Summary and Conclusion
Summary

*Chlamy dophila pneumoniae*, is one of the most important and studied gram negative bacterial strains with respect to community acquired pneumonia and other respiratory diseases like COPD, chronic asthma, Alzheimer's disease and atherosclerosis, which have a great potential to infect humans and many other mammals. It is predicted by WHO to become the third leading cause of death by 2030 and the 5th leading cause of loss of 'Disability Adjusted Life Year's (DALYs) as per the global Burden of disease study (GBDS). The present options for the early diagnosis and effective treatment are limited, so there is a need for better therapeutics. Conventional drug discovery methods are found time consuming and labor-intensive. Unfortunately, the molecular mechanisms leading to chronic infections are poorly understood and the difficulties in culturing *C pneumoniae* in experimental conditions and lack of entirely satisfactory serological methods for diagnosis, have been a hurdle for drug discovery and development. In contrast, availability of complete genome sequences, in combination with cheminformatics, metabolomics and computational biology, has paved a alternative way for the identification of novel therapeutic candidates worthy of experimental research.

A computational, Synteny based comparative genomics workflow was designed for the identification of novel therapeutic candidates against *C pneumoniae*, with the aim that a selected target should possess essential to the pathogen, and must not have homology in the human host. Initially, prediction of orthologous and non-orthologous syntenic blocks from eight bacterial pathogens which belongs to Chlamydiaceae family was carried out to identify the specific genes/proteins involved in pathogenesis to cause Chronic obstructive pulmonory disease (COPD) in *Homosapiens*, where other Chlamydial organisms lack the potential to cause COPD though some are involved in human pathogenesis.
In the current study we have been able to classify 1128 Synteny blocks present in the C. pneumoniae genome into orthologous and non-orthologous Synteny blocks, and we have specifically used the strategy of considering the different data sets to identify the putative drug targets from C. pneumoniae genome throughout the thesis. Firstly we have considered orthologous synteny blocks in the genome, the predicted 460 synteny blocks which are highly conserved in all the Chlamydial organisms compared were considered and these were exploited in step by step analysis, 414 proteins were identified as non-homologous (non-host) to the human proteome, 58 proteins were predicted as essential genes, In subcellular localization analysis step the short listed essential proteins were classified into 19 cytoplasmic proteins, 2 outer membrane proteins and 1 extracellular protein, further classification was based on functional class into transmembranes, lipoproteins, signal peptides and GPCR yielded 4 transmembranes, 4 signal peptides, 3 Lipoproteins and 0 GPCR’s, among these 9 proteins were predicted as exotoxins and 3 as endotoxins. The short listed drug targets were prioritized based on molecular weight and availability of 3D structure. DNA-directed RNA polymerase subunit beta was identified as putative potential drug target based on orthologous syntons.

Secondly a total of 354 proteins were predicted as non-orthologous to other Chlamydial organisms, except hypothetical proteins 70 were found functional out of which 60 are non homologous(non-host) to human proteome and among them 18 proteins were found to be essential for the survival of C. pneumoniae. Further subcellular localization analysis resulted in 52.5% enzymes which are found to be cytoplasmic, Among which 5 were identified and classified as bacterial exotoxins and 2 as bacterial endotoxins remaining 11 proteins were found as proteins involved in crucial functionalities like DNA binding, RNA binding, catalytic activity, ATP binding, oxidoreductase activity, hydrolase activity and proteolysis activity.
Thirdly, a computational comparative metabolic pathway analysis of the host *H.sapiens* and the pathogen *C. pneumoniae* AR39 has been carried out at three level analyses. In the first level, metabolic pathway analysis was performed to identify unique metabolic pathways and non-homologous proteins were identified. In the second level, essentiality of the proteins was analyzed, where these proteins contribute to the growth and survival of the organism. Finally these proteins were further subjected to predict protein interaction networks. Among the total 65 pathways in the *C. pneumoniae* AR39 genome 10 were identified as the unique metabolic pathways where the genes members of these pathways are associated with Lipopolysaccharide biosynthesis, Peptidoglycan biosynthesis, Polycyclic aromatic hydrocarbon degradation, Biosynthesis of secondary metabolites, Microbial metabolism in diverse environments, Two-component system, Flagellar assembly, Phosphotransferase system (PTS), Bacterial secretion system and Methane metabolism which were not found in the human host, later manual sorting of pathways was performed to get an more insight into pathways, then only proteins from pathogen-specific pathways were subjected to BLASTP analysis, 32 enzymes were predicted as essential and these proteins were considered for protein interaction analysis, finally by considering various criteria's like molecular weight, toxicity and localization, absence and presence of transmembrane, availability of 3D structure and druggability we have prioritized ribonucleotide-diphosphate reductase subunit beta (RNR) as a potential drug target which facilitate for the successful entry into drug designing pipeline.

