# PUTATIVE DRUG TARGET IDENTIFICATION FOR *CHLAMYDIA TRACHOMATIS*: AN INSILICO PROTEOME ANALYSIS

**A.Praveena***, R.Sindhuja, V.Anuradha, S.K.M. Habeeb

Department of Bioinformatics, Guru Nanak College, Velachery, Chennai- 600 042, India

**Corresponding author*: praveena_bioinfo@yahoo.com**

This article is available online at **www.ssjournals.com**

## ABSTRACT

The whole genome sequences of pathogenic bacteria and the host genome such as human has provided a Subtractive genomic approach, which can be used to identify potent vaccine and drug targets. In the present study subtractive genomic approach has been used to identify therapeutic target in *Chlamydia trachomatis*. *C.trachomatis* infection is now the most common sexually transmitted disease worldwide. The BlastP search against *Homo sapiens* revealed 551 non-homologous protein sequences out of 874 in *C.trachomati*. Further analysis of these non human homologous proteins predicted that 142 essential proteins were involved in unique metabolic pathways of *C.trachomatis*. The prediction of sub-cellular localization of the essential proteins was used to identify the membrane proteins which can be used as vaccine targets. There are 63 unique essential non-human homologous therapeutic targets found in the current study, which plays a vital role in the Peptidoglycan biosynthesis, Phosphotransferase system, Fatty acid biosynthesis and Bacterial secretion system of *C.trachomatis*.

**KEY WORDS:** Subtractive genomics, Therapeutic drug target, Unique pathways, Homologs.

## INTRODUCTION

The completion of human genome project and pathogenic bacteria have increased the chances of identifying potent drug targets against life threatening human pathogens. A number of bioinformatics tools and public databases have been developed to facilitate *in silico* analysis of the gene sequence information [1]. Drug resistance among important microorganisms is a major challenge in modern medicine. *C.trachomatis*, an obligate intracellular human gram-negative pathogen, is one of three bacterial species in the genus *Chlamydia* [2]. *C. trachomatis* was the first chlamydial agent discovered in humans in the year of 1907 [3]. There are numerous factors that contribute to the pathogenicity of *C.trachomatis*. Colonization of Chlamydia begins with attachment to sialic acid receptors on the eye, throat, or genitalia. In humans, infection in many individuals is asymptomatic and so treating only those with clinical symptoms is not the best way of controlling the spread of infection [4].

Empirical attempts were made in the 1960s and early 1970s to prevent

trachoma caused by Chlamydial infections using vaccination. Renewed attempts to protect against *C. trachomatis* infection were made in the mid 1980s following the discovery that the major outer membrane protein (MOMP) of *C. trachomatis* [5, 6]. As yet there is no consensus as to what constitutes a protective immune response against genital Chlamydia infection.

One of the important strategies to identify the novel drug target is finding the bacterial genes that are non-homologous to human genes and important for the survival of bacteria. A subtractive genomics approach and bioinformatics provide opportunities for finding the drug targets against pathogens [7]. In the present study, the genome of *C.trachomatis* is compared with the host genome *Homo sapiens* to identify the potential unique targets for the development of effective drugs and vaccines.

## MATERIAL AND METHODS

### Sequence Collection

The Genome sequence of *C.trachomatis* (Accession number: NC_010287.1) was collected from the NCBI (National Center for Biotechnology Information) database. The total 874 protein sequences of *C.trachomatis* were downloaded in FASTA format from NCBI. The protein sequences with less than 100 amino acids were screened out for further study because coding sequences having less than 100 amino acids were less likely to represent essential genes from protein table.

### Identification of Paralogs

The paralogous sequences were identified at similarity threshold of 60% using the CD-HIT suit. The prologs were excluded and the remaining set of proteins was used for further analysis.

### Similarity Search

The non-paralogous proteins were subjected to BlastP [8] search against *Homo sapiens* using threshold expectation value of $10^{-3}$as parameter to find out the non-human homologous proteins of *C.trachomatis*. The human homologous were excluded and the list of non-homologs was compiled.

### Essential Protein Search

The selected non-human homologous proteins were then subjected to similarity search using BlastP in Database of Essential Genes (DEG) (http://tubic.tju.edu.cu/deg1). A random expectation value (E-value) cut-off of 0.001 and sequence identity of 30% and above was used to screen out proteins that appeared to represent essential proteins.

### Functional classification of hypothetical proteins

Functional family prediction of the putative uncharacterized essential proteins was done by using the SVMProt web server (http://jing.cz3.nus.edu.sg/cgi-bin/svmprot.cgi.) [9].

