibrio cholerae Targeted by Piper betel Derived Compounds, PLoS One. (2013). doi:10.1371/journal.pone.0052773.

[43] N. Ruiz, D. Kahne, T.J. Silhavy, Transport of lipopolysaccharide across the cell envelope: The long road of discovery, Nat. Rev. Microbiol. (2009). doi:10.1038/nrmicro2184.

[44] Y. Gu, P.J. Stansfeld, Y. Zeng, H. Dong, W. Wang, C. Dong, Lipopolysaccharide is inserted into the outer membrane through an intramembrane hole, a lumen gate, and the lateral opening of lptd, Structure. (2015). doi:10.1016/j.str.2015.01.001.

[45] C. Schwechheimer, M.J. Kuehn, Outer-membrane vesicles from Gram-negative bacteria: Biogenesis and functions, Nat. Rev. Microbiol. (2015). doi:10.1038/nrmicro3525.

[46] S. Okuda, D.J. Sherman, T.J. Silhavy, N. Ruiz, D. Kahne, Lipopolysaccharide transport and assembly at the outer membrane: The PEZ model, Nat. Rev. Microbiol. (2016). doi:10.1038/nrmicro.2016.25.

[47] H. Dong, Q. Xiang, Y. Gu, Z. Wang, N.G. Paterson, P.J. Stansfeld, C. He, Y. Zhang, W. Wang, C. Dong, Structural basis for outer membrane lipopolysaccharide insertion, Nature. (2014). doi:10.1038/nature13464.

[48] Z. Zha, C. Li, W. Li, Z. Ye, J. Pan, LptD is a promising vaccine antigen and potential immunotherapeutic target for protection against Vibrio species infection, Sci. Rep. (2016). doi:10.1038/srep38577.

bioRxiv preprint doi: https://doi.org/10.1101/521518; this version posted January 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1: Details of potential antigenic epitopes that can recognize B cells and T cells

Name of the Protein	B-cell epitopes predicted by Lbtope	% Probabilty of correct prediction	VaxiJen score	VirulentPred score	CTL epitopes predicted by CTLpred consensus approach	MHC allele (predicted by nHLApred)	Position
LptD	NFDATVSTKQFQVFN	81.98	0.9712	1.0581	NFDATVSTK	HLA-Cw*0401, H2-Db, H2-Dd, H2-Kb, H2-Kd, H2-Ld, HLA-G, H-2Qa, Mamu-A*01	366-374
LptE	NNVNLLDKDTTRKDV	81.66	0.9274		NLLDKDTTR	HLA-Cw*0401, H2-Db, H2-Dd, H2-Kb, H2-Kd, H2-Ld, HLA-G, H-2Qa, Mamu-A*01	60-66

bioRxiv preprint doi: https://doi.org/10.1101/521518; this version posted January 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure Legends

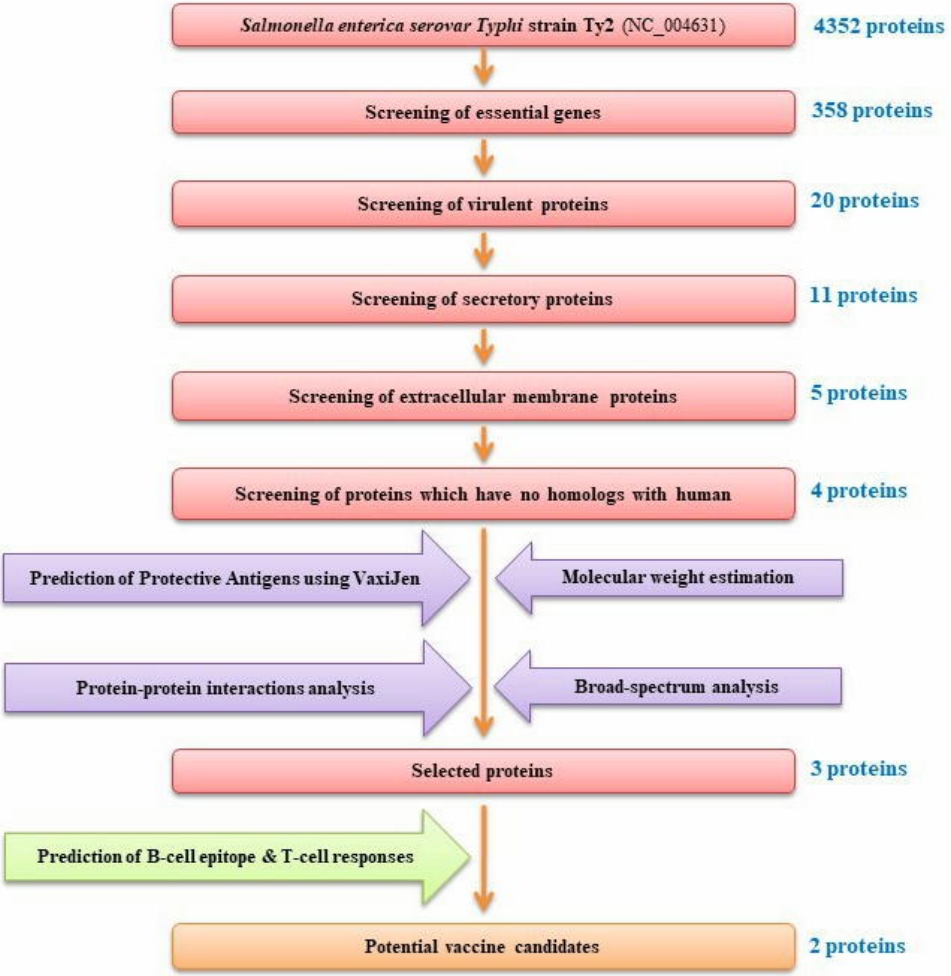
Figure 1: Computational pipelines based on reverse vaccinology and subtractive genomics based approach for screening the potential vaccine targets from *Salmonella typhi*.