Structure based drug designing strategy has been successfully adopted in the present thesis to model the protein for which structure was unavailable in any of the 3D structure databases as only four X-ray crystallographically solved structures are available for *C. pneumoniae* organism, the predicted potential drug target ribonucleotide-diphosphate reductase subunit beta structure was modelled using Modeller which had 87% identity with 3D structure 1SYY, 2ANI,
4D8G and 4D8F in RCSB PDB database, these were considered as templates to model the structure after performing multiple sequence alignment. The modelled structure was used for docking analysis by searching the specific drug in Drugbank which can inhibit the binding of the potential drug target ribonucleotide-diphosphate reductase subunit beta with ribonucleotide-diphosphate reductase subunit alpha; this (RNRs) complex leads to catalyse the conversion of ribonucleotides to deoxyribonucleotides, to inhibit this catalysis reaction, Cladribine was docked in pie stack mode at the radical binding site specifically.

Conclusions

In the current study, a critical annotation of C.pneumonia genome has been performed based on genome wide Synteny comparison between the family members of Chlamydiaceae. The genome was initially classified into orthologous and non-orthologous based on the syntons. In the present step the data set of predicted was considered to predict the drug targets strategically. Non homologous proteins were identified using BLAST analysis as the proteins present in pathogen but lack in human proteome can be used as toxins by pathogen leading to the virulence in humans. Using DEG computationally essential genes/proteins were identified as the essential genes are vital for chlamydial survival and pathogenicity. The essentiality predictions from Cpn agree well with previously reported high-throughput studies on gene essentiality. This study provides a framework for the rational identification of drug targets in C.pneumonia through Synteny based patho-genome comparison.

Further, the proteins were classified based on their functional class into transmembranes, lipoproteins, signal peptides and GPCRs which are considered as popular classes of drug targets. These proteins were prioritized based on subcellular localization, and druggability, this in silico analysis helped in
identifying and classifying proteins into cytoplasmic and periplasmic proteins based on their localization. Toxicity prediction step was crucial in identifying and classifying short listed proteins into bacterial exotoxins and endotoxins which is a directional step to the goal of identifying drug targets. Database search for the available FDA approved drugs stored in DrugBank was vital to identify the approved drugs with confirmed biological activity.

The idea of predicting possible drug targets computationally using non-orthologous gene data set was the strategy considered to predict the earlier discussed important steps were repeated to predict Non homologous proteins and essential proteins which are vital for the survival of the C. pneumonia pathogen. Later, the listed proteins resulted from earlier step were analyzed for the subcellular localization and toxicity and the need for predictions discussed in earlier steps. The prioritization of drug targets was performed using a different strategy by analyzing the key features of the proteins listed which includes prediction of domains, signatures and specific functions and the pathways in which they are involved. Finally, as an essential criteria to be met by a protein to predict drug target computationally, the availability of 3D protein structure and druggability was analyzed. The listed drug targets from this study were; 1 RNA binding protein, 6 DNA binding protein, 4 involved in catalytic activity, 3 ATP binding proteins, 1 chaperonin protein, 1 with oxydoreductase activity, 1 with hydrolase activity and 1 with proteolysis activity. In this study we recommend 4 proteins as important drug targets namely signal recognition particle protein FtsY and cysteinyll-tRNA synthetase, which are endotoxins involved in pathogenesis and serine protease and Clp protease are proteins which are ubiquitous and serves as virulence factor. The current study provides an outline for the in silico identification of anti-chlamydial drug targets through the modern strategy Synteny based comparative pathogenomics.
A critical metabolic pathway analysis has been performed by listing the *C.pneumonia* and Homosapiens pathways in KEGG database. The strategy was to identify the unique pathways present in the pathogen, as discussed earlier the pathways present in the pathogen and present in the host is ranked as unique pathways identified by manual sorting method. Later each unique pathway identified was looked for the proteins involved in the metabolic networks and these were compared against the human proteome for similar proteome. The proteins which lack the similarity percentage of ≥35% was named as non-homologous and these were further analyzed for essentiality. The prediction of signature proteins that are unique to the *C.pneumonia*, using this strategy, has helped in delineating those proteins, which are hallmarks of the unique pathways identified.

The unique pathways identified have been listed. The specific proteins/enzymes identified were 170, non host proteins were 90 and 32 proteins were essential to the organism for survival. 21 proteins were predicted as bacterial exotoxins and 1as bacterial endotoxin. These proteins were further subjected to analyze protein-protein interactions. All the proteins discussed are being reported for the first time. We recommend single potential therapeutic target, ribonucleotide-diphosphate reductase subunit beta as it meets all the criteria's considered in this study.

Finally, in this step structure based drug designing has been effectively used for modelling and docking studies, the structure for the identified potential therapeutic target ribonucleotide-diphosphate reductase subunit beta in the earlier step, due to the unavailability of the 3D structure in structure databases it was modeled using 4 templates 4D8F, 4D8G, 1SYY and 2ANI with 87% identity. This particular therapeutic target is considered as favorite drug target for the researchers in other organisms and it has been considered as favorite cancer target in humans. The identified, FDA approved drug Cladribine was docked
with the radical binding site of the modeled structure of ribonucleotide-diphosphate reductase subunit beta, which was well fitted in the binding pocket forming four hydrogen bonds between ligand and the protein with an inhibition constant of 90.91 µM and -5.51 binding energy. Modeling has been useful to elucidate the theoretical 3D structure of the nrdB protein in C. pneumonia. Structure based drug design approach has been of great importance to develop fast and accurate target identification. Docking study has been useful in studying the molecular interaction dynamics of protein and ligand interaction.