### Subcellular Localisation

Prediction of protein localization is important to identify the surface membrane proteins which could be feasible vaccine target. Sub-cellular localization analysis of the essential protein sequences has been done by PsortBTb v3.0 server (http://www.psort.org/psortb/ PSORTb).

### Therapeutic target identification

The essential proteins (enzymes) of *C.trachomatis* found in the DEG database were searched against the therapeutic

target database to identify the available related drug targets using TTD (Therapeutic Target Database, http://xin.cz3.nus.edu.sg/group/ttd/ttd.asp) . The essential proteins (enzymes) matched with the TTD were searched in the KEGG database to identify the unique pathway of the *C.trachomatis* when compared against human metabolic pathways.

## RESULTS

In the present study, a strategy for comparative metabolic pathway analysis was used to find out some potential targets against *C.trachomatis*. Only those enzymes which show unique properties than the host were selected as the target (Table 1).

The genome of *C.trachomatis* consist of 1,038,842 nucleotides, 934 genes, 874 coding proteins and the GC content of 41%. A list of 874 proteins from the genome of the *C.trachomatis* has been extensively compared with the proteins present in the genome of *H.sapiens*. Fifty five protein sequences were excluded from the total list, since coding sequences having less than 100 amino acids were less likely to represent essential genes. The CD-HIT suit results showed that there are no duplicate proteins with the threshold identity value of 60% and above. Thus, the 819 sequences were analyzed for tracing the non-Human homologous sequences using the BlastP program. The BlastP search resulted in 551 non-human homologous sequences which were short listed based on the E-Value of 0.005. These sequences were further analysed for the identification of essential genes using DEG database server and considered cutoff score was >100, to enhance the specificity of enzyme in *C.trachomatis*. A total of 142 proteins were found to be essential for *C.trachomatis* life cycle.

The functional family of 7 hypothetical proteins predicted to be essential in DEG database was identified using the SVMprot tool. The SVMprot results indicated that the hypothetical proteins are mainly from the protein families of Zinc-binding, DNA-binding, Iron-binding, lipid-binding, Metal-binding and Transmembrane proteins with high P-value, which is the expected classification accuracy in terms of percentage (Table 2). As it was suggested that membrane associated protein could be the better target for developing vaccines. The sub-cellular localization results explored that the 98 proteins were found to be located in the cytoplasm, 1 as extra cellular, 26 as membrane bound and 16 without any positive prediction (Graph1). Out of 142 unique essential proteins, 63 matched significantly and screened as the therapeutic targets.

The selected 63 putative therapeutic targets were matched with the KEGG database to identify the unique pathways. The results explored 10 unique enzymes out of 63, mainly involved in virulence pathways of bacteria such as Peptidoglycan biosynthesis, Phosphotransferase system, Fatty acid biosynthesis and Bacterial secretion system (Table 3).

## DISCUSSION

The computational approach has been used to investigate novel drug targets in pathogenic organisms such as *Pseudomonas aeruginosa* [10, 11] and *Helicobacter pylori* [12]. Anti-bacterials are essentially inhibitors of certain bacterial enzymes, all enzymes specific to bacteria

can be considered as potential drug targets[13].

Bacterial envelope and secretion systems are of particular interest in antimicrobial compound discovery. The bacterial envelope is known to actively extrude drugs via the action of an efflux pump. A unique multi-drug efflux system (BpeAB-OprB) accords *B. pseudomallei* resistance to aminoglycosides and macrolide antibiotics [14]. Thus, active compounds that can destabilize bacterial cell membranes and disrupt lipopolysaccharides are good drug candidates. In addition, the bacterial secretion machinery is also important for survival. Despite some level of similarity between prokaryotic and eukaryotic secretion systems, differences in the bacterial secretion process are sufficient to infer that these systems might be useful as drug targets without the risk of disrupting host-cell functions[15].

All the bacterial species almost share a common feature of the cell wall which is helping them to maintain their structure as well as helping the bacterium to withstand tremendous internal pressure (up to 350 lbs/cm2) to keep the bacterial cell from exploding. The cell wall is composed of peptidoglycan, teichoic acids, and proteins [16]. Chemical analysis of the cell wall indicates that more than 70% of the weight of the cell wall is peptidoglycan and that the teichoic acid is covalently bound to the peptidoglycan through a phosphodiester bond [17]. Thus the peptidoglycan biosynthesis plays a vital role in the enlargement of cell wall. The unique enzymes found in Peptidoglycan biosynthesis of *C.trachomatis* are UDP-N-acetyl muramoyl-L-alanyl-D-glutamate synthetase, N-acetyl glucosaminyl transferase, UDP-N-acetyl muramate--alanine ligase and Penicillin binding protein.