Figure 2: Structural mapping of the epitopes on the PDB structure of LptD-LptE complex.

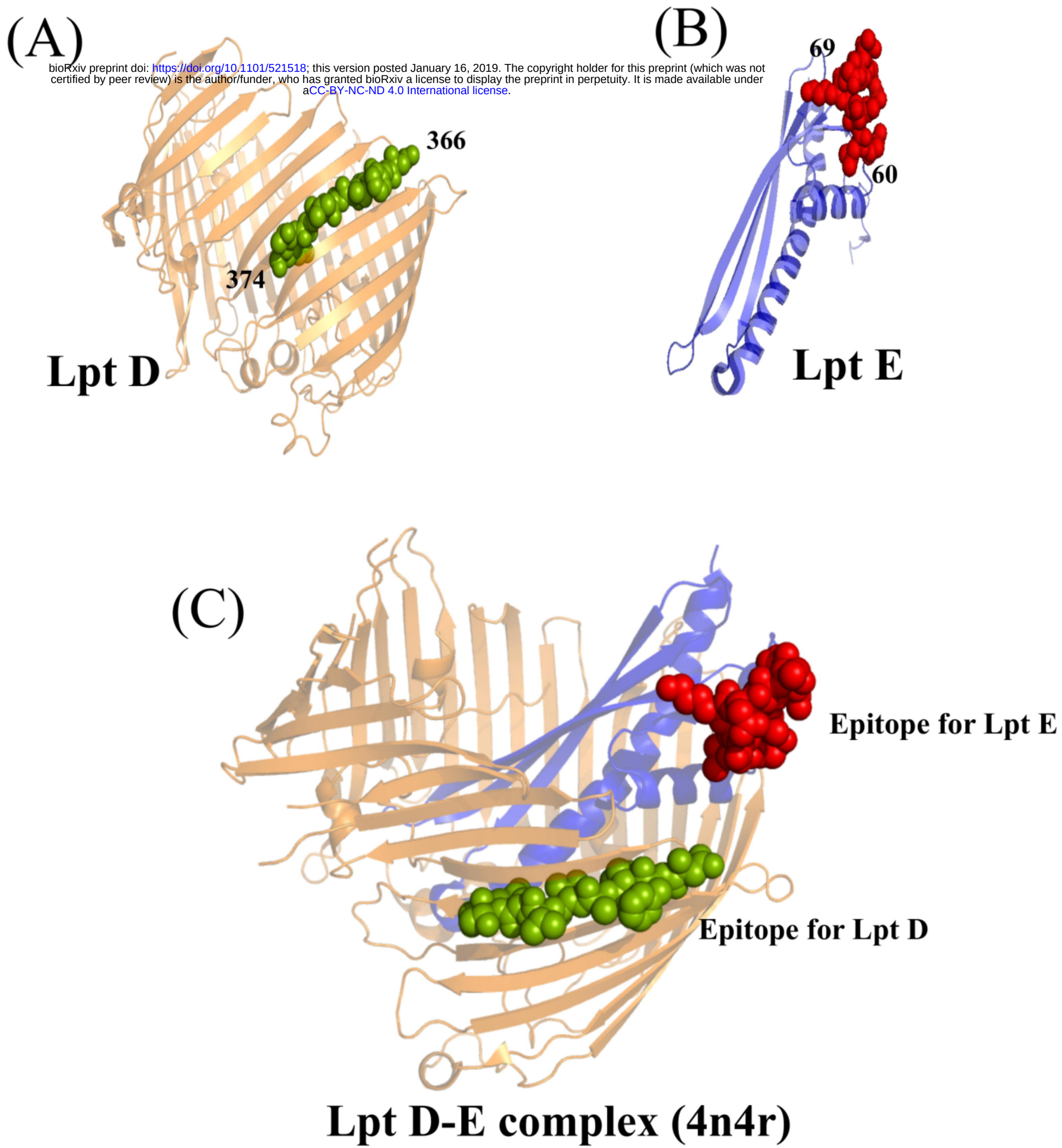
(A) Structure of LptD. The green spheres denote the epitope part of LptD. The residues mapped are from 366 to 374. (B) Structure of LptE. The red spheres denote the epitope part of Lpt E. The residues mapped are from 69 to 74. (C) Structure of LptDE complex (PDB id: 4n4r). The epitopes are mapped for LptD and LptE on the the structure with green and red respectively.