The phosphoenolpyruvate (PEP)-dependent phosphotransferase system (PTS) is a major mechanism used by bacteria for uptake of sources of energy such as carbohydrates, particularly hexoses, hexitol, and disaccharides. The unique non human homologous enzyme found in the phosphotransferase system of *C.trachomatis* is PTS family membrane transport protein IIA subunit.

Genomic analysis has revealed many novel potential targets for antimicrobial drugs and many of them are in essential and conserved metabolic pathways or cell-cell communication systems [18, 19, 20]. Such a potential mechanism which is necessary for the virulence of microbes is the bacterial secretion system. Pathogenic bacteria need virulence factors in order to infect their hosts and to survive the immune response [21, 22]. The secretion systems used for this purpose are in many cases very important or essential for bacterial virulence, and they are grouped into five classes according to their protein composition, amino acid similarities and mechanism. The unique enzymes found in the bacterial secretion system of *C.trachomatis* are preprotein translocase subunit SecA and general secretion pathway protein E.

Fatty acids are one of the most important building blocks of cellular materials. In bacterial cells, fatty acids occur mainly in the cell membranes as the acyl constituents of phospholipids. The unique enzyme of *C.trachomatis* involved in the biosynthesis of fatty acid is (3R)-hydroxy myristoyl-ACP dehydratase.

Subtractive genomics studies between the host and pathogen genome thus provides the details about the proteins likely to be essential to the pathogen but absent in the host. By applying this approach the proteome of

*C.trachomatis* was compared with the host *H.sapiens*. The present analysis traced out the essential proteins, which play a unique role in survival of bacteria. Developing drugs against the identified unique targets will be specific to the *C.trachomatis,* and therefore less or non-toxic to the host. Virtual screening against unique targets might be useful in the discovery of novel therapeutic compounds against *C.trachomatis.*

## REFERENCES


1. Kaminski N. Bioinformatics. A user's perspective. Am. J. Respir. Cell Mol. Biol. 2000; 23**:** 705-711.
2. Ryan KJ. Nosocomial infections and infection control. In: Ryan KJ, Ray CG, editors. Sherris medical microbiology an introduction to infectious diseases. 4th ed. New York: McGraw-Hill Medical Publishing Division; 2004. p. 915-21.
3. Budai I, Chlamydia trachomatis: milestones in clinical and microbiological diagnostics in the last hundred years: a review. Acta microbiologica et immunologica Hungarica 2007; 54 (1): 5–22.
4. Stagg AJ, Elsley WA, Pickett MA, Ward ME, Knight SC**.** Primary human T-cell responses to the major outer membrane protein of Chlamydia trachomatis. Immunology 1993*;* 79: 1 - 9.
5. Caldwell HD, Perry LJ. Neutralization of *Chlamydia trachomatis* infectivity with antibodies to the major outer membrane protein. Infection and Immunity *1982;* 38: 745 - 754.
6. Zhong GM, Brunham RC. Antigenic determinants of the chlamydial major outer membrane protein resolved at a single amino acid level. Infection and Immunity 1991*;* 59**:** 1141 - 1147.
7. Reddy EH, Satpathy GR. Identification of potential targets and lead molecules for designing inhibitory drugs against Chlamydophila pneumoniae. Online journal of bioinformatics 2009; 10(1):14-28.
8. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J. Mol. Biol.1990; 215: 403-410.
9. Cai CZ, Han LY, Ji ZL, Chen X, Chen YZ. SVMProt: Web-Based Support Vector Machine Software for Functional Classification of a Protein from Its Primary Sequence. Nucleic Acids. Res 2003; 31: 3692-3697.
10. Sakharkar KR., Sakharkar MK, Chow VT.A novel genomics approach for the identification of drug targets in pathogens, with special reference to Pseudomonas aeruginosa. In Silico Biol 2004; 4: 355-360.
11. Perumal D, Lim CS, Sakharkar KR, Sakharkar MK.Differential genome analyses of metabolic enzymes in Pseudomonas aeruginosa for drug target identification. In Silico Biol 2007; 7: 453-465.
12. Dutta A, Singh SK, Ghosh P, Mukherjee R, Mitter S, et al. In silico identifi cation of potential therapeutic targets in the human pathogen Helicobacter pylori. In Silico Biol 2006; 6: 0005.