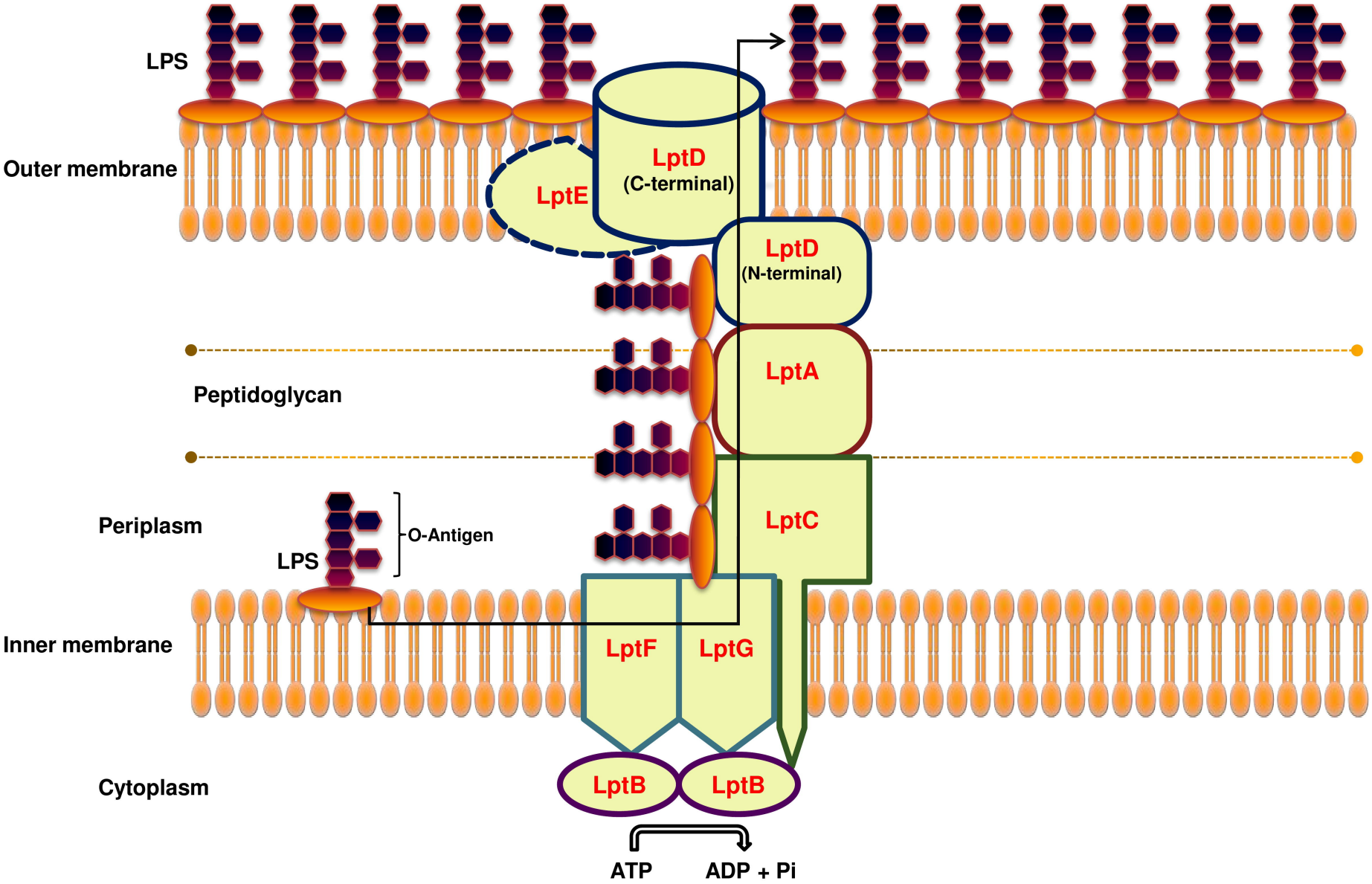
Figure 3: LPS transport system.

LptDE complex is essential for the final transport of LPS across the membrane.


Salmonella enterica serovar Typhi strain Ty2 (NC_004631)
4352 proteins
Screening of essential genes
358 proteins
Screening of virulent proteins
20 proteins
Screening of secretory proteins
11 proteins
Screening of extracellular membrane proteins
5 proteins
Screening of proteins which have no homologs with human
4 proteins
Prediction of Protective Antigens using VaxiJen
Molecular weight estimation
Protein-protein interactions analysis
Broad-spectrum analysis
Selected proteins
3 proteins
Prediction of B-cell epitope & T-cell responses
Potential vaccine candidates
2 proteins

bioRxiv preprint doi: https://doi.org/10.1101/521518; this version posted January 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.




LPS
Outer membrane
LptE
LptD
(C-terminal)
LptD
(N-terminal)
LptA
Peptidoglycan
Periplasm
LPS
O-Antigen
LptC
Inner membrane
LptF
LptG
Cytoplasm
LptB
LptB
ATP
ADP + Pi