**13.** Michael YG, Eugene VK. Searching for drug targets in microbial genomes. Curr Opin Biotechnol 1999; 10: 571-578.

**14.** Chan YY, Chua KL. The Burkholderia pseudomallei BpeAB-OprB efflux pump: expression and impact on quorum sensing. J. Bacteriol 2005; 187: 4707-4719.

**15.** Desnottes JF. New targets and strategies for the development of antibacterial agents. Trends Biotechnol1996; 14: 134-140.

**16.** Schleifer KH. The cell envelope in staphylococci and staphylococcal infections. In: Easmon CSF, Adlam C ed. London: Academic Press; 1983; 2: 387-428.

**17.** Coley LE, Tarelli AR, Archibald A, Baddiley J. The linkage between teichoic acid and peptidoglycan in bacterial cell wall. FEBS Lett 1978; 88: 1-9.

**18.** Marraffini LA, Dedent AC, Schneewind O. Sortases and the Art of Anchoring Proteins to the Envelopes of Gram-Positive Bacteria. Microbiol Mol Biol Rev 2006; 70: 192-221.

**19.** Brown ED, Wright GD. New Targets and Screening Approaches in Antimicrobial Drug Discovery Chem. Rev 2005; 105: 759-774.

**20.** Bassler BL, Losick R. Bacterially Speaking. Cell 2006; 125(2), 237-246.

**21.** Coombes BK, Valdez Y, Finlay BB.Evasive Maneuvers by Secreted Bacterial Proteins to Avoid Innate Immune Responses. Curr Biol 2004; 14: R856.

**22.** Finlay BB, McFadden G Anti-Immunology: Evasion of the Host Immune System by Bacterial and Viral Pathogens. Cell 2006; 124: 767.

**TABLE 1: Classification of proteins in *C.trachomatis.***

| | |
|---|---|
| Number of total proteins in the Genome | 874 |
| Protein sequences with >100 amino acids length | 55 |
| Number of paralogs obtained from CD-HIT suit | 0 |
| Non-Paralog sequences | 819 |
| Non-Human homologous sequences | 551 |
| Essential Proteins matched with DEG | 142 |
| Number of hypothetical proteins in essential proteins | 7 |
| Number of proteins present in membrane | 26 |
| Number of proteins matched with therapeutic targets | 63 |

**TABLE 2: Functional family prediction of hypothetical proteins**

| SI.No | Accession number of the query | Predicted protein family name | P – Value (%) |
|---|---|---|---|
| 1 | gi\|166154028\| | Zinc-binding | 99.2 |
| 2 | gi\|166154105\| | All DNA-binding | 62.2 |
| 3 | gi\|166154109\| | Iron-binding | 99.1 |
| 4 | gi\|166154269\| | All lipid-binding proteins | 97.5 |
| 5 | gi\|166154421\| | All lipid-binding proteins | 92.9 |
| 6 | gi\|166154639\| | Metal-binding | 82.2 |
| 7 | gi\|166154687\| | Transmembrane | 73.8 |

**TABLE 3: Unique pathways of the selected targets of *C.trachomatis.***

| S.No | Accession # | Description of the query | Unique pathway involved |
|---|---|---|---|
| 1 | gi\|166154786\| | General secretion pathway protein E | Bacterial secretion system |
| 2 | gi\|166154043\| | Preprotein translocase subunit SecA | Bacterial secretion system |

| | | | |
|---|---|---|---|
| 3 | gi\|166154024\| | Penicillin-binding protein | Peptidoglycan biosynthesis |
| 4 | gi\|166154100\| | UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase | Peptidoglycan biosynthesis |
| 5 | gi\|166154103\| | N-acetylglucosaminyl transferase | Peptidoglycan biosynthesis |
| 6 | gi\|166154104\| | UDP-N-acetylmuramate-alanine ligase | Peptidoglycan biosynthesis |
| 7 | gi\|166154481\| | Penicillin-binding protein | Peptidoglycan biosynthesis |
| 8 | gi\|166154450\| | 3-oxoacyl-(acyl carrier protein) synthase III | Fatty acid biosynthesis |
| 9 | gi\|166154747\| | (3R)-hydroxymyristoyl-ACP dehydratase | Fatty acid biosynthesis |
| 10 | gi\|166154501\| | PTS family membrane transport protein IIA subunit | Phosphotransferase system (PTS) |

**GRAPH 1: SUBCELLULAR LOCALIZATION OF THE UNIQUE ESSENTIAL NON-HUMAN HOMOLOGOUS PROTEINS OF *C.TRACHOMATIS***